Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.LD30577

Robotics Vision Sensor Technology and Its Current State of Development

Linjie Hu

Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Canada
tonyhuforever@gmail.com

Abstract. Robotic Vision Sensor Technology forms the basis for perception in automatic systems, making it possible for machines to interpret and interact with their surrounding environments. This article provides a systematic overview of different vision sensor technology and their operating principles, ranging from photodetector arrays to radar. The paper then analyzes the practical applications of vision sensor technology in three demanding scenarios, and suggests the need for multi-sensor fusion for more accurate vision information and robust automation. In the following section, the paper reviews the present advantages and challenges of multi-sensor systems in addition to their future developments in order to give a consolidated overview of the state of robotic vision.

Keywords: Robotic Vision, Sensor Fusion, Multimodal Perception, Sensor Technology

1. Introduction

Vision sensors are one of the most elemental components in an autonomous system because they allow the system to identify and respond to complex service environments. From conventional cameras to advanced depth-sensing devices, vision sensors can transform visual inputs into information that can be used in robotic operational tasks such as navigation, object detection and categorizations. In order for robots to operate autonomously, they must have the ability to perceive their surrounding environment accurately and reliably, which demonstrates the importance of developing vision sensor technology.

Recent research in robotic vision attempts to address inherent limitations of any single sensor through multi-sensor fusion and utilization of state-of-the-art artificial intelligence. For instance, many studies have been conducted to develop deep learning models that can combine heterogeneous signals obtained from multiple sources, such as cameras, light detection and ranging (LiDAR) and radar, to create more accurate and comprehensive environmental models [2]. Simultaneously, some research is exploring innovative sensor technologies, such as bio-inspired event-based cameras, which offer high dynamic range and low latency by asynchronously detecting per-pixel intensity changes, presenting a promising alternative for high-speed robotic applications [3]. These research aims to increase the perceptual robustness, accuracy, and efficiency of autonomous systems in complex real-world environments.

This article provides an integrated analysis of the current state of development of robotics vision sensor technologies. First of all, the working principles of some of the most important sensor types

© 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.LD30577

including photodetector arrays, Time-of-Flight (ToF) systems, triangulation-based sensors, and radio wave-based sensing are analyzed. This paper proceeds to elaborate on their practical application in real-world contexts to demonstrate the importance of multi-sensor fusion. Lastly, the discussion section describes the advantages and shortcomings of current technologies and explores future developmental trends of autonomous systems.

2. Principle of visual sensor

In the field of robotics perception, robotics visual sensors are of great importance. The operating principles of robotics visual sensors can be generally categorized into two types: passive and active, depending on whether sensors rely entirely on passive ambient signals or are able to emit their own signals for object detection. The following section describes some common types of robotics visual sensors and their operation principles.

2.1. Photodetector array based sensing

This sensing principle relies on one or two dimensional arrays of photosensitive elements that convert photons into an electrical charge. The most common implementations are Charge-Coupled Device (CCD) and Complementary Metal-Oxide-Semiconductor (CMOS) sensors.

2.1.1. 2D cameras

Digital cameras are the most common application of 2D imaging systems which rely on receiving passive lights. Specifically, when a light source emits light onto a certain object, a portion of the light is reflected by the object. The reflected light is then captured by the lens on the cameras that projects the light onto a photo detector array sensor. Each sensor pixel measures the intensity of the light and generates electrical signals that are then converted into a 2D digital image. These digital images provide good color, texture and shape information of the object but cannot directly reflect depth information. A stereo 2D camera setup or a fusion with other types of sensors is required to obtain precise depth information.

2.1.2. Hyperspectral and multispectral cameras

This spectral imaging system has become increasingly important in various fields, from precise agriculture to environmental monitoring. Hyperspectral imaging further extends the photodetector array to capture ultra-fine spectral information across various narrow wavebands, allowing material identification based on unique spectral characteristics. Though the hyperspectral imaging can achieve high spectral resolution, it requires complex data processing and longer processing time. In contrast, the multispectral imaging sacrifices some resolution by capturing few and broader wavebands from objects, in order to achieve less data processing time and higher cost-effectiveness.

2.1.3. Infrared cameras

Infrared cameras contain a passive imaging system that detects and measures the long-wave infrared radiation the objects emit. The infrared-sensitive photodetectors convert incident infrared radiation into electronic signals based on the thermal properties of the object, generating thermographic images. This sensing

DOI: 10.54254/2755-2721/2025.LD30577

technology allows for a better perception of objects in darkness and through visual obscurants such as smoke and fog.

2.2. Time-of-Flight sensing

The Time-of-Flight (ToF) sensing is an active sensing principle. Unlike the passive sensing system, the ToF sensing system emits light signals and estimates the distance to objects by measuring the time it takes for the emitted light to travel to the objects and back to the sensor.

2.2.1. ToF cameras

ToF camera systems emit a modulated near-infrared (NIR) light source onto the whole scene. A dedicated sensor array then captures the reflected light signals, measuring the phase shift between the emitted and returned signals. The measured phase difference is then converted into distance, generating a sense depth image in real time. These depth images are robust and compact, but their performance can be affected by multipath reflections and strong ambient lighting conditions.

2.2.2. LiDAR

LiDAR is a type of ToF system that emits focused laser pulses and measures the distances to objects by calculating the time it takes for the laser pulses to hit the objects and return to the sensor. The sensor then creates a high-resolution 3D point cloud that reflects highly accurate 3D structural information of an environment.

2.3. Triangulation-based sensing

Triangulation-based sensing refers to an active sensing technique that extracts depth information based on geometric triangulation; a popular application is the structured light cameras. In structured light cameras, a projector emits a predefined pattern of light onto the scene. Then an offset camera observes how this pattern deforms when it hits objects at various depths. This system then computes distance to each point by comparing the projected and observed patterns and establishes a dense depth map. While this sensor can provide high accuracy at short ranges, it also has a very high sensitivity to ambient light and a narrow operating range compared to ToF systems.

2.4. Radio wave based sensing

This principle uses radio frequency waves instead of light to detect objects and measure their distance and properties. The most common type of radio wave based visual sensor is radar (radio detection and ranging). Radar is an active sensor that transmits radio wave pulses and analyzes the reflected signals, the time delay of the reflected signals are used to determine the distances. Its key advantage is exceptional robustness to adverse weather conditions (rain, fog, snow) and its ability to directly measure velocity. However, it offers much lower spatial resolution than optical sensors, making fine-grained object recognition difficult.

3. Applications

This section analyzes three advanced applications of robotic vision sensor systems, and demonstrates how the strategic integration of complementary sensor technologies can successfully

DOI: 10.54254/2755-2721/2025.LD30577

overcome the limitations of any single type of sensor and allow robotic systems to operate in a more reliable way in various service environments.

3.1. Drone-based automated tracking

The autonomous detection and tracking of an individual by an unmanned aerial vehicle (UAV) is an advanced performance task that involves reliable perception, localization, and control mechanisms. This technology has also been increasingly used in cinematography, security monitoring, and search and rescue missions. Due to the fact that real-life environments are diverse and unpredictable, no single type of visual sensor can remain stable and accurate under all service conditions. Thus, multimodal sensor fusion has become the key to solve this difficulty.

The integration of various sensor tracking systems often starts with target detection and identification. The RGB camera in general is particularly important in this step, which provides not only high-resolution images of spatial features, but also texture information. Deep learning algorithms then provide techniques for real-time object recognition and semantic segmentation of input images. In cases of person-specific tracking, RGB imagery offers potential for advancements using applications like re-identification based on clothing properties or discriminative correlation filters. Nevertheless, 2D cameras are not built to accurately measure depth, which is essential for safely separating UAVs from the tracked subject and avoiding collisions. To compensate for this limitation, depth sensors are incorporated into the sensor unit. Specifically, ToF cameras provide some great advantages because they can yield dense depth map images at higher frame rates, and generate intensity images. By aligning depth information with normal RGB images, UAVs can identify and locate the target in 3D space. This improvement enables more responsive adjustments of the UAVs' velocity and positioning, making it possible for UAVs to maintain a predetermined distance from the target. Furthermore, the lightweight structure and minimal reliance on moving components make modern ToF cameras especially suitable for small UAV platforms where weight and energy efficiency remain critical concerns [4]. However, the ToF sensors also have shortcomings. Their sensitivity to external infrared interference can be problematic in environments with strong lightings. Under such environments, LiDAR is a less susceptible alternative since it offers coherent scene interpretation at large distances. Moreover, thermal infrared cameras can be added in the presence of adverse lighting conditions to identify a clear thermal signature from humans to facilitate consistent detection and tracking [5].

The overall performance of the tracking system is essentially dependent on the sensor fusion algorithm, frequently utilizing Kalman filter or Bayesian Inference Framework. These methods integrate positional signals from ToF or LiDAR, visual data from RGB cameras, and thermal inputs from infrared sensors to form a collective estimate of target position. This multimodal method improves the robustness against the failure of individual sensors. For example, the ToF information might be degraded when a target moves rapidly from bright sunlight into the shade, but the multimodal system can still remain consistent tracking by switching to the RGB and thermal inputs until the ToF sensor regains stability.

In summary, the utilization of UAVs to autonomously track and follow individuals underscores the synergy achievable through combined use of various types of sensing technologies. UAVs with multi-sensor systems can overcome the inherent limitations of any single sensor, achieving a high level of perceptual robustness.

DOI: 10.54254/2755-2721/2025.LD30577

3.2. Autonomous driving through multi-sensor fusion

Autonomous driving represents one of the most challenging applications in the field of robotic vision, necessitating exceptional levels of reliability, accuracy, and real-time responsiveness across a wide range of environmental contexts. A primary obstacle in achieving comprehensive situational awareness stems from the limitations of individual sensing modalities; no single sensor can independently provide a complete perception of a vehicle's surroundings. To address this inherent shortcoming, autonomous vehicles (AVs) employ an array of complementary visual sensors, integrating their outputs through sensor fusion techniques to establish a resilient and redundant model of the environment. This approach is fundamentally important for ensuring safe navigation, precise obstacle avoidance, and effective route planning.

A standard multi-sensor configuration typically incorporates cameras, LiDAR, and radar, each contributing unique functionalities. Cameras offer high-resolution semantic information crucial for understanding complex scenes, such as identifying and classifying objects - including pedestrians, other vehicles, and traffic signs - as well as recognizing text and interpreting traffic lights and lane markings. Notably, as passive devices, cameras are highly sensitive to lighting conditions. Their performance substantially compromised during nighttime or under low illumination. Furthermore, conventional cameras lack the ability to directly measure object distance.

To supplement these deficiencies, LiDAR systems, which operate based on time-of-flight principles, generate accurate and dense 3D point clouds of the surrounding environment. LiDAR excels at providing reliable distance measurements and generating detailed geometric representations that support tasks like localization and obstacle detection. Additionally, it demonstrates proficiency in outlining vehicles and infrastructure. Nevertheless, LiDAR performance decreases significantly in unfavorable weather conditions such as heavy rain or snow, or when encountering reflective or poorly reflective surfaces.

Radar complements cameras and LiDAR systems by functioning at radio frequencies, which are less affected under adverse weather conditions. Its most outstanding feature lies in its capacity to accurately determine the radial velocity of moving objects via the Doppler effect, which is particularly invaluable for planning driving routes and preventing collisions. The main drawback of radar is, however, its relatively coarse spatial resolution comparable to cameras and LiDAR, limiting its effectiveness in fine-grained object recognition and precise shape estimation.

Therefore, a multi-sensor fusion system is crucial for safe and reliable autonomous driving. Sensor fusion in autonomous vehicles may be implemented at various hierarchical levels. The data-level fusion involves combining raw sensor measurements such as mapping LiDAR points onto camera images. At the feature-level, extracted characteristics such as edges or corners are fused. The most widely adopted approach in modern systems is decision-level fusion, where each sensor processes its data independently to generate object lists that are subsequently combined by a central fusion algorithm. This methodology enables the utilization of the distinct strengths of each modality. For example, cameras might identify a distant vehicle as a truck, then LiDAR would quantify its exact location and dimensions, while radar determines its velocity. In scenarios where visibility and performance of both camera and LiDAR are compromised by conditions like thick fog, radar continues to provide reliable tracking, maintaining uninterrupted environmental perception [6].

In conclusion, the adoption of complementarity through sophisticated multi-sensor fusion frameworks, autonomous vehicles attain robust operational safety across diverse real-world circumstances. This architecture not only cross-validates perceptual data between different modalities but also maintains reliable operation even when individual sensors degrade or fail [7].

DOI: 10.54254/2755-2721/2025.LD30577

3.3. Multispectral imagery in agriculture

The use of multispectral and hyperspectral imagery in agriculture, also known as precision agriculture, has made a significant difference in crop management by surpassing the limitations of the human eye. Farmers become able to monitor crop health closely, allocate resources wisely, and increase yields significantly with the help of precision agriculture. Precision agriculture usually relies on aerial platforms like UAVs or satellites mounted with specialized sensors to measure the reflectance data of crops at specific wavelengths. These reflectance data can show symptoms of physiological stress in crops before the symptoms can be observed with human eyes.

In precision agriculture, most frequently used sensing methods are multispectral and hyperspectral. As standard RGB cameras only record visible red, green, and blue light, these sensors can also detect more spectral bands, such as near-infrared (NIR) and the red-edge regions. Because healthy crops have distinct leaf cellular structure and absorb visible red light during photosynthesis, their spectral signatures are unique and have a high reflectance in the NIR band. It's in these special characteristics that multispectral and hyperspectral spectra can be converted to vegetation indices which can objectively measure crop health. The Normalized Difference Vegetation Index (NDVI) is one of the most widely used indices. It is mathematically expressed as (NIR – Red) / (NIR + Red). Higher NDVI means healthy crops, and lower NDVI indicates stress and disease on the crop. Moreover, hyperspectral cameras are capable of recording very narrowly spaced spectral channels and, therefore, they may provide a more precise diagnosis. Their output describes low-level spectral irregularities caused by nutrient deprivation, water shortage, early-stage fungal infection and pest penetration, which can be identified by different characteristic patterns in reflected spectral properties [8].

These spectral outputs can be combined with high-resolution RGB images and accurate geographic data to generate practical agronomic information, thus most agricultural drones integrate both multispectral and traditional RGB cameras. Although the RGB images give researchers important spatial data to separate different vegetation groups, each image is also georeferenced by GPS tagging to retain accurate field alignment. During the next stages of data processing, researchers integrate RGB and multispectral datasets to produce orthomosaic plots of the whole field. NDVI maps or alternative vegetation indices of vegetated portions can be colour-coded and represent the variations in plant health across the field. These plant health maps can then further be directly interfaced with modern agriculture variable rate technology (VRT) devices. For instance, precision fertilizers or irrigation systems can use these maps to deliver selective treatment only when necessary, thereby improving input efficiency and overall productivity.

Finally, the implementation of multispectral and hyperspectral imaging has moved the direction of modern agriculture away from global interventions and solutions toward the specific interventions being used in response to plant health data. By combining spectral, visual and geospatial information, these technologies enable farmers to attain both a clearer and more precise view of crop status and, accordingly, to implement more sustainable crop management techniques.

4. Discussion

This section discusses the advantages, disadvantages, and future developments of multi-sensor systems. It aims to summarize key findings from recent studies and highlight how these insights can contribute to the improvement of sensor fusion.

DOI: 10.54254/2755-2721/2025.LD30577

4.1. Advantages

(1) The integration of complementary vision sensors shows significant benefits for robotic perception. Through the integration of RGB cameras, depth sensors, LiDAR, radar systems, and multispectral cameras, robots can obtain more accurate and detailed data regarding the targeted environments. Each sensor modality provides distinct information such as geometric characteristics, velocity of moving objects, and thermal features. When integrating these various data sources, the robot can obtain an enhanced level of perception accuracy and robustness.

(2) Further, multimodal sensor fusion efficiently overcomes drawbacks of individual sensors. This allows robots to perform reliably in adverse conditions such as poor illumination, intense glare, or inclement weather such as heavy rain. Multi-modality integration not only ensures a metric-level understanding, but also facilitates improved semantics, necessary for high-level operations including navigation, environmental mapping and object finding.

(3) In addition, multimodal representation learning enhances operation efficiency and generalization ability of machine learning schemes. Heterogeneous sensing data train models that exhibit better adaptation in new environments and a lower predisposition towards overfitting. All of these benefits make a difference in building resilient intelligent robotics systems. By leveraging complementary vision technologies, autonomous robots are better equipped for reliable operation within complex and dynamic real-world environments [10].

4.2. Disadvantages

Despite the above benefits, multimodal vision systems also encounter several significant challenges.

(1) The integration of multiple sensor modalities will result in a much higher hardware and software complexity. This process requires highly accurate data calibration and synchronization across different devices, and demands more maintenance efforts. The added sensors also result in higher costs, overall weight, and power consumption of the system, which poses substantial constraints for mobile robots or drones with limited payload capacity.

(2) Another major challenge of multi-sensor systems is their ability to accurately align spatial and temporal data from various sensor modalities. Specifically, even a small calibration error in the data fusion process can ultimately cause reduced perception accuracy and control stability. Additionally, different types of sensing technologies provide different sensor outputs. These disparities in output data add another layer of complexity to the overall system, and often require specialized preprocessing mechanisms and special neural.

(3) Furthermore, despite the fact that multi-sensor fusion typically can enhance robustness under varied environmental conditions, the overall system performance can still be severely affected under adverse environments. When multiple sensing modalities degrade simultaneously, the performance of the robotic system may be significantly compromised [11].

4.3. Future development

(1) Recent research has increasingly emphasized the importance and advantages of utilizing event-based and neuromorphic sensors in robotic perception. Specifically, those advanced sensors have high dynamic range, low latency, and great energy efficiency, which make them ideal for fast-changing environments. As a result, there has been a noticeable trend to adopt the latest sensors and combine them with traditional cameras and depth sensors to construct a new sensing system that is suitable in complex and changeable environments. The integration of new sensor technology in

DOI: 10.54254/2755-2721/2025.LD30577

robotic systems can not only increase the robustness of input sensory data, but also allow them to have better performance in a greater variety of operational scenarios.

(2) When various vision sensors are installed on a robot platform, one of the biggest issues is that it not only adds complexity but also weight to the overall system. Hence, future multi-sensor fusion systems are expected to achieve higher levels of sensor integration as well as more efficient space utilization that reduces unnecessary physical size and weight. Future multi-sensor systems can also take advantage of lower power consumption and improved deployability for functioning at diverse service environments.

(3) Modern developments in artificial intelligence and deep learning have made end-to-end, data-driven fusion systems possible. No longer dependable upon the alignment between the different vision sensors, these advanced systems learn the shared representations of the signals directly from multimodal data This performance is further boosted once they are paired with modern hardware like AI-optimized processors. Simultaneously, research is focused on uniformity in safety procedures to enhance system trustworthiness as well as facilitate implementation in practice [12].

5. Conclusion

In conclusion, integrating multiple sensor technologies significantly strengthens the adaptability of autonomous systems in complex and rapidly changing conditions. By combining complementary data sources, multi-sensor fusion helps overcome the limitations of any single sensor and leads to more robust, reliable performance. Nevertheless, several challenges remain. High computational demands and the need for more effective fusion algorithms continue to pose obstacles and must be addressed in future research. Continued innovation in robotic vision and sensing technologies will be essential for advancing the next generation of autonomous systems.

References

[1] A. Singh, V. Kalaichelvi, and R. Karthikeyan, “A survey on vision guided robotic systems with intelligent control strategies for autonomous tasks, ” Cogent Engineering, vol. 9, no. 1, art. 2050020, Apr. 2022, doi: 10.1080/23311916.2022.2050020.

[2] A. Geiger et al., "A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms, " in Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2019, pp. 1673-1680.

[3] G. Gallego et al., "Event-based Vision: A Survey, " IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 154-180, 2022.

[4] M. Camplani, S. Hannuna, M. Mirmehdi, D. Damen, A. Paiement, L. Tao, and T. Burghardt, “Real-time RGB-D tracking with depth-scaling kernelised correlation filters and occlusion handling, ” in Proc. British Machine Vision Conference (BMVC), Sept. 2015, pp. 145.1-145.11, doi: 10.5244/C.29.145.

[5] L. R. Mubarak et al, “Semantic Map Construction with Camera and LiDAR for Real-Time Object Detection, ” in Proc. 2024 IEEE International Conference on Smart Mechatronics (ICSMech), 2024.

[6] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Point Pillars: Fast Encoders for Object Detection from Point Clouds, " in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 12697-12705.

[7] S. Hacohen, O. Medina, and S. Shoval, “Autonomous driving: A survey of technological gaps using Google Scholar and Web of Science trend analysis, ” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 11, pp. 21241–21254, Nov. 2022, doi: 10.1109/TITS.2022.3172442.

[8] P. Karmakar, B. Bhattacharya, and S. Padhy, “Crop monitoring by multimodal remote sensing: A review, ” Data in Brief, vol. 49, p. 110769, Jan. 2024. doi: 10.1016/j.dib.2023.110769.

[9] C. Sun, "A review of remote sensing for potato traits characterization in precision agriculture, " International Journal of Advanced Trends in Computer Science and Engineering, vol. 11, no. 3, pp. 28–39, Jun. 2022, doi: 10.30534/ijatcse/2022/021132022.

DOI: 10.54254/2755-2721/2025.LD30577

[10] M. T. Shahria, M. S. H. Sunny, M. I. I. Zarif, J. Ghommam, S. I. Ahamed, and M. H. Rahman, “A Comprehensive Review of Vision-Based Robotic Applications: Current State, Components, Approaches, Barriers, and Potential Solutions, ”Robotics, vol. 11, no. 6, art. no. 139, Dec. 2022. doi: 10.3390/robotics11060139.
[11] Y. Wang, A. H. Abd Rahman, F. ’A. Nor Rashid, and M. K. M. Razali, “Tackling Heterogeneous Light Detection and Ranging-Camera Alignment Challenges in Dynamic Environments: A Review for Object Detection, ” Sensors, vol. 24, no. 23, art. no. 7855, Nov. 2024. doi: 10.3390/s24237855.
[12] H. Wang, J. Li, and H. Dong, “A Review of Vision-Based Multi-Task Perception Research Methods for Autonomous Vehicles, ” Sensors, vol. 25, no. 8, art. no. 2611, Apr. 2025. doi: 10.3390/s25082611.