Intelligent Product Inspection System Based on Machine Vision and Multi-Sensor Information Fusion

Wenxin Dong[1]*, Shiyuan Liu[2], Bolu Wang[3]

[1]Leicester International Institute, Dalian University of Technology, Panjin, 124200, China
[2]School of Software, Henan University of Science and Technology, Luoyang, 471000, China
[3]Institute Hohai-Lille, Hohai University, Changzhou, 213200, China
*20223291060@mail.dlut.edu.cn

Abstract. With the development of contemporary machinery and the automation industry, intelligent systems have been continuously applied in the production of products. As an important pillar of intelligent systems, machine vision, and sensor fusion technology cannot be dispensed in the process of product inspection. Machine vision technology simulates human vision through computers, extracts product information, speeds up processing efficiency, and is ultimately used in inspection. Sensor fusion technology denotes the integration and processing of data from several different sensors to enhance the reliability of the information. However, traditional research has often been confined to the isolated application of either machine vision or multi-sensor fusion technology without integrating both approaches in a combined study. In this paper, vision-related algorithms, image recognition, object detection, and multi-sensor fusion technology are the focus of the research, combined with their application in product inspection for theoretical and practical research. By cross-applying the two technologies, the intrinsic connection between them is analysed, resulting in finer data processing and good analysis and response capabilities. This study is of great value in accelerating the advancement of product inspection technology, improving the accuracy and efficiency of product inspection, and promoting the intelligent upgrading of the whole industry.

Keywords: Machine Vision, Multi-Sensor Information Fusion, Product Inspection System

1 Introduction

With the development of the global industry, the manufacturing industry is moving at an unprecedented speed in the direction of intelligence, automation, and continuous development, in which intelligent manufacturing has become a key concept to give machine intelligence, which is profoundly affecting every production link. An intelligent product inspection system is an important part of the production process, but also, to enhance product quality, enhancing the competitiveness of enterprises is the key.

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_60

Traditional manual inspection methods are not only inefficient, low precision, high cost and easy to disturb by human factors, but also in metal additive manufacturing and metallurgy and other certain industrial fields with complex structures in the field, the product is very sensitive to small process changes, so the automated inspection system has gradually become a manufacturing industry to improve productivity and ensure product quality of the key technologies.

Quality control can be significantly improved by technological innovations, and machine vision is an innovative technology that enables efficient inspection, enhances productivity, and improves the inspection ecosystem in today's manufacturing industry. The use of computer vision technology in industry, which transmits information through additional hardware and network connections, is widely used in the inspection of microprocessors, automobiles, food, pharmaceuticals, and other products. It automates the inspection process and improves inspection accuracy and productivity. When the scale of industrial production is enlarged and the production precision is improved, higher requirements are also put forward for product inspection, especially in high-precision manufacturing and harsh conditions manufacturing, product inspection is difficult or even dangerous, which is difficult for human beings to perform, and the machine vision and sensor fusion technology has great potential to play.

Machine vision and sensing technologies have made significant progress nowadays, and breakthroughs have been achieved in many fields, such as self-driving cars, healthcare, agriculture, and industrial production [1-4]. However, there are also many aspects of shortcomings, such as the differences in the types of sensors and data, as the machine vision system usually needs to be fused with other sensors (such as laser sensors, infrared sensors, ultrasonic sensors, etc.), and the data generated by different sensors varies greatly in terms of accuracy, format, timestamps, etc., how to effectively fuse the data from multiple sources to obtain more accurate detection results is a technical challenge. Similarly, there may be slight colour differences or surface texture variations in different batches of products, which poses a challenge to the inspection accuracy and stability of the vision system.

Machine vision-driven, multi-sensor information fusion product inspection techniques are not widely and intensively used in current industrial production. Currently, only a small number of special product quality inspections will use this technology, therefore, this research is currently facing several challenges, the most basic of which is that many sensors will inevitably lead to an increase in computational costs, labour consumption, and time consumption. In addition, the combined analysis of multiple data simultaneously requires greater computational power, thus increasing the difficulty and conditions of the analysis. This challenge involves cross-fertilisation and synergy between multiple disciplines and industrial fields.

This paper focuses on machine vision and multi-sensor information fusion technology as the primary approaches, systematically exploring the theoretical and practical applications of their collaborative effects in intelligent product inspection systems. The aim is to achieve efficient integration of visual information and multi-source data through various methods in combination with industrial scene requirements. This research not only provides methodological guidance for the optimal design of intelligent inspection systems but also offers practical paths to solve technical bottlenecks in actual industrial scenarios (such as data fusion delay and environmental interference), demonstrating both academic foresight and engineering application value.

2 Machine Vision Technology

2.1 Basics of Machine Vision Technology

Machine vision systems can be divided into four steps: image acquisition, processing, recognition, and classification. Objects reflect light from some kind of light source into an image sensor to generate an optical image. The optical image is subsequently transformed into a digital image by using image sensing and image data-mining techniques. To carry out subsequent analysis work, the digital image needs to be modified and adjusted in terms of pixel characteristics, area size, etc., i.e., image processing. Feature recognition techniques are often used to identify the inherent features of an object, and in recent years, deep learning (CNN) techniques have greatly improved the efficiency of feature recognition. As shown in Fig. 1, based on CNN feature extraction, target detection techniques gradually get rid of traditional manual target detection methods (e.g., HOG+SVM) for integrated feature localization and classification.

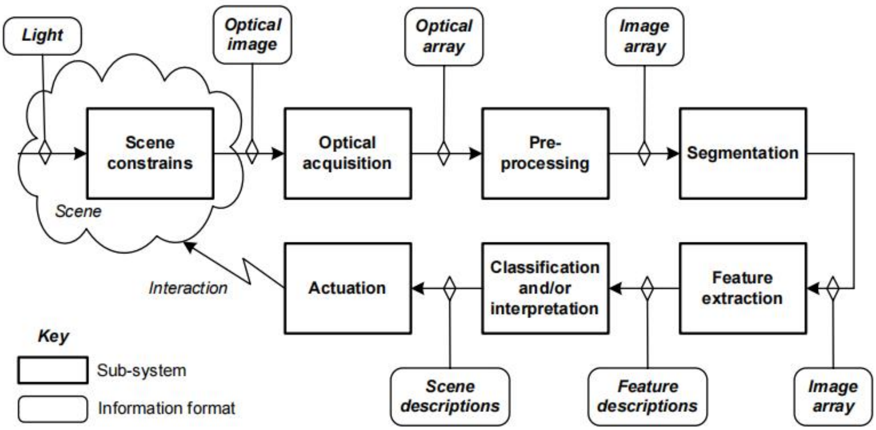

Fig. 1. A block diagram for a typical vision system operation [5]

2.1.1 Traditional image preprocessing and feature representation.

Image preprocessing includes image acquisition and processing, and commonly used techniques include grayscaling and noise removal, image enhancement and contrast adjustment, and geometric transformation and normalization. Traditional feature representation mainly uses SIFT, HOG, and other methods. The purpose is to extract differentiated information (e.g., edges, texture, semantic features) from the original data, and its research focus is mainly on how to represent the data efficiently and robustly.

Gonzalez and Woods detailed the key techniques of filtering (Gaussian and median filtering) and histogram equalization to reduce the noise of the image while retaining the edge details appropriately, in addition to enhancing the contrast to enhance the recognizability of the image [6].

Images often have a variety of information, such as affine distortion, illumination changes, 3D viewpoint changes, etc., and the image scale and rotation angle of the image will also have an impact on the recognition of the image. Lowe et al. put forward a method to extract unique invariant features (SIFT) [7]. This method requires the manual creation of feature databases to ensure the accuracy of recognition. Multiple algorithms are employed during the recognition process to compare with database information in real time, thereby mitigating the adverse effects of blurring and occlusion on recognition.

The main method for feature extraction is convolutional neural networks, of which the unsaturated neurons and very efficient GPU convolutional operations proposed by Krizhevsky et al. are the most classical [8]. This multi-layer convolution and fully connected structure enables the network to derive features from all levels of the image while having enough capacity to fit large-scale datasets. A "dropout" regularization method is used to reduce overfitting in the fully connected layers.

Before the maturity of deep learning technology, traditional image preprocessing and feature detection formed a more mature method and system, but the computational cost was high and lacked real-time performance. As deep learning technology continues to advance, image preprocessing will be more intelligent and automated.

2.1.2 Target detection based on deep learning.

Training and running deep learning models typically demand substantial computational resources. With the advancement of graphic processors and other specialized hardware in recent years, computational power has been greatly boosted, enabling deep learning models to handle large-scale image data more efficiently. This increase in arithmetic power has provided strong support for the application of deep learning in image preprocessing and feature representation, allowing complex models and algorithms to be run and optimized faster.

Girshick et al. proposed a fast region-based convolutional network for target detection, Fast R-CNN [9]. This method builds on previous research by utilizing a deep convolutional network to efficiently classify target candidate regions. Compared with previous methods, Fast R-CNN incorporates several innovations that enhance both the training and testing speed while also improving accuracy. of detection. For example, When training the very deep VGG16 network, Fast R-CNN is 9 times faster than R-CNN, and it is 213 times faster in the testing phase while achieving a higher mean average precision (mAP) on the PASCAL VOC 2012 dataset.

However, as the scenarios in which machine vision is applied become more and more widespread, the speed of target detection continues to be more demanding. Two-stage detection frameworks generally filter out most of the information in the image that may not match the target and then analyze and calculate the remaining part in a refined way, following a pattern from coarse to fine. Despite their relatively high localization and classification accuracies, they clearly have the problem of slow inference speed. The first real-time, single-stage object detector in the deep learning era (YOU ONLY LOOK ONCE, YOLO) was proposed by the Facebook Artificial Intelligence Institute, in 2016 in conjunction with the Allen Institute for Artificial Intelligence and the University of Washington [10]. YOLO only uses a single neural network for image detection, which segments the image into multiple feature regions and then removes the overlapping parts after bounding box regression, Significantly boosting the velocity of image detection. With the iteration of the product, the localization accuracy of YOLO continues to improve.

Mnih et al. pioneered a recurrent neural network model that incorporates the attention mechanism into deep neural networks to adaptively and repeatedly predict important regions, allowing the amount of computation performed to be independent of the image size, breaking the linear relationship between the amount of computation and the quantity of pixels in the image [11]. Jaderberg et al. then introduced a learnable sub-network in the convolutional neural network (Spatial Transformer), using this module to predict significant regions in the input features (possibly transforming the feature map) without training surveillance [12]. The results show that its use of the spatial transformer makes the model learning more flexible and still performs well with complex benchmarks and types. Hu et al. proposed a SENet architecture with good generalization capabilities, which can adaptively find hidden features in multiple sets of objects without providing any inputs for the features, greatly reducing computational costs [13]. Wang et al. were motivated by computer vision inspired by the classical non-local averaging method, incorporating a self-attention mechanism to weight all positional features to handle local operations, and could be inserted into various computer vision architectures to significantly enhance the throughput of video understanding and object recognition [14].

The foundation of machine vision technology covers image preprocessing, feature extraction, and multiple implementations of deep learning algorithms in target recognition, and the development of key technologies presents evolution from artificial features to automatic learning, dual optimization of computational efficiency and accuracy, and architectural innovations brought about by the attention mechanism.

But even so, there are still limitations in relying only on visual sensors for information acquisition. Although deep learning algorithms have achieved good results in target detection and image recognition, the algorithm's generalization ability is still insufficient when dealing with small differences in product appearance, especially when changes such as color and texture are involved. In addition, in complex environments such as strong light, high noise, or electromagnetic interference, the performance of vision sensors may be greatly reduced, affecting the accuracy of detection.

Multi-sensor fusion technology fuses multiple sensors to obtain multi-dimensional physical data, which can notably enhance the accuracy and robustness of recognition. Particularly in product inspection, led by machine vision, multi-sensor fusion technology has become the main direction of research.

2.2 Multi-sensor Information Fusion Technology

2.2.1 Sensor technology.

Sensors are devices that read changes in the surrounding environment and map them into readable quantitative data, which is subsequently processed further. In smart product monitoring systems, sensors act as their organs of perception, facilitating the sensing of subtle changes in the product. In multi-sensor fusion, the data collection and types of sensors play a crucial role in data input and processing.

For traditional sensors such as optical, infrared, and LiDAR sensors, the use of different sensors alone still has insurmountable drawbacks in data collection and fusion. Adarsh et al. compared infrared sensors with ultrasonic sensors and found that they have different zones of superiority, and therefore the sampling and collection of data using such sensors alone cannot fully utilise their advantages [15]. Then their advantages cannot be fully utilised. Table 1 shows the sensor-related operational variables in the experimental setup.

Table 1. The sensor-related operational variables in the experimental setup [15]

Specifications	IR Sensor (sharp gp2y0a21ykof)	Ultra Sonic Sensor (HC SR-04)
Range	10cm-80cm	2cm-10m
Beam-width	75Deg.	30Deg.
Beam Pattern	Narrow (line)	Conical
Frequency	353 THz	40KHz
Unit Cost	~750INR	~130INR

Multi-sensor fusion (MSF) is the integration of multiple sensors to process and analyse changes in the external environment efficiently. Multi-sensor fusion can be classified into three parts according to the data modality and the result of processing, the signal-pixel level (lower level), the feature-based level (mid-level), and the symbolic level (higher level) [16].

Multi-sensor fusion has significant advantages over single sensors (e.g., vision sensors) in that inaccuracies and ambiguities are reduced, signal-to-noise ratios are improved, and the accuracy and reliability of data processing and information reception are enhanced [17].

2.2.2 Multi-sensor fusion development.

In multi-sensor fusion, MSF technology plays a central role in enhancing perceptual performance in the face of multiple challenges, such as data synchronisation problems, noise suppression, and efficient management of high-dimensional information provided by multiple sensors. A. El-Dalahmeh et al. developed a novel MSF strategy that cleverly incorporates deep neural network technology, an innovation that significantly enhances vehicle detection performance [18]. The strategy combines vision camera, LiDAR and RADAR data with a specially designed uncertainty assessment architecture to achieve excellent target recognition accuracy (measured by mean average precision mAP) and fast processing rates - high scores of 80.1 and 25.1 frames per second, respectively. 25.1 frames per second, respectively. These performance metrics not only outperform the current mainstream methods but also further confirm the real-time application of this scheme in the field of self-driving navigation, which can maintain high efficiency and accuracy even in complex and changing driving scenarios, thus consolidating its technological leadership. Zhou T et al. developed a framework for perceptual data acquisition. Based on this, a deep data fusion method was innovatively designed [19]. The method is modelled on the YOLOv2 (i.e., you only look once V2) algorithm and the neural network architecture is customised to attain cross-tier data fusion across depth levels. The effectiveness of the method was validated by feeding multivariate sensor data into a convolutional neural network (CNN) using the KITTI dataset. A comparative analysis of RECOLLECT and PRECISION reveals that the proposed deep fusion scheme demonstrates higher efficiency advantages compared to traditional post-fusion algorithms.

Anon et al. proposed an indoor people dynamic monitoring scheme based on an intelligent data processing platform by fusing a multi-source sensor cooperative working mechanism with innovative machine learning algorithms [20]. The real-time capture of human movement trajectory is achieved by a Passive Infrared Sensor (PIR) and ultrasonic sensors, which in turn effectively resolves the displacement direction and relative distance of people in the room. At the data processing level, decision trees and K-Means machine learning algorithms are used to perform supervised feature extraction and unsupervised pattern analysis on the captured time-series signals to achieve accurate resolution of motion parameters. The results of multiple controlled tests in standardised experimental scenarios show that this multi-sensor fusion scheme achieves 100% interpretation accuracy in terms of recognition and prediction at 1 m, but the accuracy of recognition decreases as the distance increases or decreases. It means that the proposed system still lacks a certain degree of robustness and accuracy during high-complexity motion.

Although machine learning techniques provide an end-to-end solution to overcome the limitations of traditional methods, it is still important to explore in depth the intelligent fault diagnosis methods combining multi-sensor data. Xie et al. innovatively fused the multi-sensor fusion (MSF) technique with convolutional neural networks (CNNs) to propose a novel intelligent diagnosis framework [21]. Specifically, a strategy based on principal component analysis was first adopted to convert multi-source signal data into a three-channel red-green-blue image format for effective data integration. Subsequently, an optimised CNN architecture that embeds residual network features is designed with the aim of balancing computational complexity with diagnostic accuracy. To validate the efficacy of this method, two independent fault datasets, the KAT bearing dataset and the Gearbox dataset, were selected for testing, and the results show that the method achieves a mean experimental accuracy of 99.99% and 99.98% on the benchmark datasets. The method demonstrates significant advantages in diagnostic accuracy compared to existing deep learning methods.

2.2.3 Advantages and Challenges of Multi-Sensor Information Fusion.

Through the multi-sensor data integration technology, intelligent computing techniques are used to perform in-depth processing and fusion analysis of multi-modal time series data from multi-sensors fused efficiently and collaboratively under dynamic optimisation strategies, to generate more accurate and complete perceptual and decision-making inferences compared to the use of machines alone, and its reasoning performance is significantly better than that of traditional single-source information processing methods [22].

Although multi-sensor fusion technology has made some progress in the field of intelligent detection, it still faces several technical challenges in practical applications. The first is the complexity of multi-source data fusion. Different types of sensors differ in data format, accuracy, and timestamps, and it remains a challenge to accurately synchronise and fuse these data, especially in application scenarios with high real-time requirements. The delay in data fusion may not only affect the response speed of the system but also negatively affect the overall performance. Moreover, the consumption of computing resources and energy efficiency issues need to be addressed when dealing with large-scale data. Furthermore, adverse ambient conditions with elevated temperature parameters and high humidity may also affect the information fusion work of multiple sensors and even cause data distortion. Therefore, ensuring that systems can operate stably under these environments and guarantee high-precision detection is a major challenge at present.

2.3 Research Status of the Intelligent Detection System

Intelligent detection system realizes automatic recognition and judgment of targets by combining sensors, artificial intelligence algorithms, and data analysis technology. The system collects environmental or equipment data, pre-processes it using technologies such as image processing and signal analysis, and analyzes and makes decisions through machine learning or deep learning models. The system can quickly and accurately detect anomalies or targets in complex environments and provide real-time feedback, thereby realizing functions such as automated monitoring, fault diagnosis and early warning, and is widely used in industries such as industry, security, and medical care.

In the industrial field, Benbarrad et al. proposed using machine vision models to predict the most appropriate parameters of the production process, thereby achieving proactive defect detection in the production lines and ensuring superior product excellence [23]. When designing a defect detection architecture, it is imperative to evaluate the collaboration and optimization of various technologies within the production chain, spanning from raw materials to final products. Manufacturing process data is leveraged through predictive quality analytics, enabling closed-loop control systems to not only detect anomalies but also perform root-cause diagnostics for continuous quality enhancement. The model diagram they proposed is shown in Fig. 2.

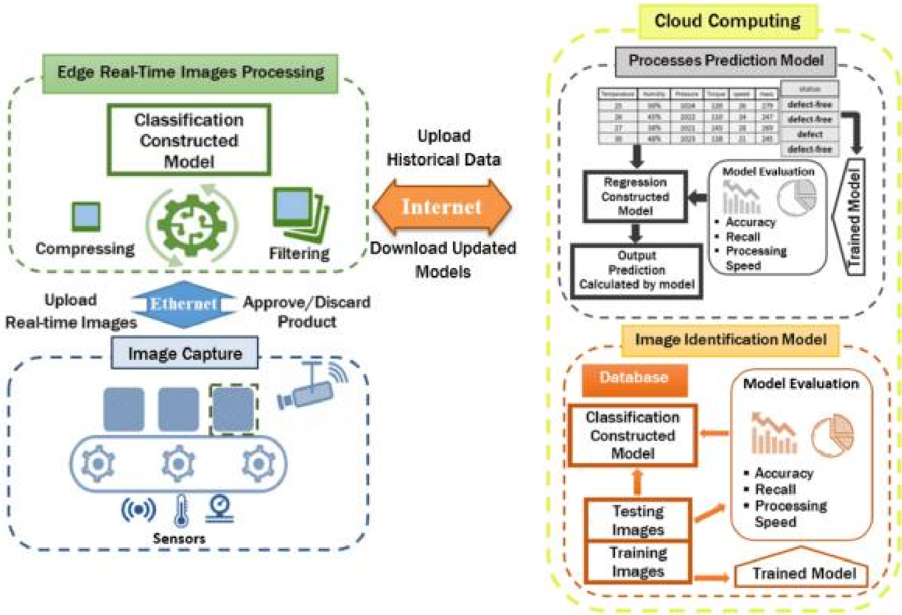

Fig. 2. An intelligent machine vision system for automated product defect detection [23]

Also in the industrial field, Wang et al. formulated a novel production management approach that uses advanced LoRa wireless communication technology to integrate dynamic process tracking and intelligent production scheduling allocation, ensuring effective and simultaneous data collection from multiple sensors, and accurately and timely capturing, transmitting and storing machine status data [24]. The data collection time was halved, and the system responsiveness was significantly improved so that machine failures can be detected and resolved on time.

In the healthcare sector, Ding et al. analyzed the functional requirements of the intelligent data transmission system of IoMT, Thoroughly examined the issues of fault diagnosis and resource allocation encountered in the IoMT model, and proposed an intelligent data transmission model that applies the wireless communication technology of the industrial Internet of Things to the medical Internet of Things (IoMT) [25]. It has high-quality data transmission, high-precision accident diagnosis, and real-time abnormality monitoring capabilities, making up for the shortcomings of the traditional IoMT model. The innovative use of the multimodal data fusion CNN algorithm has achieved the improvement of accident diagnosis classification accuracy, the improvement of IoMT data transmission rate and real-time monitoring capabilities, the reduction of data delay, and the enhancement of system stability.

In the field of food, Zhou et al. used computer vision system (CVS) and electronic nose (e-nose) technology to better detect the quality of black tea fermentation, analyzed the changes of volatile substances, tea polyphenols, image and odour feature values with fermentation time, and divided the fermentation process into three stages [26]. By using PCA and PLS-DA, 51 differential volatiles were screened out, and combined with image and odour features, a fermentation stage classification model was established through the random forest (RF), K nearest neighbour (KNN) and support vector machine (SVM), achieving an accurate detection effect close to 100%.

The intelligent detection system can automatically identify targets in complex environments and provide real-time feedback by combining sensors, artificial intelligence algorithms, and data analysis technologies. However, the existing system faces multiple bottlenecks and challenges, such as insufficient processing capabilities of multi-source data, high real-time requirements, large environmental interference, and problems with model accuracy and generalization capabilities. Particularly in the industrial sector, achieving the collaboration and optimization of different technologies throughout the production process, from raw materials to final products, remains a significant challenge. In addition, a single data source cannot fully cover complex scenarios, and these limitations need to be overcome through multimodal data fusion. By combining multi-dimensional information such as images, sensor signals, and odors, the system can more comprehensively identify and analyze targets and improve accuracy and robustness. Therefore, multimodal data fusion will become the core direction to solve the current bottlenecks and enhance the overall performance and application value of intelligent detection systems.

3 Discussion

3.1 Summary of Technical Advantages

3.1.1 Performance improvement brought by multimodal data fusion.

Machine vision and sensor fusion technology have shown significant technical advantages in the field of intelligent detection. For example, in the field of autonomous vehicles, the application of multimodal data fusion has significantly improved the detection accuracy and robustness of the system. By fusing data from different sensors, such as cameras and laser radar (LiDAR), the system can obtain information from multiple dimensions. This fusion method makes up for the constraints of using a single sensor and can provide more comprehensive detection results in complex environments. In the study of Yahya Massoud [27], a deep convolutional neural network was combined with multimodal data streams from cameras and LiDAR, and innovative fusion mechanisms (such as element-wise multiplication and multimodal decomposition bilinear pooling) were used to improve the average mean accuracy score of the bird's-eye view. Specifically, the improvements on the KITTI dataset were +4.97% and +8.35%, respectively. These methods not only improve detection but also promote the research of multimodal machine vision systems and provide new ideas for real-time high-precision detection in complex scenes. This technology is particularly suitable for application scenarios that require high precision and real-time performance in many fields, such as high-precision manufacturing, food testing, and medical diagnosis, and can significantly improve the performance of the overall system.

3.1.2 Analysis of industrial applicability brought by fusion.

Machine vision and sensor fusion technology are widely applicable in the industrial field. Tsanousa et al. shows that with the advancement of Industry 4.0, manufacturing companies are becoming more and more "intelligent", and multi-sensor systems have become a key technology to improve production efficiency, reduce costs, and improve safety [28]. Multi-sensor systems can generate a large amount of heterogeneous data, and data fusion technology solves the multimodal problem by combining data from different sensors and significantly improves the performance of the monitoring system. Data fusion technology provides strong support for the industrial field. It can adapt to extreme environmental conditions, while also efficiently handling various types of detection tasks, from simple dimensional measurement to complex surface defect detection. This is consistent with Athina's research, which emphasizes how to improve the performance of the monitoring system through data fusion technology to support the real-time and automation requirements of intelligent manufacturing. The collaborative processing of multi-sensor data enables the system to respond quickly and make efficient decisions to meet the diverse needs of modern industrial production. The combination of these technologies not only optimizes production also brings higher safety and lower costs to the industrial field, promoting the development of intelligent manufacturing.

3.2 Future Research Directions

At present, machine vision and sensor fusion technology face challenges in intelligent detection, such as the complexity of multi-source data fusion, sensor performance degradation in extreme environments, and insufficient generalization ability of deep learning algorithms. Differences in the format, accuracy, and timestamps of multi-source data lead to fusion delays, affecting the real-time performance of the system; extreme environments (such as strong light and high noise) distort sensor data and reduce detection accuracy; deep learning algorithms have limited adaptability when processing small differences such as color and texture. To address these problems and promote technological development, future research could focus on the following areas.

Cross-domain expansion and application: At present, research is primarily concentrated on the field of industrial detection, but in the future, fusion technology can be expanded to more scenarios, such as garbage sorting, medical testing, and agricultural monitoring. For example, in garbage classification, vision and touch fusion technology can help robots achieve more accurate grasping; in medical testing, an image analysis system based on multi-sensor fusion can be developed to improve diagnostic efficiency by combining pathological data. Through cross-field expansion, fusion technology can demonstrate its unique advantages in more industries.

Adaptive optimization in extreme environments: For detection needs in extreme environments, future research can focus on improving the anti-interference ability of sensors and the robustness of algorithms. For example, new sensors that resist strong light and electromagnetic interference can be developed, or fusion algorithms that can still run stably in noisy environments can be designed. In addition, the system can also introduce an adaptive learning mechanism so that parameters can be automatically adjusted in different environments, thereby improving the accuracy and stability of detection.

Multidisciplinary cross-cutting and collaborative innovation: The development of machine vision and sensor fusion technology is inseparable from the cooperation of various disciplines. In the future, interdisciplinary research in the fields of computer science, materials science, mechanical engineering, etc., can be strengthened to develop new sensors and adaptive fusion algorithms. For example, combined with the progress of materials science, sensors that can work stably in extreme environments can be developed, or mechanical structures that can be seamlessly integrated with visual systems can be designed through innovations in mechanical engineering. Through multidisciplinary collaborative innovation, we can establish a solid foundation for the technological advancement of future intelligent manufacturing systems.

4 Conclusion

This study systematically explores the theory and application of machine vision and multi-sensor fusion technology in intelligent product inspection. The foundation of machine vision technology, multi-sensor information fusion technology, and its current application in intelligent inspection are studied. The study elucidates the advantages of multimodal data fusion in improving the accuracy and robustness of points and points out the challenges faced by the current technology and future cross-disciplinary research aspects.

This includes a combination of deep learning techniques and multi-sensor data synchronisation strategies by integrating machine vision algorithms with multi-source sensor data and algorithmic summaries. The algorithms in machine vision and sensor fusion are evaluated for real-time and accuracy in complex scenarios, in view of the shortcomings of traditional detection methods in terms of computational efficiency and accuracy. Cross-scenario validation and application expansion are also carried out to verify the universality of the fusion technology through real-life cases such as industrial production lines, the Internet of Medical Things, and food fermentation monitoring. Especially in extreme environments such as high noise and strong light interference, the system still maintains stable performance, providing theoretical support and technical guarantee for actual industrial deployment.

Although current research has achieved remarkable results, there is still a need to deepen exploration in multiple directions in the future. For the current lack of adaptability of sensors to extreme environments and other issues, such as strong electromagnetic interference, extreme temperature and humidity, and other complex industrial environments, it is necessary to further optimise the sensor's noise immunity and algorithm's generalisation performance. Furthermore, the cross-cooperation and co-innovation of computer science, materials science, and mechanical engineering should be strengthened to develop new sensors and adaptive fusion algorithms and lay the technological foundation for the next-generation intelligent manufacturing system.

In conclusion, machine vision and multi-sensor fusion technology are gradually becoming one of the core technologies in the field of intelligent detection. Future research needs to continue to break through the theoretical bottlenecks, expand the application boundaries, and provide stronger technical support for industrial construction and the construction of an intelligent society.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Yeong D. J., Velasco-Hernandez G, Barry J, et al.: ‘Sensor and sensor fusion technology in autonomous vehicles: a review’, Sensors, 2021, 21, (6), 2140
2. Naqvi R A, Arsalan M, Qaiser T, et al.: ‘Sensor data fusion based on deep learning for computer vision applications and medical applications’, Sensors, 2022, 22, (20), 8058
3. Wang T, Chen B, Zhang Z, et al.: ‘Applications of machine vision in agricultural robot navigation: a review’, Computers and Electronics in Agriculture, 2022, 198, 107085
4. Liu Y, Guo L, Gao H, et al.: ‘Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture: A review’, Mechanical Systems and Signal Processing, 2022, 164, 108068
5. Golnabi, H.; Asadpour, A.: ‘Design and application of industrial machine vision system’, Robotics and Computer-Integrated Manufacturing, 2007, 23, pp. 630 637
6. Gonzalez, R. C., & Woods, R. E.: ‘Digital Image Processing’, (Prentice Hall, 3rd ed. 2008)
7. Lowe, D. G.: ‘Distinctive Image Features from Scale-Invariant Keypoints’, International Journal of Computer Vision, 2004, 60, 91-110
8. Krizhevsky A, Sutskever I, and E. Hinton G.: ‘ImageNet classification with deep convolutional neural networks’, June 2017. Commun. ACM 60, 6, 84-90
9. Girshick, R.: ‘Fast R-CNN’ 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448
10. Redmon, J., Divvala, S., Girshick, R.:‘You only look once: Unified, real-time object detection’ Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp.779-788
11. Mnih, V., Heess, N., Graves, A.: ‘Recurrent models of visual attention advances in neural information processing systems’. Proceedings. of Neural Information Processing Systems (NIPS),2014,2
12. Hu, J., Shen, L., Sun, G.:‘Squeeze-and-excitation networks’ Proceedings of the IEEE conference on computer vision and pattern recognition.2018, pp.7132-7141
13. Wang, X., Girshick, R., Gupta, A.: ‘Non-local neural networks’ Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp.7794-7803
14. Adarsh, S., Kaleemuddin, S, M., Bose, D.: ‘Performance comparison of Infrared and Ultrasonic sensors for obstacles of different materials in vehicle/robot navigation applications’.IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2016, 149(1): 012141.
15. Bosse, E., Roy, J., Grenier, D.: ‘Data fusion concepts applied to a suite of dissimilar sensors’, Proceedings of 1996 Canadian conference on electrical and computer engineering. Calgary, Canada, May,1996, pp. 692-695

16. El-Dalahmeh, A., El-Dalahmeh, M., Li, J.: ‘Enhanced Vehicle Detection through Multi-Sensor Fusion Utilizing YOLO-NAS and Faster R-CNN’, 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, June, 2024, pp.1-6
17. Zhou, T., Jiang, K., Xiao, Z.: ‘Object detection using multi-sensor fusion based on deep learning’, CICTP,2019. 2019, pp. 5770-5782
18. Jaderberg, M., Simonyan, K., Zisserman, A.: ‘Spatial transformer networks’, Advances in neural information processing systems, 2015, 28.
19. Luo, R, C., Chang, C, C.: ‘Multisensor fusion and integration: A review on approaches and its applications in mechatronics’. IEEE Transactions on Industrial Informatics, 2011, 8, (1), pp. 49-60
20. Ciuffreda, I., Casaccia, S., Revel, G, M.: ‘A multi-sensor fusion approach based on pir and ultrasonic sensors installed on a robot to localise people in indoor environments’ Sensors, 2023, 23(15): 6963.
21. Xie, T., Huang, X., Choi, S, K.: ‘Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network’, IEEE Transactions on Industrial Informatics, 2021, 18, (5), pp.3213-3223
22. Chen, G., Liu, Z., Yu, G.: ‘A new view of multisensor data fusion: research on generalised fusion’, Mathematical Problems in Engineering, 2021, (1),5471242.
23. Benbarrad, T., Salhaoui, M., Kenitar, S, B.: ‘Intelligent machine vision model for defective product inspection based on machine learning’, Journal of Sensor and Actuator Networks, 2021, 10, (1),7
24. Wang, Y., Cai, Z., Huang, T.: ‘An intelligent manufacturing management system for enhancing production in small-scale industries’, Electronics, 2024, 13(13), 2633
25. Ding, X., Zhang, Y., Li, J.: ‘A feasibility study of multi-mode intelligent fusion medical data transmission technology of industrial Internet of Things combined with medical Internet of Things’, Internet of Things, 2023,21,100689
26. Zhou, Q., Dai, Z., Song, F.: ‘Monitoring black tea fermentation quality by intelligent sensors: Comparison of image, e-nose and data fusion’, Food Bioscience, 2023,52,102454
27. Massoud, Y., Laganiere, R.: ‘Learnable fusion mechanisms for multimodal object detection in autonomous vehicles’, IET Computer Vision, 2024, 18, (4), pp. 499-511
28. Tsanousa, A., Bektsis, E., Kyriakopoulos, C.: ‘A review of multisensor data fusion solutions in smart manufacturing: Systems and trends’, Sensors, 2022, 22, (5), 1734