UNIVERSITAS DEHASEN
UNIVED BENGKULU
Achievement Through Education

e-ISSN 2828-0490

JURNAL
MEDIA COMPUTER SCIENCE

# From Sensors to Vision: Evolving Data Modalities for Intelligent IoT Systems

**Dedy Irvandy 1); Hedi Agfiria Carlusiaputri 2); Anja Wulan Sari 3); Putu Sudira 4); Pipit Utami 5)**
*1,2,3,4,5) Universitas Negeri Yogyakarta*
**Email: 1)** dedyirvandy.2024@student.uny.ac.id

***How to Cite :***
Irvandy, D., Carlusiaputri, H, A., Sari, A, W., Sudira, P., Utami, P. (2025). From Sensors to Vision: Evolving Data Modalities for Intelligent IoT Systems. Jurnal Media Computer Science, 4(2)

**ARTICLE HISTORY**
*Received [25 Juni 2025]*
*Revised [3o Juli 2025]*
Accepted [31 Juli 2025]

***KEYWORDS***
*Artificial Intelligence of Things; Computer Vision; Visual Sensing; Multimodal Sensing; Edge Computing.*

## *ABSTRACT*

*The rapid growth of the Internet of Things has increased the need for sensing systems capable of understanding complex environments beyond simple data collection. The integration of artificial intelligence and computer vision has driven a shift toward perception-oriented Artificial Intelligence of Things (AIoT) systems. This study systematically synthesizes research on vision-based sensing in AIoT, focusing on application domains, sensing modalities, AI methods, and computing architectures. A systematic literature review following PRISMA 2020 guidelines using Scopus identified 24 empirical studies published between 2018 and 2025. The findings show that AIoT vision systems are increasingly applied in manufacturing, agriculture, infrastructure, and surveillance, supporting real-time monitoring and decision-making. Core functions include detection, classification, segmentation, and activity recognition, enabled by deep learning and edge–cloud architectures. The results indicate a shift toward multimodal sensing and edge intelligence, highlighting a broader transition to perception-centric AIoT systems, with ongoing challenges in dataset generalization, multimodal integration, and efficient edge deployment.*

## INTRODUCTION

The Internet of Things (IoT) has grown rapidly over the past decade, transforming how physical environments are monitored and managed through interconnected digital systems. IoT is commonly defined as a network of interconnected objects capable of collecting, exchanging, and processing data across distributed environments (Perera et al., 2015; Ray, 2017; Syed et al., 2021; Yahya et al., 2018). Across industrial facilities, urban infrastructures, agricultural fields, and smart environments, large numbers of connected devices continuously collect data about surrounding conditions. Most early IoT deployments relied heavily on numerical sensing technologies, where environmental sensors such as temperature, humidity, vibration, and motion form the core sensing layer of many IoT monitoring systems (Adu-Manu et al., 2022; Syed et al., 2021). Sensors measuring temperature, vibration, humidity, pressure, or motion were widely used to observe physical processes and operational states. Although these measurements provide reliable quantitative information, they often capture only isolated aspects of complex environments. As intelligent and autonomous systems become more common, IoT platforms increasingly require richer environmental awareness that goes beyond simple numerical measurements.

In response to these demands, the integration of artificial intelligence with IoT infrastructures has accelerated in recent years. This technological convergence has led to what is commonly referred to as the Artificial Intelligence of Things (AIoT), a paradigm that integrates artificial intelligence techniques with IoT infrastructures to enable intelligent sensing, analytics, and automated decision-making across distributed systems (Ghoreishi et al., 2022; Haqiq et al., 2022; Miran & Ayub, 2023; Seng et al., 2022). In AIoT systems, sensing devices do more than collect data, as they increasingly support intelligent sensing, analytics, and automated decision-making across distributed environments (Miran & Ayub, 2023; Seng et al., 2022). These capabilities are further reflected in integrated AIoT architectures (Utami, Mashoedah, et al., 2024). The captured data are increasingly processed using intelligent algorithms that can identify patterns, recognize events, and support automated decision processes. One particularly visible development in this transition is the growing use of camera-based sensing and computer vision techniques in IoT environments. Cameras allow machines to observe scenes directly, enabling tasks such as object detection, activity recognition, and environmental monitoring. Several studies demonstrate how machine vision has been embedded into IoT infrastructures to support automated inspection and intelligent monitoring across various environments, including industrial inspection, surveillance, and smart monitoring systems (Hao, 2023; Silva et al., 2025; Wang et al., 2022; Yu, 2022).

The increasing adoption of visual sensing technologies also reflects the practical limitations of traditional sensor-centric monitoring. Numerical sensors typically measure one environmental variable at a time. As a result, they often provide fragmented information about complex real-world situations. Visual data, by contrast, contain spatial structure, object relationships, and contextual information that are difficult to capture using conventional sensors. Computer vision-based sensing therefore provides richer situational awareness than scalar sensors in many monitoring contexts, including smart buildings, infrastructure monitoring, and environmental observation (Cardinale, 2020; Kang et al., 2025; Mardanshahi et al., 2025; Yang et al., 2023). In manufacturing environments, for example, machine vision systems can detect subtle surface defects or irregularities that physical sensors may fail to identify (Jia & Wang, 2022; Li et al., 2022). Similar approaches have been applied in agricultural monitoring, where image-based systems analyze crop conditions and detect plant diseases using classification and detection algorithms (Chen et al., 2021; Jaramillo-Hernández et al., 2024). These examples illustrate how visual perception can significantly expand the monitoring capabilities of IoT systems.

From a broader perspective, these developments suggest a gradual evolution of sensing paradigms in IoT environments from traditional sensor-centric monitoring toward intelligent perception systems. Recent AIoT research highlights this transition as the convergence of sensing, artificial intelligence, and edge computing that enables more advanced environmental perception and analytics (Djenouri et al., 2021; Kwon & Seo, 2022; Miran & Ayub, 2023). Early IoT deployments relied mainly on sensor-based monitoring using wireless sensor networks to collect numerical environmental data. As imaging devices became more accessible and computing capabilities improved, vision-enabled sensing started to emerge. Cameras allowed IoT systems to capture spatial observations of the environment, while computer vision algorithms enabled automated interpretation of those observations. More recent research has taken this step further by integrating camera sensing with traditional sensors. These multimodal sensing architectures combine heterogeneous data sources to improve contextual awareness and monitoring reliability. For instance, smart agriculture systems may integrate environmental sensors with visual crop monitoring (Nguyen et al., 2024), while infrastructure inspection has been enhanced through deep learning-based visual analysis for crack detection and condition assessment (Matarneh et al., 2024). Taken together, these developments indicate a trend toward perception-centric AIoT systems where sensing, learning, and contextual interpretation operate together.

Despite the rapid growth of research in this area, the existing literature remains fragmented. Many studies focus on specific implementations of IoT sensing architectures or individual machine vision applications. For example, research has investigated vision-based inspection systems for manufacturing, camera monitoring systems in smart cities, and image-based surveillance technologies for safety monitoring (Lv & Chen, 2021; Silva et al., 2025; Wang et al., 2022). While these contributions provide valuable insights into particular application domains, they rarely examine the broader technological evolution of sensing modalities in IoT systems. Consequently, the literature lacks a comprehensive synthesis that explains how IoT sensing strategies are transitioning from traditional numerical sensing toward visual and multimodal perception architectures.

This fragmentation highlights an important gap in the current body of knowledge. Although visual sensing technologies are increasingly used in AIoT systems, there is still limited systematic synthesis explaining how sensing modalities have evolved within IoT environments. Existing research is also distributed across diverse application areas such as manufacturing, agriculture, urban monitoring, infrastructure management, and assistive technologies, making it difficult to clearly identify technological patterns, architectural trends, and emerging research directions in vision-enabled AIoT systems. To address this gap, the present study conducts a systematic literature review that synthesizes empirical research on visual sensing in AIoT environments, examining how vision technologies are integrated into IoT infrastructures with particular attention to sensing modalities, artificial intelligence methods, computing architectures, and application domains. By analyzing recent studies that implement vision-based IoT systems, the review clarifies how sensing paradigms are evolving within intelligent IoT ecosystems and proposes a conceptual perspective describing the progression from traditional sensor monitoring toward perception-centric AIoT architectures that combine visual perception, multimodal sensing, and distributed intelligence.

This study contributes to the literature in three main ways. First, it provides a systematic synthesis of vision-based sensing in AIoT systems across diverse application domains, consolidating fragmented research into a coherent analytical structure. Second, it develops a conceptual framework describing the evolution of sensing paradigms from sensor-based monitoring to perception-centric AIoT systems, highlighting the roles of visual sensing, multimodal integration, and distributed intelligence. Third, it identifies critical research gaps related to multimodal integration, scalability, dataset generalization, and edge deployment, and proposes a future research agenda to guide the development of adaptive and scalable AIoT systems.

To structure the analysis, the study adopts the PICo framework. Within this framework, the population refers to AIoT systems, the interest focuses on visual sensing technologies, and the context concerns intelligent monitoring and automation environments. Guided by this structure, the review investigates several key questions related to the development of vision-enabled AIoT systems. Specifically, the analysis examines in which application domains AIoT vision-based sensing has been implemented, what sensing modalities are integrated within these systems, which artificial intelligence methods are employed to perform visual sensing tasks, and how edge and cloud architectures are utilized to support AIoT vision processing. In addition, the review explores broader research trends that indicate the technological evolution from traditional sensor-based IoT monitoring toward vision-centric and perception-driven IoT systems.

## THEORETICAL BACKGROUND

### Evolution of Sensing in IoT and AIoT

The Internet of Things (IoT) describes networks of interconnected physical devices that collect, transmit, and process data from their surrounding environment. In most implementations, these devices operate within layered architectures composed of sensing, network, and application layers, where physical sensors and actuators capture environmental signals that are subsequently transmitted and processed across the system stack (Nagajayanthi, 2021; Nasiri et al., 2021). Typical deployments combine sensors, embedded controllers, communication modules, and computing platforms. Together these elements allow physical objects to interact with digital infrastructures. In practical IoT environments such as smart manufacturing systems, agricultural monitoring platforms, urban infrastructure, and smart city services, sensing devices act as the primary interface between the physical and digital worlds. They continuously capture operational conditions and environmental parameters (Chen et al., 2021; Lv & Chen, 2021; Utami, Zakarijah, et al., 2024). As IoT technologies evolved, the integration of artificial intelligence capabilities into sensing infrastructures led to the emergence of the Artificial Intelligence of Things (AIoT), where AI models process sensed data to identify patterns, recognize events, and support automated decision making across edge, fog, and cloud layers (Haqiq et al., 2022; Qiu et al., 2021; Seng et al., 2022; J. Zhang & Tao, 2021). Empirical studies demonstrate this integration in applications such as industrial defect detection, crop monitoring, and activity recognition in intelligent surveillance systems (Jia & Wang, 2022; Silva et al., 2025; Yu, 2022).

Within these intelligent environments, sensing technologies play a central role. They provide the observational capability that allows digital systems to perceive physical conditions. Early IoT deployments primarily relied on numerical sensors that measure environmental variables such as temperature, humidity, vibration, pressure, or motion. These measurements provide reliable quantitative information. However, they capture only limited contextual detail about complex environments. Consequently, researchers increasingly recognized that sensor-based monitoring alone is often insufficient for applications requiring situational awareness, spatial interpretation, or detailed visual inspection. This limitation has encouraged the integration of camera based sensing and computer vision technologies within IoT infrastructures.

The development of AIoT sensing systems can therefore be viewed as an evolutionary process in which sensing capabilities move from simple environmental measurements toward richer perception mechanisms. Early IoT systems relied on sensor-based monitoring supported by wireless sensor networks that produced structured numerical data describing physical conditions. As imaging devices became more affordable and computing capabilities improved, vision-enabled sensing emerged, allowing IoT platforms to capture spatial information through cameras and interpret it using computer vision algorithms. More recent systems combine visual data with other sensing modalities such as environmental sensors, wireless sensor networks, and motion sensing devices, enabling multimodal sensing architectures that improve contextual awareness and monitoring reliability. Examples include smart agriculture platforms that combine camera monitoring with environmental sensing (Nguyen et al., 2024) and infrastructure inspection systems that utilize camera-based analysis for crack detection (Matarneh et al., 2024).

Recent research further extends this trajectory toward perception centric AIoT architectures in which sensing, artificial intelligence, and distributed computing operate together to interpret complex environments. In these systems, deep learning models analyze visual and sensor data to extract semantic information and support automated decision making, while edge computing platforms perform real time analysis close to sensing devices and cloud infrastructures provide large scale coordination and analytics (Katsigiannis & Mykoniatis, 2024; Silva et al., 2025). This progression from sensor based monitoring to vision enabled sensing, multimodal sensing, and perception centric AIoT is summarized in Table 1. The table is presented as a conceptual synthesis derived from the reviewed studies rather than as a set of empirical findings. It organizes recurring patterns identified across the literature and frames them as an evolutionary model that links sensing modalities, AI capabilities, and system architectures in AIoT environments. The empirical distribution of these sensing approaches across the reviewed studies is analyzed in Section 4. Table 1 presents a conceptual framework derived from the synthesis of the 24 empirical studies analyzed in this review. Rather than listing individual studies, the table integrates recurring patterns identified across the dataset and organizes them into an evolutionary framework of AIoT sensing. This framework highlights how sensing paradigms gradually progress from numerical monitoring toward perception centric AIoT systems that combine multimodal sensing, artificial intelligence, and distributed edge intelligence.

**Table 1. Evolution of sensing paradigms in AIoT systems**

| Stage | Sensing paradigm | Data and AI capability | Enabling technologies and representative studies |
|---|---|---|---|
| Stage 1 | Sensor based monitoring | Numerical sensor measurements with simple rule based monitoring | Wireless sensor networks and embedded sensing nodes (Lv & Chen, 2021) |

| Stage | Sensing paradigm | Data and AI capability | Enabling technologies and representative studies |
|---|---|---|---|
| Stage 2 | Vision enabled sensing | Visual image analysis using computer vision and early deep learning detection models | Machine vision inspection systems in industrial environments (Jia & Wang, 2022; J. Zhang et al., 2022) |
| Stage 3 | Multimodal sensing | Fusion of sensor data and visual perception using machine learning models | Hybrid sensing systems combining cameras and environmental sensors (Nguyen et al., 2024) |
| Stage 4 | Perception centric AIoT | Deep learning based perception and contextual interpretation of multimodal data | Edge AI vision systems and distributed analytics platforms (Katsigiannis & Mykoniatis, 2024; Silva et al., 2025) |

### Visual Sensing and Machine Vision

Machine vision refers to the use of imaging devices together with computational algorithms to automatically analyze visual information from the physical world. In AIoT environments, machine vision functions as a perception mechanism that enables intelligent systems to observe and interpret visual scenes. Unlike traditional sensors that measure individual variables such as temperature or humidity, cameras capture spatial structures, object relationships, and dynamic activities occurring in the environment. Vision enabled IoT systems therefore operate as perception pipelines that convert raw visual observations into meaningful information. In many computer vision frameworks, this pipeline includes stages such as image acquisition, feature extraction, object detection, and visual analytics, which together allow systems to interpret scenes and extract semantic information from image or video streams (Pasolini et al., 2020).

The pipeline typically begins with visual data acquisition through imaging devices such as RGB cameras, industrial inspection cameras, surveillance cameras, depth sensors, or thermal cameras. These visual streams are then processed using computer vision algorithms that perform tasks including image classification, object recognition, segmentation, and feature extraction, with deep learning models widely used to interpret visual data, including facial expression recognition using convolutional neural network architectures (Utami et al., 2022b). Through these processes, visual sensing systems can identify patterns and objects that would be difficult to capture through numerical sensors alone. Several studies included in this review demonstrate the deployment of such vision pipelines in industrial inspection, agricultural monitoring, and infrastructure observation systems (Jaramillo-Hernández et al., 2024; Katsigiannis & Mykoniatis, 2024). Deep learning approaches are particularly common in manufacturing environments, where convolutional models are used for defect detection and segmentation of industrial products (Jia & Wang, 2022).

Visual data streams are substantially larger than conventional sensor signals. For this reason many IoT platforms perform image processing directly on edge devices. Edge computing enables lightweight deep learning models to execute inference near the sensing source, reducing latency and limiting the need to transmit large video streams to centralized servers. Several reviewed systems implement visual analytics on embedded platforms or edge computing nodes to support real time monitoring (Jaramillo-Hernández et al., 2024; Katsigiannis & Mykoniatis, 2024). The resulting visual insights can then trigger alerts, anomaly detection mechanisms, or automated responses. Examples include surveillance systems that recognize human activities, agricultural monitoring systems that identify crop diseases, and industrial inspection platforms that detect manufacturing defects. Compared with traditional sensors, vision sensing provides richer contextual awareness because cameras capture spatial relationships and interactions within complex environments. This capability significantly expands the perception capacity of modern AIoT infrastructures.

### Multimodal Sensing in Intelligent IoT

Although visual sensing significantly enhances environmental perception, many AIoT systems integrate multiple sensing modalities to achieve more reliable monitoring. Multimodal sensing refers to the integration of heterogeneous data sources including cameras, environmental sensors, wireless sensor networks, and motion detection systems. By combining complementary sensing modalities, IoT platforms can interpret both numerical measurements and spatial information simultaneously. This integration allows intelligent monitoring systems to capture richer environmental context compared with single sensing technologies.

Several reviewed studies demonstrate how multimodal sensing architectures improve system performance across different application domains. In smart agriculture environments, camera monitoring is frequently combined with environmental sensors to simultaneously track plant growth conditions and surrounding environmental parameters such as temperature and humidity (Nguyen et al., 2024). Infrastructure monitoring has also been enhanced through camera-based crack detection using deep learning techniques to improve damage assessment accuracy (Matarneh et al., 2024). In intelligent transportation environments, hybrid parking monitoring systems combine wireless sensor networks with computer vision object detection to verify vehicle presence and improve occupancy detection reliability (Hudda et al., 2024).

Multimodal sensing is also applied in industrial monitoring scenarios where machine vision systems are combined with thermal cameras or process monitoring sensors to detect anomalies in conveyor systems or industrial equipment (Kia & Leiding, 2025). Integrating multiple sensing modalities provides several advantages. It improves system robustness because complementary sensors can validate observations from different perspectives. It also enhances contextual awareness by combining spatial visual information with environmental measurements. As a result, multimodal architectures enable more reliable monitoring in complex and dynamic environments where single sensing technologies may be insufficient.

### Edge Intelligence and Perception Centric AIoT

The rapid expansion of visual sensing in IoT environments also introduces significant challenges related to data processing and communication. Image and video streams contain far larger volumes of data than conventional sensor signals. Transmitting these data continuously to centralized cloud servers can therefore create latency, bandwidth constraints, and computational bottlenecks. As AIoT applications increasingly rely on visual analytics, efficient distributed processing becomes essential for maintaining reliable monitoring and real time system responsiveness.

Edge computing has emerged as a key solution to address these limitations by performing data processing closer to sensing devices. In edge architectures, AI inference and visual data processing occur near the data source rather than exclusively in remote cloud infrastructures. This approach reduces communication overhead and enables faster response times for vision based monitoring systems (Ke et al., 2021; Ravindran, 2023; Rodríguez et al., 2023). Several studies included in this review demonstrate the practical deployment of edge based vision systems. For example, embedded edge platforms have been used for automated gauge reading and industrial inspection tasks using camera sensors (Katsigiannis & Mykoniatis, 2024). Similarly, lightweight edge AI devices have been deployed in agricultural environments to detect crops and analyze visual data directly in the field (Jaramillo-Hernández et al., 2024).

Many AIoT systems therefore adopt hybrid architectures that combine edge processing with cloud based analytics. In these architectures, edge devices handle real time inference while cloud platforms support large scale data storage, model training, and system coordination. Such edge cloud collaboration appears in multiple application domains including industrial inspection, agricultural monitoring, and urban surveillance systems (Nguyen et al., 2024; Silva et al., 2025; J. Zhang et al., 2022). This integration of visual perception, multimodal sensing, and distributed intelligence gradually leads toward perception centric AIoT systems in which sensing platforms actively interpret environmental context and support automated responses. Because existing studies remain fragmented across domains and technological approaches, a systematic synthesis is necessary to clarify how sensing modalities, AI techniques, and computing architectures evolve together in AIoT environments. The following section therefore explains the systematic review methodology used to identify, evaluate, and synthesize the empirical studies included in this review.

## RESEARCH METHODOLOGY

This review followed the PRISMA 2020 guidelines to ensure transparency and methodological rigor in the synthesis process. PRISMA provides a structured reporting framework that clarifies how studies are identified, screened, and synthesized in systematic reviews, improving transparency and reproducibility across research domains (Kahale et al., 2022; Page et al., 2021). Guided by this framework, the review process included systematic searching, study screening and selection, data extraction, quality assessment, and thematic synthesis. Each step was designed to maintain a transparent record of decisions made during the review. This structured procedure reduces bias during study identification and supports a reliable interpretation of the available evidence.

**Table 2. Operationalization of PICo into Eligibility Criteria and Review Design**

| PICo Element | Description | Role in Review Design |
|---|---|---|
| Population | AIoT systems integrating sensing and vision technologies | Defines conceptual scope and supports RQ1, RQ4 |
| Interest | Vision-based sensing or machine vision in implemented systems | Structures eligibility criteria and supports RQ2, RQ3 |
| Context | Intelligent monitoring, automation, and real-world AIoT environments | Guides screening and supports RQ1, RQ5 |

The PICo framework was applied to define the conceptual scope of the review, structure eligibility criteria, and guide the screening process. In this framework, the population referred to AIoT systems that integrate sensing technologies in real world environments. The element of interest focused on implementations of machine vision or visual sensing embedded within these systems, while the context concerned intelligent monitoring, automation, and other smart environments where sensing technologies support analytical or decision making functions. This structure ensured that the search and screening stages remained aligned with the intended research focus. Table 2 summarizes how these PICo elements were translated into practical eligibility criteria. By explicitly defining the population, technological focus, and contextual application environments, the framework helped ensure that the review consistently targeted implemented AIoT visual sensing systems rather than peripheral IoT or computer vision research. The search was performed using the Scopus database by applying the query to the article title, abstract, and keywords fields (TITLE-ABS-KEY) to ensure comprehensive retrieval of relevant studies. The initial search yielded a total of 171 records, which were subsequently subjected to the screening and selection process following the PRISMA workflow.

### Eligibility Criteria and Study Selection

To ensure that the review retained only studies directly relevant to AIoT visual sensing systems, clear inclusion and exclusion criteria were established before the screening process began. Eligible studies were required to describe AIoT environments that integrate vision based sensing technologies and provide empirical evidence through implemented systems or experimental validation. The review focused on relatively recent research to capture current technological developments, therefore the publication window was limited to studies published between 2018 and 2025. Only English language publications were considered in order to ensure consistent interpretation of methodological descriptions, and the dataset was restricted to peer reviewed journal articles and conference papers to maintain academic reliability. These criteria ensured that the review emphasized practical technological implementations rather than conceptual discussions or peripheral topics.

Studies were excluded when their primary focus did not involve operational visual sensing within AIoT environments. Articles centered solely on communication protocols or networking optimization without sensing implementations were removed, as were studies dealing with image security or privacy that did not include deployed sensing systems. Conceptual discussions, purely theoretical contributions, and review articles without implemented systems were also excluded because they did not provide empirical evidence suitable for synthesis. Applying these criteria helped refine the dataset so that the remaining studies consistently represented real AIoT visual sensing applications. The resulting corpus therefore reflects research that demonstrates implemented sensing architectures rather than speculative or purely theoretical approaches.

Following the definition of these criteria, the study selection process followed the PRISMA workflow that organizes the review into identification, screening, eligibility, and inclusion stages. Records retrieved from the database search were first compiled and duplicates were removed to establish the initial pool of candidate studies. Titles and abstracts were then examined to evaluate conceptual relevance and determine whether the studies addressed AIoT vision sensing systems. Articles that passed this stage were assessed through full text evaluation to confirm methodological relevance and compliance with the predefined criteria. When uncertainties emerged during screening, inclusion decisions were resolved through discussion to maintain consistency in the final selection of studies. Figure 1 presents the PRISMA flow diagram that summarizes how the literature was progressively refined during the review process.

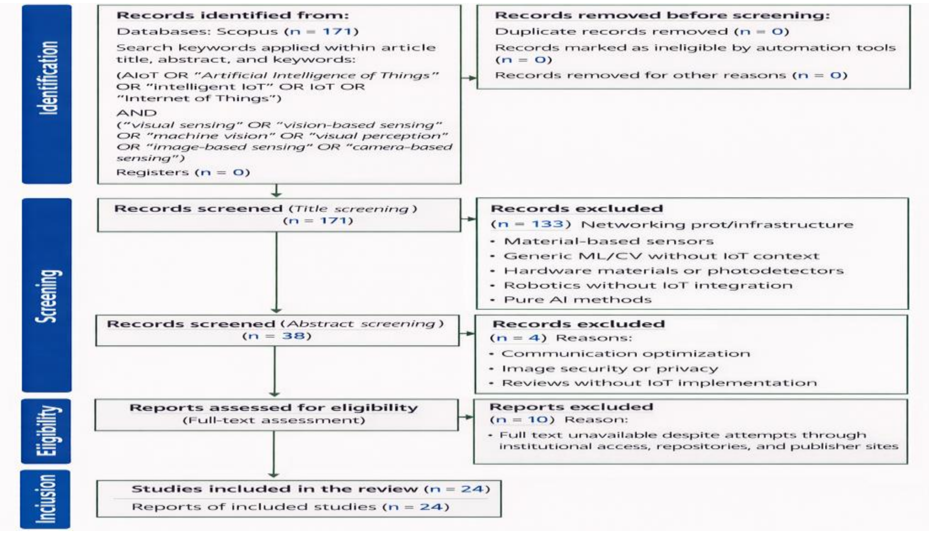

**Figure 1. PRISMA flow diagram**

Figure 1. PRISMA flow diagram illustrating the identification, screening, eligibility, and inclusion stages used to determine the final set of studies analyzed in this review. The diagram illustrates how the literature search was progressively refined as the review moved from identification to final inclusion. Through successive screening stages, the initial pool of records was filtered according to the predefined criteria until a focused set of studies relevant to AIoT visual sensing systems remained. Each stage reflects a deliberate effort to ensure conceptual relevance, methodological clarity, and alignment with the defined research scope. The diagram also highlights how multiple screening layers help reduce noise from unrelated studies while preserving research that directly contributes to understanding visual sensing in AIoT environments. This progression clarifies how the final corpus of studies was established before the synthesis stage. Through this sequential filtering process, the initial search results were gradually reduced until a final set of twenty-four empirical studies remained for analysis, which were then systematically extracted and prepared for

subsequent analysis. All retrieved records were managed and organized using Mendeley reference management software to facilitate duplicate removal, screening, and documentation of the selection process.

## RESULTS AND DISCUSSION

### Results

Beyond empirical synthesis, this review advances a conceptual understanding of sensing evolution in AIoT by framing it as a progressive transformation from numerical sensing toward perception-centric intelligence. Unlike prior studies that focus on specific applications or isolated technologies, this review integrates sensing modalities, AI methods, and computing architectures into a unified conceptual model. This model extends existing IoT and AIoT frameworks by explicitly positioning visual perception and multimodal integration as central components of intelligent sensing systems. As such, the proposed framework contributes to theory by clarifying how perception capabilities emerge through the interaction of sensing, learning, and distributed computation.

The findings can be interpreted in relation to the research questions guiding this review. First, regarding application domains, AIoT vision systems are most prominently deployed in manufacturing and agriculture, indicating that environments requiring continuous monitoring and precision analysis are primary adopters of visual sensing. Second, in terms of sensing modalities, there is a clear transition from single-modality sensing toward multimodal configurations that integrate visual and non-visual data. Third, with respect to AI methods, deep learning models dominate due to their ability to extract semantic information from complex visual data, although hybrid approaches are increasingly adopted to address deployment constraints. Fourth, in relation to system architectures, edge–cloud integration emerges as a prevalent approach that balances real-time processing and scalability. Finally, these patterns collectively suggest a broader shift toward perception-centric AIoT systems, where sensing is no longer passive but actively interprets environmental context.

### Study Characteristics and Quality Assessment

This review analyzed a final dataset of 24 studies that met the predefined inclusion criteria. These studies represent implemented vision-enabled sensing systems within Artificial Intelligence of Things environments, covering diverse application domains and system designs. As summarized in Table 3, the dataset reflects variations in sensing modalities, vision tasks, and computing architectures. The studies collectively indicate that AIoT vision systems are increasingly applied in practical environments rather than isolated prototypes (Lv & Chen, 2021; Nguyen et al., 2024; J. Zhang et al., 2022). This diversity provides a strong empirical basis for identifying technological patterns across domains.

**Table 3. Summary of Included Studies**

| Study | Domain & Modality | Vision Task & AI Method | Architecture & Key Contribution |
|---|---|---|---|
| (Lv & Chen, 2021) | Smart city, camera | Image processing, CNN | IoT architecture; robust perception |
| (J. Zhang et al., 2022) | Manufacturing, vision | Defect detection, CNN | Edge–cloud; automated inspection |
| (Saif et al., 2022) | Manufacturing, vision | Measurement, image processing | Edge–cloud; CNC inspection |
| (Katsigiannis & Mykoniatis, 2024) | Manufacturing, camera | Gauge reading, OpenCV | Edge; real-time monitoring |
| (Hudda et al., 2024) | Transportation, | Vehicle detection, | Edge; smart parking |

| Study | Domain & Modality | Vision Task & AI Method | Architecture & Key Contribution |
|---|---|---|---|
| | multimodal | Faster R-CNN | |
| (Jia & Wang, 2022) | Manufacturing, vision | Segmentation, U-Net | IIoT; defect analysis |
| (Nguyen et al., 2024) | Agriculture, multimodal | Crop detection, lightweight AI | Edge–cloud; smart farming |
| (Vukićević et al., 2021) | Warehouse, camera | QR tracking, OpenCV | Edge; inventory tracking |
| (Kia & Leiding, 2025) | Recycling, multimodal | Detection, YOLO | Edge–cloud; conveyor monitoring |
| (Wang et al., 2022) | Manufacturing, vision | NDT detection, GA+HMM | Cloud IIoT; monitoring |
| (Yu, 2022) | Human monitoring, vision | Activity recognition, 3D-CNN | Cloud IoT; motion analysis |
| (Hao, 2023) | Mechatronics, camera | Classification, SVM | Cloud IoT; inspection |
| (Chenguang, 2023) | Smart home, camera | Face recognition, YOLOv5 | Edge; energy control |
| (Sun & Lin, 2022) | Remote sensing, multimodal | Detection, feature fusion | iVIOT; target recognition |
| (Chebrolu et al., 2025) | Manufacturing, vision | Roughness estimation, ANN | IoT; quality inspection |
| (Ross et al., 2020) | Agriculture, multimodal | Detection, CV methods | Edge–cloud; pest monitoring |
| (Jaramillo-Hernández et al., 2024) | Agriculture, vision | Detection, deep learning | Edge; precision farming |
| (Matarneh et al., 2024) | Infrastructure, vision | Crack detection, CNN | Deep learning; pavement monitoring |
| (Silva et al., 2025) | Surveillance, vision | Detection, YOLO+CNN | Edge–cloud; safety |
| (Chen et al., 2021) | Agriculture, camera | Classification, EfficientNet | IoT; disease detection |
| (S. Zhang et al., 2025) | Infrastructure, embedded vision | Motion tracking | Edge; cable monitoring |
| (Rajendran et al., 2022) | Assistive, multimodal | Object recognition | Edge; navigation aid |
| (Li et al., 2022) | Manufacturing, vision | Segmentation, GAN | Edge–cloud; textile inspection |
| (J. Zhang et al., 2022) | Sports, multimodal | Action recognition | AIoT; training analytics |

Table 3 condenses the key characteristics of all 24 included studies by integrating sensing modalities, vision tasks, AI methods, and system architectures into a unified representation. The descriptive patterns indicate a recent and rapidly developing research area, with most studies published between 2020 and 2025 and a noticeable increase in the later years. This trajectory suggests growing attention to combining computer vision with IoT infrastructures for practical deployment. Across domains, applications concentrate in smart manufacturing, where vision supports defect detection and quality inspection, and in smart agriculture, where camera sensing enables crop monitoring and disease detection. Infrastructure monitoring also appears prominently, with systems designed for structural inspection and safety analysis. Additional use cases, including surveillance, assistive technologies, and sports analytics, further demonstrate how visual perception is expanding across intelligent environments.

The technological patterns reflected in Table 3 reveal a consistent shift toward data-driven and distributed intelligence. Deep learning models are widely used to interpret visual data, indicating their central role in enabling perception capabilities. Many systems adopt edge computing to support real-time processing close to sensing devices, reducing latency and communication overhead. In parallel, several studies integrate multimodal sensing by combining camera-based perception with environmental or networked sensors, which enhances contextual awareness. These developments suggest that AIoT systems are moving beyond isolated sensing toward integrated perception pipelines. Taken together, the patterns highlight a transition toward more adaptive, scalable, and context-aware vision-enabled IoT systems.

**Table 4. Aggregated Quality Assessment Across Included Studies**

| Quality Dimension | High / Moderate / Limited | Key Observation |
|---|---|---|
| Methodological Clarity | High (15), Moderate (9), Limited (0) | Most studies clearly report system design, sensing pipeline, and evaluation workflow |
| Validation | High (18), Moderate (6), Limited (0) | Experimental validation is commonly conducted using real-world or simulated deployments |
| Reproducibility | High (8), Moderate (10), Limited (6) | Reproducibility varies due to dataset accessibility and system-specific implementations |

Table 4 summarizes the quality assessment across the 24 included studies by aggregating methodological dimensions into a concise and traceable format. The evaluation is grounded in the reported clarity of system design, the presence of experimental validation, and the availability of implementation details in each study. Overall, the studies demonstrate strong methodological clarity and consistent validation, indicating a high level of technical maturity in AIoT vision systems. However, reproducibility appears more constrained in several cases, as some implementations rely on proprietary datasets or context-specific configurations, which may limit broader applicability across different environments.

### Thematic Findings

Table 5 presents a synthesis of recurring patterns identified across the 24 included studies, linking application domains, vision tasks, AI methods, computing architectures, and multimodal sensing into a coherent analytical structure. Guided by the research questions and the sensing evolution framework, this synthesis highlights five interrelated themes that consistently appear across the reviewed studies. These themes provide a structured basis for interpreting how vision-enabled AIoT systems are designed and deployed in practice. The patterns observed are not isolated cases but reflect repeatable configurations across different contexts, indicating a stable empirical foundation for the analysis.

**Table 5. Thematic Synthesis of AIoT Vision Systems**

| Theme | Synthesized Pattern (with embedded evidence) | Key Insight |
|---|---|---|
| Application Domains | AIoT vision systems are deployed across manufacturing, agriculture, infrastructure, surveillance, and assistive environments, with strong presence in industrial inspection and smart farming contexts | Vision sensing is broadly applicable across diverse intelligent environments |

| Theme | Synthesized Pattern (with embedded evidence) | Key Insight |
|---|---|---|
| Vision Tasks | Systems consistently perform detection, classification, segmentation, recognition, and activity analysis, reflecting layered perception capabilities from feature extraction to semantic interpretation | Vision tasks constitute the functional backbone of AIoT perception pipelines |
| AI Methods | Deep learning approaches such as CNN, YOLO, U-Net, and GAN dominate, often combined with classical image processing for efficiency and robustness | Hybrid AI approaches balance accuracy and deployability in real-world systems |
| Architectures | Distributed architectures including edge-only, edge–cloud, and device–edge–cloud pipelines are widely adopted to support real-time processing and scalability | Edge intelligence is a key enabler of real-time AIoT vision systems |
| Multimodal Sensing | Integration of vision with environmental sensors, WSN, or motion sensing improves contextual awareness and system robustness | Multimodal sensing enhances reliability and supports richer environmental understanding |

### Application Domains of AIoT Vision Systems

As summarized in Table 5, AIoT vision systems are deployed across a wide range of domains, with a strong concentration in industrial and environmental contexts where visual sensing directly supports monitoring and decision-making (Nguyen et al., 2024; Silva et al., 2025; J. Zhang et al., 2022). Smart manufacturing dominates, particularly for defect detection, quality inspection, and automated control processes. Smart agriculture also appears consistently, with systems leveraging image-based analysis for crop monitoring and disease identification. Infrastructure monitoring and surveillance extend these capabilities into safety and urban contexts, enabling structural assessment and activity tracking. Additional applications, including assistive systems and sports analytics, further demonstrate the adaptability of vision-enabled AIoT across heterogeneous environments.

Across these domains, a common pattern emerges in how visual sensing enhances situational awareness and operational efficiency. Rather than acting as standalone solutions, vision systems are typically embedded within broader IoT ecosystems that support continuous monitoring and automated responses. The diversity of application contexts indicates that vision sensing is no longer confined to specialized domains but is becoming a general-purpose capability within intelligent systems. At the same time, domain-specific requirements still shape system design, particularly in terms of data characteristics and performance constraints. This balance between general applicability and contextual adaptation reflects the evolving maturity of AIoT systems. Overall, application diversity reinforces the role of visual perception as a foundational component of intelligent environments.

### Vision Tasks in AIoT Systems

As reflected in Table 5, vision tasks in AIoT systems form a layered pipeline that progresses from basic detection to higher-level semantic understanding (Jia & Wang, 2022; Li et al., 2022). Object detection is the most prevalent function, supporting industrial inspection, traffic monitoring, and agricultural analysis. Image classification complements detection by assigning semantic labels to visual patterns, often used in disease identification and environmental assessment. Recognition tasks extend this capability by identifying specific entities, which is critical in surveillance and inventory tracking scenarios. In parallel, segmentation techniques enable precise localization of defects or structural features, particularly in manufacturing contexts. Together, these tasks illustrate how visual perception evolves from pixel-level analysis to structured scene interpretation. Emerging work on facial expression recognition further extends this capability by enabling systems to infer human affective states, although real-world deployment remains constrained by dataset variability, domain shifts, and environmental conditions (Abdullahu et al., 2025).

Beyond core tasks, several studies incorporate activity recognition and motion analysis to capture temporal dynamics in video streams (Sun & Lin, 2022; Yu, 2022). These capabilities are essential in applications such as behavior monitoring, sports analytics, and safety surveillance. Visual measurement and feature extraction further support quantitative analysis, including surface estimation and displacement tracking. Notably, systems rarely rely on a single task; instead, multiple functions are combined to enhance interpretability and decision support. This integration reflects a shift toward holistic perception pipelines rather than isolated image processing modules. Overall, the task composition demonstrates how AIoT systems leverage complementary vision functions to enable reliable and context-aware monitoring.

### Artificial Intelligence Methods Supporting Vision Sensing

The synthesis in Table 5 shows that AI methods are selected according to task demands and deployment constraints, with deep learning dominating most implementations (Chen et al., 2021; Kia & Leiding, 2025). Convolutional neural networks are widely adopted for classification and feature extraction due to their strong representational capacity. YOLO-based models are frequently used for real-time detection, especially in surveillance and industrial scenarios where latency is critical. U-Net architectures support fine-grained segmentation tasks that require precise localization, while GANs are applied for augmentation and defect-focused learning. These approaches collectively enable robust interpretation of complex visual data.At the same time, classical computer vision techniques remain relevant, particularly for preprocessing, feature engineering, and lightweight deployments. In several studies, traditional methods are combined with learning-based models to improve efficiency and stability. This hybrid strategy allows systems to balance accuracy with computational constraints, especially on edge devices. Model selection is therefore not purely performance-driven but also influenced by hardware limitations and real-time requirements. Such pragmatic combinations indicate a maturation of AIoT design practices. Overall, the methodological landscape reflects a convergence of deep learning and classical techniques to support deployable vision systems. In this context, lightweight and hybrid models are increasingly explored for facial expression recognition on edge devices, where computational constraints and privacy considerations require efficient and adaptive architectures (Elordi et al., 2021).

### Edge–Cloud Computing Architectures

As indicated in Table 5, AIoT vision systems increasingly rely on distributed architectures to manage computational load and latency (Nguyen et al., 2024). Edge-only configurations perform inference near the data source, enabling rapid response and reducing bandwidth usage. This approach is particularly effective for time-sensitive applications such as industrial inspection and traffic monitoring. However, edge resources are limited, which constrains model complexity and long-term analytics. To address this, many systems adopt hybrid designs that distribute processing between edge devices and cloud platforms.

Edge–cloud architectures allow preliminary analysis at the edge while delegating intensive computation and storage to the cloud. More advanced deployments implement device–edge–cloud pipelines, where data flows across multiple layers for staged processing. This structure supports scalability and facilitates integration with broader IoT ecosystems. The choice of architecture is often driven by application requirements, including latency, data volume, and system complexity. Across studies, a clear trend emerges toward flexible and adaptive computing models. These configurations enable real-time analytics while maintaining system scalability and robustness. However, deploying deep learning models at the edge remains challenging due to constraints in computation, memory, and energy, requiring optimization strategies such as pruning, quantization, and hardware-aware design.

### Multimodal Sensing Integration

Consistent with Table 5, multimodal sensing is widely adopted to overcome the limitations of single-modality perception and improve robustness (Hudda et al., 2024; Nguyen et al., 2024; Ross et al., 2020; Sun & Lin, 2022). Many systems combine visual data with environmental measurements such as temperature or humidity, enabling richer contextual interpretation. Integration with wireless sensor networks further supports distributed monitoring and cross-validation of observations. These configurations allow systems to capture complementary aspects of the environment, which enhances reliability under dynamic conditions.

In addition, some studies integrate vision with motion or structural sensors to jointly analyze visual patterns and physical signals. This approach is particularly relevant in infrastructure monitoring and assistive technologies, where multiple data sources improve situational awareness. Multimodal fusion is typically implemented through data-level or feature-level integration, depending on system design. While effective, these approaches often remain loosely coupled rather than fully unified. This indicates an opportunity for more cohesive perception frameworks that jointly model heterogeneous data. Overall, multimodal sensing emerges as a critical component of robust and context-aware AIoT systems. Recent studies also indicate that integrating visual data with contextual or physiological signals can improve facial expression recognition performance, particularly in dynamic and real-world AIoT environments (Yoon & Kim, 2025).

Taken together, the thematic patterns discussed above indicate a broader shift in how sensing is conceptualized within AIoT systems. Rather than isolated developments, these themes collectively reflect an evolution toward more integrated and perception-driven architectures. This progression can be more clearly understood through a conceptual representation of sensing modalities, as illustrated in Figure 2.

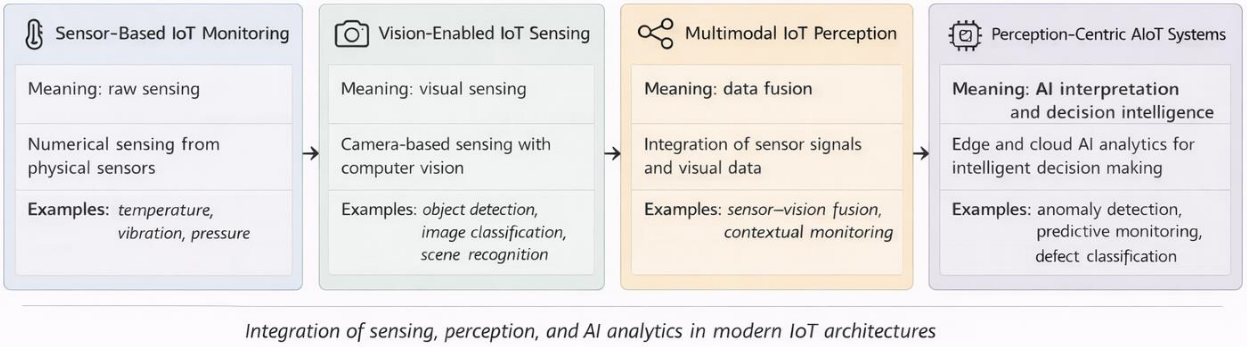

**Figure 2. Conceptual Evolution of Sensing Modalities**

Figure 2 illustrates a staged transition from sensor-based monitoring toward perception-centric AIoT systems, emphasizing how sensing capabilities evolve from numerical data acquisition to integrated perception and decision intelligence. The early stage relies on structured sensor data for basic monitoring tasks, while the introduction of vision-enabled sensing expands system capability to interpret spatial and contextual information. As systems integrate multiple data sources, multimodal sensing enables more robust and context-aware analysis by combining complementary modalities. The final stage highlights perception-centric AIoT, where edge intelligence and AI models support real-time interpretation and autonomous decision making. This progression reflects a shift from passive sensing toward active understanding of environments. Overall, the figure captures how sensing, perception, and intelligence become increasingly interconnected in modern AIoT systems.

### Discussion

Compared with previous reviews on IoT and AIoT systems, which often focus on either sensor networks, communication architectures, or specific computer vision applications, This study offers a more integrated perspective by examining the evolution of sensing modalities across domains. While earlier reviews typically analyze technological components in isolation, the present review highlights the convergence of sensing, perception, and distributed intelligence. This integrative perspective allows for a more comprehensive understanding of how AIoT systems evolve toward perception-centric architectures.

The results synthesized in Table 5 reveal several important gaps that emerge consistently across the reviewed studies. First, although multimodal sensing is increasingly adopted, many systems still process visual and non-visual data in parallel rather than within fully integrated perception pipelines. This indicates a gap in unified data fusion strategies that can jointly interpret heterogeneous inputs. Second, scalability remains a critical challenge, as vision-based IoT systems generate large volumes of data that demand substantial computational and communication resources. Third, most implementations remain domain-specific, with limited evidence of generalized perception architectures that can be transferred across application contexts. These gaps indicate that the transition toward fully perception-centric AIoT systems is still evolving rather than complete.

These findings provide several important implications for both research and practice. From a technological perspective, the dominance of deep learning methods and edge-based architectures suggests a clear shift toward real-time, distributed intelligence in IoT environments. Vision-based sensing can enable systems to move beyond simple monitoring toward contextual understanding, which is essential for intelligent decision making. At the same time, the consistent integration of multimodal sensing highlights the need for combining complementary data sources to improve robustness and reliability. For practitioners, these results suggest that effective AIoT system design should prioritize edge intelligence, scalable architectures, and multimodal integration rather than relying on single-modality sensing approaches.

From a theoretical perspective, this review contributes to the understanding of AIoT systems by conceptualizing sensing as an evolving paradigm rather than a static component. The proposed framework highlights how perception emerges from the integration of sensing modalities, artificial intelligence, and distributed computing. This perspective extends existing IoT models by incorporating perception as a core layer that bridges data acquisition and intelligent decision making. From a practical perspective, the findings suggest that effective AIoT system design should prioritize edge-enabled architectures, lightweight AI models, and multimodal data integration to ensure scalability and real-time performance. Practitioners should also consider dataset diversity and deployment conditions when developing AIoT applications, as model performance is highly sensitive to environmental variability, particularly in facial expression recognition tasks where data variability affects model performance (Utami et al., 2022a). These considerations are particularly important in real-world deployments where system robustness and adaptability are critical.

Despite these contributions, several limitations should be considered when interpreting the results. The review relied on a single database, which may limit the coverage of relevant studies. In addition, only English-language publications were included, potentially introducing language bias. The included studies also exhibit heterogeneity in datasets, experimental setups, and evaluation metrics, which constrains direct comparison across studies. These limitations indicate that the findings should be interpreted as indicative patterns rather than definitive generalizations. In addition to the identified technological gaps, several methodological limitations are evident across the reviewed studies. First, there is a lack of standardized evaluation protocols, making it difficult to compare performance across different AIoT vision systems. Second, many studies rely on domain-specific or proprietary datasets, which introduces dataset bias and limits generalizability. Third, benchmark datasets for integrated multimodal AIoT perception remain scarce, hindering reproducible research and cross-study validation. These limitations indicate the need for more standardized experimental frameworks and publicly available datasets to support robust and comparable evaluation of AIoT systems.

Building on these limitations and identified gaps, several future research directions emerge. Developing lightweight deep learning models suitable for edge deployment remains a key priority, particularly for real-time applications in resource-constrained environments. Further work is also needed to design robust multimodal sensing frameworks that can effectively integrate visual and non-visual data within unified perception pipelines. In addition, future AIoT systems should explore more advanced architectures that combine edge intelligence, cloud coordination, and collaborative learning mechanisms. Another important direction involves the development of domain-specific datasets and algorithms for advanced perception tasks, such as facial expression recognition, which require high-quality annotated data and robust models capable of handling real-world variability. Current studies in the dataset rarely address standardized datasets or cross-domain generalization, indicating an opportunity for more structured data curation and benchmarking strategies. Such developments may support scalable and adaptive perception systems capable of operating across diverse and dynamic environments. In particular, advancing facial expression recognition in AIoT requires more representative datasets, robust cross-domain generalization, and edge-optimized models that can operate reliably under real-world constraints such as occlusion, lighting variation, and hardware heterogeneity.

## CONCLUSION AND SUGGESTIONS

### Conclusion

This systematic review synthesizes evidence from 24 empirical studies to examine the evolution of sensing modalities in AIoT systems, with a particular focus on vision-based sensing. The findings show that AIoT vision systems have been increasingly implemented across domains such as manufacturing, agriculture, infrastructure, surveillance, and assistive technologies. These systems consistently rely on a combination of computer vision tasks, including detection, classification, segmentation, and activity recognition, supported by deep learning methods and distributed computing architectures. The results also indicate that edge-based and hybrid edge–cloud systems are becoming the dominant deployment paradigm, enabling real-time processing while maintaining scalability. In addition, multimodal sensing emerges as a critical strategy for improving robustness and contextual understanding by integrating visual data with complementary sensor inputs. Taken together, these findings demonstrate how AIoT systems are transitioning from isolated sensing toward integrated perception capabilities.

Beyond empirical synthesis, this review highlights a clear technological shift from traditional sensor-based monitoring to perception-centric AIoT systems. Vision-enabled sensing plays a central role in this transition by enabling systems to interpret spatial and contextual information rather than relying solely on numerical data. The integration of artificial intelligence, edge computing, and multimodal sensing further supports this evolution toward intelligent perception and autonomous decision making. The findings also indicate early developments toward higher-level perception tasks, including affective computing such as facial expression recognition, although these remain limited by dataset availability and generalization challenges. This progression underscores the need for AIoT systems that are not only capable of sensing environments but also of understanding and responding to complex real-world conditions.

### Suggestions

Based on the synthesized findings, several practical recommendations can be proposed for researchers and system developers. First, future AIoT vision systems should prioritize edge-oriented architectures combined with lightweight deep learning models to enable real-time processing under resource constraints. Second, system design should move beyond single-modality sensing by incorporating multimodal data fusion strategies that integrate visual, environmental, and contextual information to improve reliability and robustness. Third, researchers are encouraged to develop standardized and representative datasets that reflect real-world deployment conditions, particularly for advanced perception tasks such as facial expression recognition and human-centered monitoring. Finally, the design of AIoT systems should emphasize scalable and adaptive architectures that support cross-domain applicability, enabling intelligent monitoring systems to operate effectively across diverse environments such as smart farming, smart infrastructure, and human-centric applications.

## REFERENCES

Abdullahu, E., Wache, H., & Piangerelli, M. (2025). Secure and Decentralized Hybrid Multi-Face Recognition for IoT Applications. Sensors, 25(18), 5880. https://doi.org/10.3390/s25185880

Adu-Manu, K. S., Abdulai, J., Engmann, F., Akazue, M., Appati, J. K., Baiden, G. E., & Sarfo-Kantanka, G. (2022). WSN Architectures for Environmental Monitoring Applications. Journal of Sensors, 2022, 1–18. https://doi.org/10.1155/2022/7823481

Booth, A., Noyes, J., Flemming, K., Gerhardus, A., Wahlster, P., Wilt, G. J. van der, Mozygemba, K., Refolo, P., Sacchini, D., Tummers, M., & Rehfuess, E. (2018). Structured Methodology Review Identified Seven (RETREAT) Criteria for Selecting Qualitative Evidence Synthesis Approaches. Journal of Clinical Epidemiology, 99, 41–52. https://doi.org/10.1016/j.jclinepi.2018.03.003

Booth, A., Noyes, J., Flemming, K., Moore, G., Tunçalp, Ö., & Shakibazadeh, E. (2019). Formulating Questions to Explore Complex Interventions Within Qualitative Evidence Synthesis. BMJ Global Health, 4(Suppl 1), e001107. https://doi.org/10.1136/bmjgh-2018-001107

Cardinale, Y. (2020). Occupant Activity Detection in Smart Buildings: A Review. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3671533

Chebrolu, V., Koona, R., & Raju, R. S. U. (2025). Machine vision based surface roughness assessment system based on the Internet of Things and contourlet transforms. International Journal on Interactive Design and Manufacturing, 19(1), 83–98. https://doi.org/10.1007/s12008-023-01515-1

Chen, C. C., Ba, J. Y., Li, T. J., Chan, C. C. K., Wang, K. C., & Liu, Z. (2021). EfficientNet: A low-bandwidth IoT image sensor framework for cassava leaf disease classification. Sensors and Materials, 33(11), 4031–4044. https://doi.org/10.18494/SAM.2021.3526

Chenguang, L. (2023). An Efficient Vision-based Approach for Optimizing Energy Consumption in Internet of Things and Smart Homes. In IJACSA) International Journal of Advanced Computer Science and Applications (Vol. 14, Number 6). www.ijacsa.thesai.org

Djenouri, Y., Belhadi, A., Srivastava, G., Houssein, E. H., & Lin, J. C. (2021). Sensor Data Fusion for the Industrial Artificial Intelligence of Things. Expert Systems, 39(5). https://doi.org/10.1111/exsy.12875

Elordi, U., Lunerti, C., Unzueta, L., Goenetxea, J., Aranjuelo, N., Bertelsen, Á., & Arganda-Carreras, I. (2021). Designing Automated Deployment Strategies of Face Recognition Solutions in Heterogeneous IoT Platforms. Information, 12(12), 532. https://doi.org/10.3390/info12120532

Fogarty, A. E., Conger, A., & Burnham, T. (2022). Systematic Reviews: Not Always a Pain. Interventional Pain Medicine, 1, 100128. https://doi.org/10.1016/j.inpm.2022.100128

Ghoreishi, M., Treves, L., & Kuivalainen, O. (2022). Artificial Intelligence of Things as an Accelerator of Circular Economy in International Business. 83–104. https://doi.org/10.1007/978-3-030-95527-4_6

Hao, Y. (2023). Applying Mechatronics-Based Internet Of Things In Artificial Intelligence Image Detection. In Applying Mechatronics-Based Internet of Things in Artificial Intelligence Image Detection International Journal of Mechatronics and Applied Mechanics.

Haqiq, N., Zaim, M., Bouganssa, I., Salbi, A., & Sbihi, M. (2022). AIoT With I4.0: The Effect of Internet of Things and Artificial Intelligence Technologies on the Industry 4.0. Itm Web of Conferences, 46, 03002. https://doi.org/10.1051/itmconf/20224603002

Hudda, S., Barnwal, R., Khurana, A., & Haribabu, K. (2024). A WSN and vision based smart, energy efficient, scalable, and reliable parking surveillance system with optical verification at edge for resource constrained IoT devices. Internet of Things (Netherlands), 28. https://doi.org/10.1016/j.iot.2024.101346

Jaramillo-Hernández, J. F., Julian, V., Marco-Detchart, C., & Rincón, J. A. (2024). Application of Machine Vision Techniques in Low-Cost Devices to Improve Efficiency in Precision Farming. Sensors, 24(3). https://doi.org/10.3390/s24030937

Jia, L., & Wang, Y. (2022). Research on Industrial Production Defect Detection Method Based on Machine Vision Technology in Industrial Internet of Things. Traitement Du Signal, 39(6), 2061–2068. https://doi.org/10.18280/ts.390618

Kahale, L. A., El-Khoury, R., Mikati, I. E., Pardo-Hernández, H., Khamis, A. M., Schünemann, H. J., Haddaway, N., & Akl, E. A. (2022). Tailored PRISMA 2020 Flow Diagrams for Living Systematic Reviews: A Methodological Survey and a Proposal. F1000research, 10, 192. https://doi.org/10.12688/f1000research.51723.3

Kang, X., Zhu, B., Cai, Y., Xiao, Y., Liu, N., Guo, Z.-X., Wang, Q., & Luo, Y. (2025). A Concise Review of State-of-the-Art Sensing Technologies for Bridge Structural Health Monitoring. Sensors, 25(17), 5460. https://doi.org/10.3390/s25175460

Katsigiannis, M., & Mykoniatis, K. (2024). Enhancing Industrial IoT with Edge Computing and Computer Vision: An Analog Gauge Visual Digitization Approach. In Manufacturing Letters (Vol. 41). www.sciencedirect.com

Ke, R., Zhuang, Y., Pu, Z., & Wang, Y. (2021). A Smart, Efficient, and Reliable Parking Surveillance System With Edge Artificial Intelligence on IoT Devices. Ieee Transactions on Intelligent Transportation Systems, 22(8), 4962–4974. https://doi.org/10.1109/tits.2020.2984197

Kia, S., & Leiding, B. (2025). Intelligent Monitoring of BECS Conveyors via Vision and the IoT for Safety and Separation Efficiency. Applied Sciences (Switzerland), 15(11). https://doi.org/10.3390/app15115891

Kwon, H.-M., & Seo, J. (2022). Effect of Compressed Sensing Rates and Video Resolutions on a PoseNet Model in an AIoT System. Applied Sciences, 12(19), 9938. https://doi.org/10.3390/app12199938

Li, B., Zou, Y., Zhu, R., Yao, W., Wang, J., & Wan, S. (2022). Fabric Defect Segmentation System Based on a Lightweight GAN for Industrial Internet of Things. Wireless Communications and Mobile Computing, 2022. https://doi.org/10.1155/2022/9680519

Lv, Z., & Chen, D. (2021). Industrial visual perception technology in Smart City. Image and Vision Computing, 105. https://doi.org/10.1016/j.imavis.2020.104070

Mardanshahi, A., Sreekumar, A., Yang, X., Barman, S. K., & Chronopoulos, D. (2025). Sensing Techniques for Structural Health Monitoring: A State-of-the-Art Review on Performance Criteria and New-Generation Technologies. Sensors, 25(5), 1424. https://doi.org/10.3390/s25051424

Matarneh, S., Elghaish, F., Edwards, D. J., Rahimian, F. P., Abdellatef, E., & Ejohwomu, O. (2024). Automatic Crack Classification On Asphalt Pavement Surfaces Using Convolutional Neural Networks And Transfer Learning. Journal of Information Technology in Construction, 29, 1239–1256. https://doi.org/10.36680/j.itcon.2024.055

Miran, A., & Ayub, H. (2023). The Investigation and Challenges of Advanced Applications Using Artificial IoT on Edge Computing. Qalaai Zanist Scientific Journal, 8(2). https://doi.org/10.25212/lfu.qzj.8.2.55

Nagajayanthi, B. (2021). Decades of Internet of Things Towards Twenty-First Century: A Research-Based Introspective. Wireless Personal Communications, 123(4), 3661–3697. https://doi.org/10.1007/s11277-021-09308-z

Nasiri, S., Sadoughi, F., Dehnad, A., Tadayon, M. H., & Ahmadi, H. (2021). Layered Architecture for Internet of Things-Based Healthcare System: A Systematic Literature Review. Informatica, 45(4). https://doi.org/10.31449/inf.v45i4.3601

Nguyen, H. H., Shin, D. Y., Jung, W. S., Kim, T. Y., & Lee, D. H. (2024). An Integrated IoT Sensor-Camera System toward Leveraging Edge Computing for Smart Greenhouse Mushroom Cultivation. Agriculture (Switzerland), 14(3). https://doi.org/10.3390/agriculture14030489

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J., Akl, E. A., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. BMJ, n71. https://doi.org/10.1136/bmj.n71

Pasolini, G., Guerra, A., Guidi, F., Decarli, N., & Dardari, D. (2020). Crowd-Based Cognitive Perception of the Physical World: Towards the Internet of Senses. Sensors, 20(9), 2437. https://doi.org/10.3390/s20092437

Perera, C., Liu, C. H., & Jayawardena, S. (2015). The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. Ieee Transactions on Emerging Topics in Computing, 3(4), 585–598. https://doi.org/10.1109/tetc.2015.2390034

Qiu, H., Zheng, Q., Zhang, T., Qiu, M., Memmi, G., & Lu, J. (2021). Toward Secure and Efficient Deep Learning Inference in Dependable IoT Systems. Ieee Internet of Things Journal, 8(5), 3180–3188. https://doi.org/10.1109/jiot.2020.3004498

Rajendran, M., Stephan, P., Stephan, T., Agarwal, S., & Kim, H. (2022). Cognitive IoT Vision System Using Weighted Guided Harris Corner Feature Detector for Visually Impaired People. Sustainability (Switzerland), 14(15). https://doi.org/10.3390/su14159063

Ravindran, A. (2023). Internet-of-Things Edge Computing Systems for Streaming Video Analytics: Trails Behind and the Paths Ahead. Iot, 4(4), 486–513. https://doi.org/10.3390/iot4040021

Ray, P. P. (2017). Internet of Things for Smart Agriculture: Technologies, Practices and Future Direction. Journal of Ambient Intelligence and Smart Environments, 9(4), 395–420. https://doi.org/10.3233/ais-170440

Rodríguez, I., Campos, C., & Fdez-Riverola, F. (2023). Horizontally Distributed Inference of Deep Neural Networks for AI-Enabled IoT. Sensors, 23(4), 1911. https://doi.org/10.3390/s23041911

Ross, R., Parsons, L., Thai, B. S., Hall, R., & Kaushik, M. (2020). An IOT smart rodent bait station system utilizing computer vision. Sensors (Switzerland), 20(17), 1–15. https://doi.org/10.3390/s20174670

Sager, M., & Pistone, I. (2019). Mismatches in the Production of a Scoping Review: Highlighting the Interplay of (In)formalities. Journal of Evaluation in Clinical Practice, 25(6), 930–937. https://doi.org/10.1111/jep.13251

Saif, Y., Yusof, Y., Latif, K., Kadir, A. Z. A., Ahmed, M. binti Iliyas, Adam, A., Hatem, N., & Memon, D. A. (2022). Roundness Holes' Measurement for milled workpiece using machine vision inspection system based on IoT structure: A case study. Measurement: Journal of the International Measurement Confederation, 195. https://doi.org/10.1016/j.measurement.2022.111072

Seng, K. P., Ang, L.-M., & Ngharamike, E. (2022). Artificial Intelligence Internet of Things: A New Paradigm of Distributed Sensor Networks. International Journal of Distributed Sensor Networks, 18(3), 155014772110628. https://doi.org/10.1177/15501477211062835

Silva, S. V. F. da, Ferreira, D. C., Lopes, J. E. F., & Carvalho, L. M. d. (2025). Artificial Intelligence and Buying Behavior: Exploring the Balance Between Convenience and Privacy in E-Commerce From the Perspective of Regionality. Are, 7(7), 39930–39952. https://doi.org/10.56238/arev7n7-277

Sun, X. jun, & Lin, J. C. W. (2022). A Target Recognition Algorithm of Multi-Source Remote Sensing Image Based on Visual Internet of Things. Mobile Networks and Applications, 27(2), 784–793. https://doi.org/10.1007/s11036-021-01907-1

Syed, A. S., Sierra-Sosa, D., Kumar, A., & Elmaghraby, A. (2021). IoT in Smart Cities: A Survey of Technologies, Practices and Challenges. Smart Cities, 4(2), 429–475. https://doi.org/10.3390/smartcities4020024

Utami, P., Hartanto, R., & Soesanti, I. (2022a). The Development of Facial Expressions Dataset for Teaching Context: Preliminary Research. 2022 5th International Conference of Computer and Informatics Engineering (IC2IE), 263–269. https://doi.org/10.1109/IC2IE56416.2022.9970043

Utami, P., Hartanto, R., & Soesanti, I. (2022b). The EfficientNet Performance for Facial Expressions Recognition. 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 756–762. https://doi.org/10.1109/ISRITI56927.2022.10053007

Utami, P., Mashoedah, Nurkhalis, H., Nafi', M. A., Savitri, W., Prastowo, W., Diah Wulan, S., & Saputra, F. D. (2024). An Integrated IoT–AI Architecture for Precision Beekeeping: Sensing, Data Communication, Colony-State Intelligence, and Decision-Oriented Actions. Jurnal Media Computer Science, 3(2), 179–198.

Utami, P., Zakarijah, M., Fikriawan, Z., Yanti, N., Aritonang, D., Gendhis Pertiwi, G., Rizqy Arba Pratama, A., & Azra, D. (2024). Iot-Enabled Dairy Systems: From Sensing And Data Integration To Operational Evaluation How to Cite. Jurnal Media Computer Science, 3(3), 199–216. https://doi.org/https://doi.org/10.37676/jmcs.v3i2.10590

Vukićević, A., Mladineo, M., Banduka, N., & Mačužić, I. (2021). A smart warehouse 4.0 approach for the pallet management using machine vision and Internet of things (IoT): A real industrial case study. Advances in Production Engineering And Management, 16(3), 297. https://doi.org/10.14743/apem2021.3.401

Wang, H., Lu, R., & Yu, D. (2022). Machine Vision Nondestructive Inspection System Assisted by Industrial IoT Supervision Mechanism. Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/8449518

Xiao, Y., & Watson, M. (2017). Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research, 39(1), 93–112. https://doi.org/10.1177/0739456x17723971

Yahya, F. B., Lukas, C. J., & Calhoun, B. H. (2018). A Top-Down Approach to Building Battery-Less Self-Powered Systems for the Internet-of-Things. Journal of Low Power Electronics and Applications, 8(2), 21. https://doi.org/10.3390/jlpea8020021

Yang, B., Yang, S., Zhu, X., Qi, M., Li, H., Lv, Z., Cheng, X., & Wang, F. (2023). Computer Vision Technology for Monitoring of Indoor and Outdoor Environments and HVAC Equipment: A Review. Sensors, 23(13), 6186. https://doi.org/10.3390/s23136186

Yoon, S., & Kim, B. S. (2025). Multi-Scale Temporal Fusion Network for Real-Time Multimodal Emotion Recognition in IoT Environments. Sensors, 25(16), 5066. https://doi.org/10.3390/s25165066

Yu, F. (2022). Implementation of a Human Motion Capture System Based on the Internet of Things Machine Vision. Journal of Cases on Information Technology, 24(5). https://doi.org/10.4018/JCIT.302245

Zhang, J., Du, X., & Bi, R. (2022). Intelligent Recognition System of Sports Athletes' Wrong Actions Based on AI+IoT. Wireless Communications and Mobile Computing, 2022. https://doi.org/10.1155/2022/3455224

Zhang, J., & Tao, D. (2021). Empowering Things With Intelligence: A Survey of the Progress, Challenges, and Opportunities in Artificial Intelligence of Things. Ieee Internet of Things Journal, 8(10), 7789–7817. https://doi.org/10.1109/jiot.2020.3039359

Zhang, S., Ni, P., Wen, J., Zhou, R., Han, Q., Du, X., & Li, J. (2025). Embedded Vision-Based Sensing System for Noncontact Cable Vibration Monitoring With IoT Technologies. Structural Control and Health Monitoring, 2025(1). https://doi.org/10.1155/stc/6945296