Advances in Computer Vision and Pattern Recognition

ACVPR

Rui Fan
Sicen Guo
Mohammud Junaid Bocus *Editors*

Autonomous Driving Perception

Fundamentals and Applications

Springer

Advances in Computer Vision and Pattern Recognition

Titles in this series now included in the Thomson Reuters Book Citation Index!

Advances in Computer Vision and Pattern Recognition is a series of books which brings together current developments in this multi-disciplinary area. It covers both theoretical and applied aspects of computer vision, and provides texts for students and senior researchers in topics including, but not limited to:

- Deep learning for vision applications
- Computational photography
- Biological vision
- Image and video processing
- Document analysis and character recognition
- Biometrics
- Multimedia
- Virtual and augmented reality
- Vision for graphics
- Vision and language
- Robotics

Rui Fan · Sicen Guo · Mohammud Junaid Bocus
Editors

Autonomous Driving Perception

Fundamentals and Applications

Springer

Editors
Rui Fan
Control Science & Engineering, Shanghai Research Institute for Intelligent Autonomous Systems
Tongji University
Shanghai, P. R. China

Sicen Guo
Control Science & Engineering, Shanghai Research Institute for Intelligent Autonomous Systems
Tongji University
Shanghai, P. R. China

Mohammud Junaid Bocus
Electrical and Electronic Engineering
University of Bristol
Bristol, UK

ISSN 2191-6586 ISSN 2191-6594 (electronic)
Advances in Computer Vision and Pattern Recognition
ISBN 978-981-99-4286-2 ISBN 978-981-99-4287-9 (eBook)
https://doi.org/10.1007/978-981-99-4287-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Paper in this product is recyclable.

Preface

With the recent advancements in artificial intelligence, there is a growing expectation that fully autonomous driving vehicles will soon become a reality, leading to significant societal changes. The core competencies of an autonomous vehicle system can be broadly categorized into four main categories: perception, prediction, planning, and control. The environmental perception system serves as the foundation of autonomous vehicles, utilizing cutting-edge computer vision and machine learning algorithms to analyze raw sensor data and create a comprehensive understanding of the surrounding environment. Similar to the visual cognition and understanding of humans, this process allows for a deep and nuanced perception of the world.

Conventional autonomous driving perception systems are often hindered by separate sensing, memory, and processing architectures, which may not meet the demand for ultra-high raw sensor data processing rates and ultra-low power consumption. In contrast, in-sensor computing technology performs signal processing at the pixel level by utilizing the collected analog signals directly, without requiring data to be sent to other processors. This enables highly efficient and low-power consumption visual signal processing by integrating sensing, storage, and computation onto focal planes with innovative circuit designs or new materials. Therefore, the in-sensor computing paradigm holds significant potential for autonomous driving. Furthermore, fish-eye cameras have emerged as an essential sensor in the field of autonomous driving. Thanks to the unique projection principle of fish-eye cameras, they offer a significantly larger field of view (FoV) compared to conventional cameras. This distinct characteristic makes fish-eye cameras highly versatile and suitable for a wide range of autonomous driving perception applications. In addition, computer stereo vision is a cost-effective and efficient method for depth information acquisition, and it has found widespread use in 3D environmental perception. Despite the impressive results obtained by state-of-the-art (SoTA) stereo vision algorithms that utilize convolutional neural networks, their training typically necessitates a substantial amount of accurately labeled disparity ground truth data. Consequently, self-supervised or unsupervised deep stereo networks have emerged as the dominant approach in this research area.

Research on semantic segmentation has been ongoing for over a decade. However, conventional single-modal networks are unable to fully utilize the spatial information provided by range sensors, making them less effective in diverse weather and illumination conditions. To address this challenge, data-fusion semantic segmentation networks have been developed, which employ multiple encoders to extract deep features from different visual information sources. These deep features are subsequently fused to provide a more comprehensive understanding of the surrounding environment. 3D object detection is also a crucial component of autonomous driving systems that has made remarkable progress in recent years. Nonetheless, the various perceptual sensors used for object detection present their unique challenges. Cameras are vulnerable to issues such as foreshortening and flickering effects, over-exposure problems, as well as environmental variations like lighting and weather conditions. Similarly, LiDARs and RADARs suffer from low-resolution and sparse data representations. Furthermore, occlusion presents a significant challenge to object detection, leading to the partial or complete invisibility of obstructed objects. To address these challenges, collaborative 3D object detection has been proposed as an alternative to conventional approaches. Collaborative object detection facilitates information sharing between agents, enabling them to perceive the environments beyond line-of-sight and FoV. This approach holds great promise in overcoming the limitations of individual sensors and achieving more robust and accurate 3D object detection in autonomous driving systems.

The application of the simultaneous localization and mapping (SLAM) technique to autonomous driving also presents several challenges. Over the past three decades, researchers have made significant progress in addressing the probabilistic SLAM problem by developing a range of theoretical frameworks, efficient solvers, and complete systems. Visual SLAM for texture-less environments is an especially challenging task, as multi-view images cannot be effectively linked using reliable keypoints. However, researchers continue to develop new techniques and algorithms to overcome this limitation. Moreover, the enhancement of SLAM systems is also being driven by the emergence of new sensors or sensor combinations, such as cameras, LiDARs, IMUs, and other similar technologies. As these sensors become more advanced and sophisticated, they offer new opportunities to improve the accuracy and reliability of SLAM systems for autonomous driving applications.

Multi-task learning has become a popular paradigm for simultaneously tackling multiple tasks while using fewer computational resources and reducing the inference time. Recently, several self-supervised pre-training methods have been proposed, demonstrating impressive performance across a range of computer vision tasks. However, the extent to which these methods can generalize to multi-task situations remains largely unexplored. Additionally, the majority of multi-task algorithms are tailored to specific tasks that are usually unrelated to autonomous driving, posing difficulties when attempting to compare state-of-the-art multi-task learning approaches in the domain of autonomous driving.

Bird's eye view (BEV) perception involves transforming a perspective view into a bird's eye view and performing various perception tasks, such as 3D detection, map segmentation, tracking, and motion planning. Thanks to its inherent advantages in 3D space representation, multimodal fusion, decision-making, and planning, the topic of BEV perception has attracted significant interest among both academic and industrial researchers.

Road environment perception, which includes 3D geometry reconstruction of road surfaces and the intelligent detection of road damages, is also critical for ensuring safe and comfortable driving. Road surface defects can be extremely hazardous, especially when hit at high speeds, as these can not only damage the vehicle's suspension but also cause the driver to lose control of the vehicle. When one of the vehicle's tyres enters a pothole, the weight distribution across all tyres becomes unbalanced, causing the vehicle to tilt and shift more towards the tyres that are lower relative to the pothole. This uneven weight distribution can produce a considerable and focused force on the tyre when it hits the edge of the pothole, resulting in deformation, breakage, or even bending of the rim. The damage inflicted on the tyre impacts the driving experience, making it challenging to maintain a straight driving trajectory.

This book provides an in-depth, comprehensive, and SoTA review on a range of autonomous driving perception topics, such as stereo matching, semantic segmentation, 3D object detection, simultaneous localization and mapping, and BEV perception. The book's webpage can be accessed at mias.group/ADP2023.

The intended readership for this book primarily comprises of tertiary students who seek a comprehensive and yet an introductory understanding of the fundamental concepts and practical applications of machine vision and deep learning techniques. It is also directed at professionals and researchers in autonomous driving who are seeking an assessment of the current state-of-the-art methods available in existing literature, and who aspire to identify potential areas of research for further exploration. The extensive range of topics covered in this book makes it an invaluable resource for a variety of university programs that include courses related to machine vision, deep learning, and robotics.

In Chapter 1, the book discusses the use of in-sensor visual devices for autonomous driving perception. Chapter 2 provides a thorough and up-to-date review of SoTA environmental perception algorithms that are specifically designed for fish-eye cameras. In Chapter 3, the theoretical foundations and algorithms of computer stereo vision are discussed. Chapter 4 presents a review of SoTA single-modal and data-fusion semantic segmentation networks. Chapter 5 reviews 3D object detection methods for autonomous driving. Chapter 6 provides an assessment of the current SoTA collaborative 3D object detection systems and algorithms. In Chapter 7, sensor-fusion robust SLAM techniques for mobile robots are introduced. Chapter 8 discusses visual SLAM in texture-less environments. Chapter 9 presents a comprehensive survey on multi-task perception frameworks. Chapter 10 specifically covers state-of-the-art BEV perception algorithms. Finally, Chapter 11 discusses road environment perception techniques for safe and comfortable driving.

Shanghai, P. R. China — Rui Fan
Shanghai, P. R. China — Sicen Guo
Bristol, UK — Mohammud Junaid Bocus

Acknowledgements This book was supported by the National Key R&D Program of China under Grant 2020AAA0108100, the National Natural Science Foundation of China under Grant 62233013, the Science and Technology Commission of Shanghai Municipal under Grant 22511104500, and the Fundamental Research Funds for the Central Universities.

Contents