Automation and Machine Learning (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2025.060205
ISSN 2516-5003 Vol. 6 Num. 1

Review on Commodity Recognition and Inventory Counting Based on Machine Vision in Retail Scenarios

Ranning Deng

School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China
2840820473@qq.com

Keywords: Machine Vision; Commodity Recognition; Inventory Counting; Retail Scenario; Unmanned Retail; Grocery Dataset; Retail Product Checkout

Abstract: With the in-depth advancement of the digital transformation of the retail industry, machine vision technology has become a key support for optimizing the operational efficiency of supermarkets and convenience stores. This paper focuses on typical retail scenarios such as supermarkets and convenience stores, and systematically sorts out the application status of machine vision technology in three core links: commodity classification, shelf out-of-stock detection, and self-checkout recognition. It analyzes the technical characteristics of traditional manual feature extraction methods and modern deep learning methods in commodity recognition, and emphasizes the application value of the Grocery Dataset in algorithm training. Aiming at the key technical bottlenecks such as similar commodity packaging, stacked placement occlusion, and complex environmental interference, this paper summarizes the optimization strategies such as attention mechanism, multi-view fusion, and multi-modal combination, and discusses the practical experience of Retail Product Checkout (RPC) technology in complex shopping basket settlement scenarios. Finally, the paper analyzes the current challenges faced by unmanned retail in terms of environmental adaptability, computing power cost, and privacy protection, and puts forward development directions such as multi-technology integration, large-scale deployment, and human-machine collaboration. This review provides a comprehensive technical reference for the in-depth application of machine vision in the retail field.

1. Introduction

With the rapid development of new retail models, the demand for intelligent management in retail scenarios such as supermarkets and convenience stores has become increasingly urgent. Machine vision technology, by simulating human visual perception capabilities and integrating image processing, deep learning, and other algorithms, realizes accurate commodity recognition, inventory status monitoring, and shopping process automation, and has become one of the core technologies promoting the digital transformation of the retail industry. This paper focuses on the scenarios of supermarkets and convenience stores, systematically sorts out the key applications of machine vision in commodity classification, shelf out-of-stock detection, and self-checkout recognition, analyzes typical technical challenges combined with Retail Product Checkout technology and the Grocery Dataset, and looks forward to the implementation path of unmanned retail scenarios.

2. Application of Machine Vision in Commodity Recognition and Classification

2.1 Commodity Recognition Technology System

The core of machine vision-based commodity recognition is to collect commodity images through cameras, and realize category determination through preprocessing, feature extraction, and pattern matching. Traditional methods rely on manually designed features (such as SIFT, HOG) combined with classifiers like Support Vector Machine (SVM), which are suitable for commodities with obvious packaging differences but lack robustness in the face of similar packaging or complex lighting environments. In recent years, deep learning technologies (such as CNN) as well as a series of object detection algorithms (such as SSD[5]) have significantly improved recognition accuracy by automatically learning multi-level features of commodities, such as color, texture, and shape, through end-to-end training. For example, a ResNet-based model achieves an accuracy rate of 95.2% in classifying 120 types of commodities on the Grocery Dataset, which contains real shopping basket data of 7501 transactions, covering high-frequency daily necessities and long-tail commodities, providing rich scenarios for algorithm training.

2.2 Recognition Challenges of Commodities with Similar Packaging

On supermarket shelves, similar commodities from the same category often cause recognition confusion due to similarities in brand, specification, and packaging design, such as beverage bottles of different flavors and series of toiletries. In such cases, algorithms need to capture subtle differential features, such as label text layout, LOGO color gradient, and packaging size ratio. Studies have shown that introducing the Attention Mechanism[3] can guide the model to focus on key areas, and combining with Metric Learning to optimize the feature space distance can expand the inter-class interval of similar commodities by more than 30%. In addition, multi-view image fusion technology collects front, side, and top images of commodities to construct 3D feature vectors, which effectively solves the problem of insufficient information from a single perspective, especially suitable for partial occlusion scenarios during stacked placement.

2.3 Commodity Classification and Inventory Initialization Management

In the commodity shelving process, the machine vision system can automatically scan the incoming commodities, generate electronic files containing information such as category, specification, and shelf life, and connect to the inventory management system. For example, a visual classification system deployed in a convenience store uses dual verification of commodity barcodes and appearance features, increasing the warehousing efficiency by 40% and reducing the error rate to below 0.3%. For unbarcoded commodities (such as bulk fruits and fresh produce), Optical Character Recognition (OCR) technology is used to identify price tags, or pre-trained models are used to learn the natural attributes of commodities (such as fruit shape and color distribution) for classification, laying a data foundation for subsequent inventory counting.

3. Shelf Out-of-Stock Detection and Dynamic Inventory Monitoring

3.1 Real-Time Perception Technology of Shelf Status

The core of shelf out-of-stock detection is to use visual algorithms to determine whether commodities are missing or improperly displayed. Traditional methods are based on background subtraction, which identifies out-of-stock areas by comparing the current shelf image with the standard display image, but they are sensitive to interference such as lighting changes and commodity displacement. Modern solutions adopt object detection algorithms (such as YOLO[1], Faster R-CNN[2]) to detect and count commodity instances on the shelf in real time. For example, in the pilot system of 7-Eleven convenience stores, the YOLOv5-based detection model achieves a positioning accuracy of 92.7% for shelf commodities, which can complete a single shelf scan within 200ms and support real-time monitoring of 10 frames per second.

3.2 Detection Optimization for Stacked Placement Scenarios

Supermarket commodities are often placed in multi-layer stacks and multi-row displays, resulting in bottom commodities being occluded and top commodities being deformed due to inclined viewing angles, which increases the difficulty of detection. The solutions include: (1) Deploying a multi-camera array to collect shelf images from different angles and restore the complete commodity contour through image fusion technology; (2) Introducing 3D vision technologies (such as structured light, TOF cameras) to obtain depth information, and distinguishing the spatial positions of occluded commodities combined with point cloud processing algorithms; (3) Using prior knowledge of commodity display to establish a hierarchical spatial model of "shelf - shelf board - commodity", and determining the out-of-stock situation by detecting the commodity occupancy status on the shelf board. Experimental data shows that the multi-modal fusion scheme can increase the detection accuracy of stacked commodities by 18% and reduce the missed detection rate to below 1.5%. In addition, algorithms for dense object detection such as BorderDet[7] further improve the detection performance in occluded scenarios through boundary feature optimization.

3.3 Dynamic Inventory Analysis and Replenishment Decision-Making

Based on the shelf detection results, the system can update the inventory quantity in real time, and analyze the commodity co-occurrence pattern through association rule mining (such as Apriori algorithm) combined with historical sales data such as the Grocery Dataset to optimize the replenishment strategy. For example, it is found that the combined purchase frequency of "mineral water + chocolate" reaches 2.3%. When one of the commodities is out of stock on the shelf, the system automatically triggers a replenishment reminder and adjusts the display position of adjacent commodities to improve associated sales. After applying this technology, a certain supermarket reduced the out-of-stock rate of best-selling commodities from 8% to 2% and increased the inventory turnover rate by 15%.

4. Self-Checkout Recognition Technology and Shopping Process Automation

4.1 Core Recognition Schemes of Self-Checkout Systems

The self-checkout scenario requires fast and accurate recognition of commodities selected by users, and the main technical paths include:

1) Barcode/QR Code Recognition: High-speed cameras scan commodity barcodes to retrieve commodity prices and inventory information in real time. Machine vision technology can process inclined and stained barcodes with a recognition speed of 5 times per second and an accuracy rate of over 99.9%, which is a basic configuration in mainstream convenience stores currently.

2) Visual Recognition Checkout: Realizing checkout directly through commodity appearance recognition without relying on barcodes. For example, Amazon Go's "Just Walk Out" technology uses shelf cameras and deep learning models to track users' commodity picking and placing behaviors in real time, and automatically settles the payment when users leave the store. This scheme has strong capabilities in handling scenarios such as simultaneous picking of multiple commodities and occlusion overlap, but it requires intensive camera deployment and solves the computing power bottleneck.

3) Multi-Modal Fusion Recognition: Verifying commodity recognition results by combining visual images and weight sensor data. For example, a self-checkout counter uses weighing to verify the number of commodities on the basis of visual recognition, reducing the misrecognition rate from 0.8% to below 0.1%.

4.2 Practical Application of Retail Product Checkout Technology

Retail Product Checkout (RPC) technology focuses on fast checkout in complex shopping basket scenarios, and its core challenge is to solve the recognition problems of commodity overlap, occlusion, and different placement postures. Studies have shown that the visual model based on Transformer (such as DETR[4]) improves the bounding box positioning accuracy of overlapping commodities by 22% compared with traditional CNN in handling multi-object detection tasks. In addition, open-source detection toolkits such as MMDetection[8] provide rich algorithm support for the rapid implementation of RPC technology. Combining weakly supervised learning technology and using the corresponding relationship between transaction data and images in the Grocery Dataset can train efficient recognition models with a 50% reduction in labeling costs, which is suitable for system deployment in small and medium-sized retailers.

4.3 Anti-Fraud and User Experience Optimization

The problems of "missing scan" and "wrong scan" in self-checkout need to be solved by technical means. The machine vision system can identify abnormal operations through behavioral analysis, such as long-term occlusion of the scanning area and frequent commodity position replacement, and trigger manual review alerts. At the same time, it optimizes the interactive interface based on face detection and gesture recognition technology. For example, users can switch commodity categories by waving their hands, which improves checkout efficiency and convenience.

5. Technical Bottlenecks and Implementation Prospects of Unmanned Retail Scenarios

5.1 Analysis of Technical Challenges

1) Complex Environment Adaptability: Strong lighting, specular reflection, and dynamic passenger flow in supermarkets lead to unstable image quality, requiring the development of more robust preprocessing algorithms and lighting compensation models.

2) Balance between Computing Power and Cost: End-to-end deep learning models have high computing power requirements for edge computing devices, and the accuracy and speed optimization of lightweight models (such as MobileNet[6], Tiny YOLO) remains a research focus.

3) Privacy Protection and Data Security: User images collected by cameras involve privacy,

requiring the use of technologies such as federated learning and differential privacy to complete data processing locally and avoid uploading sensitive information.

5.2 Future Development Directions

1) Intelligent Retail Ecosystem Integrating Multiple Technologies: Machine vision is deeply integrated with RFID, sensor networks, and the Internet of Things (IoT) to build a fully perceived retail environment. For example, shelf pressure sensors combined with visual data can more accurately determine the picking and placing status of commodities; electronic price tags are linked with visual recognition to update commodity prices and promotional information in real time.

2) Large-Scale Implementation of Unmanned Retail: With the decrease in the cost of edge computing devices and the popularization of 5G networks, the deployment cost of small convenience stores and unmanned shelves is expected to be reduced by more than 30%. The improvement of public datasets such as the Grocery Dataset will accelerate algorithm iteration and promote the technology from pilot to large-scale application.

3) Hybrid Management Model of Human-Machine Collaboration: The machine vision system undertakes tasks with high repeatability and standardization (such as inventory counting and checkout recognition), while humans focus on personalized services (such as commodity recommendation and complaint handling), forming a collaborative system of "efficient machine execution + human flexible decision-making".

6. Conclusion

Machine vision technology has reshaped the operation models of supermarkets and convenience stores through core functions such as commodity recognition, shelf monitoring, and self-checkout, significantly improving retail efficiency and user experience. Although it faces challenges such as similar packaging recognition, stacked occlusion detection, and computing power costs, with the optimization of deep learning algorithms, the upgrading of hardware equipment, and the enrichment of industry datasets, the technical implementation of unmanned retail scenarios has entered an acceleration period. In the future, machine vision will be integrated with more emerging technologies to promote the retail industry from "informatization" to "intelligentization" and "unmanned operation", bringing more convenient and efficient shopping experiences to consumers.

References

[1] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arxiv preprint arxiv:1804.02767 (2018).

[2] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.

[3] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[4] Xizhou Zhu, et al. "Deformable detr: Deformable transformers for end-to-end object detection." arxiv preprint arxiv:2010.04159 (2020).

[5] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Cham: Springer International Publishing, 2016.

[6] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arxiv preprint arxiv:1704.04861 (2017).

[7] Qiu, Han, et al. "Borderdet: Border feature for dense object detection." European Conference on Computer Vision. Cham: Springer International Publishing, 2020.

[8] Chen, Kai, et al. "MMDetection: Open mmlab detection toolbox and benchmark." arxiv preprint arxiv:1906.07155 (2019).