IEEE
Signal Processing
MAGAZINE
Volume 37 | Number 4 | July 2020

AUTONOMOUS DRIVING

Part 1: Sensing and Perception

MIMO Radar for Advanced Driver-Assistance Systems and Autonomous Driving

Machine Learning Requires Probability and Statistics

Neural Networks, Hypersurfaces, and Radon Transforms

Observer-Based Adaptive Fourier Analysis

IEEE
Signal
Processing
Society

IEEE

Call for Papers

IEEE Journal of Selected Topics in Signal Processing

Special Issue on Recent Advances in Automotive Radar Signal Processing

We witness today an enormous amount of activities in the automotive industry, in particular the development of Advanced Driver Assistance Systems, with the goal to make driving safer and more comfortable. Moreover, the introduction of the Highly Automated Driving is considered a topical technology challenge. The performance and reliability of these systems strongly depends on the capabilities of the environmental sensing, for which radar technology is considered as indispensable.

Advantages of radar technologies when compared with LiDAR and camera technologies clearly are the robust and preferred operation in adverse weather conditions, higher range, direct measurement of relative velocity, and affordability. An active field of research and development is to mitigate limitations of automotive radar, which are mainly angular resolution and object classification capabilities. For instance, high-performance radar sensors can be developed using MIMO radar techniques and a plurality of highly integrated radio frequency components. Another trend is to exploit specific data features, e.g. micro-Doppler signatures, to improve the detection and classification of vulnerable road users. In particular, machine learning and deep learning applications are increasingly used.

This special issue targets novel technical contributions. Topics of interest include, but are not limited to:

- MIMO radar
- 4D imaging radar
- Model-based high-resolution parameter estimation
- Online methods for calibration, performance monitoring and fault detection
- Interference detection and mitigation
- Enhanced object and scene classification
- Machine learning and deep learning applications (point cloud segmentation, occupancy grid calculation, 3D object detection)
- Hardware acceleration and integration
- Radar and communications
- Radar networks
- Data fusion and target tracking
- Prototyping, measurements and experimentation

Guest editors:

- Prof. Abdelhak Zoubir (Professor at Technische Universität Darmstadt)
- Dr. Philipp Heidenreich (Advanced Technology Specialist, Opel Automobile GmbH)
- Dr. Igal Bilik (Manager of Smart Sensing and Vision Systems Group, RnD Center Israel, General Motors)
- Prof. Maria Greco (Professor at University of Pisa)
- Prof. Murat Torlak (Professor at University of Texas at Dallas)

Milestone dates:

Manuscript submission due: 8-Aug-2020
First review completed: 02-Nov-2020
Revised manuscript due: 04-Jan-2021
Second review completed: 15-Feb-2021
Final manuscript due: 15-Mar-2021

Digital Object Identifier 10.1109/MSP.2020.2999221

Contents

Volume 37 | Number 4 | July 2020

SPECIAL SECTION

AUTONOMOUS DRIVING

PG. 7

ON THE COVER

This special issue on autonomous driving will be presented in two parts. Part 1: Sensing and Perception aims to provide researchers and professionals with tutorial-style articles that cover the current state of the art as well as emerging trends in the design, development, and deployment of sensing and perception technologies for autonomous and automated driving. Part 2: Learning and Cognition is scheduled for the January 2021 issue.

COVER IMAGE: ©ISTOCKPHOTO.COM/JUST_SUPER

FEATURE

COLUMNS

PG. 148

IEEE SIGNAL PROCESSING MAGAZINE (ISSN 1053-5888) (ISPREG) is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA (+1 212 419 7900). Responsibility for the contents rests upon the authors and not the IEEE, the Society, or its members. Annual member subscriptions included in Society fee. Nonmember subscriptions available upon request. **Individual copies:** IEEE Members US$20.00 (first copy only), nonmembers US$246 per copy. Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright Law for private use of patrons: 1) those post-1977 articles that carry a code at the bottom of the first page, provided the per-copy fee is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA; 2) pre-1978 articles without fee. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. **For all other copying, reprint, or republication permission,** write to IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 USA. Copyright © 2020 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved. Periodicals postage paid at New York, NY, and at additional mailing offices. **Postmaster:** Send address changes to IEEE Signal Processing Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188 **Printed in the U.S.A.**

Digital Object Identifier 10.1109/MSP.2020.2984834

DEPARTMENTS

©ISTOCKPHOTO.COM/GIVAGA

The 2020 IEEE Radar Conference will take place in Florence, Italy, 21–25 September 2020.

SUSTAINABLE FORESTRY INITIATIVE
Certified Chain of Custody
Promoting Sustainable Forestry
www.sfiprogram.org
SFI-01681

IEEE prohibits discrimination, harassment, and bullying.
For more information, visit
http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

IEEE Signal Processing Magazine

EDITOR-IN-CHIEF
Robert W. Heath Jr.—The University of Texas at Austin, U.S.A.

AREA EDITORS

Feature Articles
Matthew McKay—Hong Kong University of Science and Technology, Hong Kong SAR of China

Special Issues
Namrata Vaswani—Iowa State University, U.S.A.

Columns and Forum
Rodrigo Capobianco Guido—São Paulo State University (UNESP), Brazil
Roberto Togneri—The University of Western Australia

e-Newsletter
Ervin Sejdic—University of Pittsburgh, U.S.A.

Social Media and Outreach
Tiago Henrique Falk—INRS, Canada

Special Initiatives
Andres Kwasinski—Rochester Institute of Technology, U.S.A.
Nuria Gonzalez Prelcic—Universidade de Vigo, Spain

EDITORIAL BOARD
Daniel Bliss—Arizona State University, U.S.A
Danijela Cabric—University of California, U.S.A.
Volkan Cevher—École polytechnique fédérale de Lausanne, Switzerland
Mrityunjoy Chakraborty—Indian Institute of Technology, Kharagpur, India
George Chrisikos—Qualcomm, Inc., U.S.A.
Elza Erkip—New York University, U.S.A.
Alfonso Farina—Leonardo S.p.A., Italy
Yifan Gong—Microsoft Corporation, U.S.A.
B. Vikrham Gowreesunker—Texas Instruments Inc., U.S.A.
Joseph Guerci—Information Systems Laboratories, Inc., U.S.A.
Nageen Himayat—Intel, U.S.A.
Clem Karl—Boston University, U.S.A.
Erik Larsson—Linköping University, Sweden
David Love—Purdue University, U.S.A
Maria G. Martini—Kingston University, U.K.
Helen Meng—The Chinese University of Hong Kong, China
Meinard Mueller—Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
Phillip A. Regalia—U.S. National Science Foundation
Alejandro Ribeiro—University of Pennsylvania, U.S.A.
Douglas O'Shaughnessy—INRS Universite de Recherche, Canada
Osvaldo Simeone—Kings College London, U.K.
Milica Stojanovic—Northeastern University, U.S.A
Ananthram Swami, Army Research Labs, U.S.A.
Jong Chul Ye—KAIST, South Korea
Qing Zhao—Cornell University, U.S.A
Josiane Zerubia—INRIA Sophia-Antipolis Mediterranee, France

ASSOCIATE EDITORS—COLUMNS AND FORUM
Ivan Bajic—Simon Fraser University, Canada
Balázs Bank—Budapest University of Technology and Economics, Hungary
Panayiotis (Panos) Georgiou—University of Southern California, U.S.A.
Hana Godrich—Rutgers University, U.S.A.
Yuan-Hao Huang—National Tsing Hua University, Taiwan
Euee Seon Jang—Hanyang University, South Korea
Piya Pal—University of California San Diego, U.S.A.
Vishal M. Patel—Rutgers University, U.S.A.
Christian Ritz—University of Wollongong, Australia
Changshui Zhang—Tsinghua University, China
H. Vicky Zhao—Tsinghua University, China

ASSOCIATE EDITORS—e-NEWSLETTER
Tamir Bendory—Tel Aviv University, Israel
Behnaz Ghoraani—Florida Atlantic University, U.S.A.
Anubha Gupta—IIIT Delhi, India
Alessio Medda—Georgia Tech Research Institute, U.S.A.
Irena Orovic—University of Montenegro, Podgorica
Sarah Ostadabbas—Northeastern University, U.S.A.

ASSOCIATE EDITOR—SOCIAL MEDIA/OUTREACH
Guijin Wang—Tsinghua University, China

IEEE SIGNAL PROCESSING SOCIETY
Ahmed Tewfik—President
Athina Petropulu—President-Elect
Fernando Pereira—Vice President, Conferences
Shrikanth Narayanan—VP Education
K.V.S. Hari—Vice President, Membership
Sergio Theodoridis—Vice President, Publications
Tülay Adalı—Vice President, Technical Directions

IEEE SIGNAL PROCESSING SOCIETY STAFF
William Colacchio—Senior Manager, Publications and Education Strategy and Services
Rebecca Wollman—Publications Administrator

IEEE PERIODICALS MAGAZINES DEPARTMENT
Jessica Welsh, *Managing Editor*
Geraldine Krolin-Taylor, *Senior Managing Editor*
Janet Dudar, *Senior Art Director*
Gail A. Schnitzer, *Associate Art Director*
Theresa L. Smith, *Production Coordinator*
Mark David, *Director, Business Development - Media & Advertising*
Felicia Spagnoli, *Advertising Production Manager*
Peter M. Tuohy, *Production Director*
Kevin Lisankie, *Editorial Services Director*
Dawn M. Melley, *Staff Director, Publishing Operations*

Digital Object Identifier 10.1109/MSP.2020.2984855

SCOPE: *IEEE Signal Processing Magazine* publishes tutorial-style articles on signal processing research and applications as well as columns and forums on issues of interest. Its coverage ranges from fundamental principles to practical implementation, reflecting the multidimensional facets of interests and concerns of the community. Its mission is to bring up-to-date, emerging, and active technical developments, issues, and events to the research, educational, and professional communities. It is also the main Society communication platform addressing important issues concerning all members.

Robert W. Heath Jr. | Editor-in-Chief | rheath@utexas.edu

Communications and Sensing: An Opportunity for Automotive Systems

Like many of you, I am still working remotely, due to COVID-19, while writing this editorial. As in the past two years, I was planning to give an update on the magazine from our editorial board meeting. However, since ICASSP was remote, we have not yet scheduled the board meeting. Instead, I have decided to talk about a topic of personal interest: connections between communications and sensing in the context of vehicular systems. I believe that this is an important signal processing topic that brings together researchers from different technical committees and societies. This editorial is also relevant given the special issue articles and feature article found in this issue.

In past *IEEE Signal Processing Magazine* (*SPM*) editorials, I have discussed vehicular applications of signal processing, going toward 6G cellular communications, and new opportunities in communications as realized during a pandemic. In each of these articles I hinted at potential opportunities related to communications and sensing (especially radar). These topics have been featured in other *SPM* content including a two-part series on advances in radar systems for modern civilian and commercial applications (issues 4 and 5 in 2019) and a two-part series on autonomous vehicles (this issue and the upcoming January 2021 issue). There is also alignment with the IEEE Signal Processing Society (SPS) Autonomous Systems Initiative. In this editorial, I expand on the potential of combining communications and sensing in different ways.

Vehicles are being equipped with more sensors to support higher levels of automation. These sensors form a network on the vehicle, whose information is fused for tasks like trajectory planning and obstacle avoidance. Perhaps even more interesting, though, is the combination of sensing and communications that turns a network of vehicles into a cooperative perception system. The sensor data from each vehicle can be exchanged and fused to create a more accurate picture of the environment, leveraging the multimodality of the sensors and their different perspectives. Imagine the contrast with sensor networking research from two decades ago [1], which envisioned networks of low-power, low-cost sensors with limited communication capability. The networks of sensor networks in a transportation system has vastly more capable sensors, highly advanced signal processing, significant computational resources, automation, and much more communication capability.

For some of you, if you have read this far, you may be wondering why I have written another editorial that discusses communications. Where is the speech signal processing? Where is the biological signal processing? What about fast implementation of algorithms? In short, despite what I usually claim to my family, I do not know everything. On the aforementioned topics, I am working with *SPM*'s area editors to encourage more content in those areas as well. Apologies for—yet again—not straying from my core research area. This is an interesting time where communications and sensing are being combined in new ways in automotive, aerial, and other applications for consumers.

Using communication signals for radar

Radar and wireless communications share the same electromagnetic spectrum and have some common features in their waveforms. The purpose though is drastically different. In radar, the information (e.g., about a target) is captured in the system that transforms the transmitted signal to the observed received signal. For example, an observed Doppler shift may be related to a target's velocity. In wireless communication, the information (e.g., what is known at the transmitter but unknown to the receiver) is encoded into the transmit waveform, which is disrupted by a system (propagation channel, circuit impairments, noise) and observed at a receiver. The receiver removes the effects of the system (which normally involves tasks like channel estimation and equalization) with the objective of discovering what was transmitted. While both radar and communications may estimate aspects of the propagation environment, they each use them for a different purpose. Further, a communication waveform does not make the best radar waveform and vice versa.

Digital Object Identifier 10.1109/MSP.2020.2990328
Date of current version: 26 June 2020

It is natural to consider ways that one type of signal can be used for the other type's purpose. Here I highlight some examples where a communication signal is used for radar. In essence, the objective is to exploit the known parts of the communication signal to estimate the unknown parameters related to the environment. It is an old idea. Early examples of bistatic radar using television signals are reviewed in [3]. What makes the topic current, though, is the use of low-power (relative to TV) access points, base stations, and devices for this purpose.

There are a number of recent examples involving Wi-Fi signals, much of it backed up with experimental data. For example, in [4] a Wi-Fi signal is used to perform through-wall imaging, for example, to detect the presence of a person. In [5], a Wi-Fi signal was used for gesture recognition. In essence, the micro-Doppler characteristics were used to train a machine learning method to classify different hand movements into one of a set of gestures. An application of this technology is whole-house audio volume control. In our work [6], we considered the use of the Wi-Fi for vehicles, known as *dedicated short-range communication*, for radar. The performance was not as good as what can be achieved by millimeter-wave automotive radar, but the cost is potentially much lower. In our other work in [7], we considered the use of the IEEE 802.11ad millimeter-wave communication signals for both communication (between two vehicles) and radar. The system gives high data rates and good radar performance, which gives extra robustness and provides additional security through diverse sources of information (the receiving vehicle broadcasting back its position, and the transmitting vehicle measuring the position of the receiving vehicle with radar).

Combining communications and radar together

In the aforementioned examples, a radar was designed to work with a given communication waveform. While it might share the same hardware as communication, no attempt was made to modify the communication waveform to better suit radar. It is possible to imagine, though, a system where tradeoffs are made, for example, that reduce communication throughput but increase radar system performance. Such tradeoffs have been a topic of recent work [8]. It is also the topic of a recent *SPM* article [9] as well as the article "Joint Radar-Communications Strategies for Autonomous Vehicles" by Ma et al. in this issue.

There are different ways that tradeoffs could be accommodated in a system. One approach would be to time multiplex between a communication waveform and a radar waveform. The time duration could be varied depending on the desired operational performance targets. The advantage of this approach is hardware reuse and potential ease of dealing with the full duplex problem (since radar waveforms are usually designed with that in mind). Another approach is a bit more integrated into the communication. The frequency of training symbols or pilots could be varied, for example to improve velocity estimation. The quality of the radar parameter estimates could improve and the channel estimate would be better for communications, but the data rate would decrease due to fewer data symbols. Systems in the future could be designed with both communication and radar sensing combined together in different combinations of joint, active, and passive radar to create for example a perceptive cellular network [10].

Leveraging sensors to aid communications

There has been a lot of work on using communication waveforms for radar. But can radar, or other sensors found in automated vehicles like cameras or lidar, be leveraged to support communications? This is interesting because such sensors use different spectrum than communications, and thus their use does not consume the limited and valuable communication resources. Further, they are already present on automated vehicles to support other automation tasks so their use does not necessarily have a cost or power penalty. The key question then is whether a millimeter-wave radar, a visible light lidar, or a visible light camera can infer something about the environment of relevance to the radio frequencies used by wireless communications?

A millimeter-wave radar has the potential to provide relevant information about a millimeter-wave communication link given the proximity of the frequencies. One early approach along these lines was presented in [11], where the radar was used to make an inference about good (or bad) communication directions to aid in millimeter-wave beam training. Intuitively, a radar should be able to help in other ways as well. For example, in a cellular communication scenario, a radar could track a vehicle that is communicating with the base station, reducing the overheads due to channel tracking. This makes sense because a radar at the base station can provide situational awareness that can be broadcast to surrounding vehicles to improve vehicle automation [12].

In contrast to bistatic radar with television transmissions, radar could also be used as a signal of opportunity. For example, work in [13] shows how millimeter-wave radar emissions from vehicles can be collected at the cellular base station and used for millimeter-wave beam training, which is a completely passive approach.

Deep learning is a valuable tool for uncovering the correlations between the sensed environment and the communication actions. Besides radar, automated vehicles have other sensors like lidar. Despite the drastic frequency differences, lidar data though can be used by a deep learning engine to detect if a link is in the line-of-sight state and also to reduce beam training overheads [14]. One could imagine that object detection and tracking via cameras (another topic with a significant deep learning component) could also be used in the same way.

Wrapping up

I believe that the interplay between sensing and communications has many opportunities for signal processing researchers. Fundamentals play a role in developing intuition and building algorithms to make good tradeoffs. Experimental work takes the spotlight given

(continued on page 13)

PRESIDENT'S MESSAGE

Ahmed Tewfik | IEEE Signal Processing Society President | tewfik@austin.utexas.edu

Audere est Facere

We continue to live through a unique experience in history. Out of concern for each other, we have voluntarily participated in essentially shutting down economic activities across the globe. We have discovered the interdependencies and precariousness of our lives and livelihoods. We have learned who and what is essential or important and have simplified our lives. We have realized the virtue of patience and self-kindness as we navigate the tremendous challenges of working from home and balancing our work obligations and family needs. Above all, we have surprised ourselves by implementing innovations that we had previously deemed too risky. Even more surprisingly, some of these innovations have allowed us to thrive in new ways and even enhance our effectiveness.

Our flagship ICASSP conference, which concluded a couple of weeks before I wrote this column, is just an example of such innovation. For a long time, we debated adding a virtual component to our conferences but shied away from taking action. The COVID-19 pandemic forced us to turn ICASSP into a virtual event in a very short time. We all missed the impromptu discussions, networking, and chance encounters that ignite our creativity and make physical meetings so enriching. However, there is no question that the virtual experience was a success with surprising advantages. The following two quotes from our colleagues summarize the overwhelming feedback we received:

- Naturally I am sure the value of meeting colleagues in-person will never change even after the pandemic, but I think allowing participants to choose an option to attend conferences virtually would solve the inequality problem for those colleagues disadvantaged geographically and/or financially.
- In the opening ceremony, Dr. Tewfik mentioned that in the future, when real (i.e., not virtual) ICASSP becomes possible again, they might consider making virtual attendance also available as an option. So, the entire conference can be attended either physically or virtually. Having such a choice is a great idea…With talks and papers available online, the learning experience is so great.

These quotes eloquently make the case that the virtual aspect of ICASSP needs to become permanent and adopted by all our meetings.

> **I am happy to report that the Society is actively reworking the physical aspect of our future conferences and adding a virtual component.**

I am happy to report that the Society is actively reworking the physical aspect of our future conferences and adding a virtual component. Specifically, the Society is considering plans that aim to create a richer, more diverse, and connected innovation environment that can supercharge our creativity and enhance our professional development.

Virtual conferences aren't the only aspect of the Society that we need to fundamentally rethink. Publications, our conference offerings, membership diversity, and services—including differentiated educational offerings and business models—are among the many challenges that we need to reinvent to build a true innovation ecosystem. I briefly discuss publications and diversity in this column. We will visit other aspects in future columns.

As a community, we have debated open-access publications for almost two decades. Our sister society, ACM, has demonstrated unambiguously during the crisis how open access can unleash creativity and increase the visibility of peer-reviewed publications. It is time for us to implement open access across our publications and for our volunteer leaders to devise and implement a solid financial plan to support open access. This will necessarily involve support from industry and perhaps governments. Industry support will force the Society to create a powerful platform that stimulates rich interactions between industry, start-ups, and academia. Such an ecosystem

Digital Object Identifier 10.1109/MSP.2020.2990329
Date of current version: 26 June 2020

can turbocharge academic and industrial creativity and innovation. It will also force the Society to seriously address the needs of the 50% of its members and 60% of its U.S. members, who are currently underserved.

> **Innovation is the product of exchanges between brilliant minds with dissimilar experiences, backgrounds, genders, and belief systems.**

But we can't always lag behind technology and trends. The challenge with publications isn't only peer-reviewed open access. The success of arXiv established the value of rapid open access to preprints that have undergone moderation but not full peer review. Our open-access strategy will need to judiciously incorporate preprints that aren't peer reviewed and anticipate, as opposed to wait for, the evolving nature of peer reviewing and authorship.

Finally, there can be no innovation ecosystem without diversity. Innovation is the product of exchanges between brilliant minds with dissimilar experiences, backgrounds, genders, and belief systems. This is the time to address our unacceptably low racial and gender diversity. This is also the time to engage physically with our worldwide membership and expand in areas that we have neglected. We can only eliminate these fundamental weaknesses by innovating and implementing strategies that we had deemed hitherto unworkable or unrealistic. We need to empower and unleash our younger members, aggressively recruit members outside of academic circles, and adopt local and industry sector dependent strategies. We must strive to be known as a welcoming and embracing Society that will always be there to give your professional career a boost.

In the next months, the Society will reach out to you to get your innovative suggestions, ideas, and feedback on plans before they are executed. This is YOUR Society and we all need to participate in its governance. In the meantime, please email me your ideas and feedback at sps-president@ieee.org.

Ahmed Tewfik

SP

Join the 5-Minute Video Clip Contest (5-MICC) at ICIP 2020

Fight the Pandemic

The IEEE Signal Processing Society is pleased to announce the **5-Minute Video Clip Contest (5-MICC)** at ICIP 2020 in United Arab Emirates (virtual conference) 25-28 October, 2020.

- Topic: image and video processing tools for fighting the coronavirus (COVID-19) pandemic. The submitted video can cover any role that image and video processing may have in management of pandemics. Examples: detection of coronavirus, identifying and tracking infected people, COVID-19 detection from chest X-rays and CT images, and patients monitoring. The contest will also accept "open topic" video submissions, regarding image and video processing for medicine.
- Eligibility: Any team composed of one faculty member, at most one graduate student and 3-5 undergraduate students is welcomed to join the open competition
- Website: The detailed guidelines are available on the official website: https://signalprocessingsociety.org/get-involved/five-minute-video-clip-contest
- Prize: The three teams with highest performance in the open competition will be selected as finalists and will be invited to participate in the final competition at ICIP 2020 (virtual). The champion team will receive a grand prize of 5,000$. The first and the second runner-up will receive a prize of 2,500$ and 1,500$, respectively.

IEEE Signal Processing Society

Important dates:
1. **Submission of 30-Second Trailers**: August 20, 2020.
2. **Announcement of the best 10 teams**: August 31, 2020.
3. **Submission of the Full 5-Minute Video**: September 30, 2020.
4. **Final Contest at ICIP**: 25-28 October, 2020.

An initiative of **The IEEE Technical Committee for Signal Processing Theory and Methods (SPTM)**
Sponsored by the IEEE Signal Processing Society

Digital Object Identifier 10.1109/MSP.2020.2999215

John Edwards

Signal Processing Advances Undersea Research

Subsea research presents unique challenges that signal processing is helping to address

Oceans cover approximately 71% of Earth's surface yet remain difficult to explore and monitor remotely. Communication challenges, including lengthy propagation delays, Doppler effects due to vehicle and water movement, and the highly dynamic multipath nature of the undersea environment, can result in significant errors and outliers in transmissions, received data measurements, and image analysis.

Global research teams are now investigating methods incorporating signal processing that promise to make undersea communication and data interpretation faster and more reliable.

An undersea software-defined network

Dynamic, programmatically efficient software-defined networks (SDNs) are already widely deployed terrestrially. Now, researchers at Florida Atlantic University (FAU) are working to design, deploy, and test an SDN that reliably supports real-time undersea wireless communications—data, voice, and video streaming—for monitoring, data collection, safety, surveillance, and other services. The FAU Underwater Autonomous Network (UAN) project aims to create the foundation for a heterogeneous network architecture that can serve autonomous undersea vehicles and enable seamless connectivity with autonomous water-surface vehicles, unmanned aerial vehicles (UAVs), and satellites.

Dimitris Pados, the project's coprincipal investigator, a professor in FAU's Department of Computer and Electrical Engineering and Computer Science, and director of the FAU Center for Connected Autonomy and Artificial Intelligence, says UAN is being designed to support a variety of different underwater acoustic and ultrasonic network devices and interface in radio frequency for over-the-air networking. In all domains, underwater or over air, the overall network system intelligently avoids intra- and internetwork spectrum interference, and, at the same time, optimizes data routing around physical interference and network bottlenecks to maximize data throughput.

The team, including coprincipal investigator Stella N. Batalama, dean of FAU's College of Engineering and Computer Science, also aims to improve existing methods of undersea localization and tracking, which tend to be costly and power intensive. "Underwater wireless communications and networking has a wide range of applications, yet remains a daunting task due to the very nature of the water propagation medium," Pados observes. Lengthy propagation delays, Doppler effects created by vehicle and water movement, and the highly dynamic multipath nature of the undersea environment frequently result in significant errors and outliers in received data measurements. "There is a great need for modular software and hardware architectures that enable rapid deployment and evaluation of new robust protocol designs," he says.

Pados notes that at least one aspect of the project can be classified as an underwater-connected autonomy initiative. "Imagine a school of robotic fish, each with a tiny UAN modem on board its robotic body," he suggests. "The fish can find each other and talk with each other to pursue multiagent robotic planning tasks underwater," Pados explains, adding that project researchers are currently working with mechanical and ocean engineering colleagues at FAU to make schools of connected fish a reality (Figure 1).

Digital signal processing (DSP) theory and practice is essential to the UAN operations in two important ways, Pados observes. For underwater network initialization and neighbor discovery, robust localization is necessary. "In the UAN, we carry out localization by novel L1-norm (absolute value) signal subspace analysis, followed by a conventional signal angle-of-arrival estimation—as has been in use for over 30 years in the signal processing literature—known as multiple signal classification (MUSIC)-type estimation," he explains.

Physical signal waveform shape optimization, which is used to avoid sensed interference in space-time, is also accomplished by DSP signal/noise subspace analysis methods. "At each receiver

---

Digital Object Identifier 10.1109/MSP.2020.2982856
Date of current version: 26 June 2020

1053-5888/20©2020IEEE

location, an estimate of the local in-space and time disturbance autocorrelation matrix is carried out," Pados explains. "Then, conventional DSP eigenvector analysis of the disturbance matrix can point toward optimal interference-avoiding waveforms." More robust analysis of the disturbance autocorrelation matrix by means of the L1-norm absolute value metric is also possible to guard against occasional faulty measurements, he notes.

Subspace signal processing and tracking, using underwater miniaturized microphone arrays for localization, and subspace signal processing of disturbance measurements for maximum signal-to-interference-plus-noise waveform design are used to support the UAN's interference-avoiding properties.

Pados says the goal is to create the ultimate autonomous, artificially intelligent network structure, one that can maintain end-to-end connectivity at the highest possible data rate in extreme environments. "UAN can be the first-ever, underwater autonomous cognitive network," he states.

The project is already attracting significant commercial interest, particularly from the diving community. "The FAU UAV can support diver-to-diver-to-surface-boat multihop connectivity with profound operational and safety implications," Pados notes.

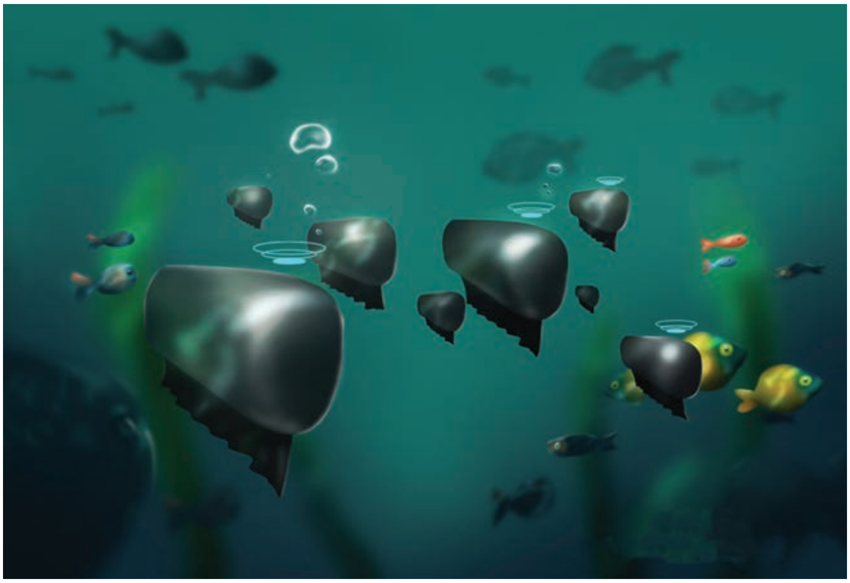


FIGURE 1. A representation of undersea robotic probes monitoring a marine environment using a new type of SDN currently being developed by researchers at FAU. (Source: Florida Atlantic University.)

FIGURE 2. The exploration vessel *Falkor* launches a probe designed to acquire images from the seabed for a research project aiming to show how combining autonomous robotics and AI technologies can dramatically accelerate the exploration and study of hard-to-reach deep-sea ecosystems. (Source: Schmidt Ocean Institute.)

Improved seabed mapping

Satellite technology has led to precise, constant mapping of Earth's entire surface. Accurate undersea mapping, however, remains frustratingly challenging. Research led by Blair Thornton, an associate professor at both the University of Southampton and the Institute of Industrial Science at the University of Tokyo, aims to show how combining autonomous robotics and artificial intelligence (AI) technologies can dramatically accelerate the exploration and study of hard-to-reach deep-sea ecosystems.

In collaboration with researchers at the University of Southampton's Southampton Marine and Maritime Institute, the Institute of Industrial Science at the University of Tokyo, the Australian Center for Field Robotics at the University of Sydney, and the University of the Balearic Islands, Thornton has developed a method to identify and correlate features in georeferenced seafloor images by using autoencoders, an artificial neural network technology. "This is an unsupervised approach that allows us to cluster images into groups of similar-looking images and also query images based on their relative similarity to each other," he explains. "If we find something that's interesting on the seafloor, we can very quickly generate maps of where similar-looking areas exist in our maps." The approach promises to help scientists respond to dynamic environmental changes and to pinpoint areas leading to the most promising scientific, operational, or environmental-management gains (Figure 2).

One of the key drawbacks with current seabed imaging technologies is that observations are limited by the strong light attenuation that occurs underwater.

"We started working on extending the range of seafloor imaging from one or two meters to more like ten meters," Thornton says. "We then developed a system that uses a combination of laser line projections and strobed stereo imagery that can map about 50-times-larger areas of the seafloor than conventional visual mapping systems." Meanwhile, Kalman filters are used to help accurately position the project's sensor-equipped robots, a particularly important task given the fact that GPS signals cannot penetrate deeply underwater.

Many existing image processing techniques excel at identifying patterns or objects within a single image frame yet are far less effective at handling features that are only apparent at larger spatial scales. "Underwater, the footprint of individual images is small compared to the types of geological and ecological features we are interested in characterizing, so we needed to develop new methods to identify these [objects]," Thornton notes.

Automating image analysis is a key project goal. "We use processing to correct the appearance of images that have been affected by strong, wavelength-dependent attenuation, the scattering of light, and uneven illumination patterns to produce images that look like the water has been drained away and removed," Thornton explains. "The automated interpretation of images, using autoencoders that we have adapted to take into account the location of observations, is critical in helping us quickly understand patterns in the data that occur on various spatial scales so that we can focus our observational efforts."

Underwater images are strongly affected by the local environment, meaning that observations are frequently at the mercy of factors that lie beyond human control. "For instance, the range we image the seafloor from is relatively near at about 10 meters," Thornton states. At the same time, the seafloor terrain is often rugged, with protrusions of several meters, and strong currents mean that the project's robots are easily displaced from their ideal trajectories by up to several meters in all directions.

Even a small variance in range to the seafloor creates a dramatic difference to an image footprint's relative size, as well as the color balance and backscatter. The types of features within the environment being observed, such as coral reefs, mineral deposits, and artificial infrastructure, also play a major role in determining image size and characteristics. "Therefore," Thornton maintains, "it's important that the techniques we use are data driven to make our processing methods robust to all these factors."

Thornton notes that the algorithms the researchers are developing to correct image appearance, and an unsupervised interpretation of imagery based on feature learning, are all data driven. They basically self-calibrate to adapt to environmental conditions, such as the clarity of the seawater, and adapt to the types of features we see in the environment to form useful information summaries. The technology allows maps—which would ordinarily take up to several months to process and complete—to be rendered onboard a vessel within days to provide accurate and current seabed views.

The researchers are aiming to produce a system that is highly reliable, even in the most demanding marine environments. "It's happened in the past that we've deployed imaging systems where the images weren't being properly saved, or that there has been an issue with the strobes we use to illuminate the seafloor not firing correctly," Thornton explains. "We try to have our systems self-diagnose and report their status, where possible, to avoid these things happening."

While the team's focus is on acquiring high-quality seabed imagery, Thornton says that the methods being developed to rapidly interpret large-scale visual reconstructions are also readily applicable to terrestrial and extraterrestrial imaging.

Thornton adds that the team also spends a considerable amount of time

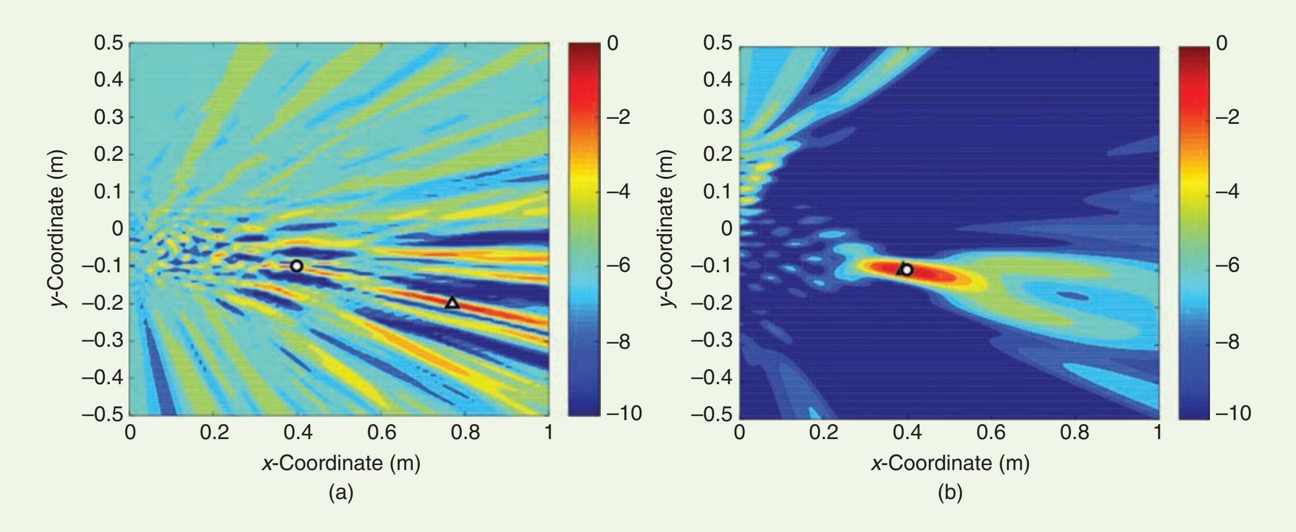


FIGURE 3. An example of what an acoustic system detects: (a) with traditional signal processing, and (b) with the new technique developed by UM engineering researchers. Both images were created from the same sound recordings. (Source: University of Michigan Engineering.)

pondering what types of hardware issues might occur while out at sea and what kinds of spare parts may be needed. “We often go out to sea for periods of several weeks to months, often with no access to the Internet,” he explains. Under such conditions, problems that can be addressed relatively easily on land by simply rushing off to a hardware store or downloading new software are unsolvable. “You have to be meticulous and well prepared and, even then, things still go wrong,” he notes.

Revealing the hidden information in sound waves

By lowering the pitch of sound waves, University of Michigan (UM) engineering researchers have developed a technique that is designed to unlock greater amounts of data from acoustic fields. The additional information that was revealed promises to enhance the performance of passive sonar and echolocation systems, both of which are used for detecting and tracking seagoing enemy vessels, as well as improve the performance of medical imaging devices, seismic surveying systems for locating oil and mineral deposits, and, potentially, radar systems.

David Dowling, a professor in UM’s Department of Mechanical Engineering, observes that many important modern remote sensing technologies are based on the analysis of recorded propagating wave signals generated by, or scattered from, items of interest. “Modern implementations of these technologies primarily rely on signal processing schemes that extract information from within the frequency bandwidth of the recorded signals,” he explains. “The unconventional approach I’ve been working on for the last 10 or so years can extend the signal processing frequency range to frequencies below and above the in-band range.”

With the help of a Fourier transform, sound amplitude versus time can be converted to sound amplitude versus frequency. The approach allows the researchers to combine any two frequencies within the signal’s recorded frequency range to reveal information outside that range at a new, third frequency that is the sum or difference of the two input frequencies. When used with a naval vessel’s sonar array, the additional information could reveal the presence of an enemy ship or underwater craft or device many kilometers away—even over the horizon (Figure 3).

Dowling says the technique he has been working on for the past decade or so can extend the signal processing frequency range to frequencies below and above the in-band range. “These techniques are based on the formation and utilization of quadratic—nonlinear—products of recorded signal-spectrum amplitudes from two different in-band frequencies,” he notes. Once formed, the autoproducts may be used in conventional-array signal processing schemes for remote sensing as if they were recorded wave signals at the difference or sum frequencies, even though the difference and sum frequencies may lie outside the bandwidth of the original recordings.

Both frequency-difference and frequency-sum approaches can be used whenever underwater acoustic signals are recorded. “For nearly all of the work my group has undertaken, we have assumed that the acoustic signal of interest has been recorded simultaneously at multiple locations [via a receiving array] and that some information about the absolute or relative location of the receivers is available,” Dowling reports. “This location information and relative timing information between receivers may be uncertain, too.”

Dowling believes that the frequency-difference and frequency-sum approaches could eventually benefit all propagating wave-based remote sensing technologies, including sonar, ultrasonic imaging and inspection, radar, seismology, and structural health monitoring. Yet, he acknowledges, more work lies ahead.

Author

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area. Follow him on Twitter @TechJohnEdwards.

Lina J. Karam, Jay Katupitiya, Vicente Milanés, Ioannis Pitas, and Jieping Ye

Autonomous Driving: Part 1—Sensing and Perception

The integration of advanced sensing, signal processing, artificial intelligence, and controls technologies into vehicles is enabling intelligent automated vehicles that can navigate autonomously in various environments. In particular, autonomous driving and, more generally, automated driving are receiving more attention, with significantly increasing resources deployed to enable safe, reliable, and efficient automated mobility in complex, uncontrolled real-world environments and for various applications ranging from automated transportation, and farming to public safety and environmental exploration. Signal processing is a critical component of automated driving. Some of the needed enabling technologies include affordable sensing platforms that can acquire useful data under varying environmental conditions; reliable simultaneous localization and mapping; machine learning that can effectively handle varying real-world conditions and unforeseen events; "machine learning- friendly" signal processing to enable more effective classification and decision making; hardware and software codesign for efficient real-time performance; resilient and robust platforms that can withstand adversarial attacks and failures; and end-to-end system integration of sensing, signal processing, machine learning, and controls.

This special issue on autonomous driving will be presented in two parts: Part 1—Sensing and Perception and Part 2—Learning and Cognition [scheduled for publication in the January 2021 issue of *IEEE Signal Processing Magazine* (*SPM*)].

In this issue

The goal of Part 1 is to provide researchers and professionals with tutorial-style articles covering the current state of the art as well as emerging trends in the design, development, and deployment of sensing and perception technologies for autonomous and automated driving. Such technologies include camera, ultrasound, Global Navigation Satellite System-, lidar-, and radar-based platforms integrating signal processing components to process the acquired data and extract information to be used for recognition, navigation, and situational awareness. Despite recent advances in such sensing platforms, the performance of these sensors can be significantly constrained by their quality–cost tradeoff, excessive energy consumption, and inconsistency under varying environmental conditions. Key concepts and the latest advances underlying the operation of such sensing technologies are discussed in Part 1 of this special issue. This special issue also sheds light on remaining challenges that need to be addressed to enable further performance improvements.

> **The integration of advanced sensing, signal processing, artificial intelligence, and controls technologies into vehicles is enabling intelligent automated vehicles that can navigate autonomously in various environments.**

Overview

Part 1 contains seven articles covering various aspects of sensing and perception for autonomous driving. The first two articles deal with problems related to robust sensing for autonomous driving, whereas the remaining five articles are each focused on a particular sensing modality (camera, lidar, or radar).

Among the robust sensing articles, the first article, "Toward Robust Sensing for Autonomous Vehicles," by Modas et al., addresses the topic of adversarial attacks that take the form of crafted alterations of the physical environment or of the sensory measurements with the objective of attacking and defeating the autonomous vehicle. The authors provide an overview of adversarial attacks for various sensing modalities and discuss countermeasures and research directions to build and deploy safer autonomous driving systems. The second article, "Automated Vehicular Safety Systems," by Stöckle et al., presents a methodology for jointly designing the functions and sensors of automated vehicular safety systems,

---

Digital Object Identifier 10.1109/MSP.2020.2990330
Date of current version: 26 June 2020

while accounting for both sensor measurement errors and customers' requirements.

Among the sensing modality articles, Chen et al. provide an overview of the emerging bio-inspired neuromorphic vision sensing in "Event-Based Neuromorphic Vision for Autonomous Driving," including key concepts, underlying signal processing algorithms, application in autonomous driving, and remaining challenges. In "Lidar for Autonomous Driving," Li and Ibanez-Guzman address the topic of automotive lidar. They introduce the main components of automotive lidar systems and present a review of the state of the art as well as challenges and trends. Rapp et al. present the working principles of single-photon lidar in "Advances in Single-Photon Lidar for Autonomous Vehicles" and discuss recent advances in signal processing techniques for this modality, applications in autonomous vehicles, and challenges for vehicular lidar. Aydogdu et al. address the topic of automotive radar interference in "Radar Interference Mitigation for Automated Driving" and discuss methods to mitigate such interference with a focus on frequency-modulated continuous wave (FMCW) radar. The article also provides a review of automotive radar and an introduction to the basics of FMCW radar. In "Joint Radar-Communications Strategies for Autonomous Vehicles," Ma et al. present a survey of dual-function radar-communications methods within the context of autonomous vehicles. Main challenges and potential research directions are also discussed.

Acknowledgments

We extend our appreciation to *SPM*'s Editor-in-Chief Robert Heath and Area Editor, Special Issues Namrata Vaswani, for their valuable input. We also thank IEEE Signal Processing Society (SPS) Publications Administrator Rebecca Wollman, and IEEE Magazines Managing Editor Jessica Welsh, for their support. Last but not least special thanks go to the contributors and reviewers without whom this special issue would not have come to fruition. This special issue is technically sponsored by the IEEE Autonomous Systems Initiative of the IEEE SPS.

Guest Editors

Lina J. Karam (lina.karam@lau.edu.lb) is the dean of the School of Engineering and a professor at the Lebanese American University. She is also a professor emerita at Arizona State University. She is the editor-in-chief of *IEEE Journal on Selected Topics in Signal Processing*, and a member of the IEEE TechRxiv Advisory Board, the IEEE Access Editorial Board, and IEEE Signal Processing Society (SPS) Awards and Publications Boards. She served as general chair of IEEE ICIP 2016 and as general cochair of the IEEE International Conference on Multimedia and Expo 2019. She is a recipient of the NSF CAREER, NASA Technical Innovation, IEEE Region 6, IEEE SPS Best Paper, and Intel Outstanding Researcher Awards. She has authored more than 240 technical publications and is an inventor on seven issued U.S. patents. She is a Fellow of the IEEE.

Jay Katupitiya (j.katupitiya@unsw.edu.au) received his B.Sc. degree in engineering from the University of Peradeniya, Sri Lanka and his Ph.D. degree from the Catholic University of Leuven, Belgium. He is currently an associate professor at the University of New South Wales (UNSW) in Sydney, Australia. Previously, he was the deputy head of the School of Mechanical and Manufacturing Engineering at UNSW and completed pioneering work in establishing the Mechatronic Engineering degree program at UNSW. Subsequently, he served as the head of the Mechatronic Engineering program at UNSW. His research focus is on unmanned field vehicles, and he has contributed to the development of a number of field-scale unmanned systems for agriculture, mining, and road construction. As a secondary area of research, he conducts space robotics research developing space robots to capture foreign objects in orbit.

Vicente Milanés (vicente.milanes@renault.com) received his Ph.D. degree in electronic engineering from the University of Alcalá, Madrid, Spain, in 2010. Previously, he was with the AUTOPIA program at the Center for Automation and Robotics (UPM-CSIC, Spain) from 2006 to 2011 and was then awarded a two-year Fulbright fellowship at California PATH, University of California, Berkeley. In 2014, he joined the RITS team at INRIA, France. Since 2016, he has been with the Research Department at Renault, France. He is the author or coauthor of more than 120 refereed publications in international journals, book chapters, and conference proceedings and has more than 10 industrial patents. His research interests focus on multiple aspects in the autonomous vehicle field.

Ioannis Pitas (pitas@aiia.csd.auth.gr) is a professor in the Department of Informatics at AUTH and the director of the Artificial Intelligence and Information Analysis lab. He served as a visiting professor at several universities. His current interests focus on computer vision, machine learning, autonomous systems, and intelligent digital media. He has published more than 900 papers, contributed to 47 books in his areas of interest, and edited or coauthored 11 books. He has also been the general or technical chair of four international conferences. He participated in 70 R&D projects, primarily funded by the European Union and is or was principal investigator in most of them. He has more than 30,800 citations to his credit and an h-index of over 83. He is the chair of the IEEE Signal Processing Society Autonomous Systems Initiative, a Fellow of the IEEE and of EURASIP, and an IEEE Distinguished Lecturer.

Jieping Ye (yejieping@didichuxing.com) is the head of Didi AI Labs and vice president of Didi Chuxing. He is also a professor at the University of Michigan, Ann Arbor. His research interests include big data, machine learning, and data mining with applications in transportation and biomedicine. He has served as an associate editor of *Data Mining and Knowledge Discovery*, *IEEE Transactions on Knowledge and Data Engineering*, and *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He won the NSF CAREER Award in 2010, and the INFORMS Wagner Prize in 2019. His papers have been selected for the outstanding student paper at the International Conference on Machine Learning in 2004, the KDD Best Research Paper runner-up in 2013, and the KDD Best Student Paper Award in 2014. He is a Fellow of the IEEE.

SP

FROM THE EDITOR

(continued from page 4)

the lack of models that connect different communication and sensing mechanisms. Many kinds of sensors may play a role beyond just radar, cameras, or lidar. Biosensors could be used to adapt communication in a wearable communication network. I hope to see many contributions to *SPM* in the future.

References

[1] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proc. ACM Int. Conf. Mobile Computing and Networking*, Aug. 1999, pp. 263–270.

[2] IEEE Signal Processing Society autonomous systems initiative. [Online]. Available: https://ieeeasi.signalprocessingsociety.org

[3] H. D. Griffiths and N. R. W. Long, "Television-based bistatic radar," *Proc. Inst. Elect. Eng.*, vol. 133, Part F, pp. 649–657, Dec. 1986. doi: 10.1049/ip-f-1.1986.0104.

[4] F. Adib and D. Katabi, "See through walls with WiFi," in *Proc. ACM SIGCOMM Conf.*, 2013, pp. 75–86.

[5] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Computing and Networking (MobiCom)*, 2013, pp. 27–38. doi: 10.1145/2500423.2500436.

[6] R. C. Daniels, E. R. Yeh, and R. W. Heath, "Forward collision vehicular radar with IEEE 802.11: Feasibility demonstration through measurements," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1404–1416, Feb. 2018. doi: 10.1109/TVT.2017.2758581.

[7] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018. doi: 10.1109/TVT.2017.2774762.

[8] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Radar-communications convergence: Coexistence, cooperation, and co-design," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 1, pp. 1–12, Mar. 2017. doi: 10.1109/TCCN.2017.2666266.

[9] K. V. Mishra, M. R. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 100–114, Sept. 2019. doi: 10.1109/MSP.2019.2913173.

[10] J. A. Zhang, A. Cantoni, X. Huang, Y. J. Guo and R. W. Heath, "Framework for an innovative perceptive mobile network using joint communication and sensing," in *Proc. IEEE 85th Vehicular Technology Conf. (VTC Spring)*, Sydney, Australia, 2017, pp. 1–5.

[11] N. Gonzalez Prelcic, R. Mendez-Rial, and R. W. Heath Jr., "Radar aided beam alignment in mm-wave V2I communications supporting antenna diversity," in *Proc. Information Theory and Applications*, San Diego, CA, Jan. 31–Feb. 5, 2016, pp. 1–7.

[12] A. Ali, N. González-Prelcic, R. W. Heath Jr., and A. Ghosh, "Leveraging sensing at the infrastructure for mmWave communication," *IEEE Commun. Mag.*, to be published.

[13] A. Ali, N. Gonzalez-Prelcic and A. Ghosh, "Automotive radar radiations as signals of opportunity for millimeter wave V2I links," in *Proc. 2019 53rd Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, pp. 554–558. doi: 10.1109/IEEECONF44664.2019.9048967.

[14] A. Klautau, N. González-Prelcic, and R. W. Heath, "LIDAR data for deep learning-based mmWave beam-selection," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 909–912, June 2019. doi: 10.1109/LWC.2019.2899571.

SP

Apostolos Modas, Ricardo Sanchez-Matilla,
Pascal Frossard, and Andrea Cavallaro

Toward Robust Sensing for Autonomous Vehicles

An adversarial perspective

©ISTOCKPHOTO.COM/OONAL

Autonomous vehicles (AVs) rely on accurate and robust sensor observations for safety-critical decision making in a variety of conditions. The fundamental building blocks of such systems are sensors and classifiers that process ultrasound, radar, GPS, lidar, and camera signals [1]. It is of primary importance that the resulting decisions are robust to perturbations, which can take the form of different types of nuisances and data transformations and can even be adversarial perturbations (APs).

Adversarial perturbations are purposefully crafted alterations of the environment or of the sensory measurements, with the objective of attacking and defeating the autonomous systems. A careful evaluation of the vulnerabilities of their sensing system(s) is necessary to build and deploy safer systems in the fast-evolving domain of AVs. To this end, we survey the emerging field of sensing in adversarial settings: after reviewing adversarial attacks on sensing modalities for autonomous systems, we discuss countermeasures and present future research directions.

Robust sensing

The robustness of a decision system refers to its capability of making the correct decision, even when testing conditions are degraded. In particular, it is important that the decisions are not changed when small perturbations alter the input signal. The lack of robustness poses critical safety threats for AVs, and it could even allow an attacker to design APs that defeat their sensing systems.

To formalize the concept of APs, let f_m be a decision-making system (e.g., a classifier or detector) for a sensing modality m, $\boldsymbol{x}$ an input signal of the same modality belonging to a data distribution $\mathcal{X}$, g_v a function that applies a perturbation $\boldsymbol{v}$ on the input signal $\boldsymbol{x}$, and $\mathcal{D}$ a function that measures the distortion between the input and the perturbed signal. Then, $\boldsymbol{v}$ is called an *AP*, and, consequently, $g_v(x)$ is an adversarial attack that generates adversarial examples, if

$$\begin{aligned} f_m(g_v(x)) &\neq f_m(x), \\ \text{subject to} \quad \mathcal{D}(g_v(x),x) &< \epsilon, \\ g_v(x) &\in \Phi, \end{aligned} \tag{1}$$

Digital Object Identifier 10.1109/MSP.2020.2985363
Date of current version: 26 June 2020

where Φ is the legitimate domain of the original signal x, and ϵ limits the distortion of the input signal. This limitation relates to the most common definition of adversarial attacks and implies that the information of the sensed signal is not considerably changed, or even destroyed, by an attack. For example, it is common to control the distortion with the constraint that the original (clean) and the adversarial signals should be close enough in an ℓ_p-norm sense, $\|g_v(x) - x\|_p < \epsilon$. While several adversarial methods focus on additive APs, where $g_v(x) = x + v$, the adversarial attack g_v can be any function that computes and applies a perturbation on a signal (e.g., transformation-based APs).

In general, adversarial attacks can be physical or on signal (see Figure 1). Physical adversarial attacks are perturbations that modify the environment in an adversarial manner, for example, to mislead in the interpretation of traffic signs [2]. Physical attacks can hide objects (i.e., causing false-negative detections), generate observations of inexistent objects (i.e., producing false-positive detections), induce misclassification of detected objects, or counterfeit signals. Physical APs are generally constructed under transformation-based constraints to be invariant to transformations that take place when captured in the physical world, such as the rotations, scaling, and illumination variations that can occur in cameras, for example.

On-signal adversarial attacks modify the data captured by sensors within the decision system. These attacks, which expose the vulnerabilities of the classifiers or detectors, have been mostly used to evaluate, understand, and improve decision systems [3], [4]. However, the ability to design an on-signal attack implies that the attacker already has complete access to the system and to the captured signal. As this scenario relates to the (cyber)security of the AV, it is out of the scope of this article. An extensive overview of on-signal attacks is provided in [5].

We focus now on physical attacks that alter the behavior of AVs exclusively by external actions. For example, attacks that hide objects (jamming) may prevent AVs from detecting pedestrians, possibly causing a collision. Attacks that add objects (spoofing) force AVs to detect inexistent objects, possibly causing the AVs to suddenly stop. Misclassification attacks aim to change the decision of a classifier, such as interpreting a stop sign as a speed limit sign, thus inducing wrong actions. Finally, counterfeiting attacks mimic the properties of the original signals with the aim of delivering adversarial measurements to a sensor (e.g., delayed or wrong GPS signals).

We can distinguish attacks based on the means used by the attackers. Physical adversarial attacks can be perpetrated using devices, objects, or patches (see Figure 2). Devices (e.g., a laser pointer) directly target the sensor. An adversarial object, such as an adversarial traffic sign, can be constructed to mislead a classifier, even if its appearance may be similar to that of a benign object. Finally, a patch, such as a crafted sticker, can be placed on objects in the environment to mislead classifiers.

Next, depending on the attacker's knowledge of the underlying system, attacks can be categorized as white, gray, or black box. In white-box settings, the attacker has complete knowledge of the system (e.g., its architecture and parameters). When machine learning models are used, the attacker also knows the training data used to generate the model. In black-box settings, the attacker has access only to the output of the system. Finally, in gray-box settings, the attacker has only partial information of the system, such as the architecture but not the parameters or the training data, or the architecture and the training data but not the parameters.

APs can also be categorized as data specific or data agnostic. A data-specific AP is a perturbation $\boldsymbol{v}$ computed for a specific signal $\boldsymbol{x}$ of a sensing modality [10], and it can efficiently apply to that particular signal only (e.g., a specific stop sign). An attacker can also construct data-agnostic APs, commonly referred to as *universal APs*, which are applicable to any signal, $\forall x \sim \mathcal{X}$, of a specific modality [11]. An example here is a single perturbation for any street sign that deceives a sensing system with high probability [2].

A subset of data-agnostic perturbations, called *class universal*, are data agnostic only for intraclass signals [12] (i.e., a single perturbation for every stop sign), that is, $\forall x \sim \mathcal{X}_c$, where c is the corresponding class. Data-agnostic and class-universal APs are critical for the robustness of AVs, since a single AP added to the physical world could potentially mislead any classifier of a specific modality, thus posing a safety threat. This potential threat derives from the fact that a strong property of data-agnostic and class-universal APs is their transferability: APs computed on a known model can also mislead an unknown model [11].

In the next sections, we analyze physical adversarial attacks and how they affect different sensors. We characterize each attack based on its type, goal (Figure 3), data dependency, system knowledge, and evaluation. To evaluate the attacks, we use the term *attack success rate* to denote the percentage of adversarial examples that successfully mislead the system.

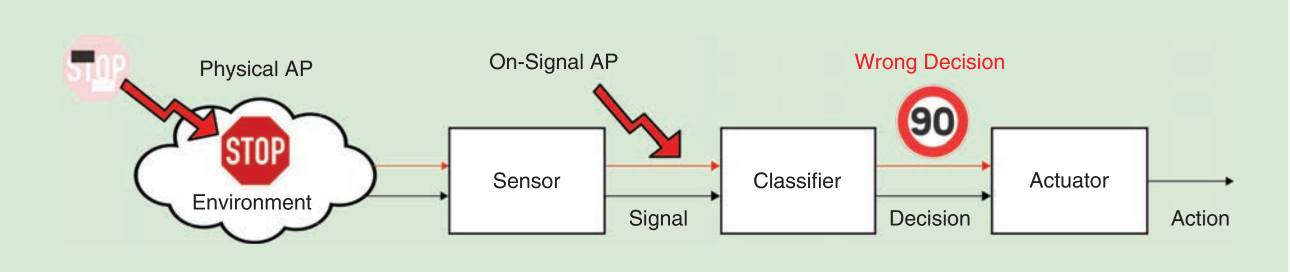


FIGURE 1. The physical and on-signal adversarial attacks.

State-of-the-art methods are compared in Table 1 using the proposed taxonomy.

Attacks using devices

In this section, we discuss physical adversarial methods that use devices to attack ultrasonic sensors, radar, GPS, lidar, and cameras.

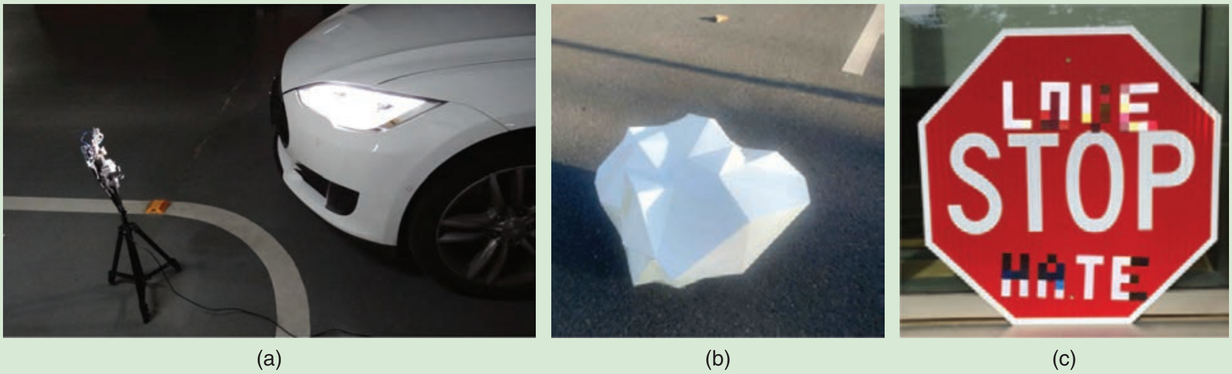
Ultrasonic and radar sensors

Black-box and data-agnostic jamming and spoofing attacks for both millimeter-wave radars and ultrasonic sensors aim to hide or add objects in tasks, such as automatic parking or moving in reverse gear [6]. The jamming transducer generates ultrasonic signals approximately around the sensor's resonant frequency. The jamming signal exceeds the sensor's threshold for the detection of returning echoes (it lowers the signal-to-noise ratio) and prevents the sensor from detecting a parked vehicle. The spoofing attack must be executed in the interval between the end of the transmitted pulse and the start of the first echo; thus, the echoes are injected with a cycle time that is several milliseconds shorter than the sensor's.

This may cause unstable spoofed sensor readings, but it guarantees successful injection in the attack slot. During parking, these attacks lead to collisions, as objects are successfully hidden. When AVs are moving in reverse gear, these attacks are successful at a distance of up to 10 m. However, these methods are applicable only to slowly moving vehicles.

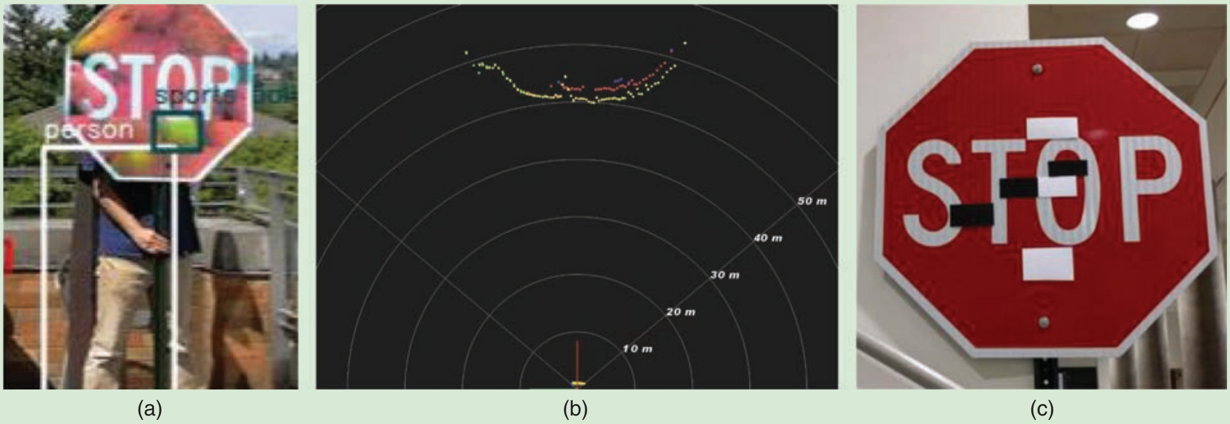
Black-box and data-agnostic attacks for ultrasonic sensors include random spoofing, adaptive spoofing, and jamming [14]. These attacks aim to hide or add objects. Emitting spoofed signals once every few milliseconds causes the AV to stop moving as it detects inexistent objects. These false positives can only be created at distances closer than the distance between the spoofer device and the vehicle.

To create inexistent objects at an arbitrary yet stationary distance, the adaptive spoofer pings a signal, listens and adjusts to the concurrent sensor signals adaptively, eliminates the existing echoes, and transmits the spoofed signal. Similarly to [6], the jamming attack continuously emits ultrasound waves around the resonant frequency of the sensors. Adaptive spoofing can deceive the AV into detecting a nonexistent object at any distance within the sensor range. Jamming prevents an AV



(a) (b) (c)

FIGURE 2. Examples of the three types of adversarial attacks: (a) with a spoofing device to hide objects from an ultrasonic sensor [6], (b) with an adversarial object that is not detectable by lidar [7], and (c) with a patch on a traffic sign that causes misclassification with cameras [2].



(a) (b) (c)

FIGURE 3. Examples of three objectives for adversarial attacks: (a) a sign that is not visible to a camera (hidden object) [8], (b) a spoofing attack that adds nonexistent obstacles for lidar [9], and (c) patches on a traffic sign that cause a camera to misclassify the traffic sign [2].

Table 1. The physical adversarial attacks per sensing modality.

		Type			Goal				Data Dependency			Knowledge			Evaluation	
Sensor	**Reference**	**Device**	**Object**	**Patch**	**H**	**A**	**M**	**C**	**DS**	**CU**	**DA**	**WB**	**GB**	**BB**	**Physical**	**AVs**
Ultrasonic	[6]	✓			✓	✓					✓			✓	✓	✓
	[14]	✓			✓	✓					✓			✓	✓	✓
Radar	[6]	✓			✓						✓		✓		✓	✓
GPS	[15]	✓						✓			✓	✓			✓	
	[13]	✓						✓	✓			✓				
Lidar	[9]	✓			✓	✓					✓		✓		✓	
	[16]	✓			✓	✓					✓	✓			✓	
	[17]	✓				✓					✓	✓				✓
	[7]		✓		✓		✓		✓		✓	✓		✓	✓	✓
Camera	[9]	✓			✓						✓		✓		✓	
	[6]	✓			✓						✓			✓	✓	
	[18]		✓				✓		✓			✓			✓	
	[10]		✓				✓		✓			✓			✓	✓
	[19]		✓				✓			✓		✓			✓	✓
	[2]		✓	✓			✓			✓		✓			✓	✓
	[8]		✓	✓	✓	✓				✓		✓			✓	✓
	[20]			✓			✓				✓	✓			✓	
	[12]			✓	✓					✓		✓			✓	
	[21]			✓	✓						✓	✓			✓	

A: adding objects; AVs: evaluation on AV settings; BB: black box; C: counterfeit signal; CU: class universal; DA: data agnostic; DS: data specific; GB: gray box; H: hiding objects; M: misclassifying objects; WB: white box. Note that [13] is evaluated only in simulations.

from detecting objects, thus potentially causing collisions during automatic parking.

Finally, gray-box jamming and spoofing attacks can target radar sensors to hide objects (i.e., generate false negatives) or to add pseudo-objects (i.e., produce false positives) [6]. Using a signal analyzer to identify the center frequency of the (known) radar sensor, an attacker can generate a jamming waveform, the frequency of which is close to the center frequency of the sensor, thus preventing the detection of neighboring vehicles. For the spoofing attack, the adversarial signals are modulated as those of the automotive radar.

However, since the spoofing attack must be executed in the interval between the end of the transmitted pulse and the start of the first echo, the slope of the ramp is tuned back and forth in a higher value range on the signal. This produces periodic distance changes in the estimations that lead to false identification of pseudo-objects. Similar to the attacks on ultrasonic sensors, these attacks are applicable only against slowly moving vehicles.

GPS

Attacks on GPS sensors usually aim to counterfeit the GNSS transmitted signals with an adversarial transmitter that emits signals identical to those sent by the satellites so that the location computed by the GNSS receivers (i.e., GPS) is incorrect. Such attacks are data agnostic and, usually, white box, since the attacker must know the GNSS implementation of the receivers and transmitters.

The most common adversarial techniques are spoofing and replay attacks. An attacker must first transmit a sufficiently powerful jamming GNSS signal to force the receivers to lose their lock on the satellite. Then, the attacker can forge and transmit spoof GNSS signals at the same frequency and with a higher power, at the receivers' antennas, than those of the legitimate GNSS signals. If the lock of the attacked receivers on the spoofed signal persists, then the location estimation is under the influence of the attacker. An attacker can spoof a varying number of GPS receivers to any location while preserving their mutual time offsets. An analysis of the requirements, architecture, and application of a receiver-spoofing attack is given in [15].

Replay attacks can be characterized by the capability of the adversarial node to receive, record, and replay GNSS signals and to delay the time between reception and retransmission [13]. Since the attacker can replay recorded signals with any additional delay, the receivers will start receiving the replayed navigation messages after some adversary-specific signal propagation delay, which is usually negligible. The delay, the most important feature, is used to control the shift in the position, velocity, and time (PVT) computations by the receivers. The randomized replay attack can create (in simulations) location offsets on the order of hundreds of kilometers.

Lidar

We can identify three types of attacks [9], namely, front/rear side (the attacker is an adversarial vehicle that moves close to the AV), roadside (the attacker installs malicious devices on the roadside), and mechanic (the attacker has limited time to place a malicious device on the AV). Relay and spoofing data-agnostic and gray-box attacks can be used to hide or to add objects. Note that these attacks assume that the technical specifications and datasheets of the sensors are publicly known.

During the relay attack, the original signal sent from the target vehicle lidar is relayed from another position to create fake echoes and make real objects appear closer or farther than their actual locations. The spoofing attack creates inexistent objects: the original signal is used as a trigger to actively spoof the lidar sensor, with the intention to relay or replay objects and control their positions. The relay attack can hide objects that are 1 m away while generating false-positive detections 20–50 m away. The success of the replay attack is limited when generating pseudopoints (on the processed point clouds generated by the lidar sensor) closer than the attacker position. Spoofing attacks can also make an object appear closer or farther away, even beyond the range of the sensor, so that the actual object will not be considered as an obstacle.

In contrast to the replay attack of [9], the white-box and data-agnostic attack of [16] can add pseudopoints (i.e., false positives) closer than the spoofer location. An adversarial device receives the lidar pulse signal, adds the required delay, and fires the delayed laser pulse back to the lidar sensor. Moreover, a saturation attack can blind the sensors, thus hiding objects by illuminating the lidar with a light of the same wavelength as that of the sensor. A weak light source produces randomly located pseudopoints, whereas a strong light source directed to the lidar completely blinds the sensor or a part of its field of view. In addition, a strong light source oblique to the lidar generates fake points in a direction other than that of the light of source.

Finally, the adversarial robustness to spoofing attacks on lidar sensors with deep neural network (DNN) detectors is studied in [17]. A white-box and data-agnostic adversarial attack can strategically control the fake points to fool the machine learning model in the object-detection step. The method adds adversarial fake points at a specific location (distance, altitude, azimuth) with respect to the lidar sensor, with an attack success rate of 75% in generating front-near false positives (5 m in front of an AV). Note that, in this case, all spoofed objects are detected as vehicles, even though this was not in the specific objectives of the attack's formulation.

Camera

We can identify two types of gray-box and data-agnostic attacks that emit light [9], namely, camera blinding and auto control confusion. The blinding attack shines light on the camera sensor, thus hiding any objects within its field of view. Blinding, which can be full or partial, occurs when the camera is unable to tune down the autoexposure or the gain, thus generating overexposed images. The effectiveness of the blinding attack depends on the environmental light, the light source used to blind (i.e., wavelength), and the distance between the adversarial light source and the camera. Experiments in bright and dark environments with different light sources at multiple distances show that 650-nm (red) lasers can efficiently hide objects.

Regarding auto control confusion, in which the goal of the attack is to influence the autoexposure control of the camera, the attacker emits bursts of light toward the camera, and, during the stabilization to the new environmental conditions (typically between 1 and 6 s), the AV cannot detect objects. The attacker continuously switches the light on and off to confuse the auto controls, and, therefore, this attack differs from situations in which the camera can adapt more gradually to the new conditions, such as when driving out of a tunnel. While the blinding attack could successfully blind—in a simulated evaluation with low-quality images—a widely used camera system at specific locations for up to 6 s when the attack is performed with burst lights from a 50-cm distance, this attack is only effective within a limited range (50–200 cm). Under similar settings as [9], the effect of different light sources on hiding objects and other useful visual information is studied in [6]. Although attacking with infrared LED light does not affect the camera, laser beams pointed directly to the camera cause complete blindness for approximately 3 s.

To summarize, adversarial attacks with devices can cause significant problems in various tasks like automatic parking (ultrasound, radar), obstacle detection (lidar), or capturing of the surroundings (camera). However, they usually require the hardware specifications of the individual sensors to be known, while their effectiveness depends on the speed of the vehicles, the range of the attack, and the precision of the used devices.

Attacks with objects and patches

In this section, we discuss physical adversarial methods that use objects and patches to attack lidars and cameras.

Lidar attacks with physical objects

Adversarial methods can generate physical 3D adversarial objects to mislead DNN-based lidar detection systems [7]. In the black-box settings, a data-specific evolutionary-based attack is used to hide a generated adversarial object that is placed in the environment. An iterative procedure is used to generate the object: from an initial mesh of vertices, a new population of vertices is generated at each iteration step by adding random perturbations drawn from a Gaussian distribution. However, when the object is placed in the environment, low attack success rates are obtained.

A data-agnostic method to hide an object or to misclassify a detected one is also proposed for white-box settings. The objective of the method is to generate a synthetic adversarial object by perturbing the vertices of an original one, such that the AV system makes incorrect predictions. For synthesizing the adversarial objects, a differentiable lidar renderer is first simulated; the feature aggregation is, then, formulated with a differentiable proxy function; and, finally, the smoothness of the generated adversarial objects is ensured by devising different loss functions in the detection model. In white-box settings, the hiding attack creates an adversarial object that is not detected 71% of the time, while the misclassification attack creates adversarial objects that are detected as another class (e.g., pedestrian) with high confidence.

Camera attacks with physical objects

The first white-box method for generating adversarial examples to misclassify the decisions of an image classifier even when transferred to the physical world was based on the expectation over transformation (EoT) framework [18]. EoT constructs APs that are robust to a set of physical transformations, such as random rotation, translation, additive noise, or 3D rendering of a texture. The generated 3D objects remain adversarial even when changing the viewpoint, translation and rotation of the object, lighting conditions, camera noise, and other physical world factors, with a classification accuracy dropping from 68.8% to only 1.1%.

Adversarial traffic signs that cause misclassification can be generated with an out-of-distribution attack or a lenticular attack [10]. An out-of-distribution adversarial example can be generated in white-box and data-specific settings, starting from an arbitrary non-traffic-sign image (i.e., a fast food chain logo), such that the arbitrary sign is classified as a traffic sign with high confidence. The AP is forced to lie only on the sign and not on the background and to be invariant to different transformations, such as rotation, shearing, resizing, and randomized brightness. The lenticular attack relies on an optical phenomenon to deceive traffic sign DNN-based classifiers. The idea is based on the fact that the camera will capture the images from a different height and angle than the passengers within the vehicle; thus, the AP might be perceived only from the camera's angle but not from the passengers' points of view.

A white-box, class-universal (universal for any stop sign), object-based attack may target DNN-based classifiers by generating visual APs that are robust to different physical conditions, such as different distance, viewpoint, illumination, and rotation [2]. To ensure that such conditions exist in the image samples, real images were collected from the road, while synthetic ones were simulated using a set of transformations. To guarantee that the perturbations are only applied to the surface of the target object, a mask is used to project the computed perturbations to a physical region on the surface of the object (the road sign). In addition, the mask generates APs that are visible but inconspicuous to human observers (i.e., the mask looks like a graffiti).

Also, since the position of the mask has an impact on the effectiveness of the attack, sparsity is enforced on the generated perturbations so that they are concentrated on regions where the system is more vulnerable. Finally, to account for fabrication errors, a term is added in the objective function to model color reproduction inaccuracies of the printing process. The generated adversarial stop signs were evaluated on a vehicle, with an initial distance from the sign of 76 m, the speed of the vehicle varying up to 32 km/h, and the vehicle moving toward the sign. For two classifiers trained on different traffic sign data sets (achieving 91% and 95.7% classification accuracy), their accuracies dropped to zero.

A white-box and class-universal attack can generate robust physical APs against a DNN for object detection [19], in the sense that the generated APs remain adversarial under various physical condition changes (e.g., distance, angle). This attack focuses on the region proposal stage of the detector, which produces various candidate detection regions that are fed into the classifier stage. The method simultaneously attacks the classification within each proposed candidate region by approximately maximizing, through backpropagation, the classification adversarial loss for every proposed region. After a forward pass on the region proposal network, the pruned region proposals are treated as constants for the classification. Then, the EoT framework is adopted for creating APs that are robust to physical transformations. Finally, the generated robust physical APs can be printed and applied on physical stop signs.

With the goal of misclassification at the corresponding part of the model, the generated signs are printed and placed in an indoor scene with such distance and angle for mimicking driving conditions. When the attack aims at a specific misclassification label, such as a person (targeted attack), the success rate is 87%. When the attack does not target a specific label (untargeted attack), the success rate is 93%. The robustness of the untargeted attack is mostly affected by the distance increment, whereas the robustness of the targeted attack is affected by the view angle. Although the attack focuses on the classification task in a drive-by evaluation (a vehicle approaches the sign from 60.9 m away and with a speed between 8 and 24 km/h), the adversarial examples are also able to affect the detection. A clean stop sign is always detected and classified correctly, but the adversarial signs, on most occasions, are not detected; and, when they are detected, they are misclassified most of the time.

Finally, based on [2], an adversarial attack to object detectors is proposed in [8] with the goal of hiding (preventing the detection of) stop signs. By defining a loss function that outputs the maximum probability of a stop sign occurring within the scene, the extended white-box and class-universal method minimizes this probability until it falls below the detection threshold of the classifier. The method is evaluated indoors and outdoors, with recordings beginning at 9.1 m away from a stop sign and ending when the sign is outside of the camera's field of view. The attack success rate is 85.6% indoors and 72.5% outdoors, and, when the adversarial examples are evaluated on a different model, the attack success rate (transferability) for the indoor case remains high at 85.9%, while, for the outdoor case, it degrades and drops to 40.2%.

Camera attacks with patches

Physical adversarial methods may also use a specific subcategory of objects, namely, patches, to attack cameras for an image-classification or an object-detection task. A data-agnostic, targeted adversarial attack can generate APs that force a DNN to misclassify objects by localizing a single perturbation on every image such that the resulting patch can be printed and installed on objects [20]. The patch is the result of training over a variety of images, where, to encourage the trained patch to work regardless of the background, multiple transformations based on the EoT framework are applied on the patch in each image. Furthermore, an ensemble patch method is used in which a single patch is jointly crafted across five models. In simulated physical attacks (an adversarial patch is added to the image but not printed), for patches occupying only 10% of the image size, the single patch achieves an attack success rate of around 89%, while the ensemble patch leads to an attack success rate of around 95% (averaging across all ensemble classifiers).

The patch-based method introduced in [2] (see the "Attacks With Objects and Patches" section) follows the same approach as the object-based one, with the difference that the AP is masked to take the form of black-and-white patches that are placed on traffic signs. The generated adversarial patches are evaluated on a vehicle with two classifiers, with the vehicle moving toward the sign from an initial distance from the sign of 76 m at a speed varying to up to 32 km/h. As the vehicle approaches the sign, the accuracies of both classifiers show a significant drop.

A white-box and data-agnostic physical patch generation method can add false-positive stop signs to the detector output [8]. To this end, a composite loss function is used such that it first creates a new object localization, followed by a targeted misclassification. A state-of-the-art detector identifies a nonexistent stop sign between 25% and 79% of the time when these patches are placed on a clean wall, where no actual stop sign exists, at 3 m from the camera.

Unlike attacks that focus on targets with no intraclass variety (i.e., only stop signs), target types with large intraclass variety (persons) are considered in [12]. The loss function to help hide people from detectors considers three factors, namely, a nonprintability score (how well the colors of a patch can be reproduced by a printer); the total variation of the image (favoring a patch with smooth color transitions); and the maximum objectiveness score in the image (i.e., the effectiveness on hiding a person), which aims to minimize the object or class score output by the detector. Operating in white-box settings and constraining the patch to lie in the neighborhood of the predicted bounding boxes of persons, the method is able to generate a class-universal adversarial patch for hiding persons. This adversarial example patch decreases the recall of the detector to 26.4% on a data set of images in which people were always detected in clean conditions.

Finally, a patch-based attack for object detectors can potentially hide all of the objects to be detected in a scene without overlapping with any of them [21]. The patch can be placed anywhere in the field of view of the camera and cause all objects to disappear, even if they are far from the patch itself. This white-box and data-agnostic method generates the patch by maximizing the loss for the original targets, given a set of physical transformations that are applied on the patch using the EoT framework. The patch can hide most of the objects within the field of view: the mean average precision drops from 40.9% to 7.2% when the patch is randomly placed. The patch also disables the detection of objects that are moving as long as the patch itself is static with respect to the camera. The patch is, to some extent, invariant to location, but it has weaker influence on objects that are farther away.

Overall, adversarial attacks with objects pose a very serious threat for AVs, in particular, for the common cases of attacking lidar and camera sensors that use DNN-based classifiers and detectors. These attacks can generate 3D adversarial objects that are placed in the physical world or even take the form of small 2D stickers (patches) that can be placed on any object. They have the ability to cause wrong classification or false detection, while they can also be used to hide all of the other objects that are around them (even pedestrians). However, despite their success, such attacks are usually applied in white-box settings and, generally, show low transferability to unknown (black-box) models.

Countermeasures

In this section, we discuss countermeasures against adversarial physical attacks. We also discuss general methods, such as adversarial training (AT), that are applied to mitigate the effects of on-signal attacks, and their adaptation to physical attacks is straightforward.

Ultrasonic and radar sensors

A method against spoofing and jamming attacks can use each ultrasonic sensor to separate spoofed echoes from real ones and report the real distance (resilient obstacle detection) [14]. Ultrasonic sensors transmit pings of the same signal throughout their lifetimes and search for only the first echo. Since there is no bond between a ping and its echoes, to detect attacks and possibly reject spoofed echoes, a physical signal authentication is achieved by shifting the signal parameters through a challenge-response scheme: first, by customizing the ping signal, and, then, by correlating the received echoes with the pings. In addition, a set of ultrasonic sensors can also collaboratively detect an attack using a single-transmitter, multiple-receiver sensor structure for resilient obstacle detection and attacker localization [14]. These approaches can be also adapted for radar sensors.

GPS

Detection mechanisms may defend against attacks on the location, time, and Doppler shift parameters [22]. For a given parameter, the receiver first collects data during periods of time that are in normal mode (training data). Then, based on the normal-mode data, the receiver predicts the future values of the parameter and compares the predicted values with the ones obtained from the GNSS. If the difference with the predicted values exceeds a selected threshold, the receiver deems itself to be under attack, and all PVT solutions are discarded.

Two more spoofing countermeasures are described in [15]. The first one focuses on the latency between the spoofing and original data bitstreams, where the receiver looks for a data bit sign change between consecutive accumulations at the coarse acquisition code-length interval. If a sign change is detected anywhere other than at an expected data bit boundary, the target receiver raises a flag. The second one focuses on detecting the vestigial authentic signal. The receiver copies the incoming digitized front-end data into a buffer used only for vestigial detection and selects one of the GPS signals being tracked, removes it from the data in the buffer, and performs acquisition for the same signal (the same pseudorandom noise identifier) on the buffered data. These steps are repeated for the same GPS signal, and the results are summed noncoherently until a probability of detection threshold is met. If a significant vestigial signal is present in the data, this technique will reveal it.

Finally, real-time GNSS spoofing detection can be achieved by processing beat carrier-phase measurements from an antenna pair [23]. Considering a spoofed and a nonspoofed signal model, the observed differential beat carrier phases are fit to either of the two models, and the estimates (along with their associated fit error costs) are used to develop a spoofing hypothesis test. Then, real-time spoofing detection can be achieved through a switched-antenna version of the two-antenna system that determines a single-differenced beat carrier phase. Such a system can use a radio-frequency (RF) switch between the two antennas, a single RF front end, and a single receiver channel per tracked signal.

Lidar

Countermeasures can be applied at the hardware or at the software level. At the hardware level, using multiple lidar sensors of different wavelengths, which collaboratively provide the scanned information, makes it harder for the attacker to manipulate every signal at the same time [9]. At the software level, an option is to (non)predictably skip certain emitting pulses, which is similar to varying the scan speed. If the system notices a response that corresponds to a skipped pulse, then this is an indication of a possible attack. Another option is to shorten the pulse period, which consequently reduces the attack window. However, lowering the pulse period also shortens the range of the sensor.

Defenses for DNNs can operate at the AV system, sensor, or machine learning model levels [17]. At the AV level, by studying the spoofed 3D point clouds, it can be observed that points from ground reflection are clustered into obstacles due to the information loss introduced in the preprocessing phase. Mapping a 3D point cloud into a 2D matrix results in height information loss, which facilitates an attack. To mitigate this effect, the ground reflection can be filtered out in the preprocessing phase, or, to reduce the information loss, the point cloud should not be transformed into an input feature matrix. At the sensor level, similarly to [9], a possible defense consists of adding some randomness in the lidar pulses: by firing a random grouping of laser pulses at each cycle, an attacker would not know which reflections the sensor would be expecting. Finally, at the machine learning model level, an option is to perform AT that, as we will discuss later, improves the robustness of DNN-based systems against adversarial attacks.

Camera

Increasing the robustness of DNN-based vision systems against APs is a very challenging and widely studied problem. We focus here on the tasks of image classification and object detection.

Most defenses are designed for specific or weak attacks (and, thus, stronger attacks can bypass them [24]) or just obscure the model (gradient masking) rather than making the classifier robust against all attacks [25]. A defense that is empirically robust against all designed attacks is AT. AT aims to train robust models rather than defend against specific attacks: instead the model being trained only with the original (clean) data, it is trained along with the adversarial examples. AT has not been studied for physical APs yet, and we, therefore, provide here the results for on-signal attacks. Adaptation to physical APs is straightforward.

Formally, for a signal $x \in \Phi \sim \mathcal{X}$, we define a set of allowed perturbations $\mathcal{S}_v \subseteq \Phi$ that formalizes the manipulative power of the attacker; that is, as described in the "Robust Sensing" section, the distortion caused by g_v shall not considerably change or destroy the information of the signal. Assuming data targets $y \in \mathcal{Y}$, a loss function $\mathcal{L}$, and model parameters $\theta \in R^p$, the goal of AT is to solve the following minimax problem [26]:

$$\min_{\theta} \mathbb{E}_{(x,y)\sim(\mathcal{X},\mathcal{Y})}[\max_{v \in \mathcal{S}_v} \mathcal{L}(g_v(x), y, \theta)], \tag{2}$$

which corresponds to learning a model using signals (adversarial examples) that maximize the adversarial loss of the model.

For the simple case of the MNIST image data set [34], AT is effective since the adversarial accuracy of the model (the accuracy of the model when given adversarial examples as inputs) remains close to 90%, without a significant impact on the accuracy with clean examples, which is almost 98%. For the slightly more complex CIFAR-10 image data set [35], though, the adversarial accuracy is almost 52%, while the clean accuracy drops to almost 80%. Robustness is, at the moment, at odds with accuracy [27], [28]: to increase robustness, one might have to sacrifice accuracy. In addition, the cost of computing adversarial examples for the whole training procedure is very high, especially for high-dimensional data sets, such as ImageNet [36].

However, as AT decreases the curvature of the decision boundaries of the classifier [29], a regularization term can be added to the training loss to impose smoothness. The minimization of this loss during training produces results comparable to AT on MNIST and CIFAR-10 but much faster. Moreover, a fast AT process [30] updates both the model parameters and image perturbations using one simultaneous backward pass, thus, with almost no additional cost to natural training, resulting in ImageNet models with around 44% adversarial and 65% clean accuracy.

For the task of object detection, the authors in [31] focus on the design principles of existing adversarial attacks; an attack to a detector can be achieved with variants of individual task losses or their combinations (localization and/or classification), and, in fact, adversarial attacks dedicated to one of the tasks can reduce the performance of the model on the other task. Exploiting these interactions between different task losses, they generalize AT from classification to detection to form a new minimax training scheme, which combines task-oriented adversarial examples that maximize the adversarial loss of the detector. This scheme improves by 20–30% both the adversarial classification accuracy and the adversarial object-detection accuracy, which is the ratio between the correctly detected objects and the actual total number of objects.

Discussion and outlook

In this section, we summarize the findings from the evaluation of the adversarial attacks and countermeasures, and identify four main research directions toward designing robust sensing for autonomous driving systems.

Evaluation

The evaluation of adversarial attacks has been limited to specific sensors or to machine learning models. When the decision process is based on machine learning models, many attacks are evaluated on a few models or on similar architectures and/or with models trained on the same or similar data sets. The corresponding evaluation is, therefore, limited, as an attack that works well on a given data set and architecture might not generalize.

In fact, increasing the variability of the data sets and architectures is important to evaluate the transferability of the attack, since the generated adversarial example might not transfer to models of different configurations. Also, most methods operate under white- or gray-box settings, which are unrealistic, as an attacker might not have (full) knowledge of the underlying sensory systems or the machine learning models. To satisfactorily evaluate the robustness of AVs to APs, it is desirable to design attacks in black-box settings that are invariant to physical transformations, such as rotations, changes in viewing angle, and illumination variations. Simulators provide a promising direction to test a wide variety of situations [32].

Fusion

While adversarial vulnerabilities of individual sensory systems have been thoroughly investigated, the effect on fused signals is still not studied. In adversarial settings, determining the modalities that contribute more to incorrect inferences could provide new insights on how the information provided by the nonadversarial sensors can support the systems and help them to remain robust (adversarial robustness through signal fusion). Moreover, from the perspective of the robustness of the fused signal, the consideration of multimodal adversarial methods (attacking the fusion itself) is an exciting research area that could open new directions in understanding vulnerabilities in adversarial settings. It is, therefore, desirable to study the effect of each individual sensor (under adversarial attacks) to the fusion-based decision-making process.

AT

AT has been mostly studied for image classification and object detection, especially for on-signal APs. AT is a general framework that can also be exploited to achieve robustness against any type of AP (e.g., patch based), including physical APs, by replacing on-signal adversarial examples with physical adversarial examples. Moreover, the principles of AT are applicable to any sensing modality when the underlying system is a machine learning model. However, current versions of AT schemes might saturate to some suboptimal solution, which could explain the increasing gap between adversarial and standard accuracy [27]. Finally, it is also important to focus on the dynamics of AT and explore how and why it improves adversarial robustness [29].

Robustness

When evaluating the robustness of deep models against adversarial attacks, it is not only important to ensure that the selected attacks are strong (to avoid overfitting to weak attacks [24]) but also to test the models using gradient-free black-box attacks [33] to ensure that the models are actually robust and that the gradients of the models are not just obfuscated (gradient masking), which gives a false sense of robustness [25]. Ultimately, a key question is to determine why APs exist. Their existence is expected to be due to a combination of causes related to the data, the architecture, and the training/learning scheme.

With AT, robustness is an effect of the modified data and learning scheme. When studying the architectures, increasing the capacity of the model seems to help AT to increase robustness [26]. Important open questions include the following. What would be the effect of other network elements (e.g., the size of the learned filters)? What are the interactions of all of these entities that cause (or can prevent) such adversarial vulnerabilities? In that sense, for building and deploying safer systems, instead of black-box ones, more transparent models that allow us to interpret and explain their functionalities and reasoning are certainly desirable. Understanding how these systems work might, indeed, shed light on the underlying causes of their adversarial vulnerabilities.

Acknowledgments

This work is supported by the CHIST-ERA program through the project Collaborative object recognition, shared manipulation and learning (CORSMAL), under U.K. Engineering and Physical Sciences Research Council grant EP/S031715/1 and Swiss National Science Foundation grant 20CH21_180444. Apostolos Modas and Ricardo Sanchez-Matilla contributed equally to this article.

Authors

Apostolos Modas (apostolos.modas@epfl.ch) is a Ph.D. candidate with the Signal Processing Laboratory 4 in Ecole Polytechnique Fédérale de Lausanne, Switzerland, under the supervision of Prof. Pascal Frossard. He received his B.Sc. and M.Sc. degrees in informatics and communications from National and Kapodistrian University of Athens, Greece, in 2017. His research interests include computer vision, deep learning, and explainable machine learning.

Ricardo Sanchez-Matilla (ricardo.sanchezmatilla@qmul.ac.uk) is a research assistant for the Collaborative object recognition, shared manipulation and learning project and a Ph.D. candidate with the Centre for Intelligent Sensing at Queen Mary University of London. He received his B.Sc. and M.Sc. degrees in telecommunication engineering from the Universidad Autonoma de Madrid, Spain, in 2015. His research interests include computer vision, deep learning, and privacy.

Pascal Frossard (pascal.frossard@epfl.ch) has been a faculty member at Ecole Polytechnique Fédérale de Lausanne (EPFL) since 2003, and he heads the Signal Processing Laboratory 4 in EPFL, Switzerland. His research interests include network data analysis, image representation and understanding, and machine learning. Between 2001 and 2003, he was a member of the research staff at the IBM T.J. Watson Research Center, Yorktown Heights, New York. He received the Swiss National Science Foundation Professorship Award in 2003, IBM Faculty Award in 2005, IBM Exploratory Stream Analytics Innovation Award in 2008, Google Faculty Award 2017, IEEE Transactions on Multimedia Best Paper Award in 2011, and IEEE Signal Processing Magazine Best Paper Award in 2016. He is a Fellow of the IEEE.

Andrea Cavallaro (a.cavallaro@qmul.ac.uk) is a professor of multimedia signal processing and the founding director of the Centre for Intelligent Sensing at Queen Mary University of London; a Turing Fellow at the Alan Turing Institute, the U.K. National Institute for Data Science and Artificial Intelligence; and a fellow of the International Association for Pattern Recognition. He is the editor-in-chief of *Signal Processing: Image Communication*; senior area editor for *IEEE Transactions on Image Processing*; chair of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee; and an IEEE Signal Processing Society distinguished lecturer. His research interests include robotic perception, camera networks, multimodal processing, and machine learning.

References

[1] A. Khan, B. Rinner, and A. Cavallaro, “Cooperative robots to observe moving targets: A review,” *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 187–198, 2018. doi: 10.1109/TCYB.2016.2628161.

[2] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno et al., “Robust physical-world attacks on deep learning visual classification,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, June 2018, pp. 1625–1634.

[3] A. Fawzi, S.-M. Moosavi-Dezfooli, and P. Frossard, “Robustness of classifiers: From adversarial to random noise,” in *Proc. Advances Neural Information Processing Systems*, Barcelona, Spain, Dec. 2016, pp. 1632–1640.

[4] A. Fawzi, S.-M. Moosavi-Dezfooli, and P. Frossard, “The robustness of deep networks: A geometrical perspective,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 50–62, 2017. doi: 10.1109/MSP.2017.2740965.

[5] X. Yuan, P. He, Q. Zhu, and X. Li, “Adversarial examples: Attacks and defenses for deep learning,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2805–2824, Sept. 2019. doi: 10.1109/TNNLS.2018.2886017.

[6] C. Yan, W. Xu, and J. Liu, “Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle,” in *Proc. DEF CON*, Las Vegas, NV, Aug. 2016, pp. 1–49.

[7] Y. Cao, C. Xiao, D. Yang, J. Fang, R. Yang, M. Liu, and B. Li, Adversarial objects against LiDAR-based autonomous driving systems. 2019. [Online]. Available: arXiv:1907.05418

[8] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, F. Tramèr, A. Prakash, T. Kohno et al., “Physical adversarial examples for object detectors,” in *Proc. 12th USENIX Workshop on Offensive Technologies*, Baltimore, MD, Aug. 2018.

[9] J. Petit, B. Stottelaar, M. Feiri, and F. Kargl, “Remote attacks on automated vehicles sensors: Experiments on camera and LiDAR,” in *Proc. Black Hat Europe*, Amsterdam, The Netherlands, Nov. 2015, vol. 5.

[10] C. Sitawarin, A. N. Bhagoji, A. Mosenia, M. Chiang, and P. Mittal, DARTS: Deceiving autonomous cars with toxic signs. 2018. [Online]. Available: arXiv:1802.06430

[11] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, “Universal adversarial perturbations,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Honolulu, HI, July 2017, pp. 86–94. doi: 10.1109/CVPR.2017.17.

[12] S. Thys, W. van Ranst, and T. Goedeme, Fooling automated surveillance cameras: Adversarial patches to attack person detection. 2019. [Online]. Available: arXiv:1904.08653

[13] P. Papadimitratos and A. Jovanovic, “Protection and fundamental vulnerability of GNSS,” in *Proc. IEEE Int. Workshop on Satellite and Space Communications*, Toulouse, France, Oct. 2008, pp. 167–171. doi: 10.1109/IWSSC.2008.4656777.

[14] W. Xu, C. Yan, W. Jia, X. Ji, and J. Liu, “Analyzing and enhancing the security of ultrasonic sensors for autonomous vehicles,” *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5015–5029, 2018. doi: 10.1109/JIOT.2018.2867917.

[15] T. Humphreys, B. Ledvina, M. Psiaki, B. O’Hanlon, and P. Kintner, “Assessing the spoofing threat: Development of a portable GPS civilian spoofer,” in *Proc. ION GNSS Conf.,* Savannah, GA, Sept. 2008, pp. 2314–2325.

[16] H. Shin, D. Kim, Y. Kwon, and Y. Kim, “Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications,” in *Cryptographic Hardware and Embedded Systems (Lecture Notes in Computer Science,* vol. 10529*)*, W. Fischer and N. Homma, Eds. Cham, Switzerland: Springer-Verlag, 2017, pp. 445–467.

[17] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu et al., Adversarial sensor attack on LiDAR-based perception in autonomous driving. 2019. [Online]. Available: arXiv:1907.06826

[18] A. Athalye, L. Engstrom, A. Ilyas, and K. Kwok, Synthesizing robust adversarial examples. 2017. [Online]. Available: arXiv:1707.07397

[19] S.-T. Chen, C. Cornelius, J. Martin, and D. H. Chau, ShapeShifter: Robust physical adversarial attack on faster R-CNN object detector. 2018. [Online]. Available: arXiv:1804.05810

[20] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, “Adversarial patch,” in *Proc. Machine Learning and Computer Security Workshop, Neural Information Processing Systems*, Long Beach, CA, Dec. 2017.

[21] M. Lee and Z. Kolter, On physical adversarial patches for object detection. 2019. [Online]. Available: arXiv:1906.11897

[22] P. Papadimitratos and A. Jovanovic, “GNSS-based positioning: Attacks and countermeasures,” in *Proc. IEEE Conf. Military Communications*, San Diego, CA, Nov. 2008, pp. 1–7. doi: 10.1109/MILCOM.2008.4753512.

[23] M. Psiaki, B. O’Hanlon, S. Powell, J. Bhatti, K. Wesson, T. Humphreys, and A. Schofield, “GNSS spoofing detection using two-antenna differential carrier phase,” in *Proc. ION GNSS Conf.*, Jan. 2014, vol. 4, pp. 2776–2800.

[24] N. Carlini and D. Wagner, “Adversarial examples are not easily detected: Bypassing ten detection methods,” in *Proc. ACM Workshop on Artificial Intelligence and Security*, TX, Nov. 2017, pp. 3–14. doi: 10.1145/3128572.3140444.

[25] A. Athalye, N. Carlini, and D. Wagner, Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. 2018. [Online]. Available: arXiv:1802.00420

[26] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” in *Proc. Int. Conf. Learning Representation*, Vancouver, Canada, Apr. 2018, pp. 1–23.

[27] D. Tsipras, S. Santurkar, L. Engstrom, A. Turner, and A. Madry, “Robustness may be at odds with accuracy,” in *Proc. Int. Conf. Learning Representation*, May 2019, pp. 1–23.

[28] A. Fawzi, O. Fawzi, and P. Frossard, “Analysis of classifiers’ robustness to adversarial perturbations,” *Mach. Learn.*, vol. 107, no. 3, pp. 481–508, Mar. 2018. doi: 10.1007/s10994-017-5663-3.

[29] S.-M. Moosavi-Dezfooli, A. Fawzi, J. Uesato, and P. Frossard, “Robustness via curvature regularization, and vice versa,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Long Beach, CA, June 2019, pp. 9078–9086.

[30] A. Shafahi, M. Najibi, A. Ghiasi, Z. Xu, J. Dickerson, C. Studer, L. S. Davis, G. Taylor et al., “Adversarial training for free!” in *Proc. Advances Neural Information Processing Systems*, Vancouver, Canada, Dec. 2019, pp. 3358–3369.

[31] H. Zhang and J. Wang, “Towards adversarially robust object detection,” in *Proc. IEEE Int. Conf. Computer Vision*, Seoul, South Korea, Oct. 2019, pp. 421–430.

[32] A. Boloor, X. He, C. D. Gill, Y. Vorobeychik, and X. Zhang, Simple physical adversarial examples against end-to-end autonomous driving models. 2019. [Online]. Available: arXiv:1903.05157

[33] J. Uesato, B. O’Donoghue, P. Kohli, and A. van den Oord, “Adversarial risk and the dangers of evaluating against weak attacks,” in *Proc. 35th Int. Conf. Machine Learning*, Stockholm, Sweden, July 2018, pp. 5025–5034.

[34] Y. LeCun, C. Cortes, and C. J. C. Burges, “The MNIST database of handwritten digits.” Accessed on: May 1, 2020. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[35] A. Krizhevsky, V. Nair, and G. Hinton, “The CIFAR-10 dataset.” Accessed on: May 1, 2020. [Online]. Available: https://www.cs.toronto.edu/~kriz/cifar.html

[36] Stanford Vision Lab, “ImageNet,” 2016. Accessed on: May 1, 2020. [Online]. Available: http://www.image-net.org/

Christoph Stöckle, Stephan Herrmann,
Tobias Dirndorfer, and Wolfgang Utschick

Automated Vehicular Safety Systems

Robust function and sensor design

©ISTOCKPHOTO.COM/OONAL

Automated vehicular safety functions that intervene in dangerous driving situations, e.g., by emergency braking, use sensor measurements for the interpretation of the driving situation. As a consequence, they are typically very vulnerable to sensor imperfections, and unavoidable measurement errors have a negative impact on both the safety and the satisfaction of the customer, which has to be taken into account when designing automated vehicular safety systems, such as an automatic emergency braking (AEB) system, for example.

The first approach for a general design methodology with which both sensors and functions for a variety of automated vehicular safety systems can be designed, while taking both the unavoidable sensor measurement errors and the customer satisfaction into account, tries to benefit from integrated circuit design. This article reviews a well-established worst-case design approach for integrated circuits and explains how ideas from integrated circuit design can be transferred to the design of automated vehicular safety systems. It also discusses how the resulting methodology allows the systematic design of both functions and sensors such that the customer requirements are fulfilled in a robust manner despite unavoidable sensor measurement errors.

Introduction

Although much research is still required to make the vision of a fully autonomous car come true, several advanced driver assistance systems (ADASs) that perform, at least temporarily, single aspects of the driving task are already available. Automated vehicular safety functions that intervene in dangerous driving situations the driver is not able to control form an important class of those ADASs. As depicted in Figure 1, they use the measurements of sensors perceiving the environment of the vehicle to interpret the driving situation and trigger appropriate actions in dangerous driving situations, e.g., an emergency brake intervention. Consequently, they are, typically, very vulnerable to sensor imperfections, and unavoidable measurement errors have a negative impact on both the safety and

---

Digital Object Identifier 10.1109/MSP.2020.2984788
Date of current version: 26 June 2020

the satisfaction of the customer. Therefore, these unavoidable sensor measurement errors have to be taken into account.

The following works deal with the uncertainty vehicular safety systems are exposed to as caused by sensor measurement errors and errors in the prediction of the unknown future evolution of a situation, and they represent the state of the art. For calculating the collision probability under measurement and prediction uncertainty, [1]–[6] present Monte Carlo-based methods, whereas [7] replaces the expensive Monte Carlo sampling by the unscented transformation (UT), and [8] and [9] provide analytical solutions.

Collision-detection algorithms using the collision probability computed by Monte Carlo-, UT-, and machine learning-based methods are compared in [10]. Hruschka et al. [11] present a criticality measure that can be used for trajectory planning, including collision avoidance in the presence of measurement and prediction errors; in [12], a method for motion planning in automated driving under measurement and prediction uncertainty is discussed, while [13]–[15] discuss collision avoidance systems whose decision for an intervention explicitly takes the measurement and prediction uncertainty into account. Methods for identifying the optimal emergency maneuver considering the pedestrian's injury risk, positioning errors, and uncertain future movements are described in [16] and [17]. On the other hand, [18] tackles the problem of predicting the location uncertainty resulting from the sensor measurement errors in the form of covariance matrices based on a neural network.

Based on Monte Carlo simulations, [19] investigates the effect of sensor measurement errors on the uncertainty of collision-warning criteria used in collision-warning systems, [20] examines their impact on the performance of a situation assessment algorithm for a collision-prevention assistant, and [21] analyzes their influence on the accuracy of predicted collision parameters, such as the time to collision (TTC) used in predictive passive safety systems. The authors of [7] also use the UT for computing the probability distribution of the TTC under measurement uncertainty in [22].

Closed-form expressions for the probability distributions of criticality measures used in ADASs, such as the TTC, which are subject to prediction and measurement uncertainty, are derived in [23]. The framework presented there is applied in [24] to analyze the impact of prediction uncertainty and sensor measurement errors on the performance of AEB systems. While closed-form expressions for the worst-case performance of a collision avoidance system in the presence of prediction and measurement errors are derived in [25], [26] presents a statistical analysis of the vehicle motion estimation from features in the environment under the influence of sensor measurement errors using closed-form expressions, and [27] obtains analytical results for the relationship between sensor measurement errors and the resulting errors in the pose of the vehicle determined by odometry.

A methodology for setting the parameters of rear-end collision-warning and avoidance algorithms based on statistical performance metrics that allow the measurement and prediction uncertainty to be taken into account is proposed by the authors of [28]. In [29], a measure for the robustness of decision functions used in active safety systems for deciding on interventions to sensor measurement errors is introduced, which can also help to derive requirements for the errors and tune the decision functions, i.e., adjust the values of their parameters. In [30], Stellet et al. use analytic statistical modeling to investigate object detection with stereo vision for collision warning, which makes it possible to predict the uncertainty in object location and obtain optimal thresholds in the presence of measurement errors. In [31] and [32], sensor parameters are derived from requirements for the vehicle localization accuracy based on a probabilistic model taking sensor measurement errors into account.

However, there is still no general methodology with which both sensors and functions for a variety of automated vehicular safety systems can be designed while taking both the unavoidable sensor measurement errors and the customer satisfaction into account. In other areas of engineering science, e.g., civil and mechanical engineering as well as integrated circuit design, general design procedures were introduced early on to deal with unavoidable parameter tolerances and deficiencies and their influence on the system performance [33]–[39]. In contrast, the first approach for such a design methodology in the area of automated vehicular safety systems, is proposed in [40] and [41] and is based on the well-established worst-case design approach for integrated circuits presented in [35]–[39]. Although applied to an AEB system, it can be generalized easily to design other automated vehicular safety systems as well.

This article starts by reviewing the design approach for integrated circuits in [35]–[39] and explains, in a tutorial-style way, using a simple AEB system as an illustrative example, how ideas from integrated circuit design, a completely different application area, can be transferred to the design of automated vehicular safety systems, which marks a paradigm shift. It also describes how the resulting methodology in [40] and [41] allows the systematic design of both functions and sensors such that the customer requirements are fulfilled in a robust

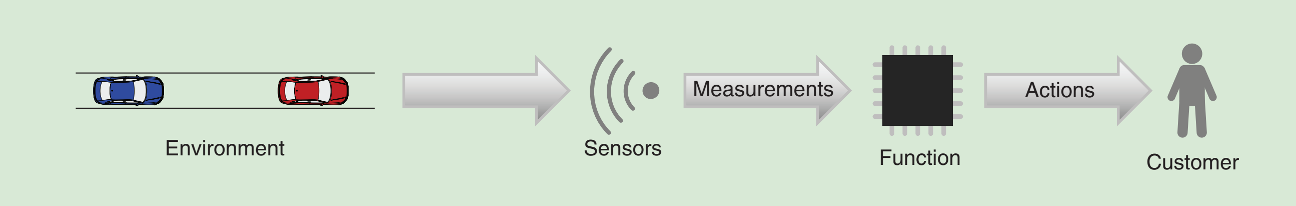


FIGURE 1. The general setting in which a vehicular safety function is embedded.

manner, despite unavoidable sensor measurement errors, by solving optimization problems. Furthermore, the article explains how the design task can be performed solely based on simulations of the automated vehicular safety system under design, which makes the design methodology easily applicable to various automated vehicular safety systems, and, finally, concludes with open research questions and challenges to be addressed by future work.

Integrated circuit design

Problem formulation for integrated circuit design

The well-established worst-case design approach for integrated circuits presented in [35]–[39] measures the performance of an integrated circuit in terms of performance properties $\boldsymbol{f}$, which are quantities of interest that have to lie in certain acceptance intervals such that the integrated circuit is considered to operate in a proper or acceptable way, e.g., the minimum output voltage. The performance properties $\boldsymbol{f}$ are dependent on the parameters $\boldsymbol{p}$ of the integrated circuit (e.g., the oxide thickness and length and width of the transistors) and, thus, are a function of them. For this function, there is usually no closed-form expression due to the high complexity of the integrated circuit, such that the evaluation of it requires a numerical simulation of the integrated circuit. It maps the given parameter values $\boldsymbol{p}$ to the respective values of the performance properties $\boldsymbol{f}$.

Due to unavoidable manufacturing tolerances, the values of some parameters $\boldsymbol{p}$ fluctuate. These parameters, such as the oxide thickness and perturbations of the length and width of transistors, are modeled by Gaussian random variables and are, thus, called *statistical parameters* $\boldsymbol{s}$, in contrast to the deterministic design parameters $\boldsymbol{d}$, which include the length and width of transistors. As a consequence, the performance properties $\boldsymbol{f}$ are random variables as well and may or may not lie in the acceptance intervals such that the performance specifications defined by the acceptance intervals might be fulfilled or violated. The probability that a manufactured integrated circuit fulfills the given performance specifications, i.e., the yield Y, can be used as the quality measure to be maximized for circuit optimization, i.e., for determining the optimal values for $\boldsymbol{d}$.

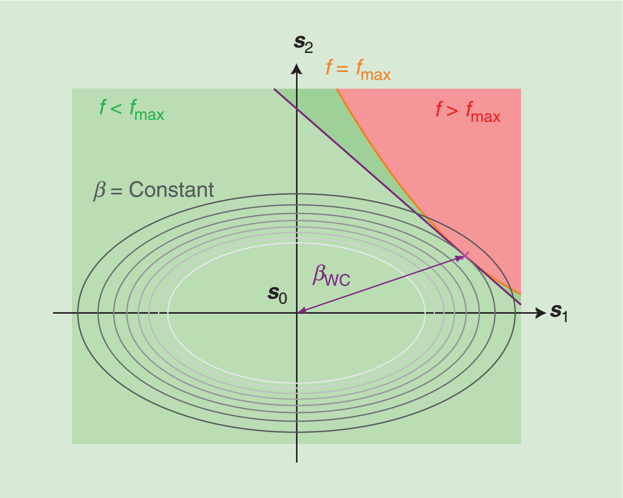


FIGURE 2. The acceptance region (green area including orange boundary) and worst-case distance β_{WC}.

Besides the classes of deterministic design parameters $\boldsymbol{d}$ and statistical parameters $\boldsymbol{s}$, there is a third class of parameters, namely, operating parameters $\boldsymbol{\theta}$, e.g., the temperature, which might have an influence on the performance of integrated circuits and make the yield Y a function of them. In contrast to the statistical parameters $\boldsymbol{s}$, the operating parameters $\boldsymbol{\theta}$ are ranging parameters for which no statistical knowledge in the form of a probability distribution is available, but only a range $T_{\boldsymbol{\theta}}$. To take this into account, the worst-case yield, i.e., the minimum yield $\min_{\boldsymbol{\theta} \in T_{\boldsymbol{\theta}}} Y$ in the whole range $T_{\boldsymbol{\theta}}$, can be used as the quality measure to be maximized for circuit optimization.

Simulation-based integrated circuit design

The maximization of the quality measure requires several evaluations of the yield Y. As there is usually no closed-form expression for the performance properties $\boldsymbol{f}$ as a function of the parameters $\boldsymbol{p}$, and it can be evaluated only by numerical simulations of the integrated circuit, this is also the case for the yield Y as a function of the deterministic design parameters $\boldsymbol{d}$ and operating parameters $\boldsymbol{\theta}$. A straightforward possibility for obtaining a value of the yield Y, at least approximately, is estimating it by a Monte Carlo simulation.

In the Monte Carlo simulation, M realizations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M$ of all involved random variables, i.e., the statistical parameters $\boldsymbol{s}$, are generated at random according to their probability distribution. Each realization $\boldsymbol{s}_i$ together with given deterministic design parameters $\boldsymbol{d}$ and operating parameters $\boldsymbol{\theta}$ is mapped to the respective values $\boldsymbol{f}_i$ of the performance properties by a circuit simulation. In these M random experiments, it is counted how often the values $\boldsymbol{f}_i$ of the performance properties lie in the acceptance intervals to obtain the frequency of fulfilling the performance specifications, which is an estimate for the probability of fulfilling the performance specifications, i.e., the yield Y.

To obtain an accurate estimate for this probability, a large number M of circuit simulations must be performed, which might lead to a prohibitively large computational complexity in practice. However, it is equivalent to the probability that the statistical parameters $\boldsymbol{s}$ lie in the acceptance region, i.e., the set of all $\boldsymbol{s}$ mapped to performance properties $\boldsymbol{f}$ lying in the acceptance intervals. An example of such an acceptance region in the case of one performance property f that is a function of only two statistical parameters s_1 and s_2, and has to lie in the acceptance interval $(-\infty, f_{\max}]$, is the green area in Figure 2, where $f < f_{\max}$, including the orange boundary, where $f = f_{\max}$.

Although the boundary is typically considered to be smooth, this does not have to be the case. The probability that the statistical parameters $\boldsymbol{s}$ lie in the acceptance region could theoretically be obtained by integrating their probability density function (pdf) in the acceptance region. Unfortunately, the boundary of the acceptance region can only be determined by many circuit simulations, and the integration of the pdf in the acceptance region with a possibly highly nonlinear boundary can only be performed numerically, which comes with a high computational complexity. Both problems can be overcome by so-called worst-case distances.

The worst-case distance β_{WC}, which is also illustrated in Figure 2, is defined as the smallest distance between the mean $\boldsymbol{s}_0$ of the statistical parameters $\boldsymbol{s}$ and the boundary of the acceptance region. The Mahalanobis distance is chosen as distance measure β. As a consequence, the set of all $\boldsymbol{s}$ that have the same distance $\beta =$ constant from their mean $\boldsymbol{s}_0$ is the set of all $\boldsymbol{s}$ for which their pdf has the same value. In Figure 2, these sets are the ellipses with the center $\boldsymbol{s}_0$. The points representing the statistical parameter values s_1 and s_2 on the largest of the shown ellipses, which intersects the boundary of the acceptance region in two points, have the largest distance β from their mean $\boldsymbol{s}_0$. If the ellipse is decreased such that the resulting smaller ellipse still intersects the boundary in two points, these two points have a smaller distance β from $\boldsymbol{s}_0$. If this procedure is continued, the two points where the decreasing ellipse and the boundary intersect converge to one point, where they touch. The magenta cross represents this point, which is the point on the boundary with the smallest distance β from $\boldsymbol{s}_0$, the worst-case distance β_{WC}.

The boundary of the acceptance region can be approximated by linearizing it with a tangential hyperplane that touches it where the statistical parameters $\boldsymbol{s}$ have the smallest distance from their mean $\boldsymbol{s}_0$, namely, the worst-case distance, and the probability that the statistical parameters $\boldsymbol{s}$ lie in the acceptance region by the probability that they lie in the resulting approximate acceptance region bounded by the tangential hyperplane. In Figure 2, the tangential hyperplane is the violet tangential line, and the resulting approximate acceptance region is the light green area bounded by this line.

The peak of the multivariate Gaussian pdf of the statistical parameters $\boldsymbol{s}$ at their mean $\boldsymbol{s}_0$ has to lie in the acceptance region, and it must be rather concentrated around its peak to be able to achieve a large probability that the statistical parameters $\boldsymbol{s}$ lie in the acceptance region, and, thus, the performance specifications are fulfilled. In addition, its maximum along the boundary of the acceptance region occurs exactly where the hyperplane touches it, and, therefore, the approximation error between them is zero, as at all other possible points where they touch or intersect. Hence, the pdf is small where the approximation error between them is large, such that the approximation of the probability that the statistical parameters $\boldsymbol{s}$ lie in the acceptance region by the probability that they lie in the approximate acceptance region is accurate. The integration of the multivariate Gaussian pdf in this approximate acceptance region to obtain the approximate probability of fulfilling the performance specifications simplifies to one evaluation of the standard normal cumulative distribution function (cdf) at the worst-case distance.

This worst-case distance can be determined by an optimization minimizing the distance between the statistical parameters $\boldsymbol{s}$ on the boundary of the acceptance region and their mean $\boldsymbol{s}_0$ without determining the whole boundary. This optimization requires just a few circuit simulations, which are automatically chosen by an appropriate optimization method in a smart way, serving the achievement of the optimization goal, and, thus, replaces a computationally expensive Monte Carlo simulation, which simply chooses an extensive amount of simulations according to the underlying probability distribution in a brute-force way for estimating the probability of interest.

Robust design of automated vehicular safety systems

Mathematical model of automated vehicular safety systems

A general mathematical model of automated vehicular safety systems including sensor measurement errors is depicted in Figure 3. The state vector $\boldsymbol{x}[n]$ sufficiently describes the state of the dynamic system in the considered driving scenario, which consists of the ego vehicle equipped with the automated vehicular safety system under design and one or several standing or moving objects, such as other vehicles, at the time instant t_n. Figure 4 shows a relevant driving scenario for the design of a simple AEB system, where the ego vehicle approaches an object, namely, another vehicle driving in the same lane in front of it, with constant relative velocity. In this example, the state vector $\boldsymbol{x}[n]$ consists of the distance and the relative velocity between the ego vehicle and the object. In general, such a driving scenario can be characterized by some scenario parameters $\boldsymbol{x}_0$, e.g., the constant relative velocity v_0 between the ego vehicle and the object in the example.

As the state vector $\boldsymbol{x}[n]$ is not directly available and can only be reconstructed from measurements, sensors are used to take measurements with a sampling rate f_s. At the time instant t_n, the sensors deliver the measurements $\boldsymbol{y}[n]$ of the quantities observed by them based on the state vector $\boldsymbol{x}[n]$ under the influence of the measurement errors $\boldsymbol{\epsilon}[n]$ made by them. These unavoidable sensor measurement errors $\boldsymbol{\epsilon}[n]$ can be modeled as Gaussian random variables in certain scenarios [19].

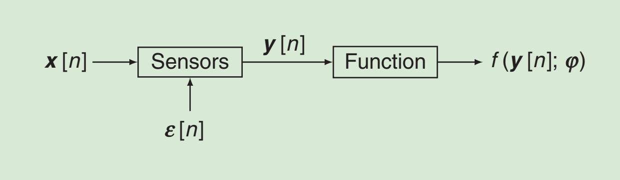


FIGURE 3. The general mathematical model of automated vehicular safety systems, including sensor measurement errors.

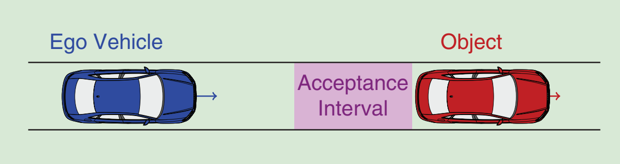


FIGURE 4. The relevant driving scenario for the design of a simple AEB system.

Even if the Gaussian assumption is not valid in a specific scenario, one can still proceed with Gaussian random variables, as they can be transformed such that the resulting random variables have the probability distribution of the actual sensor measurement errors, and this transformation can be considered to be part of the sensors. In case of the simple AEB system, the measurements $\boldsymbol{y}[n]$ are the measured distance and relative velocity, which are subject to additive independent and identically distributed zero-mean Gaussian measurement errors $\boldsymbol{\epsilon}[n]$. In such a mathematical model, the sensors can be described by parameters $\boldsymbol{\sigma}$, such as their sampling rate f_s and the standard deviations of their measurement errors $\boldsymbol{\epsilon}[n]$.

The function derives safety-relevant information from the sensor measurements $\boldsymbol{y}[n]$ at the time instant t_n to interpret the current driving situation and uses a decision rule to decide on triggering an appropriate action for mitigating a dangerous driving situation, e.g., an emergency brake intervention. Such a decision rule can be represented by a decision function $f(\cdot;\boldsymbol{\varphi})$, with adjustable parameters $\boldsymbol{\varphi}$, that maps the measurements $\boldsymbol{y}[n]$ to the function value $f(\boldsymbol{y}[n];\boldsymbol{\varphi})$, which can only be zero or one. As soon as the decision rule is fulfilled, i.e., $f(\boldsymbol{y}[n];\boldsymbol{\varphi})=1$, the respective action is triggered. In case of the simple AEB system, an example for such a decision rule $f(\cdot;\boldsymbol{\varphi})$ is to trigger the emergency brake intervention when the TTC computed from the measured distance and relative velocity for the interpretation of the current driving situation does not lie above a threshold τ anymore. The threshold τ of this TTC-based decision rule $f_{\mathrm{TTC}}(\cdot;\tau)$ is a typical example for the adjustable function parameters $\boldsymbol{\varphi}$.

Once an action is triggered, this intervention is performed automatically, as it is assumed that the driver cannot handle the detected dangerous driving situation alone. If an emergency brake intervention is triggered by the function of the simple AEB system, for example, the velocity of the ego vehicle is reduced with constant deceleration a until the relative velocity is zero.

Problem formulation for robust design of automated vehicular safety systems

In the development of automated vehicular safety systems, application engineers are typically confronted with the three basic design problems, illustrated in Figure 5. In the function design, the goal is to adapt the function to given sensors such that it meets the requirements of the customers in a robust manner despite the unavoidable sensor measurement errors. Similar to the integrated circuit design, where the optimal values for the deterministic design parameters $\boldsymbol{d}$ are determined by maximizing the quality measure, the best decision rule f_{opt} and the optimal function parameter values $\boldsymbol{\varphi}_{\mathrm{opt}}$ can be determined by maximizing a quality measure Q, which takes both the unavoidable sensor measurement errors and the customer satisfaction into account. Formally, this systematic approach for the function design can be formulated as the following optimization problem:

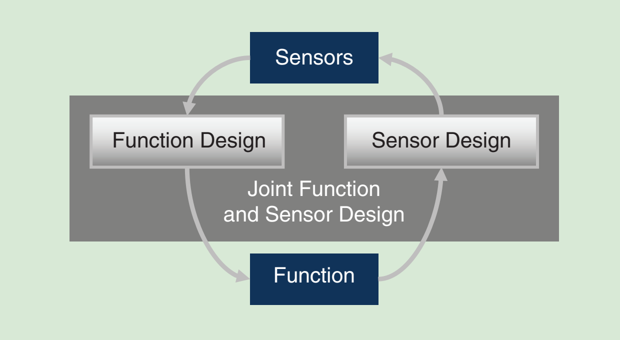


FIGURE 5. The basic design problems in the development of automated vehicular safety systems.

$$(f_{\mathrm{opt}},\boldsymbol{\varphi}_{\mathrm{opt}}) = \underset{f,\boldsymbol{\varphi}}{\operatorname{argmax}}\, Q. \tag{1}$$

Here, all predefined decision rules f are ranked with respect to the maximal quality level Q that they can achieve so that the best of them is chosen. At the top of Figure 6, an exemplary quality measure Q for the simple AEB system is plotted over the threshold τ of the TTC-based decision rule $f_{\mathrm{TTC}}(\cdot;\tau)$ for various standard deviations σ of the sensor measurement errors in the measured distance. The function is designed for given sensors, i.e., a given standard deviation σ, by maximizing the quality measure Q with respect to the threshold τ, the only adjustable parameter $\boldsymbol{\varphi}$ of the considered TTC-based decision rule $f_{\mathrm{TTC}}(\cdot;\tau)$. This maximization of the quality measure Q results in the optimal threshold τ_{opt}, which is highlighted in Figure 6(a) for the standard deviation $\sigma = 0.3$ m as an example.

In the sensor design, the goal is to determine which requirements the sensors have to fulfill such that a given function meets the specifications of the customers in a robust manner despite the unavoidable sensor measurement errors and appropriate sensors fulfilling them. The desired extent of this robust fulfillment can be expressed in the form of a required minimum quality level $Q_{\min}$, and the set of all $\boldsymbol{\sigma}$ for which the quality measure Q is not smaller than $Q_{\min}$ is the design space $\mathcal{S}$ from which the application engineer has to choose the sensor parameter values $\boldsymbol{\sigma}$ and that, thus, represents the sensor requirements.

An example of such a design space $\mathcal{S}$ can be seen in Figure 6(b), where the exemplary quality measure Q for the simple AEB system is plotted over the standard deviation σ of the sensor measurement errors in the measured distance for various thresholds τ of the TTC-based decision rule $f_{\mathrm{TTC}}(\cdot;\tau)$. This design space $\mathcal{S}$ is the interval of the standard deviation σ, the only considered sensor parameter $\boldsymbol{\sigma}$, where the quality measure Q is not smaller than the required minimum quality level $Q_{\min} = 0.95$ for the function given by the threshold $\tau = 0.54$ s.

Inside such a design space $\mathcal{S}$ costs C can be minimized to optimize a relevant objective, e.g., to minimize the costs of the sensors, and determine the optimal sensor parameter values $\boldsymbol{\sigma}_{\mathrm{opt}}$. Formally formulating this systematic approach for the sensor design leads to the following optimization problem, where the design space $\mathcal{S}$ is incorporated by the constraint:

$$\boldsymbol{\sigma}_{\text{opt}} = \underset{\boldsymbol{\sigma}}{\operatorname{argmin}}\, C \quad \text{s.t.} \quad Q \geq Q_{\min}. \tag{2}$$

In case of the simple AEB system with the standard deviation σ of the measurement errors in the measured distance as the only considered sensor parameter $\boldsymbol{\sigma}$, the costs C might be chosen to be $-\sigma$ such that they increase if the standard deviation σ is reduced. Alternatively, the costs C could also be chosen to be $1/\sigma$, which would lead to the same design results, as both cost functions C are minimal if and only if σ is maximal. This converts the minimization of C into the maximization of σ. The sensors are designed for a given function, i.e., the given TTC-based decision rule $f_{\text{TTC}}(\cdot;\tau)$ and a given threshold τ, by determining the largest σ inside the design space $\mathcal{S}$, the maximal tolerable standard deviation $\sigma_{\max}$ of the sensor measurement errors, which is highlighted in Figure 6(b) for $\tau = 0.54$ s as an example.

The goal of the joint function and sensor design is to determine both the sensors and the function such that the function meets the requirements of the customers in a robust manner despite the unavoidable sensor measurement errors. Formally, the joint function and sensor design can also be formulated as an optimization problem:

$$(\boldsymbol{\sigma}_{\text{opt}}, f_{\text{opt}}, \boldsymbol{\varphi}_{\text{opt}}) = \underset{\boldsymbol{\sigma}, f, \boldsymbol{\varphi}}{\operatorname{argmin}}\, C \quad \text{s.t.} \quad Q \geq Q_{\min}. \tag{3}$$

Solving this extension of the optimization problem for the sensor design determines the best decision rule f_{opt} by ranking all predefined decision rules f with respect to the minimal costs C that they allow for a given required minimum quality level $Q_{\min}$ along with the optimal function parameter values $\boldsymbol{\varphi}_{\text{opt}}$ in addition to the optimal sensor parameter values $\boldsymbol{\sigma}_{\text{opt}}$ such that the costs C are minimal while its constraint is fulfilled. The constraint again defines the design space $\mathcal{S}$, which is the set of all sensor parameter values $\boldsymbol{\sigma}$, predefined decision rules f, and function parameter values $\boldsymbol{\varphi}$ for which the quality measure Q is not smaller than $Q_{\min}$ and the customer requirements are fulfilled to the desired extent.

Besides the contour lines of the exemplary quality measure Q for the simple AEB system, whose only sensor parameter $\boldsymbol{\sigma}$ is the standard deviation σ and whose only function parameter $\boldsymbol{\varphi}$ is the threshold τ of the considered TTC-based decision rule $f_{\text{TTC}}(\cdot;\tau)$, an example of such a design space $\mathcal{S}$ is illustrated in Figure 6(c). It is the region where all parameter value pairs (σ,τ) correspond to a quality measure Q that does not lie below $Q_{\min} = 0.95$. If the costs C are chosen to be $-\sigma$ or $1/\sigma$ again, the minimization of C is converted to a maximization of σ. Hence, the optimal parameter value pair $(\sigma_{\max},\tau_{\text{opt}})$ highlighted by the green cross is the parameter value pair (σ,τ) inside $\mathcal{S}$ with the largest σ. This is the maximal tolerable standard deviation $\sigma_{\max}$ of the sensor measurement errors, and the corresponding τ is the optimal threshold τ_{opt} for the TTC-based decision rule $f_{\text{TTC}}(\cdot;\tau)$.

Quality measure for robust design of automated vehicular safety systems

In the integrated circuit design considering manufacturing tolerances, the yield is used as a quality measure. This can be translated to the robust design of automated vehicular safety systems considering sensor measurement errors to come up with an appropriate quality measure Q, which takes both the sensor measurement errors and the satisfaction of the customers into account and can be used in the formulated optimization problems (1)–(3).

When an automated vehicular safety system intervenes by triggering an action, relevant quantities $\boldsymbol{q}$ have to lie in certain acceptance intervals such that the customer is satisfied with the system. These acceptance intervals can be chosen in a user-specific way. The customer is satisfied with an AEB system as applied in the scenario from Figure 4, for example, if the final distance x_{end} between the ego vehicle and the object after an emergency brake intervention is neither too small nor too large. In other words, the final distance x_{end}, which is one example

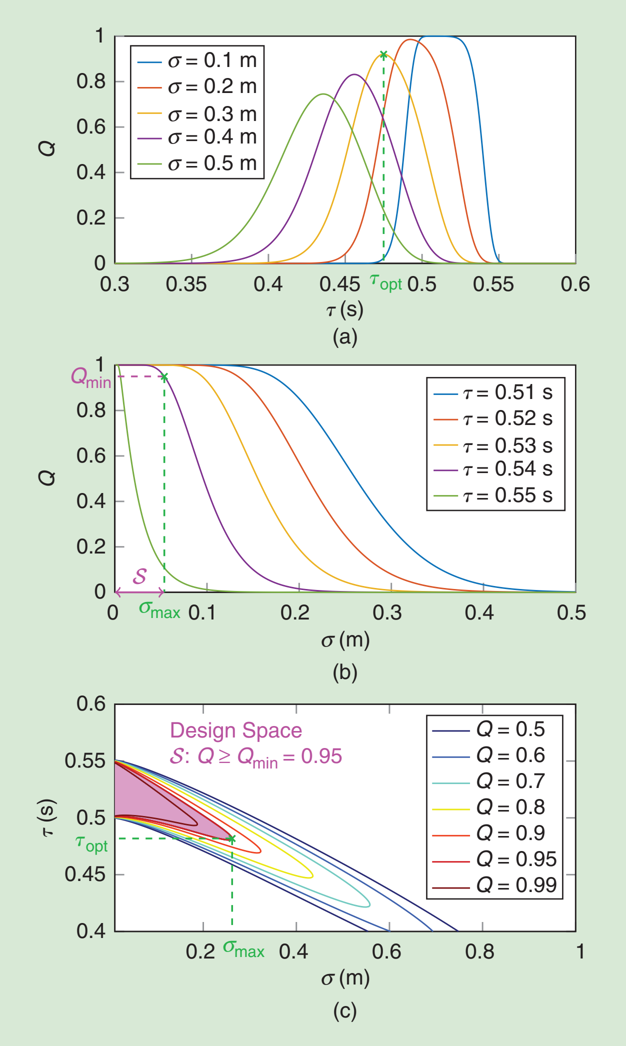


FIGURE 6. The (a) function, (b) sensor, and (c) joint function and sensor designs for a simple AEB system at a glance.

for such customer satisfaction properties $\boldsymbol{q}$, has to lie in an acceptance interval from the minimal tolerable final distance $x_{\min}$ to the maximal tolerable final distance $x_{\max}$ shown by the violet area.

Similarly to the performance properties $\boldsymbol{f}$ of integrated circuits, which are subject to manufacturing tolerances modeled by Gaussian random variables, namely, the statistical parameters $\boldsymbol{s}$, and have to lie in certain acceptance intervals, the customer satisfaction properties $\boldsymbol{q}$ of automated vehicular safety systems are subject to sensor measurement errors modeled by Gaussian random variables $\boldsymbol{\epsilon}[n]$, and the specifications for them in the form of acceptance intervals are conceptually the same. As a consequence, the customer satisfaction properties $\boldsymbol{q}$, which are a function of these random sensor measurement errors $\boldsymbol{\epsilon}[n]$, are random variables as well and might lie inside or outside the acceptance intervals such that the specifications defined by the acceptance intervals are fulfilled or violated.

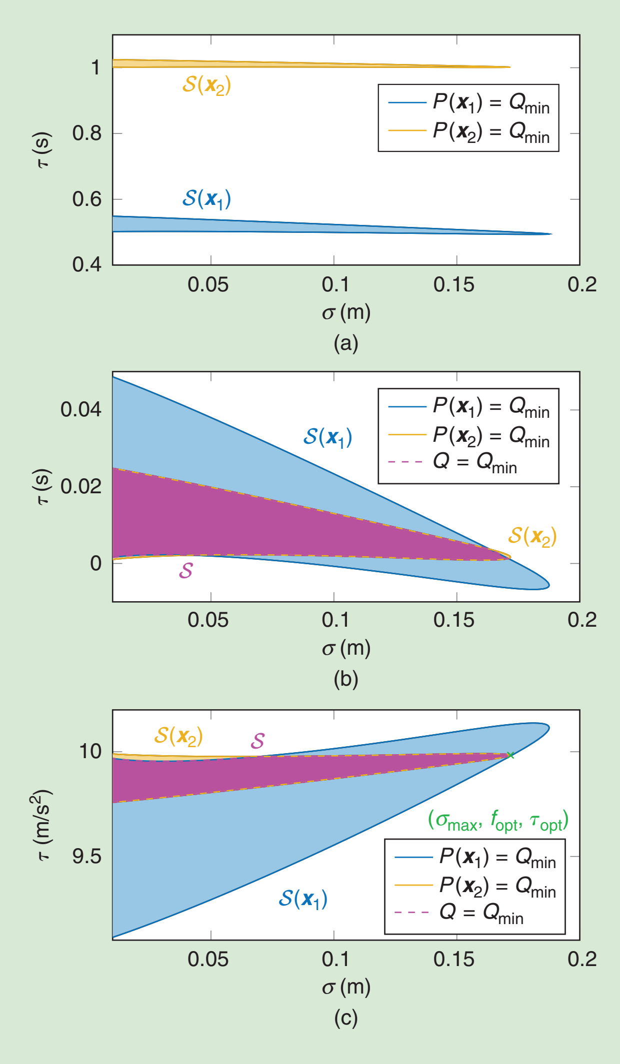


FIGURE 7. The joint function and sensor design for several scenarios and the required minimum quality level $Q_{\min} = 0.99$ in the case of a simple AEB system: (a) $f = f_{\text{TTC}}(\cdot;\tau)$, (b) $f = f_{\text{adv.TTC}}(\cdot;\tau)$, and (c) $f = f_{\text{BTN}}(\cdot;\tau)$.

Finally, the probability P of fulfilling these specifications for the customer satisfaction by an automated vehicular safety system corresponds to the probability of fulfilling the performance specifications by a manufactured integrated circuit, i.e., the yield Y. As the yield Y is used as a quality measure for deciding the optimal values for the deterministic design parameters $\boldsymbol{d}$ of an integrated circuit, the probability P of fulfilling the specifications for the customer satisfaction can be used as quality measure Q for determining the best decision rule f_{opt}, the optimal function parameter values $\boldsymbol{\varphi}_{\text{opt}}$, and the optimal sensor parameter values $\boldsymbol{\sigma}_{\text{opt}}$ of an automated vehicular safety system. The exemplary quality measure Q for the simple AEB system that has been plotted in Figure 6 for the TTC-based decision rule $f_{\text{TTC}}(\cdot;\tau)$ and the constant relative velocity $v_0 = -10$ m/s is the probability P that the respective specification for the customer satisfaction is fulfilled, i.e., the final distance x_{end} lies in the acceptance interval from the minimal tolerable final distance $x_{\min}$ to the maximal tolerable final distance $x_{\max}$.

Which values the customer satisfaction properties $\boldsymbol{q}$ have and whether they fulfill the specifications depends not only on the sensor parameters $\boldsymbol{\sigma}$, the sensor measurement errors $\boldsymbol{\epsilon}[n]$, the decision rule $f(\cdot;\boldsymbol{\varphi})$, and the function parameters $\boldsymbol{\varphi}$ but also on the scenario parameters $\boldsymbol{x}_0$ characterizing the driving scenario, such as the constant relative velocity v_0 in the scenario from Figure 4, in which the simple AEB system is applied. This makes the probability P of fulfilling the specifications for the customer satisfaction a function of the scenario parameters $\boldsymbol{x}_0$. An automated vehicular safety system is supposed to handle several driving scenarios described by the scenario parameters $\boldsymbol{x}_0$ from a set $\mathcal{X}$, which might be a whole range of such scenario parameters $\boldsymbol{x}_0$.

Conceptually, these scenario parameters $\boldsymbol{x}_0$ can be handled like the operating parameters $\boldsymbol{\theta}$ from the range $T_{\boldsymbol{\theta}}$ in the integrated circuit design. As the worst-case yield $\min_{\boldsymbol{\theta} \in T_{\boldsymbol{\theta}}} Y$, i.e., the minimum yield in the whole range $T_{\boldsymbol{\theta}}$ of the operating parameters $\boldsymbol{\theta}$, is used as a quality measure for the integrated circuit design, the worst-case probability $\min_{\boldsymbol{x}_0 \in \mathcal{X}} P$ of fulfilling the specifications for the customer satisfaction, i.e., the minimum of the probability P that these specifications are fulfilled in all scenarios $\boldsymbol{x}_0$ from the scenario set $\mathcal{X}$, can be used as quality measure Q. Consequently, the design space $\mathcal{S}$, in which this quality measure Q does not lie below the required minimum quality level $Q_{\min}$, is the intersection of the design spaces $\mathcal{S}(\boldsymbol{x}_0)$ for the individual scenarios $\boldsymbol{x}_0$ from $\mathcal{X}$, in each of which the probability $P(\boldsymbol{x}_0)$ of fulfilling the specifications for the customer satisfaction in the respective scenario $\boldsymbol{x}_0$ does not lie below $Q_{\min}$.

This is illustrated in Figure 7 for two scenarios $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, corresponding to the scenario from Figure 4 that the simple AEB system is supposed to handle after the joint function and sensor design with constant relative velocity $v_0 = -10$ m/s and $v_0 = -20$ m/s, respectively. Besides the TTC-based decision rule $f_{\text{TTC}}(\cdot;\tau)$, two further decision rules—namely, an advanced TTC-based decision rule $f_{\text{adv.TTC}}(\cdot;\tau)$ taking the critical TTC at the latest time at which the emergency brake intervention must be triggered to avoid a collision into account and a decision rule $f_{\mathrm{BTN}}(\cdot;\tau)$ based on the brake-threat-number, i.e., the deceleration required for collision avoidance—are considered as decision rule f with a threshold τ as the only function parameter $\boldsymbol{\varphi}$. The blue curve is the contour line along which the probability $P(\boldsymbol{x}_1)$ of fulfilling the specification for the customer satisfaction in the scenario $\boldsymbol{x}_1$ is equal to the required minimum quality level $Q_{\min} = 0.99$ and the boundary of the design space $\mathcal{S}(\boldsymbol{x}_1)$ highlighted by the blue area, in which $P(\boldsymbol{x}_1)$ does not lie below $Q_{\min}$. Analogously, the yellow curve is the contour line along which the probability $P(\boldsymbol{x}_2)$ of fulfilling the specification for the customer satisfaction in the scenario $\boldsymbol{x}_2$ is equal to $Q_{\min}$ and the boundary of the design space $\mathcal{S}(\boldsymbol{x}_2)$ highlighted by the yellow area, in which $P(x_2)$ does not lie below $Q_{\min}$.

> **As in integrated circuit design, the Monte Carlo simulation can be used as an estimator for the probability *P* that the specifications for the customer satisfaction are fulfilled.**

The intersection of the individual design spaces $\mathcal{S}(\boldsymbol{x}_1)$ and $\mathcal{S}(\boldsymbol{x}_2)$ yields the overall design space $\mathcal{S}$ highlighted by the magenta area, in which the quality measure Q, i.e., the minimum of the probabilities $P(\boldsymbol{x}_1)$ and $P(\boldsymbol{x}_2)$ of fulfilling the specification for the customer satisfaction in the two scenarios $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, does not lie below $Q_{\min}$. The solution $(\sigma_{\max}, f_{\mathrm{opt}}, \tau_{\mathrm{opt}})$ of the optimization problem (3) for the joint function and sensor design with the costs $C = -\sigma$ or $C = 1/\sigma$ consisting of the maximal tolerable standard deviation $\sigma_{\max}$, the optimal decision rule $f_{\mathrm{opt}} = f_{\mathrm{BTN}}(\cdot;\tau)$, and the optimal threshold τ_{opt} is also highlighted in Figure 7.

The constraint of the optimization problems (2) and (3) in the robust sensor as well as joint function and sensor design guarantees that the designed automated vehicular safety system fulfills the specifications for the customer satisfaction at least with the probability $Q_{\min}$ in all considered scenarios, which determines the reliability and safety of the system and, thus, the customer satisfaction. Increasing the required minimum quality level $Q_{\min}$ increases the reliability, safety, and customer satisfaction but might also increase the costs C. Therefore, a tradeoff has to be found by choosing $Q_{\min}$ appropriately. This is a decision that must be made by the management, which can be based on user studies or on the desired probability of fulfilling the specifications in system assessment tests like those of the New Car Assessment Program.

Simulation-based robust design of automated vehicular safety systems

The customer satisfaction properties $\boldsymbol{q}$ are dependent on the sensor parameters $\boldsymbol{\sigma}$, the sensor measurement errors $\boldsymbol{\epsilon}[n]$, the decision rule $f(\cdot;\boldsymbol{\varphi})$, the function parameters $\boldsymbol{\varphi}$, and the scenario parameters $\boldsymbol{x}_0$ and, thus, are a function of them. Apart from simple automated vehicular safety systems like the simple AEB system, there is usually no closed-form expression for this function due to the high complexity of automated vehicular safety systems in practice. As a consequence, the evaluation of this function requires a numerical simulation of the automated vehicular safety system.

This is the reason why there is usually no closed-form expression for the probability P that the customer satisfaction properties $\boldsymbol{q}$ lie in the acceptance intervals and, thus, fulfill the specifications for the customer satisfaction as a function of the sensor parameters $\boldsymbol{\sigma}$, the decision rule $f(\cdot;\boldsymbol{\varphi})$, the function parameters $\boldsymbol{\varphi}$, and the scenario parameters $\boldsymbol{x}_0$, which has to be evaluated several times for solving the optimization problems (1)–(3) of the robust function and sensor design. Only for simple automated vehicular safety systems can such closed-form expressions be derived as in [40] and [41] for the simple AEB system applied in the idealized scenario from Figure 4, but not in general.

As in integrated circuit design, the Monte Carlo simulation can be used as an estimator for the probability P that the specifications for the customer satisfaction are fulfilled. In the Monte Carlo simulation, M realizations $\boldsymbol{\epsilon}_1[n], \boldsymbol{\epsilon}_2[n], \ldots, \boldsymbol{\epsilon}_M[n]$ of all involved random variables, i.e., the sensor measurement errors $\boldsymbol{\epsilon}[n]$ at all time instants t_n, are generated at random according to their probability distribution. Each realization $\boldsymbol{\epsilon}_i[n]$ of the sensor measurement errors $\boldsymbol{\epsilon}[n]$ at all time instants t_n, together with given sensor parameters $\boldsymbol{\sigma}$, a given decision rule $f(\cdot;\boldsymbol{\varphi})$, given function parameters $\boldsymbol{\varphi}$, and scenario parameters $\boldsymbol{x}_0$, is mapped to the respective values $\boldsymbol{q}_i$ of the customer satisfaction properties by a simulation of the automated vehicular safety system. In these M random experiments, it is counted how often the values $\boldsymbol{q}_i$ of the customer satisfaction properties lie in the acceptance intervals to obtain the frequency of fulfilling the specifications for the customer satisfaction, which is an estimate for the probability P of fulfilling the specifications for the customer satisfaction.

Challenges and future work

To obtain an accurate Monte Carlo-based estimate for the probability P of fulfilling the specifications for the customer satisfaction, a large number M of simulations of the automated vehicular safety system has to be performed, which might lead to prohibitively large computational complexity in practice. Reducing the computational complexity and guaranteeing a high accuracy of the obtained approximate value for the probability P are difficult tasks. These challenges have to be addressed by future work.

A promising future research direction is to learn from the integrated circuit design and consider acceptance regions, as illustrated by the green area including the orange boundary in Figure 2, for two statistical parameters s_1 and s_2 of an integrated circuit corresponding to two sensor measurement errors $\epsilon_1[n_1]$ and $\epsilon_2[n_2]$ at time instants t_{n_1} and t_{n_2} in the case of an automated vehicular safety system. In this acceptance region, a performance property f of the integrated circuit or a customer satisfaction property q of the automated vehicular safety system lies in the acceptance interval $(-\infty, f_{\max}]$. In general, the acceptance region is the set of all sensor measurement errors $\boldsymbol{\epsilon}[n]$ at all time instants t_n corresponding to the statistical parameters $\boldsymbol{s}$ of integrated circuits that are mapped to the customer satisfaction properties $\boldsymbol{q}$ lying in the acceptance intervals such that the specifications for the customer satisfaction are fulfilled. Consequently, the probability that the sensor measurement errors $\boldsymbol{\epsilon}[n]$ lie in the acceptance region is equivalent to the probability P that the automated vehicular safety system fulfills the specifications for the customer satisfaction.

Hence, this probability P could theoretically be computed by integrating the pdf of the sensor measurement errors $\boldsymbol{\epsilon}[n]$ in the acceptance region. Unfortunately, this integration is intractable in practice due to the high computational complexity of numerically determining the boundary of the acceptance region by many simulations of the automated vehicular safety system and numerically integrating the multivariate Gaussian pdf of the sensor measurement errors $\boldsymbol{\epsilon}[n]$ in the acceptance region, especially for a large number of sensor measurement errors $\boldsymbol{\epsilon}[n]$, at all time instants t_n resulting from a high sampling rate f_s of the sensors. As in the integrated circuit design, this problem can be overcome by approximating the boundary of the acceptance and integration regions by just a few simulations of the automated vehicular safety system such that the integration of the multivariate Gaussian pdf in the resulting simplified integration region becomes easy while the approximation error is acceptable.

To this end, the boundary of the acceptance region is linearized by using worst-case distances introduced in the integrated circuit design, as illustrated by the violet line in Figure 2, for two statistical parameters s_1 and s_2 corresponding to two sensor measurement errors $\epsilon_1[n_1]$ and $\epsilon_2[n_2]$. This leads to an approximate acceptance region as highlighted by the light green area in Figure 2. The integration of the multivariate Gaussian pdf in such approximate acceptance regions to obtain an approximate probability of fulfilling the specifications for the customer satisfaction simplifies to evaluating the standard Gaussian cdf at worst-case distances. Thus, most of the effort lies in the determination of worst-case distances by minimizing the distance of the sensor measurement errors $\boldsymbol{\epsilon}[n]$ on the boundary of acceptance regions from their mean, which, however, might require significantly fewer simulations of automated vehicular safety systems under design than using expensive Monte Carlo simulations. For the simple AEB system applied in the scenario from Figure 4, [42] demonstrates that using worst-case distances can reduce the required number of simulations of the AEB system under design by several orders of magnitude as compared to using Monte Carlo simulations with a comparable approximation accuracy.

Conclusions

Reviewing a well-established worst-case design approach for integrated circuits has shown that it has some parallels to the robust design of automated vehicular safety systems considering unavoidable sensor measurement errors, for which no general design methodology exists so far, although it is of high importance. The explanation of the first approach for such a design methodology using a simple AEB system as an illustrative example has demonstrated how ideas from integrated circuit design, a completely different application area, can be transferred to the design of automated vehicular safety systems, which marks a paradigm shift, and how it allows the systematic design of both functions and sensors of automated vehicular safety systems such that the customer requirements are fulfilled in a robust manner despite unavoidable sensor measurement errors by solving optimization problems and providing design spaces.

The design task can be performed solely based on simulations of the automated vehicular safety system under design, which makes the design methodology easily applicable to various automated vehicular safety systems. One of the biggest challenges to be addressed by future work is making this first approach for such a general design methodology more efficient by reducing the required number of simulations of the automated vehicular safety system under design, for which insights gained in integrated circuit design can be utilized once again.

Authors

Christoph Stöckle (christoph.stoeckle@tum.de) received his B.Sc. and M.Sc. degrees in electrical engineering from Technische Universität München (TUM), Germany, in 2013 and 2015, respectively, where he is currently pursuing the Dr.-Ing. degree. From 2016 to 2020, he has been with the Professorship of Signal Processing Methods at TUM and, since 2020, with Audi AG, Ingolstadt, Germany. In 2015, he received the VDE Award for his master's thesis. His current research interests include signal processing for automated driving with a focus on the design of automated vehicular safety systems. He is a Member of the IEEE.

Stephan Herrmann (stephan.herrmann@audi.de) received his Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Technische Universität München (TUM), Germany, in 2013 and 2019, respectively. He was with the Professorship of Signal Processing Methods at TUM from 2013 to 2016, working on applications of machine learning and optimal control to automotive safety. Since 2016, he has been with Audi AG, Ingolstadt, Germany, working on crash sensing, advanced driver assistance systems, and highly automated driving. His current research interests include signal processing, situation interpretation, and trajectory planning.

Tobias Dirndorfer (tobias.dirndorfer@audi.de) received his doctoral degree in mechanical engineering in 2015 from Technische Universität München (TUM) with a dissertation on predictive passive safety systems in cooperation with Audi AG. Since 2012, he worked on the industrial implementation of anticipatory passive safety algorithms for Audi AG and was responsible for the active warning and braking functions and pre sense front for customers. Currently, he is responsible for the preliminary design and concept development of predictive safety functions at Audi AG and has diverse patents and publications in that field of research.

Wolfgang Utschick (utschick@tum.de) completed several years of industrial training programs before he received his doctoral degree in electrical engineering from Technische Universität München (TUM) in 1998 with a dissertation on machine learning. Since 2002, he has been a professor at TUM, where he is chairing a signal processing group. He teaches courses on signal processing, stochastic processes, optimization theory, and machine learning in the fields of various applications. Since 2011, he has been a regular guest professor at TUM Asia and the Singapore Institute of Technology, and, since 2017, he has been serving as the dean of the TUM Department of Electrical and Computer Engineering.

References

[1] A. Lambert, D. Gruyer, G. S. Pierre, and A. N. Ndjeng, “Collision probability assessment for speed control,” in *Proc. 11th Int. IEEE Conf. Intelligent Transportation Systems*, Oct. 2008, pp. 1043–1048. doi: 10.1109/ITSC.2008.4732692.

[2] J. Ward, G. Agamennoni, S. Worrall, and E. Nebot, “Vehicle collision probability calculation for general traffic scenarios under uncertainty,” in *IEEE Intelligent Vehicles Symp. Proc.*, June 2014, pp. 986–992. doi: 10.1109/IVS.2014.6856430.

[3] C. Braeuchle, J. Ruenz, F. Flehmig, W. Rosenstiel, and T. Kropf, “Situation analysis and decision making for active pedestrian protection using Bayesian networks,” in *TÜV SÜD, 6. Tagung Fahrerassistenz—Der Weg zum automatischen Fahren*, München, 2013, pp. 1–5.

[4] A. Houénou, P. Bonnifait, and V. Cherfaoui, “Risk assessment for collision avoidance systems,” in *Proc. 17th Int. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2014, pp. 386–391. doi: 10.1109/ITSC.2014.6957721.

[5] M. Schreier, V. Willert, and J. Adamy, “Bayesian, maneuver-based, long-term trajectory prediction and criticality assessment for driver assistance systems,” in *Proc. 17th Int. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2014, pp. 334–341. doi: 10.1109/ITSC.2014.6957713.

[6] D. Åsljung, M. Westlund, and J. Fredriksson, “A probabilistic framework for collision probability estimation and an analysis of the discretization precision,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2019, pp. 52–57. doi: 10.1109/IVS.2019.8813853.

[7] A. Berthelot, A. Tamke, T. Dang, and G. Breuel, “Handling uncertainties in criticality assessment,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2011, pp. 571–576. doi: 10.1109/IVS.2011.5940483.

[8] N. E. Du Toit and J. W. Burdick, “Probabilistic collision checking with chance constraints,” *IEEE Trans. Robot.*, vol. 27, no. 4, pp. 809–815, Aug. 2011. doi: 10.1109/TRO.2011.2116190.

[9] S. Patil, J. van den Berg, and R. Alterovitz, “Estimating probability of collision for safe motion planning under Gaussian motion and sensing uncertainty,” in *Proc. IEEE Int. Conf. Robotics and Automation*, May 2012, pp. 3238–3244. doi: 10.1109/ICRA.2012.6224727.

[10] M. Metro, J. Ghosh, and C. Bhat, “Optimal alarms for vehicular collision detection,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2017, pp. 277–282. doi: 10.1109/IVS.2017.7995732.

[11] C. M. Hruschka, D. Töpfer, and S. Zug, “Risk assessment for integral safety in automated driving,” in *Proc. 2nd Int. Conf. Intelligent Autonomous Systems (ICoIAS)*, Feb. 2019, pp. 102–109. doi: 10.1109/ICoIAS.2019.00025.

[12] Ö. S. Tas and C. Stiller, “Limited visibility and uncertainty aware motion planning for automated driving,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2018, pp. 1171–1178. doi: 10.1109/IVS.2018.8500369.

[13] R. Karlsson, J. Jansson, and F. Gustafsson, “Model-based statistical tracking and decision making for collision avoidance application,” in *Proc. American Control Conf.*, June 2004, vol. 4, pp. 3435–3440. doi: 10.23919/ACC.2004.1384441.

[14] M. Brännström, F. Sandblom, and L. Hammarstrand, “A probabilistic framework for decision-making in collision avoidance systems,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 637–648, June 2013. doi: 10.1109/TITS.2012.2227474.

[15] S. Lefèvre, R. Bajcsy, and C. Laugier, “Probabilistic decision making for collision avoidance systems: Postponing decisions,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Nov. 2013, pp. 4370–4375. doi: 10.1109/IROS.2013.6696983.

[16] C. Braeuchle, F. Flehmig, W. Rosenstiel, and T. Kropf, “Maneuver decision for active pedestrian protection under uncertainty,” in *Proc. 16th Int. IEEE Conf. Intelligent Transportation Systems (ITSC 2013)*, Oct. 2013, pp. 646–651. doi: 10.1109/ITSC.2013.6728304.

[17] P. Themann, J. Kotte, D. Raudszus, and L. Eckstein, “Impact of positioning uncertainty of vulnerable road users on risk minimization in collision avoidance systems,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2015, pp. 1201–1206. doi: 10.1109/IVS.2015.7225846.

[18] F. Zhang, C. M. Martinez, D. Clarke, D. Cao, and A. Knoll, “Neural network based uncertainty prediction for autonomous vehicle application,” *Front. Neurorobot.*, vol. 13, Art. no. 12, pp. 1–17, 2019. doi: 10.3389/fnbot.2019.00012.

[19] P. Zheng and M. McDonald, “The effect of sensor errors on the performance of collision warning systems,” in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, Oct. 2003, vol. 1, pp. 469–474. doi: 10.1109/ITSC.2003.1251998.

[20] J. Hillenbrand, K. Kroschel, and V. Schmid, “Situation assessment algorithm for a collision prevention assistant,” in *Proc. IEEE Intelligent Vehicles Symp.*, June 2005, pp. 459–465. doi: 10.1109/IVS.2005.1505146.

[21] T. Dirndorfer, M. Botsch, and A. Knoll, “Model-based analysis of sensor-noise in predictive passive safety algorithms,” in *Proc. 22nd Enhanced Safety Vehicles Conf.*, 2011, pp. 1–14.

[22] A. Berthelot, A. Tamke, T. Dang, and G. Breuel, “A novel approach for the probabilistic computation of time-to-collision,” in *Proc. IEEE Intelligent Vehicles Symp.*, June 2012, pp. 1173–1178. doi: 10.1109/IVS.2012.6232221.

[23] J. E. Stellet, J. Schumacher, W. Branz, and J. M. Zöllner, “Uncertainty propagation in criticality measures for driver assistance,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2015, pp. 1187–1194. doi: 10.1109/IVS.2015.7225844.

[24] J. E. Stellet, P. Vogt, J. Schumacher, W. Branz, and J. M. Zöllner, “Analytical derivation of performance bounds of autonomous emergency brake systems,” in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2016, pp. 220–226. doi: 10.1109/IVS.2016.7535389.

[25] J. Nilsson, A. C. E. Ödblom, and J. Fredriksson, “Worst-case analysis of automotive collision avoidance systems,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1899–1911, Apr. 2016. doi: 10.1109/TVT.2015.2419196.

[26] J. E. Stellet, C. Heigele, F. Kuhnt, J. M. Zöllner, and D. Schramm, “Performance evaluation and statistical analysis of algorithms for ego-motion estimation,” in *Proc. 17th Int. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2014, pp. 2125–2131. doi: 10.1109/ITSC.2014.6958017.

[27] A. Kelly, “Linearized error propagation in odometry,” *Int. J. Robotics Res.*, vol. 23, no. 2, pp. 179–218, 2004. doi: 10.1177/0278364904041326.

[28] L. Yang, J. H. Yang, E. Feron, and V. Kulkarni, “Development of a performance-based approach for a rear-end collision warning and avoidance system for automobiles,” in *Proc. IEEE Intelligent Vehicles Symp. (IV2003)*, June 2003, pp. 316–321. doi: 10.1109/IVS.2003.1212929.

[29] J. Nilsson and M. Ali, “Sensitivity analysis and tuning for active safety systems,” in *Proc. 13th Int. IEEE Conf. Intelligent Transportation Systems*, Sept. 2010, pp. 161–167. doi: 10.1109/ITSC.2010.5625103.

[30] J. E. Stellet, J. Schumacher, O. Lange, W. Branz, F. Niewels, and J. M. Zöllner, “Statistical modelling of object detection in stereo vision-based driver assistance,” in *Intelligent Autonomous Systems 13. Advances in Intelligent Systems and Computing*, E. Menegatti, N. Michael, K Berns, and H. Yamaguchi, Eds. Cham, Switzerland: Springer-Verlag, 2016, pp. 749–761.

[31] J. Rohde, J. E. Stellet, H. Mielenz, and J. M. Zöllner, “Model-based derivation of perception accuracy requirements for vehicle localization in urban environments,” in *Proc. IEEE 18th Int. Conf. Intelligent Transportation Systems*, Sept. 2015, pp. 712–718. doi: 10.1109/ITSC.2015.121.

[32] J. Rohde, J. E. Stellet, H. Mielenz, and J. M. Zöllner, “Localization accuracy estimation with application to perception design,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, May 2016, pp. 4777–4783. doi: 10.1109/ICRA.2016.7487681.

[33] A. M. Hasofer and N. C. Lind, “Exact and invariant second-moment code format,” *J. Eng. Mech. Div.*, vol. 100, no. 1, pp. 111–121, Feb. 1974.

[34] R. Rackwitz and B. Fiessler, “Structural reliability under combined random load sequences,” *Comput. Struct.*, vol. 9, no. 5, pp. 489–494, 1978. doi: 10.1016/0045-7949(78)90046-9.

[35] W. Utschick, “Ein sequentiell quadratisches Optimierverfahren zur Toleranzanalyse integrierter Schaltungen,” Diplomarbeit, TUM, Munich, 1993.

[36] H. E. Graeb, C. U. Wieser, and K. J. Antreich, “Improved methods for worst-case analysis and optimization incorporating operating tolerances,” in *Proc. 30th ACM/IEEE Design Automation Conf.*, June 1993, pp. 142–147. doi: 10.1145/157485.164641.

[37] K. J. Antreich, H. E. Graeb, and C. U. Wieser, “Practical methods for worst-case and yield analysis of analog integrated circuits,” *Int. J. High Speed Electron. Syst.*, vol. 4, no. 3, pp. 261–282, 1993. doi: 10.1142/S0129156493000121.

[38] K. J. Antreich, H. E. Graeb, and C. U. Wieser, “Circuit analysis and optimization driven by worst-case distances,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 1, pp. 57–71, Jan. 1994. doi: 10.1109/43.273749.

[39] H. E. Graeb, *Analog Design Centering and Sizing.* Dordrecht, The Netherlands: Springer-Verlag, 2007.

[40] C. Stöckle, W. Utschick, S. Herrmann, and T. Dirndorfer, “Robust design of an automatic emergency braking system considering sensor measurement errors,” in *Proc. 21st IEEE Int. Conf. Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 2018–2023. doi: 10.1109/ITSC.2018.8569495.

[41] C. Stöckle, W. Utschick, S. Herrmann, and T. Dirndorfer, “Robust function and sensor design considering sensor measurement errors applied to automatic emergency braking,” in *Proc. 30th IEEE Intelligent Vehicles Symp. (IV)*, June 2019, pp. 2284–2290. doi: 10.1109/IVS.2019.8814142.

[42] M. L. Leyrer, C. Stöckle, S. Herrmann, T. Dirndorfer, and W. Utschick, “An efficient approach to simulation-based robust function and sensor design applied to an automatic emergency braking system,” to be published.

Guang Chen, Hu Cao, Jörg Conradt, Huajin Tang,
Florian Röhrbein, and Alois Knoll

Event-Based Neuromorphic Vision for Autonomous Driving

A paradigm shift for bio-inspired visual sensing and perception

©ISTOCKPHOTO.COM/OONAL

As a bio-inspired and emerging sensor, an event-based neuromorphic vision sensor has a different working principle compared to the standard frame-based cameras, which leads to promising properties of low energy consumption, low latency, high dynamic range (HDR), and high temporal resolution. It poses a paradigm shift to sense and perceive the environment by capturing local pixel-level light intensity changes and producing asynchronous event streams. Advanced technologies for the visual sensing system of autonomous vehicles from standard computer vision to event-based neuromorphic vision have been developed. In this tutorial-like article, a comprehensive review of the emerging technology is given. First, the course of the development of the neuromorphic vision sensor that is derived from the understanding of biological retina is introduced. The signal processing techniques for event noise processing and event data representation are then discussed. Next, the signal processing algorithms and applications for event-based neuromorphic vision in autonomous driving and various assistance systems are reviewed. Finally, challenges and future research directions are pointed out. It is expected that this article will serve as a starting point for new researchers and engineers in the autonomous driving field and provide a bird's-eye view to both neuromorphic vision and autonomous driving research communities.

Introduction

Over the past few decades, the rapid development of electronics, information technologies, and artificial intelligence have made great progress in artificial visual sensing and perception systems. For example, the vision system of an autonomous vehicle becomes more intelligent by using deep learning technology. However, it still has some shortcomings compared with biological counterparts, such as the human and animal visual systems. Even small insects, such as bees, outperform the most advanced artificial vision systems such as high-quality cameras in routine functions, including real-time sensing and processing, low-latency motion control, and so on. More importantly, such biological neural systems can well perform

---

*Digital Object Identifier 10.1109/MSP.2020.2985815
Date of current version: 26 June 2020*

tasks with small energy consumption. In fact, biological neural systems usually consist of a large number of relatively simple elements. They operate in a massively parallel principle, which is different from the most common type of vision sensors such as CMOS cameras. Thus, some researchers and engineers have tried to mimic the working principles of the biological visual systems and come up with a new artificial visual system.

Recently, the developments of material technologies, lithographic processes, very large-scale integration (VLSI) design techniques, neuroscience, and neuromorphic technologies have enabled the novel conception and fabrication of bio-inspired visual sensors and processors. These new sensors and processors provide different methods to sense and perceive the world. The event-based neuromorphic vision sensor is such a bio-inspired vision sensor mimicking biological retina from both the system level and element level; it poses a paradigm shift in the way of visual information acquisition, processing, and modeling. The dynamic vision sensor (DVS) proposed by the group of Tobi Delbruck [1] is the first practicable event-based neuromorphic vision sensor based on the biological principle. DVS captures the per-pixel brightness changes (called *events*) asynchronously instead of measuring the absolute brightness of all pixels at constant rate, resulting in promising properties compared to standard frame-based cameras, such as low power consumption and low latency (in the order of microsecond), HDR (120 dB), and high temporal resolution [2]. Thus, an alternative visual sensing and perception system for autonomous vehicles is provided in challenging scenarios that state-of-the-art standard frame-based cameras cannot well perform [3], [4], such as high-speed scenes of the autonomous highway driving, low latency of motion control, and low power consumption of the vehicle onboard system.

It is well known in the research of autonomous driving that radar, lidar, ultrasound, and cameras form the backbone of sensor systems of the autonomous vehicle [5]–[7]. These sensors acquire the visual data as a sequence of snapshots recorded at discrete time stamps; therefore, visual information is compressed and quantized at a predefined frame rate. Consequently, a problem that is often known from the signal processing domain (undersampling) arises due to the timescale of motions in the observed scenes and the frame-rate of the recording camera. Things occurring between the adjacent frames, along with the consequent information, would get lost. Generally, the advanced algorithms with multiple-sensor fusion are usually developed to compensate single-sensor shortcomings in demanding applications such as highly piloted driving systems with low-latency motion control and visual feedback loops. Rather than solving this problem from an algorithmic perspective, it is better to explore alternative methods from a novel sensing perspective, such as event-based neuromorphic vision sensors. This results in providing great value for promoting subsequent tasks to become more robust, accurate, and complementary together with advanced algorithm development.

As an emerging sensing technology, the algorithms and applications of event-based neuromorphic vision are in the preliminary stage. Some works have been summarized in [8]. Unlike [8], this article aims to provide a thorough overview of the event-based neuromorphic vision for autonomous driving, from a signal processing perspective with a focus on visual perception algorithms and applications (see Figure 1). Specifically, the introduction starts from the operation principle of this bio-inspired neuromorphic vision sensor; then, the unique advantages of the sensor and its connection with the perception system of autonomous vehicles are discussed. Taking these promising properties into consideration, the signal processing techniques about event noise processing, event data representation, and meaningful event-based neuromorphic vision algorithms of given autonomous driving tasks are illustrated. Afterward, the works of event-based neuromorphic vision that are dedicated to specific applications in autonomous driving are reviewed. Finally, we address the problems remaining to be tackled and the directions for future research.

As an emerging sensing technology, the algorithms and applications of event-based neuromorphic vision are in the preliminary stage.

Bio-inspired vision

A biological retina

The retina of vertebrates, such as humans, is a highly developed multilayer neural system consisting of light sensitive cells which contain millions of photoreceptors. It is the place where the acquisition and preprocessing of the visual information happen. As shown in Figure 2(c), the retina has three primary layers including the photoreceptor layer, the outer plexiform layer, and the inner plexiform layer.

The photoreceptor layer consists of light-sensitive cells that convert incoming light into electrical signals and drive the horizontal cells and bipolar cells in the outer plexiform layer. There are two major types of bipolar cells: ON- and OFF-bipolar cells. The ON- and OFF-bipolar cells are responsible for coding the bright and dark spatial-temporal contrast changes, respectively. Particularly, the firing rate of the ON-bipolar cells will increase while the OFF-bipolar cells will no longer generate spikes if the illumination is increasing. This, in turn, increases the firing rate of OFF-bipolar cells in the case of illumination decreasing (such as getting darker). In the absence of a light stimulus, both cells generate few random spikes. This phenomenon is achieved by comparing the photoreceptor's signals with the spatial-temporal values, which are determined by the mean value of the horizontal cells, facilitating the connection between photoreceptors and bipolar cells laterally. In the outer plexiform layer, the ON- and OFF-bipolar cells synapse onto the amacrine cells and ON- and OFF- ganglion cells in the inner plexiform layer. The amacrine cells mediate signal transmission between bipolar cells and ganglion cells. The ganglion cells carry information along with different parallel pathways in the retina, which is conveyed to the visual cortex. Thus, the retina is responsible for converting spatial-temporal illumination changes into pulses, which is transmitted to the visual cortex via the optic nerve.

Silicon retina

Silicon retinas are visuals that model the biological retina and follow neurobiological principles. Pioneers of silicon retinas are Mahowald and Mead, who introduced their silicon VLSI retina in 1991 [9]. This kind of sensor is equipped with adaptable photoreceptors and a chip with a 2D hexagonal grid of pixels. It replicates parts of cell types of biological retinas, including the photoreceptors, bipolar cells, and horizontal cells. Therefore, this kind of sensor represents merely the photoreceptor layer and the outer plexiform layer. Later, Zaghloul and Boahen built the Parvo-Magno retina, which is superior to the silicon VLSI retina, by modeling five retina layers.

Despite the promising structure, many of the early silicon retinas originate from the biological sciences community and are mainly used to demonstrate neurobiological models and theories without considering real-world applications. Recently, an increasing amount of effort from Tobi Delbruck's team has been put into the development of practicable silicon retina DVS based on biological principles [1]. In Figure 2, the three-layer model of a human retina [Figure 2(c)], and corresponding DVS pixel circuitry [Figure 2(a)] are presented. Typical signals of the pixel circuits are displayed in Figure 2(b). The upper trace denotes a voltage waveform at the node $v_{\log}$, which tracks

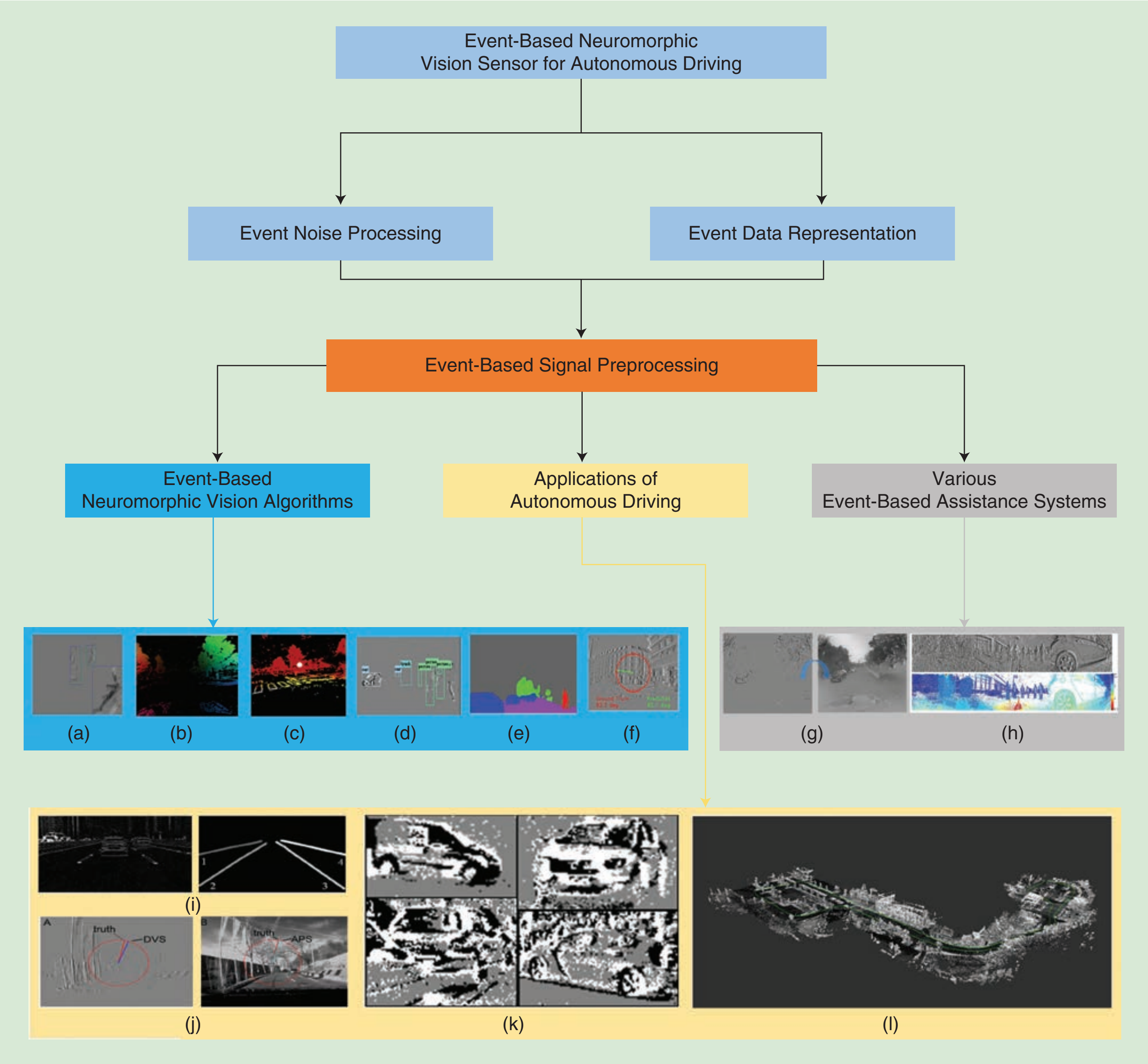


FIGURE 1. An overview of event-based neuromorphic vision sensors for autonomous driving, with representative examples for emerging systems and applications: (a) tracking (adapted from [22]), (b) optical flow (adapted from [29]), (c) depth estimation (adapted from [29]), (d) object detection (adapted from [52]), (e) semantic segmentation (adapted from [35]), (f) steering prediction (adapted from [3]), (g) image reconstruction (IR) (adapted from [45]), (h) panoramic stereo vision (adapted from [47]), (i) DET data set (adapted from [16]), (j) DDD17 data set [18] (adapted from [3]), (k) N-Cars data set (adapted from [17]), and (l) MVSEC data set (adapted from [4]).

the photocurrent through the photoreceptor layer circuit. The outer plexiform layer circuit responds with spike events (v_{diff}) of different polarities to positive and negative changes of the photocurrent. Spikes are transported to the next processing stage by the inner plexiform layer circuit. A large number of log-intensity changes are encoded in the events. Figure 2(d) illustrates the accumulated events including ON event (illumination increased) and OFF event (illumination decreased) that are drawn as white and black dots.

Today's representatives of silicon retinas are mainly from pioneers Tobi Delbruck and Christoph Posch and represent a compromise between biological and technical aspects. In their development, one prominent challenge posed is usually regarded as a wiring problem, indicating that each pixel of the silicon retina needs its own cable, which is impossible for chip wiring. A key technique for the solution, named *address event representation* (*AER*) was originally from the Caltech group of Carver Mead; it is used as an event-controlled and asynchronous point-to-point communication protocol for prototypes of the silicon retina.

As illustrated in Figure 3, the basic functionality of AER is implemented by an address encoder (AE), an address decoder (AD), and a digital bus. All neurons and pixels could transmit the time-coded information on the same line because the digital bus implements a multiplex strategy. The AE of the sending chip generates a unique binary address for each neuron or pixel in case of a change. The bus transmits the address at high speed to the receiver chip. Then, the AD determines the position and generates a spike on the receiver neuron. Event streams are employed in AER to communicate among chips. An event is a tuple (x, y, t, p); x and y are pixel addresses; t is the time stamp; and p represents the polarity. The polarity indicates the increase and decrease in the lighting intensity, corresponding to an ON event and OFF event, respectively.

This article focuses mainly on the first practically usable silicon retina, the DVS, which follows the natural, frame-free, and event-driven approach that triggers a plethora of research in event-based neuromorphic vision and autonomous driving. [A recent approach by Tobi Delbruck is the so-called dynamic and active pixel vision sensor (DAVIS) that combines dynamic and static visual information into a single pixel.] The DVS pixel models a simplified three-layer biological retina by mimicking the information flow of the photoreceptor–bipolar–ganglion cells (see Figure 2). Pixels operate independently and attach special importance to the temporal development of the local lighting intensity. The DVS pixel would automatically trigger an event (either ON event or OFF event) when the relative change in intensity exceeds the threshold. Therefore, the working principle of the DVS is fundamentally different from the

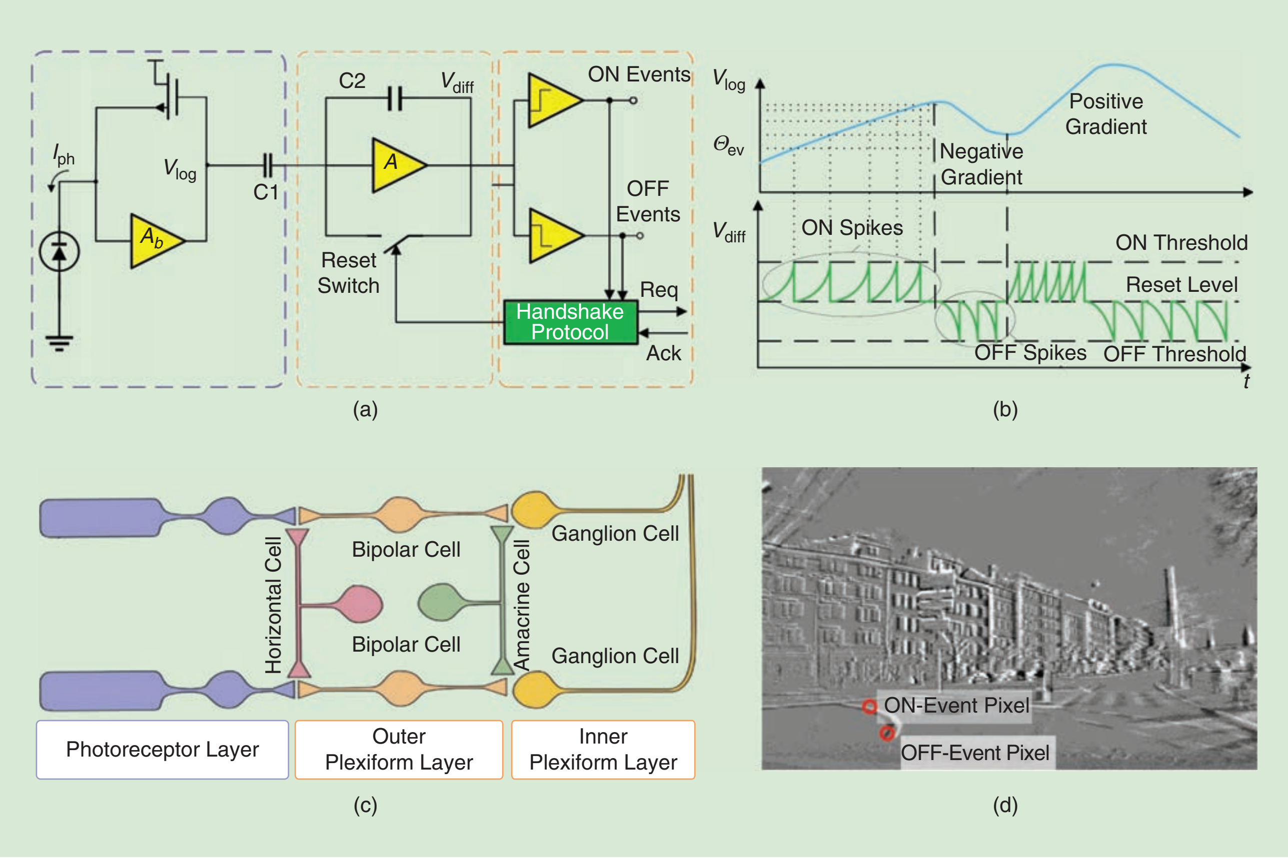


FIGURE 2. A practicable silicon retina DVS based on biological principles (adapted from [53]): (a) DVS pixel circuitry, (b) typical signals of the pixel circuits, (c) a three-layer model of a human retina, and (d) the accumulated events from a DVS. The accumulated event map has ON event (illumination increased) and OFF event (illumination decreased) drawn as white and black dots.

frame-based camera. There are three key properties of biological vision that are kept in this silicon retina: the relative illumination change, the sparse event data, and the separate output channels (ON/OFF). The major consequence of the DVS is that the acquisition of visual information is no longer controlled by any form of external timing signals such as frame clock or shutter, while the pixel itself controls its own visual information individually and autonomously.

Advantages of bio-inspired vision sensors

Due to the fundamentally different working principle and the mimicking of the biological retina, the event-based neuromorphic vision sensors have several advantages over standard frame-based cameras.

- *Energy-friendly properties*: Since event-based neuromorphic vision sensor transmits only events and autonomously filter redundant data, power is only used to process active pixels (e.g., the events triggered by illumination changes). Particularly, an energy-friendly sensor is more important than advanced algorithms for the onboard computers and devices in autonomous vehicles.
- *Low latency*: There is no need for the global exposure of the frame because each pixel works independently. Ideally, the minimal latency is 10 μs. The low-latency control of the autonomous vehicle is highly dependent on the perception systems. A low-latency perception system such as an object-detection system based on an event-based neuromorphic vision sensor would save lots of time in avoiding obstacles for the control systems.
- *HDR*: The event-based neuromorphic vision sensor such as DVS has an HDR (120 dB), which far exceeds that of the frame-based cameras (60 dB). Event-based neuromorphic vision sensors such as the DVS can simultaneously adapt to very dark and bright stimuli ensuring a highly robust perception system even in a light-changing scene such as an autonomous vehicle driving through a tunnel.
- *Microsecond resolution*: The brightness changes can be captured quickly in analog circuitry. With a 1-MHz clock, events can be detected and time-stamped with microsecond resolution. Considering the fast response requirement of the controller in autonomous vehicles in emergency driving scenes, this property is quite useful in autonomous driving.
- *No motion blur*: In the high-speed driving scenario, the motion blur problem occurs when the motion of the moving objects is beyond the sampling frequency of the frame-based camera; this may cause the failure of the perception system. An event-based neuromorphic vision sensor can capture dynamic motion precisely with no motion blur; it is of great value to autonomous driving community.

Event noise processing

The preprocessing of the raw data is essential for extracting meaningful information for sensor systems. An event-based neuromorphic vision sensor not only captures the change in the light intensity caused by moving objects, it also generates some noise activities due to the movements of background objects and the sensor noise such as temporal noise and junction leakage currents [10]–[12]. As shown in Figure 4, the event noise processing technique is responsible for excluding the event noises from the event stream. Two commonly used methods in the literature, namely the spatial-temporal correlation filter and the motion consistency filter, are illustrated as follows.

Spatial-temporal correlation filter

For a newly incoming event $e_i = (x_i, y_i, t_i, p_i)$, the spatial-temporal filter searches the most recent neighborhood event around the current pixel location (x_i, y_i) within a distance D. The incoming event would be regarded as a nonnoise event if the time difference meets:

$$t_i - t_n < d_t, \qquad (1)$$

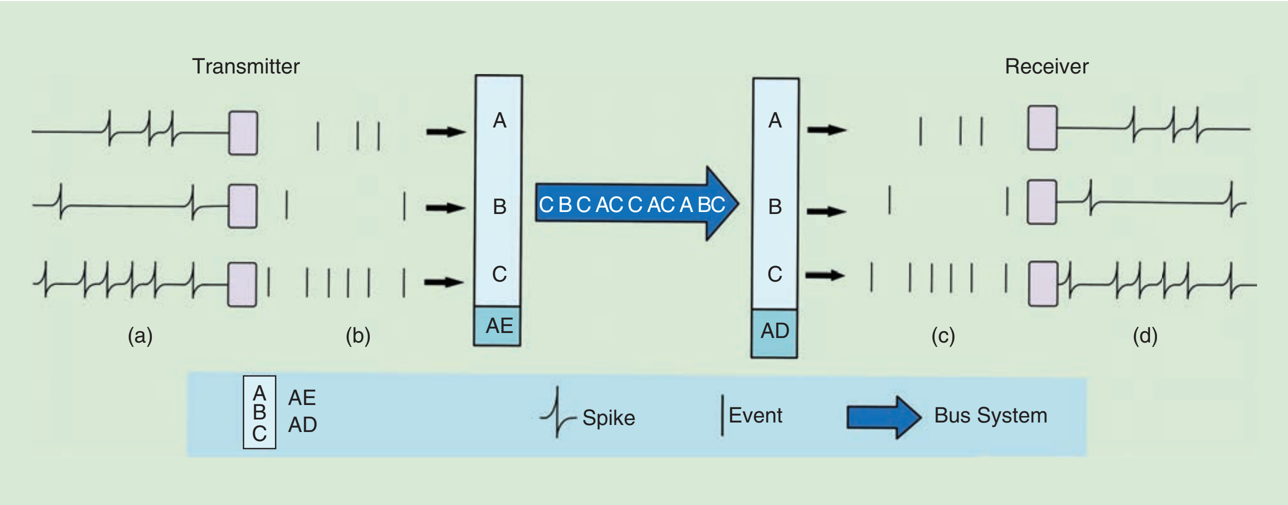


FIGURE 3. The AER communication protocol: (a) three neurons on the sending chip generate spikes; (b) spikes are interpreted as binary events. A binary address is generated by the AE and transmitted to the receiver chip by the bus line; (c) the binary address is decoded to the binary event by the AD; and (d) spikes are emitted on the corresponding neurons of the receiver chip where the positions of the neurons are determined by the AD.

where t_i is the time stamp of the event; t_n is the time stamp of the most recent neighborhood event; and d_t is the predefined threshold. The search for the most recent event checks eight neighborhood pixels around (x_i, y_i), as shown in Figure 4. It lacks temporal correlation with events in their spatial neighborhood because the event noise occurs randomly. Hence, the spatial-temporal correlation filter can effectively filter out event noise.

Motion consistency filter

In Figure 4, the principle of the motion consistency filter [13] is depicted. The blue dot denotes an incoming event caused by the object motion and the black dot represents an event noise. In the spatial-temporal domain, a newly incoming event should be consistent with the previous events (represented by red dots) caused by the same moving object. In a local region, the incoming event can be modeled as a consistent "moving plane" M. In this way, the velocity (v_x, v_y), can be used to assess the motion consistency, and the event noise can be removed because the previous events (the red dots, signal) and the black dot are not on the same plane. Concretely, the motion consistency plane for each active event e_i can be formulated as

$$ax_i + by_i + ct_i + d = 0, \tag{2}$$

where $(a, b, c, d) \in \mathbb{R}^4$ defines the plane M; (x_i, y_i) is the coordinate of event e_i; and t_i is the time stamp of event e_i. The event noise processing is an essential step to extract useful information from unwanted noise data for bio-inspired visual sensing and perception tasks of autonomous driving; it can promote the accuracy and speed of subsequent algorithms.

Event data representation

As an emerging sensing modality, event-based neuromorphic vision sensors only transmit local pixel-level changes caused by movement or light intensity change in a scene. The output data are sparse and asynchronous event streams which cannot be directly processed by standard vision pipelines, such as convolutional neural network (CNN)-based architecture. Therefore, encoding methods are utilized to convert asynchronous events into synchronous image- or grid-like representations for subsequent tasks such as object detection and tracking. According to whether or not the methods contain temporal information in the converted representations, we introduce two state-of-the-art encoding methods: spatial encoding and spatial-temporal encoding methods.

Spatial encoding

The spatial encoding methods convert event streams into event frames by storing event data at pixel location (x_i, y_i) with either fixed-time interval (e.g., 30 ms, constant time frame) or fixed number of events (e.g., 500 events, constant count frame). For an event frame, the value of the pixel is usually represented by the polarity of the last event (the positive event is 1 and the negative event is –1) or the statistical characteristics (such as the event count in a fixed-time interval, event count frame) of the events in the fixed interval. Assuming that $e_i(x_i, y_i, t_i, p_i)_{i \in [1,N]}$ represents event stream, typical approaches based on spatial encoding can be defined as follows:

1) *Constant time frames*:

$$F_j^t = \mathbf{card}(e_i | T \cdot (j-1) \le t_i \le T \cdot j), \tag{3}$$

where F_j^t represents the jth frame of time interval T; **card**() is the cardinality of a set; and e_i is the ith event of the event stream.

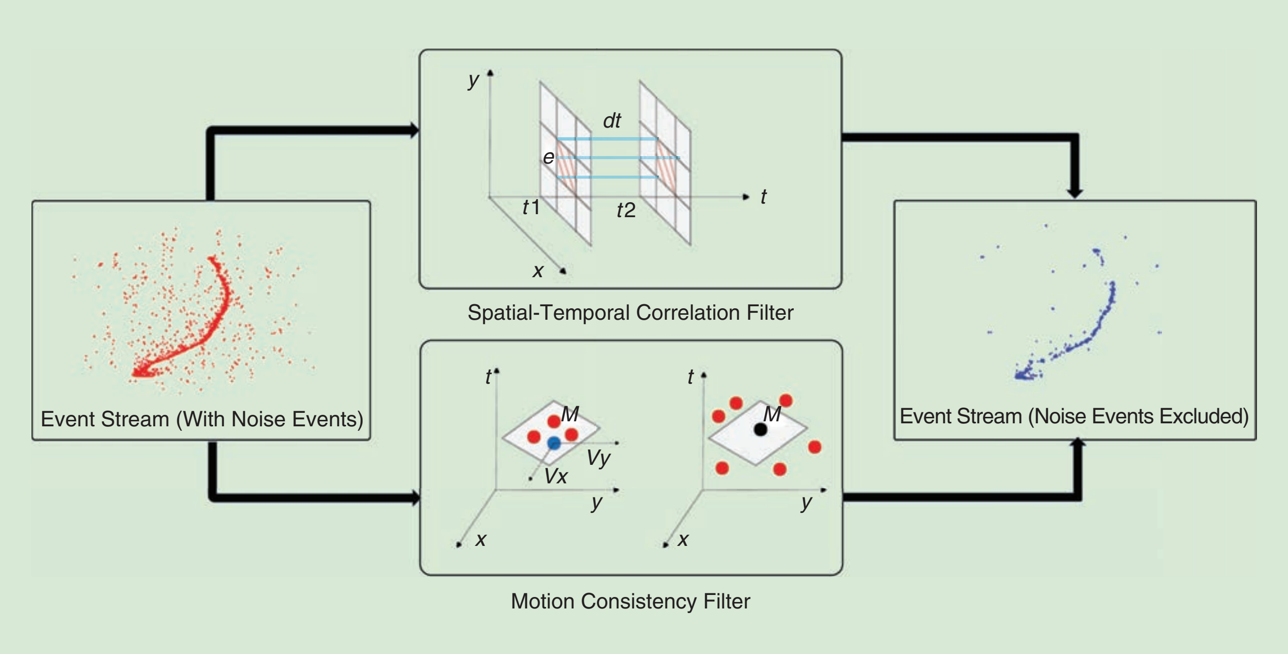


FIGURE 4. Event noise processing. The top branch is the spatial-temporal correlation filter; the bottom branch is the motion consistency filter.

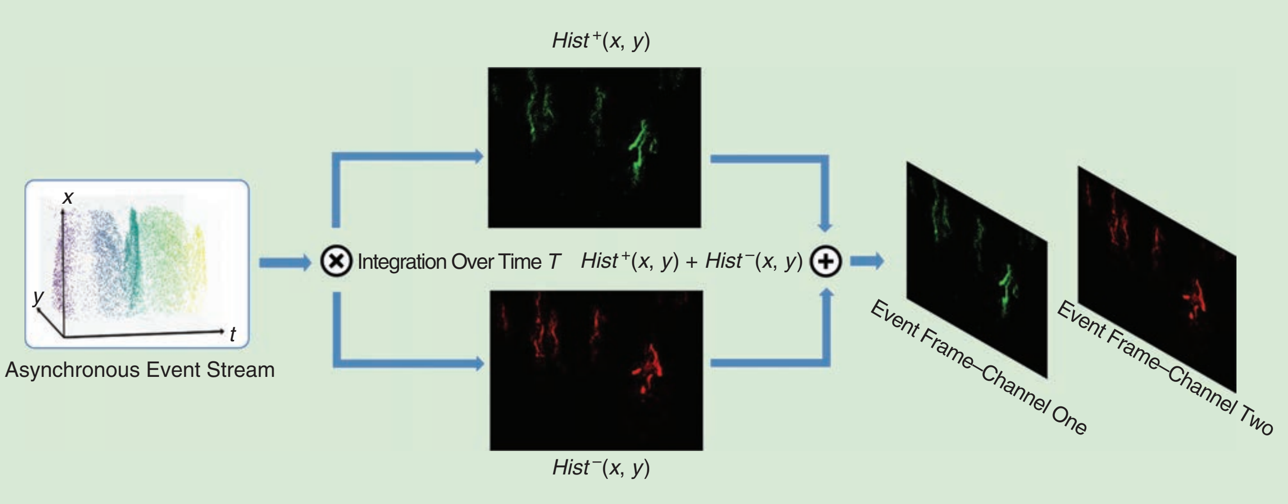


FIGURE 5. The process of converting asynchronous event data into an event frame. An event frame consists of two histograms from the positive events and negative events, respectively.

2) *Constant count frames*:

$$F_j^e = \mathbf{card}(e_i \,|\, E\cdot(j-1) \le i \le E\cdot j). \tag{4}$$

The constant count frame is defined similarly to constant time frame. F_j^e is the jth frame that contains E events.

3) *Event count frames*:

$$Hist^+(x, y) = \sum_{p_i=+1, t_i \in T} \delta(x - x_i, y - y_i). \tag{5}$$

Two separate histograms for positive and negative events are generated in a fixed-time interval T. $Hist^+(x, y)$ denotes the histogram for positive events, where δ is the Kronecker delta function. The same goes for the negative-events histogram, which is represented by $Hist^-$ with $p_i = -1$. The final representation of the events in the fixed-time interval T is an event frame, which consists of two histograms $Hist^+$ and $Hist^-$, as shown in Figure 5. Since the principle of the spatial encoding method is to project the events onto the spatial plane ($x - y$ plane), it loses the temporal information of all of the events.

Spatial-temporal encoding

The microsecond temporal resolution of the event stream provides a highly precise recording and description of the scene dynamics, which is valuable in many perception tasks such as high-speed moving object detection (e.g., vehicles). Spatial-temporal encoding methods combine spatial and temporal information of the events and convert events into a compact representation. A comparison of spatial-temporal encoding methods is presented in Table 1. A detailed description of these methods is displayed as follows.

Surface of active events

The surface of active events (SAE) uses time-stamp values instead of intensity values to represent the pixel values. For each incoming event e_i:

$$SAE: t_i \mapsto P(x_i, y_i), \tag{6}$$

where t_i is the time stamp of the most recent event at each pixel, the pixel value P at (x_i, y_i) is directly determined by the occurrence time of the events. The disadvantage of the SAE method is that it completely ignores the information of previous events happening at (x_i, y_i) and only uses the time stamp of the most recent event.

Leaky integrate-and-fire

Leaky integrate-and-fire (LIF) is an artificial neuron inspired by biological perception principles and computation primitives. A neuron receives input spikes (events) generated from a DVS, which modifies its membrane potential. If the membrane potential exceeds a predefined threshold, a spike stimulus will be sent to the output. The LIF neuron can be modeled as

Table 1. The comparison of different event data representations of spatial-temporal encoding.

Representation	Dimensions	Polarity Channel	Intensity	Weakness
SAE	H × W	2	Time stamp of the most recent event	Without temporal history
LIF	H × W	1	Event spikes	Without polarity information
Voxel grid	B × H × W	1	Sum event polarities	Without polarity information
EST	B × H × W	2	Sample event point-set into the grid	Without the least amount of information

The polarity channel is 2 if the encoding method considers the polarities of events; otherwise, it is 1. H and W represent the image height and width dimensions, respectively; B denotes the number of temporal bins.

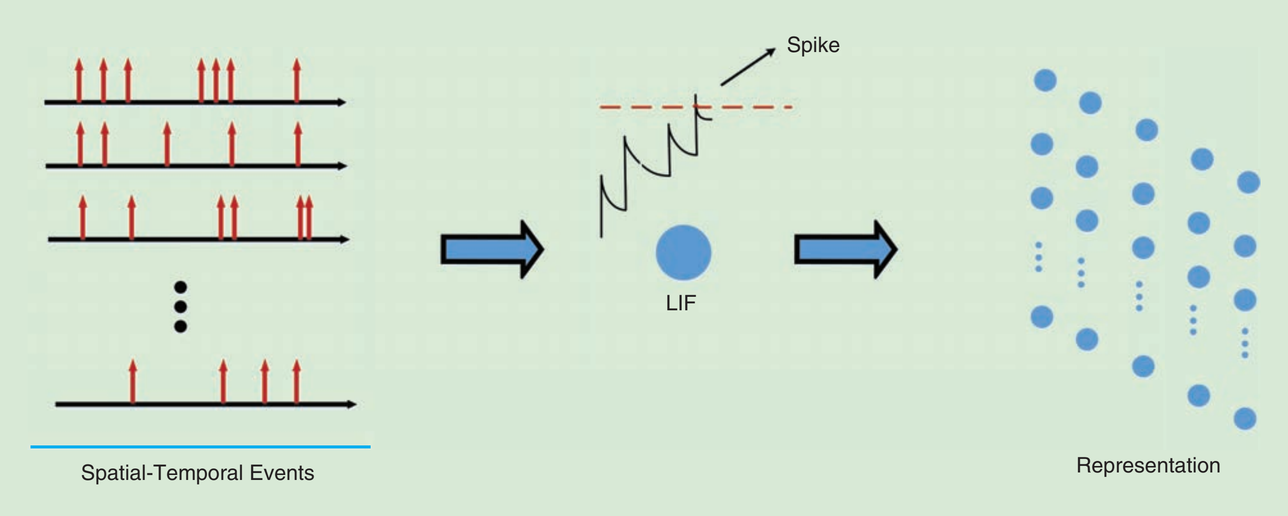


FIGURE 6. An LIF representation: Asynchronous spatial-temporal events are converted into event data representation by LIF neurons.

$$\tau \frac{dV}{dt} = -(V(t) - V_{\text{reset}}) + RI(t), \tag{7}$$

where, $V(t)$ is the membrane potential, which is a function across time; $I(t)$ is the total synaptic current; R is the membrane resistance; and τ is the membrane time constant. The neuron fires (produces an output spike) when the membrane potential reaches the threshold voltage (V_{th}) and then resets to reset voltage (V_{reset}). As shown in Figure 6, the spatial-temporal events are encoded by an LIF neuron, in which each event updates membrane potential of the neuron and the final converted representation is composed of the output spikes. An LIF neuron can not only transform event data into representation, it also serves as the basic unit of a spiking neural network (SNN) (see the section "SNNs").

Voxel grid

Voxel grid is a novel event representation aiming to improve the resolution of event stream in the temporal domain. Given a set of N events $(x_i, y_i, t_i, p_i)_{i \in [1,N]}$, B bins are used to split the time dimension; then, the time stamps of events are scaled to the range of $[0, B-1]$. The event voxel grid is defined as

$$\hat{t} = (B-1)(t_i - t_1)/(t_N - t_1), \tag{8}$$

$$V(x, y, t) = \sum_{i}^{N} p_i k(x - x_i) k(y - y_i) k(t - \hat{t}), \tag{9}$$

$$k(z) = \max(0, 1 - |z|), \tag{10}$$

where, $k(z)$ is the trilinear voting kernel, which is equivalent to the definition in [14]. As shown in Figure 7, events are converted into voxel grid representation with the fixed kernel. This representation retains the distribution of the events across the spatial-temporal dimensions.

Event spike tensor

Event spike tensor (EST) is an end-to-end learned representation [15]. In a given time interval T, EST can be formed by sampling the convolved signal,

$$S_{\pm}[x, y, t] = \sum_{e_i \in p_{\pm}} f_{\pm}(x_i, y_i, t_i) k_c(x - x_i, y - y_i, t - t_i), \tag{11}$$

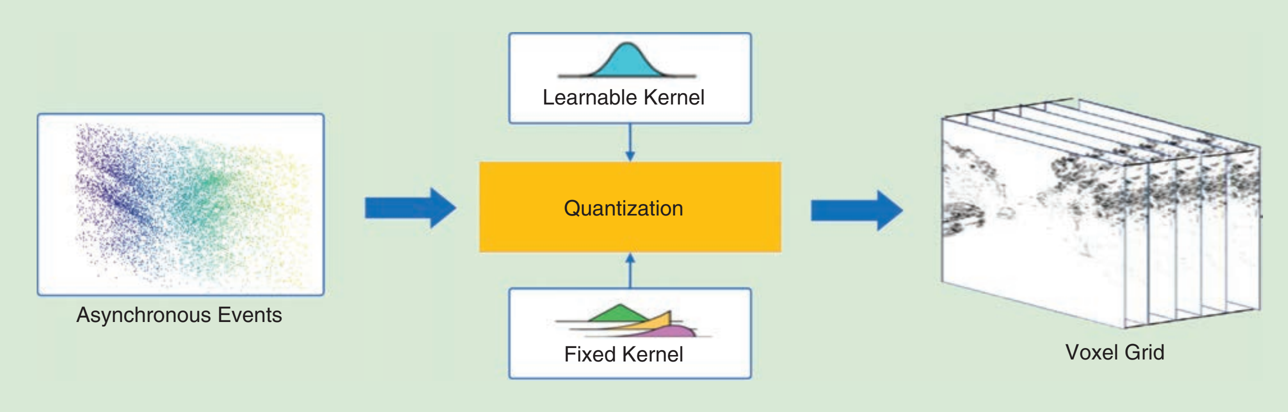


FIGURE 7. The process of converting asynchronous event data into grid-based representation with a fixed kernel [14] and a learnable kernel [15].

where, $f_{\pm}(x_i, y_i, t_i)$ is a measurement assigned to each event to represent the corresponding intensity value at the pixel location; k_c is the kernel convolution function to derive meaningful signal from the event stream. Generally, both the measurement and kernel are handcrafted functions in previous works, as illustrated in Figure 7. Particularly, the EST deploys a multilayer perception replacing the handcrafted kernel function in (11) to fit the data with the purpose of finding the best function for event streams. Simultaneously, the measurement function is chosen from a set of fixed functions. Examples of such function are the event polarity $f_{\pm} = \pm 1$; the event count $f_{\pm} = 1$; the time stamp $f_{\pm} = t$; and the normalized time stamp $f_{\pm} = (t - t_0)/T$.

There are many data sets that are built to promote the research of event-based neuromorphic vision, neurorobotics, and autonomous vehicles.

Event-based neuromorphic vision algorithms and applications of autonomous driving

The fundamental algorithms are the basis of the perception system of autonomous driving. For emerging systems and applications of bio-inspired vision, event-based neuromorphic vision algorithms are designed to extract features from event streams to fulfill given tasks. These methods can run directly on the event stream or take event representations as input (see the section "Event Data Representation"). They have been applied successfully in many vision tasks.

Event-based data sets of autonomous driving

In recent years, researchers have started to investigate the usage of event-based neuromorphic vision sensor such as DVS and DAVIS in the visual sensing and perception system of the autonomous driving system. There are many data sets that are built to promote the research of event-based neuromorphic vision, neurorobotics, and autonomous vehicles. In this section, four public event-based data sets dedicated to autonomous driving are discussed.

DET data sets

The performance of conventional lane extraction algorithms is limited because a frame-based camera cannot work well when the light is extremely dark or changes rapidly. To tackle this problem, [16] uses event-based neuromorphic vision sensors to build a high-resolution data set, called the *DET data sets*, for lane extraction. The DET data set containing various traffic scenes is collected by driving on tunnels, bridges, overpasses, and urban areas. The data set includes 5,424 event frames of 1,280 × 800 pixels with corresponding labels and consists of a training set of 2,716 frames, a validation set of 873 frames, and a test set of 1,835 frames. Two kinds of labels (per-pixel label without distinguishing lanes and per-pixel label with distinguishing lanes) are provided. The DET data set is the first bio-inspired vision data set for lane detection—a fundamental problem in autonomous driving.

N-CARS data sets

The N-CARS data set introduced by [17] provides recording cars in urban environments with a DVS. The data set consists of 12,336 car samples and 11,693 noncar (background) samples. Specifically, 7,940 car samples and 7,842 background samples are training samples, and others are testing samples. Each example is labeled by semiautomatic protocol with manual correction of the wrong one.

MVSEC data sets

In [4], the multivehicle stereo event camera data set (MVSEC) created for 3D perception with multiple sensors was presented. The MVSEC is the first data set with a synchronized stereo event-based neuromorphic vision system. The ground-truth depth data are generated from a calibrated lidar system contributing to stereo depth estimation with the event-based vision sensor. The MVSEC data set consists of long outdoor sequences in a variety of illuminations and driving speeds, which can be used for the evaluation of event-based visual odometry, localization, obstacle avoidance, and 3D reconstruction in challenging and real-word driving scenes.

DDD17 data set

For self-driving applications, end-to-end learning of the control model is a fascinating direction. The DDD17 data set [18] is the first large-scale public data set with a DAVIS sensor. The data are recorded in highway and city scenes driving from Switzerland to Germany. It has more than 12 h of data collected under different weather, road, and light conditions, covering the distance of more than 1,000 km. Furthermore, vehicle data, such as speed, GPS position, driver steering, throttle, and brake are also recorded.

Handcrafted feature

The concept of time surface is proposed to track the activity of the object due to the lack of effective low-level feature representations and descriptors for an event-based vision mission. It represents temporal characteristics and describes the spatial-temporal context around an event. For an event $e_i = (x_i, y_i, t_i, p_i)$, the time surface S_i of dimension $2R \times 2R$ is defined as

$$S_i = \begin{cases} e^{-\frac{t_i - T(C_i+R,P)}{\tau}}, & \text{if } p_i = P;\\ 0, & \text{otherwise,} \end{cases}$$

where $C_i = (x_i, y_i)$ is the pixel coordinates of the incoming event e_i, R is the radius of the spatial neighborhood around e_i, $T(C_i + R, P)$ is the time stamp of the last event with polarity P received from pixel $C_i + R$, and τ is a constant decay factor. The exponential decay expands the activity of past events and records history information of the activity in the neighborhood. Time surface has been effectively used in various vision tasks, such as object recognition and feature tracking. Further, a hierarchy of time surface is introduced for object recognition [19]. Relying on a time-oriented approach, this model is used to extract valuable spatial-temporal features from event streams. Based on the findings in [19], a sparse coding basis decomposition was used to reduce the number of prototypes in a hierarchy structure for lowering computational cost and memory need [20]. However, they only achieved better recognition performance for simple shapes, such as numbers and letters, while they cannot well perform for complex objects, such as cars. Inspired by the histogram of orientation gradient feature widely used in frame-based vision, an effective event descriptor named *histogram of averaged time surfaces* (*HATS*) was constructed [17]. Then, better classification performance and real-time computation were obtained. HATS convert event streams into local memory time surfaces and computes the histograms to formulate the final descriptor. After these features are extracted from event streams, a simple linear support vector machine classifier is used to recognize objects in the N-CARS data set.

Clustering

A classical unsupervised learning approach is clustering. Given a set of data, the clustering algorithm can be used in this study to generate different groups. The data with different characteristics are grouped into different clusters. The clustering methods can be applied directly to generate object proposals because the event stream from the DVS can be treated as sparse point cloud data where each point is an event. For example, a clustering method named *Gaussian mixture models* (*GMMs*) is used to track the pedestrian [22]. The method achieves accurate detection and tracking of pedestrian objects by extending GMMs with a stochastic prediction of objects' states. The goal of tracking is to estimate the state of one or multiple objects over time. In case of a possible collision with other traffic participants, the autonomous vehicle requires sufficient reaction time to ensure a safe brake distance. It is difficult to track a pedestrian because a pedestrian can suddenly change his or her moving direction. The results in [22] indicated that applying clustering to spatial-temporal event data has a large potential for robust object tracking.

Bio-inspired feature learning

SNNs

An SNN is a bio-inspired approach that can operate directly on spatial-temporal event data. The computational pattern of SNNs mimics the working principle of receptive fields in the primary visual cortex. As basic building blocks of SNNs, LIF and adaptive exponential are both inspired by the biological neurons found in the visual cortex of mammalians, which encode temporal information and make them naturally fit asynchronous event streams. The basic principle of SNN is that a neuron will not emit any spike if it has not received any input spike from the preceding SNN layer. Moreover, the corresponding neuron will generate spikes that are fed to the next layer only if the membrane voltage caused by received spikes exceeds a predefined threshold. The predefined network units, such as the difference of Gaussians or Gabor filters, are usually used in the first layer of SNN to extract features. Features are transmitted from the first layer of SNN to the deeper layers in parallel [23]–[25]. The major disadvantage of conventional SNNs is not differentiable, causing the popular training methods to be inapplicable. In the context of autonomous driving, a SNN architecture consisting of refractory, convolution, and clustering layers was presented [26]. It was designed with biorealistic LIF neurons and synapses. The LIF neurons are used as basic building blocks in the proposed algorithm, where the refractory layer filters off fraction of the input events to generate spike. Then, the spikes are convolved by convolution layer to produce region proposal boxes. Moreover, the clustering layer combines these boxes to cluster together to form the shapes of objects. This method is validated on object detection with real traffic scenes including humans, bikes, cars, vans, trucks, and buses.

SNN with backpropagation

SNN with handcrafted feature extractors (such as Gabor filters) cannot learn weights naturally from the data. To overcome this drawback, researchers established a novel architecture of SNN with LIF neuron and winner-takes-all (WTA) circuits [21]. The LIF neuron uses dynamic weights rather than a simpler refractory mechanism to update its membrane potential. In a WTA circuit, it would inhibit other neurons from spiking once an output spike occurs in a neuron. Furthermore, the lateral inhibition is employed to put the dynamic weights of all inhibited neurons in the WTA circuit into the refractory state. The differentiable transfer functions are derived in the WTA configuration to make SNN trainable with backpropagation; moreover, the performance of SNN architecture is also improved. In Figure 8, an SNN network with backpropagation is illustrated. However, trainable SNN is only tested on simple data sets (such as MNIST) and has not been applied in specific autonomous driving scenarios. As the output of event-based neuromorphic vision sensor is a spatial temporal event stream which is fundamentally different from frame-based camera, it requires the design of specifically tailored algorithms to accommodate the nature of events, and [21] indicates the prospect of implementing deep SNNs.

CNN

CNN is a popular feature extraction architecture, which is composed of three types of layers, including a convolutional layer, a pooling layer, and a fully connected layer. It uses spatially localized convolutional filtering to capture local features of input image. Basic visual features, such as lines, edges, and corners, are learned in the first few layers, while more abstract features are learned in deeper layers. For an input image matrix I, the correspondence activation map M is computed in the nth neuron of the CNN as follows

$$M[i, j] = \sigma\left(\sum_{x=-2k-1}^{2k+1} \sum_{y=-2k-1}^{2k+1} W[x, y] I[i-x, j-y] + b \right), \quad (12)$$

where the image size is $2k+1$, W is the nth convolutional filter, and σ is the nonlinear activation function. Generally, a max pooling layer follows each convolutional layer, in which the local maximum is used to reduce the dimension of the matrix and prevent overfitting. Moreover, fully connected layers are usually added to learn the nonlinear combination of extracted features from previous layers. Over the decades, many variants of CNNs, such as fully CNNs and encoder–decoder networks, have emerged. These networks have different structures from traditional CNNs, such as removing the full connection layer. The performance of CNNs has surpassed traditional machine learning methods in many vision tasks, relying on successful training algorithms and large amounts of data.

CNNs for optical flow, depth, and egomotion

Known as a 2D motion estimation, the optical flow is defined as the distribution of apparent velocities of movement of brightness patterns between two images. It provides valuable information about the scene and serves as input for several tasks, such as tracking and visual odometry. In the neuromorphic vision research community, some works attempt to estimate optical flow by taking advantage of high temporal resolution of event-based sensors [27]. EV-FlowNet, a self-supervised deep learning architecture for optical flow estimation for event-based sensors, is proposed in [28]. In this method, a four-channel event representation consisting of the histogram (5) and SAE (6) of different polarity is used to pass through a pipeline that is composed of four stride convolutional layers, two residual blocks, and four up-sampling convolutional layers for obtaining flow estimation. By evaluating an MVSEC data set, the network is able to accurately predict optical flow from event streams. In [29], a novel neural network framework is proposed to acquire motion information including optical flow, depth, and egomotion from a set of inputs (a voxel grid) that is an event data representation mentioned in the "Voxel Grid" section. The network architecture consists of encoder–decoder networks and pose models; among them, the encoder–decoder section is responsible for predicting optical flow and depth, while the pose model is responsible for estimating egomotion. Experimental results in the MVSEC data set indicate that the presented network can learn various motion information of events well. Recently, a lightweight evenly cascaded convolutional network (ECN) using monocular event-based sensor input for dense depth, optical flow, and egomotion estimation was introduced in [30]. ECNs use an encoder network to predict pose; meanwhile, an encoder–decoder network is applied to obtain the scaled depth. The algorithm can operate at 250 frames/s (fps) on a single NVIDIA 1,080 titanium GPU. Compared with previous works, it makes significant improvements on the performance of the MVSEC data set.

CNNs for object detection

Reliable object detection is essential to avoid accidents that might be life threatening because a self-driving car is sharing the road with many traffic participants, such as vehicles and pedestrians. For instance, a supervised learning method is applied on event data for object detection under egomotion [31]. The data set used in this article is DDD17, which is a large event-based data set applying DAVIS to record various challenging scenarios under egomotion. The DAVIS is a sensor consisting of an event-based neuromorphic sensor and a synchronized gray-scale frame-based camera. In [31], gray-scale images are fed into a state-of-the-art frame-based CNN to generate outputs (pseudolabels), which are used as ground truths for subsequent training on event-based data. This method achieves high-speed detection (100 fps) in a real outdoor scenario within various backgrounds such as day and night. As pseudolabels are not explicit enough, the authors manually labeled the DDD17 data set to explore the potential of event-based neuromorphic sensor for vehicle detection in autonomous driving [32]. A convolutional SNN is utilized to generate visual attention maps for synchronizing with the frame-based stream. Two separate event-based and frame-based streams are incorporated into a CNN detector to obtain detection output. With a joint decision model to postprocess the output, the algorithm outperforms the state-of-the-art

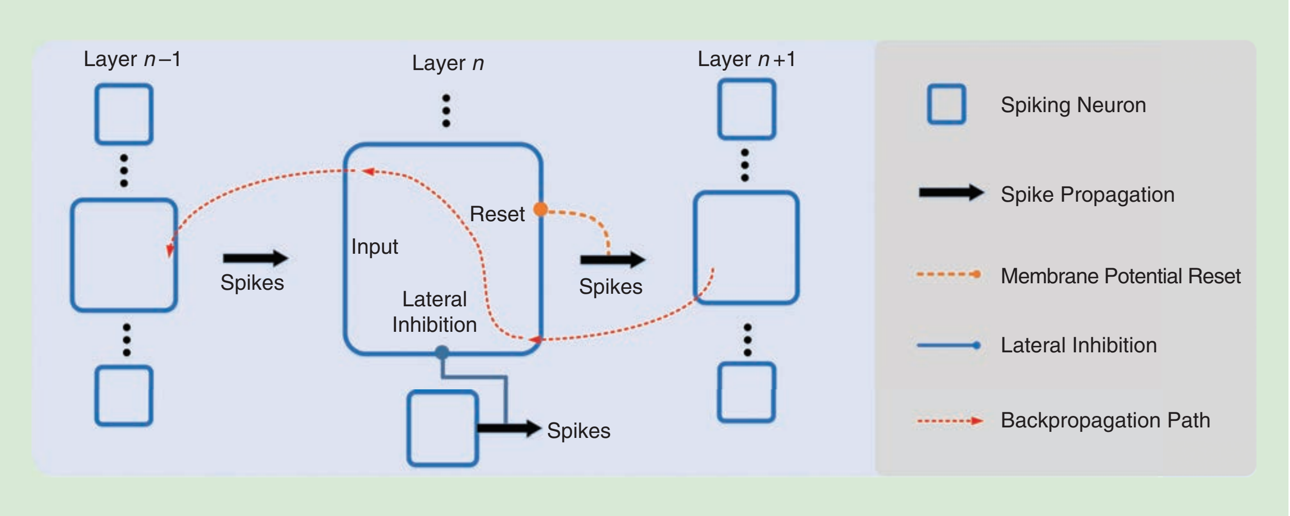


FIGURE 8. An example of how an SNN network works with backpropagation [21].

methods that only employ frame-based cameras. The detection for stationary and moving people around a self-driving car has attracted the attention of researchers. Specifically, a multicue event information fusion for pedestrian detection was proposed [33]; it was evaluated on the data set recorded by a neuromorphic vision sensor. Based on the advantages of leveraging various properties of event streams, this article performed better on positioning and recognition of pedestrians. Recently, a cross-modal approach was presented in [34]; wormhole learning was utilized to pair red, green, blue (RGB) camera and event-based neuromorphic vision sensors to improve the object detection performance under the scenario of urban driving. This method is different from transfer learning as it can be transferred back to the original domain to improve performance on the task. The experimental results of wormhole learning reveal that there are many innovative approaches to combine data from different heterogeneous sensors, such as RGB cameras, infrared cameras and neuromorphic vision sensors.

CNNs for semantic segmentation

In the sensing and perception system of autonomous driving, a comprehensive understanding of the surrounding environment is provided by semantic segmentation. The first CNN-based baseline for semantic segmentation with an event-based neuromorphic vision sensor is introduced in [35]. In this article, the authors build an event-segmentation data set (Ev-Seg) that is an extended version of DDD17 for semantic segmentation. Inspired by the study in [31], the labels of Ev-Seg are generated by running a trained CNN on gray-scale images. Then, an Xception-based CNN architecture is trained to learn generalization ability from event streams. Finally, the complementarity between the frame-based camera and event-based neuromorphic vision sensor is presented through comparing the semantic segmentation results produced from event data and corresponding gray-scale images.

CNNs for active perception

Controlling the autonomous vehicle in challenging scenes such as highway driving requires a low-latency perception system. Hence, researchers try to tackle this tough problem by unlocking the low-latency potential of event-based neuromorphic vision sensors. An end-to-end autonomous driving system, mapping from the event streams to the driving actions, is proposed in [3]. This system converts events to event count frames (histogram of different polarities) mentioned in (5), which are fed into a residual neural network (ResNet)-inspired network to predict the steering angle of the vehicle. The proposed method can accurately predict the steering angle of vehicles and performs better on DDD17 data sets than the state-of-the-art systems using gray-scale images.

CNNs-to-SNNs

CNNs have demonstrated their ability to deal with many difficult vision problems, such as object detection. SNNs have presented their potential for low-power event-driven neuromorphic hardware. However, the applications of SNNs are limited due to their shallow neural network architecture. Furthermore, the CNN-to-SNN model is developed to combine the benefits of deep architecture in CNNs with the bio-inspired mechanism of SNNs. References [36]–[38] illustrate that widely used CNNs, such as VGG, ResNet, and Inception-V3 can be converted into spiking networks. It is worth mentioning that the network can achieve a more robust performance via conversion from CNNs, although the conversion process would lose some precision and increase computation. Some works have been reviewed in [39].

Transfer learning via pretrained network

Transfer learning is a very effective method to improve the training performance of the deep neural network. Knowledge learned from a different domain can be exploited to initialize the weights of a deep neural network. The availability of event-based data sets collected with a DVS sensor is limited compared with the data set recorded by frame-based cameras. Thus, by starting the supervised training process from a better set of initial weights, the requirement of the training data can be reduced, and the generalization ability of the network can be improved. Pretrained models, such as VGGnet and ResNet, can be applied to bio-inspired sensing and perception tasks of autonomous driving. Specifically, event streams can be transformed into a three-channel image-like representation to serve as input to pretrained CNNs. In [40], the authors combined an inceptive event time surface (IETS) with transfer learning to improve performance of object classification. IETSs are generated to utilized transfer learning from the GoogLeNet that is pretrained on ImageNet, including the millions of real-world images. Nearly 100% classification accuracy on the event-based N-CARS data set is achieved by the algorithm. In [41], a robust event stream object tracking method is presented. A VGG-16 model pretrained on ImageNet is used to extract features to represent the appearance of the object. Based on correlative filter mechanism, the correlation response map is computed on the extracted features. The proposed approach performs well in various challenging visual scenarios.

Event-based assistance systems

After the basics of event-based perception system of autonomous driving are covered, the event-based assistance systems are discussed.

Image reconstruction

The event-based neuromorphic vision sensor generates HDR event data even in extreme illumination conditions and also avoids motion blur under rapid motion. Reconstructing HDR intensity images from event streams facilitates the adoption of mature computer vision techniques. Previous works focus on exploiting the low latency of neuromorphic vision sensor by directly processing event data (such as SNNs) or transferring events to image-like or grid-like representations as mentioned in the section “Event Data Representation.” However, the deep neural network trained on real image data (such as ImageNet) cannot be effectively transferred to these representations, even though it achieves some performance improvements (see the section "Transfer Learning via Pretrained Network"). As an alternative method, image reconstruction (IR) from event streams is first proposed in [42]. IR can achieve both high frame-rate images and high-quality images with no motion blur. In [43], the authors utilize the time stamp of new events to define a manifold for IR. With considering IR as an energy minimization problem, the proposed method is optimized and achieves real-time performance on a GPU. Furthermore, an asynchronous complementary filter is presented to reconstruct event streams for continuous-time intensity estimate [44]. In this article, the gray-scale frames and events produced by DAVIS are fused into an image with high temporal resolution and HDR. In addition, a new framework for IR, named *E2VID*, is introduced in [45]. E2VID converts event stream into 3D spatial-temporal voxel grid sequences (see the "Voxel Grid" section), which are taken as the input of the network. The algorithm is trained on a large synthetic event data simulated with ESIM [46] to generate reconstructed image frames. The reconstructed image data from event streams can be used for various applications such as object recognition, SLAM, and optical flow estimation.

Event-based neuromorphic vision is an emerging technique in the era of mature sensor hardware of autonomous driving.

Panoramic stereo vision

Panoramic vision in 3D offers a full 360° surrounding view which facilitates the navigation and localization tasks for autonomous driving. A novel multiperspective panoramic stereo event-based vision system is proposed in [47]. It is composed of a pair of line event-based neuromorphic vision sensors. The authors present a novel event-driven stereo matching approach for 3D panoramic vision. The process steps of the event-driven stereo matchintg algorithm include event map generation, event distribution measure, cost calculation, disparity estimation and refinement. The experimental results indicate that the tailored event-driven stereo method achieves accurately 3D reconstruction in real time out of 360° panoramic views.

Visual odometry

The goal of the visual odometry is to estimate the position and orientation of a vehicle with vision sensors. The visual odometry system of an autonomous vehicle with a traditional frame-based camera has been developed for many years, while the method based on an event-based neuromorphic vision sensor is still in the preliminary stage. For example, an event-based visual odometry system for intelligent vehicle applications is proposed in [48]. The events generated from a DAVIS sensor are aggregated into constant time frame defined in (3) to serve as input to subsequent algorithms. The feature tracking is used by visual odometry system to develop parallel pose estimation and mapping. The feasibility of event-based neuromorphic vision sensors for bio-inspired visual odometry systems in real-word outdoor driving scenes is confirmed by the results of their experiment on the MVSEC data set.

Drowsiness driving monitoring

Drowsiness driving monitoring is important to ensure that the autonomous driving vehicle is under the supervision of the drivers. In [49], an event-based drowsiness driving detection system is proposed. The event-based neuromorphic vision sensor is considered as an efficient and effective detector for the drowsiness driving-related motions due to the unique output. [49] proposes to recognize and localize the driver's eyes and mouth motions from event streams, and extracts event-based drowsiness-related features directly from the event streams caused by eye and mouth motions. Experiments in [49] demonstrate the high efficiency and accuracy under different illumination conditions such as subjects wearing sunglasses.

Spike compression

The event data compression is particularly important for maintaining the real-time performance of the sensing system of autonomous vehicles because both the data storage and transmission bandwidth of on-board event-based neuromorphic visions sensors equipped on the autonomous vehicles are limited. To address this problem, a cube-based spike coding framework is proposed by [50]. In the spatial-temporal dimension, an octree-based structure is put forward to adaptively cut the event (spike) stream into coding cubes, then address-prior mode and time-prior mode are designed to exploit the spatial and temporal characteristics of events for data compression. The proposed spike coding framework is evaluated on the DDD17 data set. Experimental results indicate that it can achieve a better compression ratio against the raw event data. Reference [51] proposes to use mixture density autoencoder to learn a low-dimensional representation from an event stream, which preserves the nature of event-based data better while being easy to feed to a sequence classifier.

Challenges and future directions in autonomous driving

Event-based neuromorphic vision is an emerging technique in the era of mature sensor hardware of autonomous driving. Comparing it with lidar, radar, and cameras is unfair because event-based sensors such as DVS are not at the same maturity level as others. Conversely, there is substantial room for the development and improvement in the cross-research of event-based neuromorphic vision and autonomous driving. Challenges and future directions closely related to autonomous driving are pointed out in numerous opportunities, as described later.

Sensor fusion in perception system of autonomous driving

To fuse the event-based neuromorphic vision sensor with others, there is an unavoidable problem that the sensor fusion brings back the disadvantages of providing a redundant, sampled intensity output with linear encoding of intensity. On the contrary, the advantages are also obvious; that is, different kinds of sensors are complementary. For example, DVS contains no color information, which is provided by frame-based cameras. The distance and speed information can be provided by lidar and radar. It remains to be seen whether the DVS output can be used to trigger frame captures of other sensors. If it is, the DVS and other sensors can operate together with mixed conventional machine vision, bio-inspired, and event-based neuromorphic vision-based approaches. Therefore, some of the limitations of a traditional sensor-based perception system may be overcome; moreover, new scenarios that were previously inaccessible in the visual sensing and perception of autonomous vehicles might be reached.

There is an emerging need for high-quality benchmarks in the fields of event-based neuromorphic vision and autonomous driving.

Active vision system of autonomous driving

In robotics, the ability to directly fuse the perception with its motoric ability is often referred to as *active perception*. In autonomous driving, it is found that the perception and action are often kept in separated spaces; this is a consequence of state-of-the-art sensors equipped on the autonomous vehicle being frame-based. The sensing and perception only exist in a discrete moment while the motion is a continuous entity. It can be argued that the event-based neuromorphic vision sensor can see the motion, which has the potential to cross the bridge between perception and motor control. New methods of encoding perceptions and actions could be meaningful to the active perception system of autonomous driving. Moreover, this would create new opportunities for real-time navigation and obstacle avoidance for autonomous driving if the visual perception can be bound with the system dynamic to enable dynamic environment perception.

Large-scale autonomous driving benchmark based on an event-based neuromorphic vision sensor

It is well known that rapid development of autonomous driving is promoted by standardized benchmarks. For example, the growing popularity of deep neural networks in intelligent vehicles and large-scale benchmarks such as KITTI, Cityscale, and ImageNet, is interconnected and mutually reinforced. In the earlier days of event-based neuromorphic vison, most of the research work was done in an indoor environment due to the low resolution of sensors. Until recently, the event-based neuromorphic vision sensor has been expanded to outdoor scenarios, such as autonomous driving, by the teams of Tobi Debruck, Kostas Daniilidis, and David Scaramuzzsa. There is an emerging need for high-quality benchmarks in the fields of event-based neuromorphic vision and autonomous driving. A standard platform would bring the mainstream of computer vision-based intelligent vehicle research to pay attention to event-based neuromorphic vision; furthermore, the unique strengths of bio-inspired vision would be leveraged to attract research interests in new sensing techniques for autonomous driving.

From simulated event data to real-world autonomous driving

Labeling the asynchronous event data is always a challenging problem because almost all of the annotation tools are developed for frame-based cameras. Additionally, there is not a standard format for the annotations. From one perspective, developing an easy-to-use tool for recording and labeling event data would make a significant contribution to the community; from another perspective, the adoption of event-based neuromorphic vision technology would also be facilitated by developing simulators. Particularly, the corresponding event streams, intensity frames, and depth information could be generated by a simulator based on the working principle of the sensor. Simultaneously, the basic facts of all recording data including the trajectory of the sensor, the label of the object, and even the optical flow are also generated without the need for annotation. With photo-realistic virtual driving scenes and realistic sensor models, the development of event-based visual sensing and perception system in autonomous vehicles will be accelerated by prototyping on simulated event data with transfer learning methods in the future.

Limitations that may exist as event-based neuromorphic vision sensors mature

There is no appearance feature such as color and texture because an event-based neuromorphic vision sensor only transmits local pixel-level changes, making it perform poorly in some applications with high requirements for appearance features. Although researchers have used the method of IR (mentioned in the section “Spatial Encoding”) to reconstruct image frames from event streams, the quality of reconstructed image frames is still not comparable to the output data produced by RGB cameras. The application of an event-based neuromorphic vision sensor is limited in some scenarios where energy, latency, and dynamic range are not important, especially in high-resolution complex scenarios.

Conclusions

Innovative solutions will emerge due to the challenges remaining on the road to fully autonomous driving. Concurrently, sophisticated signal processing techniques have been successfully applied to autonomous driving hardware such as cameras, lidars, and radars. Exploring alternative methods of visual sensing such as event-based neuromorphic vision is promising for promoting subsequent tasks to be more robust and complementary. It is reasonable to say that the research and development of an event-based neuromorphic vision for autonomous driving is still in its infancy. In this article, the advantages, signal processing techniques, emerging applications and systems, and future directions of an event-based neuromorphic vision for autonomous driving have been introduced and analyzed. This article helps researchers and engineers take the first step in developing innovative signal processing techniques toward bio-inspired visual sensing and perception of autonomous vehicles.

Acknowledgments

This research has received funding from the Shanghai Automotive Industry Sci-Tech Development Program according to grant agreement 1838, from the Shanghai AI Innovation Development Program 2018, and from the European Union's Horizon 2020 Framework Program for Research and Innovation under the specific grant agreement 785907 (Human Brain Project SGA2).

Authors

Guang Chen (guangchen@tongji.edu.cn) received his B.S and M.Eng. degrees from Hunan University and his Ph.D. degree from the Technical University of Munich (TUM). He is a research professor at Tongji University, where he leads the Intelligent Perception and Intelligent Computation group. He is also a senior research associate (guest) at TUM. He was a research scientist at fortiss GmbH from 2012 to 2016, and a senior researcher at the Chair of Robotics, Artificial Intelligence and Real-Time Systems at TUM from 2016 to 2017. He was named the Tongji Hundred Talent Research Professor 2018. His research interests include computer vision, machine learning, and bio-inspired vision with applications in robotics and autonomous vehicles. He is a Member of the IEEE.

Hu Cao (hu.cao@tum.de) received his B.S. degree in vehicle engineering from Anhui University of Technology, China, in 2017 and his M.Eng. degree in vehicle engineering from Hunan University in 2019. He is currently a Ph.D. candidate at the Technical University of Munich. His research interests include neuromorphic engineering, robotics, and deep learning.

Jörg Conradt (conr@kth.se) received his M.S. degree in robotics from the University of Southern California and a diploma in computer engineering from the Technical University of Berlin. He received his Ph.D. degree from ETH Zurich. He an associate professor at the KTH School of Electrical Engineering and Computer Science in Stockholm, Sweden. Before joining KTH, he was a W1 professor at the Technical University of Munich (TUM). He was the founding director of the Elite Master Program Neuroengineering at TUM. He is a Senior Member of the IEEE.

Huajin Tang (htang@zju.edu.cn) received his B.Eng. degree from Zhejiang University, his M.Eng. degree from Shanghai Jiao Tong University, and his Ph.D. degree from the National University of Singapore. Since 2008, he has been the head of the Robotic Cognition Lab, Institute for Infocomm Research, A*STAR, Singapore. Since 2014, he has been a professor at Sichuan University and is now a professor at Zhejiang University, China. He received the 2016 IEEE Transactions on Neural Networks and Learning Systems Outstanding Article Award and 2019 IEEE Computational Intelligence Magazine Outstanding Article Award. His research interests include neuromorphic computing, neuromorphic hardware, and robotic cognition, among others. He has served as an associate editor of *IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cognitive and Developmental Systems, Frontiers in Neuromorphic Engineering*, and *Neural Networks*. He is a Board of Governors member of the International Neural Networks Society. He is a Member of the IEEE.

Florian Röhrbein (florian@gmx.org) received his diploma and Ph.D. degree from the Technical University of Munich (TUM) and the venia legendi for computer science from the University of Bremen. He is responsible for the development and implementation of the artificial intelligence (AI) strategy for a world-leading company and is the chief editor of *Frontiers in Neurorobotics*. He was also the managing director in the Human Brain Project at TUM. He has international experience in various projects on AI, computational neuroscience and brain-inspired cognitive systems. Research stays include the MacKay Institute of Communication and Neuroscience (United Kingdom), the Honda Research Institute Europe, and the Albert Einstein College of Medicine (New York). He is a Senior Member of the IEEE.

Alois Knoll (knoll@in.tum.de) received his M.S. degree in electrical/communications engineering from the University of Stuttgart, Germany, in 1985 and his Ph.D. degree (summa cum laude) in computer science from the Technical University (TU) of Berlin, Germany, in 1988. He served on the faculty of the Department of Computer Science of TU Berlin until 1993. He joined the University of Bielefeld as a full professor and was the director of the research group Technical Informatics until 2001. Since 2001, he has been a professor at the Department of Informatics, Technical University of Munich (TUM), and was executive director of the Institute of Computer Science at TUM from 2004 to 2006. He was the program chair of IEEE Humanoids2000, general chair of IEEE Humanoids2003, program chair of IEEE-IROS 2015, and editor-in-chief of *Frontiers in Neurorobotics*. He is a Senior Member of the IEEE.

References

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 × 128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008. doi: 10.1109/JSSC.2007.914337.

[2] S. Liu, B. Rueckauer, E. Ceolini, A. Huber, and T. Delbruck, "Event-driven sensing for efficient perception: Vision and audition algorithms," *IEEE Signal Process. Mag.*, vol. 36, no. 6, pp. 29–37, Nov. 2019. doi: 10.1109/MSP.2019.2928127.

[3] A. I. Maqueda, A. Loquercio, G. Gallego, N. García, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2018, pp. 5419–5427. doi: 10.1109/CVPR.2018.00568.

[4] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2032–2039, 2018. doi: 10.1109/LRA.2018.2800793.

[5] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. N. Clark, J. Dolan, D. Duggins et al., "Autonomous driving in urban environments: Boss and the urban challenge," *J. Field Robot.*, vol. 25, no. 8, pp. 425–466, 2008. doi: 10.1002/rob.20255.

[6] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. F. R. Jesus, R. F. Berriel et al., Self-driving cars: A survey. 2019. [Online]. Available: arXiv:1901.04407

[7] E. Guizzo, "How Google's self-driving car works," IEEE Spectrum, Oct. 18, 2011. [Online]. Available: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works

[8] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison et al., Event-based vision: A survey. 2019. [Online]. Available: arXiv:1904.08405

[9] M. A. Mahowald and C. Mead, "The silicon retina," *Sci. Amer.*, vol. 264, no. 5, pp. 76–82, 1991. doi: 10.1038/scientificamerican0591-76.

[10] H. Liu, C. Brandli, C. Li, S. Liu, and T. Delbruck, "Design of a spatiotemporal correlation filter for event-based sensors," in *Proc. 2015 IEEE Int. Symp. Circuits and Systems (ISCAS)*, pp. 722–725. doi: 10.1109/ISCAS.2015.7168735.

[11] V. Padala, A. Basu, and G. Orchard, "A noise filtering algorithm for event-based asynchronous change detection image sensors on TrueNorth and its implementation on TrueNorth," *Front. Neurosci.*, vol. 12, p. 118, 2018. doi: 10.3389/fnins.2018.00118.

[12] A. Khodamoradi and R. Kastner, "O(n)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors," *IEEE Trans. Emerg. Topics Comput.*, to be published. doi: 10.1109/TETC.2017.2788865.

[13] Y. Wang, B. Du, Y. Shen, K. Wu, G. Zhao, J. Sun, and H. Wen, "EV-gait: Event-based robust gait recognition using dynamic vision sensors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 6351–6360. doi: 10.1109/CVPR.2019.00652.

[14] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Advances Neural Information Processing Systems 28*, 2015, pp. 2017–2025.

[15] D. Gehrig, A. Loquercio, K. G. Derpanis, and D. Scaramuzza, "End-to-end learning of representations for asynchronous event-based data," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 5632–5642, doi: 10.1109/ICCV.2019.00573.

[16] W. Cheng, H. Luo, W. Yang, L. Yu, S. Chen, and W. Li, "DET: A high-resolution DVS dataset for lane extraction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019, pp. 1–10.

[17] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: Histograms of averaged time surfaces for robust event-based object classification," in *Proc. 2018 IEEE/Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2018, pp. 1731–1740. doi: 10.1109/CVPR.2018.00186.

[18] J. Binas, D. Neil, S. Liu, and T. Delbrück, "DDD17: End-to-end DAVIS driving dataset," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, 2017, pp. 1–9.

[19] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "HOTS: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1346–1359, July 2017. doi: 10.1109/TPAMI.2016.2574707.

[20] G. Haessig and R. Benosman, "A sparse coding multi-scale precise-timing machine learning algorithm for neuromorphic event-based sensors," in *Proc. Micro- and Nanotechnology Sensors, Systems, and Applications X*, 2018, vol. 10639, pp. 289–296. doi: 10.1117/12.2305933.

[21] J. H. Lee, T. Delbruck, and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *Front. Neurosci.*, vol. 10, p. 508, 2016. doi: 10.3389/fnins.2016.00508.

[22] E. Piatkowska, A. N. Belbachir, S. Schraml, and M. Gelautz, "Spatiotemporal multiple persons tracking using dynamic vision sensor," in *Proc. 2012 IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*, pp. 35–40. doi: 10.1109/CVPRW.2012.6238892.

[23] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman, "HFirst: A temporal approach to object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2028–2040, Jan. 2015. doi: 10.1109/TPAMI.2015.2392947.

[24] F. Folowosele, R. J. Vogelstein, and R. Etienne-Cummings, "Towards a cortical prosthesis: Implementing a spike-based HMAX model of visual object recognition in silico," *IEEE J. Emerg. Select. Topics Circuits Syst.*, vol. 1, no. 4, pp. 516–525, Dec. 2011. doi: 10.1109/JETCAS.2012.2183409.

[25] R. Xiao, H. Tang, Y. Ma, R. Yan, and G. Orchard, "An event-driven categorization model for AER image sensors using multispike encoding and learning," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: 10.1109/TNNLS.2019.2945630.

[26] J. Acharya, V. Padala, and A. Basu, "Spiking neural network based region proposal networks for neuromorphic vision sensors," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, May 2019, pp. 1–5. doi: 10.1109/ISCAS.2019.8702651.

[27] R. Benosman, C. Clercq, X. Lagorce, S. Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 407–417, Feb. 2014. doi: 10.1109/TNNLS.2013.2273537.

[28] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "EV-flownet: Self-supervised optical flow estimation for event-based cameras," in *Proc. Robotics: Science and System*, 2018, pp. 1–9. doi: 10.15607/RSS.2018.XIV.062.

[29] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 989–997. doi: 10.1109/CVPR.2019.00108.

[30] C. Ye, A. Mitrokhin, C. Parameshwara, C. Fermüller, J. A. Yorke, and Y. Aloimonos, Unsupervised learning of dense optical flow and depth from sparse event data. 2018. [Online]. Available: arXiv:abs/1809.08625

[31] N. F. Y. Chen, "Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion," in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 757–766. doi: 10.1109/CVPRW.2018.00107.

[32] J. Li, S. Dong, Z. Yu, Y. Tian, and T. Huang, "Event-based vision enhanced: A joint detection framework in autonomous driving," in *Proc. 2019 IEEE Int. Conf. Multimedia and Expo (ICME)*, pp. 1396–1401. doi: 10.1109/ICME.2019.00242.

[33] G. Chen, H. Cao, C. Ye, Z. Zhang, X. Liu, X. Mo, Z. Qu, J. Conradt et al., "Multi-cue event information fusion for pedestrian detection with neuromorphic vision sensors," *Front. Neurorobot.*, vol. 13, p. 10, Apr. 2019. doi: 10.3389/fnbot.2019.00010.

[34] A. Zanardi, A. Aumiller, J. Zilly, A. Censi, and E. Frazzoli, "Cross-modal learning filters for RGB-neuromorphic wormhole learning," in *Proc. 15th Robotics: Science and System XV*, June 24, 2019, p. P45.

[35] I. Alonso and A. C. Murillo, "EV-SegNet: Semantic segmentation for event-based cameras," *2019 IEEE/Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–10.

[36] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *Int. J. Comput. Vis.*, vol. 113, no. 1, pp. 1573–1405, 2015. doi: 10.1007/s11263-014-0788-3.

[37] B. Rueckauer, I.-A. Lungu, Y. Hu, M. Pfeiffer, and S.-C. Liu, "Conversion of continuous-valued deep networks to efficient event-driven networks for image classification," *Front. Neurosci.*, vol. 11, p. 682, Dec. 2017. doi: 10.3389/fnins.2017.00682.

[38] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going deeper in spiking neural networks: VGG and residual architectures," *Front. Neurosci.*, vol. 13, p. 95, Mar. 2019. doi: 10.3389/fnins.2019.00095.

[39] M. Pfeiffer and T. Pfeil, "Deep learning with spiking neurons: Opportunities and challenges," *Front. Neurosci.*, vol. 12, p. 774, Oct. 2018. doi: 10.3389/fnins.2018.00774.

[40] R. Wes Baldwin, M. Almatrafi, J. R. Kaufman, V. Asari, and K. Hirakawa, "Inceptive event time-surfaces for object classification using neuromorphic cameras," in *Proc. Int. Conf. Image Analysis and Recognition*, 2019, pp. 395–403. doi: 07/978-3-030-27272-2_35.

[41] H. Li and L. Shi, "Robust event-based object tracking combining correlation filter and CNN representation," *Front. Neurorobot.*, vol. 13, p. 82, Oct. 2019. doi: 10.3389/fnbot.2019.00082.

[42] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, "Interacting maps for fast visual interpretation," in *Proc. 2011 Int. Joint Conf. Neural Networks*, pp. 770–776. doi: 10.1109/IJCNN.2011.6033299.

[43] G. Munda, C. Reinbacher, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1381–1393, 2018. doi: 10.1007/s11263-018-1106-2.

[44] C. Scheerlinck, N. Barnes, and R. Mahony, "Continuous-time intensity estimation using event cameras," in *Computer Vision—ACCV 2018*. New York: Springer-Verlag, 2019, pp. 308–324.

[45] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, 1–23.

[46] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: An open event camera simulator," in *Proc. 2nd Conf. Robot Learning*, Oct. 29–31, 2018, vol. 87, pp. 969–982.

[47] S. Schraml, A. N. Belbachir, and H. Bischof, "Event-driven stereo matching for real-time 3D panoramic vision," in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 466–474. doi: 10.1109/CVPR.2015.7298644.

[48] D. Zhu, J. Dong, Z. Xu, C. Ye, Y. Hu, H. Su, Z. Liu, and G. Chen, "Neuromorphic visual odometry system for intelligent vehicle application with bio-inspired vision sensor," in *Proc. IEEE Int. Conf. Robotics and Biomimetics*, Sept. 2019, pp. 2225–2232. doi: 10.1109/ROBIO49542.2019.8961878.

[49] G. Chen, L. Hong, J. Dong, P. Liu, J. Conradt, and A. Knoll, "EDDD: Event-based drowsiness driving detection through facial motion analysis with neuromorphic vision sensor," *IEEE Sensors J.*, to be published. doi: 10.1109/JSEN.2020.2973049.

[50] S. Dong, Z. Bi, Y. Tian, and T. Huang, "Spike coding for dynamic vision sensor in intelligent driving," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 60–71, Feb. 2019. doi: 10.1109/JIOT.2018.2872984.

[51] G. Chen, J. Chen, M. Lienen, J. Conradt, F. Röhrbein, and A. C. Knoll, "FLGR: Fixed Length Gists Representation learning for RNN-HMM hybrid-based neuromorphic continuous gesture recognition," *Front. Neurosci.*, vol. 13, p. 73, Feb. 2019. doi: 10.3389/fnins.2019.00073.

[52] A. Zanardi, J. Zilly, A. Aumiller, A. Censi, and E. Frazzoli, "Wormhole learning," in *Proc. 2019 Int. Conf. Robotics and Automation (ICRA)*, Montreal, May 20–24, 2019, pp. 7899–7905. doi: 10.1109/ICRA.2019.8794336.

[53] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output," *Proc. IEEE*, vol. 102, no. 10, pp. 1470–1484, Oct. 2014. doi: 10.1109/JPROC.2014.2346153.

You Li and Javier Ibanez-Guzman

Lidar for Autonomous Driving

The principles, challenges, and trends for automotive lidar and perception systems

©ISTOCKPHOTO.COM/OONAL

Autonomous vehicles rely on their perception systems to acquire information about their immediate surroundings. It is necessary to detect the presence of other vehicles, pedestrians, and other relevant entities. Safety concerns and the need for accurate estimations have led to the introduction of lidar systems to complement camera- or radar-based perception systems. This article presents a review of state-of-the-art automotive lidar technologies and the perception algorithms used with those technologies. Lidar systems are introduced first by analyzing such a system's main components, from laser transmitter to beamscanning mechanism. The advantages/disadvantages and the current status of various solutions are introduced and compared. Then, the specific perception pipeline for lidar data processing is detailed from an autonomous vehicle perspective. The model-driven approaches and emerging deep learning (DL) solutions are reviewed. Finally, we provide an overview of the limitations, challenges, and trends for automotive lidars and perception systems.

Introduction

Autonomous driving is entering a preindustrialization phase, with significant progress attained over the past years. Sensors initially capture data representations of the environment, which are processed by perception algorithms to build the vehicle's immediate environment used for autonomous vehicle navigation. A perception system for autonomous vehicle navigation consists of a combination of active and passive sensors, namely, cameras, radars, and lidars [1]. Lidars are active sensors that illuminate the surroundings by emitting lasers. Ranges are measured precisely by processing the received laser returns from the reflecting surfaces. As stated in [2], lidars are "poised to significantly alter the balance in commercial, military, and intelligence operations, as radar has done over the past seven decades."

Despite much progress in camera-based perception, image processing methods estimate distances. This approach encounters difficulties when estimating distances for cross-traffic entities, particularly for monocular solutions. The 2007 DARPA

Digital Object Identifier 10.1109/MSP.2020.2973615
Date of current version: 26 June 2020

Grand Challenge, a milestone in autonomous driving, demonstrated the potential of lidar perception systems. The top three teams were all equipped with multiple lidars. A 64-layer lidar, the Velodyne HDL64 [55], played a critical central role for the winning and runner-up teams [3], [4].

Currently, many high-level autonomous vehicles use lidars as part of their perception systems despite their high cost and moving parts. A typical example is their incorporation into the autonomous vehicles being tested as robovehicles in different countries, e.g., the Eco-Mobility by Autonomous Vehicles in the Paris-Saclay area (EVAPS) field operational test in France [56], where several participants (e.g., Renault, Transdev, Vedecom, and so forth) collaborate to exploit mobility services based on autonomous vehicles in the Saclay area of Paris. One of the prototypes developed by Renault is shown in Figure 1. As a result, there are more than 20 companies developing distinctive lidar systems for autonomous driving systems, ranging from low level to high end, a sort of "big bang." Which lidar type(s) will dominate autonomous driving in the future remains to be seen.

On the other hand, lidar-based algorithms also entered into a fast track. For an autonomous vehicle, lidars are mainly used for perception and localization. Due to space constraints, this article focuses only on perception usages. In the context of autonomous driving, a perception system provides a machine-interpretable representation of the environment around the vehicle. From a user's perspective, the output of a perception system comprises the following three levels of information.

- *Physical description*: the pose, velocity, and shape of objects
- *Semantic description*: the categories of objects
- *Intention prediction*: the likelihood of an object's behavior.

Therefore, the lidar outputs are used for the object detection, classification, tracking, and intention prediction, corresponding to the various layers of information. Due to lidar's superiority in ranging accuracy, the provided physical information is highly reliable. Although the semantic information carried by lidar is more or less difficult than that which is acquired from a camera, a contextual sensor is good at object recognition. In practice, lidars are combined with cameras to complement each other [5]: a camera is poor in distance estimation, while lidar is inadequate for object recognition. Precise physical and semantic information, together with map information, will improve intention prediction without any doubts. With many years of progress, a lidar-centric perception system will mature for model-based processing algorithms; however, emerging DL methods are changing this domain. Traditional model-based lidar data processing methods are computation friendly and explicable. Data-driven DL methods have demonstrated extraordinary capabilities in providing semantic information, which is the weak point of traditional methods.

Lidar technologies

A typical lidar operates by scanning its field of view (FoV) with one or several laser beams. This is done through a delicately designed beamsteering system. The laser beam is generated by an amplitude-modulated laser diode that emits at near-infrared (NIR) wavelength. The laser beam is reflected by the environment back to the scanner, with the returned signal received by a photodetector. Fast electronics filter the signal and measure the difference between the transmitted and received signals, which are proportional to the distance. The range is estimated from the sensor model based on this difference. The difference in variations of reflected energy due to surface materials as well as state of the milieu between the transmitter and receiver are

FIGURE 1. The autonomous vehicle prototype developed by Groupe RENAULT for the EVAPS project. The most evident sensor is the Velodyne UltraPuck lidar on top.

compensated for through signal processing. The lidar outputs include 3D point clouds that correspond to the scanned environments and the intensities that correspond to the reflected laser energies. Figure 2 shows a conceptual representation of this operating principle.

A lidar system can be partitioned into the laser rangefinder system and the scanning system. The laser rangefinder comprises 1) the laser transmitter, which illuminates the target via a modulated wave; 2) the photodetector, which generates the electronic signal from the reflected photons after optical processing and photoelectric conversion; 3) the optics, which collimate the emitted laser and focus the reflected signal onto the photodetector; and 4) signal processing electronics, which estimate the distance between the laser source and the reflecting surface, based on the received signal. The scanning system typically will steer laser beams at different azimuths and vertical angles, denoted by ϕ_i, θ_i, where i is an index that determines the direction at which the beam is being pointed.

This section initially addresses the principles of a rangefinder to understand its measurement process and limitations, then, it introduces the scanning systems that define the sensor FoV. It is then possible to classify lidars based on the technologies they use. This classification is then applied to examine the commercially available automotive type lidars.

Laser rangefinder principles

A rangefinder that measures the distance to an object by a laser beam is known as a *laser rangefinder.* The manner in which they operate depends on the type of signal modulation used in the laser beam. Pulsed lasers are used so that their time of flight (ToF) can be measured, these are known as *direct-detection* laser rangefinders. The laser signal can also be a frequency-modulated continuous wave (FMCW), which indirectly measures the distance and velocity from the Doppler effect. These are known as *coherent-detection* laser rangefinders.

Lidar power equation

The transmitted laser is first attenuated through the transmission medium, then diffused as it reflects from the target surface. It is partially captured by the receiving optics and finally transformed into an electrical signal by a photodetector. For a target at distance r, the amount of received power P_r by the photodetector based from a pulsed-laser emitter can be approximately modeled as [6]

$$P_r = E_p \frac{c\eta A_r}{2r^2} \cdot \beta \cdot T_r, \tag{1}$$

where E_p is the total energy of a transmitted pulse laser and c is the light speed. A_r represents the area of receive aperture at range r, η is the overall system efficiency, and β is the reflectance of the target's surface, which is decided by the surface properties and the incident angle. In a simple case of Lambertian reflection with a reflectivity of $0 < \Gamma < 1$, it is given by $\beta = \Gamma/\pi$. The final part, T_r, denotes the transmission loss through the transmission medium. When the lidar works under adverse conditions (e.g., fog, rain, dust, snow, and so on), the particles in the air scatter and absorb the photons. Equation (1) reveals that the received power p_r decreases quadratically with respect to distance r: an object at hundreds of meters away is order of magnitudes "darker" than at tens of meters. Simply increasing the power of the laser transmitter is restricted by eye-safety standard IEC 60825 [7]. To overcome this, the overall system efficiency must be improved through optics, photodetectors, and more advanced signal processing algorithms. For an FMCW laser, (1) still holds, except for slight differences.

ToF

A ToF laser rangefinder measures the range by calculating the time difference between the transmitted and received lasers

$$r = \frac{1}{2n} c \Delta t, \tag{2}$$

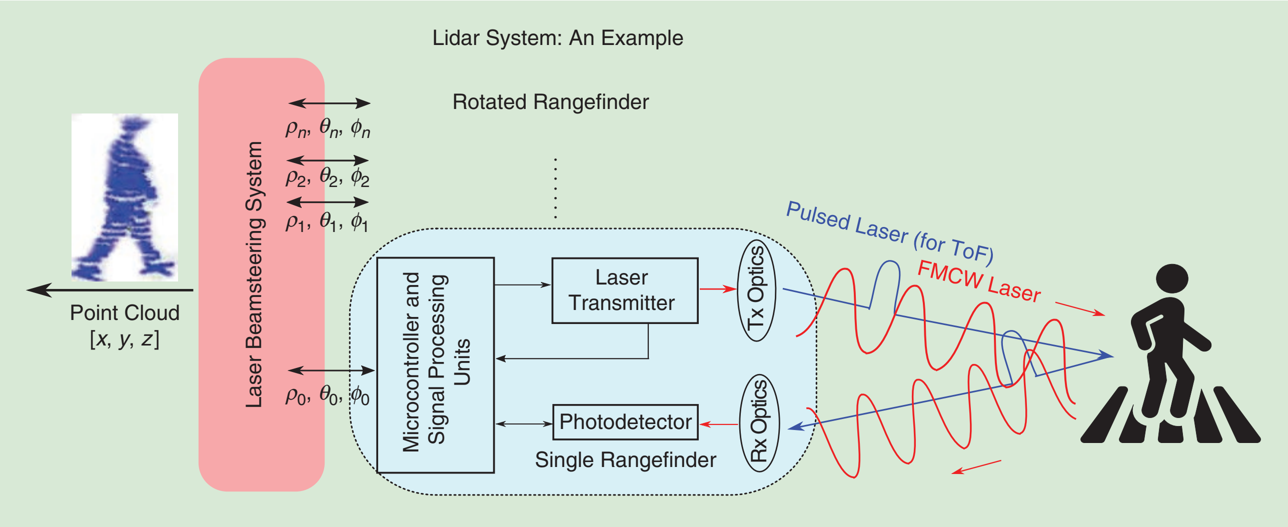


FIGURE 2. An example of a ToF laser rangefinder. The rangefinder uses either a direct or coherent method to measure the distance at a certain direction controlled by the scanning system. Tx: transmitter; Rx: receiver.

where c is the light speed and n is the index of refraction of the propagation medium ($\approx$ 1 for air). Δt is the time gap between the transmitted and received lasers. ToF lidars prevail in the current automotive lidar market because of their simple structure and signal processing methods. However, the potential for increasing their maximum range is constrained by limited transmit power due to eye-safety requirements. In ToF lidar, the return signal can be interference from strong sunlight or laser beams from other ToF lidars.

Coherent detection

By mixing the local carrier signal with the received signal, it is possible to demodulate the received signal and thus it is possible to obtain the phase and frequency shift of the laser signal, hence acquiring the distance and velocity from the reflecting surface. This can be regarded as the optical version of FMCW radars, which are popular in today's passenger vehicles' advanced driver assistant systems. FMCW lidars continuously emit a frequency-modulated laser signal (e.g., linearly chirped laser) to a target, while keeping a reference signal, also known as a *local oscillator*. The common modulation functions are sawtooth or triangle waves.

Because a FMCW lidar continuously illuminates objects using less emitted power for this purpose, thus complying with eye-safety requirements and opening the possibility to use more power to extend their FoV. The signals used for coherent detection are shown in Figure 3. The intermediate frequency [(IF) in red] can be generated by mixing the local oscillator signal (a linearly chirped triangular modulation function) from the laser transmitter (in green) with the laser signal reflected from observed surfaces (in blue). The processing of the resulting signal generates IF f_{if}, shown in red in Figure 3. By assuming that the Doppler frequency shift f_d is less than IF f_{if}, we have

$$f_{\text{if}} = \frac{4rB}{ct} = \frac{f_{\text{if}}^{+} + f_{\text{if}}^{-}}{2}, \quad f_d = \frac{f_{\text{if}}^{+} - f_{\text{if}}^{-}}{2} (\text{when} \quad f_d < f_{\text{if}}), \quad (3)$$

where B, r, and t are the modulation bandwidth, waveform period, and light speed, respectively. The velocity is obtained as

$$v = \frac{f_d \lambda}{2}, \quad (4)$$

where λ is the laser wavelength.

FMCW lidar is able to directly measure the distance and velocity at the same time, while for ToF lidar, the speed is obtained from indirect estimation through several consecutive sensor readings. By using a FMCW laser signal, it is possible to reduce the interference effect from other laser sources and strong sunlight; however, FMCW lidars require high-quality laser generators that possess long coherent distances.

Laser transmission and reception

The generation of the laser signals and their emission as well as the receiver electronics of the reflected signals also characterize the performance and cost of the laser rangefinders.

Laser sources

ToF lidars need a pulsed- (amplitude-modulated) laser signal, which is generated using a pulsed-laser diode or a fiber laser. A diode laser causes laser oscillation by flowing an electric current to the diode's junction. Diode lasers can be grouped into two classes: edge-emitting lasers (EELs) and vertical-cavity surface-emitting lasers (VCSELs). EELs have been applied in the telecommunication industry for a long time. VCSELs output a circular beam, while EELs transmit an elliptical laser beam, requiring additional beamshaping optics. In a VCSEL, forming a 2D laser array on a single chip is easier than for an EEL, which is important because it increases the lidar resolution. By contrast, the range for a VCSEL is shorter due to power limitations.

The pulsed-laser diodes used in automotive applications are hybrid devices; that is, a laser chip is mounted with capacitors that are triggered by a MOSFET transistor. Thus, at every gate opening, the electric charge accumulated in the capacitors will be discharged into the chip, which emits the optical pulse in a controlled manner. These sources are cost-effective, as their 905-nm output can be detected by economical silicon (Si) detectors; however, these diodes have a limited pulse-repetition rate and lower peak power, possibly requiring cooling. Laser diode sources for 3D flash lidar use diode stack technology, with several edge-emitting bars assembled into a vertical stack. Heat dissipation becomes an issue, hence the need for heat sinks as well as the accumulation of emitted power beyond eye-safe requirements. Fiber lasers can have higher output power, which is useful when operating at high wavelengths. Their output beams can be split and routed to multiple sensor locations using optical fiber, resulting in better pulse-repetition frequency, better beam quality, and so forth. However, they can be bulky and thus resulting in noncompact systems that are difficult to be integrated in vehicles.

Laser wavelength

Selecting an appropriate wavelength of laser should have a comprehensive consideration of atmospheric windows, eye-safety requirements, and cost. The 850–950-nm (NIR) and

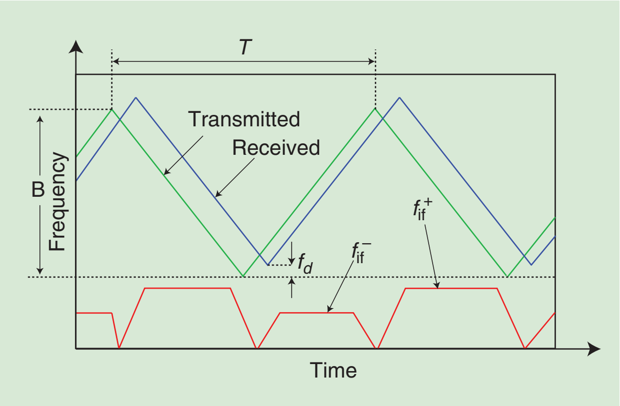


FIGURE 3. The principle of coherent detection: the distance is estimated by the intermediate frequency (red line) generated by mixing transmitted and received light waves [8].

1,550-nm [short-wave infrared (SWIR)] lasers are mostly utilized because of their popularity in industry. Either a low-price diode laser or a more powerful fiber laser at a wavelength of 850–950 nm or 1,550 nm is easily purchased from the market. The maximum power permitted by eye-safety standards for a 1,550-nm laser is higher than that of lasers in the 850–950-nm range, which means a larger range could be achieved. However, expensive indium gallium arsenide (InGaAs)-based photodiodes are required to detect laser returns at 1,550 nm. The efficiency of InGaAs-based photodiodes is lower than that of mature Si ones for NIR lasers. In addition, the atmospheric water absorption for 1,550 nm is stronger than that of 850–950 nm; therefore, lidar systems at NIR wavelengths (905 nm, for instance) are still commonplace.

Photo detector

A photo detector converts optical power to electrical power using the photoelectric effect. Photosensitivity that describes a photodetector's response when receiving photons is one of the most critical characteristics. The photosensitivity depends on the wavelength of the received laser; therefore, selecting a photodetector for a lidar system is closely related to the choice of laser wavelength. The most popular detectors are p-i-n photodiodes, avalanche photodiodes (APDs), single-photon avalanche diodes (SPADs), and Si photomultipliers (SiPMs).

P-i-n photodiodes

These diodes are formed by a p-i-n junction, which creates a depletion region free from mobile charge carriers. By applying a reverse bias to a photodiode, absorbing a photon will generate a current flow in the reverse-biased photodiode.

APD

An APD is a photodiode that applies reverse voltage to multiply photocurrent through the avalanche effect. The APD's ability to multiply signals reduces the effect of noise and achieves a higher internal current gain (roughly 100) and signal-to-noise ratio (SNR) than does the p-i-n photodiode. As a result, APDs are quite common in contemporary lidar systems. Si-based APDs are sensitive through the visible spectral region until the NIR is approximately 1,000 nm. At longer wavelengths up to 1,700 nm, InGaAs APDs are available, although at a higher cost.

SPAD

A SPAD is an APD that is designed to operate with a reverse-bias voltage above the breakdown voltage (Geiger mode), which allows the detection of very few photons in a very short time. SPADs can achieve a gain of 10^6, which is significantly higher than that of APDs; this characteristic allows the SPAD to detect extremely weak light at long distance. Furthermore, the CMOS technology used for a SPAD fabrication enables an integrated array of photodiodes on one chip. This is desirable for increasing lidar's resolution, while cutting the cost and power consumption.

SiPM

An SiPM is based on a SPAD and enables photon counting. The Geiger mode, in which a SPAD operates, is a photon-trigger mode, from which a SPAD cannot distinguish the magnitude of received photo flux. To overcome this issue, the SiPM integrates a dense array of "microcells" (a pair of SPADs and a quench resistor) working identically and independently. The SiPM's output is, in essence, a combination of the photocurrents detected from each microcell. Using this approach, an SiPM is capable of giving information on the magnitude of an instantaneous photon flux.

Scanning system

A scanning system (or beamsteering system) is designed to enable the transmitted lasers to rapidly explore a large area. The existing scanning approaches are usually classified as either mechanical spinning or solid state. The former usually contains a bulky rotating mirror system like the Velodyne HDL64 [55] used in the early stages of autonomous driving history. *Solid state* refers to a scanning system without moving parts (even though some are still steered by micromirrors), which is preferred by the automotive industry.

Mechanical spinning

Currently, the most popular scanning solution for automotive lidar is the mechanical spinning system [10], which steers the laser beams through a rotating assembly (e.g., mirror, prism, and so on) controlled by a motor to create a large FoV. Conventionally, nodding-mirror and polygonal-mirror systems [11] are the main types applied. For example, for the mechanical spinning scheme shown in Figure 4(a), an embedded nodding-mirror system tilts the lasers to generate a vertical FoV. Then, a 360° horizontal FoV is achieved by rotating the lidar base. State-of-the-art lidars use multiple beams to reduce the movable mechanism. For instance, the Velodyne VLP series uses arrays of laser diodes and photodiodes to increase point-cloud densities. The mechanical spinning system offers the advantage of a high SNR over a wide FOV; however, the rotating mechanism is bulky for integration inside a vehicle and is fragile in harsh conditions such as during vibration, which is quite common in automotive applications. A typical product example is Velodyne's HDL64.

Microelectromechanical systems microscanning

Microelectromechanical systems (MEMS) technology allows for the fabrication of miniature mechanical and electromechanical devices using Si fabrication techniques. In essence, a MEMS mirror is a mirror embedded on a chip [12]. The MEMS mirror is rotated by balancing between two opposing forces: an electromagnetic force (Lorentz force) produced by the conductive coil around the mirror and an elastic force from a torsion bar, which serves as the axis of rotation. This principle is shown in Figure 4(b). The MEMS mirrors can either be single axis for 1D movement [13] or dual axis for 2D movement. Also, a MEMS mirror can work in resonant mode at its characteristic oscillation frequency to obtain a large deflection angle and high operating frequency. In nonresonant mode, a MEMS scanning mirror can be controlled to follow a programmed scan trajectory. For example, for a MEMS-based AEye lidar, the lidar can dynamically change the FoV and scanning path to focus on some critical parts. Although MEMS lidars still contain moving parts, this near-solid-state technology is still promising because the mature techniques in the integrated circuit industry are able to satisfy the strict cost requirements.

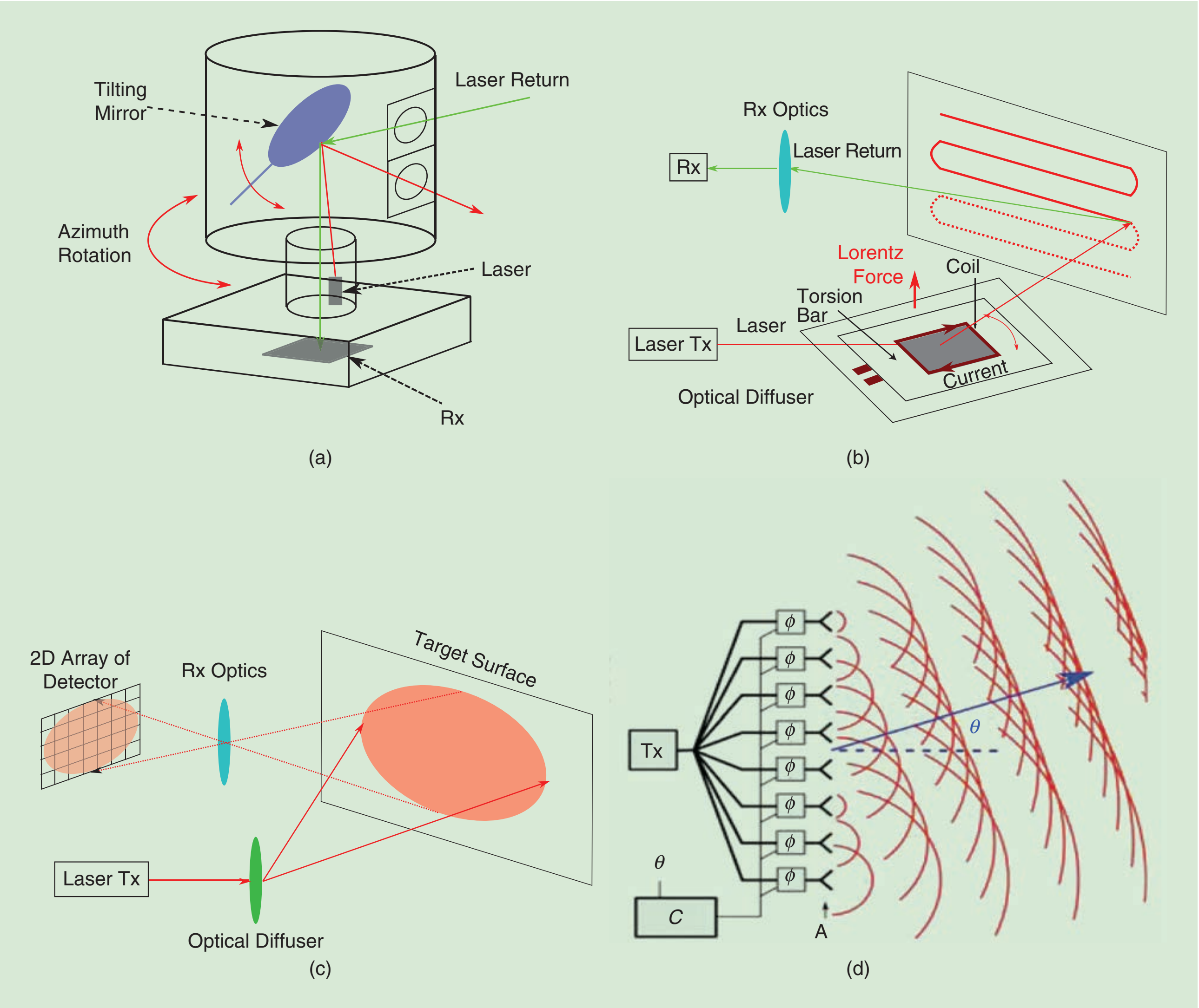
Flash

Originally applied for spacecraft in autonomous landing and docking with satellites, 3D flash lidars [14] totally remove the rotating parts within scanning systems. Hence, they are truly solid state. A flash lidar behaves as a camera, in that a single laser is spread by an optical diffuser to illuminate the whole scene at once. Then, it uses a 2D array of photodiodes (similar to the CMOS/charge-coupled device for the camera) to capture the laser returns, which are finally processed to form 3D point clouds, as shown in Figure 4(c).

Because all the pixels of flash lidar measure the ranges simultaneously, the issue of movement compensation caused by platform motion is avoided. In addition, the semiconductor-based 3D flash lidars facilitate fabrication and packaging for massive production, which leads to lower cost. However, the critical issue of 3D flash lidar is its limited detecting range (usually <100 m), because a single diffused laser is responsible for detecting the whole area under a small power threshold for eye safety. Another disadvantage is its limited FoV because it cannot rotate and scan the surroundings the way a scanning-type lidar does.

Optical phased array

As a type of true solid-state lidar, optical phased-array (OPA) lidars [15], [16] do not comprise moving components. Similar to phased-array radar, an OPA is able to steer the laser beams through various types of phase modulators. The speed of light can be changed using the optical phase modulators when the lasers are passing through the lens, as illustrated in Figure 4(d). Consequently, different light speeds in different paths allow for



(a) (b) (c) (d)

FIGURE 4. Lidar systems categorized by scanning approaches. (a) The principle of mechanical spinning lidar, (b) the principle of microelectromechanical systems lidar, (c) the principle of flash lidar, and (d) the principle of OPA lidar [9].

control of the optical wavefront shape and hence, the steering angles. Although OPA was once seen as a promising technology, there is not yet a commercial product in the market.

Current status of automotive lidar

Mechanical spinning lidar is the first venture into mass-produced cars. Announced in 2017, Audi released its latest luxury sedan, the A8, which equipped Valeo's SCALA lidar for automated driving functions, the first commercially available vehicle carrying automotive-grade lidar in the world. Valeo's SCALA [57] is a four-layer mechanical spinning lidar similar to its cousin, the IBEO Lux4. Empowered by SCALA, the A8 is able to achieve L3-level automated-driving functions without needing hands on the steering wheel (which requires legislative approval). In 2019, Valeo gained a €500 million order of its next-generation lidar, the SCALA2, from several car manufacturers.

At the same time, to reduce cost and improve robustness, many companies focus on solid-state scanning systems. As listed in Table. 1, Innoviz, Continental, and Quanergy are developing MEMS, flash, and OPA lidars, respectively. In 2018, Bavarian Motor Works announced a collaboration with Innoviz for series production beginning in 2021. To increase the maximum detection range, some in the industry have used SPAD arrays working in a single-photon detection mode (Geiger mode). The Ouster OS-1 64 [58] adopted CMOS-based SPADs to detect 850-nm lasers emitted by a 2D VCSEL array. Toyota made a lidar prototype containing a CMOS SPAD array (202 × 96 pixels) for receiving 905-nm lasers [17]. Princeton Lightwave, Inc. (acquired by Argo.ai) also realized a SPAD lidar prototype [59], although little information has been disclosed. As for SiPMs, a lidar prototype [18] was made by SensL (acquired by OmmniVision), while the commercial products are still under development.

Some companies have switched to SWIR lasers (e.g., 1,550 nm) because of its higher power threshold allowed than that of NIR lasers such as Luminar (who announced a collaboration with Toyota) and AEye. Coherent detection-based FMCW lidars are being targeted by car manufacturers and investors as well. Strobe and Blackmore, two representative FMCW lidar start-ups, were quickly acquired by Cruise and Aurora, respectively. In Table 1, we classify and list several representative automotive lidar suppliers and their disclosed technologies.

Another lidar trend is that of overcoming adverse weather conditions, such as rain, fog, snow, dust, and so forth [19], [20]. According to (1), adverse weather conditions increase the transmission loss T_r and weaken the reflectivity of an object β so that the received energy becomes less. Because SWIR lasers (e.g., 1,550 nm) can achieve higher transmission power, lidars belonging to this wavelength are expected to have better performances in harsh weather.

Lidar perception system

For an autonomous vehicle, its perception system classifies the perceived environment into hierarchical object descriptions (i.e., physical, semantic, and intention awareness) from the perception sensor outputs, its localization, and map data. As depicted in Figure 5, a traditional pipeline [4], [21] of processing lidar data consists of four steps: object detection, tracking, recognition, and motion prediction. The recent rise of DL technologies is changing this classic flow, and we will introduce it after the classic approaches. Due to the popularity of Velodyne lidars in research communities, the reviewed data processing methods are mainly based on this mechanical spinning lidar.

Object detection

Object-detection algorithms extract the object candidates and estimate their physical information, namely, the positions and shapes of the detected objects. Because in most traffic scenes the targets are perpendicular to flat ground, object-detection algorithms usually comprise ground filtering and clustering. Ground filtering labels a point cloud either ground or nonground. Then, nonground points are grouped into different objects using clustering methods.

In early research [4], the point clouds from lidar are projected into polar grids, subdividing 360° around the lidar. The points inside each grid cell are treated consecutively to generate a virtual scan, which specifies the region as free, occupied, and occluded. Occupied virtual scans are grouped into object clusters. The authors in [21] followed this method while utilizing a grid-based local plane-fitting approach instead of processing every point, as was done in [4]. The grids that are able to be fitted as a plane are classified as ground grids, and the remaining nonground grids are clustered by the connected component labeling. However, the polar grid-based methods always need a projection of 3D lidar points into discrete grids, which lose raw information from lidar measurements.

Processing lidar signals in spherical coordinates (r, φ, θ) provides a better approach. In this article, we used the Velodyne UltraPuck, where the vertical angle for each laser beam is fixed and the azimuth angle is decided by the scanning time and

Table 1. Representative lidar manufactures and the adopted technologies.

		Mechanical Spinning	MEMS	Flash	OPA	Undisclosed
ToF lidar	NIR	Velodyne, IBEO, Valeo, Ouster,* Hesai, and Robosense	Innoviz Robosense	Continental Xenomatix	Quanergy	
	SWIR	Luminar	AEye and Hesai	Argo* (Princeton Lightwave)		
FMCW lidar					Cruise (Strobe)	Aurora (Blackmore, 1,550 nm)

*The manufacturers that utilize single-photon Geiger-mode SPAD as a photodetector.

motor speed. Therefore, every range reading can be represented by $P_{i,j} = (\rho_{i,j}, \varphi_i, \theta_j,)$, where i refers to a certain laser beam and j is the azimuth-angle index, as shown in Figure 6. This approach naturally fills the range readings into a predefined data buffer (range image) and thus allows for fast access to a point and its neighbors. Processing lidar data in range view has become popular in recent years; for instance, based on a range image, the authors in [22] segmented the ground points in each column. The remaining nonground points are grouped easily through criterions of distance and angle. For a 32-beam lidar, they reached 4 ms in an Intel i5 processor.

Zermas et al. [23] processed the range image row by row. They applied the clustering in each scan line (actually, the row in the range image) and then merged the clusters scan line by

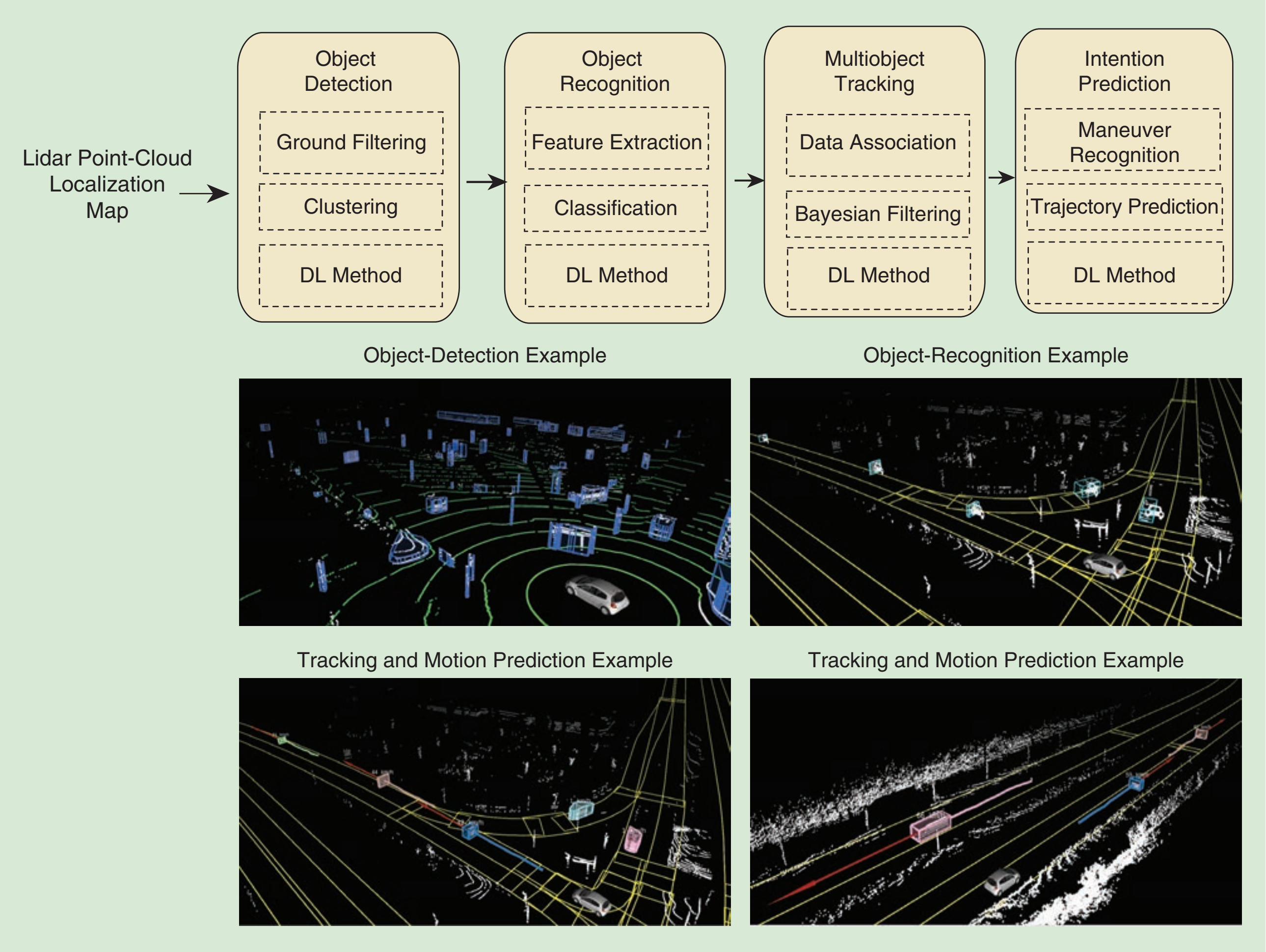


FIGURE 5. The pipeline of a classic lidar perception system, with the exemplary outputs of each step. The examples are from the authors' platform, as shown in Figure 1. Note that after object detection, we only process the objects within the road (as denoted by the yellow lines).

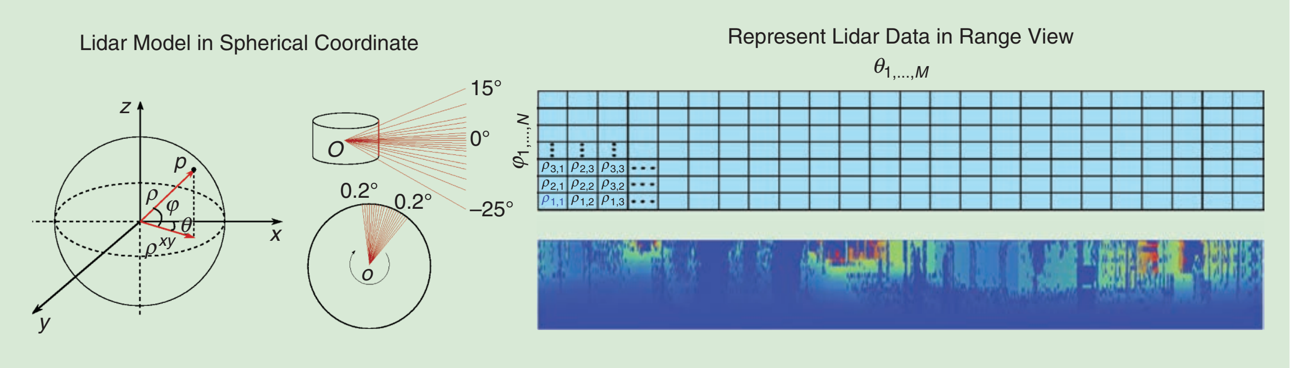


FIGURE 6. The range view of spinning lidar (Velodyne UltraPuck) for further processing. The range image (32 × 1,800) in pseudocolor facilitates the subsequent processing.

scan line. Figure 5 shows a sample result of ground filtering and clustering from our implementation based on the range image as well. The green lines are ground points, and the nonground points are grouped into object candidates (in blue polygons). Object detection provides the initial physical information, e.g., the position of an object. The following steps, such as recognition and tracking, complement semantic and more physical information, e.g., the heading and speed to the detected objects.

Object recognition

Machine learning (ML)-based object-recognition methods furnish the semantic information (e.g., the classes of pedestrian, vehicle, truck, tree, building, and so on) to the detected objects. A typical recognition process employed in [21] comprises a feature-extraction step, calculating compact object descriptors, and a classification step, where pretrained classifiers predict the categories of objects based on the extracted features. As summarized in [24], the features proposed in literature can be roughly divided into two classes: the global features for the whole object or the local features for each point. An object's size, radius, central moments, or maximum intensity [21], [25] are the most basic global features.

Applying principal component analysis (PCA) in 3D point clouds is another effective method used to acquire global shape features. As adopted in [26], three salience features (surfaceness, linearness, and scatterness) can be acquired by analyzing the eigenvalues acquired from the PCA. As for local features, the authors in [21] calculated the three salience features for each point and its neighboring points. Three histograms, each containing four bins spaced between 0 and 1 for the three salience features, are extracted as local features. A more complicated feature is the spin image (SI), introduced by Wang et al. [27]. An SI is created by spinning a grid around the normal surface n of a given point p. The virtual pixels of an SI are the distances to either line through the n or to the plane defined by p and n. The authors in [26] transformed this individual pointwise feature into a global feature: for an object, only the SI of its central point is utilized as an object descriptor. In the literature, there are more sophisticated features, such as a global Fourier histogram [24] descriptor; however, real-time requirements restrain the complexity of its features.

After feature extraction, the following classification is a typical supervised ML process: a classifier trained by a ground-truth data set predicts the class of input objects. Well-known data sets such as KITTI [60] provide abundant resources. From the arsenal of ML, plenty of ML tools such as naive Bayes [25], support vector machines (SVMs) [21], [24], [28], k-nearest neighbor (NN) [26], random forest (RF), and gradient boosting trees [29] can be applied. An SVM with a radial basis function (RBF) kernel is still the most popular method due to its speed and accuracy. Figure 5 shows the recognition results on the detected on-road objects based on our implementations (an SVM with an RBF kernel). Recently, Capellier et al. [30] applied an evidential neural network to classify the lidar object. Evidential classifiers can better handle the unknown classes that are frequently encountered in practice.

Object tracking

Multiple object-tracking (MOT) algorithms correlate and locate the detected/recognized objects through spatiotemporal consistency. MOT maintains the identities of detected objects and yields their physical states, e.g., trajectories, poses, and velocities. MOT is a classic engineering problem [31] that has been researched for a long time. A basic architecture mainly includes a single-object tracker, which "optimally" estimates the state of the tracked object, and data association, which assigns new detections to the trackers.

A single-object tracker models the movement as a dynamic state-space model and estimates its state under the Bayesian filtering framework. The Kalman filter (KF) family, the classic KF under Gaussian-linear assumption and its variants, the extended KF (EKF), and the unscented KF (UKF) make up the popular toolbox. The authors in [32] employed a KF with a constant velocity model to track lidar detections. As a nonlinear version of KF, EKF is utilized for lidar object tracking in [33]. Extending the single dynamic model to multiple maneuver models, the interacting multiple model (IMM) filter is able to handle more complicated cases. The IMM filter consists of several filters running in parallel, and each filter uses a different motion model. For a single object, an IMM-UKF filter was applied in [34], where three UKFs work for three motion models: constant velocity, constant turn rate, and random motion.

As another common approach, a particle filter (PF) is designed for more general cases that do not meet the Gaussian-linear assumption. The application of PFs in lidar data processing can be traced back to the 2017 DARPA Grand Challenge [4], where a Rao–Blackwellized PF was used. However, PFs require a large number of particles, especially for high-dimensional state space. Hence, the KF family is more popular in real-time perception systems.

Data association connects the detections with the tracks. The simplest method used is the NN filter (implemented in [32]), which assigns the detections to their closest tracks based on the Euclidean or Mahalanobis distance between the detection and the track. The NN filter is insufficient for clutter scenarios. In contrast, the joint probabilistic data association filter (JPDAF) offers a soft, probabilistic approach for detection-track association. JPDAF considers all of the possible detections (including no detection) in a gating window and estimates their assignment probabilities to the tracks, taking the weighted average of all the association hypotheses. In [34], JPDAF was applied for data association and an IMM-UKF filter was used to track an individual object.

In contrast to radar-based MOT, in which all of the detections are usually modeled as points, lidar-based MOT is distinctive in that it should track the shapes of detections as well. The simplest shape model is a 2D bounding box [4], which assumes that the detections are car-like objects. L-shape fitting [35] is the most common approach used to estimate the bounding box's center, width, height, and heading; however, a 2D bounding box is insufficient for more general objects, such as a pedestrian, tree, building, and so on. A more sophisticated method [36] implemented multiple shape models: points, polygons, L-shape, and lines for various objects. When tracking a moving object, its shape varies with the changes of pose and sensor viewpoint. Kraemer et al. [37] implemented a tracking method that simultaneously estimates the states of both poses and shapes represented by 2D polylines.

Object-intention prediction

The previously introduced modules provide the past and current information of detected targets. While in autonomous driving systems the decision-making and path-planning algorithms require future motion of the tracked targets, previous works based on certain kinematic models that are assumed to perfectly fit the detected objects are not applicable for long-term predictions. To address this shortcoming, maneuver or behavior recognition is proposed based on ML methods. Common maneuvers for a vehicle are cutting in, changing lanes, braking, overtaking, and so forth. The authors in [38] modeled the behaviors of car following and lane changing using a Gaussian mixture model (GMM) or hidden Markov model (HMM). Based on the maneuver classification realized by the HMM, the authors in [39] predicted the vehicles' motions by VGMMs (variational GMMs) under the constraints of vehicle-interaction models. With the success of recurrent neural networks (RNNs) in modeling temporal sequential data, long short-term memory (LSTM)-based methods are becoming more popular. Phillips et al. [40] used LSTM to classify the drivers' intentions at intersections, and the results show that LSTM outperforms other traditional ML methods. The authors in [41] proposed an encoder–decoder LSTM model to recognize maneuvers and predict trajectory. Beyond recognizing the maneuvers of a single object, social LSTM [42] was proposed to capture the interactions of all the objects. This is achieved by social pooling, which downsamples a target's neighboring object's LSTM states into a social tensor. The authors in [43] and [44] applied and improved the original social-pooling part for the purpose of vehicle-trajectory prediction.

Emerging DL methods

After huge successes in computer vision and speech recognition, waves of DL arrived in lidar data processing as well. DL [45] is a subset of ML algorithms that mainly uses multilayer neural networks. In contrast to traditional ML methods such as SVMs, DL technologies are able to automatically extract features from the raw input. Convolutional neural networks (CNNs) and RNNs such as LSTM are the most frequently used tools.

The basic components of perception system, ground segmentation, object detection, tracking, and recognition can all be realized using deep neural networks (DNNs). For instance, in [46], ground points were segmented by applying CNNs to lidar points represented by multichannel range images. In contrast to the object detection based on clustering in which arbitrary targets can be detected, DNN-based solutions achieve object detection using recognition, thanks to the paradigm of supervised learning. As in [47], vehicles can be detected using CNN-based neural networks in a bird's-eye view (BEV) representation of lidar points. A more complex neural network was proposed in [48], where CNNs are utilized in both a range image and BEV of lidar data and then fused with camera detections. However, due to the physical limitations of lidar, only vehicles can be effectively detected by lidar, with the best achieving results of only 52.4% average precision pedestrian detection in a KITTI benchmark of (using the DENFIDet method when writing this article) [61].

The authors in [49] integrated evidential theory into DL architecture for lidar-based road segmentation and mapping. Object tracking has been realized by DL as well. Different from the tracking-by-filtering framework described in traditional tracking algorithms, the authors in [50] proposed a deep structural model under tracking by a detection framework. A detection net first processes a sequence of lidar data and images to generate detection proposals. Then, tracks are estimated by finding the best associations of detections, which are achieved by a marching net and scoring net.

Apart from improving the traditional perception components, pointwise semantic segmentation, which was hard to realize before, is now achievable using deep learning. A well-known method, PointNet, was proposed in [52] to semantically segment 3D point clouds for general use. However, due to the sparsity of lidar data with respect to distance, the method does not work well for autonomous driving scenarios. SqueezeSeg [53] achieved real-time segmentation by applying CNNs in the range view of lidar points. Due to a lack of massive annotated data sets, the performances of these two methods are not ready for deployment in real situations, although this situation has been changed by SemanticKITTI [54], [62], the latest and biggest pointwise annotated data set based on KITTI. Based on this data set, RangeNet [51] demonstrated a fascinating performance and speed from a simple DNN structure. Figure 7 shows its structure and a sample result of RangeNet. With more annotated data sets, we have sufficient reason to expect lidar-based semantic segmentation will have better performance.

Conclusions and future directions

In this article, a review of lidar technologies was presented. How a lidar "sees" the world and what constitutes a lidar were introduced, and the main development directions of lidar technologies were analyzed as well. In summary, current automotive lidars face the following constraints or challenges: 1) cost; 2) meeting automotive reliability and safety standards (e.g., ISO 26262 and IEC 61508); 3) a long measuring distance (e.g., >200 m for highway applications); 4) adverse weather conditions, e.g., rain, fog, snow, and so on; 5) image-level resolution; and 6) its smaller size, which facilitates integration. At present, all of the possible solutions, varying from laser sources (905 nm versus 1,550 nm), scanning methods (spinning/MEMS/OPA/flash), or ranging principles (ToF or FMCW) have been exploited to overcome several or all of these difficulties. Although it is very difficult to predict which automotive lidar solution(s) will dominate in the future, one thing is certain: automotive lidars are walking out of experimental platforms, entering more and more mass-produced cars.

Then, a compact tutorial of the lidar-based perception systems for autonomous driving was presented. Three levels of

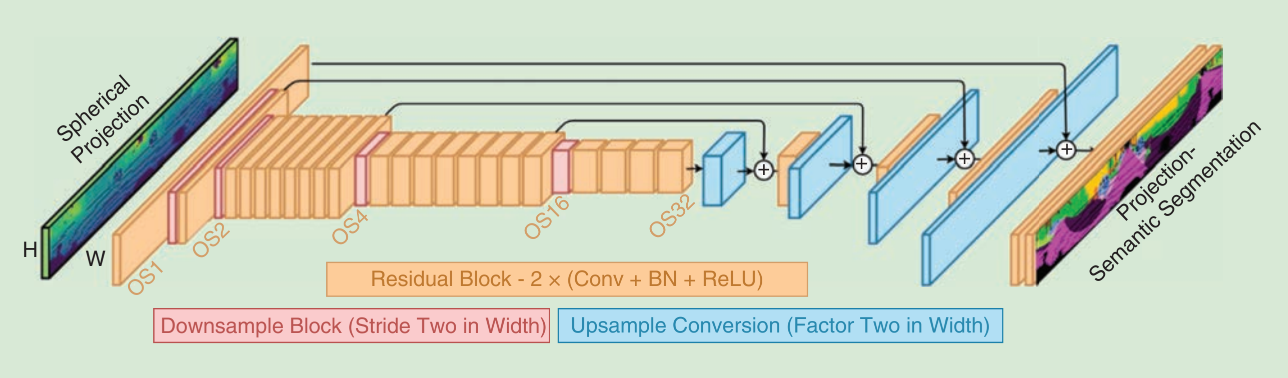


FIGURE 7. The neural network structure proposed in RangeNet [51]. Conv: conventional; ReLU: rectified linear unit; BN: batch normalization.

information providing by-perception systems along with typical processing pipelines were introduced. Generally, compared to cameras or radar, lidar is the most precise sensor used to measure range; therefore, the physical information (the objects' positions, headings, shapes, and so on) evaluated by lidar-based algorithms is highly reliable. However, semantic description is the shortcoming of lidar; this is caused by the lidar's poor resolution and its essence as a distance-measuring sensor, not a contextual sensor. Its fusion with cameras remedies lidar's weakness in recognition.

The intention-prediction level is independent of specific sensors, although it is strengthened by the precise physical information brought by lidar. Applying DL in lidar's 3D data will be one of the most important directions in the future. Lacking a huge number of annotated 3D point-cloud data sets was the bottleneck for successfully applying DL methods; however, things are changing. The aforementioned SemanticKITTI initiated a good start, and the results achieved by RangeNet++ are quite impressive. From our point of view, the algorithms used for extracting more accurate physical information and squeezing lidar's potential in semantic estimation are future directions. And, of course, with the swift progress of new lidars, new algorithms adapted to specific lidars will emerge.

Authors

You Li (you.li@renault.com) received his B.S. degree from Nanjing University of Science and Technology, China, in 2006, and his M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2010. He completed his Ph.D. thesis on machine perception applied to intelligent vehicles at the Université de Technologie de Belfort Montbéliard, France, in 2013. Since 2017, he has been a senior research engineer in the research division of Groupe RENAULT, Guyancourt, France. Since 2014, he has worked as a postdoctoral researcher at the French National Center for Scientific Research, studying unmanned surface vehicles. In 2016, he joined The Netherlands Organisation for Applied Scientific Research as a research scientist working on sensor fusion for autonomous vehicles. His main research interest is lidar-based perception systems for autonomous vehicles.

Javier Ibanez-Guzman (javier.ibanez-guzman@renault.com) received his M.S.E.E. degree from the University of Pennsylvania, Philadelphia, as a Fulbright Scholar and his Ph.D. degree from the University of Reading, Pennsylvania, on an SERC-UK fellowship. In 2011, he was a visiting scholar at the University of California, Berkeley (CITRIS), working on connected vehicle applications. He is currently a corporate expert in autonomous systems at Groupe RENAULT, Guyancourt, France, as well as a member of the technical staff at Renault S.A., working on autonomous vehicle navigation technologies and driving assistance systems. Previously, he was a senior scientist at a national research institute in Singapore, where he spearheaded work on autonomous ground vehicles operating in unstructured environments. He has several publications and patents in the robotics and automotive domains and has successfully supervised several Ph.D. students. He is a Chartered Engineer and a fellow of the Institute of Engineering Technology. He is a Member of the IEEE.

References

[1] J. L. Leonard, "A perception-driven autonomous urban vehicle," *J. Field Robot.*, vol. 25, no. 10, pp. 727–774, 2008. doi: 10.1002/rob.20262.

[2] National Research Council. *Laser Radar: Progress and Opportunities in Active Electro-Optical Sensing*. Washington, D.C.: National Academies Press, 2014.

[3] C. Urmson, A. Joshua, B. Drew, B. Christopher, B. Robert, M. N. Clark, J. Dolan, D. Duggins et al., "Autonomous driving in urban environments: Boss and the urban challenge," *J. Field Robot.*, vol. 25, no. 8, pp. 425–466, 2008.

[4] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Auton. Robots*, vol. 26, nos. 2–3, pp. 123–139, 2009. doi: 10.1007/s10514-009-9115-1.

[5] Y. Li, "Stereo vision and LIDAR based dynamic occupancy grid mapping: Application to scenes analysis for intelligent vehicles," Ph.D. dissertation, Univ. Technol. Belfort-Montbeliard, 2013.

[6] U. Wandinger, "Introduction to LiDAR," in *Springer Series in Optical Sciences*, C. Weitkamp Ed. New York: Springer-Verlag, vol. 102, 2005.

[7] *Interpretation sheet 1-Safety of laser products—Part 1: Equipment classification and requirements*. IEC Standard 60825-1, 2017.

[8] D. F. Pierrottet, F. Amzajerdian, L. Petway, B. Barnes, G. Lockard, and M. Rubioand, "Linear FMCW laser radar for precision range and vector velocity measurements," *MRS Proc.*, vol. 1076, 2008.

[9] S. Piatek, "Lidar and other techniques," Hamamatsu Photonics, Japan, Tech. Rep., 2017. [Online]. Available: https://www.hamamatsu.com/sp/hc/osh/lidar_webinar_12.6.17.pdf

[10] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Mach. Vis. Appl.*, vol. 27, no. 7, pp. 1005–1020, 2016. doi: 10.1007/s00138-016-0784-4.

[11] J. O'Neill, W. T. Moore, K. Williams, and R. Bruce, "Scanning system for LiDAR," U.S. Patent 8 072 663B2, 2011.

[12] H. W. Yoo and G. Schitter, "MEMS-based LiDar for autonomous driving," *Elektrotech. Inftech.*, vol. 135, no. 6, pp. 408–415, 2018. doi: 10.1007/s00502-018-0635-2.

[13] N. Druml, I. Maksymova, T. Thomas, D. Lierop, M. Hennecke, and A. Foroutan, “1D MEMS micro-scanning LiDar,” in *Proc. 9th Int. Conf. Sensor Device Technologies and Applications*, Venice, Italy, 2018.

[14] F. Amzajerdian, V. E. Roback, A. Bulyshev, P. F. Brewster, and G. D. Hines, “Imaging flash LiDar for autonomous safe landing and spacecraft proximity operation,” in *Proc. AIAA SPACE*, Long Beach, CA, 2016. doi: 10.2514/6.2016-5591.

[15] C. V. Poulton, M. J. Byrd, E. Timurdogan, P. Russo, D. Vermeulen and M. R. Watts, “Optical phased arrays for integrated beam steering,” in *Proc. 2018 IEEE 15th Int. Conf. Group IV Photonics (GFP)*, pp. 1–2. doi: 10.1109/GROUP4.2018.8478729.

[16] P. F. McManamon, T. A. Dorschner, D. L. Corkum, L. J. Friedman, D. S. Hobbs, M. Holz, S. Liberman, H. Q. Nguyen et al., “Optical phased array technology,” *Proc. IEEE*, vol. 84, no. 2, pp. 268–298, 1996. doi: 10.1109/5.482231.

[17] I. Takai, H. Matsubara, M. Soga, M. Ohta, M. Ogawa, and T. Yamashita, “Single-photon avalanche diode with enhanced NIR-sensitivity for automotive LIDAR systems,” *Sensors (Basel)*, vol. 16, no. 4, pp. 459–468, 2016. doi: 10.3390/s16040459.

[18] S. Gnecchi and C. Jackson, “A 1×16 SiPM array for automotive 3D imaging LiDAR systems,” in *Proc. Int. Image Sensor Workshop (IISW)*, Hiroshima, Japan, 2017, pp. 133–136.

[19] S. Michaud, J.-F. Lalonde, and P. Giguère, “Towards characterizing the behavior of LiDARs in snowy conditions,” in *Proc. IEEE/RSJ IROS Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, 2015.

[20] M. Kutila, P. Pyykonen, H. Holzhuter, M. Colomb, and P. Duthon, “Automotive LiDAR performance verification in fog and rain,” in *Proc. Int. Image Sensor Workshop (IISW)*, pp. 1695–1701. doi: 10.1109/ITSC.2018.8569624.

[21] M. Himmelsbach, A. Müller, T. Lüttel, and H.-J. Wünsche, “LIDAR based 3D object perception,” in *Proc. 1st Int. Workshop on Cognition for Technical Systems*, 2008.

[22] I. Bogoslavskyi and C. Stachniss, “Fast range image-based segmentation of sparse 3D laser scans for online operation,” in *Proc. 2016 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 163–169. doi: 10.1109/IROS.2016.7759050.

[23] D. Zermas, I. Izzat, and N. Papanikolopoulos, “Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications,” in *Proc. 2017 IEEE Int. Conf. Robotics and Automation (ICRA)*, Singapore, pp. 5067–5073. doi: 10.1109/ICRA.2017.7989591.

[24] T. Chen, B. Dai, D. Liu, and J. Song, “Performance of global descriptors for velodyne-based urban object recognition,” in *Proc. 2014 IEEE Intelligent Vehicles Symp.*, pp. 667–673. doi: 10.1109/IVS.2014.6856425.

[25] C. Premebida, O. Ludwig, and U. Nunes, “Exploiting LIDAR-based features on pedestrian detection in urban scenarios,” in *Proc. 2009 IEEE Int. Conf. Intelligent Transportation Systems (ITSC)*, pp. 1–6. doi: 10.1109/ITSC.2009.5309697.

[26] B. Douillard, J. Underwood, V. Vlaskine, A. Quadros, and S. Singh, “A pipeline for the segmentation and classification of 3D point clouds,” in *Experimental Robotics (Springer Tracts in Advanced Robotics)*, O. Khatib, V. Kumar, and G. Sukhatme, Eds. Berlin, Heidelberg: Springer-Verlag, 2014.

[27] A. E. Johnson and M. Hebert, “Using spin images for efficient object recognition in cluttered 3D scenes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, 1999. doi: 10.1109/34.765655.

[28] D. Z. Wang, I. Posner, and P. Newman, “What could move? Finding cars, pedestrians and bicyclists in 3D laser data,” in *Proc. 2012 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4038–4044. doi: 10.1109/ICRA.2012.6224734.

[29] Y. Li and Y. Ruichek, “Moving objects detection and recognition using sparse spacial information in urban environments,” in *Proc. 2012 IEEE Intelligent Vehicles Symp.*, pp. 1060–1065. doi: 10.1109/IVS.2012.6232205.

[30] E. Capellier, F. Davoine, V. Cherfaoui, and Y. Li, “Evidential deep learning for arbitrary LIDAR object classification in the context of autonomous driving,” in *Proc. 2019 IEEE Intelligent Vehicles Symp.*, pp. 1304–1311. doi: 10.1109/IVS.2019.8813846.

[31] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans, *Fundamentals of Object Tracking.* Cambridge, U.K.: Cambridge Univ. Press, 2011.

[32] F. Moosmann and C. Stiller, “Joint self-localization and tracking of generic objects in 3D range data,” in *Proc. 2013 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 1146–1152. doi: 10.1109/ICRA.2013.6630716.

[33] H. Cho, Y. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, “A multi-sensor fusion system for moving object detection and tracking in urban driving environments,” in *Proc. 2014 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 1836–1843. doi: 10.1109/ICRA.2014.6907100.

[34] A. A. Rachman, “3D-LIDAR multi object tracking for autonomous driving,” Master’s thesis, Delft Univ. Technol., Delft, The Netherlands, 2017.

[35] X. Zhang, W. Xu, C. Dong, and J. M. Dolan, “Efficient L-shape fitting for vehicle detection using laser scanners,” in *Proc. 2017 IEEE Intelligent Vehicles Symp. (IV)*, pp. 54–59. doi: 10.1109/IVS.2017.7995698.

[36] T. Johansson, “LiDAR clustering and shape extraction for automotive applications,” Master’s thesis, Chalmers Univ. Technol., Gothenburg, Sweden, 2017.

[37] S. Kraemer, C. Stiller, and M. E. Bouzouraa, “Lidar based object tracking and shape estimation using polylines and free-space information,” in *Proc. 2018 IIEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 4515–4522. doi: 10.1109/IROS.2018.8593385.

[38] C. Miyajima and K. Takeda, “Driver-behavior modeling using on-road driving data: A new application for behavior signal processing,” *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 14–21, 2016. doi: 10.1109/MSP.2016.2602377.

[39] N. Deo, A. Rangesh, and M. M. Trivedi, “How would surround vehicles move? A unified framework for maneuver classification and motion prediction,” *IEEE Trans. Intell. Veh.*, vol. 3, no. 2, pp. 129–140, 2018. doi: 10.1109/TIV.2018.2804159.

[40] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, “Generalizable intention prediction of human drivers at intersections,” in *Proc. 2017 IEEE Intelligent Vehicles Symp. (IV),* pp. 1665–1670. doi: 10.1109/IVS.2017.7995948.

[41] N. Deo and M. M. Trivedi, “Multi-modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs,” in *Proc. 2018 IEEE Intelligent Vehicles Symp. (IV)*, pp. 1179–1184. doi: 10.1109/IVS.2018.8500493.

[42] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human trajectory prediction in crowded spaces,” in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971. doi: 10.1109/CVPR.2016.110.

[43] N. Deo and M. M. Trivedi, “Convolutional social pooling for vehicle trajectory prediction,” in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1468–1476. doi: 10.1109/CVPRW.2018.00196.

[44] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, “Non-local social pooling for vehicle trajectory prediction,” in *Proc. 2019 IEEE Intelligent Vehicles Symp. (IV),* pp. 975–980. doi: 10.1109/IVS.2019.8813829.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA: MIT Press, 2016.

[46] M. Velas, M. Spanel, M. Hradis, and A. Herout, “CNN for very fast ground segmentation in velodyne LiDAR data,” in *Proc. 2018 IEEE Int. Conf. Autonomous Robot Systems and Competitions (ICARSC)*, pp. 97–103. doi: 10.1109/ICARSC.2018.8374167.

[47] B. Yang, W. Luo, and R. Urtasun, “Pixor: Real-time 3D object detection from point clouds,” in *Proc. 2018 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 7652–7660. doi: 10.1109/CVPR.2018.00798.

[48] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3D object detection network for autonomous driving,” in *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 6526–6534. doi: 10.1109/CVPR.2017.691.

[49] E. Capellier, F. Davoine, V. Cherfaoui, and Y. Li, “Transformation-adversarial network for road detection in LIDAR rings, and model-free evidential road grid mapping,” in *Proc. 11th PPNIV - IROS Workshop on Planning, Perception, Navigation for Intelligent Vehicle*, 2019.

[50] D. Frossard and R. Urtasun, “End-to-end learning of multi-sensor 3D tracking by detection,” in *Proc. 2018 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 635–642. doi: 10.1109/ICRA.2018.8462884.

[51] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, “RangeNet++: Fast and accurate LiDAR semantic segmentation,” in *Proc. 2019 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS),* pp. 4213–4220. doi: 10.1109/IROS40897.2019.8967762.

[52] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3D classification and segmentation,” in *Proc. 2017 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85. doi: 10.1109/CVPR.2017.16.

[53] B. Wu, A. Wan, X. Yue, and K. Keutzer, “SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud,” in *Proc. 2018 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 1887–1893. doi: 10.1109/ICRA.2018.8462926.

[54] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, “SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019, pp. 9297–9307.

[55] “HDL-64-E,” Velodyne Lidar, San Jose, CA. [Online]. Available: https://www.velodynelidar.com/hdl-64e.html

[56] “EVAPS,” Mov’eo, Rouen. [Online]. Available: https://pole-moveo.org/projets/evaps/

[57] “Valeo Scala,” Valeo, Paris. [Online]. Available: https://www.valeo.com/en/valeo-scala/

[58] A. Pacala, “How multi-beam flash lidar works,” Ouster, San Francisco, Nov. 8, 2018. [Online]. Available: https://www.ouster.io/blog-posts/2018/11/8/how-multi-beam-flash-lidar-works

[59] Youtube, “GeigerCruizer demo system images to 300 meters.” [Online]. Available: https://www.youtube.com/watch?v=K2YkrFSJc3A

[60] The KITTI Vision Benchmark Suite. [Online]. Available: http://www.cvlibs.net/datasets/kitti/

[61] The KITTI Vision Benchmark Suite, “Bird’s eye view evaluation 2017.” [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=bev

[62] Semantic KITTI. [Online]. Available: http://www.semantic-kitti.org/

SP

Joshua Rapp, Julián Tachella, Yoann Altmann,
Stephen McLaughlin, and Vivek K Goyal

Advances in Single-Photon Lidar for Autonomous Vehicles

Working principles, challenges, and recent advances

©ISTOCKPHOTO.COM/OONAL

The safety and success of autonomous vehicles (AVs) depend on their ability to accurately map and respond to their surroundings in real time. One of the most promising recent technologies for depth mapping is single-photon lidar (SPL), which measures the time of flight of individual photons. The long-range capabilities (kilometers), excellent depth resolution (centimeters), and use of low-power (eye-safe) laser sources renders this modality a strong candidate for use in AVs. While presenting unique opportunities, the remarkable sensitivity of single-photon detectors introduces several signal processing challenges. The discrete nature of photon counting and the particular design of the detection devices means the acquired signals cannot be treated as arising in a linear system with additive Gaussian noise. Moreover, the number of useful photon detections may be small despite a large data volume, thus requiring careful modeling and algorithmic design for real-time performance. This article discusses the main working principles of SPL and summarizes recent advances in signal processing techniques for this modality, highlighting promising applications in AVs as well as a number of challenges for vehicular lidar that cannot be solved by better hardware alone.

Introduction

Humans are not particularly good drivers. When operating multiton vehicles at high speeds, any impairment, fatigue, or other diversion of the driver's attention can lead to catastrophic results. More than 35,000 deaths and 2.4 million injuries resulted from automotive accidents in 2015 in the United States alone [1]. AVs thus offer the promise of a revolution in transportation, with the significant potential benefit of improved safety on the road. Countless other aspects of society will likewise be affected by no longer needing human drivers to control vehicles. Mobility can be improved for those unable to drive themselves due to age, inexperience, medical conditions, or lack of confidence. The economics of vehicle ownership will likely be transformed as humans transition from operators to passengers.

Before AVs can reshape the transportation landscape, numerous challenges have to be addressed. Engineers must first

Digital Object Identifier 10.1109/MSP.2020.2983772
Date of current version: 26 June 2020

1053-5888/20©2020IEEE

demonstrate the ability of AVs to perform basic driving functions safely and reliably. While driving, AVs must translate high-level goals—such as route planning, safe driving, efficient energy consumption, and adherence to rules of the road—into low-level decisions about the mechanics required to follow a planned trajectory. To achieve these goals, the complex control systems determining driving decisions must rely on information from sensing systems that are as good as, or preferably better than, human perception.

Autonomous system designs usually propose some form of sensor fusion, combining the strengths of various sensing modalities to overcome their individual limitations [2], [3]. In addition to classical GPS receivers, future commercial AVs are likely to include video cameras for identifying road signs and objects, short-range ultrasound sensors as currently used for parking assistance, and weather-robust radar for low-resolution position and velocity estimation. However, the centerpiece of most AV sensing systems is the lidar unit. Just as radar detects echoes from radio-frequency electromagnetic waves, lidar detects reflections from optical-frequency laser illumination to generate a long-range, high-resolution point cloud corresponding to the positions and reflectivity values of millions of points in the surrounding environment. Since optical wavelengths can be easily focused into narrow beams, lidar can distinguish much smaller objects than radar, which is crucial for navigating alongside pedestrians, cyclists, and other potential hazards. Although laser ranging has been under development since the 1960s, mostly for military use, terrain mapping, and atmospheric monitoring [4], commercial lidar development has greatly accelerated since 2005, when all vehicles that completed the DARPA autonomous driving Grand Challenge employed lidar for depth mapping.

More than 35,000 deaths and 2.4 million injuries resulted from automotive accidents in 2015 in the United States alone.

Sensing systems for AVs are evaluated on several criteria considered necessary for reliable real-time performance, including the maximum operating range, transverse and longitudinal resolution, field of view, refresh rate, transmit power (especially with respect to eye safety), robustness to ambient light and weather conditions, processing requirements, and cost [2], [5]. Many of these factors are determined, in part, by hardware and manufacturing constraints, such as the quality and capability of lasers, detectors, and scanning mechanisms, which have improved through continuous investment and refinement. For instance, the cost of a lidar unit was long considered a barrier to widespread deployment in AVs since the Velodyne systems originally used in the Grand Challenge cost upward of US$75,000 [6], but after more than a decade of development, lidar manufacturers such as Luminar have announced units priced at less than US$1,000.

On the other hand, the maximum range is one of the most crucial aspects that cannot be solved through hardware alone. The range capability is important because it affects driving performance directly: early identification of a potential hazard gives the vehicle more time to make safe driving decisions. At American highway speeds of approximately 105 km/h, for instance, a range extension of 30 m provides an additional second of reaction time. The baseline performance necessary for AVs is generally listed as the ability to detect dark objects with a reflectivity of 10% from a distance of at least 200 m. Amplitude-modulated continuous-wave and frequency-modulated continuous-wave lidar methods have fundamental challenges restricting the maximum range of most implementations to fewer than 100 m [5]. Although the maximum range of pulsed lidar can be extended by simply increasing the laser power, the vehicle power and eye-safety constraints limit the feasibility of this approach.

The rapidly evolving technology of SPL has the potential to overcome the range problem. Instead of trying to sample the full-waveform transient response from a pulsed lidar illumination, SPL uses detectors that are sensitive to individual photons, enabling the sensing of objects from extremely weak reflections. Such extreme sensitivity enables SPL to use eye-safe lasers and fast acquisitions while tolerating large attenuations due to very long distances or fog. For example, experiments with SPL have been used to form point clouds from standoff distances of several hundred meters up to 45 km [7], [8]. SPL is currently under commercial development for AVs by companies such as Ouster and Argo AI, which has partnered with both Ford and Volkswagen.

Beyond the specialized hardware required to detect and time stamp individual photons, imaging with so little signal relies heavily on advanced signal processing techniques tailored to the acquisition system. Importantly, SPL systems do not conform to standard signal processing assumptions: SPL measurements are neither linear nor time-invariant, and the inherent randomness of the observations cannot be simply described by Gaussian noise models. In 2014, Kirmani et al. [9] introduced the concept of photon-efficient imaging, showing that it is possible to form accurate 3D and reflectivity maps from only one photon detection for each transverse pixel location, even with half of the detected photons due to uninformative ambient light. From a computational viewpoint, successful point cloud reconstruction from very few photons requires two main components:

1) an acquisition system model, used to determine the probabilistic nature of any photon detection time
2) a scene model, employed to reconstruct a point cloud by taking advantage of useful priors.

This article focuses on these two components in the context of SPL for AVs. We begin with a description of the physical hardware, which is used to inform the acquisition system model. We then survey a number of challenges for vehicular lidar that cannot be solved by better hardware alone, including low photon counts, obscuring media, partial occlusions, and so forth, and we present state-of-the-art signal processing solutions. Finally, we look ahead to functionalities that can be developed for SPL beyond the capabilities of the human visual system, such as anticipatory route planning enabled by seeing around corners.

Basics of SPL systems

SPL comprises three main components: an illumination source, a singe-photon detector, and fast timing electronics. Diode lasers are relatively inexpensive to manufacture and can achieve root-mean-square pulse widths on the order of a few dozens of picoseconds. The most common choice of detector for SPL is the solid-state single-photon avalanche diode (SPAD), which consists of a reverse-biased photodiode biased above the breakdown voltage so that an individual photon incident on the SPAD can cause an avalanche of electrical charge carriers that is directly detectable as a digital signal.

SPL is performed using a technique called *time-correlated single-photon counting* (*TCSPC*), originally developed for fluorescence lifetime imaging microscopy. The basic idea of TCSPC is that of a stopwatch: the laser starts a timer with each illumination pulse, and the timer is stopped with the detection of a photon. The time difference between the stop and start signals gives the photon's time of flight, which is easily converted to a measurement of the round-trip distance through multiplication by the speed of light. Due to timing uncertainty and the presence of nuisance detections of ambient light, illumination is repeated to build up a histogram of photon detection times, from which more reliable depth estimates can be determined.

Transverse resolution is achieved using the illumination or detection or both, each approach presenting different advantages and drawbacks. SPL systems have conventionally employed raster-scanned illumination, as illustrated in Figure 1. A laser aimed at one spot in a scene repeatedly pulses for a certain dwell time before being redirected to the next spot. While laboratory experiments typically use a simple pair of *XY* galvo mirrors, and the original Velodyne systems physically rotate the lasers and detectors for a 360° field of view, rugged beam steering with fewer mechanical components is an area of ongoing industrial development, with current approaches employing spatial light modulators, such as optical phased arrays, microelectromechanical systems mirrors, and liquid crystal metasurfaces [6]. Scanned illumination enables the use of a single-pixel or bucket detector, which is often in a confocal configuration (focused and coaxially aligned with the laser) to limit the number of photons undergoing multiple bounces or originating from ambient sources from being detected.

Unfortunately, raster scanning is an inherently slow, serial process. A more desirable approach is to broadly illuminate a swath of the scene and achieve spatial resolution with an array of single-photon-sensitive elements [10]. While detector arrays promise faster, parallelized acquisition, existing arrays are still limited in their resolution, and the laser power broadly diffused across a larger area further reduces the signal strength received at each pixel location. One compromise is the use of a line illumination and line array detector, which reduces the spatial scanning to a single dimension and limits the diffusion of the laser power [11].

Acquisition modeling

Despite the variations in acquisition hardware, each pixel, in general, will be associated with a stream of photon detection times. The time of arrival of a photon at the detector and whether that photon will be registered are inherently random processes, so the first step in making use of SPL data is to understand the probability distribution of the detection times.

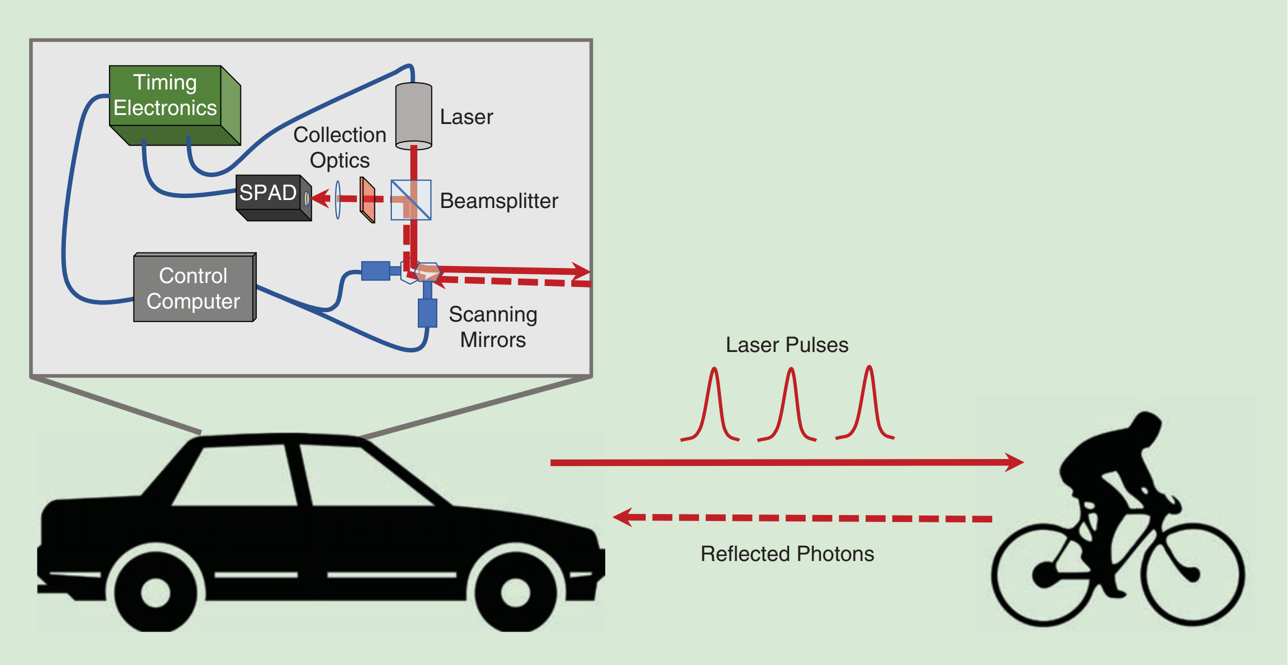


FIGURE 1. An example SPL system in a raster-scanned confocal configuration. The laser illuminates one point in the environment at a time, directed by a pair of scanning mirrors. Light reflecting back from illuminated surfaces is directed toward a single-photon-sensitive SPAD detector. The time difference between the illumination and the photon detection is recorded and processed.

Factors affecting the detection time distribution can be divided into two groups based on optical or electrical interactions.

Optical interactions

Optical factors encompass all influences on the light reaching the sensor, including the laser illumination, atmospheric attenuation, surface reflections, and ambient light. We consider the optical model for a single point in a raster-scanned lidar implementation, which is mostly analogous to light reaching a single pixel in a detector array. The laser illuminates a short-duration pulse with shape $h(t)$. Any surface within the beamwidth will reflect some light back, delayed by a time dependent on the distance d_m to surface m and attenuated by the reflectivity r_m, which includes the effects of the surface albedo, view angle, and optical illumination intensity. In addition, ambient light at the same wavelength as the laser (or more generally, within the spectral range of the detector) will reach the detector, although the timing of that incident light does not contain any depth information. The timing of photons incident on a detector is well understood to be described by a Poisson process. Because the illumination is periodic, with the period denoted t_r, the process intensity in one period is approximately given as

$$\lambda(t) = \lambda_b + \sum_{m=1}^{M} r_m h(t - 2d_m/c), \quad \text{for } t \in [0, t_\mathrm{r}), \tag{1}$$

where λ_b is the intensity of the ambient light at the illumination wavelength (assumed to be constant for the duration of the acquisition), M is the number of surfaces present in the field of view, and c is the speed of light (roughly 3×10^8 m/s in air). We refer to the first term as uninformative *background* noise, while the contributions combined in the sum are denoted as the informative *signal* intensity.

Acquisition electronics

The acquisition electronics determine which photon detection events are registered and what time information is stored for each event. For instance, not all photons incident on the SPAD actually cause the generation of a photoelectron, whereas thermal noise may cause spurious detections known as *dark counts*. Furthermore, the timing jitter of the circuitry in response to an incident photon will effectively increase the pulsewidth of $h(t)$. These effects tend to be minor in practice and do not affect the processing approach. We thus typically consider $\lambda(t)$ to describe the detection process. The photon flux, or the expected number of photons incident in one cycle, is given by $\Lambda = \int_0^{t_\mathrm{r}} \lambda(t)\,dt$, so the number of photon detections after n_r illumination cycles is a Poisson random variable $N \sim \mathcal{P}(n_\mathrm{r}\Lambda)$. As described in the "Detector Dead Times" section, this model is only sufficiently accurate in the low-flux regime (i.e., $\Lambda \ll 1$).

More fundamental to the acquisition process is the effect of temporal quantization. Each photon detected by the SPAD is assigned a time stamp, where the resolution Δ_t is dictated by the TCSPC electronics. Rather than processing each of the N photon detection times $t_1, \ldots, t_N$ separately, a histogram $[y_1, \ldots, y_{N_k}]$ of N_k time bins is typically constructed using the detection times at each pixel (i, j). Then the observation model for each time bin is a Poisson random variable, where the parameter integrates the Poisson process intensity across that bin:

$$y_k \sim \mathcal{P}\left(\int_{(k-1)\Delta_t}^{k\Delta_t} n_\mathrm{r}\lambda(t)\,dt\right), \quad \text{for } k = 1, \ldots, N_k. \tag{2}$$

When considered together, forming a histogram for each of the $N_i \times N_j$ transverse pixel locations yields a 3D data cube.

Estimation basics

Reflectivity, depth, and background information can be estimated from the photon detections according to model (2). The standard estimation procedure assumes a single surface within each pixel ($M = 1$) and known background rate $b = \lambda_b \Delta_t$. Under these assumptions, the maximum likelihood (ML) reflectivity estimate $\hat{r}$ is easily given as

$$\hat{r} = \frac{N - bN_k}{\int_0^{t_\mathrm{r}} h(t)\,dt}. \tag{3}$$

Finally, the ML depth estimator $\hat{d}$ is given by cross-correlating the detection time histogram with the logarithm of $h(t)$, also known as the *log-matched filter*:

$$\hat{d} = \frac{c}{2} \operatorname*{argmax}_{\tau \in [0, t_\mathrm{r})} \sum_{k=1}^{N_k} y_k \log\left[\hat{r}h(k\Delta_t - \tau) + b\right]. \tag{4}$$

Challenging estimation conditions

Unfortunately, the widely varying environments encountered by an AV and the need for fast acquisition often result in conditions that can make reflectivity and depth estimation more challenging. In many cases, the conventional pixel-wise estimation (3) and (4) fails under these conditions, motivating the algorithms described later in this article. A number of these conditions are depicted in Figure 2 and described as follows.

Few photons

The number of detected photons N may be small or even zero for several reasons: the number of illuminations n_r is kept low for real-time acquisition, the surface reflects very little light because it is weakly reflective or far away, and so on. Demonstrations of photon-efficient SPL used exactly one photon per pixel (ppp) [9] or an average near 1 ppp [12], [13], resulting in many pixels with no detections.

Strong ambient light

Estimation is particularly challenging if the ratio between the number of photons due to the laser and ambient illuminations, referred to as the *signal-to-background ratio* (*SBR*), is low. Even though optical methods (e.g., confocal configurations and bandpass filters) are used to limit the amount of ambient light that reaches the detector, strong daylight, especially when combined with a weak surface reflection, can result in far more detection events associated with background photons than from signal photons.

Absence of surfaces

The most basic 3D reconstruction methods assume a single surface at each pixel location. If a pixel has no object in its line of sight ($M = 0$), then the histogram contains only background detection events.

Multiple surfaces

On the other hand, there may be reflections from multiple surfaces present at one pixel ($M > 1$). This may occur because the light passes through a semitransparent material, such as glass. Alternatively, the pixel size or field of view increases with distance (e.g., due to the laser divergence in a scanned setting), so the spot is more likely to cover multiple surfaces. This same principle is often used in foliage-penetrating airborne lidar for terrain mapping.

Pulse broadening

Surfaces are generally assumed to be opaque and approximately normal to the illumination beam so that the reflected temporal response closely resembles the shape of the illumination pulse. However, subsurface scattering or oblique-angled surfaces, especially at long distances, will return broadened pulse profiles [14].

Attenuating media

Particles in the beam path, such as fog, smoke, rain, and snow, affect the acquired light by scattering photons in different directions after both the illumination (forward path) and reflection (return path). To some extent, the result is similar to that of a signal weakened by additional attenuation and increased background due to scattered photons [15], although the near-range effects of scattering also reshape the temporal distribution of the background, with more detections at earlier times [16].

Challenges due to nonideal acquisition electronics

Although the effects of the electronics have thus far been treated as minor, further challenges requiring more careful modeling arise if the actual electronics significantly deviate from the ideal.

Coarse temporal quantization

The ability to accurately resolve transient information depends on the duration of the histogram bins. For raster-scanning systems, the bin resolution that can be achieved currently is on the order of picoseconds, which is typically much less than the duration of pulse $h(t)$, so quantization effects on the depth estimation can be neglected. However, the timing resolution of detector arrays is usually coarser for each element than for a single-pixel device due to hardware and readout constraints. Whereas the single-photon-sensitive elements and timing electronics for a single pixel can easily be constructed as separate elements, the timing electronics in 2D arrays must be integrated on-chip for each pixel, resulting in a tradeoff between the fill factor of the photo-sensitive detector and timing components.

This becomes particularly problematic if the bin size Δ_t, which is the least significant bit of the temporal quantization, is larger than the duration of pulse $h(t)$. In that case, the depth resolution that can be achieved is quantization limited and can make object detection and recognition more difficult [see Figure 2(h)]. One proposed solution is to use longer-duration illumination pulses, which distributes photon detection events from the same surface across multiple time bins [10]. This acts as a sort of nonsubtractive dithering, which, through the averaging of enough detections, can provide sub-bin resolution. A more photon-efficient strategy is to take advantage of the short pulse durations and implement subtractive dithering via time shifts in the system synchronization. By inserting sub-bin

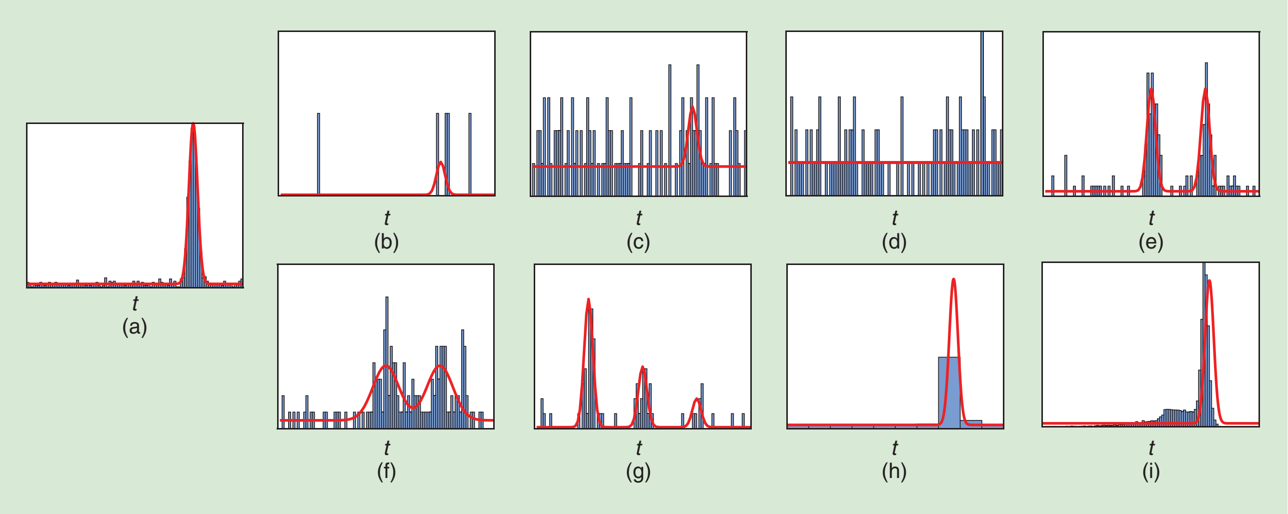


FIGURE 2. Examples of recorded histograms: (a) the ideal case, (b) few photons, (c) strong background illumination, (d) the absence of a target, (e) multiple surfaces per imaged pixel, (f) the broadening of the impulse response, (g) highly attenuating media, (h) coarse quantization, and (i) dead-time effects. The observed photon counts are shown in blue, whereas the a priori unknown Poisson intensity from (1) is in red.

delays into the synchronization signal between the laser pulse and the detection timing, efficient estimators can recover depths with sub-bin resolution [17].

Additional array considerations

Another current limitation of array manufacturing constraints is spatial nonuniformity. While raster-scanning imaging with a single-pixel detector has essentially identical system properties for each laser location, array elements have neither the same light sensitivity nor identical noise characteristics across the device. In particular, arrays often present hot pixels with overwhelming numbers of dark counts or dead pixels with inadequate light sensitivity. The inputs from these pixels must be omitted, or at least accounted for, in the reconstruction process.

Detector dead times

Unfortunately, the circuitry required for SPADs to be single-photon sensitive both precludes the ability to resolve numbers of photons and requires a reset period, or dead time, following each detection. Photons incident on a SPAD during a dead time are missed, and the dependence of a detection on the most recent detection time means the sequence of detection times can no longer be described by a Poisson process. Missed detections occur most frequently in the high-flux regime when $\Lambda \ll 1$ is no longer valid (e.g., when imaging bright objects, such as retroreflective street signs that reflect many photons), and they cause distortions of the detection time histogram that may result in erroneous depth and reflectivity estimates, thereby making accurate localization and object recognition more difficult.

The simplest way to avoid dead-time distortions is to attenuate the incident light to reduce the probability of missed photons, but attenuation is impractical for automotive lidar since it is already a challenge to recognize dark and distant objects in real time without additionally weakening the input signal. Instead, dead time can be mitigated by recognizing that the sequence of detection times forms a Markov chain [18]. A good approximation of the dead-time distorted detection time distribution is given by the stationary distribution of the Markov chain, which can be precomputed and tabulated for different light-intensity values. Additional methods exist for correcting histograms when range gating can also be used [19]. Thus, the distributions for both bright and dark objects can be accommodated and correctly matched, resulting in accurate point clouds.

Scene reconstruction

So far, we have discussed the acquisition process and associated challenges for individual SPL histograms. Here, we consider the problem of processing the complete data cube of $N_i \times N_j$ pixels with N_k histogram bins, indicated in tensor form as $\boldsymbol{y} \in \mathbb{Z}_+^{N_i \times N_j \times N_k}$. We denote the set of depth positions, reflectivity values, and background levels as $\boldsymbol{d}$, $\boldsymbol{r}$, and $\boldsymbol{b}$. The 3D reconstruction task consists of recovering a 3D point cloud $(\boldsymbol{d}, \boldsymbol{r})$ and background levels $\boldsymbol{b}$ from the photon detection events $\boldsymbol{y}$, as illustrated in Figure 3. This task is a difficult inverse problem due to the random noise affecting the measurements, the depth uncertainty related to the breadth of the impulse response, and potentially missing data issues, such as the presence of hot or dead pixels. The classical ML estimate using cross correlation (4) does not provide robust estimates when very few photons are collected by the detector. To improve these estimates, we can exploit a priori knowledge on the scene's structure, restricting the reconstructed scene to a set of plausible point clouds. This knowledge is incorporated using regularization functions by solving

$$(\hat{\boldsymbol{d}}, \hat{\boldsymbol{r}}, \hat{\boldsymbol{b}}) = \underset{\boldsymbol{d},\boldsymbol{r},\boldsymbol{b}}{\operatorname{argmin}} - \log p(\boldsymbol{y}|\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b}) + \rho_1(\boldsymbol{d}) + \rho_2(\boldsymbol{r}) + \rho_3(\boldsymbol{b}), \tag{5}$$

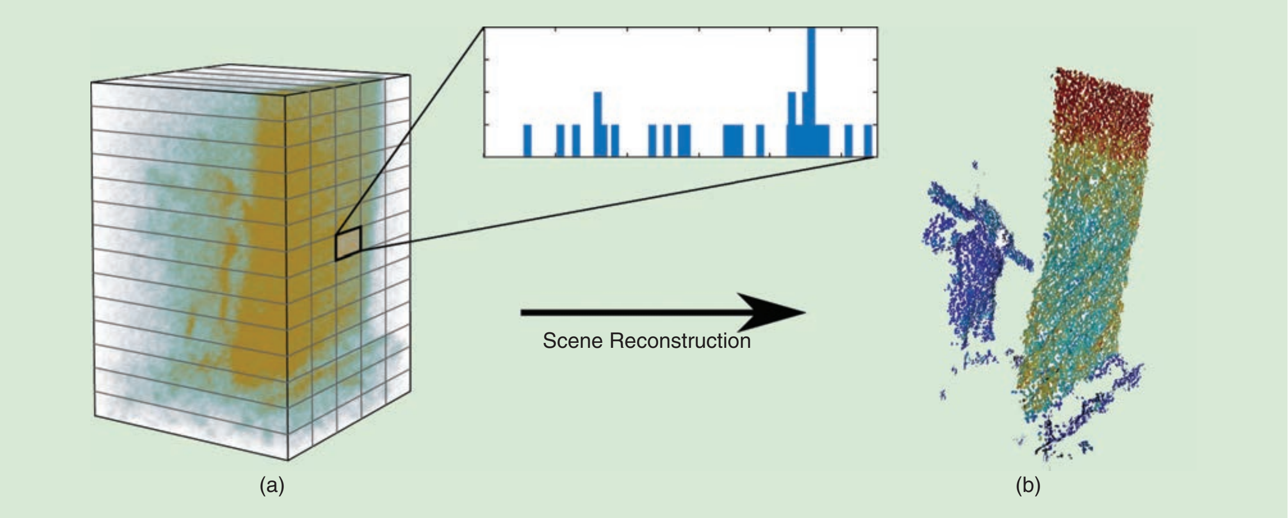


FIGURE 3. An SPL data set containing two surfaces (the man behind camouflage netting from [20]). The (a) photon detections and (b) 3D point cloud. The graph on the left shows the histogram of a given pixel. The limited number of collected photons and the high background level make the reconstruction task extremely challenging. In this case, processing the pixels independently yields poor results, while it can be improved by considering a priori knowledge about the scene's structure. Figure adapted from [24]; used with permission.

where $p(\boldsymbol{y}|\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b})$ is given by the observation model [e.g., (2)] and ρ_1, ρ_2, and ρ_3 are interpreted as the regularization terms for the depth, reflectivity, and background, respectively. In the context of Bayesian statistics, these terms are interpreted as negative log-prior distributions. The objective in (5) assumes that depth, reflectivity, and background parameters are a priori independent; however, some models also capture dependencies between them by using a nonseparable $\rho(\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b})$.

The 3D reconstruction task is challenging for several reasons. First, the cost function in (5) is generally nonconvex and even multimodal. Second, meaningful and well-defined priors that capture correlations of 2D manifolds (surfaces) are difficult to construct. Third, making an all-encompassing algorithm that can handle varying scenarios, such as those shown in Figure 2, is a difficult task. Finally, the algorithms have to be able to process very large volumes of data while providing fast estimates for any posterior decision making.

The SPL literature contains a wide variety of 3D reconstruction algorithms, differing both in the assumptions about the signal model (1) and the regularization assigned to the unknown parameters (5). We distinguish two main families of algorithms. The first group assumes exactly one object per pixel $(M = 1)$, reducing the 3D reconstruction problem to the estimation of depth, reflectivity, and background images. Most methods in this group use total variation regularization for both the depth and reflectivity images [9], [12], [21]. The second group, namely, multidepth methods, relaxes this assumption and attempts to infer a more general 3D point cloud, relying on priors defined in 3D space. In the following, we discuss different approaches to reconstruction, from robustness to background illumination and the real-time scene reconstruction of multiple surfaces per pixel. Figure 4 compares classical cross-correlation estimation and advanced 3D reconstruction algorithms in several challenging scenarios.

Robust imaging

SPL systems in AVs must be able to handle conditions with both weak signals and strong ambient light simultaneously. A fundamental goal of scene reconstruction algorithms is thus to provide accurate depth estimates for a low SBR. Initial photon-efficient methods emphasized censoring background detections by comparing detection times at neighboring pixels before applying background-free estimation techniques [9], [12]. More recent algorithms take advantage of two additional

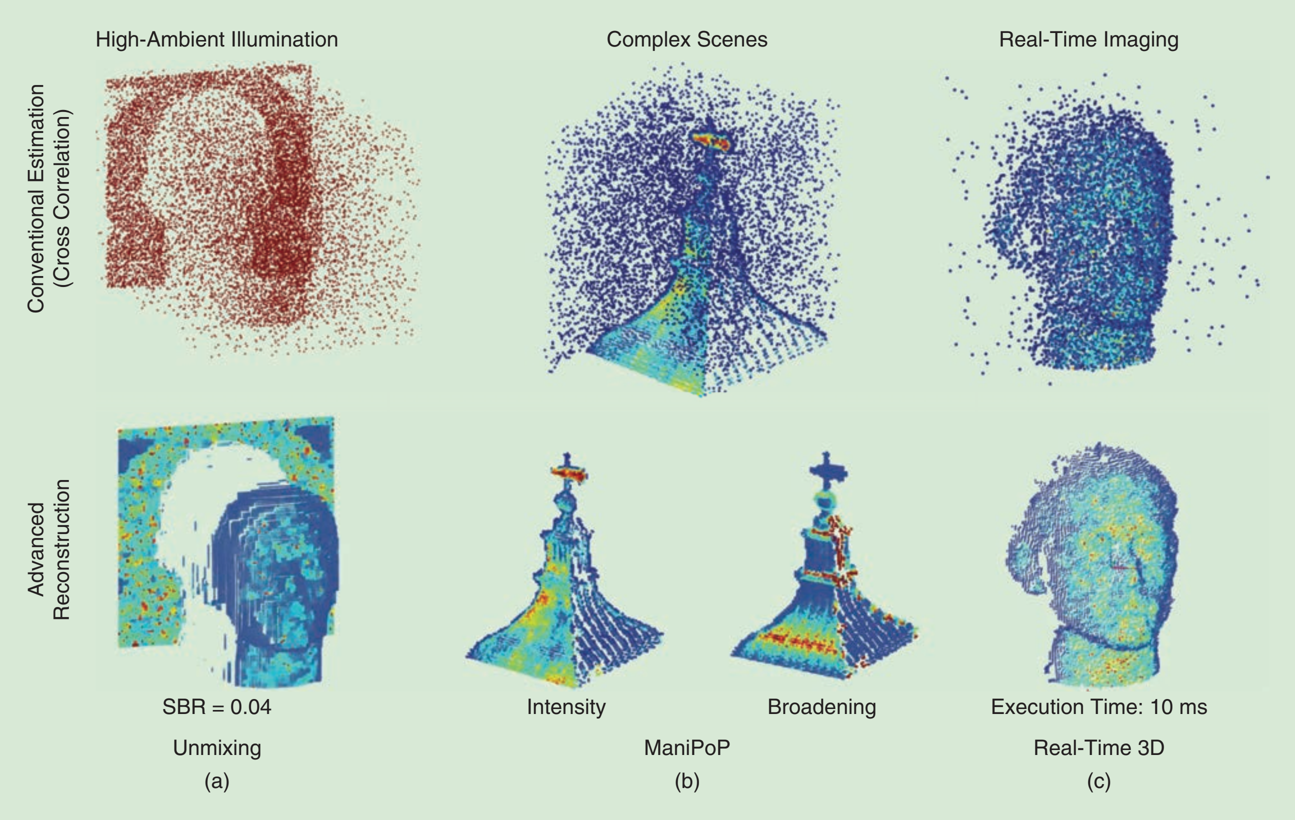


FIGURE 4. Examples of 3D reconstructions in challenging imaging conditions: The top row shows ML estimates by cross correlation, whereas the second row shows restorations using advanced methods that also incorporate spatial priors. See [7] and [23] for acquisition details. (a) High-ambient illumination. Robust single-depth algorithms (e.g., [21]) can form accurate point clouds for targets such as the mannequin head at 325 m, even when short acquisition times and strong ambient illumination result in few signal photons and an extremely low SBR. (b) Complex scenes. Multidepth algorithms, including ManiPoP [14], can handle complex scenes, such as in the 3-km acquisition depicted here, which has some pixels with zero or multiple surfaces. The estimated degree of impulse response broadening could be used in computing surface normals. (c) Real-time imaging. A real-time multidepth algorithm [24] is designed to process 3D data at video rates.

observations. At a pixel level, photons due to ambient illumination are uniformly distributed across the histogram, whereas photons due to a target are clustered following the shape of the impulse response of the device. At an image level, background intensities resemble a passive image of the scene, enabling reconstruction algorithms to benefit from additional spatial correlation. The single-depth method introduced in [21] uses the first idea of discarding nonclustered photon detections. Following an iterative multiscale approach, pixels with a very low SBR are integrated with neighboring pixels, increasing the probability of finding a cluster of signal photons [see Figure 4(a)]. Alternatively, an SPL data cube can be fused with a high-resolution 2D photograph via a convolutional neural network to improve the detail of depth and reflectivity reconstructions [22].

Imaging complex scenes

In a general setting, an a priori unknown, but usually reduced, number of surfaces can be found at each imaged pixel. Target detection algorithms focus on cases where at most one surface is present ($M \leq 1$), which encompass a wide range of practical scenes. In contrast, multidepth algorithms relax assumptions about the number of surfaces, estimating the quantity of objects M at each pixel.

In the target detection setting, simply thresholding the reflectivity estimates obtained by a single-depth algorithm is generally not robust to background illumination. Hence, specific target detection algorithms estimate an additional binary image indicating the per-pixel presence or absence of a target. In [23], spatial correlations are promoted using an Ising model, and inference is performed using a reversible-jump Markov chain Monte Carlo algorithm (RJ-MCMC). While being fully unsupervised (the hyperparameters are estimated within the algorithm), this method requires execution times on the order of hours. Recent alternatives achieve faster results by computing a per-pixel posterior probability of the target presence and processing the pixels independently in parallel.

The general multidepth assumption includes the previously discussed algorithms as special cases, at the expense of solving a harder problem. We identify two main strategies. The first approach aims at estimating a 3D volume of reflectivity values, where only a few nonzero values correspond to the 3D points, rewriting (2) for each pixel as

$$\boldsymbol{y} \sim \mathcal{P}(\boldsymbol{H}\boldsymbol{r} + \mathbf{1}_{N_k} b), \tag{6}$$

where $\boldsymbol{H}$ is the convolution matrix of $h(t)$, $\mathbf{1}_{N_k}$ is a length-N_k vector of ones, and $\boldsymbol{r} = [r_1, \ldots, r_{N_k}]^T$ is the reflectivity vector, which has small ℓ_0 pseudonorm $\|\boldsymbol{r}\|_0 = M$. In this direction, different convex relaxations combine the ℓ_1 and ℓ_2 norms to promote sparsity within histograms and take advantage of correlations across pixels (e.g., [20]). These formulations have the advantage of possessing a unique solution, which can be obtained using the standard proximal-gradient or alternating-direction method of multipliers optimization techniques. The downsides are the expensive memory requirements (at least one dense 3D cube has to be stored in memory) and the lack of exactly sparse solutions, generally relying on a postprocessing step to further sparsify the output.

A second strategy directly estimates a 3D point cloud, where the dimension of the parameter space (i.e., the number of 3D points) is a priori unknown. The first step in this direction was the algorithm in [25], which infers the point positions using an RJ-MCMC algorithm to handle the varying dimensions but does not consider spatial correlations between the point positions. The more recent ManiPoP model [26] includes spatial correlations, using a spatial point process prior to model the manifold structure of 3D surfaces. Inference is again performed with an RJ-MCMC algorithm, but the overall computational cost is significantly reduced by the use of carefully tailored proposals. In [14], ManiPoP was extended to an even more general observation model, including the effects of attenuating media and impulse response broadening [see Figure 4(b)]. While being robust to a wide range of conditions, these algorithms still require execution times on the order of seconds or minutes, hindering any real-time decision making in AVs. This shortfall is addressed by algorithms tackling real-time performance.

Real-time imaging

While the offline processing of scanned stationary targets has offered compelling evidence for the potential of SPL systems, practical vehicular deployment requires real-time performance. Crucial to speeding up the point cloud formation is the parallelization of both the acquisition and processing procedures. Initial real-time lidar development at the Massachusetts Institute of Technology Lincoln Laboratory required expensive computation to be run on a large server cluster [27]. More recently, a single-pixel SPL device illuminating short-range scenes with structured (Hadamard) patterns from a digital micromirror device was shown to be capable of processing up to 12 frames/s enabled by direct inversion of the patterns [28]. As SPAD arrays have improved to capture very high rates of time-stamped photon detections in parallel, the bottleneck for real-time performance has moved to the reconstruction algorithms.

New algorithms are designed to take advantage of the parallel computations in general-purpose GPUs, which are compact and inexpensive enough for use in AVs. Under a single-depth assumption, the algorithm in [29] avoids the construction of histograms, updating the depth and reflectivity estimates with each new photon arrival. The multidepth algorithm in [24] uses off-the-shelf point cloud denoisers from the computer graphics community as part of a proximal-gradient optimization algorithm similar to plug-and-play strategies in image processing. The combination of fast parallel denoisers and parallel gradient updates enables the acquisition and processing of up to 50 frames/s [see Figure 4(c)]. In this direction, future methods will have to incorporate other prior knowledge without slowing down the reconstruction process significantly, relying on an efficient combination of algorithmic structure and computing hardware.

Anticipatory imaging

In addition to the fast and precise formation of point clouds, which is the primary goal of lidar systems, SPL has the potential to extend its functionality beyond the capacity of human vision. A growing research area is non-line-of-sight (NLOS) imaging: forming images of, or at least detecting, objects outside the direct field of view (see Figure 5). The principle behind such work is that, even after multiple bounces, the intensity and transient measurements of light may retain some information about each interaction between the source and detector. Thus, diffuse surfaces, such as roads and walls, could act as mirrors, enabling vision around obstacles.

There are two main challenges to NLOS imaging: 1) diffuse reflectors scatter light in all directions, scrambling directional information about the light paths, and 2) the light returning to a detector after multiple diffuse bounces is extremely weak. The single-photon sensitivity of SPL systems has made them a popular choice for tackling the second challenge. The first problem continues to be a hot topic of research, with novel computational strategies (e.g., [31] and [32]) employed to make the most of the weak signals.

One recent approach uses a standard coaxial SPL configuration to scan a grid of points on a visible wall [30]. While the strong first peaks contain the usual LOS lidar signal, later parts of the temporal histogram contain contributions from light that has undergone multiple bounces before reaching the detector. Clever manipulation of the image formation model reveals that the photon detection time histogram measurements can be expressed as a 3D convolution, enabling a straightforward, closed-form solution that can be implemented with a fast Fourier transform. The resulting algorithm has lower computational and memory requirements than previous approaches using back projection or linear inverse formulations, thus even enabling some real-time demonstrations. The method is especially powerful for use with hidden retroreflective surfaces, which return much more light than diffuse surfaces do. Since retroreflective coatings are already in widespread use on roads, including on street signs and within bicycle reflectors, NLOS imaging with SPL may one day be capable of alerting an AV to potential road hazards occluded from view.

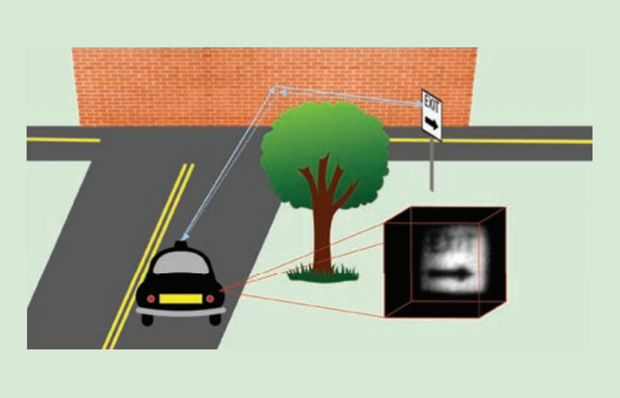


FIGURE 5. A possible use case for incorporating NLOS imaging into lidar for AVs. The reconstruction of the exit sign uses real NLOS data from [30].

Future directions for SPL

SPL is already a promising tool for providing fast, high-resolution depth sensing for AVs. Ongoing hardware development of the illumination and detection systems will continue to bring down costs and improve acquisition performance. Nonetheless, signal processing for future SPL systems will still need to address the main questions of how to model the acquisition process and how to incorporate prior knowledge of 3D structures for point cloud formation. Future extensions to SPL will incorporate more information from alternative lidar acquisitions, e.g., using multiple wavelengths or polarizations, or from other sensing modalities, such as standard red–green–blue images and radar. These other sources of information could improve the robustness of the imaging system, generally at the cost of longer or more complex acquisitions and reconstructions. Deep neural networks will likely play an important role in identifying correlated information and fusing the various modalities.

The ever-growing demand for better spatial (array size and timing) and spectral resolution can only be met with highly-scalable signal acquisition and processing techniques. Developing compressed sensing techniques for photon-counting lidar, following single-pixel SPL, wavelength-time coding, and color-coded apertures, could potentially reduce both the acquisition and reconstruction time. These methods will continue to require a better understanding of the tradeoffs between the acquisition (photon budget, laser power, and acquisition model), reconstruction complexity (memory requirements, parallel or serial architecture, and execution time) and estimation performance. Given the success demonstrated thus far and the available opportunities for improvement, SPL will ultimately be one of the factors enabling autonomous driving and the transformation of the transportation industry and beyond.

Authors

Joshua Rapp (jrapp@stanford.edu) received his B.S. degree in electrical engineering from Tufts University, Medford, Massachusetts, in 2014 and his M.S. and Ph.D. degrees in electrical engineering from Boston University, in 2016 and 2020, respectively. He is currently a postdoctoral researcher at Stanford University, California. From 2015 to 2019, he was a fellow at Draper, Cambridge, Massachusetts. His research interests include computational imaging, lidar, and statistical signal processing. In 2018, he was the recipient of a Best Student Paper Award at the IEEE International Conference on Image Processing. He is a member of the Optical Society and a Student Member of the IEEE.

Julián Tachella (julian.tachella@ed.ac.uk) received his degree (with honors) in electronic engineering from Instituto Tecnológico de Buenos Aires, Argentina, in 2016 and his Ph.D. degree (with honors) from Heriot–Watt University, Edinburgh, Scotland, in 2020. Since 2020, he has been working as a postdoctoral research associate with the School of Engineering, University of Edinburgh. His research focuses on statistical image and signal processing, with a particular interest in computational imaging applications. He is a Student Member of the IEEE.

Yoann Altmann (y.altmann@hw.ac.uk) received his Ph.D. degree in signal and image processing from the National Polytechnic Institute of Toulouse, France, in 2013. From 2014 to 2016, he was a postdoctoral researcher at Heriot–Watt University, Edinburgh, Scotland. Since 2017, he has been an assistant professor at the Institute of Sensors, Signals, and Systems, School of Engineering and Physical Sciences, Heriot–Watt University. He is currently a research fellow of the Royal Academy of Engineering (2017–2022), developing new computational methods for low-illumination imaging and sensing. He is a Member of the IEEE.

Stephen McLaughlin (s.mclaughlin@hw.ac.uk) received his Ph.D. degree from the University of Edinburgh, Scotland, in 1990. From 1991 to 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques at Edinburgh University, where he was awarded a Chair in electronic communication systems in 2002. In 2011, he joined Heriot–Watt University, Edinburgh, as a professor of signal processing and head of the School of Engineering and Physical Sciences. His research interests are computational imaging and signal processing for communications. He is a Fellow of the IEEE, Royal Academy of Engineering, and Royal Society of Edinburgh.

Vivek K Goyal (v.goyal@ieee.org) received his mathematics and electrical engineering degrees from the University of Iowa, Iowa City, where he was the recipient of the John Briggs Memorial Award for the top undergraduate across all colleges. He received his graduate degrees in electrical engineering from the University of California, Berkeley, where he was the recipient of the Eliahu Jury Award. He is a professor of electrical and computer Engineering at Boston University. He is a Fellow of the IEEE and the Optical Society.

References

[1] “Traffic safety facts 2015: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system,” National Highway Traffic Safety Administration, Washington, D.C., 2015. [Online]. Available: https://www-nrd.nhtsa.dot.gov/Pubs/812384.pdf

[2] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, “Three decades of driver assistance systems: Review and future perspectives,” *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 6–22, 2014. doi: 10.1109/MITS.2014.2336271.

[3] R. O. Chavez-Garcia and O. Aycard, “Multiple sensor fusion and classification for moving object detection and tracking,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 525–534, Feb. 2016. doi: 10.1109/TITS.2015.2479925.

[4] V. Molebny, P. McManamon, O. Steinvall, T. Kobayashi, and W. Chen, “Laser radar: Historical prospective—From the East to the West,” *Opt. Eng.*, vol. 56, no. 3, p. 031220, Dec. 2016. doi: 10.1117/1.OE.56.3.031220.

[5] B. Behroozpour, P. A. Sandborn, M. C. Wu, and B. E. Boser, “Lidar system architectures and circuits,” *IEEE Comm. Mag.*, vol. 55, no. 10, pp. 135–142, Oct. 2017. doi: 10.1109/MCOM.2017.1700030.

[6] J. Hecht, “Lidar for self-driving cars,” *Opt. Photon. News*, vol. 29, no. 1, pp. 26–33, Jan. 2018. doi: 10.1364/OPN.29.1.000026.

[7] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, “Single-photon three-dimensional imaging at up to 10 kilometers range,” *Opt. Expr.*, vol. 25, no. 10, pp. 11,919–11,931, 2017. doi: 10.1364/OE.25.011919.

[8] Z.-P. Li, X. Huang, Y. Cao, B. Wang, Y.-H. Li, W. Jin, C. Yu, J. Zhang et al., Single-photon computational 3D imaging at 45 km. Apr. 2019. [Online]. Available: arXiv:1904.10341

[9] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, “First-photon imaging,” *Science*, vol. 343, no. 6166, pp. 58–61, 2014. doi: 10.1126/science.1246775.

[10] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. Wong et al., “Photon-efficient imaging with a single-photon camera,” *Nature Commun.*, vol. 7, 2016, Art. no. 12046. doi: 10.1038/ncomms12046.

[11] D. B. Lindell, M. O’Toole, and G. Wetzstein, “Towards transient imaging at interactive rates with single-photon detectors,” in *Proc. IEEE Int. Conf. Comput. Photog.*, May 2018, pp. 1–8. doi: 10.1109/ICCPHOT.2018.8368466.

[12] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, “Photon-efficient computational 3D and reflectivity imaging with single-photon detectors,” *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 112–125, June 2015. doi: 10.1109/TCI.2015.2453093.

[13] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, “Lidar waveform-based analysis of depth images constructed using sparse single-photon data,” *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1935–1946, May 2016. doi: 10.1109/TIP.2016.2526784

[14] J. Tachella, Y. Altmann, S. McLaughlin, and J.-Y. Tourneret, “3D reconstruction using single-photon lidar data exploiting the widths of the returns,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2019, pp. 7815–7819. doi: 10.1109/ICASSP.2019.8683075.

[15] R. Tobin, A. Halimi, A. McCarthy, M. Laurenzis, F. Christnacher, and G. S. Buller, “Three-dimensional single-photon imaging through obscurants,” *Opt. Expr.*, vol. 27, no. 4, pp. 4590–4611, Feb. 2019. doi: 10.1364/OE.27.004590.

[16] G. Satat, M. Tancik, and R. Raskar, “Towards photography through realistic fog,” in *Proc. IEEE Int. Conf. Comput. Photog.*, May 2018, pp. 1–10. doi: 10.1109/ICCPHOT.2018.8368463.

[17] J. Rapp, R. M. A. Dawson, and V. K. Goyal, “Improving lidar depth resolution with dither,” in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 1553–1557. doi: 10.1109/ICIP.2018.8451528.

[18] J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, “Dead time compensation for high-flux ranging,” *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3471–3486, July 2019. doi: 10.1109/TSP.2019.2914891.

[19] A. Gupta, A. Ingle, A. Velten, and M. Gupta, “Photon-flooded single-photon 3D cameras,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, June 2019, pp. 6770–6779.

[20] A. Halimi, R. Tobin, A. McCarthy, S. McLaughlin, and G. S. Buller, “Restoration of multilayered single-photon 3D lidar images,” in *Proc. Eur. Signal Proc. Conf.*, Aug. 2017, pp. 708–712. doi: 10.23919/EUSIPCO.2017.8081299.

[21] J. Rapp and V. K. Goyal, “A few photons among many: Unmixing signal and noise for photon-efficient active imaging,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 3, pp. 445–459, 2017. doi: 10.1109/TCI.2017.2706028.

[22] D. B. Lindell, M. O’Toole, and G. Wetzstein, “Single-photon 3D imaging with deep sensor fusion,” *ACM Trans. Graph.*, vol. 37, no. 4, 2018, Art. no. 113. doi: 10.1145/3197517.3201316.

[23] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, “Robust Bayesian target detection algorithm for depth imaging from sparse single-photon data,” *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 456–467, Dec. 2016. doi: 10.1109/TCI.2016.2618323.

[24] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, “Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers,” *Nature Commun.*, vol. 10, Nov. 2019, Art. no. 4984. doi: 10.1038/s41467-019-12943-7.

[25] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, “Bayesian analysis of lidar signals with multiple returns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2170–2180, 2007. doi: 10.1109/TPAMI.2007.1122.

[26] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, and J.-Y. Tourneret, “Bayesian 3D reconstruction of complex scenes from single-photon lidar data,” *SIAM J. Imaging Sci.*, vol. 12, no. 1, pp. 521–550, Jan. 2019. doi: 10.1137/18M1183972.

[27] P. Cho, H. Anderson, R. Hatch, and P. Ramaswami, “Real-time 3D ladar imaging,” *Lincoln Lab. J.*, vol. 16, no. 1, pp. 147–164, 2006. doi: 10.1117/12.664904.

[28] M.-J. Sun, M. P. Edgar, G. M. Gibson, B. Sun, N. Radwell, R. Lamb, and M. J. Padgett, “Single-pixel three-dimensional imaging with time-based depth resolution,” *Nature Commun.*, vol. 7, July 2016, Art. no. 12010.

[29] Y. Altmann, S. McLaughlin, and M. E. Davies, “Fast online 3D reconstruction of dynamic scenes from individual single-photon detection events,” *IEEE Trans. Image Process.*, vol. 29, pp. 2666–2675, Nov. 2019. doi: 10.1109/TIP.2019.2952008.

[30] M. O’Toole, D. B. Lindell, and G. Wetzstein, “Confocal non-line-of-sight imaging based on the light-cone transform,” *Nature*, vol. 555, pp. 338–341, Mar. 2018. doi: 10.1038/nature25489.

[31] S. Xin, S. Nousias, K. N. Kutulakos, A. C. Sankaranarayanan, S. G. Narasimhan, and I. Gkioulekas, “A theory of Fermat paths for non-line-of-sight shape reconstruction,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2019, pp. 6793–6802.

[32] X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. Huu Le, A. Jarabo, D. Gutierrez et al., “Non-line-of-sight imaging using phasor-field virtual wave optics,” *Nature*, vol. 572, pp. 620–623, Aug. 2019. doi: 10.1038/s41586-019-1461-3.

SP

Canan Aydogdu, Musa Furkan Keskin, Gisela K. Carvajal,
Olof Eriksson, Hans Hellsten, Hans Herbertsson, Emil Nilsson,
Mats Rydström, Karl Vanäs, and Henk Wymeersch

Radar Interference Mitigation for Automated Driving

Exploring proactive strategies

©ISTOCKPHOTO.COM/OONAL

Autonomous driving relies on a variety of sensors, especially radars, which have unique robustness under heavy rain/fog/snow and poor light conditions. With the rapid increase of the amount of radars used on modern vehicles, where most radars operate in the same frequency band, the risk of radar interference becomes a compelling issue. This article analyzes automotive radar interference and proposes several new approaches that combine industrial and academic expertise toward the goal of achieving interference-free autonomous driving (AD).

Introduction and motivation

Radar is becoming standard equipment in all modern cars, supporting, e.g., cruise control and collision avoidance in most weather conditions while providing high-resolution detections on the order of centimeters in the millimeter-wave (mm-wave) band. The next generation of advanced driver assistance (ADAS) and AD vehicles will have a multitude of radars covering numerous safety and comfort applications such as crash avoidance, self-parking, in-cabin monitoring, cooperative driving, collective situational awareness, and so on. Because automotive radar transmissions are uncoordinated, there is a nonnegligible probability of interference among vehicles, as shown in Figure 1. Although current automotive radars have already been impacted by interference to some extent, today, they are unlikely to raise customer awareness because state-of-the-art automotive radars are continuously updated and improved upon at many system levels.

However, the mutual interference problem is expected to become more challenging if not properly handled, as more vehicles are equipped with a greater number of radars providing 360° situational awareness at various distances that enable more advanced future ADAS and AD functionalities. This is evidenced by multiple international studies, such as the MOSARIM project [1] and the more recent IMIKO radar project. All of the major players in the automotive sensor market, such as Volvo and Veoneer, are studying the next generation of "interference-free radars." This includes, for example,

Digital Object Identifier 10.1109/MSP.2020.2969319
Date of current version: 26 June 2020

enhancing models to determine the impact of a larger density of radars, simulating new interference scenarios, and investigating different medium access control (MAC) models and methods to coordinate radar transceivers, both decentralized and centralized.

At this point, the automotive industry is ready to consider novel designs and approaches, which may impact standardization bodies before a new frequency spectrum is made available in the higher radio-frequency (RF) bands. Signal processing can provide ways to reduce or mitigate interference, both at the raw signal level as well as at the postdetection/target-tracking level. The particular properties and requirements of automotive radar impose significant challenges in terms of signal processing. This includes the combination of radar and communication waveforms, which brings up further possibilities regarding ultrareliable low-latency communications in vehicular ad hoc networks (VANETs). It is therefore important and timely to review what has been done, what the reasonable approaches are, and what the future holds.

The focus of this article is on frequency-modulated continuous wave (FMCW) radar because it is the most common and robust automotive radar. We provide an analysis of the impact of interference in FMCW both quantitatively and qualitatively, in terms of their probability, severity, and effects. Then, we cover different ways to mitigate interference, ranging from changing FMCW parameters to new signal structures and the explicit coordination between vehicles. We also study new techniques that are potentially more robust toward interference, including stepped-frequency orthogonal frequency-division multiplexing (OFDM). Finally, we describe what we believe will be the long-term evolution of automotive radar and its relation to mobile communication.

Automotive radar

History and future of automotive radar

Radar has been used in automotive applications for a long time. In 1949, unfortunate car drivers were issued speeding tickets based on speed measurements obtained from the radar speed gun, recently invented by John L. Barker. However, on-board automotive radar was not made commercially available until 1999, when it was introduced for collision warning and automatic cruise control (ACC). See [2] for an early history of automotive radar with some entertaining vintage photography. Over the years, there has been a strong push to increase the integration level of mm-wave electronics used for automotive radar and industrial radar sensors. The early discrete hardware designs have been replaced by a few chips in III–V-materials, and now, CMOS single-chip solutions are available. CMOS technology provides the ability to fully integrate analog and digital electronics, making very advanced protocols and detection schemes possible at low cost and low power. Consequently, radar is becoming more and more common for supporting various automotive applications. ADAS systems based on radar are today standard equipment in most new vehicles. Vehicles capable of some level of AD are also expected to rely, at least to some degree, on radar systems for monitoring vehicle surroundings. The number of radar transceivers operating throughout the traffic environment is foreseen to increase rapidly over the coming years. As the number of radar transceivers in the traffic infrastructure increases, radar interference is also expected to increase.

Today, most radar transmissions are uncoordinated, meaning that there is no a priori agreement on who is allowed to transmit and when. A number of recent studies have identified the interference situations that are likely to arise as the automotive radar transceiver market penetration increases [1], [3]. FMCW waveforms can, up to a point, be relatively easily repaired in the event they are intermittently corrupted by interference [4], which is why they are still operational. Future radar systems are expected to occupy frequency bands that are higher and higher in frequency. Transceivers operating around 77 GHz are available today, and transceivers operating at carrier frequencies beyond 100 GHz are expected. Frequencies as high as 300 GHz and beyond are being considered for some applications, such as synthetic aperture radar mapping. Operation at such high frequencies brings the obvious benefits of improved miniaturization, but also presents challenges in terms of hardware complexity and signal attenuation. Moreover, interference-free operation will require radar transmission standardization. A standardized transmission scheduling system resembling today’s cellular communication system

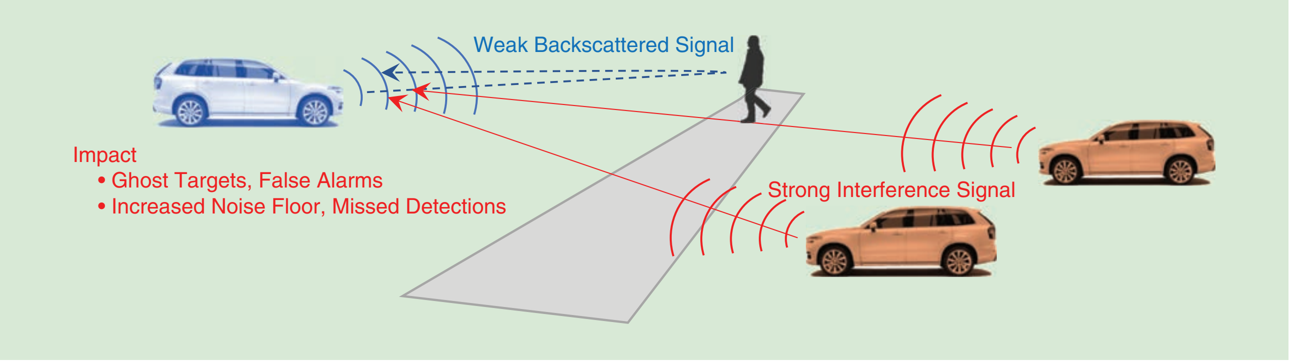


FIGURE 1. Interference is generally much stronger than the desired radar signal, due to its one-way propagation. Interference increases with more interfering radars and leads to false alarms and missed detections.

would present a solution to the interference problem, but it is not without challenges, both technical and political.

Basics of FMCW radar

In a general FMCW radar, a frequency sweep, i.e., a chirp, is generated by a voltage-controlled oscillator operated by a digital synthesizer. The generated chirp signal is split and sent into two different signal paths: one path is directed to the transmitter (Tx) antenna, while the other path is directed to the mixer correlator. Before the chirp is sent out on the Tx antenna it passes a power amplifier (PA), which boosts the transmitted energy. Here we assume that there are no idle periods between consecutive chirps. It is also possible to introduce random interchirp idle periods to reduce the probability of mutual interference between FMCW radars. The transmit waveform of an FMCW radar with K consecutive linear FM chirps (or sweeps) can be expressed as [5]

$$s(t) = \sum_{k=0}^{K-1} x(t-kT), \tag{1}$$

where the individual chirps are given by

$$x(t) = e^{j\varphi(t)}\mathrm{rect}_T(t), \quad \varphi(t) = 2\pi(f_c t + 0.5\alpha t^2). \tag{2}$$

Here, $\alpha = B/T$ is the chirp slope, B denotes the sweep bandwidth, T represents the chirp duration, f_c is the carrier frequency, and $\mathrm{rect}_T(t)$ is square pulse of duration T with an amplitude of 1. The received reflected signal from a target is very weak due to its two-way free-space propagation path loss and the losses incurred during reflection and thus needs to be amplified with a low-noise amplifier (LNA) to maintain an acceptable signal-to-noise-ratio (SNR). The amplified signal from the target reflection is correlated with the Tx signal in the mixer correlator, which is called *dechirping*. The low-pass filter at the output of the correlating mixer offers some interference rejection. The round-trip delay and Doppler shift caused by the relative velocity of the target shifts the frequency of the received signal compared to that of the transmitted signal. As a result, the mixer creates a beat signal that will pass through a low-pass filter and be digitized, yielding delay and Doppler estimates after matched filtering. In modern automotive radars, it is also possible to estimate azimuth and elevation of targets using multiple antennas.

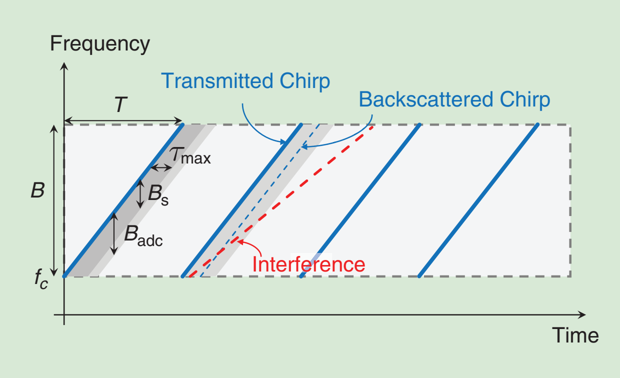


FIGURE 2. Four consecutive chirps in time-frequency representation, with several key notations included. Backscattered and interfering signals are shown at the second chirp.

FMCW radar signal processing chain

Suppose there exists a single target of interest acting as a point scatterer, characterized by a complex channel gain γ [including the effects of path loss, antenna gain, and radar cross section (RCS)], an (initial) round-trip propagation delay $\tau = 2R/c$, and a normalized relative Doppler shift $\nu = 2v/c$, where R and v denote, respectively, the distance and relative radial velocity between the radar and target, and c is the speed of wave propagation. The received backscattered signal is now processed in three stages.

Stage 1: Dechirping

Under the stop-and-hop assumption [6, Ch. 2.6.2], the kth chirp of the received signal is given by

$$r_k(t) = \gamma x(t + (t+kT)\nu - \tau) + w_k(t), \tag{3}$$

where $0 \le t \le T$ denotes the time relative to the beginning of the kth chirp, and $w_k(t)$ is measurement noise. To obtain the beat signal at the intermediate frequency, the received signal $r_k(t)$ in (3) is dechirped through conjugate mixing with the transmitted signal $x(t)$ in (2). Here we ignore the terms whose total phase progression over a coherent processing interval (CPI) of K chirps is smaller than $\pi/4$ for typical automotive FMCW settings [6, Ch. 2.6.3]:

$$y_k(t) = r_k(t)x^*(t) = \gamma e^{j2\pi f_c \nu (t+kT)} e^{-j2\pi\alpha\tau t}\mathrm{rect}_T(t-\tau) + w_k(t)x^*(t). \tag{4}$$

Let $\tau_{\max}$ denote the round-trip delay (see Figure 2) that corresponds to a maximum target range of interest (i.e., $\tau_{\max} \ge \tau$), which is related to the radar bandwidth of interest B_s as $\tau_{\max} = B_s/\alpha$. The analog-to-digital converter (ADC) bandwidth $B_{\mathrm{adc}} \ge B_s$ imposes a limit on B_s and thus the maximum detectable range $\tau_{\max}$. After low-pass filtering the beat signal in (4) with bandwidth B_s, sampling with a period of T_s for $\tau_{\max} \le t \le T$, we rearrange into a slow-time fast-time data matrix, where the kth row contains the samples of the kth chirp (fast time), while the nth column contains the nth sample of each chirp (slow time). In other words, we have

$$y_{k,n} = \gamma e^{j2\pi(-\alpha\tau + f_c\nu)nT_s} e^{j2\pi f_c \nu kT} + w_{k,n}, \tag{5}$$

for $k = 0, \dots, K-1$ and $n = n_{\max}, \dots, N-1$, $n_{\max} = \lfloor \tau_{\max}/T_s \rfloor$, $N = \lfloor T/T_s \rfloor + 1$, and $w_{k,n}$ are independent and identically distributed complex Gaussian noise samples with variance σ^2.

Stage 2: Target range-velocity estimation

To provide an estimate of target range and velocity, a 2D discrete Fourier transform (DFT) can be applied across slow- and fast-time dimensions of the beat signal in (5), which yields the

FMCW delay-Doppler spectrum evaluated at a given delay-Doppler pair $(\hat{\tau}, \hat{\nu})$:

$$z(\hat{\tau}, \hat{\nu}) = \sum_{k=0}^{K-1} \sum_{n=n_{\max}}^{N-1} y_{k,n} e^{j2\pi\alpha\hat{\tau}nT_s} e^{-j2\pi f_c \hat{\nu} kT}. \tag{6}$$

The periodogram $|z(\hat{\tau}, \hat{\nu})|^2$ corresponding to (6) yields a dominant target peak at $(\hat{\tau}, \hat{\nu}) = (\tau - f_c\nu/\alpha, \nu)$, which can be recovered using, for example, constant false alarm rate (CFAR) detectors [6, Ch. 6]. The peak value of $|z(\hat{\tau}, \hat{\nu})|$ is proportional to the processing gain $G_{\mathrm{p}} = K(N - n_{\max})$. This frequency identification method is referred to as the *periodogram spectral estimator* [7, Ch. 2.2.1]. Here, the shift $f_c\nu/\alpha$ in delay stems from range-Doppler coupling inherent in the FMCW waveform [6, Ch. 4.6.4]. To compensate for the coupling effect, the Doppler-dependent term $f_c\nu/\alpha$ can be added back to the delay estimate after delay-Doppler retrieval. When there are multiple objects in the radar field of view (FOV), (6) will have multiple peaks. To distinguish the different objects, each object pair must be separated by a certain minimum gap in delay and Doppler domains, which is determined by the radar resolution. The range and velocity resolution of an FMCW radar can be derived from (6) (assuming $n_{\max} \ll N$) as $\Delta R = c/(2B)$ and $\Delta v = \lambda_c/(2KT)$, where λ_c is the carrier wavelength [7, Ch. 2.4]. Therefore, higher sweep bandwidth leads to better range resolution, while a longer CPI duration means improved velocity resolution.

Stage 3: Tracking filter

At the final stage of the signal processing chain, the delay-Doppler detections $\{(\tau_p, \nu_p)\}_{p=0}^{P-1}$ (along with the corresponding azimuth-elevation pairs in the case of multiple antennas [5]) are fed to a data association and tracking filter to provide filtered 3D positions and the velocities of surrounding objects. Here, P denotes the number of targets seen during one scan.

Is interference really a problem?

In the next section, we provide a theoretical analysis for the impact of interference on the radar signal processing. We start by studying a single link and then extend to a network of vehicles on a multilane highway to assess the impact of interference as a function of the vehicle and radar density as well as the deployment scenario. Our focus will be on direct interference from one radar to another. Indirect interference (i.e., scattered on objects) will be weaker and is ignored for conceptual simplicity.

Single-link interference

The interfering radar employs the FMCW waveform $s_{\mathrm{int}}(t) = \sum_{k=0}^{K-1} x_{\mathrm{int}}(t - k\tilde{T})$ where $x_{\mathrm{int}}(t) = e^{j2\pi(f_c t + 0.5\tilde{\alpha}t^2)}\mathrm{rect}_{\tilde{T}}(t)$, while the victim radar utilizes the same waveform as specified in (1) and (2). Here, $\tilde{\alpha} = \tilde{B}/\tilde{T}$, $\tilde{B}$ and $\tilde{T}$ denote, the chirp slope, sweep bandwidth, and chirp duration of the interfering radar, respectively. The samples (5) then become

$$y_{k,n} = \gamma e^{j2\pi(-\alpha\tau + f_c\nu)nT_s} e^{j2\pi f_c \nu kT} + \gamma_{\mathrm{int}} x_{\mathrm{int},k,n} + w_{k,n}.$$

Interference is generally much stronger than the desired backscattered signal because they are governed by the Friis free-space propagation equation and the radar equation, respectively [8]:

$$|\gamma_{\mathrm{int}}|^2 = P\frac{G_{\mathrm{trx}}\lambda^2}{(4\pi)^2 r^2}, \quad |\gamma|^2 = P\frac{G_{\mathrm{trx}}\sigma\lambda^2}{(4\pi)^3 d^4}, \tag{7}$$

where r is the distance between the interferer and the victim radar, d is the distance between the radar and the target, P is the transmit power, G_{trx} is the combined transmit and receive antenna gain, and σ is the RCS of the target. Hence, for similar d and r and typical values of σ, $|\gamma_{\mathrm{int}}|^2 \gg |\gamma|^2$. The nature of the interference depends on the total interference power (i.e., the aggregate power of the interference samples) and the level of coherence between the victim and interfering radar [9].

The total interference power, which is a function of the radar waveform parameters and signal delays, depends on the statistics of the samples $x_{\mathrm{int},k,n}$. The samples satisfy $|x_{\mathrm{int},k,n}|^2 \in \{0, 1\}$ depending on whether or not the interference signal at time (k, n) is in the bandwidth of interest of the victim radar. As a result, the overall power of the interference is $\mathbb{E}\{\sum_{k,n} |\gamma_{\mathrm{int}}|^2 |x_{\mathrm{int},k,n}|^2\} = f|\gamma_{\mathrm{int}}|^2 G_{\mathrm{p}}$, where G_{p} is the radar processing gain and f is the average interference probability.

How this total interference power manifests itself depends on its radar parameters, and interference can be classified as *coherent*, *incoherent*, or *partially coherent* [10]. Coherent interference occurs when the interferer uses the same parameters (α, T, and B) as the victim radar. In that case, the interfering radar signal leaking into the bandwidth of interest B_{s} of the victim radar (i.e., $f = 1$) leads to a ghost target, i.e., a peak in the delay-Doppler spectrum with very high power [11]. Ghost targets lead to false detections, which in turn may cause incorrect behavior of safety systems. Incoherent interference occurs when samples $x_{\mathrm{int},k,n}$ are independent random variables due to the interferer using very different waveform parameters (e.g., a different chirp pattern) so that the total interference power $f|\gamma_{\mathrm{int}}|^2 G_{\mathrm{p}}$ ends up as an increased noise floor. A noise floor increase resulting from interference can lead to more severe degradation in detection performance than an equivalent increase in thermal noise floor due to the sidelobes of the interference spectrum [10]. In between these two extreme cases, partially coherent interference occurs due to a slight mismatch in chirp slope or chirp duration or in the presence of phase noise (PN). This causes the energy of the ghost target peak (caused by coherent interference) to spread over the delay-Doppler domain. For incoherent or partially coherent interference, where the interference manifests itself as a noise floor, the overall interference power [or, equivalently, the average signal-to-interference-plus-noise ratio (SINR)] is a reasonable performance metric, while for fully coherent interference, percentiles are more meaningful.

To illustrate how a ghost target is spread out depending on the relative waveform parameters, Figure 3 shows the fast-time fast Fourier transform (FFT) output, i.e., the range FFT, corresponding to an interfering radar signal as a function of distance. The larger the difference in the chirp slope, the more the interference is spread out, due to the decrease in coherence; this affects the detection of targets in various ways. Incoherent interference may hinder the detection of low RCS targets (such as pedestrians and cyclists) over a large fraction of the delay-Doppler domain, whereas a (partially) coherent interference can mask even high RCS targets (vehicles) but in a smaller fraction of the delay-Doppler domain.

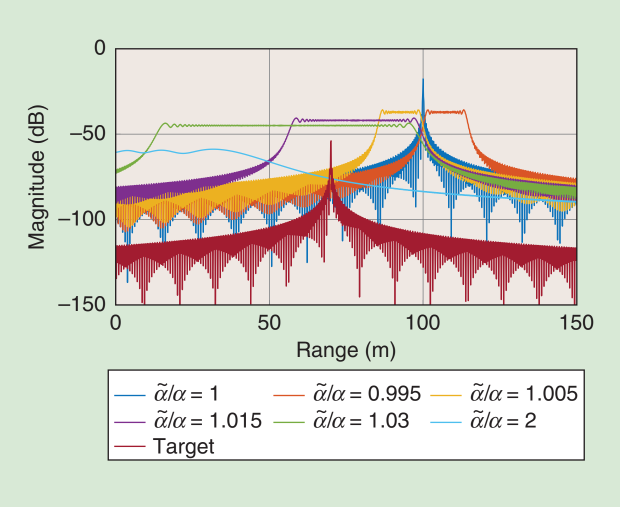


FIGURE 3. The FMCW range profiles in the presence of an interfering radar for various values of chirp-slope ratios, where $\tilde{\alpha}$ and α represent the chirp slopes of the interfering radar and the victim radar, respectively. The victim radar waveform parameters are $f_c = 77$ GHz, $B = 1$ GHz, and $T = 20\ \mu$s. The interfering radar has an identical chirp duration of $\tilde{T} = 20\ \mu$s but a varying sweep bandwidth of $\tilde{B}$ (thus, $\tilde{\alpha}$). The interference signal has a one-way propagation delay τ_{int} corresponding to a range of $R_{\text{int}} = 100$ m, while the desired target is located at $R = 70$ m. Due to an increased noise floor, the target may not be detected depending on its range and the chirp-slope mismatch between the victim and interfering radars.

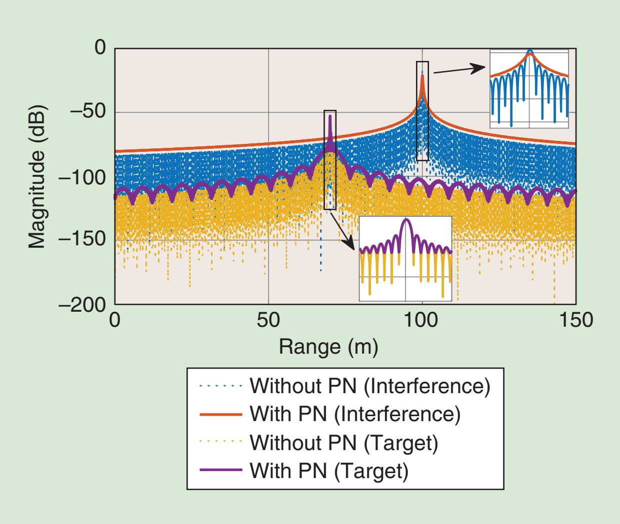


FIGURE 4. The range profiles of a victim FMCW radar in the presence of an interfering radar with identical chirp parameters (coherent interference), where both radars' oscillators have PN processes (the profiles without PN are also shown for comparison). The same parameters as those that are shown in Figure 3 are used with $\tilde{\alpha}/\alpha = 1$. The perfect range decorrelation of the interfering signal with the victim radar signal (due to independent PN processes at the victim and interfering radars) makes the spectral smearing effect more pronounced in the interference profile than in the target profile [15].

In practice, oscillators in FMCW radars do not have an ideal, impulse-like RF spectrum due to phase and frequency instabilities [12], [13]. In Figure 4, we demonstrate the effect of oscillator PN on the averaged range response of a victim FMCW radar when the oscillators of both the victim and interfering radars are subjected to phase noise processes with parameters $L_p = -70$ dBc/Hz (pedestal height) at $W_p = 200$ kHz (pedestal width) [13]. Range spectra are derived by computing the range FFTs of signal and interference powers averaged over the randomness of PN which is modeled as a zero-mean, wide-sense stationary random process under the assumption of white FM PN in the oscillator [14, Sec. V]. As observed from Figure 4, the oscillator PN induces the spectral smearing of target and interference profiles, thereby causing a loss of details in the spectrum, which deteriorates detection performance and leads to the masking of weak targets.

Network interference

The aforementioned interference analysis can be extended to a complete network, for instance, on a multilane road, by employing a stochastic geometry approach [8]. As shown in Figure 5, consider a victim radar surrounded by L lanes of traffic, with lane separation R, each modeled as a 1D Poisson point process $\Phi(\mathbf{x})$ with intensity $1/\Delta$ (so that Δ is the expected distance between vehicles and $\mathbf{x}$ is a vehicle location along a road) [3]. Radars (here, one per side of the vehicle) are incoherent and can have different chirp durations, but otherwise share the same bandwidth B, duty cycle $u \in [0,1]$ (i.e., the fraction of time the radar is transmitting), and FOV.

We recall that for antennas with a narrow FOV, the antenna gain is $G \approx 4\pi/(\phi\theta)$, where ϕ is the beamwidth in the elevation domain and θ the beamwidth in the azimuth domain. This means that a radar with 1 and 30° elevation and azimuth beamwidth, respectively, will have a gain of approximately 31.3 dBi. The expected value of the interference probability f is easily found to be $f = u\alpha\tau_{\max}/B$. Here we made use of the following asymptotic results: 1) when $\tilde{\alpha} = \alpha$, the probability of interference is uB_s/B and the interference lasts an entire chirp duration; 2) when $\tilde{\alpha} \ll \alpha$, the probability of interference is u and the duration is $\tilde{\alpha}\tau_{\max}/(\alpha - \tilde{\alpha})$;

FIGURE 5. Network interference on a six-lane highway. Interference from interfering radars is aggregated and depends on the properties of the individual radars as well as on the placement of vehicles on the road.

and 3) when $\tilde{\alpha} \gg \alpha$, there are $\tilde{\alpha}/\alpha$ simultaneous interferers, each lasting for a duration of $\alpha\tau_{\max}/(\tilde{\alpha}-\alpha)$. As a result, the aggregate interference seen by the victim radar due to interference from lane $\ell \in \mathbb{Z}$ (indexed with reference to the victim radar) is

$$I_{\mathrm{p}}(\ell) = \sum_{\mathrm{x} \in \Phi(\mathrm{x}) \cap \mathrm{x} \in \mathrm{FOV}} P f \frac{G_{\mathrm{trx}}\lambda^2}{(4\pi)^2 r^2(\mathbf{x})}, \tag{8}$$

where $r(\mathbf{x}) = \sqrt{\ell^2 R^2 + x^2}$, where x is the 1D position along the road ranging from $\ell R/(\tan\theta_{\mathrm{p}}/2)$ to $+\infty$. Here, θ_{p} is the minimum of the forward and backward FOV. Accordingly (with a slight abuse of notation), the interference averaged over the locations of the interferers is

$$\begin{aligned}\mathbb{E}\{I_{\mathrm{p}}(\ell)\} &= P\frac{G_{\mathrm{trx}}\lambda^2}{(4\pi)^2}\frac{f}{\Delta}\int_{\frac{\ell R}{\tan\theta_{\mathrm{p}}/2}}^{+\infty}\frac{1}{\ell^2R^2+x^2}\mathrm{d}x \\ &= P\frac{G_{\mathrm{trx}}\lambda^2}{(4\pi)^2}\frac{f}{\Delta}\frac{1}{\ell R}\frac{\theta_{\mathrm{p}}}{2},\end{aligned} \tag{9}$$

while for $\ell = 0$, $r(\mathbf{x}) = x$, where we need a certain safety margin to avoid singularities, we set x from Δ to $+\infty$, leading to $I_{\mathrm{p}}(\ell = 0) = P(G_{\mathrm{trx}}\lambda^2/(4\pi)^2)\, f\,(1/\Delta^2)$. An example of network interference on a six-lane highway is shown in Figure 6, as a function of the average intervehicle spacing Δ for a vehicle target 150-m away with an RCS of 10 m^2. The analytical result shows the impact of the interference of nearby vehicles, leading to an orders of magnitude reduction of the SINR. For small Δ, the interference is larger from passing lanes, while for large Δ, oncoming traffic dominates. We also observe that, even though the interference power can be large, factor f reduces its impact significantly. The example $f \approx 0.01$ leads to a 20-dB reduction in interference. In this example, the target can be detected in spite of incoherent network interference, while a pedestrian target with a smaller RCS (such as 0.1–1 m^2) farther away than 50 m would be difficult to detect.

Intermediate conclusion

From the aforementioned analysis, we found that interference can manifest itself in different ways and increase the occurrence of both missed detections and false alarms. Due to the nature of FMCW signals (in particular, the small instantaneous bandwidth), there is a natural robustness to interference. Both the total received interference power and the mutual coherence between victim and interfering radar play an important role. As a rule of thumb, the signal-to-interference ratio (SIR) for a target at distance d due to the power transferred by an interferer at distance r to a victim radar can be determined as follows: the useful signal power (the peak of the periodogram) and the interference power are

$$S = |\gamma|^2 G_{\mathrm{p}}^2, \quad I \le f\,|\gamma_{\mathrm{int}}|^2 G_{\mathrm{p}} G_{\mathrm{I}}, \tag{10}$$

where $G_{\mathrm{I}} \in [1, G_{\mathrm{p}}]$ depending on the level of coherence of the interference (i.e., $G_{\mathrm{I}} = 1$ for incoherent interference and $G_{\mathrm{I}} = G_{\mathrm{p}}$ for coherent interference). The level of coherence can be characterized through signal-to-interference mitigation gain, which is a function of FMCW waveform parameters of victim and interfering radars [10], [16]. Hence,

$$\mathrm{SIR} \ge \frac{|\gamma|^2 G_{\mathrm{p}}^2}{f\,|\gamma_{\mathrm{int}}|^2 G_{\mathrm{p}} G_{\mathrm{I}}} = \frac{\sigma r^2}{(4\pi) d^4}\,\frac{G_{\mathrm{p}} B}{u\alpha B_{\mathrm{s}} G_{\mathrm{I}}}. \tag{11}$$

The first factor is out of the designer's control, while the second factor can be optimized via the duty cycle (small u), chirp slope (small B_{s}/B), FOV (thereby reducing f), and effective processing gain (increase $G_{\mathrm{p}}/G_{\mathrm{I}}$) to ensure that the SIR is much large than 1. Our results indicate that incoherent interference leads to a significant increase in the noise floor (tens of dBs) so that it can reduce the ability to detect weak targets. Nevertheless, for nearby targets or targets with a high RCS, the SIR margin is sufficient to allow for reliable detection. When the interference is partially coherent this margin drops significantly.

Interference mitigation strategies

The impact of interference includes ghost targets and increases in noise floor. Both are detrimental to radar operations. Approaches that deal with interference can be grouped as either reactive, which aim to reduce the impact of interference after it has occurred, or proactive, which aim to avoid or reduce interference by design. In the next section, we describe various reactive strategies as well as three such proactive strategies: a quasi-orthogonal FMCW waveform, a low-rate data communication between radar Txs, and an OFDM radar approach. Each of these approaches lead to high interference suppression, with limited impact on radar performance. In addition, the OFDM

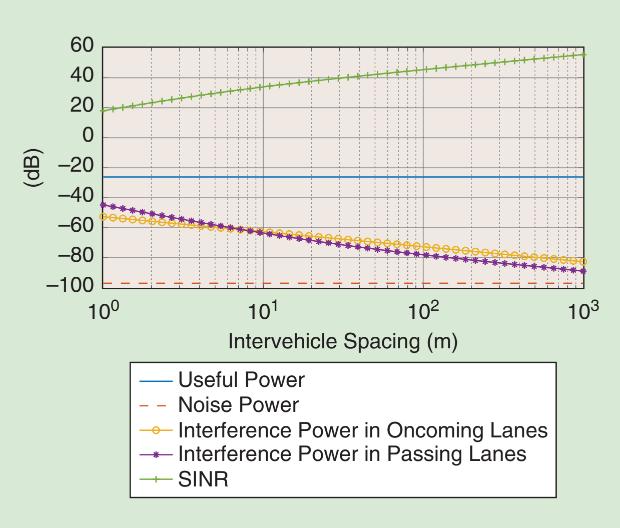


FIGURE 6. An incoherent network interference on a six-lane highway as a function of average vehicle spacing. The distance to the target is 150 m, with a 10-μs chirp time, $\sigma = 10$ m^2, and 10 dBm of transmit power. It features a 1-GHz bandwidth, $T = 30$ μs, a 150-m maximum range, a 50-MHz ADC bandwidth, 10 dBm of transmit power (the same for the front and back ends), a 20% duty cycle, a 100-μs frame duration, and 30 and 90° FOVs in the forward and backward directions, respectively.

radar approach can enable high-rate communication links between vehicles.

Standard (reactive) approaches

Extensive studies have been conducted in the context of the MOSARIM project, where a broad range of time- or frequency-domain signal processing techniques were proposed to mitigate FMCW and pulsed-radar interference. These techniques are capable of deleting instantaneous interference, which exists for a limited time, or bandlimited interference, which pollutes a specific portion of the whole radar band (while no solution is offered for the worst-case recurring or wideband interference [1]).

The current attitude toward interference mitigation in the industry focuses on various techniques, including pulse-to-pulse processing and removing polluted pulses, sniffing and avoiding used frequencies, using frequency diversity, using narrow main beams, or sidelobe null steering [3]. Generally, these techniques are reactive strategies that focus on eliminating interference after it occurs, making it infeasible for highly dynamic VANETs, which require for ultralow latencies. Other reactive strategies exploit the sparsity of useful signal and interference components in different transform domains, namely, the DFT and short-time Fourier transform domains, respectively, to extract the desired signal component [17] or to solve a sparse recovery problem that helps reconstruct the intervals in the range spectrum spoiled by interference [18], which belongs to a general class of time-domain excision approaches [19]. Such strategies can retain the desired target response while eliminating the interference, albeit at the cost of losing information and introducing artifacts [19].

In addition to these reactive strategies, there also exist cognitive approaches that utilize idle time-frequency resources for FMCW transmission via slope and bandwidth adaptation [20]. Interference-avoidance techniques can also be more invasive, such as notifying the driver, disabling the sensor feature, shifting functionality to another sensor, and reducing CFAR detection sensitivity [3]. However, these avoidance mechanisms either decrease radar detection performance or disable the radar completely.

Quasi-orthogonal waveforms

Concept

From the interference analysis, we established that the interference is proportional to $f = u\alpha\tau_{\max}/B$, where u is the radar duty cycle, α is the chirp slope, $\tau_{\max}$ is the maximum target round-trip time, and B is the radar bandwidth. Accordingly, by decreasing the chirp slope or, equivalently, increasing the chirp duration, interference can be mitigated. We consider a class of signals containing linear chirps with the same chirp rate, a fixed frame rate, and a fixed duty cycle. A chirp $x(t)$ from (2) and a delayed chirp $x(t-\Delta\tau)$ repeated with period T have the following power leakage/coupling for $\Delta\tau \neq 0$:

$$C(\Delta\tau) = \left|\frac{1}{T}\int_0^T e^{j[\varphi(t)-\varphi(t-\Delta\tau)]}dt\right|^2 \leq \frac{1}{\pi^2 B^2 (\Delta\tau)^2}. \tag{12}$$

Because $B\Delta\tau/T$ is the instantaneous frequency difference between the chirps, the coupling is the same as two sinusoids with frequency difference $B\Delta\tau/T$. As a result, we conclude that these waveforms are quasi-orthogonal. As $\Delta\tau$ was arbitrary, this property is maintained under random starting and arrival times of FMCW waveforms. From (12), it is possible to derive the number of signals N that cause acceptable interference, i.e., smaller than the power backscattered from a typical target. Suppose chirp duration T is divided into N segments of duration $\Delta\tau$, so that $\Delta\tau = T/N$; then, using (7),

$$P\frac{G\sigma\lambda^2}{(4\pi)^3 d^4} \geq PC(\Delta\tau)\frac{G\lambda^2}{(4\pi)^2 r^2} = P\frac{N^2}{(\pi BT)^2}\frac{G\lambda^2}{(4\pi)^2 r^2}$$

$$\Rightarrow N \leq TB\frac{\sqrt{\sigma}}{r},$$

if r and d take on similar values. For $r = 500$ m, $B = 1$ GHz, and $\sigma = 0.1\ \mathrm{m}^2$, using one long chirp of duration $T = 10$ ms leads to more than 6,000 quasi-orthogonal waveforms (compared to 12 for a conventional chirp with $T = 20\ \mu$s). The challenges of retrieving velocity and range data as well as physically realizing such a radar is briefly described in the next section.

Signal processing

In (5), even though the speed and range appear in an ambiguous combination in a single chirp, this expression is only an approximate representation of a Doppler shift because we neglected several constants and small terms, which are negligible for small T. For long chirps, these neglected values should be considered so that the range-speed ambiguity can be resolved within a single chirp. Formally,

$$\begin{aligned} y(t) &= r(t)x^*(t) \\ &= \tilde{\gamma}e^{j2\pi f_c \nu t}e^{-j2\pi\alpha\tau t}\mathrm{rect}_T(t-\tau)\,e^{j\varphi_0}e^{j2\pi t^2\nu\alpha} + w(t)x^*(t), \end{aligned} \tag{13}$$

where $r(t) = r_0(t)$ with $r_k(t)$ being defined in (3), $\tilde{\gamma}$ is a real quantity denoting the target reflectivity, and $\varphi_0 = \pi\alpha\tau(\tau - 2f_c/\alpha)$ is an absolute phase. The Fourier transform $Y(f)$ of $y(t)$ will have a maximum at $f^* = f_c\nu - \alpha\tau$, with

$$Y(f^*) = \tilde{\gamma}e^{j\varphi_0}\int_0^T e^{j2\pi t^2\nu\alpha}\mathrm{d}t. \tag{14}$$

Hence, finding the maximum of $|Y(f)|^2$ yields an estimate of $f_c\nu - \alpha\tau$, which allows us to write $\varphi_0 = \pi((f_c\nu - f^*)/\alpha) \times (f_c(\nu - 2) - f^*)$. We can then invert (15) to solve for $\nu = 2v/c$. The solution must be found numerically, making use of the fact that the target reflectivity $\tilde{\gamma}$ is real and positive and that the complex angle of the integral increases monotonically with velocity within a significant velocity span.

When comparing the outlined approach to that of the slow-/fast-time FMCW radar case mentioned previously, note the principal difference is that instantaneous velocity is determined by measuring the complex amplitude. For pulse-Doppler radar, velocity is found by locating the target peak in the Doppler spectrum. The present approach assumes that just one target is present for any beat frequency, whereas pulse-Doppler radar may handle several targets within the same range/Doppler resolution cell. On the other hand, the attained beat-frequency resolution is refined in proportion to the prolonged sweep. In effect, both methods exhibit the same low probability that two separate targets should be superimposed within the same resolution bin.

The single slow-sweep method for obtaining nearly orthogonal waveforms contains a further challenge compared to that of pulse-Doppler radar, apart from the novel signal processing required. Indeed, the steep ramps of pulse-Doppler FMCW frequency offset the target response as compared to those of the transmit signal. Leakage between the Tx and receiver (Rx) will manifest itself at the mixer output as a high-amplitude low-frequency peak. In a typical FMCW radar, this unwanted response is removed using a configurable analog electronic filter; however, the implementation of analog high-pass filters with very low cutoff frequencies at the chip level is not feasible. Presently, radar is based on the subtraction of leakage appearing as a dc component with a feedback loop rather than filtering with respect to range and Doppler-frequency offsets. The conceived radar scheme in Figure 7 indicates that (15) can be slightly modified by digitally modifying the signal phase by some offset, thereby changing the unambiguous speed range for (15). To effectively cover a sufficiently large velocity span, two or three such velocity channels should be processed in parallel. The overall processing burden still remains fully reasonable.

Implications

The freedom of multiple orthogonal radar channels can be brought into practice in different ways. The considered case with a very large number of channels allows for the convenient method of simply not requiring any common scheduling of the channels adopted (apart from the several radars that may be located in the same vehicle, in which case channel coordination is easy to achieve). The number of vehicles in such close proximity to each other that an interference conflict is imminent will be much smaller than the number of available channels. Just by selecting the radar channels randomly, the chances are good that there will be no interference. In the case of interference, the individual vehicle will then randomly pick another channel and, with high probability, the conflict is thereby resolved.

Coordinated transmission via wireless communication

Another way to mitigate interference is by coordinating automotive radars through communications so that radars are assigned disjointed frequency-time-space resources, making use of the fact that 1) radars receive over a small fraction B_s/B of the available bandwidth, 2) radars are only active a fraction u of the time, and 3) radar signals are blocked (mostly by other vehicles) and limited by the FOV. The assignment requires coordinating the allocation of frequency-time resources through distributed network communication. Such communication can be achieved either via a dedicated technology, such as the 802.11p standard or cellular vehicle-to-everything (C-V2X) communication.

Alternatively, one can exploit the similarity of the radar circuitry to standard communication hardware and upgrade automotive radars to joint radar communication units (RCUs), which use the same hardware both for radar and communications

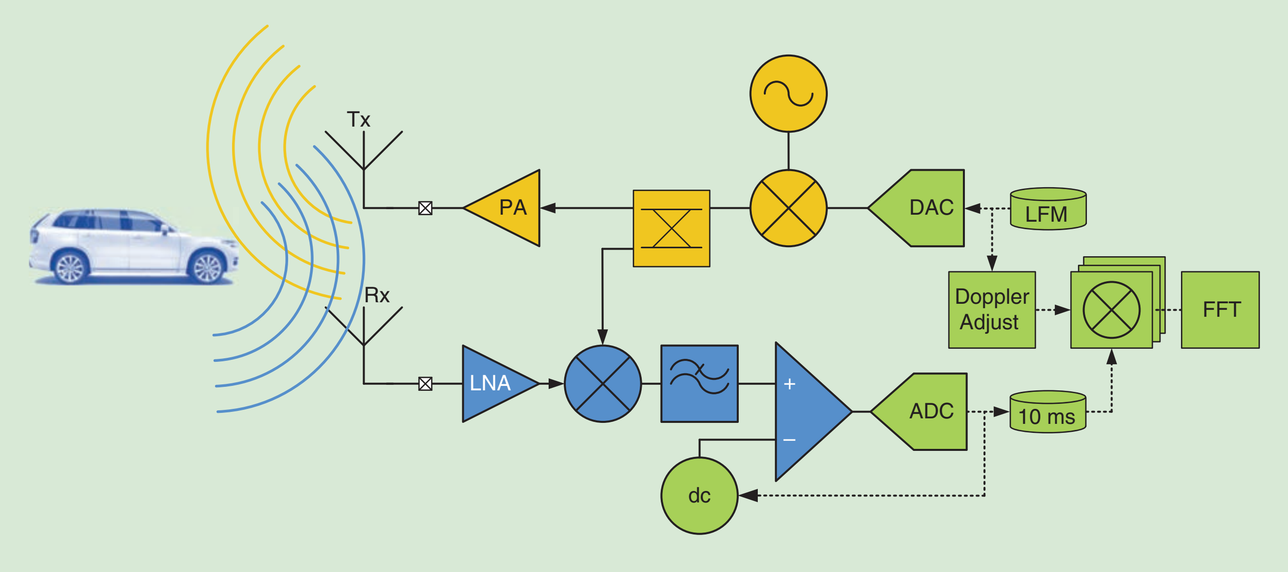


FIGURE 7. The simplified scheme of a single slow-sweep FMCW radar, which allows for many orthogonal signals. The linear FM (LFM) waveform is obtained digitally at baseband and is upconverted in the analog domain. This upconverted signal is mixed with the receive signal yielding downconversion to a signal with bandwidth essentially set by 50 MHz. The dominant power in the downconverted signal will come from the transmit signal which, however, appears as a dc component. This is removed by a dc canceler stage after which the signal is digitized. The canceler operates on the principle of minimizing at ADC output at some rate much slower than the sweep time and has the vital function of avoiding ADC saturation. In the digital domain further mixing with Doppler offsets may be required to cover a very wide range of target speeds. DAC: digital–analog converter.

(implementing data link and physical layer functions). The radar and communication functionality are time multiplexed, where communication occurs over a fixed and dedicated communication bandwidth (with the bandwidth limited by the ADC), which is free of radar transmissions. The time multiplexing is possible because of the idle period. As a result, when $u \approx 1$, RCUs cannot be used and 802.11p or C-V2X is more appropriate. Nevertheless, in this article, we consider the RCU approach, as it is readily modified when using another dedicated communication technology. As is typical in VANETs, communication is unacknowledged when using distributed MAC-based carrier-sense multiple access (CSMA).

The goal of the communication is to assign radars to time slots so that different radars remain quasi-orthogonal. In contrast to the long chirps described previously, we consider standard short chirps, thus limiting the number of transmissions per chirp period. The frequency-time resources are shared for three RCUs, namely, r_i, r_j, and r_k, as illustrated in Figure 8. The basic principle is as follows. Each vehicle initially assigns starting times to the automotive RCUs mounted on this vehicle through a central processor. These starting times are broadcast to neighboring vehicles during a communication slot. All the RCUs on a vehicle broadcast short control communication packets at the same time over the communication band. The broadcast communication packet includes information about the starting times used by all the RCUs on that vehicle. The other RCUs or vehicles that receive this information store it in a database and allocate themselves nonoverlapping starting times based on the stored information. A priority index is used to prioritize the dissemination of the resource allocation of a larger group of vehicles to avoid fluctuations in the distributed VANET.

A practical implementation requires synchronization among radars, which can be achieved through GPS or via a dedicated synchronization protocol. Using reasonable synchronization requirements (around a 1–2-μs error), the authors in [21]

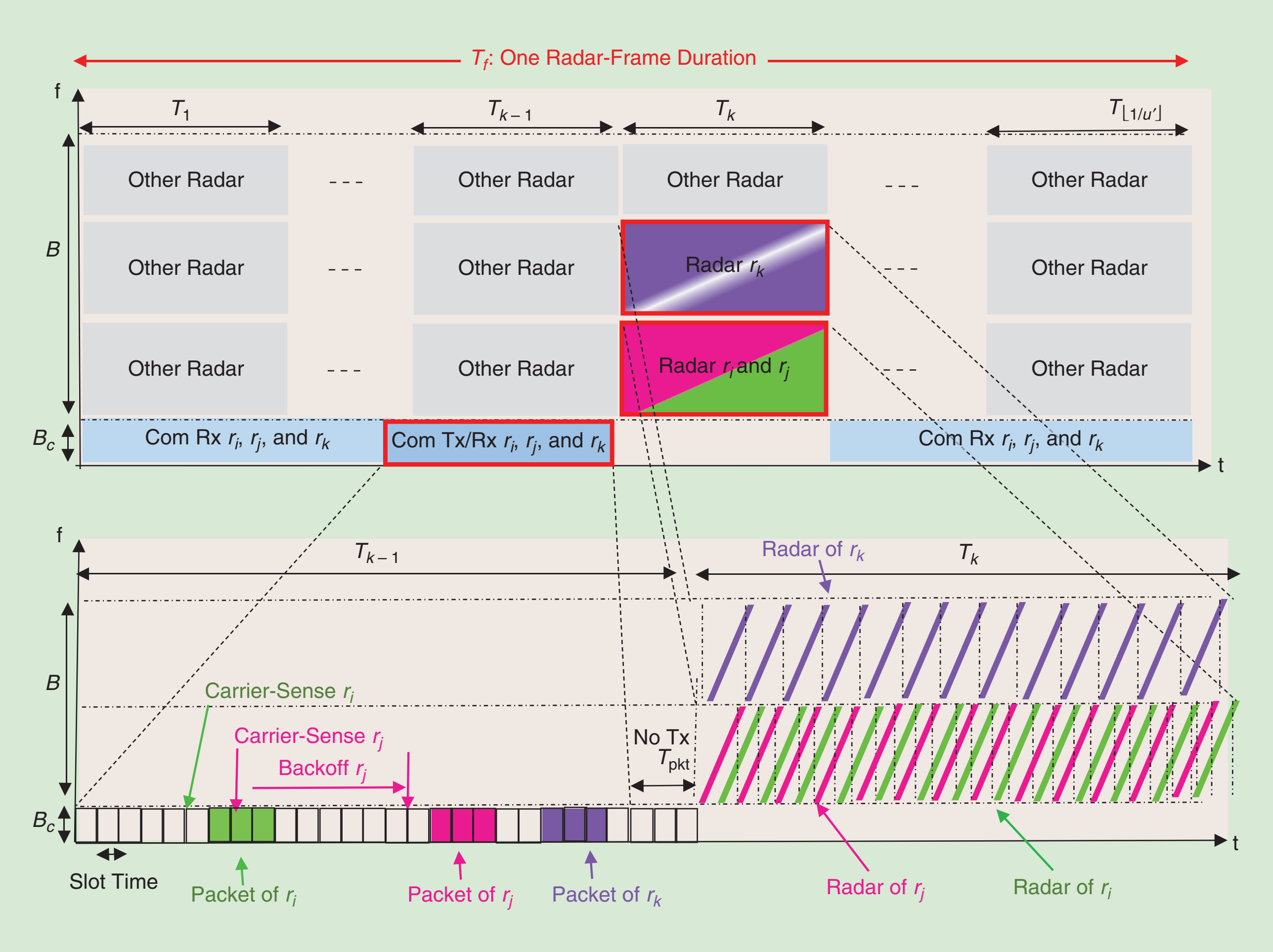


FIGURE 8. The sharing of frequency-time resources by three RCUs, r_i, r_j, and r_k. The whole bandwidth is divided into a communication channel of bandwidth B_c and sweep bandwidth B, where various types of automotive radars may use portions of B. One radar frame T_f is divided into time slots T_k, $1 \le k \le \lfloor 1/u' \rfloor$, where $u' = (N+1)T/T_f$ is the modified radar duty cycle. The radar frame is divided into time slots, where radar transmission/reception takes place in T_k, communication transmission is done prior to radar at T_{k-1}, and the reception at any time except T_k. Each RCU broadcasts its packet over the communication channel through carrier-sense multiple access, which includes the starting time and frequency band information of its radar transmission. The interference is resolved because r_i and r_j are allocated to quasi-orthogonal time slots, whereas r_k is allocated to a different frequency band. Any conflicts in frequency-time resource sharing that occur in one frame T_f are resolved in the following frames. Com: communication.

demonstrated significant reductions in radar interference within a few tens of ms. Figure 9 shows the interference probability (i.e., the expected value of f) as a function of the number of interfering radars, with and without coordination. The potential of such protocols to adapt radar signals according to changing traffic conditions via communications offers intelligent radar sensing strategies, such as cooperative localization, disabling unnecessary sensing, and so on.

Joint radar and communication

A third, more forward-looking alternative is to exploit the fact that radar and communication systems operate in similar frequencies and develop a system that can perform the dual role of radar and communication [22], called *RadCom*. Both the pilot and the data from the transmitted signal can be exploited for radar functions when processing the backscattered communication signal. A prominent candidate for this is OFDM, which is the de facto waveform for all cellular and Wi-Fi-based standards, due to its flexibility and robustness to wireless propagation effects. OFDM has also been studied extensively as a radar waveform [24]–[28] but is limited by its ADC bandwidth (generally orders of magnitude smaller than the radar bandwidth), which in turn limits radar resolution. Assuming a high-FMCW chirp slope, the OFDM and FMCW radars with identical waveform parameters exhibit similar performance in terms of accuracy and resolution [23]. OFDM provides the additional capability of communications at the expense of increased hardware complexity due to OFDM modulator/demodulator operations [23].

A way around this problem is the use of stepped-frequency OFDM, which involves consecutive OFDM frames, each transmitted with a different carrier frequency [29], [30]. The main rationale behind the use of stepped-frequency OFDM as a RadCom waveform is to surpass the range-resolution limitation of conventional OFDM radar (which is imposed by the ADC bandwidth) via frequency hopping across individual OFDM frames with low baseband bandwidth [30], while maintaining standard wideband OFDM as a special case.

Figure 10(a) illustrates the exemplary time-frequency plot of a stepped-frequency OFDM waveform. To avoid interference, different vehicles are assigned orthogonal resources, as shown for three vehicles. Hence, the stepped-frequency OFDM can exploit high resolution offered by the large total bandwidth $MN\Delta f$ using the joint processing of M individual OFDM frames on different carriers, while simultaneously requiring a low-rate ADC to sample small baseband bandwidth $(N\Delta f)$ OFDM blocks. For each carrier, L OFDM symbols of duration T_{sym} are sent, constituting a frame. The choice of N, M, L, and Δf and the hopping pattern provides flexibility in the RadCom waveform and enables us to provide radar performance similar to that of an equivalent wideband OFDM radar but with low-rate, low-cost ADCs [30].

Signal processing and resource allocation

Under standard assumptions (i.e., a cyclic prefix longer than τ [26], [28] and a small Doppler approximation [24]), the received symbol on the nth subcarrier for the ℓth symbol of the mth frame [considering the same radar environment with a single target as specified in (3)] can be written as [30]

$$y_{m,\ell,n} = \gamma x_{m,\ell,n} e^{-j2\pi(f_m+n\Delta f)\tau} e^{j2\pi f_0(mL+\ell+1)T_{\text{sym}}\nu} + w_{m,\ell,n}, \quad (15)$$

where $x_{m,\ell,n}$ denotes the complex data or pilot symbol, γ is the complex channel gain, and $w_{m,\ell,n}$ is the additive noise term with variance σ^2. Delay estimation can be performed by matched filtering the data cube $y_{m,\ell,n}$ across frame-frequency dimensions (m and n directions), while processing along frame-time dimensions (m and ℓ directions) can provide a Doppler estimate. Frequency hopping across consecutive OFDM frames introduces delay-Doppler coupling, which can be overcome by incorporating phase correction terms in the DFT implementation of matched filtering [30].

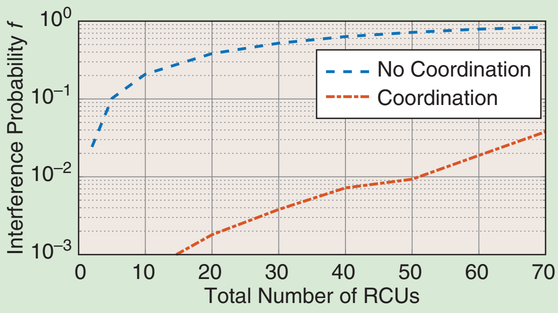


FIGURE 9. The radar interference probability f after one time frame (20 ms) with a coordinated transmission for various numbers of radars for 10,000 Monte Carlo simulations, $B = 1$ GHz, $B_s = 50$ MHz, $T = 20$ μs, $f_c = 79$ GHz, and $K = 99$.

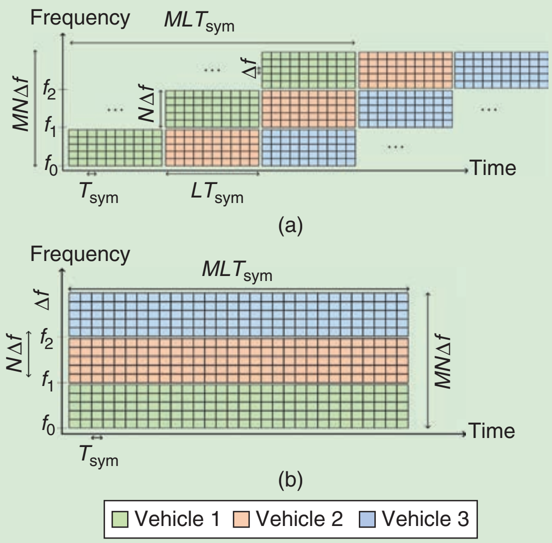


FIGURE 10. A time-frequency resource coordination in a joint radar communications vehicular network for (a) a stepped-frequency OFDM and (b) a narrow-band OFDM.

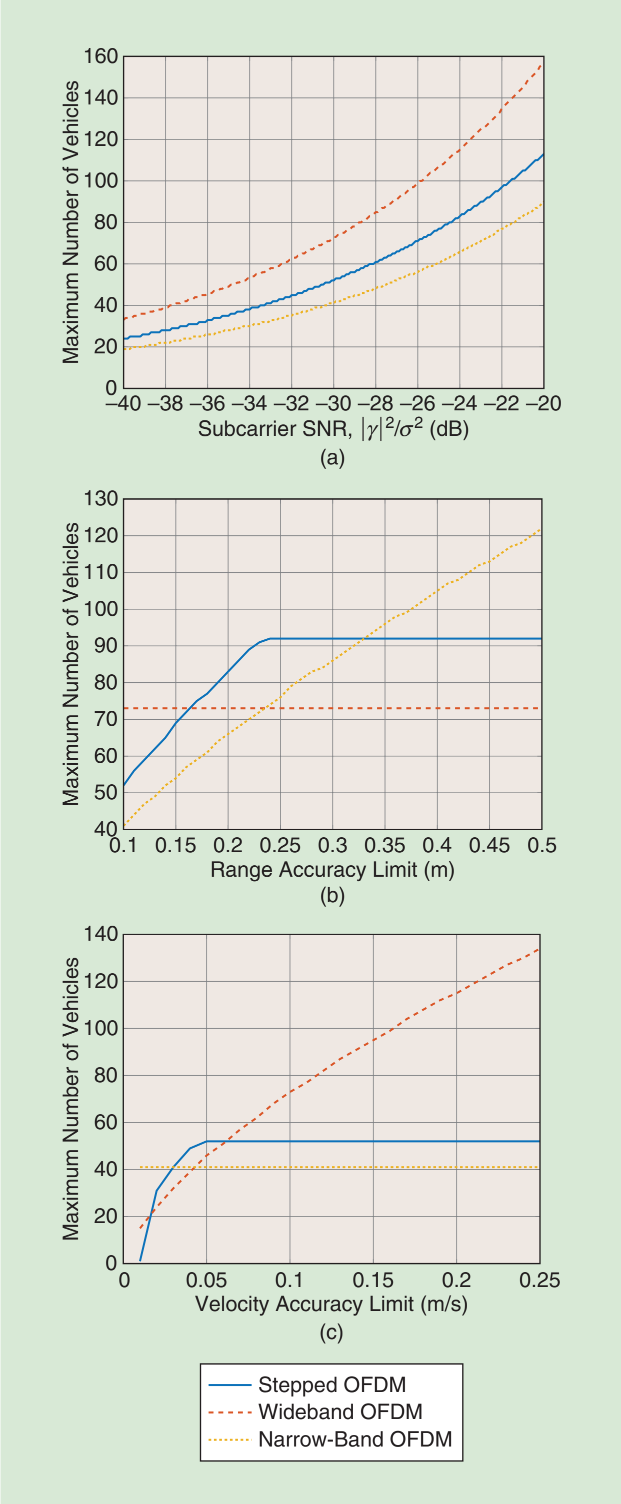


FIGURE 11. The maximum number of vehicles that can be accommodated within a time-frequency resource defined by $B_{\text{tot}} = 1\,\text{GHz}$ and $T_{\text{tot}} = 30\,\text{ms}$ for the three different OFDM waveforms, where $f_0 = 77$ GHz, $\Delta f = 500$ KHz, and $T_{\text{cp}} = 400\,\text{ns}$. Stepped OFDM and narrow-band OFDM schemes require an ADC with a 50-MHz rate, whereas wideband OFDMs require a 1-GHz-rate ADC. (a) The subcarrier SNR, where range and velocity accuracy limits are 0.1 m and 0.1 m/s, respectively. (b) The range-accuracy limit, where the subcarrier SNR is –30 dB and the velocity accuracy limit is 0.1 m/s. (c) The velocity accuracy limit, where subcarrier SNR is –30 dB and the range accuracy limit is 0.1 m.

A time-frequency resource allocation scheme coordinated by a central unit (e.g., a 5G base station) helps to alleviate mutual interference among radars on different vehicles, similar to that of conventional OFDM radar networks [26, Ch. 4]. Resources can be assigned to maximize the number of vehicles that can be fit into a given time-frequency block such that each vehicle meets preset radar accuracy requirements. To characterize radar accuracy, we employ the Cramér–Rao bound [26] on variances of unbiased estimates of delay and Doppler parameters using the signal model in (15).

Figure 11(a)–(c) presents exemplary results for the three different OFDM schemes [i.e., stepped frequency, narrowband, and wideband where, in the latter case, each vehicle uses the total bandwidth (and thus requires an ADC with a 1-GHz sampling rate)] for duration LT_{sym} and then remains silent for duration $(M-1)LT_{\text{sym}}$). As shown in the figure, the stepped-frequency OFDM radar can support more vehicles in a given spectral resource than it can in the conventional narrow-band OFDM radar using the same hardware requirements because the former offers the flexibility to trade off a decrease in Doppler accuracy for an improvement in ranging accuracy (frequency hopping increases ranging accuracy and reduces Doppler accuracy). In Figure 11(b), because wideband OFDM is essentially limited by the velocity accuracy constraint, relaxing the range accuracy constraint does not further improve its performance [similar to the narrow-band OFDM in Figure 11(c)].

Finally, we note that the stepped OFDM provides a design tradeoff between the narrow-band and wideband OFDM schemes, retaining the improved resolution and accuracy properties of the wideband OFDM using significantly reduced hardware requirements, as in the case of the narrow-band OFDM.

Outlook and challenges

In this section, we consider the primary research and development challenges for the coming years. For communication-based interference mitigation strategies, the coexistence between radar and communication signals is an important challenge. For joint radar and communication signals, there is the potential of a revolution of cellular-type signals (e.g., 5G New Radio) to be reused for radar purposes [22], opening new synergies and avoiding the need for dedicated RF hardware all together. The extended frequency bands made available for 5G are interesting by themselves due to the possible improvement in radar resolution, which, together with the already standardized orthogonal signaling, establishes an exciting area for automotive sensing. The main challenge is to find solutions that will enable radar and communication functionalities with such a low information latency that vehicle safety is not compromised in any traffic scenario. These solutions should include techniques for a fair distribution of the available time and frequency space for all users. It must also secure a low data loss for both radar and communication, which, of course, is the aim of minimizing the possibility of interference.

For the generation and detection of slow chirps, new hardware architectures will be necessary; this will push a migration from analog toward digital electronics and signal processing, which will pave the way for technologies such as imaging radar. The other modulation waveforms proposed in this article will also require hardware that differs from the current radar designs. Analog-to-digital (and vice versa) conversion will be close to the RF front end, making more complex digitally generated and filtered waveforms possible. To move this development forward, we will implement a demonstrator platform, complete with mm-wave front ends, high-speed digital signal generation and acquisition, and the independent generation of arbitrary interference. The methods outlined in this article will be tested and evaluated on the demonstrator platform in a realistic environment. The intention is to use this demonstrator (see Figure 12) to verify the theoretical analysis regarding interference probability and SINR for different types of modulation, and verify the speed measurement method used for slow-ramp modulation.

Further development would include the integration of critical electronic components in CMOS technology to ensure that a complete solution is feasible to implement commercially. Advanced CMOS technologies can also facilitate the implementation of alternative waveforms on automotive radars, such as phase-modulated continuous wave (PMCW). Compared to the widely used FMCW radar, the PMCW waveform has the major disadvantage of requiring very high-rate ADCs to sample wideband code sequences. Conversely, it possesses several advantages that make it attractive for future deployments, including improved robustness to interference via proper code design, not requiring a highly linear frequency ramp synthesizer, and the inherent applicability to multiple-input, multiple-output radar configurations through code orthogonality across multiple antennas. From the perspective of radar interference mitigation and radar communications convergence, we expect that the main focus of the automotive industry in the coming years will be on the cost and integration of both analog and digital functions on the same silicon chip to reduce the likelihood of hardware failure.

Vehicle radars can also be expected to operate in higher frequency bands such as 100–300 GHz to enable more bandwidth, reduce costs, and miniaturize hardware. It is possible to influence regulators to include some level of standardization in automotive radars, which is needed for mitigating interference among different automobile brands [3], before a new frequency spectrum is made available in the higher RF bands. It is therefore timely to conduct research and discuss the developmental challenges of automotive radar interference before it becomes a problem.

Acknowledgment

This research was supported by Vinnova grant 2018-01929.

Authors

Canan Aydogdu (canan@chalmers.se) received her B.S. degree from the Electrical and Electronics Engineering Department, Bogazici University, Istanbul, Turkey, in 2001 and her M.S. and Ph.D. degrees from the Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey, in 2003 and 2011, respectively. She is currently a Marie Skłodowska-Curie Fellow at Chalmers University of Technology, Gothenburg, Sweden, where she works on cross-layer multihop sensing systems for future energy-efficient self-driving car networks. Previously, she was an associate professor at Izmir Institute of Technology, Turkey.

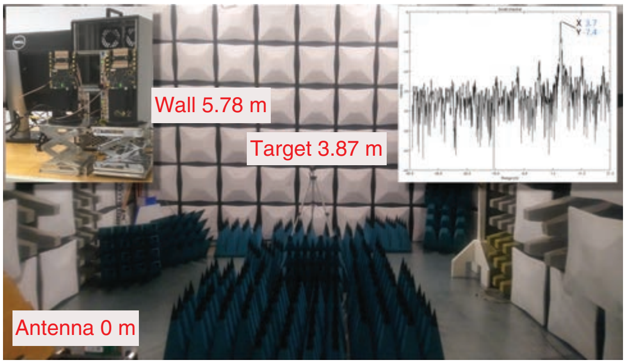


FIGURE 12. A radar test setup in an anechoic chamber with a trihedral target seen from the radar position. The upper-left inset shows the experimental radar demonstrator, while the upper-right inset shows a typical acquired range profile.

Musa Furkan Keskin (furkan@chalmers.se) received his B.S., M.S., and Ph.D. degrees from the Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey, in 2010, 2012, and 2018, respectively. He is currently a postdoctoral researcher at Chalmers University of Technology, Gothenburg, Sweden. His research interests include intelligent transportation systems, radar signal processing, and radar communications coexistence.

Gisela K. Carvajal (gisela.carvajal@qamcom.se) received her electrical engineering degree from the Universidad de Carabobo, Valencia, Venezuela, in 2005, her M.S. degree through the Erasmus Mundus Merit program (master of research on information and communication technologies) from both the Universitat Politècnica de Catalunya, Barcelona, Spain, and Politecnico di Torino, Italy, in 2008, and her Ph.D. degree from Chalmers University of Technology, Gothenburg, Sweden, in 2014. Currently, she is with QAMCOM Research, Gothenburg, where she works on the simulation and calibration of radar systems. Her research interests include algorithm development, signal and image processing for radar remote sensing, and scattering analysis.

Olof Eriksson (olof.eriksson@veoneer.com) received his M.Sc. degree in electrical engineering from the Faculty of Engineering LTH at Lund University, Sweden, in 1979. He is a research specialist at Veoneer Sweden AB, Vårgårda, where he works on advanced electromagnetic sensor technologies, including radar systems.

Hans Hellsten (hans.hellsten@saabgroup.com) received his Ph.D. degree in theoretical physics from Stockholm University, Sweden, in 1980. He is currently a senior radar expert with Saab Surveillance, Linköping, Sweden, and an adjunct professor at Halmstad University, Sweden. Previously, he was the director of research at the Swedish Defence Research Agency, Linköping. He has written numerous books and received many patents. His research interests include radar system design and electromagnetic modeling and signal processing.

Hans Herbertsson (hans.herbertsson.external@veoneer.com) is a research project leader at Veoneer Sweden AB, Vårgårda, where he leads concept development activities with a focus on electromagnetic sensor technologies. Previously, he managed microwave development activities in telecommunications as a product owner and product manager at Ericsson AB.

Emil Nilsson (emil.nilsson@hh.se) received his M.S. and Ph.D. degrees in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 1996 and 2013, respectively. Currently, he is a professor in the School of Information Technology, Halmstad University, Sweden. He is also with Raytelligence AB, Halmstad, Sweden, where he provides microwave and radar expertise. Previously, he was a radio-frequency integrated circuit designer at ARM. His research interests include working on cohabitant radar technologies and vital sign sensors.

Mats Rydström (mats.rydstrom.external@veoneer.com) received his M.S. degree in computer science from Chalmers University of Technology, Gothenburg, Sweden, and his Ph.D. degree in electrical engineering from Chalmers University of Technology, in 2008, where he worked on wireless positioning systems and signal processing for wireless communications. Currently, he is with Veoneer Sweden AB, Vårgårda, where he works on technologies related to signal processing for future radar systems in automotive applications.

Karl Vanäs (karl.vanas@volvocars.com) received his M. Sc. degree in electrical engineering from the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. Currently, he works as a technical expert in radar systems at Volvo Car Corporation, Gothenburg. He has 20 years of experience working with radars, within both the automotive and the defense industries. During the last 10 years, he has worked on various radar projects at Volvo Cars.

Henk Wymeersch (henkw@chalmers.se) received his M.S. degree in computer science in 2001 and his Ph.D. degree in applied sciences/electrical engineering in 2005, both from Ghent University, Belgium. Currently, he is a professor of communication systems at Chalmers University of Technology, Gothenburg, Sweden. Previously, he was a postdoctoral researcher at the Massachusetts Institute of Technology, Cambridge, from 2005 to 2009. His research focuses on algorithm design and analysis for transportation systems.

References

[1] I. M. Kunert, "MOSARIM: More safety for all by radar interference mitigation," Project Final Rep., 2012. [Online]. Available: http://cordis.europa.eu/docs/projects/cnect/1/248231/080/deliverables/001-D611finalreportfinal.pdf

[2] H. Meinel and J. Dickmann, "Automotive radar: From its origin to future directions," *Microw. J.*, vol. 56, no. 9, pp. 24–40, 2013.

[3] W. Buller, B. Wilson, J. Garbarino, J. Kelly, N. Subotic, B. Thelen, and B. Belzowski, "Radar congestion study," National Highway Traffic Safety Administration, Washington, D.C., Tech. Rep. DOT HS 812 632, Sept. 2018.

[4] J. Moss and A. Lefevre, "Radar interference suppression," European Patent Appl. 3 489 710, May 29, 2019.

[5] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, 2017. doi: 10.1109/MSP.2016.2628914.

[6] M. A. Richards, *Fundamentals of Radar Signal Processing.* New York: McGraw-Hill, 2005.

[7] P. Stoica and R. Moses, *Spectral Analysis of Signals.* Englewood Cliffs, NJ: Prentice Hall, 2005.

[8] A. Al-Hourani, R. J. Evans, S. Kandeepan, B. Moran, and H. Eltom, "Stochastic geometry methods for modeling automotive radar interference," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 333–344, 2017. doi: 10.1109/TITS.2016.2632309.

[9] M. Goppelt, H.-L. Blöcher, and W. Menzel, "Automotive radar–investigation of mutual interference mechanisms," *Adv. Radio Sci.*, vol. 8, no. 3, pp. 55–60, 2010. doi: 10.5194/ars-8-55-2010.

[10] G. Kim, J. Mun, and J. Lee, "A peer-to-peer interference analysis for automotive chirp sequence radars," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8110–8117, Sept. 2018. doi: 10.1109/TVT.2018.2848898.

[11] M. Goppelt, H. Blöcher, and W. Menzel, "Analytical investigation of mutual interference between automotive FMCW radar sensors," in *Proc. German Microwave Conf.*, Mar. 2011, pp. 1–4.

[12] M. Gerstmair, A. Melzer, A. Onic, and M. Huemer, "On the safe road toward autonomous driving: Phase noise monitoring in radar sensors for functional safety compliance," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 60–70, Sept. 2019. doi: 10.1109/MSP.2019.2902914.

[13] K. Siddiq, M. K. Hobden, S. R. Pennock, and R. J. Watson, "Phase noise in FMCW radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 1, pp. 70–81, Feb. 2019. doi: 10.1109/TAES.2018.2847999.

[14] A. Chorti and M. Brookes, "A spectral model for RF oscillators with power-law phase noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 9, pp. 1989–1999, Sept. 2006. doi: 10.1109/TCSI.2006.881182.

[15] M. C. Budge and M. P. Burt, "Range correlation effects in radars," in *Rec. IEEE Nat. Radar Conf.*, Apr. 1993, pp. 212–216. doi: 10.1109/NRC.1993.270463.

[16] T. Schipper, M. Harter, T. Mahler, O. Kern, and T. Zwick, "Discussion of the operating range of frequency modulated radars in the presence of interference," *Int. J. Microw. Wireless Technol.*, vol. 6, nos. 3–4, pp. 371–378, 2014. doi: 10.1017/S1759078714000221.

[17] F. Uysal, "Synchronous and asynchronous radar interference mitigation," *IEEE Access*, vol. 7, pp. 5846–5852, Jan. 2019. doi: 10.1109/ACCESS.2018.2884637.

[18] J. Bechter, F. Roos, M. Rahman, and C. Waldschmidt, "Automotive radar interference mitigation using a sparse sampling approach," in *Proc. European Radar Conf. (EURAD)*, Oct. 2017, pp. 90–93. doi: 10.23919/EURAD.2017.8249154.

[19] S. Alland, W. Stark, M. Ali, and M. Hegde, "Interference in automotive radar systems: Characteristics, mitigation techniques, and current and future research," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 45–59, Sept. 2019. doi: 10.1109/MSP.2019.2908214.

[20] Z. Slavik and K. V. Mishra, "Cognitive interference mitigation in automotive radars," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6. doi: 10.1109/RADAR.2019.8835643.

[21] C. Aydogdu, M. F. Keskin, N. Garcia, H. Wymeersch, and D. W. Bliss, "RadChat: Spectrum sharing for automotive radar interference mitigation," *IEEE Trans. Intell. Transport. Syst.*, to be published. doi: 10.1109/TITS.2019.2959881.

[22] F. Liu, C. Masouros, A. Petropulu, H. Griffiths, and L. Hanzo, Joint radar and communication design: Applications, state-of-the-art, and the road ahead. June 2019. [Online]. Available: https://arxiv.org/abs/1906.00789

[23] J. Fink and F. K. Jondral, "Comparison of OFDM radar and chirp sequence radar," in *Proc. 2015 IEEE 16th Int. Radar Symp. (IRS)*, pp. 315–320. doi: 10.1109/IRS.2015.7226369.

[24] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using OFDM waveforms," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 226–238, 2010. doi: 10.1109/JSTSP.2009.2038977.

[25] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, July 2011. doi: 10.1109/JPROC.2011.2131110.

[26] M. Braun, "OFDM radar algorithms in mobile communication networks," Karlsruher Inst.s für Technologie, Karlsruhe, Germany, 2014. [Online]. Available: https://publikationen.bibliothek.kit.edu/1000038892

[27] G. Hakobyan and B. Yang, "High-performance automotive radar: A review of signal processing algorithms and modulation schemes," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 32–44, Sept. 2019. doi: 10.1109/MSP.2019.2911722.

[28] K. V. Mishra, M. R. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 100–114, Sept. 2019. doi: 10.1109/MSP.2019.2913173.

[29] G. Lellouch, A. K. Mishra, and M. Inggs, "Stepped OFDM radar technique to resolve range and doppler simultaneously," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 2, pp. 937–950, Apr. 2015. doi: 10.1109/TAES.2014.130753.

[30] B. Schweizer, C. Knill, D. Schindler, and C. Waldschmidt, "Stepped-carrier OFDM-radar processing scheme to retrieve high-resolution range-velocity profile at low sampling rate," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1610–1618, Mar. 2018. doi: 10.1109/TMTT.2017.2751463.

Dingyou Ma, Nir Shlezinger, Tianyao Huang,
Yimin Liu, and Yonina C. Eldar

Joint Radar-Communications Strategies for Autonomous Vehicles

Combining two key automotive technologies

©ISTOCKPHOTO.COM/OONAL

Self-driving cars constantly assess their environment to choose routes, comply with traffic regulations, and avoid hazards. To that aim, such vehicles are equipped with wireless communications transceivers as well as multiple sensors, including automotive radars. The fact that autonomous vehicles implement both radar and communications motivates designing these functionalities in a joint manner. Such dual-function radar-communications (DFRC) designs are the focus of a large body of recent work. These approaches can lead to substantial gains in size, cost, power consumption, robustness, and performance, especially when both radar and communications operate in the same range, which is the case in vehicular applications.

This article surveys the broad range of DFRC strategies and their relevance to autonomous vehicles. We identify the unique characteristics of automotive radar technologies and their combination with wireless communications requirements of self-driving cars. Then, we map the existing DFRC methods along with their pros and cons in the context of autonomous vehicles and discuss the main challenges and possible research directions for realizing their full potential.

Sensing and communication in autonomous vehicles

Autonomous vehicles are required to navigate efficiently and safely in a wide variety of complex uncontrolled environments. To meet these requirements, such self-driving cars must be able to reliably sense and interact with their surroundings. This acquired sensory information as well as data communicated from neighboring vehicles and roadside units are essential to avoid obstacles, select routes, detect hazards, and comply with traffic regulations, all in real time.

To reliably sense the environment, autonomous vehicles are equipped with multiple sensing technologies, including computer vision acquisition, i.e., cameras, lidar, laser-based sensors, GPS, and radar transceivers. Each of these technologies has its advantages and disadvantages. To allow accurate sensing in a broad range of complex environments, self-driving cars should simultaneously utilize all of these

Digital Object Identifier 10.1109/MSP.2020.2983832
Date of current version: 26 June 2020

aforementioned sensors. Radar, for instance, provides the ability to accurately detect distant objects and is typically more robust to weather conditions and poor visibility compared to other competing sensing technologies [1].

Radar systems, which detect the presence of distant objects by measuring the reflections of electromagnetic probing waves, have been in use for over a century. Radar has been most commonly used in military applications, aircraft surveillance, and navigation systems. The application of radar for vehicles, referred to as *automotive radar* [2], is substantially different from traditional radar systems: most notably, automotive radar systems, which are used by mass-produced vehicles, are far more limited in size, power, and cost. Furthermore, while conventional radar aims to detect a relatively small number of distant targets, e.g., airplanes, automotive radar is required to sense in complex dense urban environments in which a multitude of scatterers at close ranges should be accurately detected. Despite these differences, today automotive radar is an established and common technology, and the vast majority of newly manufactured vehicles are equipped with radar-based autonomous driving assistance systems (ADASs) [1].

In addition to their ability to sense their environment, autonomous vehicles are also required to carry out various forms of communications, as illustrated in Figure 1: vehicle-to-vehicle (V2V) transmissions allow self-driving cars to share their attributes with neighboring vehicles; vehicle-to-infrastructure (V2I) messages facilitate intelligent road management by conveying information between cars and roadside units; vehicle-to-pedestrian (V2P) communications can be used to warn or alarm nearby pedestrians; and, finally, service providers and cloud applications exchange possibly large amounts of data with self-driving cars via vehicle-to-network (V2N) and vehicle-to-cloud (V2C) links, respectively. The resulting broad range of different tasks, which substantially vary in their latency, throughput, and reliability requirements, can be implemented by using individual communications technologies for each application or by using a unified vehicle-to-everything (V2X) strategy [3], possibly building upon the cellular infrastructure.

Automated cars, thus, implement two technologies that rely on the transmission and processing of electromagnetic signals: radar and wireless communications. A possible approach in designing self-driving cars is to use individual systems for radar and communications, each operating separately. An alternative strategy is to jointly design these functionalities as a DFRC system. Such schemes are the focus of extensive recent research attention [4]–[20]. In particular, it was shown that jointly implementing radar and communications contributes to reducing the number of antennas [21], system size, weight, and power consumption [6] as well as alleviating concerns for electromagnetic compatibility and spectrum congestion [5]. Utilizing such joint designs in vehicular systems can mitigate the mutual interference among neighboring cars, facilitate coordination, and improve pedestrian detection [22]. These benefits make DFRC systems an attractive technology for autonomous vehicles.

While the conceptual advantages of joint radar-communications designs for autonomous vehicles are clear, the proliferation of different DFRC strategies makes it difficult to identify what scheme is most suitable for which scenario. For example, some DFRC methods use existing V2X communications waveforms as radar probing signals, thus allowing high communication throughput with relatively limited sensing capabilities [17], [18]. Alternative schemes embed digital messages in the radar probing signals [15], [16], thus supporting low data rates, which may be more suitable to serve as an additional channel to the standard communications functionalities of autonomous vehicles.

The goal of this article is to review DFRC technologies in light of the unique requirements and constraints of self-driving cars, facilitating the identification of the proper technology for different scenarios. We begin by reviewing the basics of automotive radar, identifying its main challenges, recent advances, and fundamental differences from conventional radar systems. We then survey DFRC methods, dividing previously proposed approaches into four main categories: coordinated signals transmission methods utilizing individual signals for each functionality; communications waveform-based schemes, which use the communications signal

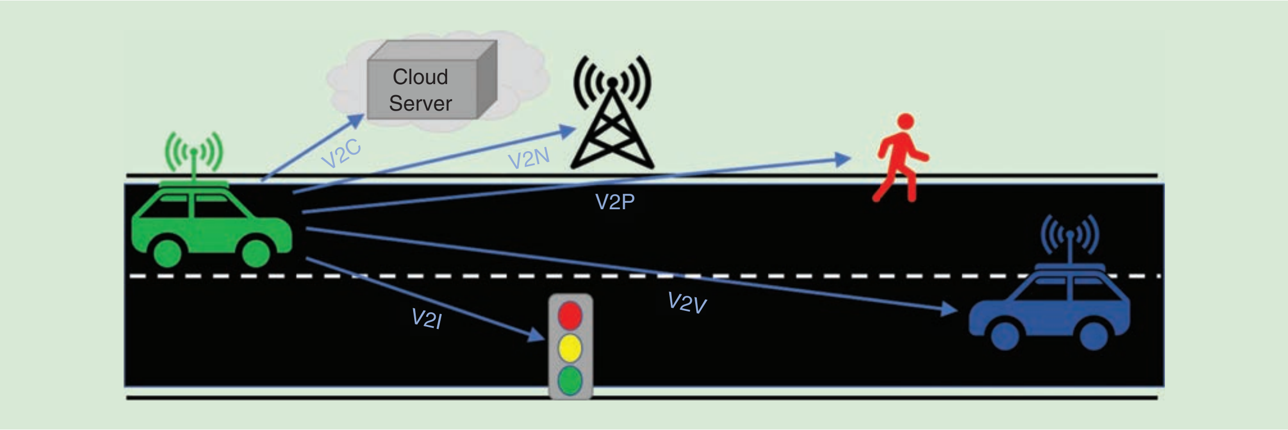


FIGURE 1. The autonomous vehicle communications links.

as a radar probing waveform; radar waveform-based techniques, which embed the digital message into the parameters of the radar signal; and the design of dedicated dual-function waveforms. We detail a representative set of DFRC methods for each category and provide a map of the existing strategies in terms of their radar capabilities, information rates, and complexity.

Basics of automotive radar

Past decades have witnessed growing interest in automotive radar to improve the safety and comfort of drivers. A typical ADAS implements various radar subsystems that enable functions including adaptive cruise control, blind-spot detection, and parking assistance [1]. To understand the benefits of combining automotive radar with digital communications, we first review the basics of automotive radar.

Automotive radars operate under different requirements and constraints compared to conventional radars, such as those utilized in military applications and air traffic control. First, conventional radar systems are required to detect a relatively small number of targets in ranges on the order of tens or hundreds of kilometers, while automotive radars must detect a multitude of objects in short ranges on the order of a few tens of meters. Furthermore, automotive radars are incorporated into mass-produced vehicles and, hence, have more strict constraints on cost, size, power consumption, and spectral efficiency compared to conventional radar. Finally, automotive radars are densely deployed in urban environments; thus, they must be robust to interference while inducing minimal interference with neighboring radar systems.

Various techniques have been proposed to overcome the aforementioned challenges. In Table 1, we summarize the main challenges along with the leading methods to tackle them. It is noted that no single radar scheme is suitable to handle the complete set of requirements. For example, the popular frequency-modulated continuous-wave (FMCW) waveform (see "Frequency-Modulated Continuous-Wave Radar"), which can be operated using simplified hardware components, suffers from high sensitivity to interference;

Table 1. Automotive radar requirements.

Requirements	Possible Solutions
Operating in short ranges	Utilize separate transmit and receive antennas to process short-range echoes.
Limited antenna size	Operate at mm-wave bands using patch antennas.
	Increase virtual aperture (see "Multiple-Input, Multiple-Output Radar").
Simplified hardware	Constant envelope signaling.
	Low-complexity dechirp recovery, e.g., FMCW (see "Frequency-Modulated Continuous-Wave Radar").
Low-power amplifiers	Continuous or high duty cycle waveform, e.g., FMCW.
Interference robustness	Divide spectrum using OFDM (see "Orthogonal Frequency-Division Multiplexing Waveform Radar").
	Introduce agility (see "Frequency Agile Radar") to increase survivability.

mm-wave: millimeter-wave.

Frequency-Modulated Continuous-Wave Radar

Frequency-modulated continuous-wave (FMCW) radar is a continuous constant modulus radar waveform with a linearly modulated frequency, which can be generated and detected using simplified hardware. To present FMCW, we consider a radar system equipped with a single transmit antenna and a uniform linear array with L_R elements for receiving. In each radar coherent processing interval, M FMCW pulses of duration T_p are periodically transmitted with a pulse repetition interval (PRI) denoted by T_{PRI}, where T_{PRI} is slightly larger than T_p. The mth pulse is given by $s_m(t) = e^{j2\pi f_c t + j\pi\gamma t^2}$, $t \in [mT_{\text{PRI}}, mT_{\text{PRI}} + T_p]$, where f_c, is the carrier frequency, and γ is the frequency modulation rate.

To formulate the received signal, assume P targets are located in the far field. The distance, velocity, and angle of the pth target are denoted as r_p, v_p, and θ_p, respectively. For the pth target, with the far-field assumption, the round time delay between the transmit antenna and the lth receiver is $\tau_{l,p} = (2(r_p + v_p t) - ld \sin\theta_p)/c$, where d is the distance between adjacent elements in the receiving array, and c is the speed of light. The radar echo received in the lth receiving antenna during the mth transmit pulse is represented as $r_{m,l}(t) = \sum_{p=0}^{P-1} \alpha_p s_m(t - \tau_{l,p}) + w_l(t)$, where α_p is the complex reflective factor of the pth target, and $w_l(t)$ is additive white Gaussian noise.

To process the received signal, $r_{m,l}(t)$ is mixed with the transmit signal. This procedure, referred to as *dechirp*, yields a demodulated signal given by $y_{m,l}(t) = r_{m,l}(t) \cdot s_m^*(t)$, i.e.,

$$y_{m,l}(t) = \sum_{p=1}^{P} \tilde{\alpha}_p e^{-j2\pi\left(\frac{2\gamma r_p}{c} + \frac{2v_p f_c}{c}\right)(t - mT_{\text{PRI}}) - j2\pi f_c \frac{2nvT_{\text{PRI}}}{c} + j2\pi \frac{ldf_c \sin\theta_p}{c}} + \tilde{w}_{m,l}(t), \quad \text{(S1)}$$

where $(\cdot)^*$ is the complex conjugate, $\tilde{\alpha}_p := \alpha_p e^{-j2\pi f_c \frac{2r_p}{c}}$, and $\tilde{w}_{m,l}(t) := w_l(t) \cdot s_m^*(t)$.

After dechirp, the waveform frequency is typically much smaller compared to the bandwidth of the transmitted waveform, and it can be sampled with low-speed analog-to-digital converters. It follows from (S1) that the targets' range, velocity, and direction can then be recovered from the 3D discrete Fourier transform of the sampled $y_{m,l}$ in the fast time domain (within a pulse), slow time domain (between pulses), and spatial domain (over antennas), respectively.

orthogonal frequency-division multiplexing (OFDM) radar (described in "Orthogonal Frequency-Division Multiplexing Waveform Radar"), which is suitable for multiuser scenarios, tends to require relatively costly hardware compared to alternative radars. An additional aspect that should be considered in selecting an automotive radar scheme is its capability to be combined with wireless communications. The fact that self-driving cars utilize both radar and digital communications motivates their joint design as a DFRC system, as discussed in the following section.

Overview of dual-function systems

Since DFRC systems implement both radar and communications using a single device, these functionalities inherently share some of the system resources, such as spectrum, antennas, and power. Broadly speaking, existing DFRC methods can be divided into four main categories as illustrated in Figure 2: coordinated separated signals transmission, communications waveform-based approaches, radar waveform-based schemes, and joint dual-function waveform designs. In the following, we review each of these categories and discuss their pros and cons in the context of autonomous vehicles. Throughout this section, we consider a DFRC system jointly implementing a radar transceiver as well as the transmission of digital messages using L_T transmit antennas (for both radar and communications) and L_R receive antennas (for radar). For simplicity, we assume a single communications receiver equipped with a single antenna.

Separate coordinated signals

A common DFRC approach is to utilize different signals for radar and communications, designing the functionalities to mitigate their cross interference, as illustrated in Figure 2(a). Here, the $L_T \times 1$ transmitted signal can be written as

$$\boldsymbol{s}(t) = \boldsymbol{s}^{(r)}(t) + \boldsymbol{s}^{(c)}(t), \tag{1}$$

where $\boldsymbol{s}^{(r)}(t)$ is the radar probing waveform, and $\boldsymbol{s}^{(c)}(t)$ is the continuous-time communications signal. The ability to jointly transmit two dedicated signals with limited cross interference is typically achieved using either orthogonality boosting by division in time and/or frequency or via spatial beamforming.

Time–frequency division

Arguably, the most simple method to mitigate cross interference is to allocate a different frequency band to each waveform, commonly dictated by regulated spectrum allocation, or, alternatively, a different time slot. In such cases, the signals $\boldsymbol{s}^{(r)}(t)$ and $\boldsymbol{s}^{(c)}(t)$ in (1) either reside in different bands (for frequency division) or satisfy $\boldsymbol{s}^{(r)}(t)(\boldsymbol{s}^{(c)}(t))^T = \mathbf{0}$ at each time instance (for time division). Since system resources are allocated between both subsystems, these strategies

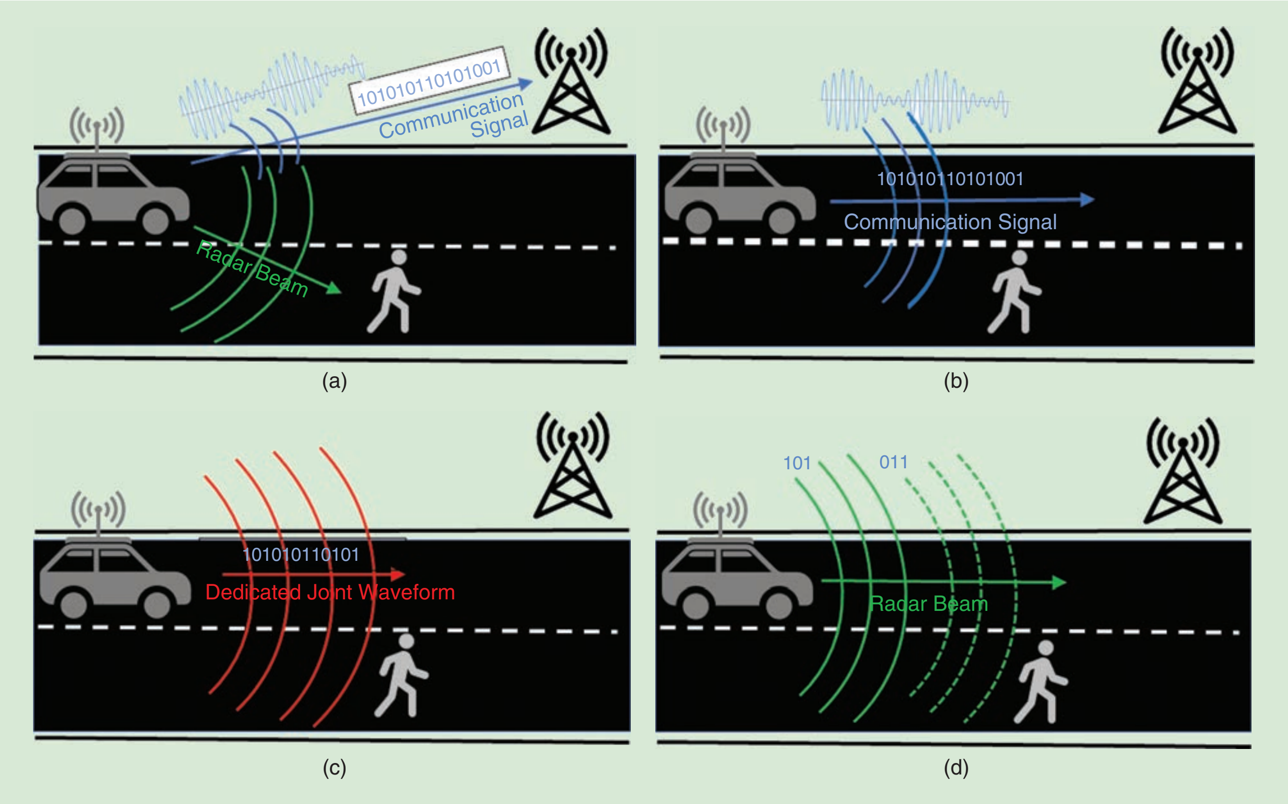


FIGURE 2. An illustration of DFRC strategies for autonomous vehicles: (a) coordinated signals, (b) communications waveform based, (c) dual-function waveform, and (d) radar waveform based. The blue, green, and red waveforms represent communications signals, radar beams, and dedicated dual-function waveforms, respectively.

inevitably result in a tradeoff between radar and communications performance [19].

A straightforward approach is to allocate the resources in a fixed or arbitrary manner. For instance, in [21], a DFRC system is achieved by using fixed nonoverlapping bands and antennas. A random antenna allocation scheme is proposed in [14], jointly enhancing the radar angular resolution and the communication rates. The work [20] proposed a media access protocol for automotive DFRC systems with time and frequency division to mitigate interference with neighboring radars. These approaches assume that each functionality has its own frequency band. Using OFDM signaling, i.e., letting the entries of $\boldsymbol{s}^{(r)}(t)$ and $\boldsymbol{s}^{(c)}(t)$ represent OFDM radar and communications waveforms (see "Orthogonal Frequency-Division Multiplexing Waveform Radar"), respectively, allows the division of the spectrum in an optimized manner, as we detail next.

Consider a frequency band divided into N subbands. The discrete-time transmitted signal from the lth transmit antenna can be written as the $N \times 1$ vector $\boldsymbol{s}_l$. Since the spectrum is divided into radar and communications, $\boldsymbol{s}_l$ is given by

$$\boldsymbol{s}_l = \boldsymbol{F}^H[\boldsymbol{U}_l\boldsymbol{s}_l^{(r)} + (\boldsymbol{I} - \boldsymbol{U}_l)\boldsymbol{s}_l^{(c)}], \tag{2}$$

where $\boldsymbol{F}^H$ is the inverse discrete Fourier transform (DFT) matrix; the $N \times 1$ vectors $\boldsymbol{s}_l^{(r)}$ and $\boldsymbol{s}_l^{(c)}$ denote the OFDM radar and communications symbols, respectively, in the frequency domain; and $\boldsymbol{U}_l$ is a diagonal matrix of size $N \times N$ with elements zero or one, representing the subcarrier selection at the lth element.

Setting the matrix $\boldsymbol{U}_l$ in (2) determines how the bandwidth is divided. The work [23] showed that when $\boldsymbol{U}_l$ represents spectral interleaving, i.e., the support of its diagonal consists of multiple bulks of zeros and ones, radar resolution is comparable to that using the complete spectrum. When the DFRC system has a priori knowledge of the statistical model of the radar target response and the communications channel, the subcarrier selection matrix $\boldsymbol{U}_l$ can be set to optimize a linear combination of the radar target-echo mutual information and the communications input–output mutual information, as proposed in [13].

Spatial beamforming

The utilization of multiple antennas enables the mitigation of mutual interference through spatial beamforming, for example, by projecting the radar waveform into the null space of its channel to the communications receiver [24], resulting in a zero forcing beamformer. While such beamforming was originally proposed for separate systems, it can also be utilized for a DFRC system.

In this model, the communications and radar signals are beamformed using the matrices $\boldsymbol{U}^{(c)}$ and $\boldsymbol{U}^{(r)}$, respectively, to mitigate the mutual interference while satisfying the performance constraints. The signals received at the communications receiver and the radar target with direction θ are, thus,

$$\begin{aligned} y^{(c)} &= \boldsymbol{h}^T(\boldsymbol{U}^{(r)}\boldsymbol{s}^{(c)} + \boldsymbol{U}^{(r)}\boldsymbol{s}^{(r)}) + w^{(c)}, \text{ and} \\ y_\theta^{(r)} &= \boldsymbol{a}^T(\theta)(\boldsymbol{U}^{(c)}\boldsymbol{s}^{(c)} + \boldsymbol{U}^{(r)}\boldsymbol{s}^{(r)}), \end{aligned} \tag{3}$$

where $\boldsymbol{h}$ is the channel response from the DFRC transmitter to the communications receiver, and $\boldsymbol{a}(\theta)$ is the steering vector of the DFRC transmitter to the radar target in direction θ. Using (3), the beamforming matrices $\boldsymbol{U}^{(c)}$ and $\boldsymbol{U}^{(r)}$ are jointly designed to mitigate cross interference while satisfying the performance requirements, e.g., maximizing the signal-to-interference-plus-noise ratio (SINR) at the communications receiver while meeting a given radar beampattern [25].

A clear advantage of the separated signals transmission strategy is that it can provide a wide variety of possible performance combinations. For time–frequency division schemes, the performance is determined by how the system resources, such as spectrum and time slots, are allocated to each functionality. The performance tradeoffs may be potentially improved using spatial beamforming, allowing each functionality to utilize the full bandwidth and operate simultaneously at all time slots. However, the spatial beamformer is designed based on a priori channel knowledge, which may be unavailable for fast-moving vehicles. According to the previously given discussion, time–frequency division-based schemes are likely to be more attractive in automotive applications. Since properly optimizing the resource allocation to achieve a desired performance tradeoff requires considerable computation, fixed suboptimal allocations, such as spectral interleaving, may be preferable.

Communications waveform-based schemes

Another common DFRC strategy is to utilize standard communications signals for probing, as illustrated in Figure 2(b). The majority of communications waveform-based designs in the literature utilize OFDM signaling, especially for automotive applications. In the sequel, we first briefly review spread spectrum-based DFRC systems, followed by a more detailed presentation of shared OFDM waveforms and a description of how structured vehicular communications protocols can be used for sensing.

Spread spectrum waveforms

Spread spectrum techniques transmit a communications signal with a given bandwidth over a much larger spectral band, typically using spread coding or frequency hopping. The usage of spread spectrum signals for radar probing was studied in [9]. The main drawback of spread spectrum DFRC design is that the radar dynamic range is limited, which is a by-product of the imperfect autocorrelation properties of the spreading sequences [9]. In addition, accurately recovering the target velocity from spread spectrum echoes is typically computationally complex, limiting the applicability of such DFRC systems. Finally, high-speed analog-to-digital converters (ADCs) are required for wideband spectrum spread waveforms, as dechirp used in FMCW is not applicable, increasing cost and complexity.

OFDM waveforms

The most common communications waveform-based approach is to utilize OFDM signaling. OFDM is a popular digital communications scheme due to its spectral efficiency, inherent ability to handle intersymbol interference, and the fact that it can be implemented using relatively simple hardware components. Since first proposed in [29], OFDM has received extensive attention as a radar waveform, especially for automotive radar, due to its high flexibility and adaptability in transmission and since, unlike FMCW, it does not suffer from range–Doppler coupling [30]. The fact that OFDM is commonly utilized in both radar and communications indicates its potential for DFRC systems.

Compared with the case where the coefficients $\{a_{m,n}\}$ in the OFDM waveform are specifically designed for radar (see "Orthogonal Frequency-Division Multiplexing Waveform Radar"), the complex weights of the dual-function OFDM waveform are the communications symbols. The setting of the waveform parameters can have a notable effect on each functionality. The work [31] designed the subcarrier spacing according to the maximum unambiguous range and the maximum velocity. In [6], channel knowledge was used to allocate power between the subcarriers to maximize the sum of the data rate and the mutual information between the received echoes and the target impulse response. Radar processing of OFDM waveforms utilizes matched filtering, which depends on the transmitted data, causing high-level sidelobes. This data dependency can be eliminated by dividing each subcarrier by its corresponding symbol [9]. The range and velocity of each target are then estimated using a 2D DFT in the carrier domain and slow time domain (between different symbols), respectively.

OFDM can be naturally combined with multiple-input, multiple-output (MIMO) radar, which transmits orthogonal waveforms from each antenna (see "Multiple-Input, Multiple-Output Radar") by assigning different subcarriers to different transmit elements. Several works have studied how to divide the subcarriers among the elements. The proposed methods include division by equidistant subcarrier interleaving [23], nonequidistant subcarrier interleaving [32], and random assignments [33].

A drawback of using shared OFDM waveforms in vehicular systems stems from the fact that, when utilized from moving vehicles, OFDM exhibits subcarrier misalignment, degrading the maximal radar unambiguous range [30]. Additional drawbacks are related to hardware constraints: wideband OFDM waveforms require high-rate ADCs, affecting the system cost and power consumption. Another hardware limitation of OFDM compared to monotone waveforms is its high peak-to-average-power ratio, inducing distortion in the presence of practical nonlinear amplifiers. A weighted OFDM method was proposed to control the maximum peak-to-average power ratio [26], [27]. To utilize OFDM with narrowband transmissions, one can apply stepped frequency methods, as proposed in [33].

Protocol-oriented DFRC methods

An alternative strategy is to exploit the existing communications protocols, utilizing them as an automotive radar waveform. Here, there is no compromise in the communications part, and the radar functionality is a byproduct of the protocol, which is typically IEEE 802.11p or IEEE 802.11ad [17], [18], [35], [36]. The IEEE 802.11p standard focuses on vehicular communications and supports short-range device-to-device transmissions for safety applications. This protocol operates in the 5.9-GHz band and uses OFDM signaling. Consequently, its transmissions realize a DFRC system with an OFDM shared waveform, as proposed in [17].

IEEE 802.11ad is a generic standard for short-range millimeter-wave (mm-wave) communications operating at 60 GHz. Its large bandwidth enables higher data rates for communications and better accuracy/resolution for radar operation. To avoid the usage of data-dependent waveforms, it has been proposed that the a priori known IEEE 802.11ad preamble be used for radar probing [18]. As the preamble now affects radar performance, the work [36] studied the design of radar-suitable preamble sequences. In such mm-wave communications, highly

Orthogonal Frequency-Division Multiplexing Waveform Radar

For an orthogonal frequency-division multiplexing (OFDM) waveform radar with N subcarriers, the transmit signal at the mth pulse is

$$s_m(t) = \sum_{n=1}^{N} a_{m,n} \operatorname{rect}\left(\frac{t - mT_O}{T_O}\right) e^{j2\pi(f_c+f_n)t}. \tag{S2}$$

In (S2), $\{a_{m,n}\} \in \mathcal{A}$ are complex weights transmitted over the mth symbol on carrier f_n, which can be either fixed or randomized from some discrete set $\mathcal{A}$; $\operatorname{rect}(t)$ is a rectangular window of unity support; $T_O = T_S + T_{CP}$ is the OFDM symbol duration; T_S is the elementary symbol duration; T_{CP} is the duration of the cyclic prefix; and $f_n = (n-1)(1/T_S)$. Using the notations in (S1), the radar echo from P targets observed by the lth receive antenna can be approximated by

$$y_{m,l}(t) \approx \sum_{p=1}^{P} \sum_{n=1}^{N} \tilde{\alpha}_p \cdot a_{m,n} \operatorname{rect}\left(\frac{t - mT_O - \frac{2r_p}{c}}{T_O}\right) \times e^{j2\pi f_n\left(t - \frac{2r_p}{c}\right) - j2\pi f_c \frac{2v_p - ld\sin\theta_p}{c} t} + w_l(t). \tag{S3}$$

OFDM radar processing is based on matched filtering [S1]. Its performance is determined by the complex weights, which can be optimally designed according to some requirements, e.g., the maximum peak-to-average ratio of the transmit signal [26] or the Cramér–Rao bound [27].

Reference

[S1] D. Garmatyuk, J. Schuerger, and K. Kauffman, "Multifunctional software-defined radar sensor and data communication system," *IEEE Sensors J.*, vol. 11, no. 1, pp. 99–106, 2010. doi: 10.1109/JSEN.2010.2052100.

Multiple-Input, Multiple-Output Radar

Multiple-input, multiple-output (MIMO) radar uses multiple transmit and receive antennas. By transmitting orthogonal waveforms from each antenna, one can generate a virtual array with larger aperture, increasing the angular resolution without requiring additional hardware elements. While MIMO radar can also be combined with nonorthogonal waveforms, we focus on such systems transmitting orthogonal waveforms, which is the common practice in MIMO radar [34].

To formulate MIMO radar transmission, let L_T and L_R be the numbers of transmit and receive antenna elements, respectively. The adjacent distances of the transmit antenna and the receive antenna are d_T and d_R, respectively. A common practice is to set $d_T = L_R d_R$. We use $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_{L_T}(t)]^T$ for the transmit waveforms, which are orthogonal, namely, $\int s_l(t)\, s_{l'}^*(t)\, dt = \delta(l - l')$, where $\delta(\cdot)$ is the Kronecker delta. For simplicity, we consider a single pulse and targets associated with a particular range and Doppler bin. The received signal is

$$\boldsymbol{y}(t) = \sum_{p=1}^{P} \alpha_p \boldsymbol{a}^T(\theta_p)\, \boldsymbol{s}(t)\, \boldsymbol{b}(\theta_p) + \boldsymbol{w}(t), \qquad \text{(S4)}$$

where $\boldsymbol{a}(\theta) := [1, e^{j2\pi f_c d_T \sin\theta/c}, \ldots, e^{j2\pi f_c (L_T - 1) d_T \sin\theta/c}]^T$ is the transmit steering vector, $\boldsymbol{b}(\theta) := [1, e^{j2\pi f_c d_R \sin\theta/c}, \ldots, e^{j2\pi f_c (L_R - 1) d_R \sin\theta/c}]^T$ is the receive steering vector in direction θ, and $\boldsymbol{w}(t)$ is white Gaussian noise. Applying matched filtering and vectorization yields

$$\tilde{\boldsymbol{y}} = \mathrm{vec}\Big(\int \boldsymbol{y}(t)\, \boldsymbol{s}^H(t)\, dt\Big) = \sum_{p=1}^{P} \alpha_p \boldsymbol{a}(\theta_p) \otimes \boldsymbol{b}(\theta_p) + \tilde{\boldsymbol{w}}, \qquad \text{(S5)}$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator, $\tilde{\boldsymbol{w}} := \mathrm{vec}(\int \boldsymbol{w}(t)\, \boldsymbol{s}^H(t))$, and $\otimes$ is the Kronecker product. Since $\boldsymbol{a}(\theta_p) \otimes \boldsymbol{b}(\theta_p) = [1, e^{j2\pi f_c d_R \sin\theta_p/c}, \ldots, e^{j2\pi f_c (L_T L_R - 1) d_R \sin\theta_p/c}]^T$ it holds that MIMO radar achieves an equivalent angle resolution of a phased array radar with $L_T L_R$ receive antennas in this configuration, effectively enhancing the angular resolution by a factor of L_T.

directional beams are used. Once the communications data link is established, radar can only reliably detect targets located in the assigned beam direction. Several approaches have been proposed to extend the scanning area at the cost of power reduction in [36].

The main benefit of protocol-based DFRC designs is that they implement radar with minimal effect on the communications functionality. As such, their radar capabilities are quite limited. The radar coverage area is restricted by the directionally beamformed mm-wave transmission. In addition to its restricted coverage area, the scheme has a relatively low radar duty cycle as only the preamble is utilized for probing, limiting its detection range in vehicular systems operating under peak power constraints.

To conclude, communications waveform-based DFRC approaches, and particularly those using shared OFDM signaling, enable the transmission of high data rates by utilizing conventional digital communications schemes. The fact that OFDM is widely studied for both radar and communications makes it an attractive DFRC design. In the context of autonomous vehicles, several drawbacks must be accounted for: first, to radiate enough power on the target, radar waveforms are typically beamformed to be directional. The communications receivers should, thus, be located in the radar beam to observe high signal-to-noise ratios. Such transmissions may, thus, be more suitable to serve as a secondary communications channel in addition to a possible cellular-based V2X technology, which can communicate with the receivers in the omnidirection. Similarly, protocol-oriented schemes, which utilize standard communications transmission while exploiting its structure for probing, are more likely to provide additional sensing capabilities to a dedicated automotive radar. Finally, relatively costly hardware components are required for generating wideband waveforms and sampling their reflections. Despite these drawbacks, sensing using communications waveforms is considered to be a promising DFRC approach for autonomous vehicles [1].

Radar waveform-based techniques

DFRC systems can also be designed by embedding the communication message in conventional radar waveforms, as illustrated in Figure 2(d). These techniques are divided into two categories: the first approach modifies the radar waveform to incorporate digital modulations; the second method utilizes index modulation (IM), conveying data bits via the indices of certain radar parameters.

Modified radar waveforms

A possible approach to embed digital communications into an existing radar system is to modify the waveform to include modulated symbols. For example, the traditional FMCW (see "Frequency-Modulated Continuous-Wave Radar") can be modified to include phase-modulated symbols by replacing the mth pulse $s_m(t)$, defined in "Frequency-Modulated Continuous-Wave Radar," with $s_m(t) e^{j\phi_m}$, where ϕ_m encapsulates the information message in the form of, e.g., continuous phase modulation as proposed in [37]. Alternatively, the linear frequency of the pulse can convey information via frequency modulation [39], for example, by using a positive frequency modulation rate γ to transmit the bit one and a negative value for zero. While these schemes are typically power efficient [9] and have low complexity, their communication rate is very limited.

Higher communication rates can be obtained by utilizing multiple orthogonal waveforms and beamforming. Assume J

Frequency Agile Radar

A promising approach to tackle mutual interference between radars is to utilize frequency agile radar (FAR) [S2]. Here, a subband waveform (of a much narrower bandwidth compared to the available band) is transmitted in each cycle, and its central frequency varies randomly from cycle to cycle. These random variations reduce the spectral collision probability from neighboring radars.

To formulate the signal model, we use $\mathcal{F}=\{f_c+(n-1)\Delta f \mid n=1,2,\ldots,N\}$ to denote the carrier frequency set, where Δf is the carrier spacing. During the mth transmit pulse, the transmitted signal is $s_m(t)=e^{j2\pi f_m t}$, where f_m is randomly chosen from $\mathcal{F}$. After demodulation, the signal observed at the lth receive antenna can be expressed using the notations of (S1) as

$$y_{m,l}(t)=\sum_{p=1}^{P}\alpha_p e^{-j2\pi f_m\frac{2r_p}{c}-j2\pi f_m\frac{2v_p(m-1)T_{\mathrm{PRI}}}{c}+j2\pi f_m\frac{ld\sin\theta_p}{c}}+w_l(t). \tag{S6}$$

Using matched filtering, FAR can synthesize a large bandwidth and enables the generation of high range resolution profiles. However, the random changing of carrier frequency leads to a high sidelobe level, which affects the detection of weak targets. To mitigate the sidelobe problem, compressed sensing methods can be applied for range–Doppler processing [S3], while recovery guarantees for such methods are provided in [S4] under sparse and block-sparse target scenes, respectively.

References

[S2] S. R. J. Axelsson, "Analysis of random step frequency radar and comparison with experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 890–904, Apr. 2007. doi: 10.1109/TGRS.2006.888865.

[S3] T. Huang, Y. Liu, H. Meng, and X. Wang, "Cognitive random stepped frequency radar with sparse recovery," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 50, no. 2, pp. 858–870, Apr. 2014. doi: 10.1109/TAES.2013.120443.

[S4] T. Huang, Y. Liu, X. Xu, Y. C. Eldar, and X. Wang, "Analysis of frequency agile radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6228–6240, Dec. 2018. doi: 10.1109/TSP.2018.2876301

---

orthogonal waveforms $\{s_j(t)\}_{j=1}^{J}$ are simultaneously transmitted from an antenna array, and let $\{\boldsymbol{u}_j\}_{j=1}^{J}$ be the corresponding beamforming vectors. The transmit signal is expressed as $\boldsymbol{s}(t)=\sum_{j=1}^{J}\boldsymbol{u}_j s_j(t)$. In the communications receiver, the received signal is $y^{(c)}(t)=\boldsymbol{g}_c^T\boldsymbol{s}(t)+w^{(c)}(t)$, where $\boldsymbol{g}_c$ and $w^{(c)}(t)$ are the channel response and additive noise, respectively. By applying matched filtering with the orthogonal waveforms, the receiver obtains the vector $\boldsymbol{y}^{(c)}=[y_1^{(c)},y_2^{(c)},\ldots,y_J^{(c)}]^T$, where $y_j^{(c)}=\boldsymbol{g}_c^T\boldsymbol{u}_j+w_j^{(c)}(t)$. The communication data bits can be conveyed by modulating the amplitude [7] or phase [8] of $\boldsymbol{g}_c^T\boldsymbol{u}_j$. Although the communication rates are improved by transmitting multiple waveforms, the system complexity is also increased, and the transmitter must have a priori knowledge of the communications channel $\boldsymbol{g}_c$. Furthermore, it is difficult to guarantee that the envelope of the transmit signal is constant modulus, which may reduce power efficiency in transmission.

IM-based techniques

IM is a promising communications technique, gaining growing interest due to its high energy and spectral efficiency [40]. Instead of using conventional modulations, IM embeds data bits into the indices of certain transmission building blocks [40]. These building blocks, including spatial allocation and frequency division, are also important waveform parameters for radar. IM-based DFRC techniques, thus, embed the digital message into the combination of radar waveform parameters. The term *index* represents the radar parameters, such as carrier frequency, time slot, antenna allocation, or orthogonal waveforms in MIMO radar with orthogonal waveforms. Consequently, such DFRC systems use unmodified conventional radar schemes, and the ability to communicate is encapsulated in the parameters of the transmission. While IM-based DFRC schemes are the focus of ongoing research, existing methods typically build upon either MIMO radar or frequency agile radar (FAR) schemes. While MIMO radar can, in general, utilize orthogonal or nonorthogonal waveforms, we henceforth use the term *MIMO radar* for such schemes utilizing orthogonal waveforms, which is the typical approach in MIMO radar [35].

IM for MIMO radar

The work [40] proposed to combine MIMO radar with IM by embedding the bits in the assignment of the orthogonal waveforms across the transmit antennas. For a MIMO radar with L_T transmitting antennas, there are $L_T!$ possible arrangements in each pulse repetition interval (PRI), supporting a maximal rate of $\log L_T!$ bits per PRI. In [41], this approach was extended to sparse array MIMO radar configurations, where only K out of L_T transmit elements are active in each PRI. As a consequence, it requires only K transmit orthogonal waveforms, represented (with a slight notation abuse) by the vector $\boldsymbol{s}(t)=[s_1(t),s_2(t),\ldots,s_K(t)]^T$. The transmitted $L_T\times 1$ vector $\tilde{\boldsymbol{s}}(m,t)$ in the mth PRI is a permutation of $\boldsymbol{s}(t)$, i.e., it is given by $\tilde{\boldsymbol{s}}(m,t)=\boldsymbol{\Omega}_{\mathrm{M}}^T\boldsymbol{\Lambda}^T(m)\boldsymbol{s}(t)$, where $\boldsymbol{\Lambda}(m)$ is a $K\times K$ permutation matrix, and $\boldsymbol{\Omega}_{\mathrm{M}}(m)\in\{0,1\}^{K\times L_T}$ is the antenna selection matrix, which has a single nonzero entry in each row. When the channel is memoryless, the signal received at the communications receiver is

$$y^{(c)}(m,t)=\boldsymbol{g}_c^T(m)\tilde{\boldsymbol{s}}(m,t)+w^{(c)}(m,t), \tag{4}$$

where $\boldsymbol{g}_c$ is the $L_T\times 1$ channel vector, and $w^{(c)}(m,t)$, is the additive noise. After matched filtering with the orthogonal waveforms, the obtained vector can be written as

$$\boldsymbol{y}^{(c)}(m) = \int y^{(c)}(t, m)\boldsymbol{s}(t)\,dt = \boldsymbol{\Lambda}(m)\boldsymbol{\Omega}_{\mathrm{M}}(m)\boldsymbol{g}_c + \boldsymbol{w}(m). \quad (5)$$

The communication message can be embedded in $\boldsymbol{\Lambda}(m) \times \boldsymbol{\Omega}_{\mathrm{M}}(m)$ in (5), i.e., the product of the permutation matrix and selection matrix. As there are $\binom{L_T}{K}$ kinds of antenna selection patterns and $K!$ kinds of waveform permutations, up to $\log_2\binom{L_T}{K} + \log_2 K!$ bits can be encapsulated in each PRI.

IM via FAR

FAR (see "Frequency Agile Radar") is a radar scheme designed for congested environments. The carrier frequencies of FAR change randomly from pulse to pulse, allowing the achievement of an ergodic wideband coverage, while utilizing narrowband waveforms and enabling the mitigation of interference from neighboring radars. The work [42] proposed a DFRC system that embeds a digital message into the permutation of the agile carrier frequencies. For a carrier set with N different carrier frequencies, there are $N!$ different carrier frequency permutations that can be utilized for information embedding.

In [15] and [16], a DFRC system is proposed based on multicarrier agile waveforms and IM. Unlike traditional FAR, here, multiple carriers are simultaneously sent from several subarrays of transmit antennas. For a DFRC system with L_T transmit antenna elements and a possible carrier frequency set $\mathcal{F}$ of cardinality N, the corresponding information embedding consists of two stages: in the mth pulse, $K < N$ carriers, denoted by the set $\{f_{m,1}, \ldots, f_{m,K}\}$, are first selected from $\mathcal{F}$. Then, the antenna array is divided into K subarrays, where each subarray has $L_K = L_T/K$ elements. The transmitted signal of the multicarrier frequency agile DFRC system in the mth PRI is expressed as

$$\tilde{s}(m, t) = \sum_{k=1}^{K} \boldsymbol{\Omega}_{\mathrm{F}}(m, k)\boldsymbol{u}(\theta, f_{m,k})\,\mathrm{rect}\left(\frac{t - mT_p}{T_p}\right)e^{j2\pi f_{m,k}(t - mT_p)}, \quad (6)$$

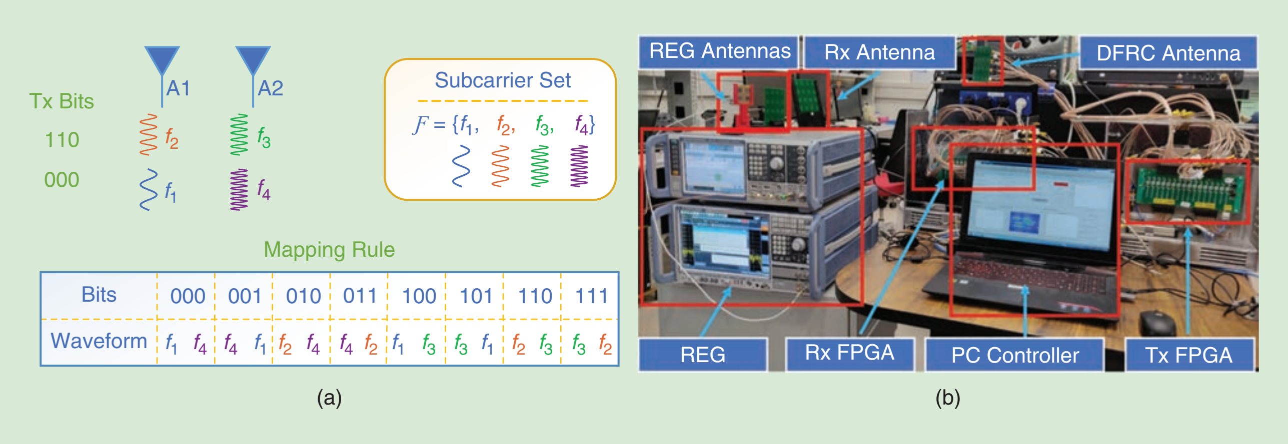
where θ is the beamsteered direction; $\boldsymbol{u}(\theta, f_{m,k})$ is the radar beamforming vector for the kth carrier with frequency $f_{m,k}$; and $\boldsymbol{\Omega}_{\mathrm{F}}(m, k)$ is the selection matrix, which determines the transmit antennas of carrier with frequency $f_{m,k}$. The communication message is embedded into the antenna allocation pattern as well as the selection of carrier frequencies. The number of antenna allocation patterns is $L_T!/(L_K!)^K$, and there are $\binom{N}{K}$ possible combinations of carrier selections. Hence, the total number of transmission patterns that can be used for information embedding is $\binom{N}{K} \cdot L_T!/(L_K!)^K$. An illustration of this scheme as well as a hardware prototype designed in [14] to demonstrate its feasibility are shown in Figure 3.

Since IM-based DFRC systems utilize conventional radar waveforms, radar detection is carried out using standard methods. For example, FAR detection is based on matched filtering followed by compressed sensing recovery [43]. Symbol detection at the communications receiver can be realized using the maximum likelihood rule or, alternatively, via a reduced complexity IM detector; see, e.g., [15].

The main advantage of radar waveforms-based DFRC methods is that they provide the ability to communicate with minimal degradation to the performance of the radar scheme from which the technique originates. For example, the radar performance of MIMO radar as well as FAR combined with IM are roughly equivalent to their radar-only counterparts [16], respectively. In particular, FAR is attractive for automotive radar due to its inherent applicability in congested setups and compliance with simplified hardware. Nonetheless, the communications functionality of radar waveform-based DFRC systems is relatively limited in throughput and typically results in increased decoding complexity, making it more suitable to serve as an alternative channel in addition to existing, e.g., cellular-based, vehicular communications, rather than replacing the latter.

Joint waveform design

The approaches detailed so far are all based on traditional radar and/or communications signaling. A DFRC system is



Bits	000	001	010	011	100	101	110	111
Waveform	f_1 f_4	f_4 f_1	f_2 f_4	f_4 f_2	f_1 f_3	f_3 f_1	f_2 f_3	f_3 f_2

(a)

(b)

FIGURE 3. (a) An illustration of IM-FAR [16]. The array consists of $L_T = 2$ elements, divided into $K = 2$ subarrays of $L_K = 2$ elements. The carrier set is $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$. The mapping rule represents the codebook. (b) A hardware prototype implementing IM-FAR [14] equipped with 16 antenna elements. Rx: receiver; Tx: transmitter.

then obtained by either designing the conventional waveforms to coexist, as detailed in the "Separate Coordinated Signals" section, or alternatively, by using only one standard waveform while extending it to be dual functional. Using traditional signaling techniques has clear advantages due to their established performance and applicability with existing hardware devices. Nonetheless, the fact that these waveforms were not originally designed for DFRC scenarios implies that one can achieve improvement by deriving dedicated dual-function waveforms, as illustrated in Figure 2(c).

Dedicated joint waveforms, which do not originate from conventional radar/communications signaling, are designed according to a dual-function objective, which accounts for the performance of both radar and communications [10]–[12]. Here, the transmitted joint signal is denoted by the $L_T \times J$ matrix $\boldsymbol{X}$, where J is the block length. We focus on a multiuser scenario with L_U single antenna receivers. The signal received at the receivers and at the radar target with direction θ can be expressed as

$$\boldsymbol{Y}^{(c)} = \boldsymbol{HUX} + \boldsymbol{W}^{(c)}, \text{ and } \boldsymbol{y}_\theta^{(r)} = \boldsymbol{a}^T(\theta)\boldsymbol{UX}, \tag{7}$$

where $\boldsymbol{H}$ is an $L_U \times L_T$ channel matrix, $\boldsymbol{U}$ is the joint beamformer, and $\boldsymbol{W}^{(c)}$ is the additive noise term.

Using (8), one can design the joint waveform $\boldsymbol{X}$ to approach some desired observations at the communications receivers as well as the radar target, as proposed in [10]. A possible drawback is that the signals received in other directions are not constrained, and, thus, the radar transmit beampattern may have a high sidelobe level outside the main lobe. This can be overcome by restricting the radar beampattern [11], [12], which is, in turn, achieved by constraining the signal covariance. In particular, [11] considered $\boldsymbol{X}$ to be a communications signal and optimized the joint precoding to approach a predefined beampattern while meeting a minimal SINR level at each receiver. The work [12] designed the joint waveform $\boldsymbol{X}$ to minimize the multiuser interference under specific radar constraints, such as omnidirectional or directional beampatterns, constant modulus designs, and waveform similarity.

Dual-function waveforms specifically designed for DFRC offer the ability to balance radar and communications in a controllable manner. Furthermore, using joint optimization, without being restricted to conventional waveforms, can potentially yield any achievable performance tradeoff between radar and communications. Despite these clear theoretical benefits, their application in an automotive DFRC system is currently still limited due to practical considerations. For example, current joint waveform designs involve solving a relatively complex optimization problem, which depends on prior channel knowledge. In fast-moving vehicles, accurate instantaneous channel knowledge is difficult to obtain, and even when it is available, the optimization process must be frequently repeated, inducing increased computational burden.

Discussion

The DFRC methods surveyed here vary significantly in their characteristics, such as radar performance, communication throughput, complexity, and hardware requirements. Although several efforts have been made in the literature to characterize the achievable radar-communications tradeoff in DFRC systems [4, Ch. 6], to date, there is no unified joint measure that allows the rigorous evaluation of different schemes.

To demonstrate the challenge in comparing DFRC methods, we numerically evaluate two promising schemes: OFDM waveforms, which utilize communications signaling for radar probing, and the radar-based IM via the FAR method. In particular, we consider a single-antenna automotive radar in the 24-GHz band divided into 1,024 bins, using the same configuration as in [9]. OFDM utilizes the complete frequency band, while IM-FAR uses a single subcarrier at each instance, embedding the message in its selected index, i.e., a total of $\log_2 1{,}024 = 10$ bits/symbol. To guarantee that both methods operate with the same data rate, we group the OFDM subcarriers into 10 distinct blocks and assign a binary phase shift keying symbol

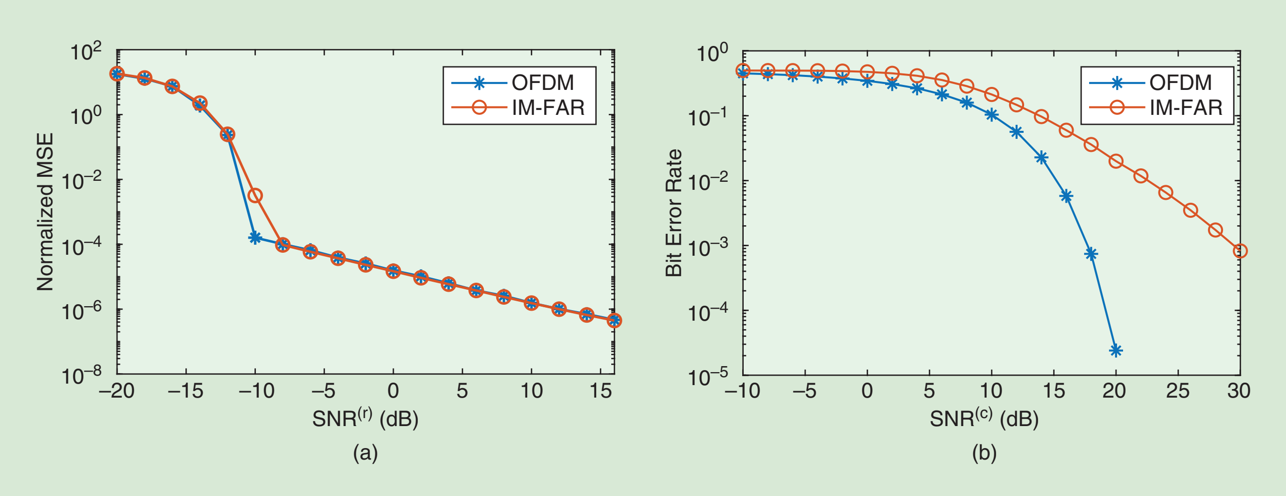


FIGURE 4. A numerical comparison of OFDM-based DFRC systems to FAR with IM. (a) The normalized MSE in range estimation. (b) The bit error rate. SNR: signal-to-noise ratio; (r): radar; (c): communications.

to each block. Both schemes use the same pulsewidth, PRI, and power, attempting to recover a point target with range 10 m and relative velocity 5 m/s, while communicating over a Rayleigh flat fading channel.

The resulting normalized mean-square error (MSE) in target range recovery as well as the communications bit error rate are depicted in Figure 4. Observing Figure 4, we note that OFDM achieves improved communications performance over IM-FAR, while their radar performance is relatively similar. The results in Figure 4, which are in favor of OFDM-based DFRC systems, are relevant for interference-free scenarios, where a single DFRC system probes the environment. In dense scenarios with multiple interfering devices, which model automotive systems in urban settings, FAR is expected to be more capable of mitigating the mutual interference due to its random spectral sparsity [16].

Due to the difficulty in comparing DFRC schemes, we schematically evaluate their radar versus communications performance tradeoff in Figure 5. Separate coordinated transmission methods, which utilize individual signals for each functionality, support a broad range of possible performance combinations, determined by how the system resources are allocated between the functionalities. In particular, beamforming techniques, which require a priori channel knowledge, allow the signals to utilize the full bandwidth and operation time and, thus, have the potential to achieve improved performance compared to time–frequency division strategies. Nonetheless, in the presence of multiple scatterers and communications receivers, which is the case in vehicular applications, obtaining accurate channel knowledge and mitigating mutual interference by beamforming may be infeasible, while spectral division can be applied with controllable complexity, regardless of the number of receivers and their physical location.

Communications waveform-based approaches, particularly when using OFDM transmission, support high data rates by utilizing conventional digital communications signals. Specifically, OFDM is a digital communications scheme that has some of the characteristics of good radar waveforms. In the context of autonomous vehicles, a major limitation of this approach is that, since a single directed beam is used, the receiver should be located in the radar search area. Furthermore, OFDM transmission requires relatively costly hardware, and its radar capabilities are degraded when utilized by a moving vehicle.

Protocol-oriented approaches, which represent an extreme case of using a communications waveform for radar probing, offer the ability to utilize existing vehicular communications protocols for sensing. They provide minimal communications degradation with limited radar capabilities. As such, these methods can be considered as an additional sensing technology, which should not replace dedicated automotive radar.

Radar waveform-based schemes, especially IM-based DFRC systems, can be naturally integrated into automotive radar systems with minimal effect on their performance. While MIMO radar implementing instantaneous wideband waveforms offers improved radar performance over frequency agile waveforms, the latter may be preferable for vehicular applications due to their robustness to congested environments and reduced complexity. Nonetheless, the limited bit rates of IM and its associated decoding complexity make such DFRC schemes more suitable to provide an additional communications channel, independent of the cellular network. The usage of such channels for safety and emergency messages can be valuable in autonomous vehicles, increasing the probability of their successful transmission.

Joint waveform design techniques optimize a dual-function waveform in light of a combined constraint on each functionality. This joint approach has the potential of achieving any given tradeoff between radar and communications performances. Nonetheless, being a relatively new field of study, current dual-function designs may not be suitable for automotive applications. In particular, current designs require instantaneous channel knowledge, limiting their application for self-driving vehicles.

To conclude, there is no single DFRC method that is suitable for all scenarios and requirements encountered in autonomous vehicle applications. Understanding the advantages and disadvantages of each approach will allow engineers to properly select the technologies incorporated into future self-driving cars.

Conclusions and future challenges

Autonomous vehicles implement wireless communications as well as automotive radar, which both require the transmission

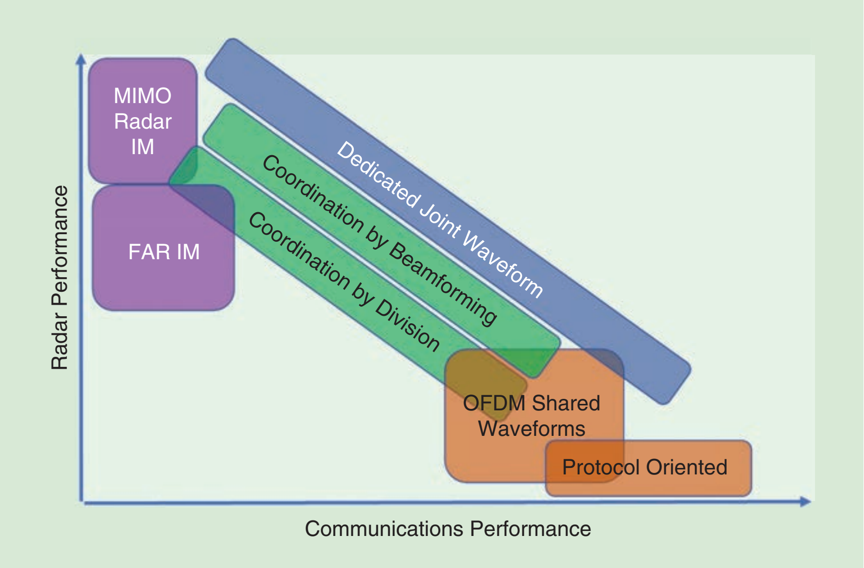


FIGURE 5. A schematic comparison between considered DFRC schemes in terms of their radar-communications tradeoff.

and reception of electromagnetic signals. Jointly designing these functionalities as a DFRC system provides potential gains in performance, size, cost, power consumption, and robustness, making it an attractive approach for autonomous vehicles. In this survey, we reviewed state-of-the-art DFRC designs, focusing on their application for autonomous vehicles. To that aim, we first reviewed the basics of automotive radars. Then, we mapped existing DFRC strategies, proposing their division into four main categories: coexistence schemes, which utilize independent waveforms for each functionality; communications waveform-based approaches, where conventional communications signals are used for radar probing; radar waveform-based schemes, which embed the digital message into standard radar technologies; and joint waveform design approaches, which achieve the DFRC system by deriving dedicated dual-function waveforms. The pros and cons of each category were analyzed according to the radar and communications requirements in vehicular scenarios. While we conclude that no single DFRC scheme is suitable for all of the scenarios in self-driving, our analysis can significantly facilitate the design of sensing and communications technologies for future autonomous vehicles.

While joint radar-communications designs have been studied for over a decade, they still give rise to a multitude of unexplored research directions, particularly in the context of autonomous vehicles. On the theoretical side, the lack of a unified performance measure makes it difficult to compare approaches, and one must resort to heuristic arguments, as was done in this article. Such an analysis will also uncover the fundamental limits of DFRC designs, characterizing their optimal gain over well-studied separate systems. From an algorithmic perspective, the utilization of joint nonstandard radar and communications waveforms, utilized in some of the aforementioned strategies, can be facilitated by the development of dedicated recovery and decoding algorithms.

For conventional waveforms, such as OFDM signals, efficient allocation of resources to optimize both functionalities is a relatively fresh area of study. Additionally, the presence of multiple sensing vehicular technologies, such as vision-based sensing and lidar, along with the ability to communicate with neighboring devices that also sense their environment, give rise to potential improved understanding of the surroundings by properly combining these technologies. Finally, on the practical side, future investigations are required to implement these strategies in vehicular platforms and test their performance in real road environments. Such combined studies should allow us to characterize the benefits and limitations of DFRC systems for self-driving cars, allowing their theoretical potential to be translated into performance gains in this emerging and exciting technology.

Acknowledgments

This work received funding from the National Natural Science Foundation of China under grants 61801258 and 61571260, the European Union's Horizon 2020 research and innovation program under grant 646804-ERC-COG-BNYQ, and the Air Force Office of Scientific Research under grant FA9550-18-1-0208.

Authors

Dingyou Ma (mdy16@mails.tsinghua.edu.cn) received his B.S. degree in aerospace science and technology from Xidian University, Xi'an, China, in 2016. He is currently pursuing his Ph.D. degree with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests are the system design and signal processing on dual-function radar-communications system.

Nir Shlezinger (nirshlezinger1@gmail.com) received his B.Sc., M.Sc., and Ph.D. degrees in 2011, 2013, and 2017, respectively, from Ben-Gurion University, Israel, all in electrical and computer engineering. From 2017 to 2019, he was a postdoctoral researcher in the Technion–Israel Institute of Technology and is currently a postdoctoral researcher in the Signal Acquisition Modeling, Processing, and Learning Lab, Weizmann Institute of Science. From 2009 to 2013, he worked as a research and development engineer at Yitran Communications. His research interests include communications, information theory, signal processing, and machine learning. He is a Member of the IEEE.

Tianyao Huang (huangtianyao@tsinghua.edu.cn) received his B.S. degree in 2009 in telecommunication engineering from the Harbin Institute of Technology, Heilongjiang, China, and his Ph.D. degree in 2014 in electronics engineering from Tsinghua University, Beijing, China. From 2014 to 2017, he was a radar researcher in the Aviation Industry Corporation of China. Since July 2017, he has been an assistant professor with the Intelligent Sensing Lab, Department of Electronic Engineering, Tsinghua University. His research interests include compressed sensing, radar signal processing, waveform design, and joint design of radar and communication. He is a Member of the IEEE.

Yimin Liu (yiminliu@tsinghua.edu.cn) received his B.S. and Ph.D. degrees (both with honors) in electronics engineering from Tsinghua University, Beijing, China, in 2004 and 2009, respectively. Since 2004, he has been with the Intelligence Sensing Lab, Department of Electronic Engineering, Tsinghua University, where he is currently an associate professor studying new concepts of radar and other microwave sensing technologies. His research interests include radar theory, statistic signal processing, compressive sensing and their applications in radar, spectrum sensing, and intelligent transportation systems. He is a Member of the IEEE.

Yonina C. Eldar (yonina.eldar@weizmann.ac.il) is a professor in the Department of Math and Computer Science at the Weizmann Institute of Science, Rehovot, Israel, where she heads the center for Biomedical Engineering and Signal Processing. She is also a visiting professor at the Massachusetts Institute of Technology and the Broad Institute and an adjunct professor at Duke University. She is a member of the Israel Academy of Sciences and Humanities and a fellow of the European Association for Signal Processing. She has received the IEEE Signal Processing Society Technical Achievement Award, IEEE/AESS Fred Nathanson Memorial Radar Award, IEEE Kiyo Tomiyasu Award, Michael Bruno Memorial Award from the Rothschild Foundation, Weizmann Prize for Exact Sciences, and the Wolf Foundation Krill Prize for Excellence in Scientific Research. She is the editor-in-chief of *Foundations and Trends in Signal Processing* and serves the IEEE on several technical and award committees. She is a Fellow of the IEEE.

References

[1] S. M. Patole, M. Torlak, D. Wang, and M. Ali, “Automotive radars: A review of signal processing techniques,” *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, Mar. 2017. doi: 10.1109/MSP.2016.2628914.

[2] H. H. Meinel, “Evolving automotive radar–From the very beginnings into the future,” in *Proc. 8th European Conf. Antennas and Propagation (EuCAP)*, 2014, pp. 3107–3114. doi: 10.1109/EuCAP.2014.6902486.

[3] X. Wang, S. Mao, and M. X. Gong, “An overview of 3GPP cellular vehicle-to-everything standards,” *GetMobile: Mobile Comput. Commun.*, vol. 21, no. 3, pp. 19–25, 2017. doi: 10.1145/3161587.3161593.

[4] B. Paul, A. R. Chiriyath, and D. W. Bliss, “Survey of RF communications and sensing convergence research,” vol. 5, pp. 252–270, Dec. 2017. doi: 10.1109/ACCESS.2016.2639038.

[5] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, “Signaling strategies for dual-function radar communications: An overview,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 10, pp. 36–45, Oct. 2016. doi: 10.1109/MAES.2016.150225.

[6] Y. Liu, G. Liao, J. Xu, Z. Yang, and Y. Zhang, “Adaptive OFDM integrated radar and communications waveform design based on information theory,” *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2174–2177, Oct. 2017. doi: 10.1109/LCOMM.2017.2723890.

[7] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, “Dual-function radar-communications: Information embedding using sidelobe control and waveform diversity,” *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2168–2181, Apr. 2016. doi: 10.1109/TSP.2015.2505667.

[8] A. Hassanien, M. G. Amin, Y. D. Zhang, F. Ahmad, and B. Himed, “Non-coherent PSK-based dual-function radar-communication systems,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2016, pp. 1–6. doi: 10.1109/RADAR.2016.7485066.

[9] C. Sturm and W. Wiesbeck, “Waveform design and signal processing aspects for fusion of wireless communications and radar sensing,” *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011. doi: 10.1109/JPROC.2011.2131110.

[10] P. M. McCormick, S. D. Blunt, and J. G. Metcalf, “Simultaneous radar and communications emissions from a common aperture, part I: Theory,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1685–1690. doi: 10.1109/RADAR.2017.7944478.

[11] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, “MU-MIMO communications with MIMO radar: From co-existence to joint transmission,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018. doi: 10.1109/TWC.2018.2803045.

[12] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, “Toward dual-functional radar-communication systems: Optimal waveform design,” *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, Aug. 2018. doi: 10.1109/TSP.2018.2847648.

[13] M. Bică and V. Koivunen, “Multicarrier radar-communications waveform design for RF convergence and coexistence,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 7780–7784. doi: 10.1109/ICASSP.2019.8683655.

[14] D. Ma, N. Shlezinger, T. Huang, Y. Shavit, M. Namer, Y. Liu, and Y. C. Eldar, Spatial modulation for joint radar-communications systems: Design, analysis, and hardware prototype. 2020. [Online]. Available: arXiv:2003.10404

[15] T. Huang, N. Shlezinger, X. Xu, Y. Liu, and Y. C. Eldar, MAJoRCom: A dual-function radar communication system using index modulation. 2019. [Online]. Available: arXiv:1909.04223

[16] T. Huang, N. Shlezinger, X. Xu, D. Ma, Y. Liu, and Y. C. Eldar, Multi-carrier agile phased array radar. 2019. [Online]. Available: arXiv:1906.06289

[17] L. Reichardt, C. Sturm, F. Grünhaupt, and T. Zwick, “Demonstrating the use of the IEEE 802.11p car-to-car communication standard for automotive radar,” in *Proc. 6th European Conf. Antennas and Propagation (EuCAP)*, Mar. 2012, pp. 1576–1580. doi: 10.1109/EuCAP.2012.6206084.

[18] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, “IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018. doi: 10.1109/TVT.2017.2774762.

[19] A. R. Chiriyath, B. Paul, and D. W. Bliss, “Radar-communications convergence: Coexistence, cooperation, and co-design,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 1, pp. 1–12, 2017. doi: 10.1109/TCCN.2017.2666266.

[20] C. Aydogdu, M. F. Keskin, N. Garcia, H. Wymeersch, and D. W. Bliss, Radchat: Spectrum sharing for automotive radar interference mitigation. 2019. [Online]. Available: arXiv:1908.08280

[21] G. C. Tavik, C. L. Hilterbrick, J. B. Evins, J. J. Alter, J. G. Crnkovich, J. W. de Graaf, W. Habicht, G. P. Hrin et al., “The advanced multifunction RF concept,” *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 3, pp. 1009–1020, Mar. 2005. doi: 10.1109/TMTT.2005.843485.

[22] C. Aydogdu, N. Garcia, and H. Wymeersch, “Improved pedestrian detection under mutual interference by FMCW radar communications,” in *Proc. IEEE 29th Annu. Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2018, pp. 101–105. doi: 10.1109/PIMRC.2018.8581028.

[23] C. Sturm, Y. L. Sit, M. Braun, and T. Zwick, “Spectrally interleaved multi-carrier signals for radar network applications and multi-input multi-output radar,” *IET Radar Sonar Nav.*, vol. 7, no. 3, pp. 261–269, Mar. 2013. doi: 10.1049/iet-rsn.2012.0040.

[24] S. Sodagari, A. Khawar, T. C. Clancy, and R. McGwier, “A projection based approach for radar and telecommunication systems coexistence,” in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2012, pp. 5010–5014. doi: 10.1109/GLOCOM.2012.6503914.

[25] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, Joint transmit beamforming for multiuser MIMO communication and MIMO radar. 2019. [Online]. Available: arXiv:1912.03420

[26] T. Huang and T. Zhao, “Low PMEPR OFDM radar waveform design using the iterative least squares algorithm,” *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1975–1979, Nov. 2015. doi: 10.1109/LSP.2015.2449305.

[27] A. Turapaty, Y. Jin, and Y. Xu, “Range and velocity estimation of radar targets by weighted OFDM modulation,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2014, pp. 1358–1362. doi: 10.1109/RADAR.2014.6875811.

[28] N. Levanon, “Multifrequency complementary phase-coded radar signal,” *IEE Proc. Radar Sonar Nav.*, vol. 147, no. 6, pp. 276–284, 2000. doi: 10.1049/ip-rsn:20000734.

[29] G. Franken, H. Nikookar, and P. V. Genderen, “Doppler tolerance of OFDM-coded radar signals,” in *Proc. European Radar Conf. (EuRAD)*, 2006, pp. 108–111. doi: 10.1109/EURAD.2006.280285.

[30] M. Braun, C. Sturm, A. Niethammer, and F. K. Jondral, “Parametrization of joint OFDM-based radar and communication systems for vehicular applications,” in *Proc. 20th Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2009, pp. 3020–3024. doi: 10.1109/PIMRC.2009.5449769.

[31] G. Hakobyan and B. Yang, “A novel OFDM-MIMO radar with non-equidistant dynamic subcarrier interleaving,” in *Proc. European Radar Conf. (EuRAD)*, Oct. 2016, pp. 45–48.

[32] C. Knill, F. Roos, B. Schweizer, D. Schindler, and C. Waldschmidt, “Random multiplexing for an MIMO-OFDM radar with compressed sensing-based reconstruction,” *IEEE Microw. Compon. Lett.*, vol. 29, no. 4, pp. 300–302, 2019. doi: 10.1109/LMWC.2019.2901405.

[33] G. Lellouch, A. K. Mishra, and M. Inggs, “Stepped OFDM radar technique to resolve range and doppler simultaneously,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 51, no. 2, pp. 937–950, Apr. 2015. doi: 10.1109/TAES.2014.130753.

[34] D. W. Bliss and K. W. Forsythe, “Multiple-input multiple-output (MIMO) radar and imaging: Degrees of freedom and resolution,” in *Proc. 37th Asilomar Conf. Signals, Systems and Computers*, Nov. 2003, pp. 54–59. doi: 10.1109/ACSSC.2003.1291865.

[35] P. Kumari, R. W. Heath, and S. A. Vorobyov, “Virtual pulse design for IEEE 802.11ad-based joint communication-radar,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 3315–3319. doi: 10.1109/ICASSP.2018.8461678.

[36] E. Grossi, M. Lops, L. Venturino, and A. Zappone, “Opportunistic automotive radar using the IEEE 802.11ad standard,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1196–1200. doi: 10.1109/RADAR.2017.7944386.

[37] C. Sahin, J. Jakabosky, P. M. McCormick, J. G. Metcalf, and S. D. Blunt, “A novel approach for embedding communication symbols into physical radar waveforms,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1498–1503. doi: 10.1109/RADAR.2017.7944444.

[38] G. N. Saddik, R. S. Singh, and E. R. Brown, “Ultra-wideband multifunctional communications/radar system,” *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 7, pp. 1431–1437, July 2007. doi: 10.1109/TMTT.2007.900343.

[39] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, “Index modulation techniques for next-generation wireless networks,” *IEEE Access*, vol. 5, pp. 16,693–16,746, Aug. 2017. doi: 10.1109/ACCESS.2017.2737528.

[40] E. BouDaher, A. Hassanien, E. Aboutanios, and M. G. Amin, “Towards a dual-function MIMO radar-communication system,” in *Proc. IEEE Radar Conf. (RadarConf)*, May 2016, pp. 1–6. doi: 10.1109/RADAR.2016.7485316.

[41] X. Wang, A. Hassanien, and M. G. Amin, “Dual-function MIMO radar communications system design via sparse array optimization,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 55, no. 3, pp. 1213–1226, June 2019. doi: 10.1109/TAES.2018.2866038.

[42] X. Wang and J. Xu, “Co-design of joint radar and communications systems utilizing frequency hopping code diversity,” in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019. doi: 10.1109/RADAR.2019.8835576.

[43] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

MIMO Radar for Advanced Driver-Assistance Systems and Autonomous Driving

Advantages and challenges

©ISTOCKPHOTO.COM/JUST_SUPER

Shunqiao Sun, Athina P. Petropulu, and H. Vincent Poor

Important requirements for automotive radar are high resolution, low hardware cost, and small size. Multiple-input, multiple-output (MIMO) radar technology has been receiving considerable attention from automotive radar manufacturers because it can achieve a high angular resolution with relatively small numbers of antennas. For that ability, it has been exploited in the current-generation automotive radar for advanced driver-assistance systems (ADAS) as well as in next-generation high-resolution imaging radar for autonomous driving. This article reviews MIMO radar basics, highlighting the features that make this technology a good fit for automotive radar and reviewing

---

Digital Object Identifier 10.1109/MSP.2020.2978507
Date of current version: 26 June 2020

important theoretical results for increasing the angular resolution. The article also describes challenges arising during the application of existing MIMO radar theory to automotive radar that provide interesting problems for signal processing researchers.

Introduction

Radar technology has been used in defense, civilian, and commercial applications since World War II. Surveillance radars have been in service for a number of decades. Ground-based air surveillance radars, such as air traffic control radar, are usually pulse radar systems with maximum detectable ranges of more than 100 km [1]. Airborne surveillance radars carried by aircraft and unmanned aerial systems, such as the joint surveillance target attack radar system, can run in ground moving target indication and synthetic aperture radar imagving modes [2] to detect moving and stationary targets on the ground at more than 250 km, respectively. To achieve a satisfactory return-signal power, the transmit power of such radar can be several kilowatts. Large phased-array antennas have been deployed to achieve electronic beam scanning. Since the late 1990s, radar sensors have found widespread applications in ADASs, such as adaptive cruise control (ACC) and automatic emergency braking (AEB) [3], [4]. More recently, radar has emerged as one of the key technologies in autonomous driving systems, providing environmental perception in all weather conditions [5], [6]. Some of today's self-driving cars, such as Zoox, have more than 10 radars, providing 360° surround sensing (see Figure 1). Differing from ground-based and airborne surveillance radars, automotive radars have a small size (multi-inch by multi-inch), short range (within multihundreds of meters), low power consumption (multiwatt), and low cost. They are integrated behind the vehicle bumper or windshield, operating in a highly dynamic environment with multipath, as shown in Figure 2.

FIGURE 1. A Zoox autonomous driving vehicle, exhibited at the 2019 Conference on Computer Vision and Pattern Recognition, is equipped with 10 millimeter-wave (mm-wave) automotive radar sensors (marked by white arrows) at the front, rear, left, and right sides.

Figure 3 presents a typical radar perception configuration for autonomous driving. Four short-range radar (SRR) sensors with an azimuth field of view (FoV) of $[-75^\circ, 75^\circ]$ and a detection range of 45 m are deployed at the four corners of the vehicle for blind-spot detection, cross traffic alerts, and so forth. Two midrange radar (MRR) sensors with an azimuth FoV of $[-40^\circ, 40^\circ]$ and a detection range of 100 m are deployed at both the front and rear sides for lane-change assistance and AEB. A long-range radar (LRR) sensor with an azimuth FoV of $[-15^\circ, 15^\circ]$ and a detection range of 250 m is deployed at the front for the ACC. All radar sensors in a vehicle are connected to an electronic control unit for further processing, such as radar tracking and sensor fusion [7].

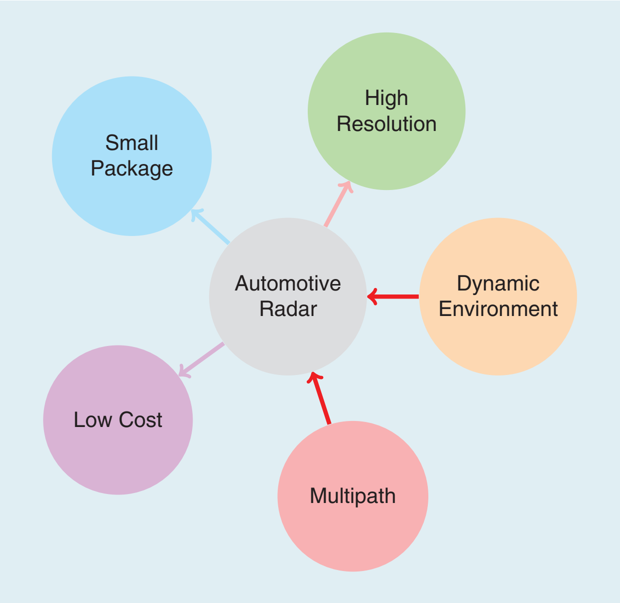


FIGURE 2. The requirements of automotive radar for ADAS and autonomous driving (indicated by outward-pointing arrows) and the challenges imposed by the environment (indicated by inward-pointing arrows).

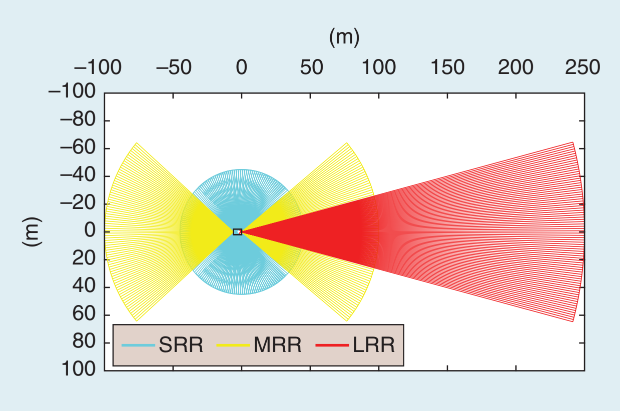


FIGURE 3. A typical radar-perception configuration [5] for autonomous driving contains four short-range radars (SRRs), deployed at the four corners of the car and providing 360° coverage; two midrange radars (MRRs) for both forward and rearward looking; and one LRR for forward looking.

For control critical functions, such as AEB, distant obstacles need to be detectable at a high angular resolution. Therefore, automotive radar for ADAS and autonomous driving needs to have a high angle discrimination capability. Employing a large antenna array would improve the angular resolution; however, the resulting large package size would make integration on the vehicle difficult. While for conventional phased-array radar, a small package size implies a low angular resolution, for MIMO radar [8]–[10], the package size is not a limiting factor. This is because MIMO radar can synthesize virtual arrays with a large aperture using only a small number of transmit and receive antennas.

Important requirements for automotive radar are high resolution, low hardware cost, and small size.

This advantage has been exploited by almost all major automotive suppliers in their different types of radar products, such as SRR, MRR, and LRR [11]–[14]. For Level 4 and Level 5 (L4/L5) autonomous driving, both the azimuth and elevation angular resolution of automotive radar need to be less than 1º. High-resolution imaging radar with hundreds of virtual array elements is currently being developed to produce the so-called point clouds, which are groups of points detected by radar that represent the object's shape [11], [15], [16]. A variety of signal processing tools, e.g., the fast Fourier transform (FFT), short-time Fourier transform, filtering, and beamforming, have been adopted in automotive radar to obtain target features, such as a micro-Doppler spectrum of pedestrians [17]–[20] and a range–Doppler spectrum of the surrounding environment. Machine learning algorithms and deep neural networks [21] have also been applied in automotive radar for target recognition and classification.

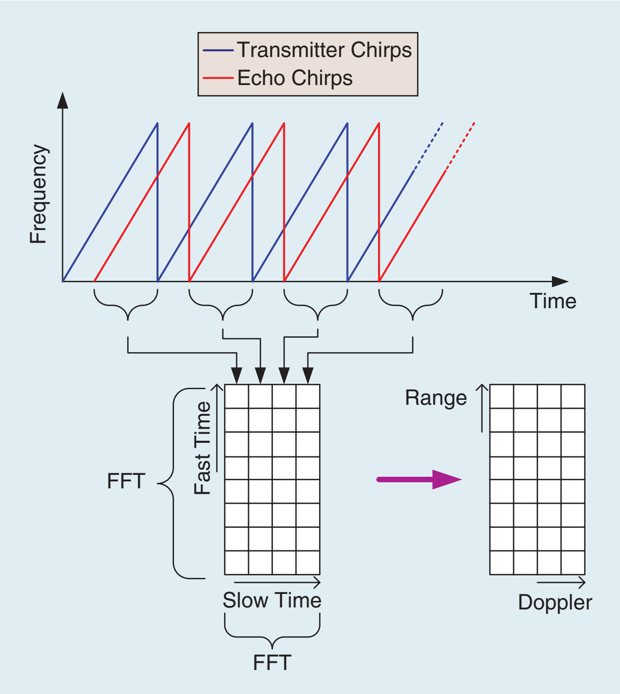


FIGURE 4. The FMCW radar chirps. The range and Doppler estimations are performed using a 2D FFT.

In summary, to meet the requirement for ADAS and especially L4/L5 autonomous driving, automotive radar needs to have a high angular resolution, small package size, and low cost. Since it is meant to operate in highly dynamic environments, it needs to detect obstacles fast, especially for time critical functions, such as AEB. In this article, we discuss MIMO radar with millimeter-wave (mm-wave) technology as a means of achieving the aforementioned requirements. We review key issues in state-of-the-art frequency-modulated continuous-waveform (FMCW) MIMO radar, different strategies for achieving waveform orthogonality, virtual array synthesis, and high-resolution angle finding methods for both uniform linear arrays and sparse linear arrays. We discuss MIMO radar in the light of high-resolution imaging radar for L4/L5 autonomous driving. We also highlight challenges in designing automotive MIMO radar, such as angle finding in the presence of multipath, waveform-orthogonality strategies in FMCW radar and pulse-modulated CW (PMCW) radar, and the mitigation of radar mutual interference. We hope this article serves as a tutorial on automotive MIMO radar and provides interesting research problems to signal processing researchers.

State-of-the-art automotive FMCW radar

Mm-wave technology has found great applicability in automotive radar. The typical frequency band of mm-wave automotive radar is 76–81 GHz. The high frequencies facilitate small enough antennas that can fit behind the bumper of the vehicle. Also, the wide available bandwidth enables a high target-range resolution [22]. State-of-the-art automotive radar transmits FMCW at mm-wave frequencies, which supports high-resolution target-range and velocity estimation at a much lower cost than lidar technology. Automotive MIMO radar uses FMCW waveforms along with some mechanism that guarantees waveform orthogonality. This section introduces the principles of FMCW. Strategies for achieving waveform orthogonality based on FMCW waveforms are discussed in the "Introduction of Automotive Radar With MIMO Radar Technology" section.

An FMCW waveform, also referred to as a *chirp*, is a complex sinusoid whose frequency increases linearly with time $t \in [0, T]$, i.e., $f_T(t) = f_c + (B/T)t$, where B is the signal bandwidth and f_c is the carrier frequency. FMCW radar transmit chirps in a periodic fashion, with a period referred to as the *pulse repetition interval* (*PRI*). The frequency of an FMCW signal through multiple periods, with a PRI equal to T, is displayed in Figure 4. The target echo at the radar receiver contains a delayed and attenuated copy of the transmitted chirp. For a target at range R, moving with a radial speed of v, the delay equals $\tau = (2(R + vt))/c$, where time t spans multiple periods and c is the speed of light. The received signal is mixed with the transmitted chirp, which results in a complex sinusoid known as the *beat signal*. The beat-signal frequency equals $f_b = f_R + f_D$, where $f_R = 2RB/(Tc)$ is the range frequency and $f_D = (2v/c)f_c$ is the Doppler frequency.

The process of obtaining the beat signal is implemented in the radio frequency domain by a mixer, followed by a bandpass filter (BPF) with the maximum cutoff frequency $f_b^{\max}$; the latter filter is used to remove signals with frequencies outside the band of interest, which also places a limit on the maximum detectable range.

The estimation of the beat frequency is implemented in the digital domain, after the sampling of the beat signal. In automotive scenarios, the maximum detectable range, $R_{\max}$, is hundreds of meters. It holds that $(2R_{\max}/c) \ll T$, and thus $f_R \ll B$. Since it typically holds that $f_D \ll f_R$, the beat frequency is much smaller than B, and therefore a low-speed analog-to-digital converter (ADC) can be used to sample the beat signal. The time during one period or chirp is usually referred to as the *fast time*, while the time across multiple periods or chirps is referred to as the *slow time*. Thus, if we sample the beat signal and put the samples of each chirp in the columns of a matrix, the row indices of that matrix correspond to the fast time and the column indices to the slow time (see Figure 4). In automotive scenarios, $f_D \ll f_R$; therefore, f_D can be taken as constant within each chirp. Thus, by applying FFTs on the sampled beat signal along the fast time, one can identify f_R based on which of the target's ranges can be obtained as $R = cf_R T/(2B)$.

Radar technology has been used in defense, civilian, and commercial applications since World War II.

To obtain the target's Doppler frequency, a second FFT operation is subsequently carried out along the slow time (the range frequency f_R is the same across the slow time). The PRF is $f_{\text{PRF}} = 1/T_{\text{PRI}}$. To avoid Doppler ambiguity, it is desired that $f_{\text{PRF}} \geq 2f_D$. Thus, the maximum unambiguous detectable radial speed of FMCW radar is $v_{\max} = c/(4f_c T_{\text{PRI}})$ [23]. The application of these two FFTs is equivalent to a 2D FFT of the beat signal in the fast and slow times, and the result is called the *range–Doppler spectrum*. Range and Doppler detection can be performed using conventional thresholding-based methods applied to the 2D range–Doppler spectrum, such as the constant false alarm rate detector [24] or the recently proposed deep neural network-based detector [25]. Via the 2D FFT, the targets can be separated in the range and Doppler domains. Since the number of targets within the same range–Doppler bin is small, angle finding can be carried out using sparse sensing techniques, such as compressive sensing. The details of angle finding approaches will be covered in the "Angle Finding in Automotive MIMO Radar" section.

The 2D FFT operation used for beat-frequency estimation can be computed with low-cost digital signal processors (DSPs) and field-programmable gate arrays. The range resolution depends on the beat-frequency resolution. Since the latter is estimated based on a signal of time duration approximately equal to T, the beat-frequency resolution is $1/T$, and thus the range resolution is $c/(2B)$. One can see how the range resolution benefits from a high bandwidth. Thus, the low hardware cost, coupled with the high range resolution that can be achieved, make FMCW radar very popular in automotive radar applications.

As an example, an LRR FMCW radar operating at carrier frequency $f_c = 77$ GHz has a typical pulse duration of $T = 50$ μs and bandwidth of $B = 150$ MHz. Assume the maximum detectable range is 250 m. Then, the maximum range frequency is $f_R^{\max} = 5$ MHz. For a moving target with a maximum speed of $v = 120$ mi/h, the maximum Doppler frequency is $f_D^{\max} = 27.4$ KHz. The beat signal can be sampled with a low-cost ADC with a sampling rate as low as $f_s = 12$ MHz. For a PRI equal to T, and based on the preceding, the PRF is $f_{\text{PRF}} = 1/T_{\text{PRI}} = 20$ KHz. It can be seen that, in this example, $f_{\text{PRF}} < 2f_D^{\max}$, and therefore the Doppler frequency will be aliased, resulting in Doppler ambiguity. In practice, chirp sequences with different PRIs are transmitted to resolve the Doppler ambiguity.

As discussed, FMCW radar has the desirable ability to estimate the target range and Doppler at a low cost. However, FMCW radar also faces several challenges. Those include frequency sweep nonlinearity [26]–[28] arising from the phase noise of the local oscillator (LO) and imperfect phase-locked loop circuits and voltage-control oscillators. Sweep nonlinearity has a significant impact on radar range estimation [26].

A challenge common to all radars is to maintain a high dynamic range, which is the ratio between the maximum and minimum possible successfully received power. Since the path loss is inversely proportional to the fourth order of the range, targets at long distances typically yield weaker reflections compared to targets at closer distances. Therefore, a high dynamic range is required to accommodate targets at a wide range of distances. In FMCW radar, because the waveforms are transmitted continuously, the receiver always receives the signals of the transmitters via a direct path, referred to as *leakage*. Leakage may saturate the low-noise amplifier in the front end of the radar receiver [29]. It can also cause high peaks at the first few bins of the range FFT, thus reducing the radar's dynamic range.

In many pulse radars, sensitivity time control has been widely used to control the gain of targets at near and far distances by exploiting the relationship between the time delay and target range. As discussed, the range frequency of the beat signal is proportional to the target range. Therefore, in FMCW radar, an analog gain-control logic can be implemented in the frequency domain to reduce the receiver gain of close-range targets while continually increasing the gain as the target's range increases [30]. Automatic gain-control logic can be implemented in a BPF so that the receiver gain increases within the detectable range [31]. Leakage between the transmit and receive antennas can be suppressed via sufficient isolation between the transmit and receive antennas through antenna design in the analog domain [30], [32]. Further, the transmit antenna leakage effect in the mixer output can be suppressed via a BPF, with the lowest cutoff frequency corresponding to a range of tens of centimeters to 1 m. The BPF also helps suppress the multibounce signals between automotive radar, the radome, and the vehicle bumper. Finally, in the complete receiving chain, an N-bit ADC can provide an additional $6N$-dB dynamic range.

In the signal processing chain of FMCW radar, a range FFT length of N_R and a Doppler FFT length of N_D can provide a signal-to-noise ratio (SNR) improvement of $10\log_{10}(N_R N_D)$ dB [33]. This SNR improvement is considered as a processing gain, which significantly benefits angle finding that will be discussed in the "Angle Finding in Automotive MIMO Radar" section.

Introduction of automotive radar with MIMO radar technology

In state-of-the-art automotive FMCW radar, the range and Doppler parameters of targets can be estimated by using a single receive antenna. However, to estimate the angle parameter of targets, a receive antenna array is needed. In MIMO radar, the transmit antennas transmit FMCW sequences in a way that guarantees their orthogonality. At each receive antenna, the contribution of each transmit antenna is extracted by exploiting waveform orthogonality. For M_t transmit antennas and M_r receive antennas, a virtual array with $M_t M_r$ elements can be synthesized. The array response of the synthesized array, i.e., the array corresponding to a MIMO radar, can be expressed as $\mathbf{a}_t(\theta_t) \otimes \mathbf{a}_r(\theta_r)$, where $\mathbf{a}_t(\theta_t)$ and $\mathbf{a}_r(\theta_r)$ are the transmit- and receive-array steering vectors corresponding to a direction of departure (DoD) θ_t and direction of arrival (DoA) θ_r, respectively. Here, $\otimes$ denotes the Kronecker product.

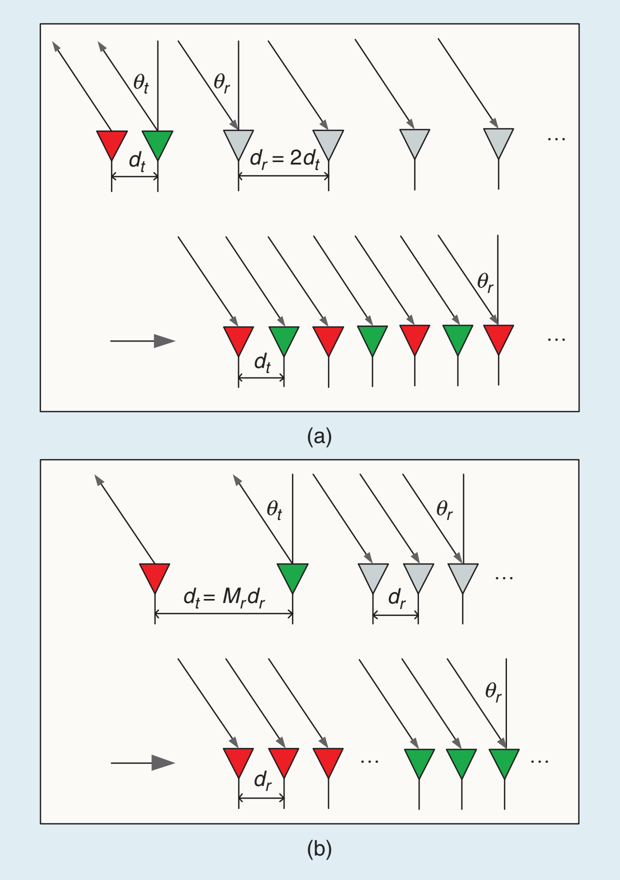


FIGURE 5. The different MIMO radar virtual array configurations [23] using the time-division multiplexing (TDM) or Doppler-division multiplexing (DDM) scheme with $M_t = 2$ transmit antennas and M_r receive antennas: (a) interleaved, with $d_r = 2d_t$ and $d_t = \lambda/2$; (b) stacked, wit $d_t = M_r d_r$ and $d_r = \lambda/2$. Different colors indicate that the transmit antenna transmits different time slots or codes.

Figure 5 shows two MIMO radar configurations with $M_t = 2$ transmit and M_r receive antennas [23]. One is in interleaved mode with $d_r = 2d_t$ and $d_t = \lambda/2$, and the other is in stacked mode with $d_t = M_r d_r$ and $d_r = \lambda/2$. Here, $\lambda = c/f_c$ is the wavelength. It can be verified that for both array configurations, the synthetic virtual arrays are equivalent to uniform linear arrays (ULAs) with $M_t M_r$ elements and a spacing of d_r when $\theta_t = \theta_r$.

Because MIMO radar transmits orthogonal waveforms, when isotropic array elements are used, the array beampattern (also referred to as the *MIMO radar array factor*) is omnidirectional. Thus, MIMO radar loses the coherent array-processing-gain advantage enjoyed by traditional phased-array radar systems [10]; the SNR of the array response at a given angular direction is less than that of phased-array radar with transmit beamforming. Still, in the automotive application scenario, the high-resolution angle finding ability of MIMO radar coupled with its low cost are viewed as more important than the loss of coherent processing gain.

Virtual array synthesis in automotive FMCW radar using MIMO radar technology relies on the separability of the transmit signals of the different antennas. The separation is easier when the transmit signals of different antennas are orthogonal. In the following, we review techniques to achieve waveform orthogonality while transmitting FMCW, such as TDM, DDM, and frequency-division multiplexing (FDM).

Waveform orthogonality via TDM

In TDM MIMO radar [23], [35]–[38], only one transmit antenna is scheduled to transmit at each time slot. In Figure 6, a signal processing example of a MIMO radar in TDM is given by NXP Semiconductors [34], where $M_t = 2$ transmit antennas emit FMCW chirps alternatively. The switch delay between transmit antennas is $\Delta t = T_{\mathrm{PRI}}$. At each receive antenna, range FFTs of length N_r are conducted for each chirp, and the FFT outputs of $2N_d$ chirps are assembled in two matrices corresponding to odd and even chirp sequences, respectively. The receive array corresponding to the odd and even chirp sequences form two subarrays, which can be used to synthesize a virtual array according to interleaved or stacked configurations.

For a moving target with a velocity of v, the switching delays of the transmit antennas introduce a target phase migration from chirp to chirp, which is defined as $\phi = 2\pi f_D \Delta t = (4\pi/\lambda) v \Delta t$. As a result, the virtual array pattern would be distorted [23]. In Figure 6, the phase difference between corresponding columns in the two matrices is $\phi = (4\pi/\lambda) v T_{\mathrm{PRI}}$. If $v = -0.5 v_{\max}$ and $v = -v_{\max}$, where $v_{\max}$ is the maximum unambiguous detectable radial speed and $v_{\max} = c/(4 f_c T_{\mathrm{PRI}})$ (see the "State-of-the-Art Automotive FMCW Radar" section), the phase shifts are $\phi = -\pi/2$ and $\phi = -\pi$, respectively. The array beampattern distortion is demonstrated in Figure 7 for a moving target

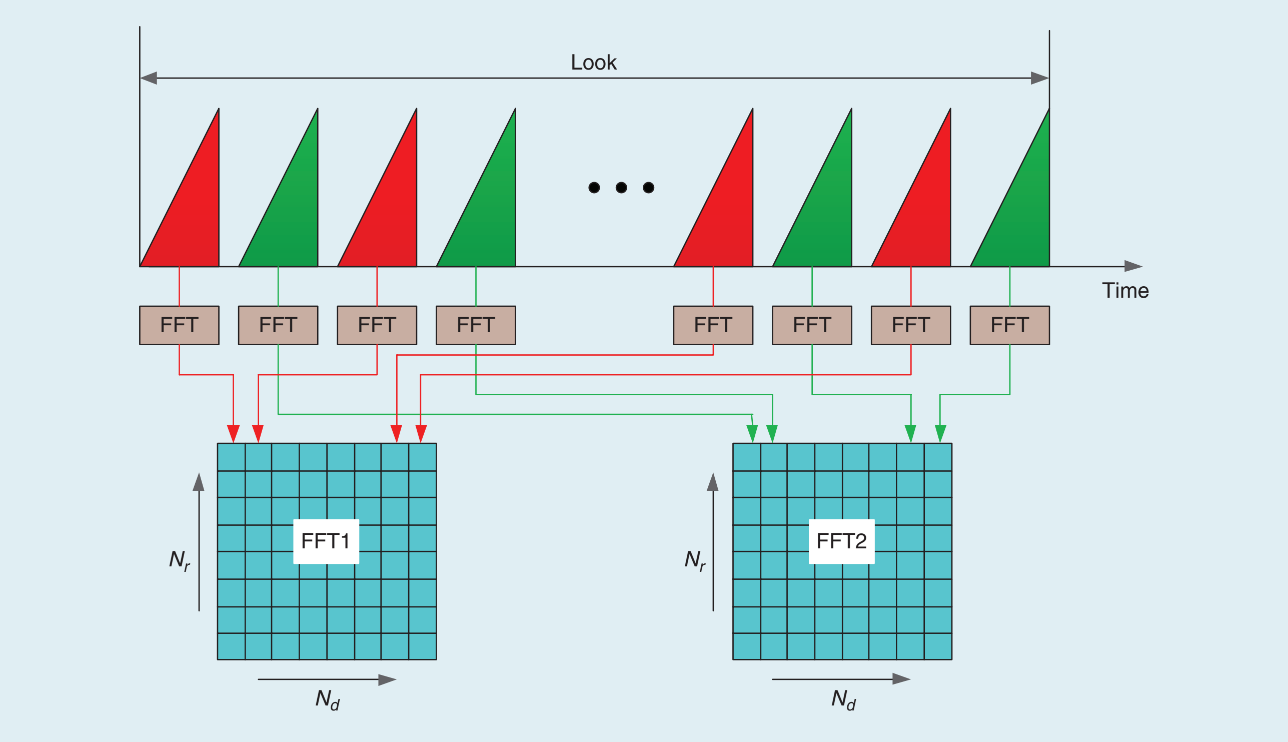


FIGURE 6. An example of radar signal processing with the TDM scheme [34], where $M_t = 2$ transmit antennas alternately transmit FMCW chirp sequences. The red and green colors denote the odd and even echo chirp sequences, respectively. The range FFTs are conducted for each chirp, and the FFT outputs are stored in two matrices corresponding to odd and even sequences, respectively, for further processing.

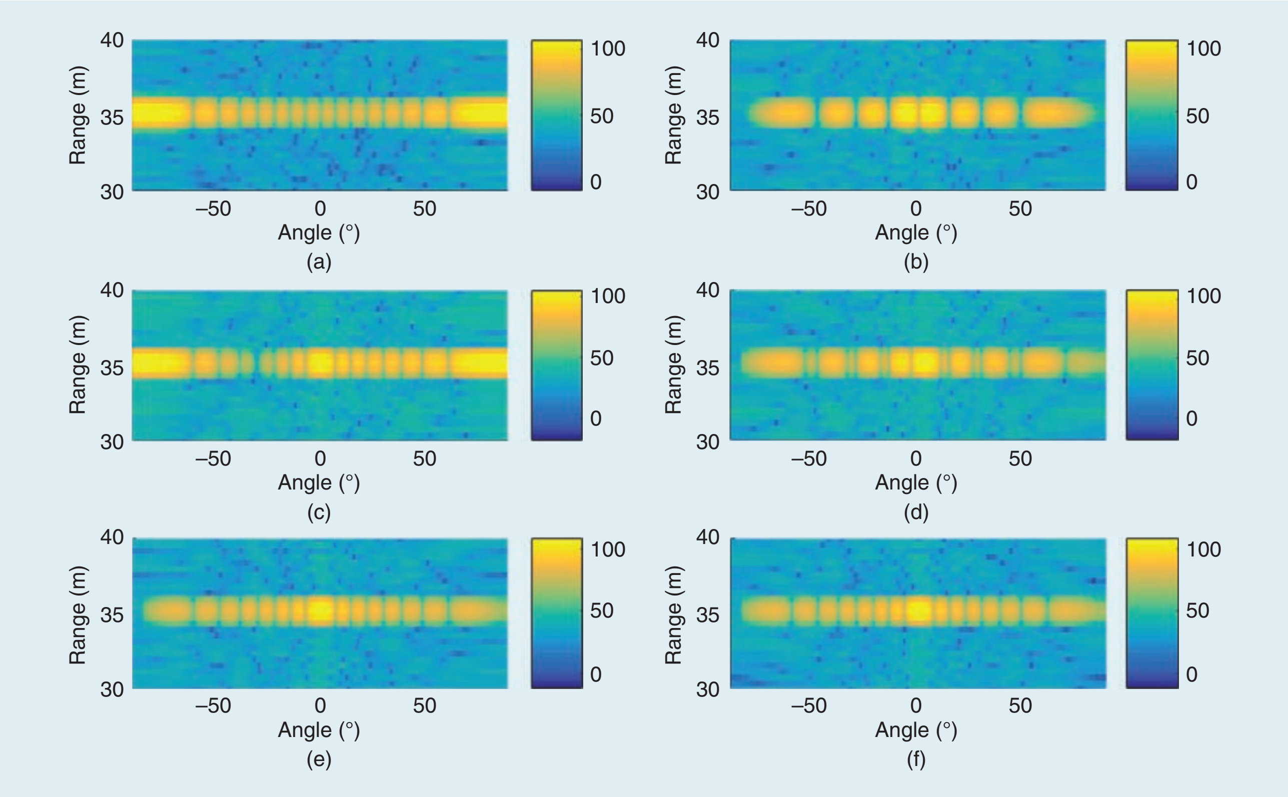


FIGURE 7. The examples of the MIMO radar range and azimuth images using TDM [23], with $M_t = 2$, $M_r = 8$ for a single target with a range of 35 m and azimuth angle of $\theta = 0°$. Two MIMO array configurations, interleaved (left column) and stacked (right column), are considered. (a) and (b) The radial velocity of $v = -v_{\max}$. (c) and (d) The radial velocity of $v = -0.5\,v_{\max}$. (e) and (f) The radial velocity of $v = 0$.

with a range of 35 m and azimuth angle of $\theta = 0°$ [23]. Here, MIMO radar with $M_t = 2$, $M_r = 8$ operates in TDM fashion. As stated in [23] and seen in Figure 7, for the interleaved MIMO array configuration, when the target velocity increases, the grating lobes at the edge of the FoV show up, while the peak at the target direction decreases and totally disappears when $v = -v_{\max}$. For the stacked MIMO array configuration, as the target velocity increases, the peak is slightly off the boresight, with a mirror grating lobe at the opposite direction.

The phase migration introduced by every moving target in the virtual array response needs to be compensated for before angle finding. The phase-shift estimate $\hat{\phi}$ can be obtained after each target velocity has been estimated based on the 2D FFT of a single receive antenna or the noncoherent 2D FFT integration of the same subarray. For instance, in the example of Figure 6, the phase in the beam vector of the subarray obtained from the even chirps needs to be compensated for by multiplying it with $e^{-j\hat{\phi}}$, while the phase in the beam vector of the subarray corresponding to the odd chirps is kept unchanged. It should be noted that in TDM MIMO radar, the pulse-repetition interval is enlarged by the transmit antenna number M_t. As a result, the maximum unambiguous detectable velocity, which was defined in the "State-of-the-Art Automotive FMCW Radar" section, will be reduced by a factor of M_t [39].

Waveform orthogonality via DDM

In one look, a total of N chirps (i.e., pulses) are transmitted sequentially with pulse-repetition interval T_{PRI}. All transmit antennas simultaneously transmit the same FMCW waveform after multiplying it with a phase code that is different for each antenna and changes between pulses, i.e., $x_m(n) = e^{j2\pi\alpha_m(n)}$, $m = 1, \ldots, M_t$, $n = 1, \ldots, N$ [40]. To separate the hth transmit signal at the lth receiver, after the range FFT, a slow-time Doppler demodulation is applied to all range bins corresponding to the same chirp. The Doppler demodulated outputs of N chirps are assembled into a vector $\mathbf{s}_l^h$. Then, the Doppler FFT is applied on the vector $\mathbf{s}_l^h$. To separate the transmit signals in the Doppler domain, one of the two methods described next can be applied.

The first approach is to design phase codes such that the Doppler FFT of the interference $e^{j2\pi(\alpha_m(n)-\alpha_h(n))}$ is shifted to a frequency that is higher than the maximum detectable Doppler frequency $f_D^{\max}$. Therefore, a low-pass filter (LPF) can be applied to remove the interference [40]. One example of such phase codes is

$$\alpha_m(n) = \alpha_m n, \quad m = 1, \ldots M_t, \quad n = 1, \ldots N, \tag{1}$$

where the starting phase α_m is linear across different transmit antennas, i.e., $\alpha_m = a_0 m$. Figure 8(a) shows the range and Doppler spectra of a target with a range of 75 m and velocity of 10 m/s. Automotive MIMO radar has two transmit antennas and the transmit phase codes given in (1), with $a_0 = 1$ and $N = 512$. It can be seen that signals from different transmit antennas are shifted to a higher Doppler spectrum, which can be removed via an LPF in the Doppler domain. With this approach, the radar pulse repetition frequency f_{PRF} should be larger than $M_t f_D^{\max}$ [41]. Thus, if the f_{PRF} is kept unchanged, the maximum detectable unambiguous Doppler frequency is reduced by a factor of M_t. In practice, a Doppler unfolding, or de-aliasing, algorithm needs to be developed with different f_{PRF} in different looks.

The second approach is to design phase codes so that the Doppler FFT of the interference can be distributed into the entire Doppler spectrum as pseudo noise. It is desired to minimize the peak interference residual (PIR) in the Doppler spectrum [42], calculated using the discrete time Fourier transform for $m = 1, \ldots, M_t$, i.e.,

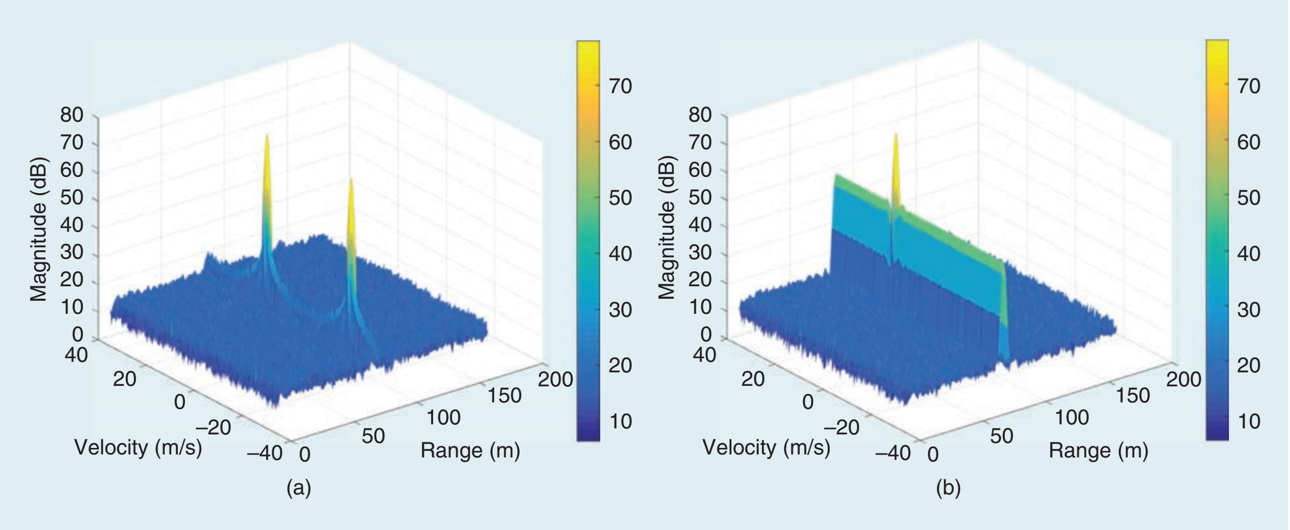


FIGURE 8. The range and Doppler spectra of a target with a range of 75 m and velocity of 10 m/s. The automotive MIMO radar has two transmit antennas, and a slow time phase coding of length $N = 512$ is applied for DDM. (a) The phase shift codes defined in (1) for the range-Doppler spectrum. (b) The two Chu sequences for the range-Doppler spectrum.

$$\mathrm{PIR} = \max_{f, m \neq h} \left| \sum_{n=1}^{N} e^{j2\pi(\alpha_m(n) - \alpha_h(n))} e^{j2\pi f n} \right|, \tag{2}$$

where $f \in [-(1/2)f_{\mathrm{PRF}}, (1/2)f_{\mathrm{PRF}}]$. Following (2), the cross correlation of the spectra of two codes needs to be flat [42] since the Fourier transform of multiplication of two codes in the time domain is equivalent to the convolution of the spectrum of one code with the time reversed and the complex conjugate of the other. The maximum autocorrelation value of an unimodular sequence of length N is N. The ideal cross correlation of two unimodular sequences of length N has a magnitude of $\sqrt{N}$. Thus, in the ideal case, according to [42], the maximum power gain of the currently transmitted signal over the other signals is $\sqrt{N}$. For example, the maximum achievable waveform attenuation is roughly 27.1 dB for a unimodular sequence set with $N = 512$.

Radar has emerged as one of the key technologies in autonomous driving systems, providing environmental perception in all weather conditions.

Constant-amplitude zero-autocorrelation codes are good candidates for DDM. The discrete Fourier transform of a constant-amplitude zero autocorrelation code also has a constant amplitude and zero autocorrelation [43]. One such example is the Chu sequence [44], which is defined as $x_m(n) = e^{(j\pi/N)m(n+1)n}$, $m = 1, \ldots, M_t$, $n = 1, \ldots, N$, where N is a prime number. In practice, the Chu sequence of the prime length is first generated and then truncated into a length for an efficient FFT. For example, we generate Chu codes of prime length 521 and truncate them to length $N = 512$. By calculation with the FFT, the peak interference residual defined in (2) is $1.08\sqrt{N}$. Therefore, the waveform attenuation for a Chu sequence of length $N = 512$ is approximately 26.4 dB. In Figure 8(b), we show the range and Doppler spectra of a target with a range of 75 m and velocity of 10 m/s. The automotive radar has two transmit antennas, and two Chu sequences of length $N = 512$ are applied for the slow-time DDM. It can be seen that the waveform attenuation is roughly 26 dB. In practice, binary phase codes are used due to hardware constraints [45]. The binary phase-code sequences are obtained via an exhaustive search such that the peak interference residual in (2) is low. As the code length increases, the search time will grow exponentially.

The benefit of slow time phase coding is that the interference from other transmitters does not affect different range bins. The range resolution is determined only by the bandwidth of the FMCW chirp. Therefore, it avoids the range sidelobe issue using fast time phase coding. However, the Doppler sidelobes would be high due to the residual of the slow time phase coding. As a result, targets with a low radar cross section (RCS), e.g., pedestrians, that are close to targets that have strong reflections, e.g., trucks, might be masked by the waveform residual. In other words, the waveform residual reduces the radar dynamic range. Given the code length, the number of phase codes with good correlation properties is limited, or, equivalently, the number of antennas that can transmit simultaneously is limited.

Waveform orthogonality via FDM

In the FDM scheme, the transmitted signals are modulated by different carrier frequencies. According to [46], the separation of multiple transmit FMCW signals is achieved by shifting the mth transmit FMCW chirp by an offset frequency $f_{\mathrm{off},m}$. If the differences between all $f_{\mathrm{off},m}$ are larger than twice the cutoff frequency of the antialiasing BPF $f_b^{\max}$, which is determined by the maximum unambiguous detectable range and Doppler, the transmitted signals can be separated at the receive end. Specifically, the received signal at each receiver is first mixed with the same starting carrier frequency f_c. The separation of transmit signals in the mixer output can be implemented by a frequency shift followed by an LPF with cutoff frequency $f_b^{\max}$ [46]. Each receiver needs to carry out such a frequency shift and filtering operation M_t times. As a result, a high range resolution can be realized using a typical FMCW chirp with a large bandwidth. Meanwhile, after the FMCW demodulation, frequency shift, and filtering operation, the FDM MIMO scheme can still utilize a low sampling rate determined by the beat signal.

Let us consider the FDM scheme in the context of the example in the "State-of-the-Art Automotive FMCW Radar" section, i.e., a FMCW LRR radar with a maximum detectable range of 250 m and a maximum detectable velocity of 120 mi/h. For bandwidth $B = 150$ MHz and chirp duration $T = 50\ \mu$s, the maximum beat frequency is $f_b^{\max} = f_R^{\max} + f_D^{\max} = 5.0274$ MHz. Therefore, the frequency shift for the mth transmit antenna in the FDM scheme can be chosen as $f_{\mathrm{off},m} = 12(m-1)$ MHz. The intermediate frequency (IF) should have a bandwidth of $12\,M_t$ MHz to hold the mixer output.

Angle finding in automotive MIMO radar

In automotive MIMO radar with M_t transmit and M_r receive antennas, a virtual uniform linear array of $M_t M_r$ elements can be synthesized with interelement spacing d. The array response can be written as

$$\mathbf{y} = \mathbf{A}(\theta)\mathbf{s} + \mathbf{n}, \tag{3}$$

where $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)]$ is the virtual array steering matrix with

$$\mathbf{a}(\theta_k) = \left[1, e^{j(2\pi/\lambda)d\sin(\theta_k)}, \ldots, e^{j(2\pi/\lambda)(M_t M_r - 1)d\sin(\theta_k)}\right]^T. \tag{4}$$

Here, $\mathbf{n}$ is a noise term, and $\mathbf{s} = [\beta_1, \ldots, \beta_K]^T$, where β_k denotes the target reflection coefficient for the kth target. The array response at a particular time instance consisting of data obtained at all the virtual receivers and corresponding to the same range–Doppler bin is defined as the *array snapshot*. In highly dynamic automotive scenarios, usually only a small number of array snapshots, or even a single snapshot in the worst case, is available [47].

In automotive MIMO radar with a virtual ULA, angle finding can be done with digital beamforming (DBF) [12], [48], [49] by performing FFTs on snapshots taken across the array elements, i.e., $\mathbf{y}$ in (3) (see Figure 9). DBF can be implemented efficiently in an embedded DSP with a single snapshot. However, DBF is not a high-resolution angle-finding method. Higher-resolution angle finding can be achieved with subspace-based methods, such as multiple signal classification (MUSIC) [50] and estimation of signal parameters via rational invariance techniques (ESPRIT) [51]–[54], sparse sensing-based methods [55]–[63], and the iterative adaptive approach (IAA) of [64] and [65]. The performance of subspace-based angle-finding methods relies on accurate estimation of the array covariance matrix with multiple snapshots, which is a challenging task in the highly nonstationary automotive radar scenarios. In such a context, spatial smoothing [66] is applied for introducing virtual snapshots for array covariance-matrix estimation. While sparse sensing-based methods and IAA have a high computational cost, they yield angle estimates based on a single snapshot, which is important for snapshot-limited automotive radar.

MIMO radar can synthesize virtual arrays with a large aperture using only a small number of transmit and receive antennas.

Achieving a high angular resolution for the L4 and L5 autonomous driving requirement using a ULA with $d = \lambda/2$ is very expensive. According to [24], the 3-dB beamwidth of an antenna array with aperture size D is $\delta_\theta = 2\arcsin(1.4\lambda/(\pi D))$. To achieve a 3-dB beamwidth of 1°, the antenna-array aperture should be roughly $D \approx 51\lambda$. If the antenna array is a ULA with its interelement spacing as one-half of a wavelength, it should be composed of approximately 100 array elements. Even with the help of MIMO radar technology, the cost of synthesizing such a large virtual ULA with half-wavelength element spacing is very high. One way to further reduce the cost without sacrificing the high angular resolution is via the use of nonuniform, or sparse linear arrays (SLAs) [67]–[72] synthesized with MIMO radar technology. In that context, selecting the locations of the array elements and carrying out angle finding with the virtual sparse array are key problems.

High-resolution angle finding with ULAs

Subspace methods with spatial smoothing

The performance of subspace-based angle finding methods requires an estimate of the array covariance matrix. Such an estimate is typically obtained based on multiple snapshots. However, in the highly dynamic automotive environment, it is not possible to obtain enough snapshots before the model of (3) changes. In such scenarios, spatial smoothing [66] can introduce virtual snapshots for array covariance-matrix estimation. In spatial smoothing, the array snapshot, $\mathbf{y}$, is divided into overlapped subarrays of length L, and a new sampled array covariance matrix $\mathbf{R} \in \mathbb{C}^{L \times L}$ is obtained based on the subarray snapshots.

The eigenvalue decomposition of $\mathbf{R}$, along with the Akaike information criteria metric [73] or the minimum description length metric [74], can be used to identify the number of targets. It should be noted, however, that many ideal assumptions in the deduction of these criteria (including additive white Gaussian noise that is uncorrelated with the source signal) and the availability of enough snapshots for an accurate covariance matrix estimation might not be satisfied in practice. The target angles can be found by identifying the locations of peaks of the MUSIC pseudospectrum [50], $P(\theta_i)$, computed at all possible θ_is, i.e.,

$$P(\theta_i) = \frac{1}{\mathbf{a}_L^H(\theta_i)\mathbf{U}_\mathbf{n}\mathbf{U}_\mathbf{n}^H\mathbf{a}_L(\theta_i)}, \tag{5}$$

where $\mathbf{U}_\mathbf{n}$ is the noise subspace of $\mathbf{R}$ and $\mathbf{a}_L(\theta_i)$ is the array steering vector of length L corresponding to search direction θ_i. The computation cost of the MUSIC algorithm is high due to the angle search process. Alternatively, the ESPRIT algorithm could be used for angle estimation [51].

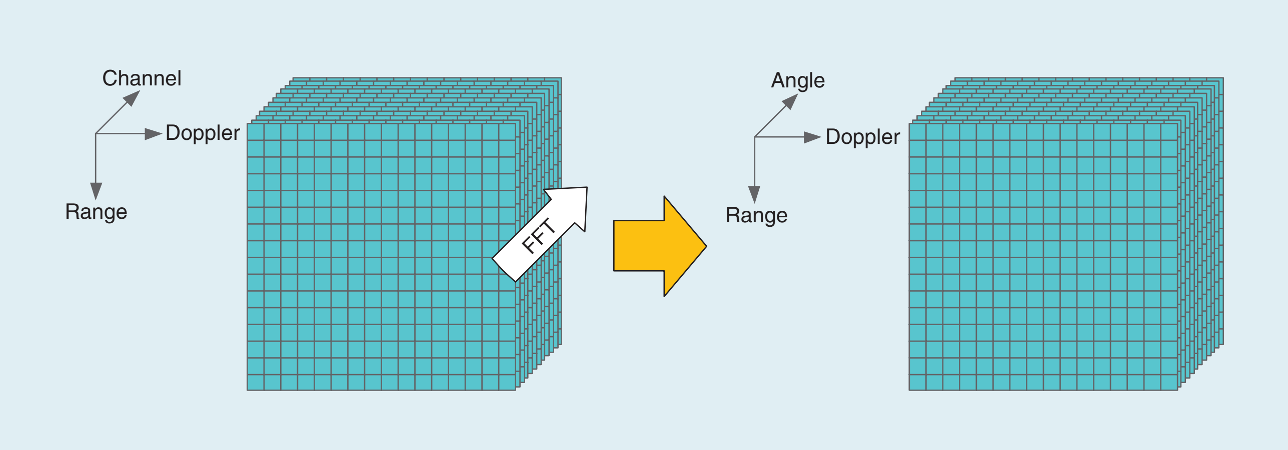


FIGURE 9. The DBF.

ESPRIT is also a subspace method, which exploits the array shift-invariance properties, and has been widely used in practice. It has a lower complexity than MUSIC, which, however, comes at the cost of reduced angular resolution. ESPRIT requires $2L$ sensors, with $L > K$. To achieve the same angular resolution as MUSIC, ESPRIT needs twice as many sensors. Since 2D arrays are needed in automotive radar to estimate both the azimuth and elevation angles, 2D ESPRIT algorithms [54] can be applied if the array element spacing is uniform rectangular.

Compressive sensing

MIMO radars equipped with mm-wave technology offer a wide bandwidth, thus achieving a high range resolution. As a result, there are only a small number of targets that fall in the same range–Doppler bin, and thus the targets are sparse in the DoA space. This property can be exploited by sparse sensing-based high-resolution methods for target angle estimation. To apply compressive sensing for DoA estimation, the whole DoA FoV is discretized into a fine grid. Assume that the DoA space is discretized on a fine grid with N points and that there are K targets on the grid. The array response in (3) can be rewritten as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{6}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_N)]$ is the basis matrix, with $\mathbf{a}(\theta_i)$ denoting the array steering vector corresponding to the ith grid point, and $\mathbf{x} = [\beta_1, \beta_2, \ldots, \beta_N]^T$ is a sparse vector with K nonzero elements. The value of β_i is nonzero if there is a target at the ith grid point. The coherence of the basis matrix, defined as

$$\mu \triangleq \max_{i \neq l} \frac{|\mathbf{a}^H(\theta_i)\mathbf{a}(\theta_l)|}{\|\mathbf{a}(\theta_i)\|_{\ell_2}\|\mathbf{a}(\theta_l)\|_{\ell_2}}, \tag{7}$$

needs to be low for obtaining uniform recovery guarantees [75]. When meeting the required coherence conditions, the DoA can be found by solving an ℓ_1-norm optimization problem, such as the Dantzig selector [76], defined as

$$\begin{aligned} &\min \|\mathbf{x}\|_{\ell_1} \\ &\text{s.t.} \|\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x})\|_{\ell_\infty} < \eta. \end{aligned} \tag{8}$$

or greedy methods, such as orthogonal matching pursuit (OMP) [77].

In the preceding formulation, targets are assumed to be on the grid, which is not always possible in practice. While one can make the grid finer to capture the targets, the coherence of matrix $\mathbf{A}$ would increase, which would make the ℓ_1-norm solution invalid [78]. Thus, the performance of compressive sensing-based methods is sensitive to targets appearing off the grid [79]. Sparse sensing and matrix completion-based methods [62], [63] can avoid grid issues without sacrificing the high-resolution performance.

IAA

The covariance matrix of M array snapshots $\mathbf{y}_l$, $l = 1, \ldots, M$, can be written as $\mathbf{R} = \mathbf{A}(\theta)\mathbf{P}\mathbf{A}^H(\theta)$, where $\mathbf{P}$ is a $K \times K$ diagonal matrix whose diagonal elements contain the power of target reflections. Angle finding in the IAA algorithm [64], [65] is carried out by iteratively estimating the reflection coefficient β_k. The estimate is found by minimizing the weighted least-square cost function $\sum_{l=1}^{M}\|\mathbf{y}_l - \beta_k(l)\mathbf{a}(\theta_k)\|^2_{\mathbf{Q}^{-1}(\theta_k)}$, where $\|\mathbf{x}\|^2_{\mathbf{Q}^{-1}(\theta_k)} = \mathbf{x}^H\mathbf{Q}^{-1}(\theta_k)\mathbf{x}$ and the interference and noise covariance matrix $\mathbf{Q}(\theta_k) = \mathbf{R} - \hat{P}_k\mathbf{a}(\theta_k)\mathbf{a}^H(\theta_k)$. The solution is given by [64]:

$$\hat{\beta}_k(l) = \frac{\mathbf{a}^H(\theta_k)\mathbf{R}^{-1}\mathbf{y}_l}{\mathbf{a}^H(\theta_k)\mathbf{R}^{-1}\mathbf{a}(\theta_k)}. \tag{9}$$

Then matrix $\mathbf{P}$ can be updated as $\hat{P}_k = (1/M)\sum_{l=1}^{M}|\hat{\beta}_k(l)|^2$. In IAA algorithm implementation, the DoA space is discretized into a fine grid of N points, and steering matrix $\mathbf{A}$ is constructed in the same way as in compressive sensing. In addition, a standard delay-and-sum beamformer is used to initialize $\mathbf{P}$:

$$\hat{P}_k = \frac{\sum_{l=1}^{M}|\mathbf{a}^H(\theta_k)\mathbf{y}_l|^2}{M|\mathbf{a}^H(\theta_k)\mathbf{a}(\theta_k)|^2}. \tag{10}$$

High-resolution angle finding with SLAs

As stated before, the cost of synthesizing a large virtual ULA of D elements with half-wavelength element spacing is very high. One way to further reduce the cost without sacrificing the high angular resolution is via the use of nonuniform or SLAs [67], [68], [80]. With MIMO radar technology, $M_t M_r < D$ virtual array elements can be synthesized. To make the SLA aperture the same as the ULA, two virtual-array elements should be deployed at the edge locations of the ULA. For the remaining virtual array elements, there are multiple possibilities to deploy. The main issue with the SLA is that the grating lobes may introduce ambiguity into angle finding. In that context, the key problems are how to select the locations of the array elements such that the peak sidelobe level (PSL) of the virtual SLA beampattern is low and how to carry out angle finding. There is no analytical solution to determining the antenna locations that achieve a minimum PSL for a given number of antennas [81]. Optimal sparse array design requires global optimization techniques, such as particle-swarm optimization [68], [82], [83].

In automotive MIMO radar with a virtual SLA, angle finding can still be done with conventional FFT or ESPRIT methods if the holes in the virtual SLA can be filled via interpolation or extrapolation techniques to mitigate the grating lobes [16], [84]. Alternatively, instead of filling the holes, angle finding of the sparse array can be done using spatial compressive sensing ideas [85]. In the SLA scenario, it can be easily verified that the coherence of the basis matrix [see (7)] is the PSL of the SLA array beampattern [67]. Therefore, the coherence (or, equivalently, the PSL) of a sparse array plays a key role in obtaining uniform recovery guarantees for compressive sensing [75]. If the PSL of the SLAs is low, angle finding using SLAs can be done via compressive sensing or IAA.

In Figure 10, we give an example of a virtual SLA with an aperture of 19λ, synthesized with MIMO radar technology using four transmit and four receive antennas. The first and fourth transmit/receive antennas are deployed at the edge of the physical aperture, while the remaining antennas are chosen such that the PSL is -9.1 dB. Angle estimation via the IAA when using the sparse linear array of Figure 10 is illustrated in Figure 11. The ground truth involves two targets with azimuths of 5 and 10°. The SNR of the received beam vector is set to 30 dB. For comparison, the FFT spectrum is also plotted. One can observe the sharper peaks around the target azimuth angles and the more attenuated sidelobe in the IAA spectrum as compared with the FFT spectrum.

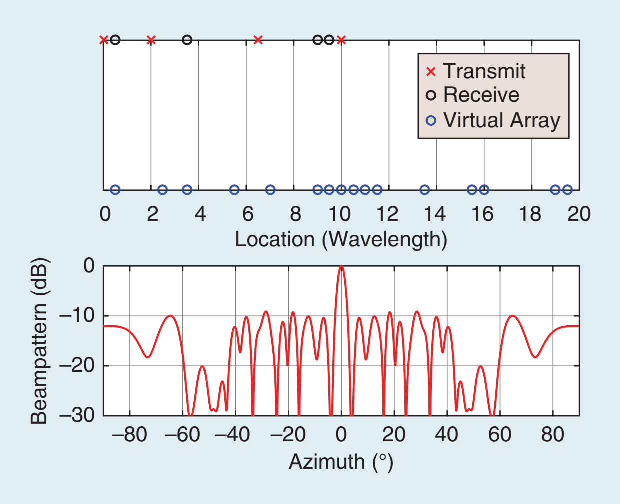


FIGURE 10. A toy example of an SLA synthesized with MIMO radar technology using four transmit and four receive antennas. The physical limitation of the array aperture is 10λ. We fix the locations of the first and fourth transmit/receive antennas at 0λ, 0.5λ and 10λ, 9.5λ, respectively, such that a maximum virtual array aperture of 19λ is achieved. The remaining transmit/receive antennas are chosen such that the PSL of the synthetic virtual array beampattern is -9.1 dB.

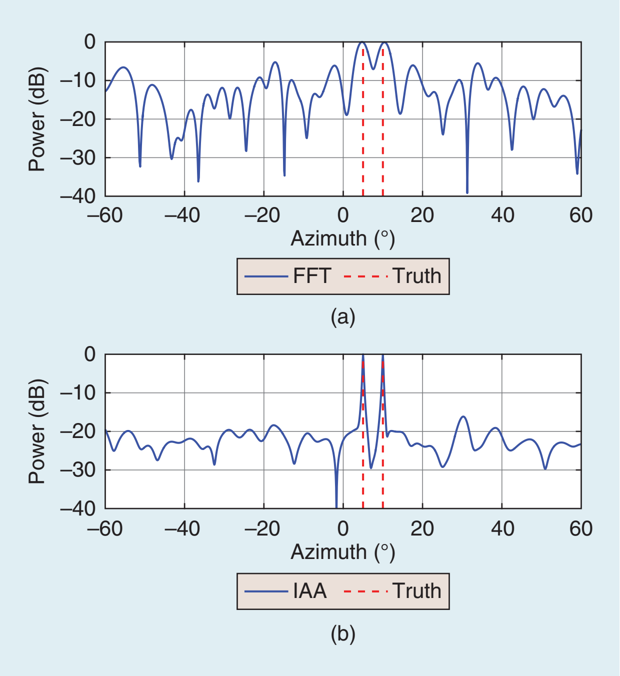


FIGURE 11. Angle finding via (a) the FFT spectrum and (b) the IAA method using the SLA of Figure 10 synthesized by MIMO radar technology. There are two targets with azimuths of 5 and 10°. The SNR is set to 30 dB.

High-resolution imaging radar for autonomous driving

Automotive radar with a small number of antennas has been used for ADAS purposes since the late 1990s. Such radar mostly sees point targets and obtains velocity information. However, the current generation of automotive radar for ADAS has a rather limited ability to resolve closely spaced targets. Lidar systems have better angular resolution (less than 1º) and have been introduced into L4/L5 autonomous driving systems. Lidar can provide point clouds. Through the use of deep neural networks, such as PointNet [86] and PointNet++ [87], the point clouds can lead to target identification. However, due to its use of the light-spectrum wavelength, lidar is susceptible to bad weather conditions, such as fog, rain, snow, and dust in the air. In addition, the cost of lidar is high. On the other hand, automotive radar with mm-waveform technology has the potential to provide point clouds at a much lower cost than lidar and with more robustness to weather conditions. Such radar is referred to as *high-end radar* or *imaging radar* [88]. Computer vision techniques [86], [87] that were previously reserved for high-resolution camera sensors and lidar systems can be applied to imaging-radar data to identify targets. For example, a car can be identified based on the 2D radar points of an imaging radar using PointNet [89]. Imaging radar has been attracting the interest of major Tier-1 suppliers and automotive radar start-ups.

MIMO radar is a good candidate for high-resolution imaging radar for autonomous driving. In MIMO radar using FMCW waveforms, the targets are first distinguished in the range and Doppler domains. Then, a large virtual array with hundreds of elements can be synthesized to provide a high resolution in both the azimuth and elevation. As a result, point clouds with performance similar to lidar can be generated at a much lower cost. In this section, we introduce the concept of imaging radars using MIMO technology; present some examples for synthesizing hundreds of virtual array elements by cascading multiple radar transceivers, with each supporting a small number of antennas; and discuss design challenges.

Cascade of multiple radar transceivers

Today, most of the automotive radar transceivers designed for ADAS functionality, such as the MR3003 from NXP Semiconductor and AWR1243 from Texas Instruments, can support up to three transmit and four receive antennas. Therefore, using a single automotive radar transceiver with MIMO radar technology, only 12 virtual array elements can be synthesized. To meet the requirement for L4 and L5 autonomous driving, multiple automotive radar transceivers would need to be cascaded together, with all the transceivers synchronized as a single unit. The received data from all the receive antennas would be processed coherently. Cascading provides a cost-effective and scalable solution to achieve a high angular resolution.

In [90], General Motors and Texas Instruments successfully demonstrated that up to four Texas Instruments AWR1243 radar chips can be cascaded together to provide 12 transmit and 16 receive antennas, enabling a synthesis of 192 virtual array elements. In [91], a prototype of five cascading Infineon radar chips was built to synthesize a virtual array of 128 × 4 elements. Such a high number of virtual array elements provides a lot of opportunities in array design. Several azimuth and elevation array configurations can be found in [16]. Usually, a tradeoff between balancing the angular resolution in the azimuth and elevation needs to be considered.

Examples of cascaded imaging radars

Figure 12 shows an imaging radar design reference board that has 12 transmit and 16 receive antennas, formed by cascading four Texas Instruments AWR1243 radar transceivers [92]. The azimuth FoV is $[-70°, 70°]$. One transceiver is selected as the master and all the others as slaves for the clock distribution. In this way, synchronization can be achieved among four transceivers, enabling coherent FMCW transmission from the 12 transmit antennas and joint data processing from the 16 receive antennas. The array configuration of cascaded imaging radar is illustrated in Figure 12. There are three transmit antennas placed along the vertical direction for elevation angle finding and nine transmit antennas placed along the horizontal direction for azimuth angle finding. The virtual array in the horizontal direction is a dense ULA with half-wavelength spacing, and it consists of 86 virtual array elements (the overlapped virtual array elements are not shown). The array aperture in the azimuth direction is $D_x = 42.5\lambda$. In antenna theory, the 3-dB beamwidth defines the angular resolution. According to [24], the 3-dB beamwidth of the azimuth angle is

$$\Delta\theta_{\mathrm{AZ}} = 2\arcsin\left(\frac{1.4\lambda}{\pi D_x}\right) \approx 1.2°. \tag{11}$$

In the vertical direction, the antennas in three elevation positions form multiple minimum redundancy arrays (MRAs) [93] along the horizontal direction. Angle finding in the MRAs requires multiple snapshots. These MRAs along the horizontal direction can be used as snapshots for elevation angle finding. The elevation array aperture is $D_y = 3\lambda$, and the 3-dB beamwidth of elevation is

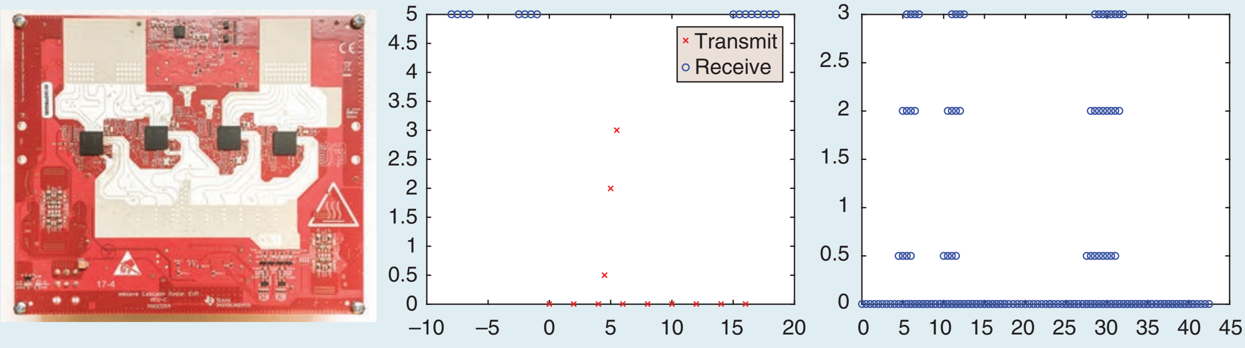
$$\Delta\theta_{\mathrm{EL}} = 2\arcsin\left(\frac{1.4\lambda}{\pi D_y}\right) \approx 17°. \tag{12}$$

In the second example [94], an imaging radar testbed with $M_t = 24$, $M_r = 24$ antennas using a TDM scheme (see Figure 13) is presented. The virtual array apertures have been doubled in both the azimuth and elevation directions. The missing elements in the vertical direction in the middle have been interpolated [84]. After filling the holes and removing the redundant elements, the virtual array is a uniform rectangular array with $25 \times 23 = 575$ elements. Wider FoVs of $[-25°, 25°]$ in both the azimuth and elevation were considered in [94]. However, for the forward-looking LRR sensors, the typical azimuth and elevation FoVs are $[-15°, 15°]$ and $[-5°, 5°]$, respectively [5]. Consequently, the array interelement spacing in horizontal and vertical directions can be set to $d_x = 1.93\lambda$ and $d_y = 5.73\lambda$, respectively. Since the interelement spacing is larger than one half wavelength, there are grating lobes in both the azimuth and elevation. However, the grating lobes out of the FoVs can be suppressed through the antenna element design. If the carrier frequency is $f_c = 77$ GHz, the real size of the physical 2D antenna array is roughly 10 × 25 cm. Then, the 3-dB beamwidth of the azimuth and elevation beampattern is [24]:

$$\Delta\theta_{\mathrm{AZ}} = 2\arcsin\left(\frac{1.4\lambda}{\pi 24 d_x}\right) \approx 1.1°, \tag{13}$$

$$\Delta\theta_{\mathrm{EL}} = 2\arcsin\left(\frac{1.4\lambda}{\pi 22 d_y}\right) \approx 0.4°. \tag{14}$$

It is worth noting that the mutual coupling between array elements is reduced significantly if the interelement spacing is larger than one half wavelength, which reduces the burden



(a) (b) (c)

FIGURE 12. (a) The Texas Instruments AWR1243P imaging-radar board [92]. Four AWR1243P radar transceivers are cascaded together, providing (b) 12 transmit and 16 receive antennas, enabling (c) the synthesis of 192 virtual array elements.

for the array calibration [94]. If we want to keep the azimuth FoV unchanged and increase the azimuth angular resolution to roughly 0.4°, each horizontal receive array needs to add 21 more antennas; i.e., the dimension of the new physical antenna array is 66 × 24 with size of approximately 25 × 25 cm. The angular resolution of 0.4° in both the azimuth and elevation directions provided by imaging radar is similar to the azimuth and elevation resolution of Velodyne's lidar products, such as the HDL-64E, with a spin rate of 20 Hz [95].

Design challenges of imaging radar

Achieving waveform orthogonality in imaging radars using FMCW with a large number of transmit antennas is quite challenging. One strategy could be to divide the transmit antennas into several subgroups. In each subgroup, the transmit antennas would transmit simultaneously with slow time phase coding (DDM), while antennas of different subgroups would be scheduled to transmit in different time slots (TDM).

Clock distribution among multiple cascaded transceivers is also challenging. For FMCW mixer operation, an LO is shared among the master and slaves, and the LO routing from the master to all the slaves in the circuit should be matched. Also, the additional ADC sampling and data transmission among different transceivers needs to be synchronized. It is desirable to develop an automotive radar transceiver that can incorporate a large number of transmit and receive antennas. For example, Uhnder has developed a radar system-on-chip (SoC) that has 12 transmit and 16 receive antennas, enabling the synthesis of 192 virtual array elements [96]. Thus, the four current automotive radar transceivers in the cascaded imaging radar shown in Figure 12 can be replaced with a single SoC radar chip. The radar on chip developed by Vayyar has 48 transceivers at 76–81 GHz, which can provide synthesis across 2,000 virtual-array elements [97].

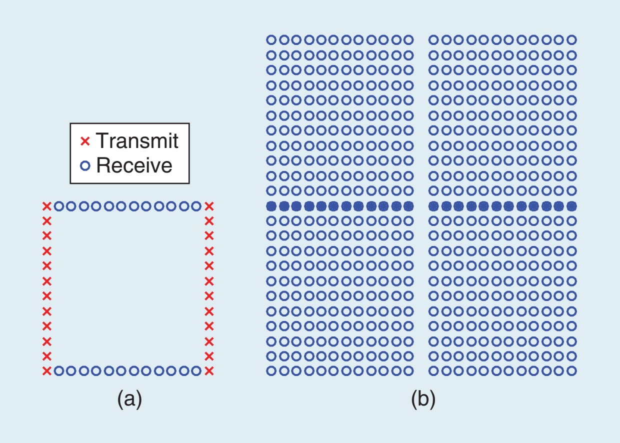


FIGURE 13. An example of imaging radar using MIMO radar technology with $M_t = 24$, $M_r = 24$. (a) The physical antenna configuration. (b) The virtual antenna array.

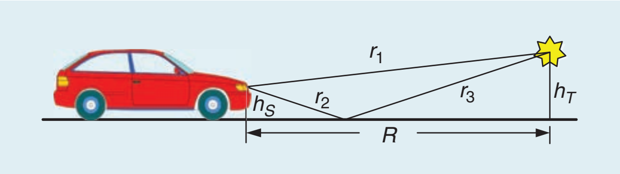


FIGURE 14. The direct path and the vertical multipath [98].

Challenges in automotive MIMO radar

In this section, we discuss signal processing challenges that cannot be addressed by the straightforward application of existing ideas and that could inspire new research.

Angle finding in the presence of multipath reflections

Automotive radar runs in multipath scenarios [98]. In general, radio propagation in the presence of multipath occurs along four possible routes, i.e., direct/direct, direct/indirect, indirect/direct, and indirect/indirect routes. Figure 14 shows a vertical multipath scenario where the height of the radar and target are h_S and h_T, respectively. The length of the direct/direct path is $d_1 = 2r_1$; the length of the direct/indirect or indirect/direct path is $d_2 = d_3 = r_1 + r_2 + r_3$; and the length of the indirect/indirect path is $d_4 = 2(r_2 + r_3)$. The received signal, having gone through the four paths, can be written as

$$y_r = \sum_{i=1}^{4} \beta_i e^{j\frac{2\pi}{\lambda}d_i}, \tag{15}$$

where the amplitude term β_i is a function of the antenna gain, path loss, road-reflection coefficient, and target RCS. The signal from the indirect paths and that from the direct path may arrive out of phase and thus add up destructively. As a result, the power of the received signal would fluctuate with distance [98], and thus angle finding at SNR nulls would be unstable.

MIMO radar with colocated transmit and receive antennas, also referred to as *monostatic MIMO radar*, is based on the assumption that the DoD and DoA are equal. However, in the presence of multipath, that assumption does not hold, and the system becomes bistatic [99], [100]; i.e., the transmit and receive antennas view the target from different aspect angles. Figure 15 shows a vehicle moving parallel to the guardrail, with an SRR sensor mounted at its front left corner. The length of the direct path of the radar signal is $d_{r_1} = 2r_1$, corresponding to $\theta_t = \theta_r = \theta_1$. There are also multipath reflections due to the guardrail. The range of the first multipath reflection is $d_{r_2} = 2(r_1 + r_2 + r_3)$, corresponding to $\theta_t = \theta_1$, $\theta_r = \theta_2$ or $\theta_t = \theta_2$, $\theta_r = \theta_1$. The range of the second multipath reflection is $d_{r_3} = 2(r_2 + r_3)$, corresponding to $\theta_t = \theta_r = \theta_2$. Compared to the direct path, multipath reflections result in longer-range and smaller Doppler. For the first type of multipath, the range and Doppler bin is the same as mirror-image target detection. However, as $\theta_t \neq \theta_r$, it turns out that the phase of each virtual array element is corrupted. In other words, the monostatic MIMO radar assumption does not hold, which results in a "ghost" target whose direction is different from the mirror target.

To solve this issue, some ideas have been proposed in [99]–[101]. For example, joint estimation of the DoD and DoA is proposed in [99]; however, by ignoring the structure of the transmit array, that method is not able to enjoy the benefit of the synthesized virtual array. Polarimetric features are exploited in [100] to separate objects in a multipath scenario. However, the approach in [100] can separate only certain real target cases from their mirror targets; for example, when the real target is known and the polarization state change of the multipaths can be recognized. The method in [100] does not work when the ghost target direction is different than that of the mirror target. Doppler information can be exploited to detect moving vehicles in urban areas under multipath [101]. However, Doppler information is not always available when both the objects and host vehicles are stationary. In general, there is a need for more research addressing the ghost target issue in MIMO radar due to multipath.

Waveform orthogonality in automotive MIMO radar

As stated in the "Introduction of Automotive Radar With MIMO Radar Technology" section, different strategies, such as TDM, DDM, and FDM, can be adopted in automotive FMCW radars to achieve waveform orthogonality. However, several challenges associated with each strategy need to be addressed. For example, in the TDM scheme, the scheduling delay between transmit antennas may introduce phase error for a moving target, which needs to be compensated for; otherwise, the synthesized array beampattern will be distorted. Further, the maximum unambiguous detectable velocity under TDM is reduced by a factor of M_t. In the DDM scheme, the Doppler sidelobes are high due to the residual of the phase coding. As a result, targets with small RCSs, e.g., pedestrians, that are close to the target with strong reflections, e.g., trucks, might be masked. The search time for phase codes using stochastic algorithms increases exponentially as the code length increases. Computationally efficient algorithms are needed to address this problem. In the FDM scheme, although a randomization of the frequency shift among transmit antennas could reduce the range-angle coupling, a large number of transmit antennas would be needed for the improvement to be notable [41].

Recently, PMCW has been proposed for achieving orthogonality [102]–[105]. Each antenna transmits a sequence of phase-coded pulses. Let $\mathbf{x}_m = [x_m(1), \ldots, x_m(N_p)]^T$ be the complex unimodular code sequence of the mth transmit antenna, where $x_m(n) = e^{j\phi_m(n)}$ is the nth code of $\mathbf{x}_m$ and N_p is the code length. Here, the phase $\phi_m(n)$ can be chosen arbitrarily in $[-\pi, \pi]$. The duration of a single code sequence is $T_p = N_p T_c$, with T_c being the duration of a subpulse. In practice, binary code sequences have been widely used due to their simplicity. The bandwidth of PMCW is $B = 1/T_c$. The time–bandwidth product of a code sequence is $BT_p = N_p$. Since the pulses are transmitted continuously, the code sequences should have good periodic autocorrelation and cross-correlation properties [102]. The periodic cross correlation of two code sequences $\mathbf{x}_m$ and $\mathbf{x}_l$ at lag k is defined as

$$r^{\mathcal{P}}_{ml}(k) = \sum_{n=1}^{N} x_m(n)\, x_l^*((n+k)\ \mathrm{mod}(N_p)), \qquad (16)$$

when $m = l$, $r^{\mathcal{P}}_{ml}(k)$ becomes the periodic autocorrelation function of $\mathbf{x}_m$. Good correlation properties require that the values of the periodic autocorrelation at nonzero lag and that the values of the cross correlation at any lag be low. The Welch lower bound on the cross correlation between any pair of binary sequences with a period of N_p in a set of M_t sequences equals [106]

$$r^{\mathcal{P}}_{ml}(k) \geq N_p \sqrt{\frac{M_t - 1}{M_t N_p - 1}} \approx \sqrt{N_p}. \qquad (17)$$

Good periodic cross-correlation properties help achieve waveform orthogonality, while good periodic autocorrelation properties make it easier to use matched filters to extract signals reflected from the range bin of interest and suppress signals reflected from other range bins.

As compared to FMCW, PMCW radar has several advantages. PMCW radar is better suited for achieving waveform orthogonality in imaging radars with a large of number of transmit antennas. PMCW radar can take advantage of existing sequences with good autocorrelation and cross-correlation properties that were previously developed for code-division multiple accessing (CDMA) communications, such as Gold, Kasami, and m-sequences [107]–[109]. Further, in PMCW radar, each automotive radar sensor can have a unique digital sequence, which may help reduce the automotive radar mutual interference. As a bonus, PMCW radar also provides certain communication capability [110] and thus can be explored as a dual-functional radar communication system [111].

However, PMCW radar has many implementation challenges. First, the sampling rate of the ADC should satisfy

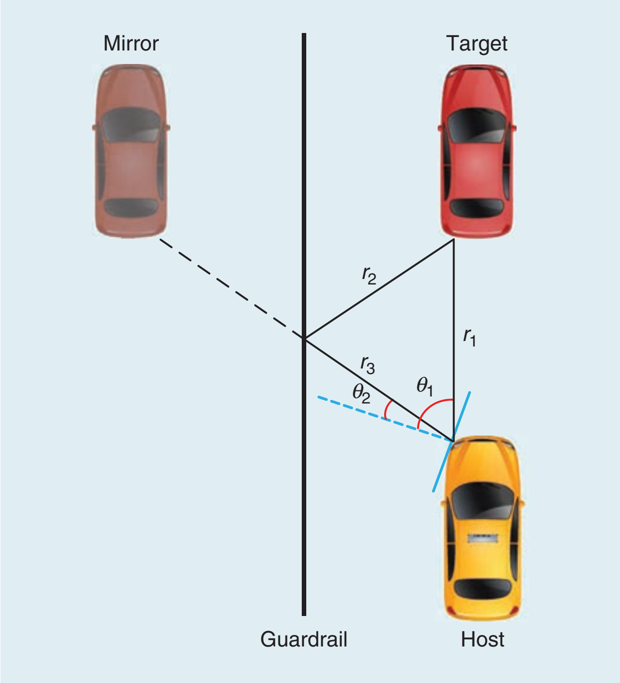


FIGURE 15. A typical multipath reflection scenario along a guardrail for an SRR sensor mounted at the vehicle's front left corner.

the Nyquist rule; i.e., $f_s \geq 2B = 2/T_c$. The high bandwidth required for a high range resolution necessitates a high-speed ADC and high-speed processing hardware. In practice, keeping the resolution of the ADC as low as possible is required [112]. Second, according to the Welch bound of (17), the cross-correlation lower bound of any pair of binary sequences is of the order of $O(\sqrt{N_p})$, which might not provide a sufficient separation of the transmit waveforms of different antennas. In practice, the autocorrelation and cross correlation of the code sequences are desired to have low sidelobes within a low correlation zone. Furthermore, because there is no mapping relationship between the range and beat signal in PMCW, it would be difficult to use high-pass analog filters to reject or attenuate ultraclose-range return signals, including direct path signals from the transmit antennas, reflections from the radome, and vehicle bumpers. This escalates the dynamic-range challenge, especially when the resolution of the ADC must be kept as low as possible [112].

Mutual-interference mitigation

Automotive radar mutual interference is a challenging issue that needs to be addressed. The use of radar for ADAS and autonomous driving is climbing rapidly. As the number of vehicles equipped with automotive radar is escalating, with each vehicle deploying up to 10 automotive radar units, the probability of mutual interference between automotive radar units increases. Figure 16 shows an example of an automotive radar interference scenario where two front-looking automotive radar sensors from two stopped cars illuminate each other. If the radar sensors operate at the same frequency band and transmit at the same time, they will interfere with each other. Without interference mitigation, automotive radar sensors suffer from performance degradation [113]. Therefore, it is desired to optimize both the radar transmit- and receive side operations to mitigate the interference.

A typical mutual interference scenario for two automotive FMCW radars is given in Figure 17. The blue line indicates the chirp of an FMCW radar mounted on the host vehicle, with a faster sweep rate S_1, referred to as the *victim radar*, while the red line denotes the chirp of an FMCW radar mounted on another vehicle, with a slower sweep rate S_2, referred to as the *interference radar*. The interference radar is set to illuminate the FoV of the victim radar directly. At the receiver of the victim radar, the duration of the corrupted samples introduced by interference in one pulse is $T_{\text{int}} = 2\left(\left| f_b^{\max}/(S_1 - S_2)\right|\right)$, where $f_b^{\max}$ is the highest cutoff frequency of the antialiasing BPF [114]. After mixing with the transmitted chirp, in addition to the beat frequency corresponding to real targets, the corrupted samples contain frequencies spanning the whole interval of

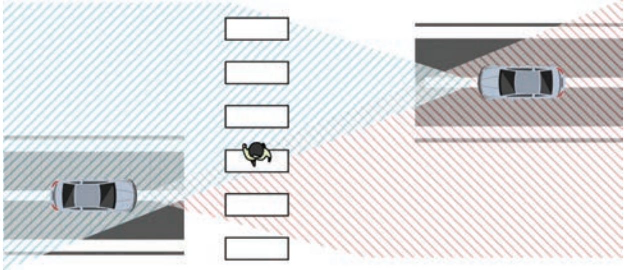


FIGURE 16. An illustrative automotive radar interference scenario, where two front-looking automotive radar sensors illuminate each other.

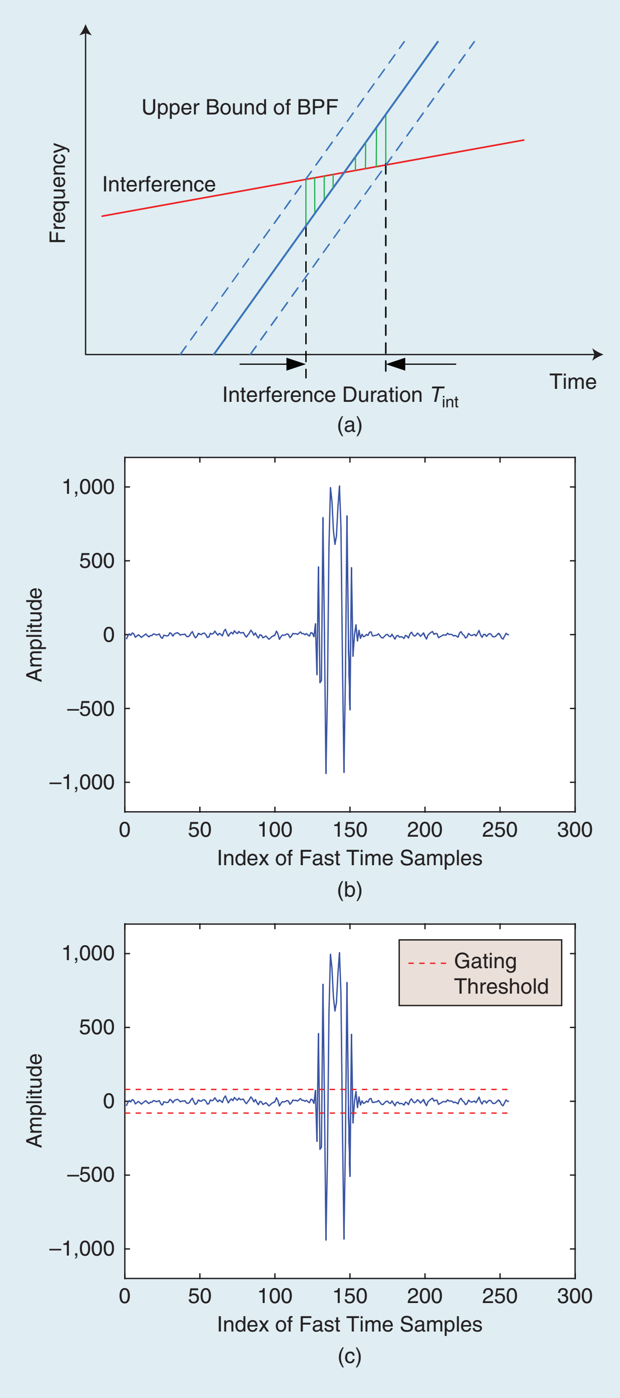


FIGURE 17. (a) Automotive radar interference exists when the victim and interference radars have different FMCW sweep rates. (b) After down-conversion and passing through the BPF, the sequence sampled from one chirp is produced. (c) Since corrupted samples introduced by interference have a much larger amplitude than the samples of the target echo, they can be clipped or gated.

the BPF (see Figure 17). Further, since the interference signal arrives via a direct path, it is much stronger than the echoes of the target.

Intuitively, this type of corrupted samples introduces extra energy that spans the whole beat frequency band of interest. As a result, after performing range and Doppler FFTs on the receive signal with interference, the noise floor in the range–Doppler spectrum would increase, as shown in Figure 18. As a result, targets with a small RCS might be masked. For example, the pedestrian in Figure 16 might be buried in the noise and may not be detected by automotive radar when the interference is strong.

Machine learning algorithms and deep neural networks have also been applied in automotive radar for target recognition and classification.

The amount of interference energy can be viewed as an integral of corrupted samples through the interference duration interval T_{int}. Therefore, the interference can be greatly suppressed by clipping, or gating, the corrupted samples. In particular, the corrupted samples with an amplitude larger than a threshold close to the amplitude of the target echo plus noise would be clipped (see Figure 17). However, gating does not completely remove the interference because the clipped corrupted samples still contain frequency components that occupy the whole beat frequency band of interest.

Figure 18 presents the simulated range and Doppler spectra of victim radar. Three scenarios have been simulated: target plus noise only, target plus noise plus interference, and target plus noise plus interference with gating. It can be seen that without gating, the target is masked by the increased noise that is due to the interference. Fast time-domain gating helps to recover the target in the range and Doppler domains. It should be noted that this type of interference needs to be mitigated in fast time after ADC sampling, and thus the mitigation algorithms not only need to be effective but also efficient. Gating has a low computational cost, and it can significantly suppress the noise level in the range and Doppler spectra.

Other interference-mitigation approaches include repairing the corrupted samples in the time domain, frequency domain [115], or spatial domain using adaptive beamforming techniques [116]. However, if the FMCW sweep rates of the interference and victim radar are the same, the interference radar will create ghost targets [117], which makes interference mitigation more challenging, since this scenario is difficult to detect.

With the introduction of PMCW radar, the mutual interference between automotive radar sensors can be greatly mitigated in the code domain. The PMCW code sequences used in different automotive radar sensors are typically not time aligned. Therefore, to suppress the mutual interference, the periodic cross correlation of any pair of sequences should be low at all lags. PMCW radar is similar to CDMA in communication systems. In other words, the interference will be a

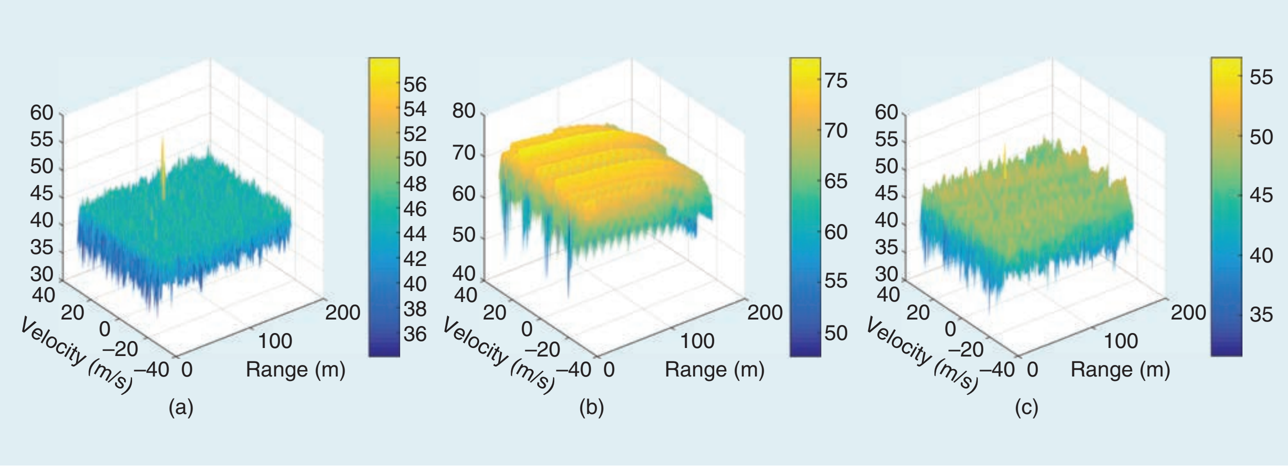


FIGURE 18. The range and Doppler spectra of the victim radar with and without time-domain gating. (a) The target plus noise. (b) The target plus interference and noise, without gating. (c) The target plus interference and noise, with gating.

Table 1. The different DoA estimation algorithms in automotive radar scenarios.

Algorithm	Resolution	Snapshot	Array	Grid-Free	Rank Estimation	Robustness	Complexity
DBF	Low	Single	ULA/SLA	No	No	Strong	Low
MUSIC	High	Multiple	ULA	No	Yes	Medium	High
ESPRIT	High	Multiple	ULA	Yes	Yes	Medium	Medium
OMP	High	Single	ULA/SLA	No	No	Medium	High
IAA	High	Single	ULA/SLA	No	No	Strong	High

wide-band pseudonoise signal. However, the mutual interference mitigation using PMCW highly depends on the periodic cross-correlation properties of the code sequence. Therefore, designing code sequences with good periodic autocorrelation and cross-correlation properties is of great interest. The other research problems include investigating the interference between FMCW and PMCW radars [118].

Efficient, high-resolution angle finding algorithms are needed

A typical duration of a look in automotive radar is roughly 50 ms, corresponding to a detection update rate of 20 Hz [119]. In such a short duration, the current generation of automotive radar for ADAS can report a maximum of 64–200 detections. With high-resolution imaging radar, the number of cells that can be selected for angle finding from the 2D range and Doppler spectrum is approximately 10,000 in a single look for a typical road scenario [15]. To achieve point clouds for autonomous driving, angle finding needs to be performed thousands of times in a single look, which is a great challenge for imaging radar with hundreds of virtual array elements. Computationally efficient, high-resolution angle finding algorithms are highly desirable for real-time implementation in automotive radar.

Achieving waveform orthogonality in imaging radars using FMCW with a large number of transmit antennas is quite challenging.

To reduce the computation complexity, beamspace ESPRIT [52] and unitary ESPRIT [53] algorithms have been proposed. The idea of beamspace ESPRIT is to decompose the original ULA vector into several low-dimensional beamspaces via a transform, such as the FFT. Then, if the beamspace transform matrix has the same shift-invariance structure, angle finding can be carried out via ESPRIT on each beamspace in parallel, with a reduced computational time [52]. The unitary ESPRIT algorithm takes advantage of the unit-magnitude property of the phase factors representing the phase delays between the two subarrays and is formulated in terms of real-valued computations. As a result, it achieves a great reduction of the computational complexity [53].

The computation cost of each IAA iteration is $2NM^2 + NM + M^3$, where M is the number of array snapshots and N is the number of discretized grids. Fast and superfast IAA algorithms have been proposed in [120]–[122], respectively. The fast IAA algorithm exploits the FFT operation as well as the Gohberg–Semencul representation of matrix $\mathbf{R}^{-1}$. As a result, the computation cost of each fast IAA iteration is $M^2 + 12\zeta(2M) + 3\zeta(N)$, where $\zeta(N)$ stands for the computation cost of performing an FFT of size N, i.e., $O(N \log N)$ [121]. The superfast IAA uses a conjugate-gradient algorithm to approximate the matrix $\mathbf{R}^{-1}$, which further reduces the computation cost.

The strengths and limitations of each DoA estimation algorithm discussed in the "Angle Finding in Automotive MIMO Radar" section when they are applied to the automotive radar scenario are summarized in Table 1. For subspace-based, high-resolution DoA estimation methods, such as MUSIC and ESPRIT, the automotive radar array needs to be a ULA, and multiple snapshots are required to estimate the array covariance matrix accurately. However, automotive radar operating in a highly dynamic environment typically relies on a single snapshot. While multiple snapshots can be generated via spatial smoothing or by dividing a chirp into subchirps, the associated cost is a respectively reduced array aperture or reduced SNR. SLAs have been widely used in automotive radar to further reduce the hardware cost. However, it is not straightforward to apply MUSIC or ESPRIT to SLAs-based automotive radar.

On the other hand, DBF and sparsity-based, high-resolution methods, such as the OMP and IAA, apply to SLA-based as well as ULA-based automotive radar and work with a single snapshot. In the DBF method, the number of targets can be estimated by counting the number of peaks in the DoA spectrum. DBF is not sensitive to coherent or correlated signals, which in subspace-based methods [123] need special preprocessing via spatial smoothing. It has been shown that DBF is robust to array-element position errors and has a low computational cost [123]. However, DBF is not a high-resolution method. Also, since the OMP and IAA are iterative schemes, they involve a high computation cost, which limits their applicability in the low-cost embedded DSPs typically used in current-generation automotive radar. Further, the methods of DBF, MUSIC, OMP, and IAA assume targets are on the grid and suffer from errors when the targets arise between grid points. In summary, more research is needed on developing computationally efficient, high-resolution DoA estimation algorithms that are robust to noise and applicable to automotive radar using SLAs with a low PSL under a single snapshot.

Summary

We have reviewed the advantages of MIMO radar technology in increasing the angular resolution of commercial automotive radar while, at the same time, offering a low hardware cost and small package size. In particular, we have shown that MIMO radar technology plays a key role in high-resolution imaging-radar systems for L4 and L5 autonomous driving. Automotive MIMO radar technologies, such as waveform orthogonality with slow time phase coding, time-division multiple access, and nonuniform linear array synthesis with low-peak sidelobes have been discussed. High-resolution angle finding literature for snapshot-limited automotive MIMO radar has been reviewed. In particular, high-resolution angle finding methods, including subspace methods with spatial smoothing, compressive sensing, and the IAA, have been discussed.

In particular, we have reviewed methods of cascading multiple automotive radar transceivers to synthesize hundreds or thousands of virtual array elements in high-resolution imaging radar systems and have discussed related design challenges. We have also addressed the challenges of MIMO radar technology in automotive applications, which would inspire further research for the signal processing community. In addition, we have discussed angle finding issues in the presence of multipath reflections, waveform orthogonality strategies, automotive mutual interference mitigation, and computationally efficient, high-resolution angle finding methods.

Authors

Shunqiao Sun (shunqiao.sun@ua.edu) received his Ph.D. degree in electrical and computer engineering from Rutgers University, New Jersey in 2016. In 2019, he joined the Department of Electrical and Computer Engineering at the University of Alabama, Tuscaloosa, as a tenure-track assistant professor. His research interests lie at the interface between statistical and sparse signal processing with mathematical optimizations, multiple-input, multiple-output (MIMO) radar, automotive radar, remote sensing, machine learning, and connected and autonomous vehicles. He was awarded the 2015–2016 Rutgers University Electrical and Computer Engineering Graduate Program Academic Achievement Award. He was also the winner of the 2016 IEEE Aerospace and Electronic Systems Society Robert T. Hill Best Dissertation Award for his thesis "MIMO Radar With Sparse Sensing." He is a Senior Member of the IEEE.

Athina P. Petropulu (athinap@rutgers.edu) received her B.S. degree from the National Technical University of Athens, Greece, and her M.S. and Ph.D. degrees from Northeastern University, Boston, all in electrical and computer engineering. She is a distinguished professor in the Department of Electrical and Computer Engineering, Rutgers University, New Jersey, which she chaired from 2010 to 2016. Her research interests include statistical signal processing, wireless communications, signal processing in networking, physical-layer security, and radar signal processing. She received the 1995 Presidential Faculty Fellow Award from the National Science Foundation and the White House. In 2005, she received the IEEE Signal Processing Magazine Best Paper Award, and in 2012, the IEEE Signal Processing Society Meritorious Service Award for exemplary service in technical leadership capacities. She is the president-elect of the IEEE Signal Processing Society and a Fellow of the IEEE.

H. Vincent Poor (poor@princeton.edu) received his Ph.D. degree in electrical engineering and computer science from Princeton University, New Jersey in 1977. He is the Michael Henry Strater University Professor of Electrical Engineering at Princeton, where he served as dean of the School of Engineering and Applied Science from 2006 to 2016. His research interests include information theory and signal processing and their applications in wireless networks, energy systems, and related fields. He received the IEEE Signal Processing Society Technical Achievement and Society Awards in 2007 and 2011, respectively; the 2017 IEEE Alexander Graham Bell Medal; the 2019 American Society for Engineering Education Benjamin Garver Lamme Award; a D.Sc. honoris causa from Syracuse University, New York, in 2017; and a D.Eng. honoris causa from the University of Waterloo, Ontario, in 2019. He is a Fellow of the IEEE.

References

[1] W. L. Melvin and J. A. Scheer, *Principles of Modern Radar*, vol. III, *Radar Applications*. Edison, NJ: SciTech, 2014.

[2] G. W. Stimson, H. Griffiths, C. Baker, and D. Adamy, *Stimson's Introduction to Airborne Radar System*, 3rd ed. Edison, NJ: SciTech, 2014.

[3] M. Schneider, "Automotive radar: Status and trends," in *Proc. German Microwave Conf. (GeMiC)*, Ulm, Germany, Apr. 2005, pp. 144–147.

[4] E. Coelingh, A. Eidehall, and M. Bengtsson, "Collision warning with full auto brake and pedestrian detection: A practical example of automatic emergency braking," in *Proc. 13th IEEE Conf. Intelligent Transportation Systems (ITSC)*, Funchal, Portugal, Sept. 2010, pp. 155–160. doi: 10.1109/ITSC.2010.5625077.

[5] S. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, 2017. doi: 10.1109/MSP.2016.2628914.

[6] F. Engels, P. Heidenreich, A. M. Zoubir, F. Jondral, and M. Wintermantel, "Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 36–46, 2017. doi: 10.1109/MSP.2016.2637700.

[7] J. Steinbaeck, C. Steger, G. Holweg, and N. Drumi, "Next generation radar sensors in automotive sensor fusion systems," in *Proc. Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, Bonn, Germany, Oct. 2017, pp. 1–6. doi: 10.1109/SDF.2017.8126389.

[8] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, 2007. doi: 10.1109/MSP.2007.904812.

[9] J. Li and P. Stoica, Eds., *MIMO Radar Signal Processing*. Hoboken, NJ: Wiley, 2009.

[10] J. Bergin and J. R. Guerci, *MIMO Radar: Theory and Application*. Boston: Artech House, 2018.

[11] I. Bilik, O. Bialer, S. Villeval, H. Sharifi, K. Kona, M. Pan, D. Persechini, M. Musni et al., "Automotive MIMO radar for urban environments," in *Proc. IEEE Radar Conf.*, Philadelphia, May 2016, pp. 1–6. doi: 10.1109/RADAR.2016.7485215.

[12] S. Alland and J. Searcy, "Radar system and method of digital beamforming," U.S. Patent 2009/0085800, Apr. 2, 2009.

[13] M. Wintermantel, "Radar system with improved angle formation," U.S. Patent 2011/0074621, Mar. 31, 2011.

[14] M. Schoor, G. Kuehnle, K. Rambach, and B. Loesch, "Method for operating a MIMO radar," U.S. Patent 2014/0347211, Nov. 27, 2014.

[15] F. Meinl, M. Stolz, M. Kunert, and H. Blume, "An experimental high performance radar system for highly automated driving," in *Proc. Int. Conf. Microwaves Intelligent Mobility (ICMIM)*, Nagoya, Japan, Mar. 2017, pp. 71–74. doi: 10.1109/ICMIM.2017.7918859.

[16] S. Alland, J. P. Bordes, C. Davis, and M. Ali, "Virtual radar configuration for 2D array," U.S. Patent 9 869 762, Jan. 16, 2018.

[17] M. Amin, Y. Zhang, F. Ahmad, and K. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 71–80, 2016. doi: 10.1109/MSP.2015.2502784.

[18] M. Amin, Ed., *Radar for Indoor Monitoring: Detection, Classification, and Assessment*. Boca Raton, FL: CRC, 2017.

[19] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018. doi: 10.1109/TAES.2017.2740098.

[20] S. Gurbuz and M. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, 2019. doi: 10.1109/MSP.2018.2890128.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. doi: 10.1038/nature14539.

[22] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 845–860, 2012. doi: 10.1109/TMTT.2011.2178427.

[23] D. Zoeke and A. Ziroff, "Phase migration effects in moving target localization using switched MIMO arrays," in *Proc. European Radar Conf. (EuRAD)*, Paris, France, Sept. 2015, pp. 85–88. doi: 10.1109/EuRAD.2015.7346243.

[24] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, McGraw-Hill, 2014.

[25] D. Brodeski, I. Bilik, and R. Giryes, "Deep radar detector," in *Proc. IEEE Radar Conf.*, Boston, Apr. 2019, pp. 1–6.

[26] S. Ayhan, S. Scherr, A. Bhutani, B. Fischbach, M. Pauli, and T. Zwick, "Impact of frequency ramp nonlinearity, phase noise, and SNR on FMCW radar accuracy," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 10, pp. 3290–3301, 2016. doi: 10.1109/TMTT.2016.2599165.

[27] K. Ramasubramanian and B. Ginsburg, "AWR1243 sensor: Highly integrated 76–81-GHz radar front-end for emerging ADAS applications," Texas Instruments

Inc., Dallas, White Paper, 2017. [Online]. Available: http://www.ti.com/lit/wp/spyy003/spyy003.pdf

[28] P. Wang, D. Millar, K. Parsons, R. Ma, and P. V. Orlik, "Range accuracy analysis for FMCW systems with source nonlinearity," in *Proc. Int. Conf. Microwaves for Intelligent Mobility (ICMIM)*, Detroit, Apr. 2019, pp. 1–5. doi: 10.1109/ICMIM.2019.8726679.

[29] K. Lin, Y. Wang, C. Pao, and Y. Shih, "A Ka-band FMCW radar front-end with adaptive leakage cancellation," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4041–4048, 2006. doi: 10.1109/TMTT.2006.885882.

[30] A. G. Stove, "Linear FMCW radar techniques," *IEE Proc. F, Radar Signal Process.*, vol. 139, no. 5, pp. 343–350, 1992. doi: 10.1049/ip-f-2.1992.0048.

[31] S. Jung, S. Kim, W. Choi, H. Kim, H. Kim, and Y. Eo, "High dynamic range Ku-band CMOS transceiver IC for FMCW radar application," in *Proc. IEEE MTT-S Int. Microwave Symp. (IMS)*, Honolulu, HI, June 2017, pp. 1415–1417. doi: 10.1109/MWSYM.2017.8058883.

[32] B. B. Adela, "Antennas for silicon-based mm-wave FMCW radars: Antenna integration and MIMO system design," Ph.D. dissertation, Technische Univ. Eindhoven (TU/e), Eindhoven, The Netherlands, 2019.

[33] K.-W. Gurgel and T. Schlick, "Remarks on signal processing in HF radars using FMCW modulation," in *Proc. Int. Radar Symp. (IRS)*, Hamburg, Germany, Sept. 9–11, 2009, pp. 1–5.

[34] F. Jansen, A. Filippi, and Z. Zivkoic, "MIMO radar system," European Patent EP3021132A1, May 18, 2016.

[35] A. Duly, D. Love, and J. Krogmeier, "Time-division beamforming for MIMO radar waveform design," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1210–1223, 2013. doi: 10.1109/TAES.2013.6494408.

[36] K. Rambach and B. Yang, "Colocated MIMO radar: Cramer–Rao bound and optimal time division multiplexing for DOA estimation of moving targets," in *Proc. IEEE 38th Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 4006–4010. doi: 10.1109/ICASSP.2013.6638411.

[37] D. Bleh, M. Rösch, M. Kuri, A. Dyck, A. Tessmann, A. Leuther, S. Wagner, B. Weismann-Thaden et al., "*W*-band time-domain multiplexing FMCW MIMO radar for far-field 3-D imaging," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3474–3484, 2017. doi: 10.1109/TMTT.2017.2661742.

[38] J. Bechter, F. Roos, and C. Waldschmidt, "Compensation of motion-induced phase errors in TDM MIMO radars," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 12, pp. 1164–1166, 2017. doi: 10.1109/LMWC.2017.2751301.

[39] F. Roos, J. Bechter, N. Appenrodt, J. Dickmann, and C. Waldschmidt, "Enhancement of Doppler unambiguity for chirp-sequence modulated TDM-MIMO radars," in *Proc. Int. Conf. Microwaves Intelligent Mobility (ICMIM)*, Munich, Apr. 2018, pp. 1–4. doi: 10.1109/ICMIM.2018.8443352.

[40] V. F. Mecca, D. Ramakrishnan, and J. L. Krolik, "MIMO radar space-time adaptive processing for multipath clutter mitigation," in *Proc. IEEE Workshop on Sensor Array Multichannel Processing (SAM)*, Waltham, MA, July 2006. doi: 10.1109/SAM.2006.1706131.

[41] H. Sun, F. Brigui, and M. Lesturgie, "Analysis and comparison of MIMO radar waveforms," in *Proc. Int. Radar Conf.*, Lille, France, Oct. 2014, pp. 1–6. doi: 10.1109/RADAR.2014.7060251.

[42] N. Madsen and S. Cao, "Slow-time waveform design for MIMO GMTI radar using CAZAC sequences," in *Proc. IEEE Radar Conf.*, Oklahoma City, OK, Apr. 2018, pp. 1456–1460. doi: 10.1109/RADAR.2018.8378779.

[43] J. Benedetto, I. Konstantinidis, and M. Rangaswamy, "Phase-coded waveforms and their design," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 22–31, 2009. doi: 10.1109/MSP.2008.930416.

[44] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, 1972. doi: 10.1109/TIT.1972.1054840.

[45] R. Feger, H. Haderer, and A. Stelzer, "Optimization of codes and weighting functions for binary phase-coded FMCW MIMO radars," in *Proc. Int. Conf. Microwaves Intelligent Mobility (ICMIM)*, San Diego, CA, May 2016, pp. 1–4. doi: 10.1109/ICMIM.2016.7533916.

[46] R. Feger, C. Pfeffer, and A. Stelzer, "A frequency-division MIMO FMCW radar system based on Delta–Sigma modulated transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 12, pp. 3572–3581, 2014. doi: 10.1109/TMTT.2014.2364220.

[47] P. Hacker and B. Yang, "Single snapshot DOA estimation," *Adv. Radio Sci.*, vol. 8, pp. 251–256, Oct. 2010. doi: 10.5194/ars-8-251-2010.

[48] P. Barton, "Digital beam forming for radar," *IEE Proc. F, Commun., Radar Signal Process.*, vol. 127, no. 4, pp. 266–277, 1980. doi: 10.1049/ip-f-1.1980.0041.

[49] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita, and H. Fujinami, "Electronically scanned millimeter-wave radar for precrash safety and adaptive cruise control system," in *Proc. IEEE Intelligent Vehicles Symp.*, Columbus, OH, June 2003, pp. 304–309. doi: 10.1109/IVS.2003.1212927.

[50] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986. doi: 10.1109/TAP.1986.1143830.

[51] R. Roy and T. Kailath, "ESPRIT: Estimation of signal parameters via rotation invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989. doi: 10.1109/29.32276.

[52] G. Xu, S. D. Silverstein, R. H. Roy, and T. Kailath, "Beamspace ESPRIT," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 349–356, 1994. doi: 10.1109/78.275607.

[53] M. Haardt and J. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1232–1242, 1995. doi: 10.1109/78.382406.

[54] M. D. Zoltowski, M. Haardt, and C. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, 1996. doi: 10.1109/78.485927.

[55] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar: Algorithms and performance," in *Proc. 43th Annu. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2009, pp. 464–468. doi: 10.1109/ACSSC.2009.5469862.

[56] Y. Yu, A. P. Petropulu, and H. V. Poor, "MIMO radar using compressive sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, 2010. doi: 10.1109/JSTSP.2009.2038973.

[57] Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing-based MIMO radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5338–5352, 2011. doi: 10.1109/TSP.2011.2162328.

[58] Y. Yu, A. P. Petropulu, and H. V. Poor, "CSSF MIMO radar: Compressive-sensing and step-frequency based MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1490–1504, 2012. doi: 10.1109/TAES.2012.6178074.

[59] Y. Yu, S. Sun, R. N. Madan, and A. P. Petropulu, "Power allocation and waveform design for the compressive sensing based MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 898–909, 2014. doi: 10.1109/TAES. 2014.130088.

[60] T. Strohmer and B. Friedlander, "Analysis of sparse MIMO radar," *Appl. Comp. Harm. Anal.*, vol. 37, no. 3, pp. 361–388, 2014. doi: 10.1016/j.acha.2013.12.005.

[61] S. Sun, "MIMO radars with sparse sensing," Ph.D. dissertation, Rutgers Univ., New Brunswick, NJ, 2016.

[62] S. Sun, W. U. Bajwa, and A. P. Petropulu, "MIMO-MC radar: A MIMO radar approach based on matrix completion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1839–1852, 2015. doi: 10.1109/TAES.2015.140452.

[63] S. Sun and A. P. Petropulu, "Waveform design for MIMO radars with matrix completion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1400–1411, 2015. doi: 10.1109/JSTSP.2015.2469641.

[64] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, 2010. doi: 10.1109/TAES.2010.5417172.

[65] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 5–20, 2010. doi: 10.1109/JSTSP.2009.2038964.

[66] T. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 506–811, 1985. doi: 10.1109/TASSP.1985.1164649.

[67] L. Carin, D. Liu, and B. Guo, "Coherence, compressive sensing, and random sensor arrays," *IEEE Antennas Propag. Mag.*, vol. 53, no. 4, pp. 28–39, 2011. doi: 10.1109/MAP.2011.6097283.

[68] C. Schmid, R. Feger, C. Wagner, and A. Stelzer, "Design of a linear non-uniform antenna array for a 77-GHz MIMO FMCW radar," in *Proc. IEEE MTT-S Int. Microwave Workshop on Wireless Sensing, Local Positioning, RFID*, Cavtat, Croatia, Sept. 2009, pp. 1–4. doi: 10.1109/IMWS2.2009.5307896.

[69] C. A. Alcalde, "Radar system and method for virtual antenna signals," U.S. Patent 9 664 775 B2, May 30, 2017.

[70] Z. Li and C. A. Alcalde, "Angle finding for a detector having a paired staggered array," U.S. Patent 2018/0267555 A1, Sept. 20, 2018.

[71] S. Alland, "MIMO antenna with improved grating lobe characteristics," U.S. Patent 2015/0253419 A1, Sept. 10, 2015.

[72] J. Searcy and S. Alland, "MIMO antenna with elevation detection," U.S. Patent 9 541 639 B2, Jan. 10, 2017.

[73] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974. doi: 10.1109/TAC.1974.1100705.

[74] Y. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, 1998. doi: 10.1109/18.720554.

[75] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inv. Problems*, vol. 23, no. 3, pp. 969–985, 2007. doi: 10.1088/0266-5611/23/3/008.

[76] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when *p* is much larger than *n*," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007. doi: 10.1214/009053606000001523.

[77] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, 2011. doi: 10.1109/TIT.2011.2146090.

[78] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013. doi: 10.1109/TIT.2013.2277451.

[79] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011. doi: 10.1109/TSP.2011.2112650.

[80] Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. Antennas Propag.*, vol. 12, no. 3, pp. 257–268, 1964. doi: 10.1109/TAP.1964.1138220.

[81] R. Feger, C. Wagner, S. Schuster, S. Scheiblhofer, H. Jager, and A. Stelzer, "A 77-GHz FMCW MIMO radar based on an SiGe single chip transceiver," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1020–1035, 2009. doi: 10.1109/TMTT.2009.2017254.

[82] N. Jin and Y. Rahmat-Samii, "Advances in particle swarm optimization for antenna designs: Real-number, binary, single-objective and multiobjective implementations," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 556–567, 2007. doi: 10.1109/TAP.2007.891552.

[83] M. Gonalez-Huici, D. Mateos-Nunez, C. Greiff, and R. Simoni, "Constrained optimal design of automotive radar arrays using the Weiss–Weinstein bound," in *Proc. Int. Conf. Microwaves Intelligent Mobility (ICMIM)*, Munich, Apr. 2018, pp. 1–4. doi: 10.1109/ICMIM.2018.8443558.

[84] T. Spreng, S. Yuan, V. Valenta, H. Schumacher, U. Siart, and V. Ziegler, "Wideband 120 GHz to 140 GHz MIMO radar: System design and imaging results," in *Proc. European Microwave Conf. (EuMC)*, Paris, France, Sept. 2015, pp. 430–433. doi: 10.1109/EuMC.2015.7345792.

[85] M. Rossi, A. Haimovich, and Y. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, 2014. doi: 10.1109/TSP.2013.2289875.

[86] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 2017, pp. 652–660.

[87] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Conf. Neural Information Processing Systems (NIPS)*, Long Beach, CA, Dec. 2017, pp. 1–10.

[88] S. Brisken, F. Ruf, and F. Hohne, "Recent evolution of automotive imaging radar and its information content," *IET Radar, Sonar Navigat.*, vol. 12, no. 10, pp. 1078–1081, 2018. doi: 10.1049/iet-rsn.2018.0026.

[89] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, 2D car detection in radar data with PointNets. July 2019. [Online]. Available: arXiv:1904.08414

[90] I. Bilik, S. Villeval, D. Brodeski, H. Ringel, O. Longman, P. Goswami, C. Y. B. Kumar, S. Rao et al., "Automotive multi-mode cascaded radar data processing embedded system," in *Proc. IEEE Radar Conf.*, Oklahoma City, OK, Apr. 2018, pp. 0372–0376. doi: 10.1109/RADAR.2018.8378587.

[91] A. Och, C. Pfeffer, J. Schrattenecker, S. Schuster, and R. Weigel, "A scalable 77 GHz massive MIMO FMCW radar by cascading fully-integrated transceivers," in *Proc. Asia-Pacific Microwave Conf. (APMC)*, Kyoto, Japan, Nov. 6–9, 2018, pp. 1235–1237. doi: 10.23919/APMC.2018.8617548.

[92] "Design guide: TIDEP-01012—Imaging radar using cascaded mmWave sensor reference design (REV. A)," Texas Instruments Inc., Dallas, June 2019. [Online]. Available: http://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf

[93] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, 1968. doi: 10.1109/TAP.1968.1139138.

[94] A. Ganis, E. M. Navarro, B. Schoenlinner, U. Prechtel, A. Meusling, C. Heller, T. Spreng, J. Mietzner et al., "A portable 3-D imaging FMCW MIMO radar demonstrator with a 24×24 antenna array for medium-range applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 298–312, 2018. doi: 10.1109/TGRS.2017.2746739.

[95] "HDL-64E S3 high definition real-time 3D LiDAR," Velodyne LiDAR, San Jose, CA, Datasheet, 2018. [Online]. Available: https://velodynelidar.com/products/hdl-64e/

[96] V. Giannini, M. Goldenberg, A. Eshraghi, J. Maligeorgos, L. Lim, R. Lobo, D. Welland, C.-K. Chow et al., "A 192-virtual-receiver 77/79GHz GMSK code-domain MIMO radar system-on-chip," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, Feb. 17–21, 2019, pp. 164–166. doi: 10.1109/ISSCC.2019.8662386.

[97] Vayyar. Accessed on: Jan. 2020. [Online]. Available: https://vayyar.com/

[98] A. Ioffe, W. Doerr, H. Yan, D. H. Vu, and A. H. Arage, "RCS characteristics of street curbs and the applications in automotive radar classification," in *Proc. European Radar Conf. (EuRAD)*, London, Oct. 2016, pp. 241–244.

[99] F. Engels, M. Wintermantel, and P. Heidenreich, "Automotive MIMO radar angle estimation in the presence of multipath," in *Proc. European Radar Conf. (EuRAD)*, Nuremberg, Germany, Oct. 2017, pp. 82–85. doi: 10.23919/EURAD.2017.8249152.

[100] T. Visentin, J. Hasch, and T. Zwick, "Analysis of multipath and DOA detection using a fully polarimetric automotive radar," in *Proc. European Radar Conf. (EuRAD)*, Nuremberg, Germany, Oct. 2017, pp. 45–48. doi: 10.23919/EURAD.2017.8249143.

[101] L. B. Fertig, J. M. Baden, and J. R. Guerci, "Knowledge-aided processing for multipath exploitation radar (MER)," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 10, pp. 24–36, 2017. doi: 10.1109/MAES.2017.160035.

[102] W. V. Thillo, P. Gioffré, V. Giannini, D. Guermandi, S. Brebels, and A. Bourdoux, "Almost perfect auto-correlation sequences for binary phase-modulated continuous wave radar," in *Proc. European Radar Conf. (EuRAD)*, Nuremberg, Germany, Oct. 2013, pp. 491–494.

[103] V. Giannini, D. Guermandi, Q. Shi, A. Medra, W. V. Thillo, A. Bourdoux, and P. Wambacq, "A 79 GHz phase-modulated 4 GHz-BW CW radar transmitter in 28 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 49, no. 12, pp. 2925–2937, 2014. doi: 10.1109/JSSC.2014.2355819.

[104] A. Bourdoux, U. Ahmad, D. Guermandi, S. Brebels, A. Dewilde, and W. V. Thillo, "PMCW waveform and MIMO technique for a 79 GHz CMOS automotive radar," in *Proc. IEEE Radar Conf.*, Philadelphia, May 2016, pp. 1–5. doi: 10.1109/RADAR.2016.7485114.

[105] D. Guermandi, Q. Shi, A. Dewilde, V. Derudder, U. Ahmad, A. Spagnolo, A. Bourdoux, P. Wambacq et al., "A 79-GHz 2×2 MIMO PMCW radar SoC in 28-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 52, no. 10, pp. 2613–2626, 2017. doi: 10.1109/JSSC.2017.2723499.

[106] L. Welch, "Lower bounds on the maximum cross correlation of signals (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, 1974. doi: 10.1109/TIT.1974.1055219.

[107] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 13, no. 4, pp. 619–621, 1967. doi: 10.1109/TIT.1967.1054048.

[108] T. Kasami, "Weight distribution formula for some class of cyclic codes," Coordinated Sci. Lab., Univ. of Illinois, Urbana, Rep. R-285, 1966, pp. 1–32.

[109] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *IEEE Commun. Mag.*, vol. 36, no. 9, pp. 48–54, 1998. doi: 10.1109/35.714616.

[110] G. Hueber and A. M. Niknejad, Eds., *Millimeter-Wave Circuits for 5G and Radar*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[111] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, to be published. doi: 10.1109/TCOMM.2020.2973976.

[112] W. V. Thillo, V. Giannini, D. Guermandi, S. Brebels, and A. Bourdoux, "Impact of ADC clipping and quantization on phase-modulated 79 GHz CMOS radar," in *Proc. European Radar Conf. (EuRAD)*, Rome, Italy, Oct. 2014, pp. 285–288. doi: 10.1109/EuRAD.2014.6991263.

[113] S. Alland, W. Stark, M. Ali, and A. Hedge, "Interference in automotive radar systems: Characteristics, mitigation techniques, and future research," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 45–59, 2019. doi: 10.1109/MSP.2019.2908214.

[114] F. Uysal and S. Sanka, "Mitigation of automotive radar interference," in *Proc. IEEE Radar Conf.*, Oklahoma City, OK, Apr. 2018, pp. 0405–0410. doi: 10.1109/RADAR.2018.8378593.

[115] M. Wagner, F. Sulejmani, A. Melzer, P. Meissner, and M. Huemer, "Threshold-free interference cancellation method for automotive FMCW radar systems," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Florence, Italy, May 2018, pp. 1–4. doi: 10.1109/ISCAS.2018.8351077.

[116] M. Rameez, M. Dahl, and M. Petterson, "Adaptive digital beamforming for interference suppression in automotive FMCW radars," in *Proc. IEEE Radar Conf.*, Oklahoma City, OK, Apr. 2018, pp. 0252–0256. doi: 10.1109/RADAR.2018.8378566.

[117] G. M. Brooker, "Mutual interference of millimeter-wave radar systems," *IEEE Trans. Electromagn. Compat.*, vol. 49, no. 1, pp. 170–181, 2007. doi: 10.1109/TEMC.2006.890223.

[118] A. Bourdoux, K. Parashar, and M. Bauduin, "Phenomenology of mutual interference of FMCW and PMCW automotive radars," in *Proc. IEEE Radar Conf.*, Seattle, WA, May 2017, pp. 1709–1714. doi: 10.1109/RADAR.2017.7944482.

[119] H. Rohling and M.-M. Meinecke, "Waveform design principles for automotive radar systems," in *Proc. CIE Int. Conf. Radar*, Beijing, Oct. 2001. doi: 10.1109/ICR.2001.984612.

[120] M. Xue, L. Xu, and J. Li, "IAA spectral estimation: Fast implementation using the Gohberg–Semencul factorization," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3251–3261, 2011. doi: 10.1109/TSP.2011.2131136.

[121] G. O. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4154–4167, 2011. doi: 10.1109/TSP.2011.2145376.

[122] G. O. Glentis and A. Jakobsson, "Superfast approximative implementation of the IAA spectral estimate," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 472–478, 2012. doi: 10.1109/TSP.2011.2170979.

[123] L. C. Godara, "Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, 1997. doi: 10.1109/5.622504.

Ulisses M. Braga-Neto and Edward R. Dougherty

Machine Learning Requires Probability and Statistics

The contemporary practice of machine learning often involves the application of deterministic, computationally intensive algorithms to iteratively minimize a criterion of fit between a discriminant and sample data. There is often little interest in using probability to model the uncertainty in the problem and statistics to characterize the behavior of predictors derived from data, with the emphasis being on computation and coding. It follows that little can be stated about performance on future data, beyond perhaps a simple error count on a given test set. In this article, we argue that the knowledge imparted by deterministic computational methods is not rigorously related to the real world and, in particular, future events. This connection requires rigorous probabilistic modeling and statistical inference as well as an understanding of the proper role of computation and an appreciation of epistemological issues.

Gauss and the least-squares method

We illustrate the issue with a brief historical excursion into the development of the least-squares method, which is usually credited to an 1809 paper by Gauss [1]—even though Legendre published it in 1805 [2]. Around 1795 (thus, before Legendre's publication), Gauss became preoccupied with the inaccuracy of the classical model of planetary motion (due to Kepler and refined by Newton) in predicting the orbit of the asteroid Ceres. The classical planetary model does not take into account the uncertainty introduced by noise in the observations and the presence of unmodeled variables. Namely, astronomical observations contain inaccuracies, such as human error, atmospheric interference, and optical imperfections in telescopes, and the orbits of planets are determined not only by the sun but also by a superposition of the effects of all the other planets, which creates an intractable analytical problem (known as the *n-body problem*). To address these issues, Gauss introduced the least-squares method, which can be summarized in the famous passage [1]:

> The most probable value of the unknown quantities will be that in which the sum of the squares of the differences between the actually observed and the computed values multiplied by numbers that measure the degree of precision is a minimum.

Thus, the "most probable value" is the one that minimizes the sum of squared deviations between the observations and a candidate in a given family (for example, the family of all ellipses). The least-squares approach has proven to be extremely influential in science and engineering. However, its basic formulation has a significant limitation—it cannot say anything about the performance of the method on future data. This is because no knowledge about the properties of the noise in the model is assumed or sought. This makes the least-squares approach essentially deterministic.

Gauss was not unaware of this issue. In a later work [3], he gave the conditions on the observation noise under which the approach is optimal: if the noise random variables are uncorrelated and have zero mean with constant variance across all observations, then the least-squares solution is unbiased and has minimum variance among all linear estimators; i.e., it is the best linear unbiased estimator (BLUE)—a result known today as the Gauss–Markov theorem. Even though Laplace also worked on the theory of the least-squares method around the same time as Gauss, and Markov later clarified many of its issues, R. Plackett credits Gauss' 1821 paper fully for the Gauss–Markov theorem (which should perhaps then be called the Gauss theorem) [4]. Gauss's 1821 approach is fully stochastic, where the unmeasurable and uncontrollable disturbances are modeled as random variables.

The contrast between the 1809 deterministic least-squares method by Gauss and his 1821 fully stochastic approach represents a significant epistemological transition and illuminates the entire issue we discuss in this article. The 1809 result appeared to be a useful computational method that produced a "good-looking" result given the data. However,

---

Digital Object Identifier 10.1109/MSP.2020.2985385
Date of current version: 26 June 2020

its properties could not be established until the result in 1821, which required conditions on the probability distribution of the noise to lead to the statistical optimality of the procedure on future data (and not optimality based merely on the minimization of a sum of squared errors on the current data). In a similar fashion, probability theory and statistics are indispensable components of statistical signal processing, stochastic control, and information theory.

Computers: Thinking the unthinkable or not thinking at all?

Computers are fascinating because of their superhuman speed and accuracy in executing rote tasks. This feeling about automation is old and precedes computers. For example, Denis Diderot, who made the mechanical arts one of the central parts of the *Encyclopedie*, includes the following quote by a certain M. Perault in the famous article on the stocking machine [5], [6]:

> When one sees stockings being knit, one marvels at the suppleness and dexterity of the worker's hand, though he only makes one stitch at a time. What then when one sees a machine that makes hundreds of stitches at once, that is, makes in one moment all the diverse movements that the human hands would take many hours to make? […] and all that without the worker who operates the machine understanding anything, knowing anything, or even dreaming of it.

The impression that automation always produces results that are not just faster but superior to manual labor is very strong. The philosopher W. Barrett put it this way in his book *The Illusion of Technique* [7]:

> In the popular imagination the faith in hardware expresses itself in the images of technological gigantism: just make the computer mammoth enough and it will solve all problems. But the intrinsic logic of a problem remains what it is even if we had at our disposal a computer gigantic enough to cover a modern city. The absence of an intelligent idea in the grasp of a problem cannot be redeemed by the elaborateness of the machinery one subsequently employs.

A purely data-driven approach is naturally computationally expensive, and this was a key reason why its use was not frequent before the advent of fast and cheap computers. Now computation and storage are relatively cheap and widely available, which makes it very attractive to apply computation indiscriminately. B. Efron put the matter thus in his paper, "Computers and the Theory of Statistics: Thinking the Unthinkable" [8]: "The 'unthinkable' mentioned in the title is simply the thought that one might be willing to perform 500,000 numerical operations in the analysis of 16 data points. Or one might be willing to perform a billion operations to analyze 500 numbers. Such statements would have seemed insane thirty years ago."

The propensity of using computation indiscriminately was very much in the minds of the pioneers of the information age. The very first manual of the BASIC programming language, invented by J. Kemeny and T. Kurz at Dartmouth in 1964, had a piece of advice from digital signal processing pioneer Richard Hamming of Bell Labs: "Typing is no substitute for thinking" [9].

Our point is not that computation should be avoided but that an exaggerated reliance on it can create an illusion of excellence and independence of human supervision. In the case of machine learning, it has created the expectation that vast amounts of computation can produce accurate predictions from data, without a specification of conditions that provide the possibility of this knowledge.

On prediction, validation, and experimental design

We arrive at the fundamental question: How does one know that one has a predictive model that is strongly connected to the real world and future events? In classification, a model is predictive if the classification error rate is small. However, how do we know that we have a predictive classifier? This question can only be answered in practice by estimating the classification error using an error-estimation rule applied to the training data, a distinct set of test data, or a mixture of training and test data [10]. The accuracy of the error estimation rule must be measured by a validity criterion, the most common one being the root-mean-square error between estimated and true error, which in turn depends on the feature–label distribution, i.e., the joint probability distribution between the feature vector and the label. Without an underlying probability model, classification validity cannot be established. One could apply any existing error estimation rule, such as cross-validation or test-set error estimation, but this would simply provide a number that relates to the sample training and testing data used and has no quantifiable relation with future performance of the classifier. One might claim to expect a proportion of errors on future applications that agree with the estimate, but this statement is not quantifiable in terms of prediction versus observation and therefore lacks scientific content.

The next logical question is: How can one know that the feature–label distribution reflects the true relationship between feature vector and label? If one has the wrong feature–label distribution, then the trained model will perform poorly on future data. In practice, it is not possible to know completely the feature–label distribution at work in a specific problem. However, assumptions about the feature–label distribution can be enforced by sound experimental design, i.e., the way the data gathering process is planned and executed. In a recent paper, F. Mazzocchi characterizes the issue as follows [11]:

> Science does not collect data randomly. Experiments are designed and carried out within theoretical, methodological and instrumental limitations. Instruments are designed based on prior theories and knowledge, which determine what these instruments indicate

> with respect to the object under investigation. Research does not examine each possible manipulation that could occur, but selects what is relevant in light of a given perspective, sometimes in order to match theoretical predictions with experience.

For instance, for the Gauss–Markov theorem to hold and the BLUE to be valid, the data must be acquired and conditioned/transformed so that the disturbances are, at least approximately, uncorrelated and zero mean with constant variance. Another often overlooked example of an important experimental design issue is the requirement by many machine learning algorithms that the training data be independent and identically distributed. For example, cross-validation is approximately unbiased only under this assumption: if it is violated, cross-validation can be grossly biased [12].

A last objection could be raised by a skeptic: How can we be sure that we can learn from the present data about events that will happen in the future? In other words, are we not always measuring performance on existing test data, as it becomes available? This is the radical empiricist challenge to science, also known as the problem of induction, which was first raised by David Hume in the 18th century [13]. Many attempts have been made to answer this question in the affirmative, but this principle cannot be proved logically. Instead, we must adopt it as a postulate. In the preface to "Scientific Inference," Sir Harold Jeffreys puts it this way [14]: "Discussions from the philosophical and logical point of view have tended to the conclusion that this principle cannot be proved by logic alone, which is true, and have left it at that. […] In the present work the principle is frankly adopted as a primitive postulate and its consequences are developed." As is the case in science at large, Jeffreys' "primitive postulate" must also be adopted in machine learning to avoid the radical empiricist perspective.

Deep neural networks

Like other machine learning methods based on optimization, neural networks learn from data by iteratively adjusting the parameters of a discriminant to fit a set of labeled data points. The justification often cited for such approaches is that the discriminant of a neural network with a sufficiently large number of parameters can produce results that are arbitrarily close to the optimal predictor in a distribution-free manner; that is, neural networks are universal function approximators. If the large number of parameters is organized over a large number of layers, one has a deep neural network.

A classical theorem by G. Cybenko [15] shows that the classifiers produced by a neural network discriminant with continuous sigmoids are dense in the space of all classifiers, and therefore can be arbitrarily close to the optimal classifier. Cybenko's theorem applies to depth-bound ("shallow") neural networks, with a depth of 2 (one hidden layer and one output layer), but the number of neurons k in the hidden layer, i.e., the width of the neural network, must be allowed to increase without bound for arbitrary approximation. A recent result by Z. Lu and collaborators [16] provides a comparable denseness result for width-bound ("deep") networks, where the maximum number of neurons per layer is fixed, but the number of layers must be allowed to increase freely.

However, these deterministic results do not address the performance of neural networks trained from data. In particular, they do not weigh directly on the issue of consistency, i.e., on the stochastic convergence of the error of the trained classifier to the optimal error as sample size increases to infinity. To bear out the promise of distribution-free classification, consistency has to be universal; i.e., it must hold under any feature–label distribution. The few universal consistency results of which we are aware apply to shallow neural networks and make unrealistic demands on training, such as the requirement of training error or L_1-error minimization (for which gradient descent cannot be applied); e.g., see Theorems 30.7 and 30.9 in [17], respectively.

It turns out that even universal consistency is not enough. It was shown by L. Devroye and collaborators [17, Th. 7.2] hat for any universally consistent classification rule, a feature–label distribution exists such that convergence of the classification error to the optimal error is as slow as desired. In other words, universal consistency can say nothing about the problem of selecting a good classifier using finite training data: under no assumptions about the feature–label distribution, any classification algorithm can be arbitrarily bad. The situation changes if assumptions are made about the feature–label distribution. For example, it was shown by N. Glick [18, Th. A] that the difference between the expected and optimal classification error rates converges exponentially fast to zero in discrete histogram classification, with a rate that depends on the feature–label distribution: the more separable the classes are under the feature–label distribution (in a precise sense), the faster the rate of convergence is guaranteed to be.

Bayesian deep learning provides an alternative to deterministic neural networks and has attracted significant attention in recent years [19], [20]. The classical approach to Bayesian neural networks has been based on placing prior distributions on the weights of the network [21], [22]. In [23], an alternative approach to Bayesian deep learning was developed where dropout training of deep neural networks was formulated as approximate Bayesian inference in deep Gaussian processes. These are welcome developments. However, we point out that these approaches to Bayesian deep learning are not probabilistically related to the feature–label distribution, nor do they even require a feature–label distribution. The mechanism has a statistical dimension relative to the prior distribution, but the output is not necessarily statistically related to nature, only to the actual data. If there is an underlying model distribution representing scientific knowledge, then the issue arises as to the connection between it and the prior distribution. In the classical Bayesian approach, there is none, the priors being chosen ad hoc, perhaps according to some general information–theoretical criteria. This disconnect has been referred to as a "scientific gap" in [24].

In that reference, an alternative approach was proposed, where the model distribution is treated as uncertain, with the uncertainty occurring in its parameters. This model uncertainty is propagated to uncertainty in the classifier or regressor, and so the latter uncertainty is directly related to the underlying model.

Conclusions

In this article, we have attempted to make the case that the success of machine learning in science and engineering depends essentially on the use of rigorous probabilistic modeling and statistics. We discussed the problem in the context of general machine learning and then looked more closely at the currently important topic of deep neural networks.

In our view, the issues discussed here concern not only the theoretician or scientist but also affect the practitioner, because practical application requires consistency with the demands of the theory. This view is expressed this way by Y. Gal and Z. Ghahramani in [23]: "Model uncertainty is indispensable for the deep learning practitioner as well," whereas N. Wiener put it this way in the context of biology [25]: "The physiologist need not be able to prove a certain mathematical theorem, but he must be able to grasp its physiological significance and tell the mathematician for what he should look for."

In the last few decades, however, there has been open opposition to the use of probability and statistics in predictive modeling. For example, this can be observed in the well-known polemic by L. Breiman [26], in which he writes: "The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems."

There is some evidence that this might be due to the lack of widespread mathematical literacy in the research community at large. For instance, J. Simon writes [27], "In the mid-1960's, I noticed that most graduate students—among them many who had had several advanced courses in statistics—were unable to apply statistical methods correctly in their social science research. I sympathized with them. Even many experts are unable to understand intuitively the formal mathematical approach to the subject. Clearly, we need a method free of the formulas that bewilder almost everyone."

The authors have pondered on such issues elsewhere [28]–[31]; others have done so as well [11], [32]–[35]. The crucial question is: Do we want knowledge about the real world in the sense of modern engineering and science, or do we merely want knowledge of specific events, the latter being more understandable and requiring simpler mathematics? Before answering the question, one should consider the enormous benefits that we, as modern engineers, have accrued from the probabilistic-statistical approach, beginning with the Wiener–Kolmogorov theory of linear systems in the 1930s and flowing forward in the development of optimal filtering, stochastic control, statistical signal processing, and information theory. Deep thought should be given as to whether abandoning this epistemology is desirable.

Authors

Ulisses M. Braga-Neto (ulisses@ece.tamu.edu) received his Ph.D. degree in electrical and computer engineering from The Johns Hopkins University, Baltimore, Maryland. He is currently a professor in the Department of Electrical and Computer Engineering at Texas A&M University, College Station. His research interests include pattern recognition, machine learning, and statistical signal processing. He is the author of the recent textbook, *Fundamentals of Pattern Recognition and Machine Learning* (Springer, 2020) and coauthor of *Error Estimation for Pattern Recognition* (IEEE-Wiley, 2015) with Edward Dougherty. He received the National Science Foundation CAREER Award for his work in this area. He is a Senior Member of the IEEE.

Edward R. Dougherty (edward@ece.tamu.edu) received his M.S. degree in computer science from Stevens Institute of Technology, Hoboken, New Jersey and his Ph.D. degree in mathematics from the Rutgers University, Camden, New Jersey. He is a distinguished professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, where he holds the R.M. Kennedy '26 Chair in electrical engineering and is the scientific director of the Center for Bioinformatics and Genomic Systems Engineering. He was awarded the Doctor Honoris Causa by the Tampere University of Technology, Finland, the SPIE President's Award, and served as the editor of the Society for Optical Engineering/Society for Imaging Science and Technology *Journal of Electronic Imaging*. At Texas A&M University, he received the Association of Former Students Distinguished Achievement Award in Research. He is a Fellow of the IEEE and SPIE.

References

[1] C. F. Gauss, *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, vol. 7. Hamburg, Germany: Perthes et Besser, 1809.

[2] S. M. Stigler, "Gauss and the invention of least squares," *Ann. Statist.*, vol. 9, no. 3, pp. 465–474, 1981. doi: 10.1214/aos/1176345451.

[3] C. F. Gauss, *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*, vol. 1. Göttingen, Germany: H. Dieterich, 1823.

[4] R. L. Plackett, "A historical note on the method of least squares," *Biometrika*, vol. 36, nos. 3–4, pp. 458–460, 1949. doi: 10.2307/2332682.

[5] D. Diderot, "Bas (bonneterie—)," in *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers, par une Société de Gens de lettres*, Autumn 2017 ed., R. Morrissey and G. Roe, Eds. Chicago: Univ. Chicago, ARTFL Encyclopédie Project, 2017, p. 2:98. [Online]. Available: https://encyclopedie.uchicago.edu/

[6] J. Stalnaker, *The Unfinished Enlightenment: Description in the Age of the Encyclopedia*. Ithaca, NY: Cornell Univ. Press, 2010.

[7] W. Barrett, *The Illusion of Technique: A Search for Meaning in a Technological Civilization*. Garden City, NY: Anchor Press, 1978.

[8] B. Efron, "Computers and the theory of statistics: Thinking the unthinkable," *SIAM Rev.*, vol. 21, no. 4, pp. 460–480, 1979. doi: 10.1137/1021092.

[9] M. J. Lorenzo, *Endless Loop: The History of the BASIC Programming Language*. Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2017.

[10] U. M. Braga-Neto and E. R. Dougherty, *Error Estimation for Pattern Recognition*. Hoboken, NJ: Wiley, 2015.

[11] F. Mazzocchi, "Could big data be the end of theory in science?" *EMBO Rep.*, vol. 16, no. 10, pp. 1250–1255, 2015. doi: 10.15252/embr.201541001.

[12] U. Braga-Neto, A. Zollanvari, and E. R. Dougherty, "Cross-validation under separate sampling: Strong bias and how to correct it," *Bioinformatics*, vol. 30, no. 23, pp. 3349–3355, 2014. doi: 10.1093/bioinformatics/btu527.

[13] D. Hume, *An Enquiry Concerning Human Understanding: A Critical Edition*, vol. 3. Oxford, U.K.: Oxford Univ. Press, 2000.

[14] H. Jeffreys, *Scientific Inference*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1973.

[15] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Math. Control, Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989. doi: 10.1007/BF02551274.

[16] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The expressive power of neural networks: A view from the width,” in *Proc. 31st Int. Conf. Advances Neural Information Processing Systems*, 2017, pp. 6231–6239. doi: 10.5555/3295222.3295371.

[17] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag, 1996.

[18] N. Glick, “Sample-based multinomial classification,” *Biometrics*, vol. 29, no. 2, pp. 241–256, 1973. doi: 10.2307/2529389.

[19] H. Wang and D.-Y. Yeung, “Towards Bayesian deep learning: A framework and some existing methods,” *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3395–3408, 2016. doi: 10.1109/TKDE.2016.2606428.

[20] A. Kendall and Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?” in *Proc. 31st Int. Conf. Advances Neural Information Processing Systems*, 2017, pp. 5574–5584. doi: 10.5555/3295222.3295309.

[21] D. J. MacKay, “A practical Bayesian framework for backpropagation networks,” *Neural Comput.*, vol. 4, no. 3, pp. 448–472, 1992. doi: 10.1162/neco.1992.4.3.448.

[22] A. Graves, “Practical variational inference for neural networks,” in *Proc. 24th Int. Conf. Advances Neural Information Processing Systems*, 2011, pp. 2348–2356. doi: 10.5555/2986459.2986721.

[23] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proc. 33rd Int. Conf. Machine Learning*, 2016, pp. 1050–1059. doi: 10.5555/3045390.3045502.

[24] X. Qian and E. R. Dougherty, “Bayesian regression with network prior: Optimal Bayesian filtering perspective,” *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6243–6253, 2016. doi: 10.1109/TSP.2016.2605072.

[25] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 1948.

[26] L. Breiman, “Statistical modeling: The two cultures (with comments and a rejoinder by the author),” *Stat. Sci.*, vol. 16, no. 3, pp. 199–231, 2001. doi: 10.1214/ss/1009213726.

[27] J. Simon, *Resampling: The New Statistics*. Arlington, VA: Resampling Stats, 1997.

[28] U. Braga-Neto, “Small-sample error estimation: Mythology versus mathematics,” in *Proc. Mathematical Methods Pattern and Image Analysis*, 2005, vol. 5916, p. 59160V. doi: 10.1117/12.619331.

[29] E. R. Dougherty and U. Braga-Neto, “Epistemology of computational biology: Mathematical models and experimental prediction as the basis of their validity,” *J. Biol. Syst.*, vol. 18, no. 14, pp. 65–90, 2006. doi: 10.1142/S0218339006001726.

[30] E. R. Dougherty and M. L. Bittner, *Epistemology of the Cell: A Systems Perspective on Biological Knowledge*, vol. 35. Hoboken, NJ: Wiley, 2011.

[31] P. V. Coveney, E. R. Dougherty, and R. R. Highfield, “Big data need big theory too,” *Phil. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 374, p. 20160153, Nov. 2016. doi: 10.1098/rsta.2016.0153.

[32] D. J. Glass and N. Hall, “A brief history of the hypothesis,” *Cell*, vol. 134, no. 3, pp. 378–381, 2008. doi: 10.1016/j.cell.2008.07.033.

[33] M. Frické, “Big data and its epistemology,” *J. Assoc. Inform. Sci. Technol.*, vol. 66, no. 4, pp. 651–661, 2015. doi: 10.1002/asi.23212.

[34] J. P. Ioannidis, “Why most published research findings are false,” *PLoS Med.*, vol. 2, no. 8, p. e124, 2005. doi: 10.1371/journal.pmed.0020124.

[35] W. Lee Kraus, “Editorial: Would you like a hypothesis with those data? Omics and the age of discovery science,” *Mol. Endocrinol.*, vol. 29, no. 11, pp. 1531–1534, 2015. doi: 10.1210/me.2015-1253.

SP

Soheil Kolouri, Xuwang Yin, and Gustavo K. Rohde

Neural Networks, Hypersurfaces, and the Generalized Radon Transform

Artificial neural networks (ANNs) have long been used as a mathematical modeling method and have recently found numerous applications in science and technology, including computer vision, signal processing, and machine learning [1], to name a few. Although notable function approximation results exist [2], theoretical explanations have yet to catch up with newer developments, particularly with regard to (deep) hierarchical learning. As a consequence, numerous doubts often accompany NN practitioners, such as How many layers should one use? What is the effect of different activation functions? What are the effects of pooling? and many others.

This didactic lecture note is meant to highlight an alternative interpretation of NN-based techniques and their use in supervised learning problems. Here, we draw a connection between classification methods in machine learning and the integration geometry of classic and generalized Radon transforms (GRTs). We explicitly show that the output distributions of the nodes in an NN are curvilinear projections of the input distribution (also known as the *slices* of the GRT). We then use these concepts to highlight different properties of NNs, which may help demystify certain of their properties as well as potentially provide a path for novel studies and developments.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. The material includes the derivations for the inverse Radon transform, the random variable transform theorem, and the isomorphism relationship between a perceptron and a standard Radon slice, along with the Python codes for generating the figures in this article. Please contact skolouri@hrl.com or gustavor@virginia.edu for further questions about this article.

For brevity and to reduce prerequisites, the derivations presented fall short of rigorous mathematical proofs. The Python code to reproduce all of the figures used here is available on GitHub [13].

Relevance

This article aims to establish a common language between NNs and signal transformation methods typically used in inverse problems (e.g., image reconstruction). More specifically, our goal is to show that, just like a linear classification boundary (e.g., linear logistic regression) can be thought of as a "straight-line projection" of the input data probability density function (PDF), certain types of NNs can be thought of as computing a "curved" projection over the input data PDF. By focusing on the geometry of integration, we also intend to visualize concepts related to nonlinearity, the role of different activation functions, and adversarial attacks. In the end, we hope that this common language facilitates the exchange of ideas between the fields of machine learning and inverse problems, to the benefit of both.

Prerequisites

Linear algebra, basic probability, and basic concepts of machine learning are helpful in understanding the content of this lecture note. Radon transforms are also helpful and are reviewed as well.

Problem statement: Statistical regression and classification

Let x refer to an input digital signal or image that can be viewed as a point within a space X (e.g., the $\mathbb{R}^d$ Euclidean) space. Now, let h be a function that takes as input x and assigns a label to it. Mathematically, we write this as $h: X \to Y$, with Y corresponding to the label space. In learning problems, $y \in \mathbb{R}^K$ is usually a (given) vector for which the value of the kth element represents the probability that the sample x belongs to the kth class, although other regression problems can also be formulated with the same approach. The goal in regression (e.g., classification) problems is to utilize input training data to estimate a model f_θ, with θ a vector of parameters, such that $f_\theta \sim h$ (f_θ closely approximates h).

Digital Object Identifier 10.1109/MSP.2020.2978822
Date of current version: 26 June 2020

Omitting here a measure of theoretic formulation (see [3] for a more complete development), let p_X, p_Y, and $p_{X,Y}$ define the PDFs for random variables X (input signal space) and Y (label space), respectively. The well-known random variable transformation (RVT) theorem [4] links p_X and p_Y via

$$p_Y(y) = \int_X p_X(x)\delta(y - h(x))\,dx, \tag{1}$$

where δ is the standard Dirac distribution (see the supplementary material for this article available in IEEE *Xplore*). Figure 1 shows the geometric interpretation of (1), which is the topic of this note. The same transformation of the random variables technique can be used to derive

$$p_{f_\theta}(z) = \int_X p_X(x)\delta(z - f_\theta(x))\,dx. \tag{2}$$

The goal in a regression task is to estimate f_θ so that it accurately "predicts" the dependent variable y for each input x. In other words, we wish to find $f_\theta \sim h$ over the distribution of the input space. To that end, goodness-of-fit measures are used to fit a model f_θ to given labeled data (supervised learning). One popular model is to find θ that minimizes the expected discrepancy between $y = h(x)$ and $f_\theta(x)$ according to a per-example dissimilarity measure $\mathcal{L}$ (i.e., loss function):

$$\min_\theta \mathbb{E}_{(x,y)\sim p_{X,Y}} \mathcal{L}(y, f_\theta(x)). \tag{3}$$

The expected dissimilarity between h and f_θ is referred to as *risk*. Given that we often have access only to samples from the distribution, $p_{X,Y}$, in practice, we use the empirical distribution, which leads to the empirical risk minimization:

$$\min_\theta \sum_{n=1}^{N} \frac{1}{N} \mathcal{L}(y_n, f_\theta(x_n)). \tag{4}$$

The empirical risk minimization can be interpreted in relation to random variables $Y = h(X)$ and f_θ and their respective distributions. For instance, the cross-entropy-minimization strategy $(-1/N)\sum_{k=1}^{N} y_k \cdot \log(f_\theta(x_k))$ can be viewed as an estimate of $\mathbb{E}_{x\sim p_X}(h(x)\cdot \log(f_\theta(x)))$, which is equivalent to minimizing the Kullback–Leibler (KL)-divergence between p_Y and p_{f_θ} [5].

Main objective

The goal of this note is to exploit the described relationships to clarify the role of an NN classifier $f_\theta(x)$ in linking the PDF of the input data p_X and the PDF of an output node p_{f_θ}. In particular, we work to show that $p_{f_\theta} = \mathcal{G}p_X(\cdot, \theta)$, where $\mathcal{G}$ stands for the GRT operation discussed in the next section. Next, we consider the formulations for the standard Radon transform and its generalized version as well as demonstrate a connection between this transformation and the reviewed statistical learning concepts.

Radon transforms

The "standard" Radon transform

The standard Radon transform, $\mathcal{R}$, maps distribution p_X to the infinite set of its 1D marginal distributions. More precisely, the standard Radon transform integrates the distribution p_X along hyperplanes of $\mathbb{R}^d$, and it is defined as

$$\mathcal{R}p_X(t, \theta) := \int_X p_X(x)\delta(t - \theta\cdot x)\,dx, \tag{5}$$

where δ is the 1D Dirac delta function. Each hyperplane is parameterized by a unit vector, $\theta \in \mathbb{S}^{d-1}$, and an intercept $t \in \mathbb{R}$, where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$. In other words, each hyperplane can be written as

$$H(t, \theta) = \{x \in \mathbb{R}^d | \theta\cdot x = t\}, \tag{6}$$

which, alternatively, could be thought of as the level set of the function $g(x, \theta) = \theta\cdot x = t$. For a fixed θ, the integrals over all hyperplanes orthogonal to θ define a continuous function, $\mathcal{R}p_X(\cdot, \theta): \mathbb{R} \to \mathbb{R}$, that is a projection/slice of p_X. Figure 2(a) provides a visual representation of the Radon transform, the integration hyperplanes $H(t, \theta)$ (i.e., lines for $d = 2$), and the slices $\mathcal{R}p_X(\cdot, \theta)$.

The Radon transform is an invertible linear transformation (i.e., linear bijection). The inverse of the Radon transform denoted by $\mathcal{R}^{-1}$ is defined as

$$\begin{aligned} p_X(x) &= \mathcal{R}^{-1}(\mathcal{R}p_X(t, \theta)) \\ &= \int_{\mathbb{S}^{d-1}} (\mathcal{R}p_X(\cdot, \theta) * \eta(\cdot)) \circ (\theta\cdot x)\,d\theta, \end{aligned} \tag{7}$$

where $\eta(\cdot)$ is a 1D high-pass filter with corresponding Fourier transform

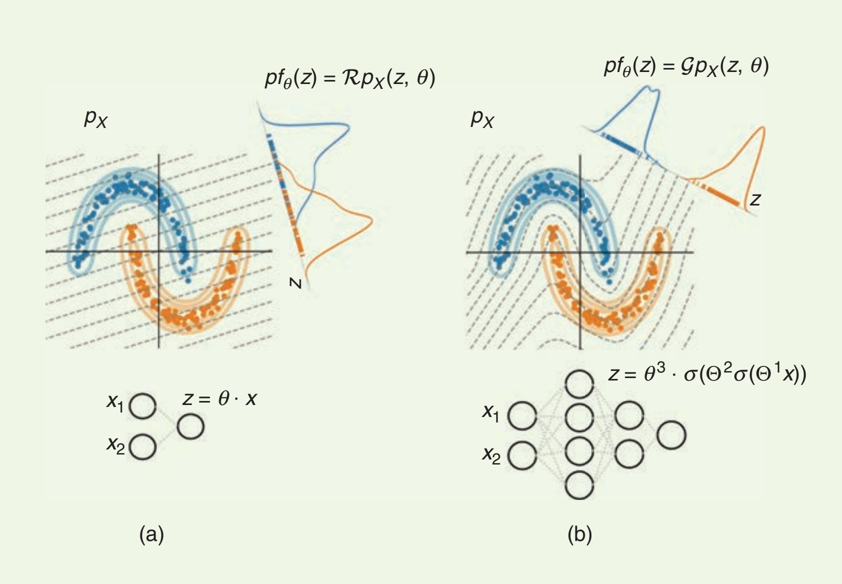


FIGURE 1. The PDFs of classifier outputs p_{f_θ} can be understood as (a) a straight-line or (b) a curvilinear projection of the input data PDF, p_X. In the case of linear classifiers (e.g., logistic regression), the PDF of the classifier's output is a projection from the standard Radon transform of the data PDF, p_X [see (a)]. In the case of nonlinear classifiers (e.g., multilayer perceptrons with nonlinear activations), the PDF of the classifier's output can be viewed as a projection obtained from the GRT of the data PDF.

$\mathcal{F}\eta(\omega) \approx c|\omega|^{d-1}$ (it appears due to the Fourier slice theorem; see the supplementary material for this article available in IEEE *Xplore*) and "*" denotes the convolution operation. This definition of the inverse Radon transform is also known as the *filtered back-projection method*, which is extensively used in image reconstruction in the biomedical imaging community. Intuitively, each 1D projection/slice, $\mathcal{R}p_X(\cdot,\theta)$, is first filtered via

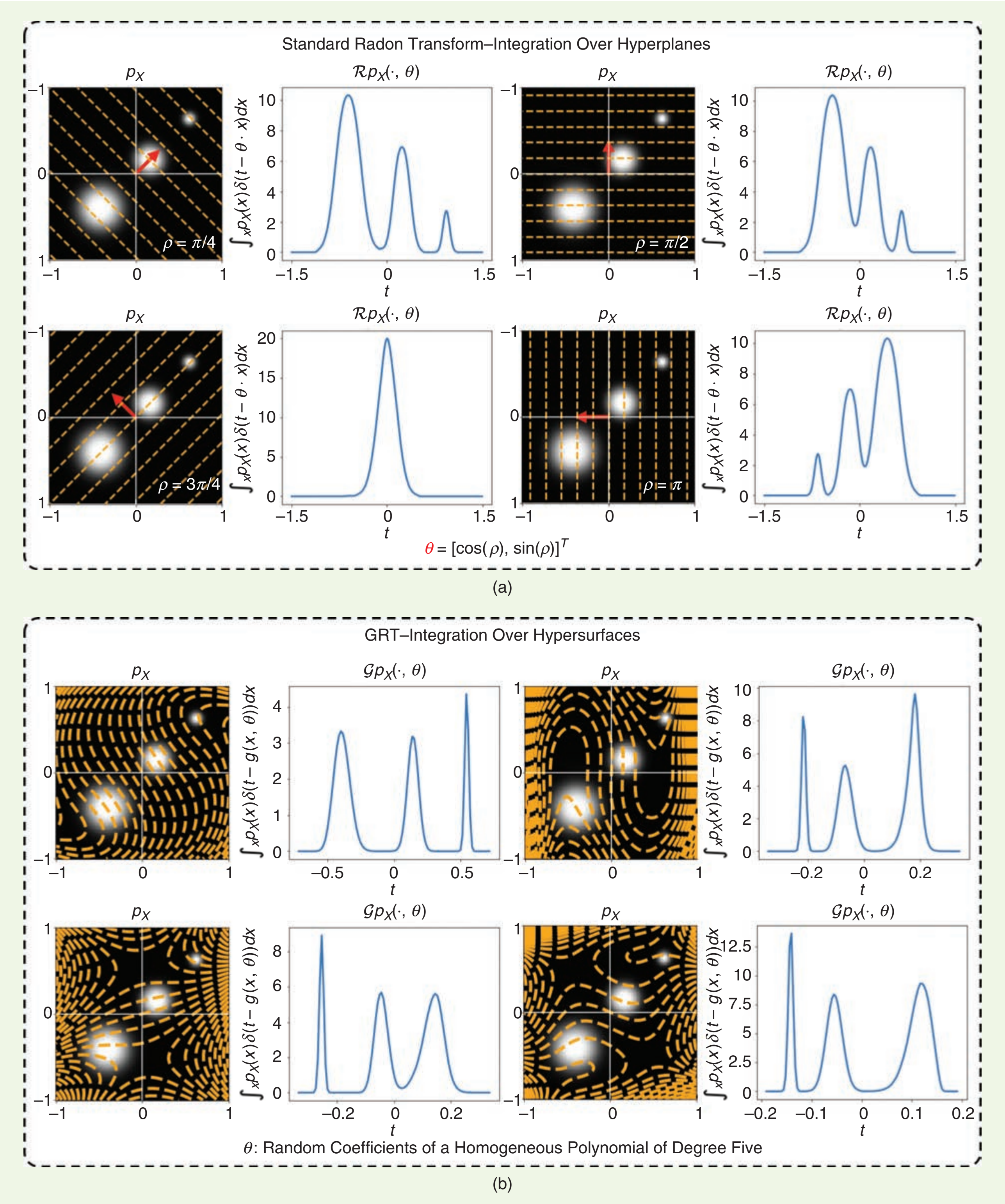


FIGURE 2. Linear versus curvilinear projections. A visualization of the Radon transforms projecting 2D Gaussian mixture density functions to 1D density functions. Each pair of figures shows the input density function p_X together with the integration lines (orange) and the resulting 1D density function, i.e., projection (blue). (a) The projections along straight lines (parameterized by different θs as indicated by red arrows). (b) The projections along polynomial curves.

a high-pass filter and then smeared back into X along $H(\cdot,\theta)$ to approximately p_X. The filtered summation of all smeared approximations then reconstructs p_X. We note that there are example works in the NN literature where these relationships are exploited to derive certain approximation results for one-hidden-layer NNs (see, for example, [6] and [7]).

We note that the classic Radon transform establishes a unique correspondence between smooth functions (PDFs) p_X and $\mathcal{R}p_X$: each function p_X gets mapped to only one function $\mathcal{R}p_X$, and vice versa (bijective map). This concept is important since it enables one to define the inverse operation, as explained earlier, which has important tomographic applications, for instance, in medical imaging and synthetic aperture radar, among others. Another interesting application of the bijectivity concept is in statistics, where one compares two distributions, p_X and q_X, based on their marginals $\mathcal{R}p_X$ and $\mathcal{R}q_X$, and invertibility indicates that if $\mathcal{R}p_X(t,\theta) = \mathcal{R}q_X(t,\theta)$ for $\forall t \in \mathbb{R}$ and $\forall \theta \in \mathbb{S}^{(d-1)}$, then $p_X = q_X$. The comparison of 1D projections of high-dimensional distributions is the basis for interesting metrics for distributions, including the sliced Wasserstein metric (see [8] for a recent example).

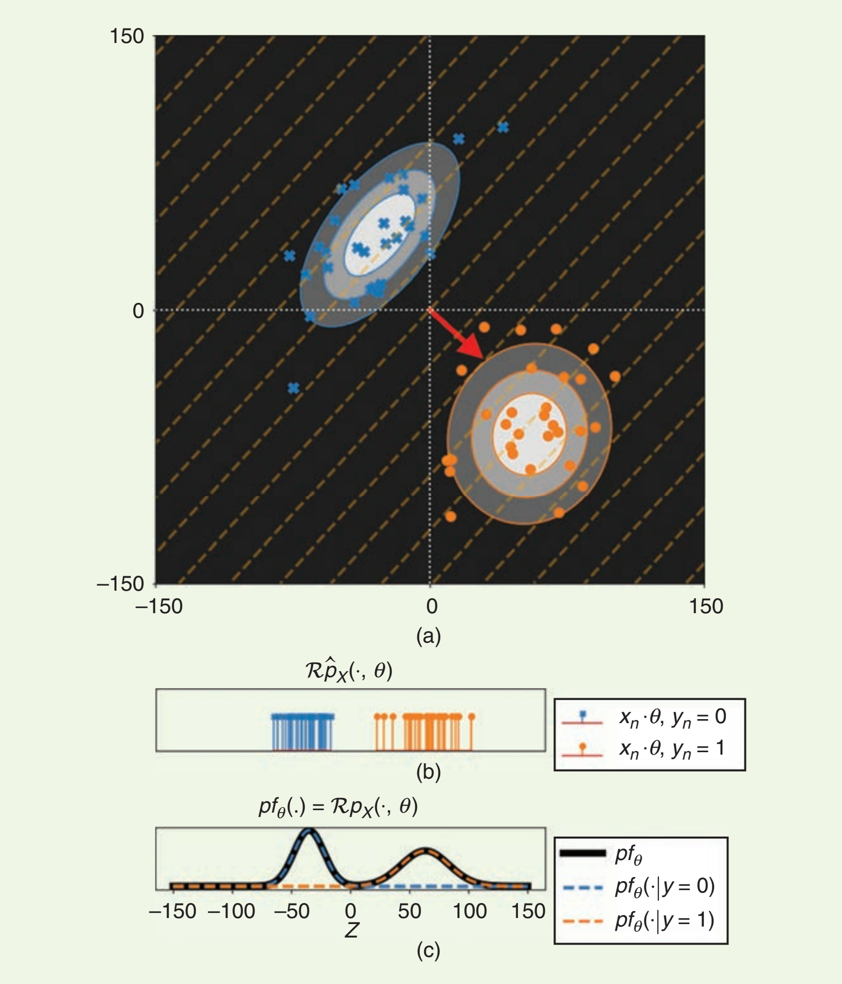


FIGURE 3. A linear classifier provides a linear projection of the distribution of the input high-dimensional data p_X at an optimal θ, for which the data can be best discriminated. Therefore, one can think of the distribution of the output of the classifier as a slice of the classic Radon transform of the distribution p_X. (a) The PDF of the input data, where the blue and orange colors denote the conditional distributions $p_X(\cdot|y=0)$ and $p_X(\cdot|y=1)$ for the two classes. The blue crosses and orange circles show samples from the conditional distributions. (b) The empirical distribution $(1/N)\sum_{n=1}^{N}\delta(z-\theta\cdot x_n)$ of the output of a linear classifier. (c) The Radon transform of p_X, $\mathcal{R}p_X(\cdot,\theta)$ at the optimal θ.

Estimating Radon projections numerically

As mentioned previously, in most machine learning applications, one does not have direct access to the actual distribution of the data but to its samples, $x_n \sim p_X$. In such scenarios, the empirical distribution of the data is used as an approximation for p_X:

$$p_X(x) \approx \hat{p}_X(x) = \frac{1}{N}\sum_{n=1}^{N}\delta(x - x_n), \tag{8}$$

where δ is the Dirac delta function in $\mathbb{R}^d$. Then, it is straightforward to show that the Radon transform of $\hat{p}_X$ is

$$\mathcal{R}\hat{p}_X(t,\theta) = \frac{1}{N}\sum_{n=1}^{N}\delta(t - \theta\cdot x_n). \tag{9}$$

See the supplementary material for this article available in IEEE *Xplore* for detailed derivations of (9).

Linear classification and the Radon transform

Now, let us consider the supervised learning of a linear binary classifier. Given the data samples $\{x_n \sim p_X\}_{n=1}^{N}$ and their corresponding labels $\{y_n \in \{0,1\}\}_{n=1}^{N}$, the task is to learn a linear function of the input samples, $f_\theta(x) = \theta \cdot x$ for $\theta \in \mathbb{S}^{d-1}$, such that

$$\theta \cdot x \underset{y=0}{\overset{y=1}{\gtrless}} b.$$

Many methods exist to obtain the optimal θ, e.g., support vector machines or logistic regression [9]. While the projection $z = f_\theta(x) = \theta \cdot x$ is applied to each sample, we can consider the distribution of the numerical output of the classifier $f_\theta(x)$. By combining the transformation of random variables formula (2) with the Radon transform formula (5), we then have

$$\begin{aligned} p_{f_\theta}(z) &= \int_X p_X(x)\delta(z - f_\theta(x))\,dx \\ &= \int_X p_X(x)\delta(z - \theta\cdot x)\,dx \\ &= \mathcal{R}p_X(z,\theta). \end{aligned}$$

It is clear that the output distribution of a particular linear classifier defined by $f_\theta(x) = x \cdot \theta$ can be viewed as a slice of the standard Radon transform of the distribution of the input data

p_X. Figure 3 depicts this relationship. Note that these results are not exclusive to a binary classifier, and one can readily extend them to multiclass classification problems via one-versus-one or one-versus-all strategies [9].

The GRT

The GRT extends the original idea of the classic Radon transform introduced by J. Radon [10] from integration over hyperplanes of $\mathbb{R}^d$ to integration over hypersurfaces [11] [i.e., $(d-1)$-dimensional manifolds]. The GRT has various applications, including thermoacoustic tomography, where the hypersurfaces are spheres, and electrical impedance tomography, where integration over hyperbolic surfaces appears. Figure 2(b) demonstrates the concept of integration along hypersurfaces (i.e., curves for $d=2$) of our distribution domain $X \subseteq \mathbb{R}^2$.

To define the GRT, we first introduce the family of real-valued parametric functions $g(\cdot,\theta)\colon X \to \mathbb{R}$ for $\forall\theta \in \{\mathbb{R}^n \backslash 0\}$ (the vector $\theta = 0$ is excluded) and with $X \subseteq \mathbb{R}^d$. Then, the GRT is a linear operator, $\mathcal{G}$, defined as

$$\mathcal{G}p_X(t,\theta) := \int_X p_X(x)\,\delta(t - g(x,\theta))\,dx. \tag{10}$$

From a geometrical perspective and for a fixed t, $\mathcal{G}p_X(t,\theta)$ is the integral of p_X along the hypersurface $H(t,\theta) = \{x \in X | g(x,\theta) = t\}$. Note that the classic Radon transform is a special case of the GRT where $g(x,\theta) = \theta \cdot x$.

As with the standard Radon transform discussed previously, a natural question to ask is whether the GRT defines a bijective map that can be used to define interesting metrics for high-dimensional distributions as well as potentially develop an inverse transformation formula. As we mentioned before, invertibility is a desirable characteristic for Radon transforms, and it significantly empowers these transformations to be used in tomographic applications and statistical analysis, e.g., defining probability metrics. Here, we enumerate some necessary conditions for invertibility of the GRT. Invertibility of $\mathcal{G}$ necessitates the class of our real-valued functions, g, to satisfy four conditions [12]:

1) $g(\cdot,\theta)$ is a real-valued C^∞ function, $\forall\theta \in \mathbb{R}^n \backslash 0$, which ensures smoothness of hypersurfaces.
2) $g(x,\theta)$ is homogeneous of degree one in θ, i.e., $\forall\lambda \in \mathbb{R}$, $g(x,\lambda\theta) = \lambda g(x,\theta)$. Homogeneity is needed to ensure unique parametrization of hypersurfaces,
3) g is nondegenerate in the sense that $\nabla_x g(x,\theta) \neq 0$ in $X \times \mathbb{R}^n \backslash \{0\}$. The nondegenerate assumption ensures that the $(d-1)$-dimensional hypersurfaces do not collapse to points and that the integrals are well defined.
4) The mixed Hessian of g is strictly positive, i.e., $\det(\partial^2 g / \partial x^i \partial \theta^j) > 0$. This condition is a local form of the Bolker's condition (see [12]), which allows one to locally identify (x,θ) with the covector $(\nabla_x g(x,\theta)) / \|\nabla_x g(x,\theta)\|$.

The investigation of the sufficient and necessary conditions for showing the invertibility of GRTs is a long-standing topic [11], [12]. Conditions 1–4 for a defining function, g, enumerated in this section, are necessary conditions for injectivity but are not sufficient. Though the topic related to inversion of the GRT is beyond the scope of this didactic note, an inversion approach is given in [12].

GRT examples

Here, we list a few examples of known defining functions that lead to injective GRTs. The circular defining function, $g(x,\theta) = \|x - r * \theta\|_2$, with $r \in \mathbb{R}^+$ and $\Omega_\theta = \mathbb{S}^{d-1}$, was shown to provide an injective GRT. More interestingly, homogeneous polynomials with an odd degree also yield an injective GRT [11], i.e., $g(x,\theta) = \sum_{|\alpha|=m} \theta_\alpha x^\alpha$, where we use the multi-index notation $\alpha = (\alpha_1, \ldots, \alpha_{d_\alpha}) \in \mathbb{N}^{d_\alpha}$, $|\alpha| = \sum_{i=1}^{d_\alpha} \alpha_i$, and $x^\alpha = \prod_{i=1}^{d_\alpha} x_i^{\alpha_i}$. Here, the summation iterates over all possible multi-indices α, such that $|\alpha| = m$, where m denotes the degree of the polynomial and $\theta_\alpha \in \mathbb{R}$. The parameter set for homogeneous polynomials is then set to $\Omega_\theta = \mathbb{S}^{d_\alpha - 1}$. We can observe that choosing $m=1$ reduces to the linear case $g(x,\theta) = \theta \cdot x$, since the set of the multi-indices with $|\alpha| = 1$ becomes $\{(\alpha_1, \ldots, \alpha_d);\ \alpha_i = 1$ for a single $i \in [\![1, d]\!]$, and $\alpha_j = 0,\ \forall j \neq i\}$, and contains d elements. Figure 2(b) shows random slices for homogeneous polynomials of degree five.

NNs and the GRT

An NN with a fixed architecture provides a family of nonlinear parametric functions. Here, we look at this family of functions through the lens of GRT. To illustrate the relationship between deep NNs and the GRT transform, we start by describing the link between the basic units of a deep NN (i.e., perceptron + activation function) and the standard Radon transform.

Single perceptron + activation

Let $z = \sigma(\theta \cdot x)$, with $\|\theta\| = 1$, define a perceptron with an added activation $\sigma(\cdot)$ for input data $x \sim p_X$, where we incorporated the bias, b, into θ. Such perceptron + activation functions construct the basic unit of NNs. Treating $z \sim p_Z$ as a random variable and using the RVT in (1), it is straightforward to show that p_Z is isomorphic to a single slice of p_X, $\mathcal{R}p_X(\cdot,\theta)$, when σ is invertible (see the supplementary material in IEEE *Xplore* for a proof). The isomorphic relationship provides a fresh perspective on perceptrons, stating that the distribution of the perceptron's output, $z = f_\theta(x)$, is equivalent to integration of the original data distribution p_X along hyperplanes $H(z,\theta) = \{x | \theta \cdot x = \sigma^{-1}(z)\}$, $-\infty < z < \infty$ [see (6)]. In addition, one can show that the distribution of the output of a perceptron is equal to the GRT with $g(x,\theta) = f_\theta(x)$:

$$\begin{aligned} p_{f_\theta}(z) &= \mathcal{G}p_X(z,\theta) \\ &= \int_X p_X(x)\,\delta(z - \sigma(\theta \cdot x))\,dx. \end{aligned} \tag{11}$$

An important and distinctive point is that, here, we are interested in the distribution of the output of a perceptron, $\mathcal{G}p_X(z,\theta)$, and its relationship to the original distribution of the data, p_X, as opposed to the individual responses of the perceptron, $z_n = g(x_n,\theta)$. Figure 4(a) and (b) demonstrates the level sets (or level curves, since $d=2$) and the line integrals for $g(x,\theta) = \theta \cdot x$ and $g(x,\theta) = \sigma(\theta \cdot x)$, where $\|\theta\| = 1$. Note that samples that lie on the same level set will be mapped to same value z.

In other words, the samples that lie on the same level sets of $g(x,\theta)$ are indistinguishable in the range of the perceptron. Next, we discuss the case of having multiple perceptrons, i.e., a multilayer perceptron with activations.

Multilayer (deep) NNs

To obtain a hierarchical (multilayer) model, the concept of a perceptron can be applied recursively. As before, let Θ^1 and Θ^2 correspond to two matrices whose rows contain a set of projection vectors (different θs in the preceding section): e.g., $\Theta^1 = [\theta_1^{1T}, \theta_2^{1T}, \ldots]$, where θ_i^{1T} is the transpose of the projection vector corresponding to the ith node/perceptron in layer 1. A two-layer NN model can be written as $\sigma(\Theta^2\sigma(\Theta^1)x)$. Expanding the idea further, we then may define a general formula for a K-layer NN as

$$g(x,\theta) = \sigma(\theta_1^K \cdot \sigma(\Theta^{K-1}\sigma(\Theta^{K-2} \times (\ldots\sigma(\Theta^1 x))))). \qquad (12)$$

Note that θ_1^K in (12) refers to a column vector, which collapses the output of the NN to one node, and that $\Theta^k = [\theta_1^k, \ldots, \theta_{L_k}^k]$, where L_k is the number of neurons in the kth layer of a deep NN.

The set of functions defined by such a multilayer perceptron, $g(\cdot,\theta)$ in (12), provides a GRT for the input distribution, p_X. Note, however, that the invertibility of such a transformation depends on the choice of the activation function σ (i.e., it should satisfy the necessary conditions enumerated before). Therefore, from the definition in (1), we have that the distribution over the output node p_{f_θ} can be considered as a projection (slice) of the GRT of p_X evaluated at θ: $p_{f_\theta} = \mathcal{G}p_X(\cdot,\theta)$, with

$$\mathcal{G}p_X(z,\theta) = \int_X p_X(x)\delta(z - \sigma(\theta_1^K \cdot \sigma(\Theta^{K-1} \ldots \sigma(\Theta^1 x))))dx. \qquad (13)$$

Figure 4 (c) and (d) demonstrates the level sets and the generalized slices (i.e., the line integrals) of pX using a multilayer perceptron as $g(x,\theta)$. Figure 4(c) is initialized randomly, and Figure 4(d) shows $g(x,\theta)$ after the network parameters are trained in a supervised classification setting to discriminate the modes of the half-moons distribution. It can be seen that, after training, the level sets $H(z,\theta)$ traverse only a single mode of the half-moons distribution, which indicates that the samples from different modes are not projected onto the same point (i.e., the distribution is not integrated across different modes).

The facility with which NNs generate highly nonlinear functions becomes readily apparent, even with relatively few parameters. (In later sections, we compare these to other polynomials.) We note that with just one layer, NNs can form nonlinear decision boundaries, as the superposition of surfaces formed by $\sigma(\theta_1^1 \cdot x) + \sigma(\theta_2^1 \cdot x) + \cdots$ can add curvature to the resulting surface. Note also that, generally speaking, the

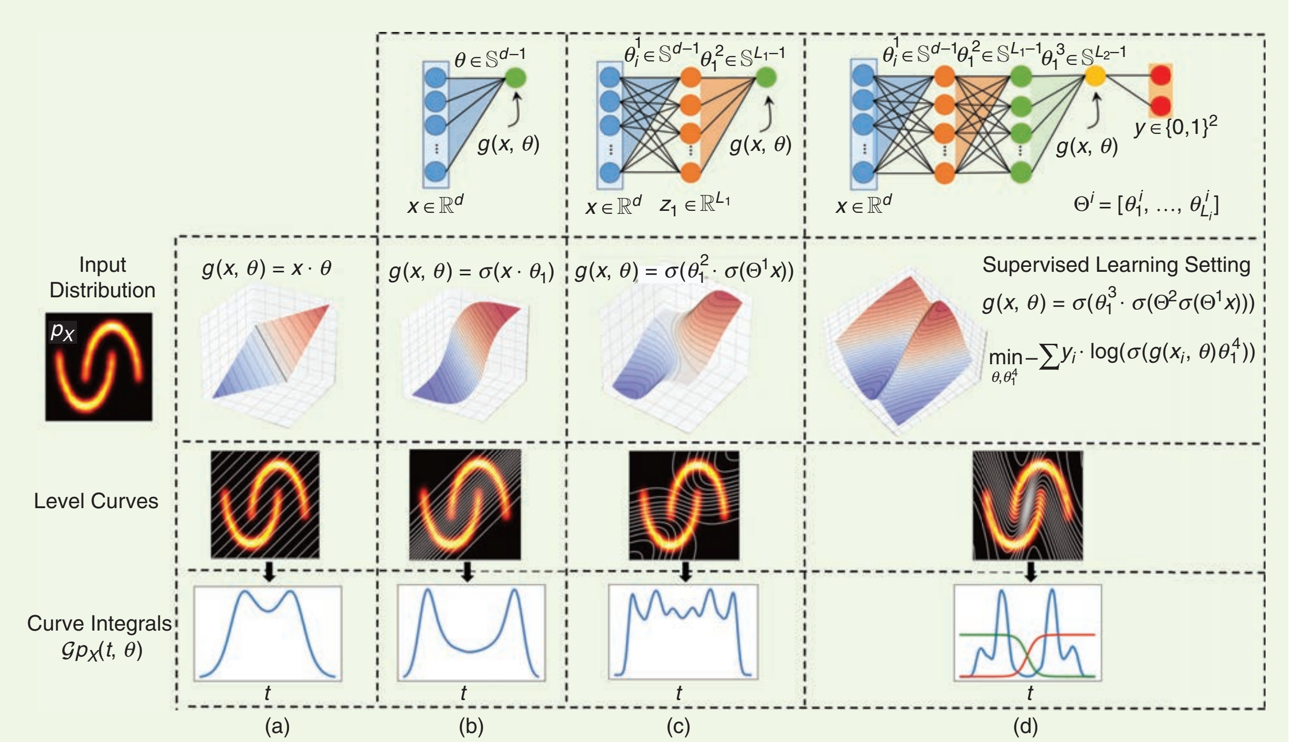


FIGURE 4. The relationship between NNs with different numbers of layers and the types of projections they compute. Curve integrals for the half-moons data set for (a) a random linear projection, which is equivalent to a slice of the classic Radon transform; (b) one-layer perceptron + activation with random initialization, which is isomorphic to the linear projection; (c) a two-layer perceptron with random initialization; and (d) a trained multilayer perceptron (with activations). The weights of all perceptrons are forced to be normalized so that $\|\theta_i^j\| = 1$. As one can see, adding nonlinear projections yields curved projection paths, and the objective of training is to discover the integration path that best separates the two distributions.

integration streamlines (hypersurfaces for higher-dimensional data) have the ability to become more curved (nonlinear) as the number of layers increases. Figure 5 compares linear projections, circular projections, a homogeneous polynomial of degree five, and an ANN of depth one, all trained to minimize the logistic regression cost function. While it is clear that linear and circular projections don't have enough "flexibility" to solve the separation problem, a polynomial degree of degree five seems to emulate the behavior of an ANN of depth one. It is possible that in the future, the point of view provided by analyzing the nonlinear projections associated with different NNs can provide inspiration for alternative models as well as insights with which to understand different properties of NNs.

Activation functions

It has been noted recently that NNs [e.g., convolutional NNs (CNNs)] can, at times,

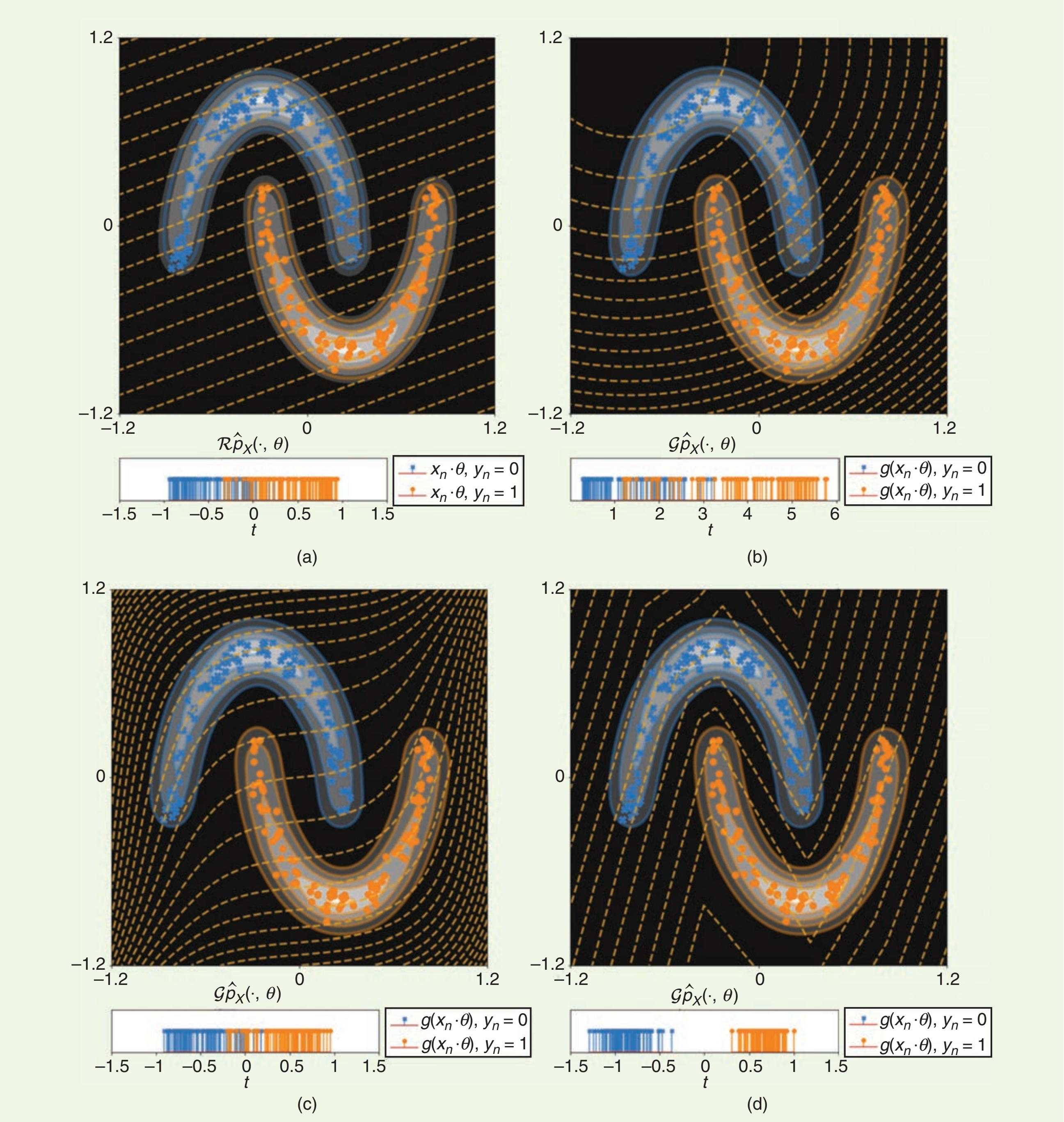


FIGURE 5. The optimally discriminant curvilinear projections, with different defining function $g(\cdot,\theta)$: (a) linear (i.e., the standard Radon transform), (b) circular, (c) homogeneous polynomial of degree five, and (d) a multilayer perceptron with LeakyReLU activations. The bottom panel for each quadrant shows the empirical generalized Radon projections, $(1/N)\sum_{n=1}^{N}\delta(t-g(x,\theta))$. Also, for an NN with LeakyReLU activations, note that the hypersurfaces (shown in orange) demonstrate some kinks, which are due to nondifferentiability at the origin.

work better when σ is chosen to be the rectified linear unit (ReLU) function, as opposed the sigmoid option [5]. The experience has encouraged others to try different activation functions, such as the LeakyReLU [5]. While theory describing which type of activation function should be used with which type of learning problem has yet to emerge, the interpretation of NNs as nonlinear projections can help highlight the differences between activation function types. Specifically, Figure 6 can help visualize the effects of different activation functions on the integration geometry over the input data space X.

First, note that the ReLU is a noninvertible map, given that negative values all map to zero. This will cause the surface generated by a perceptron constructed with ReLU to have a region over X that is flat, whereby all points in that region are integrated and mapped to the same value (zero) in the output space. This ability may provide ReLU NNs with the flexibility to construct adaptable characteristic function-type models for different regions in the data space. However, the outcome of the optimization procedure will dictate whether such regions would emerge in the final model. Finally, note that both ReLU and the LeakyReLU activation functions contain nondifferentiable points, which are also imparted on the surface function (hence, the sharp "kinks" that appear over isosurfaces lines).

Pooling

Pooling (e.g., average or maximum sampling) [5] operations are typically used

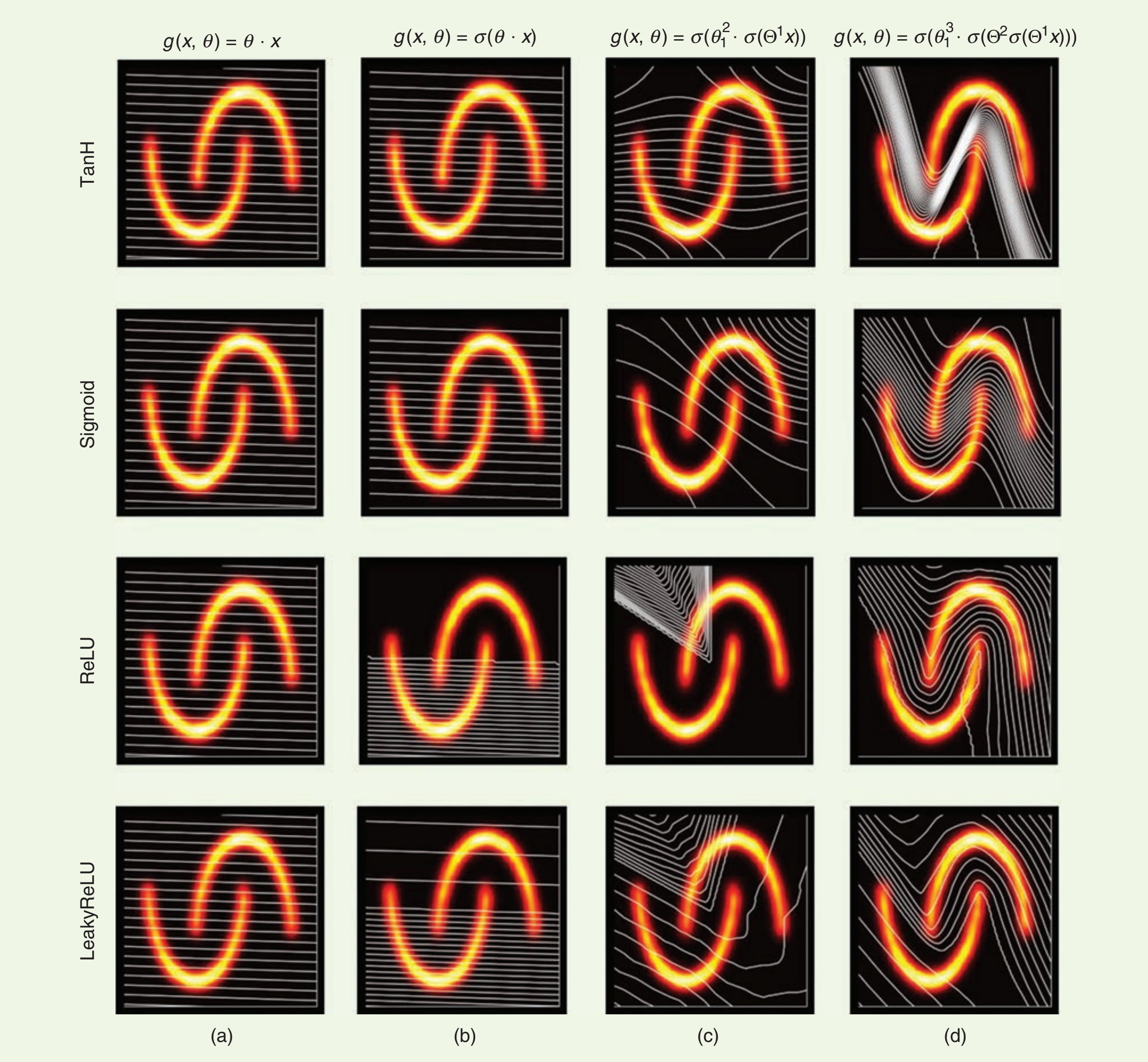


FIGURE 6. The level curves of $g(x,\theta)$ introduced by different activation functions. Parameters $\theta_1 \in \mathbb{R}^2$, $\Theta^1 \in \mathbb{R}^{50\times 2}$ and $\theta_1^2 \in \mathbb{R}^{50}$ are randomly initialized (with the same seed) for demonstrations in (a)–(c). Parameters for (d) $\Theta^1 \in \mathbb{R}^{50\times 2}$, $\Theta^2 \in \mathbb{R}^{100\times 50}$, $\theta_1^3 \in \mathbb{R}^{100}$ are optimized by minimizing a binary classification loss. It is clear that ReLU-type functions yield generating functions with sharp transitions, which, in turn, yield integration paths that are more "flexible" than sigmoid and TanH activation functions.

in large NNs, especially CNNs. The reasons are often practical, as subsampling can be used as a way to control and reduce the dimension (number of parameters) of the model. Another often-stated reason is that pooling can also add a certain amount of invariance (e.g., translation) to the NN model. In the case of average pooling, it is clear that the operation can also be written as a linear operator Θ^k in (12), where the pooling operation can be performed by replacing a particular row of Θ^k by the desired linear combination between two rows of Θ^k, for example.

Maximum pooling (max pooling), on the other hand, selects the maximum surface value (perceptron response), over all surfaces (each generated by different perceptrons) in a given layer. Figure 7 shows a graphical description of the concept, though it should be noted that, as defined earlier, the integration lines are not being added. Rather, the underlying "level" surfaces are being added. This illustration (see accompanying code) can help clarify the role of nonlinear pooling operations in introducing curves in the projection/slice being computed.

Adversarial examples

It has often been noted that highly flexible nonlinear learning systems, such as CNNs, can be "brittle" in the sense that a seemingly small perturbation of the input data can cause the learning system to produce confident, but erroneous, outputs. Such perturbed examples are often termed *adversarial examples*. The fields of adversarial attacks, i.e., generating adversarial examples, and defenses against such attacks have attracted increasing attention from the community in recent years.

Figure 8 utilizes the integral geometric perspective described previously to provide a visualization of how NNs (as well as other classification systems) can be fooled by small perturbations. To find the minimum displacement that could cause misclassification, using the blue point as the starting point x_0, we perform gradient ascent $x_{n+1} = x_n + \gamma \nabla g(x_n, \theta)$ until we reach the other side of the decision boundary (which is indicated by the orange point). We limit the magnitude of the displacement small enough so that the two points belong to the same distribution.

However, once integrated along the isosurfaces corresponding to the NN drawn in Figure 8, due to the uneven curvature of the corresponding surface, the two points are projected onto opposite ends of the output node, thus fooling the classifier into making a confident, but erroneous, prediction. This geometric interpretation further sheds light on the importance of regularization techniques on NN parameters to avoid obtaining hypersurfaces with high curvature.

What we have learned

In this note, we showed that the PDF associated with the numerical output z of any perceptron or node [call it $p_{f_\theta}(z)$] within a hierarchical NN can be interpreted as a particular hypersurface integration operation over the distribution p_X (the PDF of the input data). More specifically, we've learned that by combining (2) from the "Problem Statement: Statistical Regression and Classification" section (transformation of random variables) with (12) from the "NNs and the GRT" section (formula for an NN), we have

$$p_{f_\theta}(z) = \mathcal{G}p_X(z, \theta),$$

where $\mathcal{G}p_X(z, \theta)$ is given in (10) and specifies the GRT [12]. The conditions on the NN architecture ensuring this equation are outlined in the "The GRT" section.

Geometrically, one can interpret p_{f_θ} as a projection (path integral) of p_X, where the path is specified by the isosurfaces of $f_\theta(x)$ (node formula). The projections can be computed over straight lines (in the

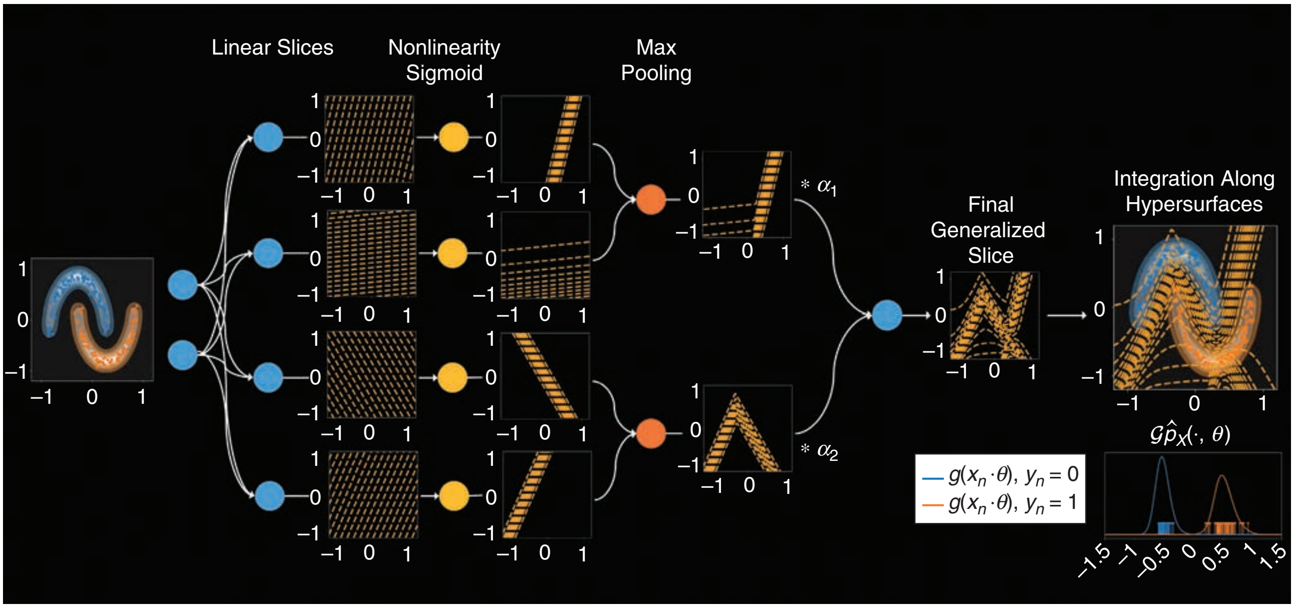


FIGURE 7. A demonstration of the max-pooling operation: an illustration of how max-pooling operations can help reduce the dimension of the output, while at the same time introducing curvature in the curvilinear projection/slice being computed. The level surfaces corresponding to perceptron outputs for a given input sample x are selected for maximum response. (See text for more details.) Also, note that the projections in inner nodes of an NN carry essential information about the input distribution; however, modern applications often discard this information.

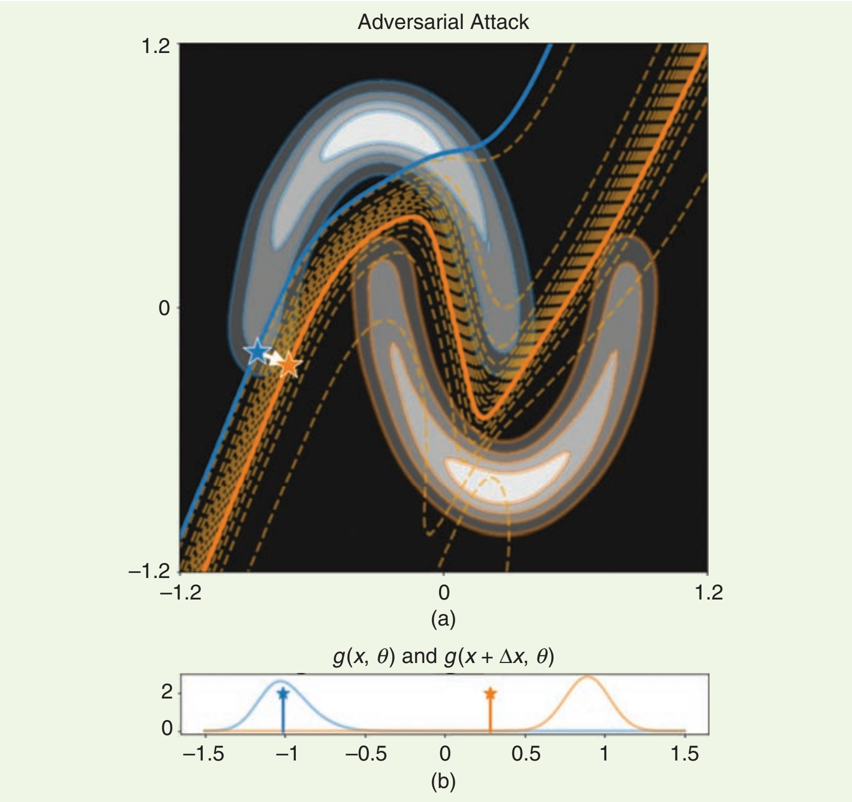


FIGURE 8. (a) The discriminant hypersurfaces for the half-moons data set are shown with dotted orange lines, where the defining function is a deep NN. The blue star indicates an unperturbed input sample from the upper half moon distribution. The white arrow indicates the adversarial perturbation, which is obtained from a gradient ascent path. We can see that the adversarial perturbation leads to a shift between hypersurfaces, such that the smallest perturbation in the input space causes a considerable change in the curvilinear projections. (b) The output distribution (i.e., a slice of the GRT), where colors indicate the conditional distributions. The blue and orange bars in show the output of the network before and after applying the adversarial perturbation.

case of simple linear logistic regression) or "curved" hypersurfaces in the case of a deep NN. The analogy has limitations, however, given that the GRT formalism [12] depends on generating functions that satisfy certain properties (e.g., nondegeneracy). Depending on the choice of activation function, for example, these may or may not be satisfied.

Despite these limitations, the analogy is useful to provide an alternative interpretation of NNs as operators that project the PDF of the input data over specific hypersurfaces generated by the level sets of the function for each node. Specifically, we highlight the following "take-home" points.

Linearity

We highlight that while the multilayer NN function $f_\theta(x)$ is nonlinear when the activation function σ being used is nonlinear, the PDF of the numerical output of a node [denoted as $p_{f_\theta}(z)$] can be obtained by applying a linear operator (the GRT) to the input data PDF $p_X(x)$. To see that this is a linear relationship, let $p_1(x)$ and $p_2(x)$ define two distinct PDFs for the same data space $x \in X$. Define a convex combination between p_1 and p_2 such as $p_X(x) = \alpha p_1(x) + (1 - \alpha) p_2(x)$, $\alpha \in [0, 1]$. It is easy to show [simply plug the definitions into equation (2)] that $p_{f_\theta}(z) = \alpha \mathcal{G} p_1(z, \theta) + (1 - \alpha) \mathcal{G} p_2(z, \theta)$. Thus, whether the classifier is considered as linear (in the case analogous to the standard Radon transform) or as a nonlinear NN (in the case analogous to the GRTs), they are both linear operations when the numerical PDF of output nodes is considered.

Geometric interpretation of NN properties

The interpretation of an ANN node output PDF as an integration over hypersurfaces defined as level sets of the node function f_θ can be used to demystify different properties of ANNs. Earlier, we showed how different types of activation functions (e.g., ReLU, sigmoid, and others) can affect the geometry of the hypersurfaces involved. ReLU-type activation functions, for instance, can cause sharp changes (kinks) along the integration path. We also discussed how the pooling operation can introduce new features in the isosurface map and can themselves add different types of nonlinearities to the aforementioned isosurfaces. Finally, we can also use this geometric interpretation to explain how opportunities for adversarial attacks may come from NNs that generate high-curvature isosurfaces that fail to approximate the input data PDF $p_X(x)$, leading to high-confidence but erroneous predictions.

Inspiring future developments

It is common to focus on NN output nodes as the main objective of an ANN, as they identify the boundaries between classes (assuming successful training). As demonstrated here, the information content of other nodes in an ANN can be interpreted geometrically. This connection could be useful, as it could lead to new methods for maximum likelihood classifiers based on density estimation as the GRT theory may provide ways to "reconstruct" information about $p_x(x)$ from multiple "projections" $p_{f_{\theta_1}}(z), p_{f_{\theta_2}}(z), \ldots$.

Yet another benefit of the link between ANNs and the mathematics of the GRT is that it can provide a useful framework for devising and applying projection-based distances of statistical distributions. Wasserstein-based distances (optimal transport, earth movers), for example, were recently proposed as metrics for a variety of data analysis problems [8]. ANNs (e.g., multilayer perceptrons, CNNs) could be used, thus, as a set of "trained" projections that one could use to compare different data distributions for solving problems related to transfer learning and others.

Acknowledgments

This work was supported in part by National Institutes of Health award GM130825.

Authors

Soheil Kolouri (skolouri@hrl.com) received his B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2010, and his M.Sc. degree in electrical engineering from Colorado State University, Fort Collins, in 2012. He received his Ph.D. degree in biomedical engineering from Carnegie Mellon University, Pittsburgh, Pennsylvania, in 2015, where he received the Bertucci Fellowship Award for outstanding graduate students from the College of Engineering in 2014 and the Outstanding Dissertation Award from the Biomedical Engineering Department in 2015. He is currently a research scientist and a principal investigator at HRL Laboratories, Malibu, California.

Xuwang Yin (xy4cm@virginia.edu) received his B.S. and M.S. degrees in computer engineering from the University of Science and Technology Beijing, China, in 2011 and 2013, respectively. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of Virginia. His research interests include adversarial machine learning and generative modeling.

Gustavo K. Rohde (gustavo.rohde@gmail.com) received his B.S. degree in physics and mathematics in 1999 and his M.S. degree in electrical engineering in 2001 from Vanderbilt University, Nashville, Tennessee. He received his Ph.D. degree in applied mathematics and scientific computation in 2005 from the University of Maryland, College Park. He is currently an associate professor of biomedical engineering and electrical engineering at the University of Virginia, Charlottesville, where he leads the imaging and data science laboratory.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539.

[2] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989. doi: 10.1007/BF02551274.

[3] F. Cucker and S. Smale, “On the mathematical foundations of learning,” *Bull. Amer. Math. Soc.*, vol. 39, no. 1, pp. 1–49, 2002. doi: 10.1090/S0273-0979-01-00923-5.

[4] D. T. Gillespie, “A theorem for physicists in the theory of random variables,” *Amer. J. Phys.*, vol. 51, no. 6, pp. 520–533, 1983. doi: 10.1119/1.13221.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[6] S. M. Carroll and B. W. Dickinson, “Construction of neural nets using the Radon transform,” in *Proc. Int. 1989 Joint Conf. Neural Networks*, pp. 607–611. vol. 1. doi: 10.1109/IJCNN.1989.118639.

[7] Y. Ito, “Differentiable approximation by means of the Radon transformation and its applications to neural networks,” *J. Comput. Appl. Math.*, vol. 55, no. 1, pp. 31–50, 1994. doi: 10.1016/0377-0427(94)90183-X.

[8] S. Kolouri, K. Nadjahi, U. Simsekli, R. Badeau, and G. K. Rohde, “Generalized sliced Wasserstein distances,” in *Proc. Advances Neural Information Processing Systems*, 2019, pp. 261–272.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin: Springer Science+Business Media, 2006.

[10] J. Radon, “Uber die bestimmug von funktionen durch ihre integralwerte laengs geweisser mannigfaltigkeiten,” *Berichte Saechsishe Acad. Wissenschaft. Math. Phys., Klass*, vol. 69, p. 262, Apr. 1917.

[11] L. Ehrenpreis, *The Universality of the Radon Transform*. London: Oxford Univ. Press, 2003.

[12] A. Homan and H. Zhou, “Injectivity and stability for a generic class of generalized Radon transforms,” *J. Geom. Anal.*, vol. 27, no. 2, pp. 1515–1529, 2017. doi: 10.1007/s12220-016-9729-4.

13. rohdelab, “radon-neural-network,” GitHub. Accessed on: 2020. [Online]. Available: https://github.com/rohdelab/radon-neural-network

---

IMAGE LICENSED BY INGRAM PUBLISHING

László Sujbert, Gyula Simon, and Gábor Péceli

An Observer-Based Adaptive Fourier Analysis

Recently, an observer-based algorithm was reviewed that is able to perform recursive sliding discrete Fourier transform (SDFT) [1]. DFT is the basic algorithm used for Fourier analysis [2], but it provides unbiased estimates only if the sampling is coherent, i.e., the ratio of the sampling frequency f_s and the input signal frequency f_{in} is a rational number. Coherent sampling of real-life periodic signals (e.g., mains voltage) can only be achieved if f_{in} and f_s are synchronized.

Noncoherent sampling results in picket fence effect and leakage. Such phenomena can be reduced by windowing [3] but not completely suppressed. The problem is especially interesting if speech or music signals are to be processed. Another important application is order analysis or order tracking [4], where the spectral components are calculated as a function of the harmonic indices, rather than frequency. Order analysis is often applied to vibration signals of rotating machinery with a changing speed of revolution. The problem is generally solved by using digital resampling [5], [6].

In this article, we show that the observer-based approach, similar to [1], can be used for the analysis of noncoherently sampled periodic signals as well as for order tracking. To achieve this, only the so-called basis vectors of the DFT are to be substituted by harmonic complex exponentials having the same fundamental frequency as that of the input signal. However, this fundamental frequency is generally unknown; thus, it has to be estimated. We show that the fundamental frequency can be estimated using the observer itself, resulting in a nonlinear observer that is able to estimate both the fundamental frequency and the spectrum of the input signal simultaneously. This observer is called the *adaptive Fourier analyzer* (*AFA*).

The AFA was first published in [7] for the case of unknown but constant fundamental frequencies. Later, in [8], the observer has been extended for the error-free reconstruction of periodic signals of linear, logarithmic as well as hyperbolic time–frequency function.

The novelty of our article is that it introduces the design of the AFA in detail, addresses its stability issues, and compares it to other methods. As an illustration, the order analysis problem is used, where periodic signals with constant and linearly changing fundamental frequencies are to be analyzed. Additionally, the analysis of a violin sound record is presented. Through the analysis of these signals, we demonstrate that the AFA is a competitive tool if periodic signals of arbitrary constant or changing frequency are to be analyzed.

Observer-based FA

Observer-based recursive SDFT

First, we briefly review the observer-based algorithm as it appears in [1]. Its detailed derivation can be found in [9]. The DFT of a signal $x[n]$ can be defined as

$$X_k = \sum_{n=0}^{N-1} x[n]\, W_N^{-kn}, \; k = 0 \dots N-1, \quad (1)$$

where the rotating unit is defined by

$$W_N = e^{j\frac{2\pi}{N}}, \quad (2)$$

with j being the imaginary unit. The SDFT can be defined using a slight modification of the indices:

$$X_k^C[n+1] = \sum_{m=0}^{N-1} x[q+m]\, W_N^{-km}, \quad k = 0 \dots N-1, \quad (3)$$

where $q = n - N + 1$. The SDFT result differs from the DFT result only in phase, and they are identical at every N step, i.e.,

$$X_k^C[n] = X_k, \; k = 0 \dots N-1, \quad n = 0, N, 2N, \dots, \quad (4)$$

The root of the observer-based approach is the so-called conceptual signal model:

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_k c_k[n], \quad (5)$$

where $c_k[n]$ is the kth basis vector and X_k is its weighting factor. The notation

---

Digital Object Identifier 10.1109/MSP.2020.2982167
Date of current version: 26 June 2020

1053-5888/20©2020IEEE

X_k is intentionally used, as it equals the result of the DFT if

$$c_k[n] = W_N^{kn}, \; k = 0 \dots N-1, \qquad (6)$$

i.e., if the basis vectors consist of coherent samples of complex exponentials. The weighting factors in the signal model are state variables, and the corresponding observer is designed to follow these state variables. The equations of the observer are

$$\begin{aligned} y[n] &= \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_k[n]c_k[n] \\ &= \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_k^C[n] \\ \hat{X}_k[n+1] &= \hat{X}_k[n] + g_k[n](y[n] - x[n]), \\ & k = 0 \dots N-1, \end{aligned} \qquad (7)$$

where $y[n]$ denotes the reconstructed signal, and the operator $\hat{}$ stands for estimation. The functions $g_k[n]$ are the reciprocal basis vectors that can be expressed easily if (6) holds:

$$\begin{aligned} g_k[n] &= \bar{c}_k[n] = W_N^{-kn}, \\ k &= 0 \dots N-1, \end{aligned} \qquad (8)$$

where the bar denotes the complex conjugation. The derived observer, shown in Figure 1, is deadbeat and reaches steady state in N steps. In steady state, $\hat{X}_k[n] = X_k$ $(k = 0 \dots N-1)$ and $y[n] = x[n]$, so the error signal $e[n]$ also equals zero. The figure clearly shows the connection between the DFT X_k and SDFT results $\hat{X}_k^C[n]$.

Observer for noncoherently sampled periodic signals

First, the signal model is to be redefined. The conceptual signal model can be described by using (5), with a slight modification of the indices:

$$x[n] = \frac{1}{N}\sum_{k=-L}^{L} X_k c_k[n]. \qquad (9)$$

In this case, the weighting factors X_k are the Fourier coefficients. The $c_k[n]$ and $g_k[n]$ values, generally, are not coherent samples:

$$\begin{aligned} c_k[n] &= e^{j2\pi f_1 kn} = W_1^{kn}, \\ g_k[n] &= \bar{c}_k[n], \; k = -L \dots L, \end{aligned} \qquad (10)$$

where f_1 is the fundamental (relative) frequency of $x[n]$. Now, the number of the components is $N = 2L+1$, and W_1 is a new rotating unit. As the periodic signal $x[n]$ is band limited, the number of harmonic components is limited in the following way:

$$Lf_1 < 0.5 < (L+1)f_1. \qquad (11)$$

Note that no component at the half of the sampling rate (relative frequency $f = 0.5$) is modeled. For real periodic signals, it is not a restriction. The structure of the observer is the same as that which can be seen in Figure 1, and the system is also described by using (7).

In steady state, the state variables of the observer equal those of the signal model, i.e., $\hat{X}_k[n] = X_k$ $(k = -L \dots L)$, so the input signal is perfectly reconstructed. Now the steady state is reached in an infinite number of steps, but if (11) also holds, the system is fairly fast.

The observer provides the Fourier coefficients of the input signal, the only requirement is that the fundamental frequency has to be known. From now on, the general observer for periodic signals without the ability to estimate the fundamental frequency will be labeled as the *FA*.

Transfer functions

There are some transfer functions that play an important role in AFA design. Each internal channel of the observer performs a down-converting—integrating—up-converting sequence and has the following transfer function:

$$\begin{aligned} H_k(z) &= \frac{\hat{X}_k^C(z)}{E(z)} = \frac{W_1^k z^{-1}}{1 - W_1^k z^{-1}}, \\ k &= -L \dots L. \end{aligned} \qquad (12)$$

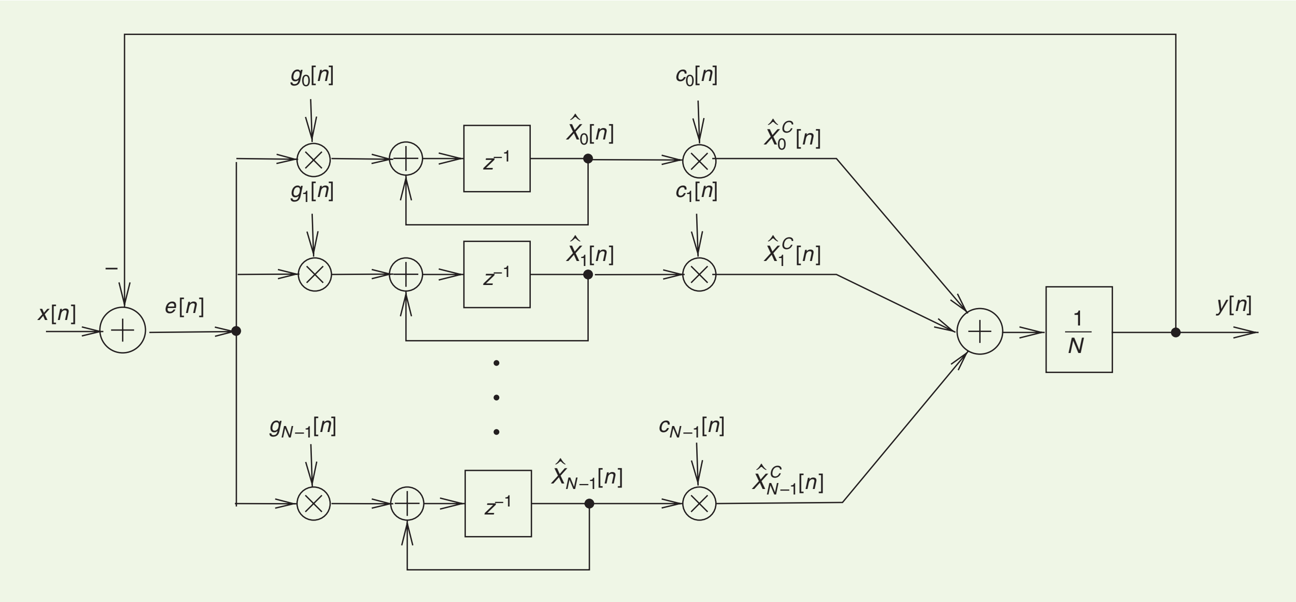


FIGURE 1. An observer for periodic signals.

The expressions $\hat{X}_k^C(z)$ and $E(z)$ denote the z transform of the signals $\hat{X}_k^C[n]$ and $e[n]$, respectively. The gain of $H_k(z)$ is infinite if the input of the channel has a frequency equal to the frequency of the basis function $c_k[n] = W_1^{kn}$. Thus each internal channel of the observer is a resonator, and the frequency of $c_k[n]$ is the so-called resonator frequency. The closed-loop transfer function from the input signal $x[n]$ to the signal component $\hat{X}_k^C[n]$ is

$$H_{k,\mathrm{obs}}(z) = \frac{\hat{X}_k^C(z)}{X(z)} = \frac{H_k(z)}{1+\frac{1}{N}\sum\limits_{i=-L}^{L} H_i(z)}, \quad k=-L\ldots L, \tag{13}$$

where $\hat{X}(z)$ is the z transform of the input signal $x[n]$. The closed-loop transfer function from the input signal $x[n]$ to the error signal $e[n]$ is also important:

$$H_e(z) = \frac{E(z)}{X(z)} = \frac{1}{1+\frac{1}{N}\sum\limits_{k=-L}^{L} H_{k,\mathrm{obs}}(z)}. \tag{14}$$

When sampling is coherent, i.e., the observer performs the SDFT, the aforementioned transfer functions have a much simpler form. Equation (13) can be reduced as follows:

$$H_{k,\mathrm{obs}}(z) = (1-z^{-N})\frac{W_N^k z^{-1}}{1-W_N^k z^{-1}}, \quad k=-L\ldots L, \tag{15}$$

and its magnitude response is

$$|H_{k,\mathrm{obs}}(f)| = \left|\frac{\sin \pi N(f-f_k)}{\sin \pi (f-f_k)}\right|, \quad k=-L\ldots L, \tag{16}$$

where f and f_k denote the frequencies relative to the sampling frequency, i.e., $f, f_k \in [0\ldots1]$. Note that the magnitude is zero at all of the resonator frequencies, with the exception of $f = f_k$, where it equals N. For coherent sampling, the error-transfer function (14) is also much simpler:

$$H_e(z) = 1 - z^{-N}, \tag{17}$$

and its magnitude response is

$$|H_e(f)| = 2|\sin \pi N f|. \tag{18}$$

Note that the gain is zero at all of the resonator frequencies, without exception. The magnitude responses of $(1/N)H_{1,\mathrm{obs}}(f)$ and $H_e(z)$ for the $N = 255$ case are depicted in Figure 2 with blue and red lines, respectively.

In the case of noncoherent sampling, no simple forms exist, but the main features of the magnitude responses are maintained: 1) $H_k(f)$ has infinite gain at the resonator frequency, 2) $H_{k,\mathrm{obs}}(f)$ has zeros at all of the resonator frequencies, with the exception of $f = f_k$, where its gain equals N, and 3) $H_e(f)$ has zeros at all of the resonator frequencies. If the condition (11) holds, the actual magnitude curves are very similar to those shown in Figure 2.

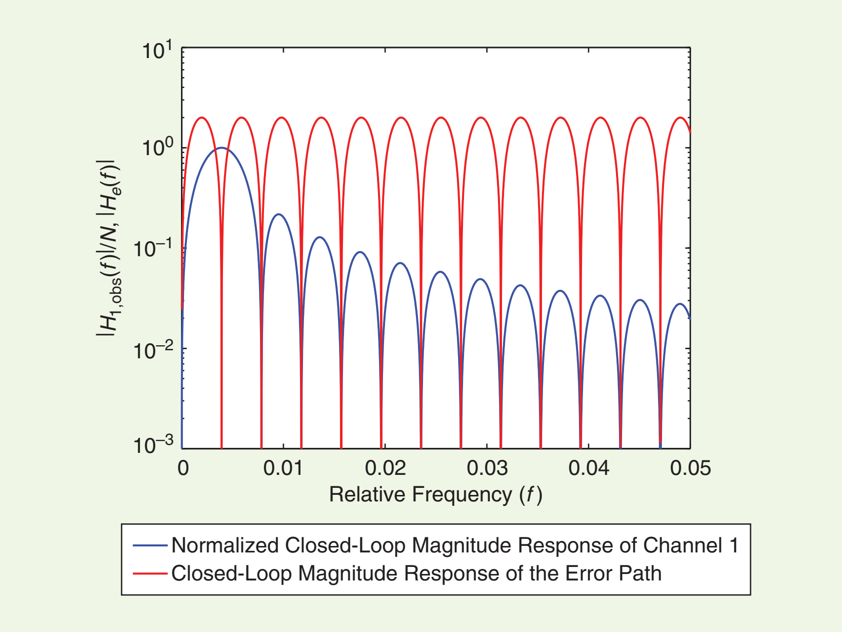


FIGURE 2. The closed-loop magnitude response of channel 1 normalized by N (blue curve) and the magnitude response of the error path (red curve) if $N = 255$ (zoomed-in view).

The AFA

Derivation of the AFA

The AFA primer with separate frequency estimator

To present the idea behind the operation of the AFA, an auxiliary system that estimates the fundamental frequency of the input signal is introduced, as shown in Figure 3. The procedure is well known and can be found, e.g., in [10]. Let us assume that we have an initial estimate of the fundamental frequency, f_1, but the input signal has a slightly different frequency, $f_{\mathrm{in}} \neq f_1$. First, the input signal is down-converted by f_1 and the resulted complex signal is filtered by $F(z)$, a narrow-band low-pass filter admitting only a single, complex exponential to the output. The operation is explained through the use of an illustrative signal $x[n]$, whose spectrum $X(f)$ is shown in Figure 4. The result of the operation is depicted in Figure 5. Down-converting shifts the spectrum to the left, and in steady state the signal components

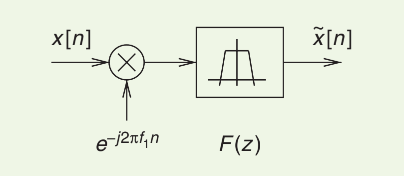


FIGURE 3. An auxiliary system for frequency estimation.

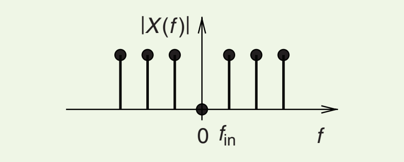


FIGURE 4. An example input signal for the derivation of the AFA.

are filtered out, with the exception of one, at $\Delta f = f_{\text{in}} - f_1$. The magnitude response of filter $F(z)$ is drawn with a dashed line, while the original and remaining amplitudes of the signal components are depicted by empty and filled circles, respectively. Under the spectrum, the signal $\tilde{x}[n]$ is plotted in the time domain. It is a rotating complex exponential with an approximate frequency of Δf:

$$\tilde{x}[n] \approx F(\Delta f)\cdot X_1 e^{j2\pi\Delta f n}, \tag{19}$$

where $F(\Delta f)$ is the filter response at $f = \Delta f$. The origin of the approximate equality is that other small signal components, attenuated in the nonideal stopband of $F(z)$, may also be present. These disturbing components are small rotating vectors superimposed to the end of the main vector, and their maximum range is depicted by a gray circle. Now the frequency estimator can be updated in the following way:

$$f_1' \approx f_1 + G\cdot\text{angle}(\tilde{x}[n], \tilde{x}[n-1]). \tag{20}$$

The function angle returns the angle between the succeeding samples of $\tilde{x}[n]$. There is a gain G that controls the speed of the adaptation; if $G = (1/2\pi)$, the frequency is updated, theoretically, in one step. Due to the disturbing signal components and inevitable noise, in practice, G is chosen to be a smaller positive constant. Now Fourier analysis can be performed in the following way: Both the auxiliary system and the FA receives the input signal $x[n]$, and the former passes the frequency estimate to the latter. Note that (19) is true in steady state only; thus, between the frequency update steps (20), enough time should pass to allow for the transient of filter $F(z)$ to decay.

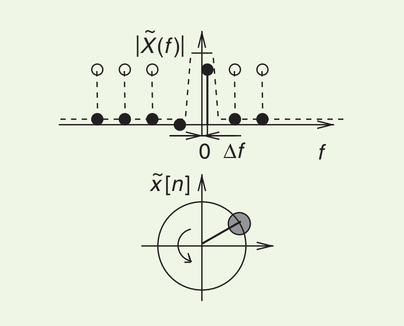


FIGURE 5. The operation of the auxiliary frequency estimator.

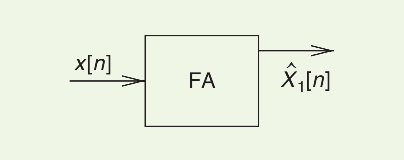


FIGURE 6. The frequency estimation by the FA.

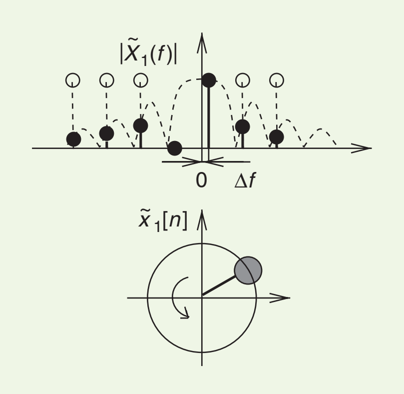


FIGURE 7. The signal transformation in the AFA.

Frequency estimation using the FA

Consider the FA derived in the "Observer for Noncoherently Sampled Periodic Signals" section with $\hat{X}_1[n]$ as the output, as presented in Figure 6. By observing the structure of Figure 1 and using (13), the output signal $\hat{X}_1[n]$ can be expressed as $\hat{X}_1[n] = H_{1,\text{obs}}(z) c_1^{-1}[n] x[n]$, where $c_1^{-1}[n]$ can be considered a down-converting function. The operation is illustrated again by the signal depicted in Figure 4, and the signal transformation is explained in Figure 7. It is assumed again that the current estimate of the fundamental frequency is f_1, but the input signal has a slightly different frequency, that is, $f_{\text{in}} \neq f_1$. The magnitude response of the filter is depicted in Figure 2 with a blue line and also appears in Figure 7 as a dashed line. Due to its down-converting, it is shifted to the left and acts as a sinc-like low-pass filter. The original and remaining amplitudes of the signal components are depicted again by empty and filled circles, respectively. Under the spectrum, the signal $\hat{X}_1[n]$ is plotted in the time domain. It is again a rotating complex exponential with an approximate frequency of Δf:

$$\hat{X}_1[n] \approx H_{1,\text{obs}}(f_{\text{in}}) \cdot X_1 e^{j2\pi\Delta f n}, \tag{21}$$

where $H_{1,\text{obs}}(f_{\text{in}})$ is the closed-loop response of the first channel at $f = f_{\text{in}}$. Due to its nonideal filtering, disturbing components may be present at the output, thus the equality is once again only approximate. The range of the disturbing components is depicted again by a gray circle. The frequency estimator can be updated in the following way:

$$f_1' \approx f_1 + G\cdot\text{angle}(\hat{X}_1[n+1], \hat{X}_1[n]). \tag{22}$$

Now the angle is calculated by the updated and previous state variable $\hat{X}_1[n]$. The frequency updates can be performed similarly to those of the AFA primer.

Frequency estimation using the FA seems to be advantageous, as no separate system is needed. However, filter $H_{1,\text{obs}}(z)$ may seem to be less selective than a well-designed low-pass filter $F(z)$, but in fact, it is very selective for the disturbing components. The (complex) spectral components of the signal $x[n]$ appear at the frequencies $k\cdot f_{\text{in}}, k = -L \ldots L$, while $H_{1,\text{obs}}(z)$ is exactly zero at the frequencies $k\cdot f_1, k = -L \ldots 0, 2, \ldots L$. As f_{in} differs only slightly from f_1, the higher harmonics of the signal are close to a zero of $H_{1,\text{obs}}(z)$, thus all of the spectral components of $x[n]$ are suppressed, with the exception of the fundamental one. Moreover, the closer the updated frequency estimator is to f_{in}, the better the suppression is of the disturbing components, and so is the estimation. Filter $F(z)$ cannot offer this feature. Nevertheless, it is obviously supposed that the estimation process is stable. For more details, see "Overcoming the Nonlinear Convergence Analysis Problem."

Algorithm of the AFA

The AFA is, in fact, the FA completed by the frequency estimation described previously, but there are some tricks in the design that make the system robust.

Overcoming the Nonlinear Convergence Analysis Problem

The adaptive Fourier analyzer (AFA) introduced in the "Algorithm of the AFA" section is a robust and reliable analysis tool used for periodic signals. Unfortunately, mainly due to the highly nonlinear nature of the AFA, until now, only partial convergence analysis results are available (see, e.g., [S1]). The FA itself is stable for any f_1, which can be proven using the Nyquist stability criterion (see, e.g., [S2]).

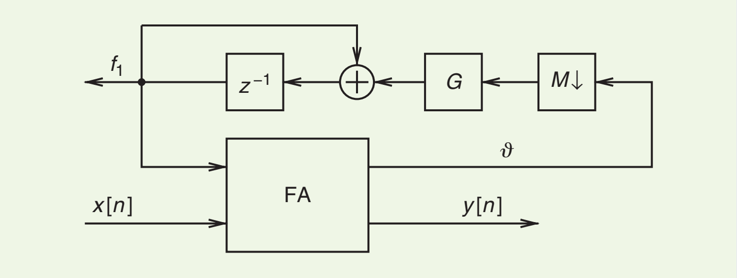


FIGURE S1. The frequency update in steady state.

In the AFA, the spectral components and frequency are estimated simultaneously. To overcome this problem, we first go back to the design described in the "Frequency Estimation Using the FA" section. There are two main assumptions: 1) The frequency update is made only in steady state of the FA and 2) the frequency of the input signal f_{in} differs only slightly from the actual estimate of f_1, i.e., $\Delta f = f_{\text{in}} - f_1 \ll f_1$. The frequency update is depicted in Figure S1. It is supposed that the FA reaches the steady state in M steps after each update of the frequency f_1. This way, the frequency update does not influence the convergence of the FA, thus the latter is stable for any f_1. In steady state, the rotation between the succeeding samples of $\hat{X}_1[n]$ gives the output ϑ. Note that ϑ is proportional to the frequency difference Δf if the estimated signal component is significantly higher than that of the other disturbing signal components. This is true if $\hat{X}_1[n]$ is not zero and if Δf is sufficiently small; thus, the disturbing signal components are suppressed by $H_{1,\text{obs}}(z)$. If the gain G is a sufficiently small positive constant, the averaging is stable.

Thus the system in Figure S1 is proved to be stable. The experiences show that the AFA, even without complying with the rules (update in steady state, and a small frequency difference) is stable in most cases. A special exception is when the fundamental component is missing or has much less power than do the other components (partials) of the signal. A signal processing engineer could be advised by displaying the error signal $e[n]$ of the observer. In the case of perfect reconstruction (i.e., a true estimate of the frequency), $e[n]$ tends to zero.

References

[S1] G. Simon and G. Péceli, "Convergence properties of an adaptive Fourier analyzer," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 46, no. 2, pp. 223–227, Feb. 1999. doi: 10.1109/82.752959.

[S2] K. J. Åström and B. Wittenmark, *Computer Controlled Systems.* Englewood Cliffs, NJ: Prentice Hall, 1990.

The first change is that the frequency estimator is updated at each time instant:

$$f_1[n+1] = f_1[n] + G\cdot\text{angle}(\hat{X}_1[n+1], \hat{X}_1[n]),\quad G = \frac{1}{2\pi N}. \tag{23}$$

The frequency estimator $f_1[n]$ now acts as a new state variable of the observer. As the frequency $f_1[n]$ is a variable, the functions $c_k[n]$ are updated differently from (10):

$$c_k[n+1] = c_k[n]e^{j2\pi f_1[n+1]k},\quad k = -L\ldots L. \tag{24}$$

Such an update ensures continuous phase transition. Function $g_k[n]$ is set accordingly:

$$g_k[n+1] = \bar{c}_k[n+1],\ k = -L\ldots L, \tag{25}$$

where the bar denotes the complex conjugation.

During the operation of the AFA, $f_1[n]$ can be far from its initial value and (11) does not hold any longer, i.e., either too many or too few components are present. To avoid this, after each frequency update, the numbers of the components are recalculated. Another trick is that the highest component cannot be close to the relative frequency $f = 0.5$, thus the limit is somewhat smaller:

$$Lf_1[n] < 0.5 - \frac{1}{2N} < (L+1)f_1[n]. \tag{26}$$

Such a rule warrants fast settling even if the highest component is near to the limit frequency. The components above the limit are simply to be canceled, while if new components are to be started, they are initialized as follows:

$$\hat{X}_L[n+1] = \hat{X}_{-L}[n+1] = 0$$
$$c_L[n+1] = c_{-L}[n+1] = 1. \tag{27}$$

The latter procedure represents an important feature of the AFA: Its structure is also adaptive, and the number of modeled components depends on the actual fundamental frequency.

Now the algorithm of the AFA can be summarized as follows:

- initialization: L is arbitrary, $N = 2L+1$, $f_1 = 1/N$
- $\hat{X}_k[0] = 0$, $c_k[0] = 1$, $g_k[0] = 1$, $k = -L\ldots L$
- operation by using (7) and (23)
- update of $c_k[n]$, $g_k[n]$, according to (24) and (25)

- cancellation or starting components, according to (27) and (26).

Download MATLAB reference implementations at http://mit.bme.hu/~sujbert/afa.

Improvements

The AFA introduced in the previous section is designed to reconstruct any periodic signal of constant fundamental frequency. If the frequency of the signal changes, there is an error in both frequency and spectral estimation. The AFA can follow the changes in frequency and the error can be small, but some applications require an exact analysis even if the frequency is not constant. The AFA has been improved to perfectly reconstruct the periodic signals of some typical time–frequency functions (sweep types) [8], such as linear, logarithmic, and hyperbolic. Linear sweep is the most common type, logarithmic sweep is useful, e.g., for acoustic measurements, while hyperbolic sweep is required by the vibration analysis of mechanical systems. Here the frequency estimation for linear sweep is recalled.

The frequency estimation (23) is completed by the estimation of the sweep rate:

$$\begin{aligned} f_1[n+0.5] &= f_1[n] + G\cdot\text{angle}\\ &\qquad(\hat{X}_1[n+1],\hat{X}_1[n])\\ v[n+1] &= v[n] + 0.4G\cdot f_1[n]\cdot\text{angle}\\ &\qquad(\hat{X}_1[n+1],\hat{X}_1[n])\\ f_1[n+1] &= f_1[n+0.5] + v[n+1], \end{aligned} \tag{28}$$

where variable $v[n]$ is the estimated sweep rate, and the coefficient 0.4 is set empirically to minimize the settling time. In this formula, $f_1[n+0.5]$ is an intermediate variable, as the frequency estimator is updated two times.

All of the AFA algorithms work properly if the frequency is constant. Unfortunately, if the time–frequency function is unknown, the reconstruction is not necessarily perfect; however, if the algorithm includes some modeling of the time–frequency function, e.g., applies linear frequency tracking, the performance will be reasonably better than it is for the constant frequency case.

Examples

First, the performance of the AFA is demonstrated by using an order-tracking example. The input signal is a bandlimited square wave made of the first four nonzero components. The duration of the signal is 3 s. In the first second, the frequency is 50 Hz, which linearly increases to 150 Hz between 1 and 2 s. In the last second, the frequency is constant, that is, 150 Hz. The sampling frequency is set to $f_s = 10$ kHz and the AFA is initialized using $f_1 = 1/255$, resulting in $L = 127$ and $N = 255$.

Alternative methods are also checked for the same problem. The order-tracking algorithm of Potter [(Hewlett–Packard (HP)] [5], [6] requires a tachometer signal that indicates the periods of the fundamental component (e.g., an impulse in each revolution of a rotating shaft), but the signal to be analyzed is sampled by a constant frequency. The spectrum is calculated by fast Fourier transform (FFT) of the signal, resampled at uniform angle increments: with N being the length of the FFT, samples at phase positions $\varphi = k(2\pi/N)$ $(k = 0 \ldots N-1)$ are calculated. The algorithm estimates a constant or linearly sweeping frequency by the tachometer pulses, while the signal is interpolated by a polynomial. In this example, $N = 256$, and the interpolation is made by a 10th-order polynomial. Note that the HP method needs an extra tachometer signal, while the AFA uses the input signal itself for frequency estimation. The performance of the AFA primer, presented in “The AFA Primer With Separate Frequency Estimator” section, is also tested using a 1,025-tap finite impulse-response low-pass filter.

Interpolated DFT can also be used as an alternative method for frequency estimation [11], [12]. Here the frequency is calculated as the linear combination of the frequencies near to the peak, weighted by the measured amplitudes. The method is also developed for windowed data; for the tests we used a Hanning window.

Adaptive comb (or adaptive notch) filters were proposed to estimate the spectra of periodic signals with an unknown frequency [13]. In our tests, the algorithm presented in *IEEE Signal Processing Magazine* [14] was used (the MATLAB code is available in [15]). The algorithm tunes the notch frequencies of cascaded second-order infinite impulse-response filters to coincide with those of the signal to be analyzed. In case of convergence, the fundamental frequency of the periodic signal is estimated by the first notch frequency. The estimation of the spectrum requires further steps, e.g., a recursive least-squares algorithm [13].

Figures 8 and 9 show the orders of the input signal calculated by the HP method and the AFA, respectively. Both methods produce a similar result: no picket fence nor leakage can be observed in steady state. Short transients can be observed in both cases at time instants $t = 0$, 1, and 2 s after initialization and when the linear sweep starts and stops. The quality of the estimation can be judged by the feedback

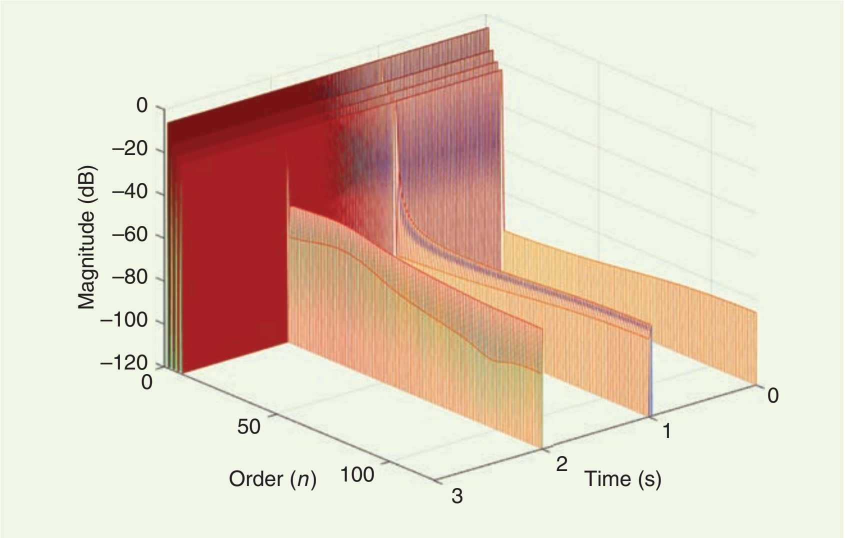


FIGURE 8. The orders estimated by the HP method.

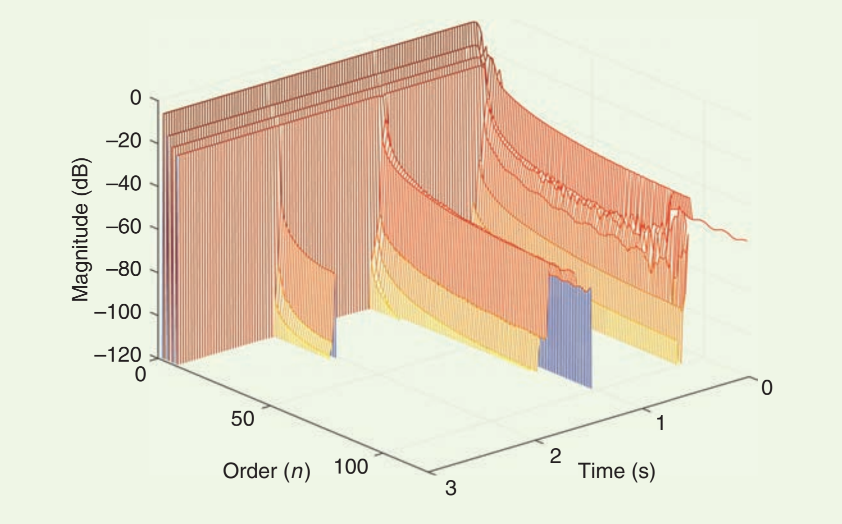


FIGURE 9. The orders estimated by the AFA.

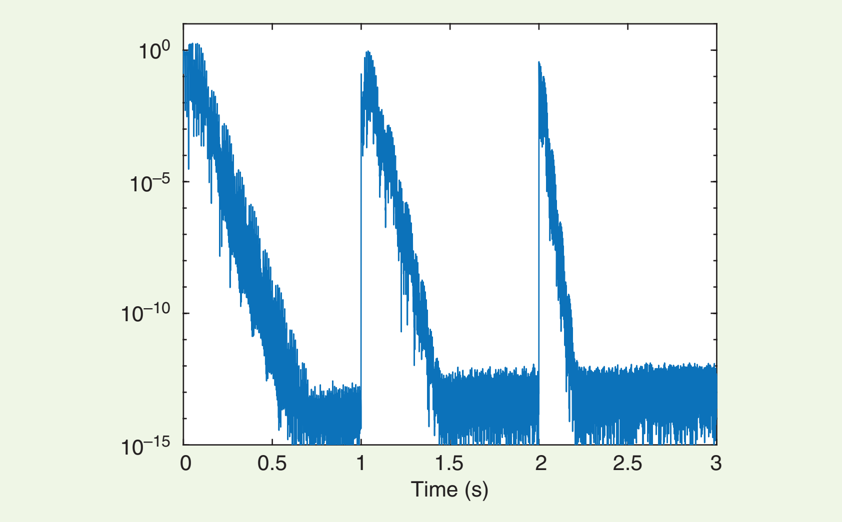


FIGURE 10. The AFA error signal.

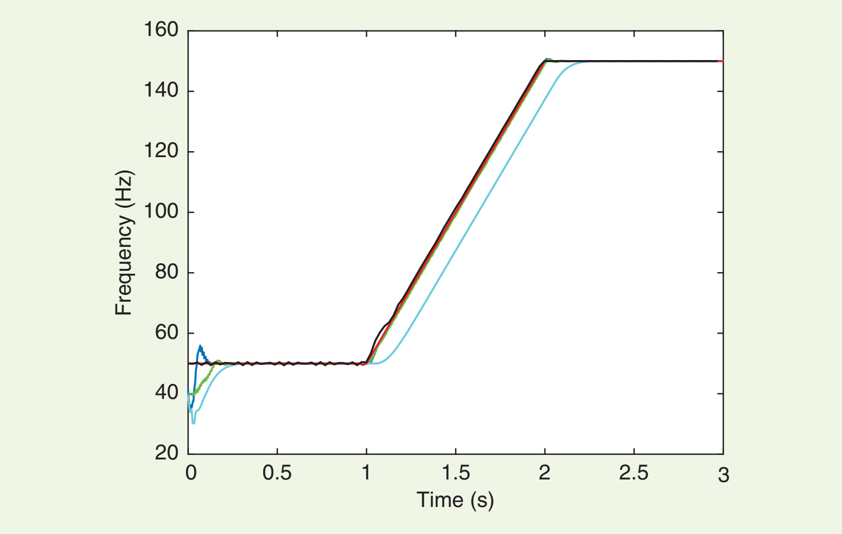


FIGURE 11. The frequency-estimation results: The HP method (red), AFA (blue), AFA primer (cyan), interpolated DFT (black), and adaptive notch filter (green).

error of the AFA depicted in the logarithmic scale shown in Figure 10.

The estimation of the fundamental frequency can be followed in Figure 11. The result of the HP method is drawn with a red line, while the estimation of the AFA is depicted by a blue line. Due to the tachometer pulses, the HP method is almost perfect. The AFA needs some time to settle, but after the transient the estimation is precise. The zoomed-in view for the $f = 50$-Hz period can be seen in Figure 12. An observant reader can also notice an estimation error of the AFA when the linear sweep starts at $t = 1$ s. During this transient the sweep rate is adapted to the correct value.

The estimations of the AFA primer and the interpolated DFT are drawn with cyan and black lines, respectively. The AFA primer is very accurate for the constant periods of frequency but has a long transient phase. This is because of the large delay of the filter, which is roughly 0.05 s. The tracking error for the linear sweep period is large, as the method is not prepared for sweeping periodic signals. The interpolated DFT has a fair estimation, but the error is greater than that of the other methods. Averaging would reduce the error, but it would result in a longer transient phase. The green line shows the frequency estimation of the adaptive notch filter. The quality of the estimation is similar to that of the AFA, but the notch filter has an observable bias in case of linearly changing frequency.

The statistical properties of the frequency estimators for the AFA, interpolated DFT, and adaptive notch filter are listed in Table 1. The mean error and standard deviation of the frequency estimation is calculated for the steady state in each time interval. In the first and third second, the adaptive notch filter and the AFA have no bias, but the interpolated DFT has a small error. In the sweeping phase, only the AFA can track the signal with no bias. As the test signal is noise-free, the standard deviation of the estimation is small for the adaptive notch filter and the AFA, while the estimation of the interpolated DFT varies slightly due to the leakage.

Next, the methods are compared to each other if the noisy input signal is to be analyzed. The input signal is burdened by Gaussian random noise with a signal-to-noise ratio of 20 dB. To be fair, the tachometer signal of the HP method is also noisy, having $dt = 20\ \mu\text{s}$ standard deviation in zero crosses. Note that this is 0.1% of the period time for the $f = 50$-Hz period. The result of the estimation can be seen in Figure 13. The transient phase of the AFA is similarly long as mentioned previously, and the estimation errors in the steady state are approximately equal for the three competitive methods. Table 2 summarizes the statistical data for the test. The standard deviation of the estimation is nearly equal for the adaptive notch filter and AFA, but the bias of the adaptive notch filter in the sweeping phase is clear.

Another example shows the analysis of a violin open E5 string sound [16] with a duration of approximately 3 s and a sampling frequency of 44.1 kHz. Both the sound level and frequency vary slightly in the record. The signal has a rich spectral content, as depicted in Figure 14. The solid black line is the result of the 1,024 FFT samples preceding time instant $t = 1.5$ s. The blue circles in the figure indicate the result of the AFA estimation at the same time instant. It can clearly be seen that the AFA has found the right magnitudes. Figure 15 shows the frequency-estimation curves. The progress of the frequency estimator of the AFA is plotted with a blue line, while that of the adaptive notch filter with a green line. Both of them could find the fundamental frequency of the sound, which is close to its nominal value, i.e., 659.25 Hz, indicated with a red line in the figure. Note that the estimator of the AFA started far from the final value, but the adaptive notch filter required a good initial guess to converge.

It can be concluded that the AFA is a competitive tool for the fundamental frequency estimation of periodic signals. Moreover, it provides the spectrum of the signal simultaneously.

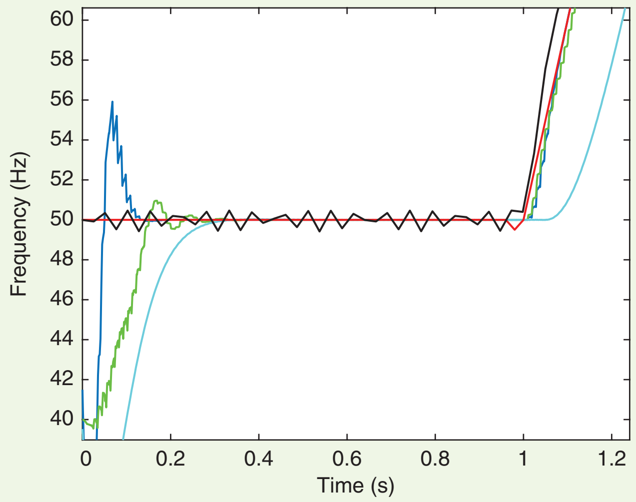


FIGURE 12. The frequency-estimation results: The HP method (red), AFA (blue), AFA primer (cyan), interpolated DFT (black), and adaptive notch filter (green) (zoomed-in view).

Table 1. The mean error and standard deviation of the frequency estimation of the noise-free test signal.

Method	Mean Error (Hz) First Second	Mean Error (Hz) Second Second	Mean Error (Hz) Third Second	Standard Deviation (Hz) First Second	Standard Deviation (Hz) Second Second	Standard Deviation (Hz) Third Second
Notch filter	1.4e-07	−6.4e-01	2.7e-10	2.7e-06	8.3e-02	2.0e-09
DFT	3.8e-02	1.1e+00	−2.6e-04	4.1e-01	2.8e-01	1.1e-02
AFA	−4.1e-16	1.6e-14	2.4e-15	3.1e-14	7.8e-13	8.3e-13

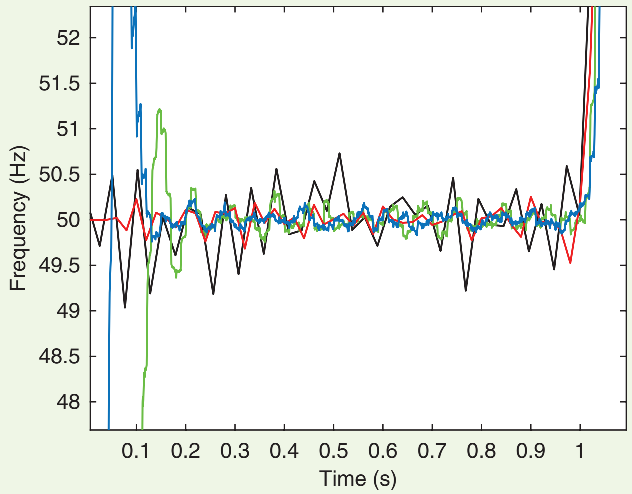


FIGURE 13. The frequency-estimation results for a noisy input signal: The HP method (red), AFA (blue), interpolated DFT (black), and adaptive notch filter (green) (zoomed-in view).

Summary

Recently, an observer-based algorithm was reviewed that is able to perform recursive SDFT. Such an observer is able to reconstruct any coherently sampled, band-limited periodic signal. Because the observer's output is identical to that of the SDFT, the reconstruction is not perfect for noncoherently sampled periodic signals.

This article began with a slight modification of the observer, which can perfectly reconstruct any band-limited periodic signal. This can be achieved simply by the proper adjustment of the basis-reciprocal basis system, i.e., complex exponentials with the proper frequencies are to be utilized.

The main portion of the article introduced the AFA. It was shown that, with an adequate modification of the observer, both the fundamental frequency and the spectral content of the signal can be estimated simultaneously. It was also demonstrated that the AFA is capable of the error-free reconstruction of periodic signals using a varying time–frequency function. The signal-tracking properties of the AFA were illustrated and compared to those of other algorithms using practical examples.

Table 2. The mean error and standard deviation of the frequency estimation of the noisy test signal.

	Mean Error (Hz)			Standard Deviation (Hz)		
Method	First Second	Second Second	Third Second	First Second	Second Second	Third Second
Notch filter	1.7e-03	−6.5e-01	1.9e-03	8.5e-02	1.1e-01	8.2e-02
DFT	−4.1e-02	1.1e+00	5.6e-02	4.2e-01	2.6e-01	2.5e-01
AFA	−5.8e-03	−2.4e-03	4.2e-03	6.1e-02	4.6e-01	3.9e-01

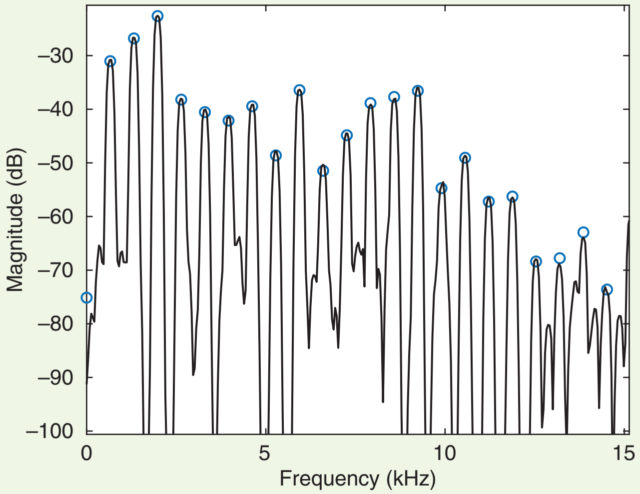


FIGURE 14. The spectral estimation of a violin open E5 string sound at time instant $t = 1.5$ s by FFT (black line) and AFA (blue circles).

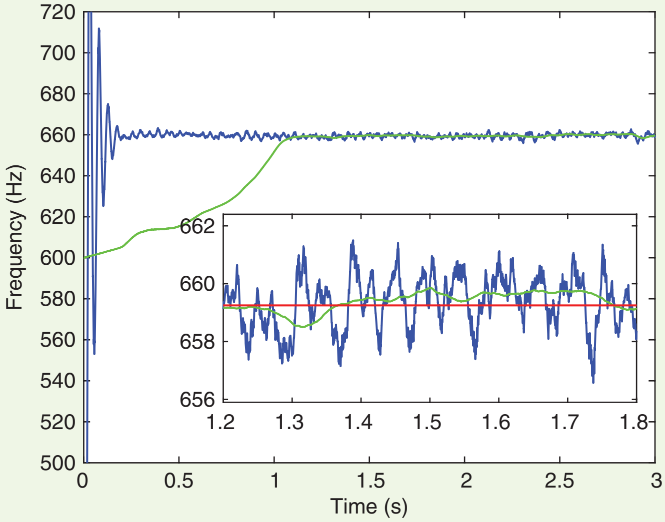


FIGURE 15. The frequency estimation of a violin open E5 string sound by the AFA (blue line) and adaptive notch filter (green line). The red line indicates the nominal frequency.

Supplementary material

This article has supplementary downloadable material provided by the author available in IEEE *Xplore*. The material includes MATLAB functions implementing several versions of the AFA algorithm with test examples. The supplementary materials can also be downloaded at http://mit.bme.hu/~sujbert/afa.

Authors

László Sujbert (sujbert@mit.bme.hu) received his M.Sc., Ph.D., and Dr. habil. degrees in electrical engineering from the Budapest University of Technology and Economics, Hungary, in 1992, 1998, and 2017, respectively. He has been with the Department of Measurement and Information Systems, the Budapest University of Technology and Economics since 1992, where he is the head of the Digital Signal Processing Laboratory. He is active in the fields of measurement, signal processing and embedded systems. He is a Senior Member of the IEEE.

Gyula Simon (simon.gyula@amk.uni-obuda.hu) received his M.Sc. and Ph.D. degrees in electrical engineering from the Budapest University of Technology, Hungary, in 1991 and 1998, respectively. Currently, he is a professor at Óbuda University, Budapest, Hungary. His research interests include adaptive signal processing, localization, and sensor networks. He is a Member of the IEEE.

Gábor Péceli (peceli@mit.bme.hu) received his electrical engineering degree from the Budapest University of Technology and Economics (BME), Hungary, in 1974, and his Candidate and Doctor of Technical Sciences degrees from the Hungarian Academy of Sciences (HAS), Budapest, in 1985 and 1989, respectively. Since 1974, he has been with the Department of Measurement and Information Systems of BME, where he served as chair for 20 years. His main research interest is related to measurement and signal processing structures and embedded adaptive systems. He is a Life Fellow of the IEEE and a member of HAS.

References

[1] Z. Kollár, F. Plesznik, and S. Trumpf, "Observer-based recursive sliding discrete Fourier transform," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 100–106, Nov. 2018. doi: 10.1109/MSP.2018.2853196.

[2] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures.* New York: Wiley, 1971.

[3] F. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978. doi: 10.1109/PROC.1978.10837.

[4] A. Brandt, T. Lagö, K. Ahlin, and J. Tuma, "Main principles and limitations of current order tracking methods," *Sound Vib.*, vol. 39, no. 3, pp. 19–22, Mar. 2005.

[5] R. Potter, "A new order tracking method for rotating machinery," *Sound Vib.*, vol. 24, no. 9, pp. 30–34, 1990.

[6] "Effective machinery measurements using dynamic signal analyzers," *ROM*, Palo Alto, CA, *Application Note*, 243-1, 1997.

[7] F. Nagy, "Measurement of signal parameters using nonlinear observers," *IEEE Trans. Instrum. Meas.*, vol. 41, no. 1, pp. 152–155, Feb. 1992. doi: 10.1109/19.126651.

[8] F. Nagy, "An adaptive Fourier analysis algorithm," presented at the 5th Int. Conf. Signal Processing Applications and Technology, Dallas, TX, Oct. 18–21, 1994, pp. 414–418.

[9] G. Péceli, "A common structure for recursive discrete transforms," *IEEE Trans. Circuits Syst.*, vol. 33, no. 10, pp. 1035–1036, Oct. 1986. doi: 10.1109/TCS.1986.1085844.

[10] P. Stoica and R. Moses, *Spectral Analysis of Signals.* Englewood Cliffs, NJ: Prentice Hall, 2005.

[11] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, Sept. 1974. doi: 10.1109/TIT.1974.1055282.

[12] X. Ming and D. Kang, "Corrections for frequency, amplitude and phase in a fast Fourier transform of a harmonic signal," *Mech. Syst. Signal Process.*, vol. 10, no. 2, pp. 211–221, 1996. doi: 10.1006/mssp.1996.0015.

[13] A. Nehorai and B. Porat, "Adaptive Comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1124–1138, Oct. 1986. doi: 10.1109/TASSP.1986.1164952.

[14] L. Tan and J. Jiang, "Novel adaptive IIR filter for frequency estimation and tracking," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 186–189, Nov. 2009. doi: 10.1109/MSP.2009.934189.

[15] S. Ikaro, "Fundamental Frequency Tracking through Comb (Notch) IIR Filtering," MathWorks, Natick, MA. Accessed on: Jan. 27, 2020. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/27057-fundamental-frequency-tracking-through-comb-notch-iir-filtering

[16] "Record of a violin E5 open string," Freesound. Accessed on: Feb. 13, 2020. [Online]. Available: https://freesound.org/people/jeudyx/sounds/332599/

DATES AHEAD

Please send calendar submissions to:
Dates Ahead, Att: Samantha Walter, E-mail: walter.samantha@ieee.org

Editor's Note

Due to changing situations around the world because of the novel coronavirus (COVID-19) outbreak, please double-check each conference's website for the latest news and updates.

2020

JULY

Virtual: IEEE International Conference on Multimedia and Expo (ICME)
6–10 July
General Chairs: Marta Mark, Ebroul Izquierdo, and Vladan Velisavljevic
URL: http://www.2020.ieeeicme.org/

SEPTEMBER

Virtual: Sensor Signal Processing for Defence (SSPD)
15–16 September
General Chairs: Mike Davies, Stephen McLaughlin, Jordi Barr, and Gary Heald
URL: https://sspd.eng.ed.ac.uk/

IEEE International Workshop on Machine Learning for Signal Processing (MLSP)
21–24 September, Espoo, Finland.
General Chair: Simo Särkkä
URL: https://ieeemlsp.cc

IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)
21–23 September, Tampere, Finland.
General Chairs: Atanas Gotchev and Dong Tian
URL: https://events.tuni.fi/mmsp2020/


©ISTOCKPHOTO.COM/GIVAGA

The 2020 IEEE Radar Conference will take place in Florence, Italy, 21–25 September 2020.

IEEE Radar Conference (RadarConf 2020)
21–25 September, Florence, Italy.
Honorary Chair: Alfonso Farina
General Cochairs: Fulvio Gini and Maria Sabrina Greco
URL: https://www.radarconf20.org/

OCTOBER

Virtual: IEEE International Workshop on Signal Processing Systems (SiPS)
20–22 October
URL: http://www.sips2020.org

Virtual: IEEE International Conference on Image Processing (ICIP)
25–28 October
General Chairs: Mohammed Al-Mualla and Moncef Gabbouj
URL: http://2020.ieeeicip.org/

NOVEMBER

54th Annual Asilomar Conference on Signals, Systems, and Computers
1–4 November, Pacific Grove, California, United States.
General Chair: Joseph R. Cavallaro
URL: www.asilomarssc.org

DECEMBER

IEEE International Workshop on Information Forensics and Security (WIFS)
6–9 December, New York, United States.
General Chairs: David Doermann and Anthony Hoog
URL: https://www.wifs2020.nyu.edu/home

2021

JANUARY

28th European Signal Processing Conference (EUSIPCO)
18–22 January, Amsterdam, The Netherlands.
General Chairs: Richard Heusdens and Cédric Richard
URL: https://eusipco2020.org

IEEE Spoken Language Technology Workshop (SLT)
19–22 January, Shenzhen, China.
General Chairs:Zhijian Ou and Lei Xie
URL: http://slt2020.org/index.html

SP

Digital Object Identifier 10.1109/MSP.2020.2985824
Date of current version: 26 June 2020

(continued from page 148)

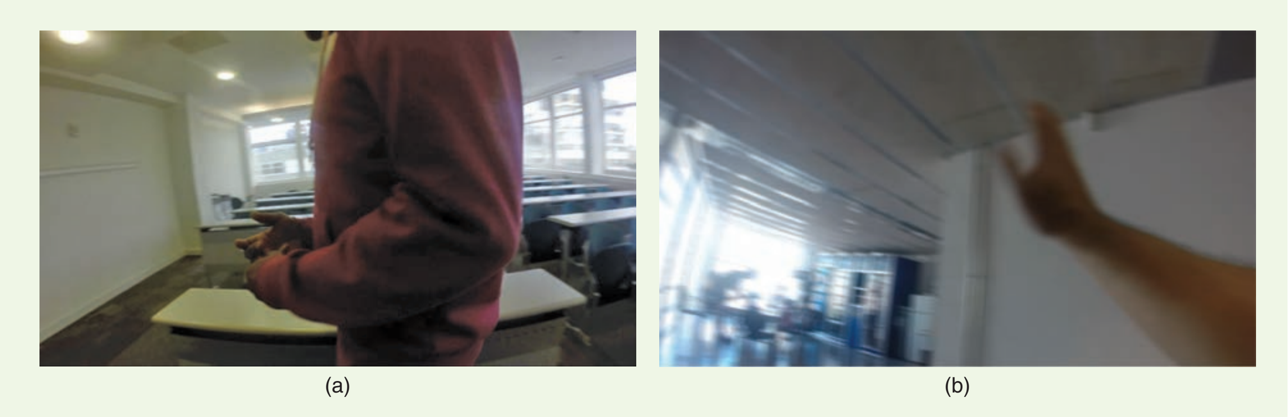

(a) (b)

FIGURE 1. The challenges of BWC video data include (a) poor framing and (b) low visual quality due, for example, to motion blur or overexposure. (Used with permission from [24] and [25].)

information without compromising subsequent services that use the data. PETs may be applied once the data are transferred to a content management system or may be integrated within BWCs and transform the data in the device itself, prior to their transfer.

BWCs capture time-varying information in audio, inertial measurements and video data, all of which are difficult to protect using standard privacy-preserving approaches, such as differentially private protocols [15]. PETs designed for BWCs should be protective, reliable, and operative (these properties are a selected subset from those listed in [16] and have been adapted to be specific to BWCs). A PET is protective if it safeguards private, secondary information that is not necessary for the service, reliable if it does not affect the service and maintains performance, and operative if it integrates easily with existing work practices and its functions are explainable.

The design of PETs also requires the identification of who or what the private information should be protected from: individuals observing the data or algorithms extracting sensitive information. Although protection from individuals can be implemented using standard access control procedures, the protection from algorithms is a recent and growing challenge.

Algorithmic inferences on audio data can reveal a wide range of potentially private information, such as one's height and weight [17], emotional state [18], [19], and health conditions [19]. Motion sensor data collected by inertial measurement units can also reveal information about an individual's physical characteristics, such as height and weight [20], level of activity [21], and changes in behavioral patterns [22]. These data may be protected with PETs developed for other types of devices and that use transformations of the data [23]. Protecting BWC video data is, however, more challenging.

To preserve the privacy of bystanders captured in BWC videos, PETs designed for traditional stationary cameras could be adopted, but additional steps are required to ensure that the privacy-preserving methods remain effective. In fact, the unique challenges associated with BWC videos [24], such as motion blur and poorly framed content (Figure 1), hinder the direct application of PETs designed for stationary cameras. Also, to preserve the privacy of the wearer, new solutions are needed for video data. This is a distinct challenge in BWCs, as they capture a unique first-person viewpoint [25], which may include the wearer handling medications or close-ups of smartphone screens (Figure 2). Furthermore, BWCs indirectly disclose the activities of the wearer, which are captured through the motion of the camera itself [26]. Such motion information is difficult to disentangle from the primary visual information needed for the intended use of the video data, making the design of obfuscation or data-minimization techniques that prevent the collection of this secondary information an important research opportunity.

In conclusion, the increasing adoption of BWCs presents the research community with several new challenges associated with the development of PETs that are specific to the unique type of first-person data collected by BWCs. These PETs should not only operate on each modality (vision, sound, and inertial) independently, but also across modalities, as cross-modal correlations heighten the threat to privacy [27].

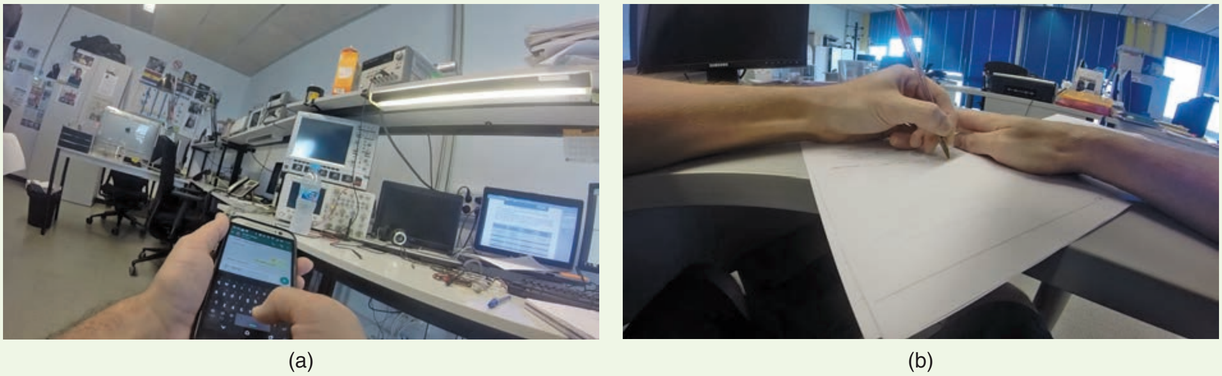
Acknowledgments

We wish to thank the Alan Turing Institute (EP/N510129/1), which is funded by the U.K. Engineering and Physical Sciences Research Council, for its support throughout the project Privacy-Preserving Multimodal Learning for Activity Recognition (PRIMULA).

Authors

Maria S. Cross (m.s.cross@se18.qmul.ac.uk) received her M.Sc. degree in artificial intelligence from Queen Mary University of London (QMUL), United Kingdom, and her Ph.D. degree in health informatics from the University College London, United Kingdom. She is currently a postdoctoral research assistant at the Centre for Intelligent Sensing, QMUL.

Andrea Cavallaro (a.cavallaro@qmul.ac.uk) is a professor of multimedia signal processing at Queen Mary University of London (QMUL), United Kingdom, and a



(a)

(b)

FIGURE 2. The unique first-person viewpoint captured by BWCs may include (a) smartphone screens and (b) personal documents. (Used with permission from [25].)

Turing fellow with the Alan Turing Institute, the U.K. National Institute for Data Science and Artificial Intelligence. He is a fellow of the International Association for Pattern Recognition; director of the QMUL Centre for Intelligent Sensing; editor-in-chief of *Signal Processing: Image Communication*; senior area editor for *IEEE Transactions on Image Processing*; chair of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee; and an IEEE Signal Processing Society Distinguished Lecturer.

References

[1] R. Adams, "Schools trial body cameras to aid safety and monitor behaviour," *The Guardian*, Feb. 7, 2020. Accessed on: Apr. 2, 2020. [Online]. Available: www.theguardian.com/education/2020/feb/07/schools-trial-body-cameras-to-aid-safety-and-monitor-behaviour

[2] C. Lum, M. Stoltz, C. S. Koper, and J. A. Scherer, "Research on body-worn cameras: What we know, what we need to know," *Criminol. Public Policy*, vol. 18, no. 1, pp. 93–118, 2019. doi: 10.1111/1745-9133.12412.

[3] B. Ariel, M. Newton, L. McEwan, G. A. Ashbridge, C. Weinborn, and H. Sabo Brants, "Reducing assaults against staff using body-worn cameras (BWCs) in railway stations," *Crim. Justice Rev.*, vol. 44, no. 1, pp. 76–93, 2019. doi: 10.1177/0734016818814889.

[4] "Gartner says worldwide wearable device sales to grow 17 percent in 2017," Gartner, Stamford, CT, Aug. 24, 2017. Accessed on: Apr. 6, 2020. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2017-08-24-gartner-says-worldwide-wearable-device-sales-to-grow-17-percent-in-2017

[5] "Wearable and body-worn cameras market: Growth, trends and forecast (2020–2025)," Mordor Intelligence, Hyderabad, India, 2019. Accessed on: Apr. 2, 2020. [Online]. Available: https://www.mordorintelligence.com/industry-reports/wearable-and-body-worn-cameras-market

[6] T. Nortcliffe, "Safeguarding body worn video data," Home Office, London, Rep. no. 011/18, Oct. 2018.

[7] Council of the European Union (EU), European Parliament. (2016, Apr. 27). 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ 2016 L 119/1, Article 4. Accessed on: May 3, 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[8] J. Finocchiaro, A. U. Khan, and A. Borji, "Egocentric height estimation," in *Proc. IEEE Winter Conf. Applications of Computer Vision (WACV)*, Santa Rosa, CA, Mar. 2017, pp. 1142–1150. doi: 10.1109/WACV.2017.132.

[9] M. Nouredanesh, A.W. Li, A. Godfrey, J. Hoey, and J. Tung, "Chasing feet in the wild: A proposed egocentric motion-aware gait assessment tool," in *Proc. European Conf. Computer Vision (ECCV)*, Munich, Sept. 2018, pp. 176–192. doi: 10.1007/978-3-030-11024-6_12.

[10] H. Pirsiavash, and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, June 2012, pp. 2847–2854. doi: 10.1109/CVPR.2012.6248010.

[11] D. Damen, H. Doughty, G. M. Farinella, S. Fidler, A. Furnari, E. Kazakos, D. Moltisanti, J. Munro et al., "Scaling egocentric vision: The epic-kitchens dataset," in *Proc. European Conf. Computer Vision (ECCV)*, Munich, Sept. 2018, pp. 720–736.

[12] A. Fathi, J. K. Hodgins, and J. M. Rehg, "Social interactions: A first-person perspective," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, June 2012, pp. 1226–1233. doi: 10.1109/CVPR.2012.6247805.

[13] C. Cadwalladr, "The great British Brexit robbery: How our democracy was hijacked," *The Guardian*, May 7, 2017. Accessed on: Apr. 6, 2020. [Online]. Available: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked- democracy

[14] A. Kofman and A. Tobin, "Facebook ads can still discriminate against women and older workers, despite a civil rights settlement," ProPublica, New York, Dec. 13, 2019. Accessed on: Apr. 6, 2020. [Online]. Available: https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement

[15] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proc. Annu. Int. Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, St. Petersburg, Russia, May–June 2006, pp. 486–503. doi: 10.1007/11761679_29.

[16] M. Hansen, J. H. Hoepman, M. Jensen, and S. Schiffner. "Readiness analysis for the adoption and evolution of privacy enhancing technologies: Methodology, pilot assessment, and continuity plan," European Union Agency for Network and Information Security (ENISA), Heraklion, Greece, Mar. 2016. Accessed on: May 3, 2020. [Online]. Available: https://www.enisa.europa.eu/publications/pets

[17] R. M. Krauss, R. Freyberg, and E. Morsella, "Inferring speakers' physical attributes from their voices," *J. Experimental Soc. Psychol.*, vol. 38, no. 6, pp. 618–625, 2002. doi: 10.1016/S0022-1031(02)00510-3.

[18] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M. A. Nicolaou, B. Schuller, and S. Zafeiriou, "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 5200–5204. doi: 10.1109/ICASSP.2016.7472669.

[19] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger et al., "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Proc. INTERSPEECH*, Lyon, France, Aug. 2013, pp. 148–152.

[20] A. Masuda and T. Maekawa, "Estimating physical characteristics with body-worn accelerometers based on activity similarities," *J. Inform. Process.*, vol. 24, no. 2, pp. 237–246, 2016. doi: 10.2197/ipsjjip.24.237.

[21] M. N. S. Zainudin, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Monitoring daily fitness activity using accelerometer sensor fusion," in *Proc. IEEE Int. Symp. Consumer Electronics (ISCE)*, Kuala Lumpur, Malaysia, Nov. 2017, pp. 35–36. doi: 10.1109/ISCE.2017.8355540.

[22] A. Gruenerbl, V. Osmani, G. Bahle, J. C. Carrasco, S. Oehler, O. Mayora, C. Haring, and P. Lukowicz, "Using smartphone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients," in *Proc. Augmented Human Int. Conf.*, Kobe, Japan, Mar. 2014, pp. 1–8. doi: 10.1145/2582051.2582089.

[23] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Privacy and utility preserving sensor-data transformations," *Pervasive Mobile Comput.*, vol. 63, pp. 1–13, Mar. 2020. doi: 10.1016/j.pmcj.2020.101132.

[24] A. Brutti and A. Cavallaro, "On-line cross-modal adaptation for audio-visual person identification with wearable cameras," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 40–51, Feb. 2017. doi: 10.1109/THMS.2016.2620110.

[25] G. Abebe, A. Catala, and A. Cavallaro, "A first-person vision dataset of office activities," in *Proc. Int. Workshop Multimodal Pattern Recognition Social Signals Human Computer Interaction*, Beijing, Aug. 2018, pp. 27–37. doi: 10.1007/978-3-030-20984-1_3.

[26] G. Abebe, and A. Cavallaro, "A long short-term memory convolutional neural network for first-person vision activity recognition," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, Venice, Italy, Oct. 2017, pp. 1339–1346. doi: 10.1109/ICCVW.2017.159.

[27] A. Cavallaro and A. Brutti, "Audio-visual learning for body-worn cameras," in *Multimodal Behaviour Analysis in the Wild*, X. Alameda-Pineda, E. Ricci, and N. Sebe, Eds. New York: Academic, 2019, pp. 103–119.

SP

© GRAPHIC STOCK

The Advertisers Index contained in this issue is compiled as a service to our readers and advertisers: the publisher is not liable for errors or omissions although every effort is made to ensure its accuracy. Be sure to let our advertisers know you found them through *IEEE Signal Processing Magazine.*

IEEE SIGNAL PROCESSING MAGAZINE REPRESENTATIVE
Erik Henson, Naylor Association Solutions, Phone: +1 352 333 3443, Fax: +1 352 331 3525, ehenson@naylor.com

COMPANY	PAGE NUMBER	WEBSITE	PHONE
Magic Data	9	http://en.imagicdatatech.com	+86 10 85527250
MathWorks	Cover 4	mathworks.com/deeplearning	

Digital Object Identifier 10.1109/MSP.2020.2984832

IMAGE LICENSED BY INGRAM PUBLISHING

Maria S. Cross and Andrea Cavallaro

Privacy as a Feature for Body-Worn Cameras

Body-worn cameras (BWCs) are becoming increasingly prevalent within today's society. These devices are now commonly seen on supermarket assistants, shopping center security guards, and public transport staff. Schools are also trialling BWCs on teachers to monitor students' behavior [1]. The use of BWCs is believed to promote the transparency and accountability of behaviors as well as the security of the wearer [2], [3]. With an expected shipment of more than 5 million units in the next year [4] and a compounded annual growth rate of 16% in the next five years [5], BWCs will become a permanent feature within everyday life. Such an uptake of BWCs marks a transition from purposive to passive data collection.

In this article, we discuss the threat to privacy that this passive data collection creates, along with opportunities to mitigate this risk. Furthermore, we argue that the use case of BWCs at work will stimulate the development of solutions that prevent the collection of data that could infringe upon the privacy of the wearer. Finally, we discuss the desirable properties of privacy-enhancing technologies (PETs) for BWCs.

BWCs record large quantities of audiovisual and inertial data that, instead of collecting only select, primary information that is relevant to the intended use, also capture secondary information that is irrelevant for the intended purpose [6]. The recording of an interaction with an aggressive customer is an example of primary information, whereas the identity of other customers waiting to be served is an example of secondary information. Collecting secondary information, which may be personally identifiable information (i.e., information relating to an identified or identifiable natural person), goes against the principle of data minimization prescribed in data-protection regulations [7].

Primary and secondary information captured by BWCs can be used to comprehensively describe the behaviors of the wearer. Such a description may be employed to profile wearers through their physical characteristics [8], [9], their activities [10], the foods they eat [11], and how they interact with others [12]. Each additional day of BWC usage marks the collation of more information, also enabling the inference of insights that are not directly observable. Examples of such insights are how active the wearers are, what motivates them, and their psychological profiles. In detailing the condition of the wearer, these profiles are, in practice, also becoming health records, with potentially greater levels of detail than obtained through clinical interaction.

The exploitation of profiles derived from users' online behaviors has already led to public controversies [13], [14]. This is due, in part, to the asymmetric power relation between users and providers who shift the responsibility of privacy choices to users through lengthy and complex privacy notices. These notices are often written more to protect the provider than to inform users, whose consent should be a "freely given, specific, informed and unambiguous indication of the data subject's wishes" [7].

The situation changes when BWCs are used by employees (e.g., shop assistants or security staff) because the dependence of the wearer upon the organizations collating their data (i.e., their employer) invalidates any consent processes, as the consent cannot be assumed to be "freely given" [7]. It is, therefore, the responsibility of the employer to protect the privacy of employees wearing BWCs and to safeguard BWC data (and any information therein) through a duty of confidentiality.

Maintaining confidentiality becomes more challenging when the data are accessed by an external company. As each individual wearer generates large volumes of data, employers (i.e., data controllers [7]) must seek ways to store and manage the BWC data produced. These solutions and their operation are often beyond the capabilities of the organizations adopting the technology, encouraging the outsourcing of the handling and curation of BWC data. This new landscape creates an urgent need for BWC solutions that offer privacy as a feature and enable employers to govern access to their employees' private, secondary information.

PETs aim to minimize access to data representing personal, secondary

(continued on page 145)

Digital Object Identifier 10.1109/MSP.2020.2989686
Date of current version: 26 June 2020

CALL FOR PAPERS

IEEE Journal of Selected Topics in Signal Processing

Special Issue on Tensor Decomposition for Signal Processing and Machine Learning

Tensor decomposition, also called tensor factorization, is very useful for representing and analyzing multidimensional data. Tensor decomposition has been applied in signal processing (speech, audio, communications, radar, biomedicine), machine learning (clustering, dimensionality reduction, latent factor models, subspace learning), and beyond. Tensor decomposition helps us to learn a variety of models, including community models, probabilistic context-free-grammars, the Gaussian mixture model, and two-layer neural networks.

The multidimensional nature of the signals and even "bigger" data provide a good opportunity to exploit tensor-based models and tensor network, with the aim of meeting the strong requirements on flexibility, convergence, and efficiency. Although considerable research has been done on this subject, there are many challenges still outstanding that need to be explored, like high computational cost of algorithms, tensor deflation, massive tensor decomposition, etc. The goal of this special issue is to attract high quality papers containing original research on tensor methods, tensor decompositions for signal processing and machine learning, and their applications in big data, social network, biomedical and healthcare, advanced data-driven information and communication technology (ICT) systems and others.

Potential topics include, but are not limited, to the following:

- New tensor decompositions and uniqueness issues of tensor models
- Low-rank approximations
- Fast and robust tensor decompositions
- Novel algorithms for existing tensor decomposition models
- Optimization problems related to tensor models
- Tensor-based detection and parameter estimation
- Tensor decomposition for 5G/B5G wireless communications
- Tensor-based data-driven networking
- Tensor processing and analysis in social networks
- Tensor decomposition for industry internet of things
- Spatial temporal data via tensor factorization
- Computer vision with tensor method
- Biomedical, healthcare, and audio signal processing with tensors
- Pattern recognition and neural networks with tensor decomposition

Submission Guidelines:

Prospective authors should follow the instructions on the IEEE JSTSP webpage https://signalprocessingsociety.org/publications-resources/ieee-journal-selected-topics-signal-processing and submit their manuscripts at http://mc.manuscriptcentral.com/jstsp-ieee.

Important Dates:

Manuscript submissions due: July 1, 2020
First review due: September 1, 2020
Revised manuscript due: November 1, 2020
Second review due: December 15, 2020
Final manuscript due: January 15, 2021

Guest editors:

Hongyang Chen (Lead GE), Fujitsu Ltd, Japan (hongyang.chen@fujitsu.com)
Sergiy Vorobyov, Aalto University, Finland (svor@ieee.org, sergiy.vorobyov@aalto.fi)
Hing Cheung So, City University of Hong Kong, China (hcso@ee.cityu.edu.hk)
Fauzia Ahmad, Temple University, Philadelphia, USA (fauzia.ahmad@temple.edu)
Fatih Porikli, Australian National University, Australia (fatih.porikli@anu.edu.au)

Digital Object Identifier 10.1109/MSP.2020.2999174