Research on the application and future development of visual recognition in modern agriculture

Yuke Li[1], Boru Pan[2,*], Yuhang Wan[3]

[1] School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an City, Shaanxi Province, 710000, China
[2] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, Guangdong Province, 510641, China
[3] European Institute, Yanshan University, Qinhuangdao City, Hebei Province, 066000, China

* Corresponding author: 202130132488@mail.scut.edu.cn

Abstract. Population ageing has become a challenge faced by most developing countries in recent years, and for developing countries where agriculture is the supporting industry, the need for a large number of labourers to carry out agricultural operations is an inevitable requirement for the long-term stable development of their agriculture. Agricultural robots based on visual recognition technology will greatly reduce the impact of the aging population on agricultural production, and at the same time can improve the efficiency and quality of agricultural production. The application and development trend of visual recognition technology in modern agriculture is examined in a comprehensive overview and analysis provided in this paper. Firstly, the application of visual recognition in modern agriculture is studied and analyzed around computer vision and related visual recognition algorithms. After that, the main problems in its application are discussed. And finally, an outlook on the future development trend of visual recognition in agriculture is presented in order to provide a reference for the future in-depth application of visual recognition in agriculture.

Keywords: visual recognition, modern agriculture, industrial upgrading, agricultural equipment

1 Introduction

Nowadays, population development in most developing countries has entered an ageing stage, and for developing countries whose main focus is on agricultural development, their agricultural workforce will undoubtedly be affected by the ageing of the population. However, the need for intense physical labor in traditional agricultural machine operations has not diminished, which makes long-term stabilization of agriculture challenged.

From the engineering aspect, we can solve this problem by developing high-efficiency and high-quality agricultural automation robots. Such that all links in

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_4

agricultural production are changed from being dominated by humans to being dominated by human-controlled robots. This reduces labor inputs and increases productivity at the same time. Among all the key technologies, visual recognition technology can not only provide technical support for agricultural mechanized production but also control the quality of agricultural products safely and efficiently. In the marketing of agricultural products, information-integrated visual recognition technology can protect the overall quality and commodity value of agricultural products.

Visual recognition early in the field of agriculture is mainly studied in the agricultural robotics-related technology, economic feasibility, potential, prospects, and other macro issues. In recent years, visual recognition technology has been applied more to crops that are difficult to grow and harvest. Visual recognition goes deeper in serving agricultural production. Therefore, the development of visual recognition technology in agriculture will become a crucial topic. The crucial role of visual recognition technology in agriculture is inextricably linked to its practicality and reliability. For example, visual recognition of plant diseases has played a great role in plant disease diagnosis, resulting in a significant increase in crop yields [1].

It can be seen that visual recognition has crucial, practical, and reliable characteristics in the field of agriculture, and it is of practical significance to research and analyze its application in agriculture at this stage and prospect the development trend. Therefore, this paper firstly focuses on computer vision and related visual recognition algorithms to research and analyze the usage of visual recognition in modern agriculture. Secondly, the main problems of the usage of visual recognition in agricultural production are discussed. Finally, an outlook on the future development trend of visual recognition in agriculture is proposed, with a view to providing references for the in-depth application of visual recognition in agriculture in the future.

2 Technical architecture

2.1 Computer vision

As an emerging inspection method, visual recognition technology has found an increasingly wide utilization in modern agriculture. By employing an image sensor that replaces the visual function of human, this technology has the capability to capture image information of the object and convert it into a data matrix. After that, the data is processed and identified computationally, with the option to magnify key information according to needs [2]. This ensures the precise analysis of target features and the extraction of effective information, making it possible to control the growing environment and state of agricultural crops.

2.2 Components of a computer vision system

A computer vision system is mainly composed of visual sensors, lighting units, an image acquisition system, a specialized image processing system, and other modules.

The visual sensor is a crucial device for capturing images. It can be classified based on the type of light-sensitive devices, such as CMOS and CCD sensors. Additionally, the captured light's wavelength can range from near-infrared to visible, ultraviolet, hyperspectral, and other categories. The lighting unit varies greatly in different environments when used. And for the visual recognition system in modern agriculture, the natural light has a significant impact on the lighting unit. Therefore, the selection of appropriate light sources is indispensable for making visual recognition in modern agriculture a success.

2.3 Categorization of computer vision systems

Monocular vision technology: As the foundation of other vision technologies, monocular vision technology refers to capturing images through a single camera. It is economical and is simple in structure. The advantage of monocular vision technology is that localization can be accomplished with only a single image, without the requirement for intricate image matching. Due to its simple structure, advanced algorithms, and minimal computation, monocular vision technology has enjoyed widespread popularity among intelligent robotics, specifically for target ranging and indoor localization with regard to monocular features. However, monocular vision technology has limitations when it comes to complex perceptual needs because it perceives a single channel of information. Therefore, in most cases, monocular cameras must collaborate with other sensors such as ultrasound and infrared.

Binocular stereo vision technology: Binocular stereo vision is a significant type of machine vision that relies on the parallax principle. This technology enhances the accuracy of visual recognition, but it requires a higher level of computational power. It captures two images of the tested object from distinct positions by using imaging equipment and calculates the positional deviation between corresponding points of the pictures to obtain its three-dimensional geometric information. When the binocular stereo vision technology is applied, the spatial position relationship between two cameras must be known accurately. At the same time, to gain access to more accurate three-dimensional information of the visual scene, two cameras at different locations must simultaneously capture the same scene and then perform a complex matching of the resulting two images. Binocular stereo vision technology has been extensively used in mobile robot localization, obstacle avoidance, and map construction. However, the main issue with binocular stereo vision technology is the matching of corresponding points, which significantly restricts its future development in robotics.

Hybrid vision technology: Based on monocular vision technology, hybrid vision technology integrates additional sensing technologies such as sound, temperature, and infrared to enhance the precision and stability of images. It is efficient and has a wide application. For instance, in the automatic driving detection system, monocular vision technology is combined with LiDAR to accurately locate the target and identify its image features, which enhances the reliability of the detection process. In certain inspection scenarios, color cameras and long-wave infrared cameras are combined to perceive dimensions, abnormal responses, high temperatures, and other conditions. The sensing capabilities of smart devices will be enhanced by combining various sensing sensors with vision technology. This approach is tailored to the needs of specific application scenarios.

2.4 Visual recognition technology

Visual recognition technology refers to using cameras and computers to identify, track and measure objects instead of relying on the human eye. Then the target graphics are turned into images that are appropriate to observe or to transmit to an equipment for detection. Figure 1 shows the process of how the visual recognition technology is implemented.

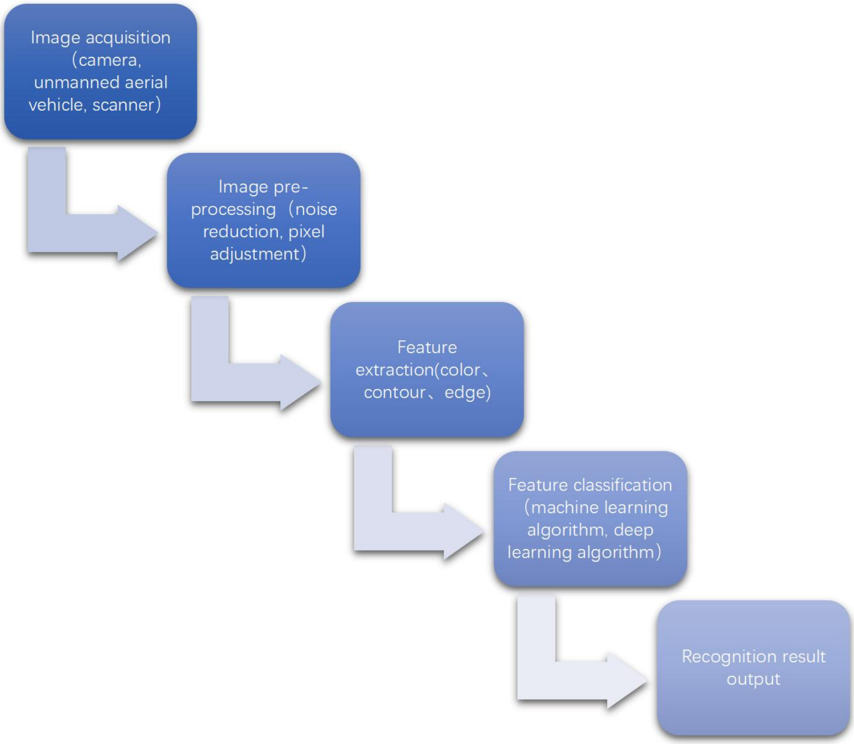

Fig. 1. The process of visual recognition technology. (Photo/Picture credit :Original)

The objective of computer vision recognition is to extract object-related information from two-dimensional images captured by the camera and then categorize it. The process of visual recognition technology is mainly composed of five steps: image acquisition, image preprocessing, feature extraction, feature classification, and recognition result output.

Image acquisition: The target image is captured through various devices and then converted into a digital format suitable for interpretation and processing via computer.

Image pre-processing: To ensure the quality of the captured images for subsequent processing, image pre-processing is imperative. Pre-processing includes noise reduction, image quality enhancement, geometric distortion correction, pixel adjustment, and so on.

Feature extraction: Feature extraction refers to extracting the representative features from an image through various image processing techniques, such as color extraction, contour extraction, edge extraction, and so on. The algorithms of extraction include Scale-Invariant Feature Transform, Speeded Up Robust Features, and Histogram of Oriented Gradient. To enhance the recognition rate of diverse targets, feature extraction must be accompanied by both localization and invariance of image features.

Feature classification: After the feature extraction, the computer matches and compares the extracted features with the pre-trained model for classification, this process is called feature classification. The algorithms of classification include the machine learning algorithm (Support Vector Machine, Random Forest) and the deep learning algorithm (Convolutional Neural Networks). During the classification, to ensure the accuracy and reliability of the identification outcomes, evaluating and verifying them is necessary.

Recognition result output: Based on the results of feature classification, output recognition outcomes for visual recognition.

3 Visual recognition technology for modern agriculture

With the development of visual recognition technology, both traditional recognition algorithms and deep learning-based recognition systems have found widespread use in a variety of fields of modern agricultural production, and have been continuously improved and promoted.

3.1 Fruit picking recognition

Visual recognition technology is widely used in agricultural picking robots to identify and distinguish fruits for automatic picking and harvesting.

In 2019, Yu et al. used the Mask-RCNN algorithm to identify mature strawberries. By using the feature pyramid (FPN) network for feature extraction, the merging of multiple-scale features of color, shape, and texture was realized, and the precision rate reached 95.78 % [3].

In 2021, algorithm YOLOv4 was enhanced by Zhang Qin et al. to recognize the picking points of tomato bunches, combining the YOLOv4 algorithm with the RGB-D depth image information and using traditional algorithms such as depth segmentation and clustering algorithms, to realize the identification and localization of tomato picking points, and the recognition success reached 93.83% [4]. To recognize apples, Yan et al. suggested an enhanced YOLOv5 method which can identify apples that can be picked and apples that cannot be picked due to branch occlusion. Its mAP reaches 86.75 %, which outperforms the original YOLOv5 algorithm by 5.05% [5].

In 2022, Xu et al. proposed an improved Mask-RCNN algorithm to identify cherry tomatoes, combining RGB and depth information, and overcoming the constraints brought on by the disparity in fruit and stem size by multi-task loss equalization and adaptive feature pooling. The recognition accuracy of the fruit reached 93.76 %, which was 11.53 % higher than the standard Mask-RCNN algorithm [6]. The precision of the algorithm keeps getting better as it is developed and improved, but it also faces problems such as too many parameters, large models that are difficult to embed, and poor real-time performance. Therefore, the lightweight of network model is a key research direction.

In 2023, Cong et al. proposed a lightweight YOLOv3 algorithm to detect shiitake mushrooms, employing the lightweight GhostNet16 backbone network as opposed to DarkNet53. This model's mAP is 97.03%, which is 2.04 % more than the original model's, and the parameter is 29.8 M, which is 2.08 times lower than that of the initial design. This offers a crucial theoretical foundation for robotically picking fresh mushrooms [7]. An enhanced YOLOv5s model was proposed by Zhang et al. to identify peppers. The size of the new model was 46.6% less than the old model, and the identification speed was increased from 29 milliseconds per frame to 14 milliseconds per frame. At the cost of reducing mAP by 1.3 %, the model running speed is greatly improved [8]. Wang et al. employed the YOLOv7-tiny model to realize the recognition of dragon fruit with different postures. The classification accuracy of dragon fruit from positive and lateral perspectives was 80.4 %. Compared with EfficientDet, SSD, Faster-RCNN, and CenterNet algorithms, it has significant advantages. It provides an effective reference for the identification of multi-pose crops [9].

3.2 Growth process detection

In addition to the recognition of fruits by picking robots, visual recognition is also used in the detection of crop growth process to achieve timely regulation of crop growth environment, disease control or timely picking according to different growth conditions of crops, so as to ensure high yield and high quality of crops.

In 2018, Long et al. used Alex Net and transfer learning methods to learn and identify the disease characteristics of Camellia oleifera. Through transfer learning, the model's categorization capability and convergence speed both increased, and the over-fitting phenomenon was avoided through data expansion. The classification's rate of accuracy reached 96.53 %, and the average F1 of the five common diseases reached 96.5 % [10].

In 2021, Mask-RCNN model was employed by Rehman et al. to identify apple leaf diseases with an optimal accuracy of 96.6% by stretching the image contrast enhancement [11]. Lu et al. used the method of combining YOLOv3 algorithm and manual features to identify the maturity of winter jujube, and mAP reached 94.78 %. The maturity classification accuracy reached 97.28 %, which can accurately identify the maturity of winter jujube and meet the requirements of timely picking [12].

In 2022, Chen et al. improved the EfficientDet algorithm to detect the ripeness of olive fruit trees by introducing the convolutional attention module (CMBA), which solved the problem that the difference between the characteristics of neighboring ripe fruits was not obvious and difficult to identify, and, in the test, the recall rate was 93.59% and the precision rate was 92.89%. [13]. Wen et al. proposed the AD-YOLOv3 algorithm to detect the disease of Panax notoginseng leaves, and replaced the original feature pyramid of YOLOv3 with the attention feature pyramid (AEP), which solves the interference issue with feature fusion. Compared with YOLOv3, the new algorithm's accuracy has increased by 2.83%, and the accuracy of F1 is improved by 1.68 %, which provides a better recognition algorithm for pest detection [14]. Fan et al. implemented ripeness grading of strawberries using YOLOv5 combined with dark channel enhancement with training accuracy above 85% and testing accuracy above 90% [15].

In 2023, Song et al. proposed an improved YOLOv7 algorithm to recognize and distinguish apples at the young fruit stage, inserting a fusion efficient channel attention (ECA) mechanism to solve the problem of small variability of fruit and leaf colors, tiny size and dense distribution, which makes it difficult to recognize. The mAP reaches 98.6% and F1 reaches 94.8% under severe occlusion, realizing better robustness to the effects of scene blurring and severe occlusion with guaranteed accuracy [16].

3.3 Agricultural machinery assistance

Vision inspection technology can also assist agricultural machinery in path planning, weeding, precise irrigation, and other operations to achieve higher operational efficiency.

In 2020, Guan et al. used a threshold algorithm to identify the area boundaries with the help of the color difference between before and after harvesting of the rice field, and provided a target path for the rice harvester for subsequent operations, which is expected to improve the operation efficiency [17]. In 2021, Chang et al. used the YOLOv3 model to realize precise irrigation of orchid seedlings, with watering success rates of 82% and 83.3% for single and multiple pots, respectively. It can effectively reduce the waste of water resources [18]. Xu et al. proposed an improved Xception model to identify weeds, and downsampled the final convolution layer using global maximum pooling layer. The average test recognition accuracy of 8 types of weeds and seedling corn under natural conditions reached 98.63%, which provided technical support for accurate weeding [19].

4 Discussion

4.1 Limitation analysis

At present, although the visual recognition technology is developing rapidly, it still faces problems such as the natural environment affecting the quality of image acquisition, the characteristics of crop growth increasing the difficulty of visual recognition, the existing algorithms are difficult to balance the accuracy, the recognition speed and lightweight, and the over-fitting phenomenon still exists.

First of all, the variations in the natural environment's wind and lighting negatively impact the speed and accuracy of visual recognition as well as image acquisition. Secondly, the crops are generally dense, the stems and leaves are intertwined, and the fruits are superimposed on each other, so that the recognition object faces various occlusions. At the same time, the complexity of crop growth posture and growth status distribution will also affect the accuracy of visual recognition. Then, with traditional recognition algorithms often struggling to meet recognition needs in complex natural environments, visual recognition based on deep learning began to be applied in large numbers. However, with the continuous improvement of various improved algorithms, many algorithms have begun to face the problem that the model is huge, which makes it difficult to deploy to embedded devices and has poor real-time performance. Lightweight has become a new practical demand. Finally, when the training set samples are too few, the noise is too large or the number of weight learning iterations is too many, the over-fitting phenomenon will occur during the test, which puts forward higher requirements for model training.

4.2 Future development

Intelligentize: In the future, agriculture is bound to benefit from the use of robots equipped with visual recognition technology, which can provide high-resolution images for agricultural production. If the robots are integrated with real-time data on crops and soil collected through sensors, they will be capable of monitoring crop growth and predicting pests and diseases accurately. Meanwhile, the robots can provide smart management and strategic decision for agriculture by analyzing crop growth, soil moisture, weather changes, and other factors. The combination of visual recognition technology and machine learning algorithm makes it possible to predict and diagnose the crops' growth status, pests, and diseases by using a large amount of historical data, which will help farmers better manage their crops.

Efficient: The implementation of visual recognition technology will greatly improve the efficiency and quality of agricultural production. Agricultural robots utilize a vision system to recognize and classify crops. They can take over various agricultural tasks in farm fields such as planting, spraying pesticides, and harvesting, which were previously performed by humans. This change will significantly alleviate farmers' labor burden and enhance the efficiency and precision of operations, making agricultural production more efficient.

Environmental: Visual recognition technology will support the implementation of precision agriculture and enhance the capacity to personalize the management of crop-specific demands for certain crops. In the future, farmers will be able to provide optimal irrigation, fertilization, spraying, and other treatments for their crops based on specific requirements. It means that visual recognition technology will not only enhance the productivity and quality of crops, but also minimize wastage of resources and mitigate adverse environmental effects.

5 Conclusion

The application and development trend of visual recognition technology in modern agriculture is examined in a comprehensive overview and analysis provided in this paper. In this paper, on the basis of introducing the system composition and classification of computer vision technology, the application of visual recognition technology in modern agriculture is researched through three application areas, namely, fruit picking recognition, growth process detection, and agricultural machinery assistance. Among them, there are more applications in recognition of fruit picking, and recognition of various types of crop fruits has been able to achieve a high accuracy rate. Applications in the field of growth process detection are mainly centered on detecting pests and diseases and identifying crop growth cycles, which are promising. The application in the field of agricultural machinery assistance is less at present, mainly focusing on helping agricultural machinery to realize precise spraying, weeding, and planning the operation path, and there is still more space for exploration. Although visual recognition technology has been applied in the above fields, it still faces problems such as the natural environment affecting the quality of image acquisition, the growth characteristics of crops increasing the difficulty of visual recognition, the difficulty of existing algorithms to take into account the accuracy, recognition speed, and lightweight, and the existence of overfitting phenomenon. Solving these problems will be the key to technological improvements in visual recognition in agriculture. In the future, visual recognition will make agricultural production present intelligent, efficient, and environmentally friendly development trend, in the field of agriculture still has a greater potential for development.

Authors contribution

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

References

1. Liu, Xinda, et al. "Plant disease recognition: A large-scale benchmark dataset and a visual region and loss reweighting approach." IEEE Transactions on Image Processing 30 (2021): 2003-2015.
2. Bagga, P. J., et al. "Tool Wear Detection Using Computer Vision System in Machining." Advances in Manufacturing Technology and Management: Proceedings of 6th International Conference on Advanced Production and Industrial Engineering (ICAPIE)—2021. Singapore: Springer Nature Singapore, 2022.
3. Yu, Yang, et al. "Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN." Computers and Electronics in Agriculture 163 (2019): 104846.

4. Zhang, Qin, et al. "Method for recognizing and locating tomato cluster picking points based on RGB-D information fusion and target detection." Trans. Chin. Soc. Agric. Eng.(Transactions CSAE) 37.18 (2021): 143-152.
5. Yan, Bin, et al. "A real-time apple targets detection method for picking robot based on improved YOLOv5." Remote Sensing 13.9 (2021): 1619.
6. Xu, Penghui, et al. "Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation." Computers and Electronics in Agriculture 197 (2022): 106991.
7. Cong, Peichao, et al. "MYOLO: A Lightweight Fresh Shiitake Mushroom Detection Model Based on YOLOv3." Agriculture 13.2 (2023): 392.
8. Zhang, Song, and Mingshan Xie. "Real-Time Recognition and Localization Based on Improved YOLOv5s for Robot's Picking Clustered Fruits of Chilies." Sensors 23.7 (2023): 3408.
9. Wang Jinpeng, et al. "Multi pose Huolongguo Detection System for Picking Robots Based on the Optimal YOLOv7 Model." Transactions of the Chinese Society of Agricultural Engineering 39.8 (2023).
10. Long, Mansheng, et al. "Image recognition of Camellia oleifera diseases based on convolutional neural network & transfer learning." Transactions of the Chinese Society of Agricultural Engineering 34.18 (2018): 194-201.
11. Rehman, Zia Ur, et al. "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture." IET Image Processing 15.10 (2021): 2157-2168.
12. Lu, Zhiheng, et al. "Design of a winter-jujube grading robot based on machine vision." Computers and Electronics in Agriculture 186 (2021): 106170.
13. Chen, F., et al. "Detection of the olive fruit maturity based on improved EfficientDet." Trans. Chin. Soc. Agric. Eng.(Trans. CSAE) 38 (2022): 158-166.
14. Othman, Suhaili, et al. "Journal of Agriculture and Food Research." Journal of Agriculture and Food Research 12 (2023): 100590.
15. Fan, Youchen, et al. "Strawberry maturity recognition algorithm combining dark channel enhancement and YOLOv5." Sensors 22.2 (2022): 419.
16. Song Huaibo, Ma Baoling, Shang Yuying, et al. "Detection of Apple Young Fruits Based on YOLO v7-ECA Model." Journal of Agricultural Machinery, 50.6 (2023):233-242.
17. Zhuohuai, G. U. A. N., et al. "Visual navigation path extraction method in rice harvesting." Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery 51.1 (2020).
18. Chang, Ting-Wei, Wei-Cheng Wang, and Rongshun Chen. "Intelligent Control System to Irrigate Orchids Based on Visual Recognition and 3D Positioning." Applied Sciences 11.10 (2021): 4531.
19. Xu Yanlei, et al. "Field Natural Environment Weed Identification Method Based on Lightweight Convolutional Networks." Journal of Jilin University (Engineering Edition) 51.06 (2021): 2304-2312.