Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD27260

Research and Application of Robot Vision Sensors

Yi Yang

University of Science and Technology Beijing, Beijing, China
u202343143@xs.ustb.edu.cn

Abstract. Vision sensors are important tools for robots to know the outside world. They have been used in many different fields. These sensors help robots know their surrounding environment clearly. This helps make robots more intelligent. Robots are changing from just following orders to making decisions on their own. Vision sensors play a very important role in this change. The development of sensor technology is closely linked to big improvements in robots' sensing ability. This paper aims to study how robot vision sensors work, their use cases and the challenges they meet. It analyzes current vision sensor technologies and their development trends. The paper puts forward future research directions. It also sums up the advantages and future uses of vision sensors in the robot field.

Keywords: robot vision sensors, imaging principles, data processing, deep learning, multi-sensor fusion

1. Introduction

In recent years, AI technology has made many big improvements. Robots are now widely used in factories, hospitals, daily services and other fields. This has become a main trend of technology development. Robot vision system is one of the most important abilities for robots to know the world. It makes robots more independent and much smarter. Vision sensors are key tools for robots to learn about the outside world. They are very important when robots finish different jobs. With vision sensors, robots can see things, tell them apart, find positions and move by themselves. They can finish many difficult tasks well. In environments that are complex and always changing, robots use vision sensors to get information around them. They can know objects, follow targets and stay away from barriers. This makes sure that robots finish tasks quickly, correctly and safely. So, keeping developing and improving vision sensor technology is one of the basic things for robots to be used widely.

As science and technology keep moving forward, vision sensor technology has new changes all the time. These new changes help robots a lot in knowing their environment. Vision sensors are very useful, especially in smart factories, self-driving cars, drones and service robots. In the field of intelligent manufacturing, robots can accurately complete assembly, inspection, handling and other tasks through vision sensors, greatly improving production efficiency and product quality. The continuous development of autonomous driving technology relies on high-precision vision sensors, which can obtain real-time road condition information and make quick decisions to ensure driving safety and smooth traffic. In the field of service robots, vision sensors help robots recognize and understand human actions and needs, enabling robots to provide intelligent services in home, medical, education and other scenarios. In-depth research on the technology and application of robot vision sensors can not only promote the further development of robot technology but also promote the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD27260

automation process of various industries, improve overall production efficiency, enhance people's quality of life, and make important contributions to the construction of an intelligent society.

At present, domestic and foreign research in the field of robot vision sensors is gradually developing from traditional image processing technology to deep learning technology and the fusion of multi-modal sensors. Traditional vision sensors are mainly based on Charge-Coupled Device (CCD) and Complementary Metal-Oxide-Semiconductor (CMOS) technologies. These sensors are widely used in robot vision systems and can perform basic image capture and processing. Domestic enterprises such as Xiling Vision have developed high-speed sensor chips with a reaction speed of 100,000 frames per second, which significantly improves real-time response capabilities. However, problems such as target tracking in dynamic scenes and adaptability to illumination changes remain to be solved, and the accuracy and robustness of multi-sensor fusion are still research hotspots [1]. In recent years, with the rapid development of deep learning and computer vision technology, image recognition and understanding technology based on deep neural networks have made significant breakthroughs, and the ability of robots to understand the environment in complex scenes has been greatly enhanced. At the same time, depth perception technologies such as Light Detection and Ranging (LiDAR) and structured light sensors have gradually moved to the forefront of applications and become important supplements to current robot vision systems. LiDAR obtains three-dimensional information of the surrounding environment by emitting laser beams and receiving reflected signals, which has high accuracy and broad application potential; structured light sensors use optical principles to achieve high-precision measurement of the surface shape of objects, and play an important role in object modeling and depth perception.

Though image recognition and scene understanding have made some progress, robots still face many problems in changing and complex environments. These problems include their ability to fit different situations, working speed and result correctness. For example, in busy indoor or outdoor places, it is hard to recognize and follow moving objects. It is also difficult to deal with light changes, blocked views and blur from moving. These are still key points of today’s research. Besides, how to mix data from vision sensors and other sensors like laser sensors and sound sensors is also a hot research topic. This can help robots know the world more stably and correctly.

This paper talks about robot vision sensors as important parts for robots to know the world. It tries to study carefully how they work, their different uses and main problems. It looks at current vision sensor technologies from old CCD and CMOS sensors to new ones like laser sensors, light pattern sensors and sensors using deep learning, as well as their development trends. The paper wants to show the good points and future uses of vision sensors in robots. It especially looks at their roles in factory robots, service robots and self-driving robots. What’s more, it gives ideas for future study to solve main problems. These problems include fitting changing complex environments, working speed, result correctness and mixing data from many sensors. The final goal is to make robot technology better and help more industries become automatic.

2. Principles of vision sensors

2.1. Definition and classification of vision sensors

Vision sensors work based on light imaging technology. Light enters the sensor through the lens. The image sensor chip then turns the light into pictures that can be used for analysis. The quality of these pictures and the sensor’s resolution directly affect how well a robot understands the environment. With different imaging methods like binocular vision and ToF technology, robots can sense depth and locate objects more accurately.

Vision sensors are very important in modern tech uses, especially for picture processing, object recognition and depth sensing. Based on how they work, vision sensors have two main types: passive

DOI: 10.54254/2755-2721/2025.LD27260

sensors and active sensors. Each kind has its own benefits and suitable uses. Passive sensors take in light from natural or man-made light sources to make pictures. They do not need extra light sources, so their working way is simple. Common passive sensors are CCD sensors and CMOS sensors. CCD sensors turn light signals into electric signals by charge coupling and deal with them through circuits. They work well in dark places and are often used in high-end cameras and monitoring systems. But CMOS sensors turn light signals into electric signals through combined circuits. They cost little, use less power and are highly integrated. So they are more widely used, like in phone cameras and monitoring cameras. Passive sensors use little power and need no extra signals. They can process pictures well in good light. But their picture quality may drop a lot in dark environments.

Active sensors send out signals like light, laser or infrared on their own and take in the reflected signals. They can get 3D information about objects' distance and shape. Common active sensors include LiDAR, structured light sensors (such as open-source real-time systems for construction robotics [2]) and ToF sensors. LiDAR sends out laser beams and measures the return time to work out object distance. It can get exact 3D space information and is widely used in self-driving cars, drone flights and other fields, especially in complex environments. Structured light sensors send fixed light shapes onto object surfaces. They check the shape change to calculate objects' 3D form and are often used for face recognition and object scanning. ToF sensors get scene depth data by measuring the time between sending out light pulses and receiving the reflected ones. This way is accurate and fast. It is mostly fit for uses that need quick depth sensing.

2.2. Imaging principles

Vision sensors take pictures by using optical imaging technology. Light goes into the sensor through the lens. The image sensor chip then processes the light and makes pictures that can be used for analysis [2]. The quality of the picture and the sensor's resolution directly influence how well a robot understands the environment. With different imaging ways like binocular vision and ToF technology, robots can know depth and find objects more exactly. Vision sensors work by catching light signals from objects. The lenses focus light onto image sensor chips to make pictures. The sensor notices the strength, color and direction of light. It then turns light into electronic signals to make image data. Different imaging technologies help vision sensors give high-quality visual information.

When making images, light passes through the lens and reaches image sensors like CMOS or CCD sensors. These sensors change light signals into electrical signals. After processing, they make digital pictures. The resolution of the sensor, how well it catches light and the quality of the lens system all affect the clearness and correctness of the final picture. This then influences how well a robot knows the environment.

Different imaging ways bring different visual results. For example, binocular vision technology uses two cameras to copy human eyes. It can get depth information of a scene and build 3D models. This makes robots better at knowing space. ToF technology gets depth information by measuring the time between sending out light pulses and receiving them back. This way is very correct and fast. It is best for uses that need real-time depth sensing.

2.3. Data processing methods

Data processing methods are important parts for vision systems to get good perception. They usually include three steps: image pre-processing, feature extraction and target recognition. These steps work together to improve the accuracy and efficiency of image analysis. They let computers understand and study images just like humans. The goal of image pre-processing is to improve the quality of original images. It can remove noise, make key features clearer and lift image contrast. This builds a good base for later feature extraction and target recognition. Common image pre-processing ways

DOI: 10.54254/2755-2721/2025.LD27260

include denoising, enhancement and edge detection. Denoising often uses filters like mean filtering and Gaussian filtering to remove image noise and make images clearer. Enhancement changes the brightness and contrast of images to show more details in pictures. Histogram equalization is a common enhancement method. Edge detection takes out edge information from images to help later analysis. Feature extraction means getting useful features from images. It usually includes edge features, texture features and color features. These features show key information such as the shape, texture and color of objects in images. They help algorithms know what objects look like. For example, the Sobel operator is used to find image edges. GLCM is used to study texture features. HSV color space is used to get color information. Target recognition means identifying objects in images through feature matching, machine learning or deep learning algorithms. Feature matching compares with the feature library of known objects to find matching areas and locate objects. Traditional machine learning algorithms like SVM and K-Nearest Neighbors use lots of training data to classify and predict objects. Deep learning, especially CNNs, has made great progress in target recognition. It can automatically get high-level features in images through multi-layer convolution and pooling operations. This helps classify and locate objects efficiently. In recent years, deep learning models like YOLO and Faster R-CNN are widely used in real-time object detection. By combining the three methods above—image pre-processing, feature extraction and target recognition—robots can get useful information from original images. They can recognize objects correctly, know the environment well and make decisions. So they can finish tasks well such as path planning and obstacle avoidance in real uses.

3. Applications of robot vision sensors

3.1. Industrial robots

In the industrial field, vision sensors are often used for positioning, assembly, welding and painting on automatic production lines. With vision sensors, robots can finish part positioning, checking and assembly jobs with high exactness. This makes sure product quality and production speed are good. For example, industrial robots can find broken parts through vision sensors to make sure products meet quality standards.

In modern manufacturing, industrial robots are one of the key technologies to make production faster and better. Using vision sensors on industrial robots helps them do many jobs in a smarter way. This is especially true for difficult assembly, checking and processing work. With vision sensors, industrial robots can do exact positioning and work without people's help. This makes production lines much more automatic [3].

On automatic production lines, vision sensors are often used to position parts exactly and put them together automatically. Industrial robots get 3D pictures of parts with cameras or laser scanners. They use computer vision rules to study these pictures and get real-time information. The information includes where parts are, how they are placed and how big they are. With this information, robots can change their work ways to make sure parts are put in the right place correctly. Vision sensors are very exact, so robots can make few mistakes in assembly. This improves production speed and makes products more consistent. This is supported by good calibration ways for line structured light vision sensors in robot systems [4].

Besides, industrial robots use vision sensors to find broken parts and control quality. This is very important to make sure products meet standards. With high-resolution cameras or infrared imaging tools, robots can find small problems on or inside parts. These problems include cracks, small holes and surface scratches. Vision sensors help robots watch every production step in real time. They can find and fix possible problems in time. This technology is very important in car making, electronic

DOI: 10.54254/2755-2721/2025.LD27260

product making and metal processing. It can greatly reduce quality problems from people's mistakes and make products more reliable [5].

Welding and painting are common jobs for industrial robots. These jobs need very exact work and stable performance. Vision sensors help robots watch where workpieces are and their surface conditions in real time. Then robots can do exact welding or painting. For example, vision sensors help robots find where welding starts and ends. This makes sure welding quality is good. In painting, vision sensors can check color and shine of the painted area. This makes sure painting is even. So the product surface looks good and works well.

As manufacturing keeps developing, industrial robots are becoming smarter and more modular. Vision sensors are an important part of them. They make robots more flexible when doing different production jobs. For example, robots can change task ways by themselves when production needs change. This makes production lines more flexible and reduces time for people to step in and adjust. This is especially useful for making many kinds of products in small numbers. Vision sensors help robots fit different product production needs quickly.

3.2. Service robots

Service robots are a new kind of smart machine. In recent years, they have got much attention and research in many areas, especially homes, hospitals and business places. Vision sensors become more and more important for service robots. They make service robots smarter and work more efficiently. With vision sensors, service robots can know the complex world and talk with it right away. So they can finish tasks better and give right services. In home cleaning robots, vision sensors help robots make good maps, plan cleaning ways and stay away from obstacles. This is supported by low-power image sensor technologies for vision-based IoT devices [6]. The robot gets real-time information about the room. Then it can plan the cleaning path correctly to cover every corner. Vision sensors help robots know furniture, obstacles and walls. They also stop the robot from hitting things or missing places. This makes cleaning faster and better. As deep learning and image recognition technology develop, cleaning robots can use right cleaning ways for different floors and dirt. For example, robots can tell carpets and hard floors apart. They choose different cleaning ways for different surfaces. They can even know what kind of dirt it is to clean better.

In medical care, service robots become more and more important. Today, people are getting older and older. So we need more medical care robots. Vision sensors help robots know how healthy patients are and do some simple care work. For example, vision sensors can look at patients' faces and body changes. They can tell if patients feel pain or uncomfortable. Then the robot tells nurses or doctors in time. Besides, robots can know things like medicine bottles and medical tools with vision sensors. This makes robots work more on their own and makes medical services more automatic [5].

In business and service jobs, vision sensors also help make service robots smarter. Take stores, hotels and airports for example. Service robots can guide customers and send things with vision sensors. They know where customers are and what they need right away to give more personal services. At the same time, vision sensors help robots watch for safety. They can find dangerous things in time, like strange people or sudden accidents. Then the robot tells workers to keep public places safe. In hotels, robots can talk with customers through vision sensors. They give suggestions and answer questions to make customers feel better.

3.3. Autonomous driving robots

Self-driving technology is a big progress in modern robot technology. It is slowly changing the transportation field. Vision sensors are very important for self-driving robots. They help robots know the world around them correctly. These sensors include high-definition cameras, LiDAR and

DOI: 10.54254/2755-2721/2025.LD27260

millimeter-wave radar. They work together to make sure the self-driving system can make fast and right decisions in different difficult environments. High-definition cameras are the most common vision sensors. They can tell traffic signs, lane lines, people, obstacles and other cars' positions and speeds. They give necessary road information to the system. LiDAR uses laser beams to scan the world around and makes 3D point cloud data. It gives correct depth information and fixes the weak points of vision sensors in dark or complex environments. LiDAR can still make high-precision environment models, especially in bad weather like haze and night. Besides, millimeter-wave radar uses electromagnetic waves to find objects that reflect signals. It has strong anti-interference ability and can go through heavy rain and fog. It helps the system know how far and how fast the obstacles in front are. It plays an important role in stopping crashes and avoiding obstacles [7].

4. Future development analysis

4.1. Technical advantages

Vision sensors have many clear strengths in robot technology. But when people use them in real life, they still meet many problems and limits that need to be fixed. Here are some main points.

4.1.1. Core role in robotic perception

As a key part of a robot's sensing system, vision sensors give exact and enough information about the environment. Their main technical strengths include clear picture quality, fast response, and measuring without touching objects. In difficult environments, they catch and process picture data well. This helps robots know the environment and make decisions.

4.1.2. Superior high-dimensional perception

Compared with old sensors, vision sensors get more kinds of information from picture details. These include object shape, color, surface feel and moving path. This is very important for robot navigation, target recognition and task completion.

4.1.3. High resolution

This clear feature helps sensors catch small objects and tiny changes. In object handling and assembly work, it finds the position and state of things exactly to make sure work is done well. In self-driving cars, clear pictures catch details like traffic signs, people, road lines and other road users. They give quick feedback to keep driving safe.

4.1.4. Strong real-time performance

They take pictures faster than old sensors like radar and ultrasonic sensors. They show changing situations in the environment quickly and give clear, immediate feedback. This is especially useful in changing scenes. For example, when meeting fast-moving targets, vision sensors can react and analyze quickly. They help robots make decisions in time.

4.1.5. Non-contact measurement

This is a special good point. Different from touch and force sensors that need to touch objects directly, vision sensors finish measuring and recognizing by taking pictures from far away. They do not bring trouble from real touch. They are more flexible and safe when measuring easy-to-break things and doing monitoring work.

DOI: 10.54254/2755-2721/2025.LD27260

4.1.6. Synergy with deep learning and AI

Working closely with deep learning and AI makes vision sensors more useful. With the help of deep neural networks like CNNs, vision sensors can learn from lots of data. They get useful information and recognize targets more exactly and quickly. They can tell many targets in pictures at the same time and deal with fast-changing things in changing environments. This lets robots track and recognize still and moving objects. People use this widely in industrial robots, self-driving cars, smart monitoring and other fields.

4.1.7. High-level scene understanding

When working with AI, vision sensors can do more than just recognize objects. They can study how things work with each other and how they connect with the environment. This “smart vision” creates more difficult and flexible use scenes. For example, in intelligent home systems, vision sensors can tell home things and know if they are being used. Then they can make smart control and interaction possible.

4.2. Existing problems

4.2.1. Environmental sensitivity

Vision sensors are very easy to be influenced by light changes, blocked views and reflected light. This is a main problem. Different lighting conditions greatly affect image quality and make recognition results unstable. It is hard to keep stable work in outdoor or complex indoor places. For example, strong direct light, shadows or dark environments will lower detection accuracy and make sensors less fit for the environment [8].

4.2.2. Excessive demand for computing resources

This is a key limit. Processing visual data needs a lot of image analysis and complex calculations. Fast-response situations need strong processing power and high-speed working ability. Most vision sensors need high-performance processors and good algorithms to work quickly. But in small, low-power machines with limited resources, processing often slows down and works poorly. This makes the whole system less efficient.

4.2.3. Insufficient adaptability to dynamic environments

This is a big problem when objects move fast and the environment changes often. These situations need more stable work and more exact recognition. Deep learning has made progress, but sensors still react slowly and recognize poorly when facing fast-changing targets or messy backgrounds. They often recognize wrongly or fail to track things in time. This is especially true for fast-moving objects or sudden events.

4.2.4. Multi-target recognition and accuracy issues

There are clear problems when many objects appear at the same time. It is hard to tell and recognize these targets correctly. Present algorithms and sensors often make wrong detections, miss detections or recognize unclearly. This happens when targets are close to each other and backgrounds are messy. It stops robots from knowing the environment well and finishing tasks correctly.

DOI: 10.54254/2755-2721/2025.LD27260

4.2.5. Hardware costs and maintenance difficulties

These problems stop vision sensors from being widely used. Even if their work ability gets better, they still cost a lot. They need high-resolution cameras, good lenses and strong processing hardware. These things make the whole system costly. Also, sensors have strict rules for using environments. They need regular check and reset. This raises daily costs and makes maintenance more complex.

4.3. Development trends

As technology keeps improving, robot vision sensors will develop in several important ways in the future.

4.3.1. High integration and multi-functionality

This is a main development direction. Future vision sensors will not only take pictures as a single function. They will combine different sensing technologies such as infrared, LiDAR and sound sensing to build multi-modal sensing systems. This combined design helps robots know the world correctly in more complex environments. It also makes robots better at collecting and processing different kinds of information.

4.3.2. Combination of low power consumption and edge computing

This will make vision sensors more useful in real life. With the development of IoT and edge computing, data processing will not totally depend on cloud servers. Some tasks will be done on local machines instead. This cuts down delay, makes response faster and saves energy effectively. Low-power vision sensors will be more fit for portable machines or robot systems that work for a long time. This is especially true for robots that need to work on their own for a long time.

4.3.3. Deep integration of AI and deep learning

This will make vision sensors much smarter. Future deep learning models will become smaller and more adaptable. On one hand, people can make models smaller by knowledge distillation and pruning. For example, YOLOv8n has 30% fewer parameters than the old version. It can be used on small processors. On the other hand, online learning will be added. Sensors can change their detection rules and feature extraction ways according to real-time situations. Test data show that dynamic adaptive models are 22% better at tracking targets in complex blocked situations than static models. They make recognition much more stable in environments with many moving targets [9].

4.3.4. Emergence of new sensor technologies

New sensor technologies may bring new development chances. For example, brain-like vision sensors copy the human visual system. They make processing speed and energy use much better. Quantum sensors use quantum rules. They can find very small changes with high sensitivity. They have a bright future in precise measurement in the future. This is in line with the development of event-based vision [10].

4.3.5. Expansion of intelligent-automated application scenarios

This will help vision sensors be used more widely. From self-driving cars and drones to industrial robots and smart homes, vision sensors are becoming more and more important. By working with other sensors and smart algorithms, robots will understand and fit complex changing environments better. They can make smart decisions more efficiently.

DOI: 10.54254/2755-2721/2025.LD27260

5. Conclusion

The application of vision sensors in robot technology has made significant progress, and they have great potential in industrial automation, smart homes, medical services and other fields in the future. With the continuous development of technology, vision sensors will empower robots, enabling them to achieve more efficient and accurate operations in more complex and changeable environments. However, in the face of problems such as environmental adaptability and computational complexity, in-depth technical research and innovation are still needed.

References

[1] Yasuda, Y. D. V., Martins, L. E. G., & Cappabianco, F. A. M. (2024). Vision-Based Autonomous Navigation of Mobile Robots: A Review. Journal of Intelligent & Robotic Systems.

[2] Lam, T. F., Blum, H., Siegwart, R., & Gawel, A. (2022). SL Sensor: An open-source, real-time and robot operating system-based structured light sensor for high accuracy construction robotic applications. Automation in Construction.

[3] Li, Y., Zhu, J.-Y., Tedrake, R., & Torralba, A. (2020). Connecting Touch and Vision via Cross-Modal Prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Wang, Z., Fan, J., Jing, F., Deng, S., Zheng, M., & Tan, M. (2020). An Efficient Calibration Method of Line Structured Light Vision Sensor in Robotic Eye-in-Hand System. IEEE Sensors Journal,20(11),6200-6208.

[5] Li, B., Bai, J., Qiu, S., Wang, H., & Guo, Y. (2023). VITO-Transformer: A Visual-Tactile Fusion Network for Object Recognition. IEEE Transactions on Instrumentation and Measurement.

[6] Maheepala, M., Joordens, M. A., & Kouzani, A. Z. (2021). Low Power Processors and Image Sensors for Vision-Based IoT Devices: A Review. IEEE Sensors Journal,21(2),1172-1186.

[7] Li, Y., & Liu, Y. (2021). Vision-based Obstacle Avoidance Algorithm for Mobile Robot. 2020 Chinese Automation Congress (CAC).

[8] He, B., Wang, Z., Zhou, Y., Chen, J., Singh, C. D., Li, H., Gao, Y., Shen, S., Wang, K., Cao, Y., Xu, C., Aloimonos, Y., Gao, F., & Fermüller, C. (2024). Microsaccade-inspired Event Camera for Robotics. Science Robotics.

[9] Ren, H., & Qureshi, A. H. (2023). Robot active neural sensing and planning in unknown cluttered environments. IEEE Transactions on Robotics and Automation, 39(4), 2738-2750.

[10] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., & Censi, A. (2019). Event-based Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.