



## Views &amp; Comments

# Vision Sensing for Intelligent Driving: Technical Challenges and Innovative Solutions

Xinle Gong<sup>a,b</sup>, Zhihua Zhong<sup>c,\*</sup>

<sup>a</sup>School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>b</sup>School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China

<sup>c</sup>School of Automotive Studies & College of Transportation, Tongji University, Shanghai 200092, China



## 1. Introduction

With the rapid development of automotive intelligent technologies, in-vehicle vision sensors have gained traction as essential components of intelligent driving systems, providing critical road condition data, assisting driving decisions, and enabling autonomous driving functions. Compared with other in-vehicle perception sensors, vision sensors deliver more detailed and comprehensive environmental information, offering vital data to support intelligent driving decision-making and control [1]. Therefore, the accuracy and reliability of visual perception directly influence the safety and performance of intelligent driving systems, making it a central factor in achieving truly intelligent and autonomous vehicles. Fig. 1 illustrates the vision sensing in intelligent driving.

Compared with industrial and consumer cameras, automotive-grade cameras are subject to more complex and stringent technical requirements, necessitating targeted and innovative engineering solutions [2]. First, in terms of function and application, automotive-grade cameras must adapt to variable driving environments, maintaining high performance in image processing, hardware robustness, and system integration under varying lighting conditions. Second, in design and manufacturing, these cameras require specialized materials and advanced packaging technologies, and must undergo rigorous testing for environmental adaptability and resistance to vibration impact to ensure good imaging quality even under adverse weather conditions. Additionally, automotive-grade cameras encounter significant technical challenges in lenses, complementary metal oxide semiconductor (CMOS) image sensors, and image signal processors. These challenges include demands for high dynamic range, adaptation to both strong and low light environments, vibration-induced blur compensation, low latency, high resolution, and reliable performance [3]. Meeting these technical requirements is essential for achieving safe and precise environmental perception and target recognition in intelligent driving scenarios.

## 2. Key techniques and challenges

### 2.1. Structure and key techniques

In-vehicle vision sensors primarily consist of a lens, CMOS image sensor, image signal processor, and serializer [4]. The lens refracts light reflected from objects and focuses it onto the CMOS image sensor. The CMOS image sensor utilizes the photoelectric conversion effect of photodetectors to convert the optical image on the photosensitive surface into an electrical signal, maintaining a proportional relationship with the incident light intensity. The image signal processor subsequently preprocesses the raw image data, performing functions such as image scaling, auto exposure (AE), auto white balance (AWB), auto focus (AF), and image noise reduction while also converting the data into an appropriate format. Finally, the serializer transmits the processed image data. Fig. 2 illustrates the structure and working principle of vision sensors.

### 2.2. Main challenges

In-vehicle vision sensors are essential in real-time target detection and scene interpretation for intelligent driving systems, and significant progress has been achieved in recent years. However, owing to the complexity and variability of driving environments, current vision sensing technology continue to encounter limitations in balancing high resolution with high frame rate, capturing high dynamic range images, resisting intense light, maintaining sensitivity in low-light environments, providing quick response and low latency, and ensuring safety and reliability [5]. Moreover, vibrations and jolts during vehicle operation can significantly impact the performance of vision sensors. These limitations result in reduced environmental perception accuracy and delays in target recognition under sudden lighting changes, extreme weather, and special driving scenarios, thereby increasing driving safety risks and reduce the rapid development of intelligent driving technology [6]. This paper focuses on three core components of vision sensors—optical lens, CMOS image sensor, and image signal processor—and discusses the primary technical challenges associated with each of them.

\* Corresponding author.

E-mail address: [zzh@cae.cn](mailto:zzh@cae.cn) (Z. Zhong).

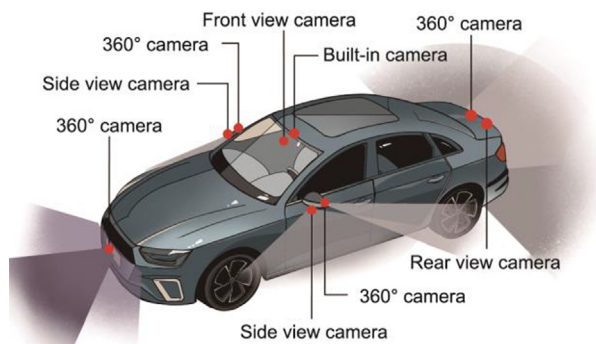


Fig. 1. Vision sensing for intelligent driving.

(1) Optical lens. The parameters of a lens, such as aperture size and focal length range, are constrained by physical factors like depth of field, aperture diffraction, and lens size, which influence image quality under varying lighting conditions [7]. Owing to these optical limitations, current lenses often fail to preserve critical details in high-contrast scenes containing both strong and low light, thereby hindering accurate recognition of road conditions and traffic signs [8]. Although automotive wide-angle lenses extend the field of view, they also introduce optical distortion, resulting in substantial errors in object recognition and distance estimation. For instance, intelligent vehicles typically construct bird’s-eye view (BEV) images using 190° fisheye cameras; however, substantial distortion errors at the periphery causes localization deviations of obstacles, particularly in narrow road scenarios. Additionally, because traditional lens hydrophobic coatings lack sufficient dirt resistance, residual water droplets blur images, preventing BEV systems from detecting lane markings on reflective road surfaces during rainy conditions.

The sharpness of lenses is limited by the diffraction limit. Current lens technology struggles to improve resolution while maintaining miniaturization, which limits the ability to capture and recognize details of distant or small objects such as distant traffic signs or road hazards. The materials used in current lenses are sensitive to temperature, causing focal length drift and inaccurate color reproduction during extreme temperature changes, thereby degrading imaging quality in extreme weather conditions. Additionally, lens performance deteriorates over prolonged use. Furthermore, mechanical vibrations and jolts during vehicle operation can result in defocusing, reducing image clarity and stability [9]. Automotive tests have indicated that when a vehicle traverses a bump at speeds exceeding 60 km·h<sup>-1</sup>, the camera vibration frequencies can reach 30 Hz, causing image blur. This subsequently delays the activation of the automatic emergency braking (AEB) system.

(2) CMOS image sensor. High spatial resolution increases the number of pixels and the amount of image data collected. However, owing to the readout speed and analog-to-digital conversion efficiency of CMOS image sensors, this lengthens data processing

time and reduces temporal resolution (i.e., frame rate) [10]. As a result, current technologies often compromise temporal resolution to enhance spatial resolution, making it difficult for in-vehicle vision sensors to simultaneously detect small distant targets and track fast-moving objects in real time. The human eye has a dynamic range of approximately 160 dB; however, current automotive-grade CMOS sensors typically achieve only 120–140 dB. Although this range is gradually improving, it remains challenging to fully accommodate variations in lighting intensity across all scenarios encountered in intelligent driving [11]. Moreover, increasing resolution reduces pixel size, decreases saturation charge capacity, and further limits the dynamic range of CMOS image sensors. This makes them prone to local overexposure or underexposure in complex scenes containing both bright and dim lighting, such as tunnel entrances or nighttime environments with strong light sources. These limitations can result in the loss of critical detail and affect the simultaneous recognition of high-reflectivity objects (e.g., traffic signs) and shaded areas. For instance, backlit conditions during sunrise or sunset significantly reduce license plate recognition rates.

Limited by characteristics such as dynamic range, electronic overflow, lens scattering, analog-to-digital converter bit depth, and exposure control, current CMOS image sensors exhibit a limited ability to suppress strong light sources (e.g., direct sunlight or vehicle headlights), resulting in glare effects that influence imaging in areas around these sources [12]. This can create detection blind spots, such as temporary “blinding” from oncoming high-beam headlights during nighttime driving. High-resolution vision sensors decrease pixel size, which results in smaller photodetectors. Owing to physical limitations, this reduces the photoelectric conversion efficiency of individual pixels and increases readout noise, thereby negatively impacting low-light performance. In low-light conditions such as nighttime or dusk, image quality deteriorates, reducing detection accuracy for distant or small targets. Additionally, it increases the computational burden for noise reduction, limiting the real-time performance of image processing [13]. For instance, existing vision sensors experience a significant decrease in pedestrian detection recall rates during dusk. The primary reason is the sharp decline in the signal-to-noise ratio (SNR) of CMOS sensors under low-light conditions.

Current CMOS image sensor technology primarily employs Bayer filter arrays to differentiate color sources, resulting in color resolution being only one-fourth of the original resolution [14]. Color noise becomes significant under high sensitivity, hindering consistent appearance features of targets in varying lighting conditions and influencing color-based target recognition, such as traffic signs and signals. The dark current of CMOS image sensors also increases with temperature, which negatively impacts image quality. Therefore, current technology requires recalibration of CMOS image sensors at different temperatures; however, extreme temperature conditions, such as intense heat or cold, can still degrade sensor performance. Additionally, prolonged operation can cause a temperature rise, which influences sensor stability.

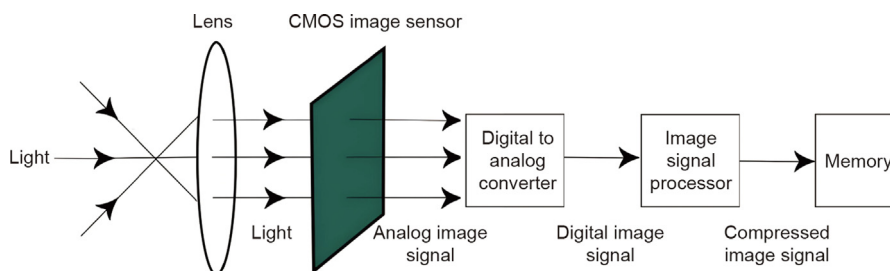


Fig. 2. Structure and working principle of vision sensors.

(3) Image signal processor. The on-chip and parallel processing capabilities of current image signal processors are limited, affecting the speed of image processing, especially for high-resolution and high-frame-rate image data [15]. This limitation directly impacts the real-time response capability of intelligent vehicles. Processing delays in high-speed driving or complex traffic environments can hinder timely identification and response to sudden situations, compromising the safety and reliability of intelligent driving systems. In low-light conditions, sensor noise increases, and the noise reduction and signal enhancement algorithms of current image signal processors require improvement [16]. These shortcomings can result in deteriorated image quality and a reduced ability to recognize and track vehicles and pedestrians in nighttime or tunnel environments.

Current image signal processors struggle to effectively handle noise and glare in environments containing strong light, glare, rain, snow, and road surface reflections. Under these extreme conditions, image clarity and usability are reduced, leading to misjudgments or failures in accurately identifying obstacles [17]. Many image signal processors rely on hard-coded technology and lack programming flexibility, making it difficult to adapt to new processing algorithms or specific intelligent driving needs through software updates. Additionally, they lack scene understanding and contextual awareness, reducing optimization for specific driving scenarios, such as tunnels or nighttime environments, and limiting the intelligence of the vision perception system. Current image signal processors also face challenges in handling image blur and distortion caused by vehicle vibrations and jolts, further reducing image quality and affecting the accuracy of target detection and tracking, particularly during high-speed driving or on uneven roads. Table 1 summarizes the technical challenges in automotive vision sensors.

### 3. Innovative solutions

Based on the technical challenges faced by intelligent vehicle vision sensors, this study proposes targeted ideas and recommendations, emphasizing innovative solutions and interdisciplinary integration.

#### 3.1. New materials and technologies

To address the performance bottlenecks of conventional vision sensors—particularly in resolution, sensitivity, and light-sensing capabilities—studies must explore new high-efficiency photosensitive materials, such as quantum dots and perovskites. Recent research [18] has highlighted the potential of nanostructured materials in revolutionizing image-sensing technologies by enhancing photon absorption and improving sensor efficiency. These advanced photoelectric materials can significantly enhance key functions of intelligent driving systems, including environmental perception, obstacle detection, and nighttime imaging. To further improve light-capturing efficiency and overall image

quality, leveraging nanostructure design to optimize optical properties is an effective approach. For instance, the study presented in Ref. [19] introduces an all-analog photoelectronic chip designed for high-speed vision tasks, demonstrating how optimized material properties can substantially enhance imaging performance, particularly in high-speed and low-light conditions. Specifically, enhancing photoelectric conversion efficiency via novel approaches based on the photoelectric effect can significantly improve sensor performance. Additionally, investigating nano-optical technologies, such as surface plasmon resonance, offers a viable strategy for increasing light absorption and boosting image clarity under challenging lighting conditions. By integrating these advancements, vision sensors can achieve higher imaging performance, ensuring more reliable perception for autonomous driving in complex real-world environments.

#### 3.2. Architectural evolutions

To address the challenge of balancing resolution and frame rate in vision sensors, it is essential to explore neuromorphic vision perception paradigms inspired by the human visual system. By designing adaptive, bio-inspired vision sensors, perception capabilities can be significantly enhanced in complex environments, enabling more sensitive and reliable visual sensing for autonomous vehicles. For instance, the vision chip described in Ref. [20] features a complementary pathway structure that mimics the human visual system, achieving both high dynamic range and rapid response times, making it suitable for real-time perception in dynamic driving environments. To further improve high-speed imaging performance, researchers have explored ultra-high-speed dynamic imaging technology based on continuous burst photography principles [21]. This approach enables high-resolution imaging of fast-moving targets, enhancing crucial functions such as obstacle detection, dynamic scene perception, and motion trajectory tracking. These high-speed imaging methods enable capturing transient motion patterns with extreme precision, which is critical for autonomous driving scenarios requiring rapid response and high reliability. To optimize image processing efficiency and flexibility, the development of specialized integrated circuits is crucial. For instance, the study presented in Ref. [22] demonstrates an integrated imaging sensor designed for aberration-corrected 3D photography, highlighting how advanced sensor architectures can improve image processing efficiency while reducing distortions. By designing reconfigurable image signal processor architectures and optimizing data transmission performance, vision sensors can achieve higher computational efficiency and adapt to varying processing demands. Additionally, addressing the impact of vehicle vibrations and sudden movements on image quality necessitates the integration of advanced mechanical and electronic image stabilization techniques. Specifically, combining hybrid optical image stabilization and electronic image stabilization systems can dynamically compensate for motion-induced distortions, ensuring stable and high-quality imaging for autonomous driving applications.

**Table 1**  
Summary of technical challenges in automotive vision sensors.

Component	Key challenges	Impact on performance
Optical lens	Diffraction limits, distortion, temperature sensitivity, and vibration-induced defocus	Loss of detail; inaccurate object recognition and distance estimation
CMOS image sensor	Trade-off between resolution and frame rate, limited dynamic range, glare effects, increased noise in low light, and reduced color resolution	Delayed target detection; degraded performance in extreme lighting conditions
Image signal processor	Limited processing speed, inadequate noise reduction, and lack of algorithmic flexibility and anti-shake capability	Reduced real-time performance; misjudgment in dynamic scenarios

### 3.3. Intelligent image processing algorithms

To address the computational load and energy constraints of conventional computing architectures in real-time image processing, it is essential to explore new computational paradigms, such as neuromorphic computing and quantum computing. These emerging technologies present new avenues to develop lightweight, low-power artificial intelligence algorithms that enable efficient, real-time image processing for intelligent driving applications. For instance, the study in Ref. [23] presents a probabilistic model for high-frequency periodic signal detection utilizing event cameras. By leveraging the asynchronous nature of event cameras, this approach efficiently distinguishes target signals from random noise, overcoming the sampling rate limitations of conventional frame-based cameras. To enhance visual perception in complex and dynamic environments, adaptive exposure and color correction algorithms must be explored to mitigate the impact of variable lighting conditions. Furthermore, investigating physics-based de-hazing and de-rain algorithms can improve image quality under adverse weather conditions. Research in Ref. [24] proposes a multi-scale frequency separation network that significantly improves image deblurring efficiency. The network can be extended by incorporating rain droplet refraction models to improve lane detection. For enhanced perception in nighttime and low-visibility scenarios, integrating thermal imaging technology can provide complementary information beyond the visible spectrum, thereby increasing the robustness of autonomous driving perception systems. Furthermore, deep learning techniques can be leveraged to refine image processing and improve recognition accuracy, particularly in object detection and classification tasks under challenging conditions such as high-speed motion and occlusions. To enhance image stability and clarity, real-time image stabilization algorithms based on artificial intelligence must be developed [25]. By analyzing motion patterns between consecutive frames, these algorithms can dynamically compensate for image distortion caused by vehicle vibrations and jolts, ensuring high-quality visual perception in autonomous driving systems.

## 4. Conclusion

In this paper, the key structures, core technologies, and technical challenges of in-vehicle vision sensors for intelligent driving were systematically analyzed. The study revealed that current automotive vision sensors face significant limitations in optical lenses, CMOS image sensors, and image signal processors, such as diffraction constraints, optical distortion, limited dynamic range, glare sensitivity, reduced color resolution, and insufficient processing speed. These limitations collectively hinder the accuracy, stability, and reliability of environmental perception in intelligent driving systems, particularly under extreme lighting conditions, adverse weather, and complex driving scenarios. To overcome these challenges, this paper proposed several innovative solutions: exploring new materials and high-efficiency photoelectric conversion technologies to enhance sensing performance; adopting architectural evolutions inspired by the human visual system to achieve higher dynamic range and ultra-high-speed imaging; and developing intelligent image processing algorithms based on neuromorphic and quantum computing paradigms to enable real-time, energy-efficient, and robust perception in complex environments. Through these advancements, in-vehicle vision sensing is expected to continue evolving towards greater intelligence, reliability, and integration, ultimately enabling safer, more efficient, and fully autonomous intelligent driving in the future.

## CRedit authorship contribution statement

**Xinle Gong:** Writing – original draft, Formal analysis, Methodology. **Zhihua Zhong:** Supervision, Writing – review & editing, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (52102438), the China Postdoctoral Science Foundation (2022M711803 and 2024T170482), and the State Key Laboratory of Intelligent Green Vehicle and Mobility.

## References

- [1] Gouveia LCP, Choubey B. Advances on CMOS image sensors. *Sens Rev* 2016;36(3):231–9.
- [2] Innocent M, Velichko S, Lloyd D, Beck J, Hernandez A, Vanhoff B, et al. Automotive 8.3 MP CMOS image sensor with 150 dB dynamic range and light flicker mitigation. In: Proceedings of the 2021 IEEE International Electron Devices Meeting (IEDM); 2021 Dec 11–16; San Francisco, CA, USA. Piscataway: IEEE; 2021. p. 30.2.1–4.
- [3] Gove RJ. CMOS image sensor technology advances for mobile devices. In: *High Performance Silicon Imaging*. Amsterdam: Elsevier; 2020. p. 185–240.
- [4] Fossum ER. CMOS image sensors: electronic camera-on-a-chip. *IEEE Trans Electron Devices* 1997;44(10):1689–98.
- [5] El-Desouki M, Deen MJ, Fang Q, Liu L, Tse F, Armstrong D. CMOS image sensors for high speed applications. *Sensors* 2009;9(1):430–44.
- [6] Gehrig D, Scaramuzza D. Low-latency automotive vision with event cameras. *Nature* 2024;629(8014):1034–40.
- [7] Liu Z, Hu G, Ye H, Wei M, Guo Z, Chen K, et al. Mold-free self-assembled scalable microlens arrays with ultrasmooth surface and record-high resolution. *Light Sci Appl* 2023;12(1):143.
- [8] Gamal AE, Eltoukhy H. CMOS image sensors. *IEEE Circuits Devices Mag* 2005;21(3):6–20.
- [9] Xu JT, Wang XY, Wang TD, Chen X, Song Z, Lei H, et al. Review on optical visual sensor technology. *J Image Graph* 2023;28(6):1630–61.
- [10] Sukhavasi SB, Sukhavasi SB, Elleithy K, Abuzneid S, Elleithy A. CMOS image sensors in surveillance system applications. *Sensors* 2021;21(2):488.
- [11] Takayanagi I, Kuroda R. HDR CMOS image sensors for automotive applications. *IEEE Trans Electron Devices* 2022;69(6):2815–23.
- [12] Scott-Thomas J. Trends and developments in state-of-the-art CMOS image sensors. Piscataway: IEEE; 2023.
- [13] Theuwissen A. CMOS image sensors: state-of-the-art and future perspectives. In: Proceedings of the ESSCIRC 2007–33rd European Solid-State Circuits Conference; 2007 Sep 11–13; Munich, Germany. Piscataway: IEEE; 2007. p. 21–7.
- [14] Ohta J. *Smart CMOS image sensors and applications*. 2nd ed. Boca Raton, FL: CRC Press; 2020.
- [15] Maheepala M, Joordens MA, Kouzani AZ. Low power processors and image sensors for vision-based IoT devices: a review. *IEEE Sens J* 2021;21(2):1172–86.
- [16] Wang X, Wong W, Hornsey R. A high dynamic range CMOS image sensor with inpixel light-to-frequency conversion. *IEEE Trans Electron Devices* 2006;53(12):2988–92.
- [17] Bandoh Y, Qiu G, Okuda M, Daly S, Aach T, Au OC. Recent advances in high dynamic range imaging technology. In: Proceedings of the 2010 IEEE International Conference on Image Processing; 2010 Jun 26–29; Hong Kong, China. Piscataway: IEEE; 2010. p. 3125–8.
- [18] Iqbal MA, Malik M, Le TK, Anwar N, Bakhsh S, Shahid W, et al. Technological evolution of image sensing designed by nanostructured materials. *ACS Mater Lett* 2023;5(4):1027–60.
- [19] Chen Y, Nazhamaiti M, Xu H, Meng Y, Zhou T, Li G, et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* 2023;623(7985):48–57.
- [20] Yang Z, Wang T, Lin Y, Chen Y, Zeng H, Pei J, et al. A vision chip with complementary pathways for open-world sensing. *Nature* 2024;629(8014):1027–33.
- [21] Huang TJ. Spiking continuous photographing principle and demonstration on ultrahigh speed and high dynamic imaging. *Acta Electron Sinica* 2022;50(12):2919–27.

- [22] Wu J, Guo Y, Deng C, Zhang A, Qiao H, Lu Z, et al. An integrated imaging sensor for aberration-corrected 3D photography. *Nature* 2022;612(7938):62–71.
- [23] Ben-Ezra DEC, Arad R, Padowicz A, Tugendhaft I. Probabilistic approach for detection of high-frequency periodic signals using an event camera. *Nat Comput* 2026;22:99–110.
- [24] Zhang Y, Li Q, Qi M, Liu D, Kong J, Wang J. Multi-scale frequency separation network for image deblurring. *IEEE Trans Circuits Syst Video Tech* 2023;33(10):5525–37.
- [25] Li H, Hua Q, Shen G. Tianmouc vision chip designed for open-world sensing. *Sci China Mater* 2024;67(9):3046–8.