Theory and Practice of Science and Technology
2026, VOL. 7, NO. 2, 62-67
DOI: 10.47297/taposatWSP2633-456912.20260702

Research on Parcel Sorting System Fusing RGB-D and Force Sensing in Complex Scenarios

Hongyi Zhang

School of Resources and Materials, Northern University at Qinhuangdao, Qinhuangdao, Hebei, 066004, China

ABSTRACT
With the rapid development of e-commerce and cross-border logistics, logistics centers have increasingly higher requirements for the efficiency and accuracy of parcel sorting. Traditional manual sorting has obvious shortcomings in efficiency and stability, so automated sorting systems have gradually become an important direction for the development of intelligent logistics. Most existing automated sorting systems rely on machine vision technology, which uses industrial cameras to acquire images of target objects on the conveyor belt, and then industrial computers analyze and generate control commands to realize automatic grasping and transfer of parcels. However, in actual logistics environments, parcels usually present characteristics such as stacking, occlusion, and irregular shapes, making it difficult for a single vision system to maintain stable recognition performance.To address the problem of insufficient sorting stability in complex logistics environments, this paper proposes a parcel sorting method fusing RGB-D vision and force feedback. The system uses an RGB-D camera to obtain the spatial pose information of parcels, and combines the real-time grasping state feedback from the force sensor to realize collaborative perception of visual and force information, thereby improving the stability of parcel recognition and grasping in complex scenarios. Experimental results show that the system can achieve relatively stable parcel recognition and grasping under complex environmental conditions, providing a feasible solution for the design of intelligent logistics sorting systems.

KEYWORDS
RGB-D Vision; Force sensing; Robotic sorting; Multi-modal fusion; Intelligent logistics

1 Introduction

In recent years, with the rapid development of e-commerce and cross-border trade, the logistics industry has entered a stage of rapid growth, and the requirements for the efficiency and accuracy of parcel sorting are constantly improving. Traditional manual sorting has obvious shortcomings in efficiency and stability, which is difficult to meet the demand for large-scale parcel processing. Therefore, automated sorting systems have gradually become an important development direction of intelligent logistics.

In automated sorting systems, machine vision technology plays a key role. Machine vision simulates human visual functions through cameras and computers to realize the recognition and positioning of target objects, and is widely used in industrial production and logistics sorting fields. A typical visual sorting system uses industrial cameras to collect image information, obtains the position coordinates of target objects through computer processing, and then the robot completes grasping and handling to realize automatic sorting.

However, in actual logistics environments, visual sorting systems still face certain difficulties. For example, factors such as large changes in illumination in the warehousing environment, obvious reflection on the packaging surface, and dust interference will reduce the recognition accuracy of RGB vision. At the same time, parcels are often in a random stacking state, and special-shaped parcels and flexible packaging are common. Traditional vision algorithms are prone to false detection or missed detection under occlusion conditions.

In addition, it is difficult to judge the contact state between the robot and the parcel only by visual information, which is prone to slippage or grasping failure during the grasping process. With the development of robot technology, combining vision sensors with force sensors has become an important research direction of sorting systems. Multi-sensor fusion can improve the stability and reliability of the system.

Therefore, the research on parcel sorting methods fusing RGB-D vision and force sensing is of great significance for improving the performance of robot sorting systems in complex scenarios.

1.1 Domestic and International Research Status

1.1.1 Overseas Research Status

In overseas research, robot automatic sorting systems have become an important research direction in the field of intelligent logistics. Sorting robots are relatively mature in structural design and motion control, with the characteristics of high speed and high positioning accuracy. At the same time, researchers have gradually combined vision systems with robot control systems, analyzed the information obtained by cameras through vision software, and guided robots to complete sorting operations to realize automated production processes.

In recent years, foreign scholars have proposed a variety of robot sorting methods. For example, some studies have realized automatic classification and stacking of different types of parcels through robot systems, and used statistical methods to process parcels of different sizes to improve sorting efficiency. [8] At the same time, some studies have

optimized sorting task allocation through algorithms to improve system throughput and reduce waiting time.

Overall, foreign countries have made great progress in robot structural design and control, but there are still some deficiencies in complex environment perception, especially in the recognition of stacked parcels and stable grasping, which need further research.

1.1.2 Domestic Research Status

Domestic scholars have carried out a lot of research on machine vision sorting systems. Early research mainly focused on vision algorithms, for example, the recognition and positioning of regular workpieces were realized by combining corner features and contour features.

With the development of robot technology, the research focus has gradually shifted to the design of robot vision sorting systems. A typical system uses an industrial camera to obtain images of parcels on the conveyor belt, generates control commands after computer processing, and the robot completes grasping and transfer according to the coordinates of visual feedback to realize automatic sorting.

At the same time, some studies have applied machine vision technology to the logistics sorting process to extract information such as parcel position, size and packaging type, thereby improving the operation efficiency of the logistics system.

Overall, domestic research on visual sorting systems has achieved certain results, but most systems still rely on single visual information. Under complex environmental conditions, single-modal vision systems are difficult to meet the demand for stable grasping, so fusion of multi-modal perception technology has become an important direction of current research.

1.2 Research Content

Aiming at the problem of insufficient grasping stability of robots in complex warehousing environments, this paper proposes a multi-modal grasping and sorting method fusing RGB-D vision and force feedback. The main research contents include:(1) Establish a parcel detection and positioning method based on RGB-D vision. Use depth cameras to obtain 3D scene information, and realize parcel detection and grasping area determination through image processing algorithms. (2) Establish a force sensing model to realize contact state detection and slip judgment during grasping. Evaluate the grasping state and adjust it in time by collecting force sensor data in real time.(3) Design a visual and force fusion decision method. Correct the grasping action in real time by fusing multi-source information to improve grasping stability. (4) Build a robot sorting experimental platform and verify the effectiveness of the proposed method through experiments.

2 Relevant Theoretical Basis

2.1 RGB-D Visual Perception

The RGB-D vision system can simultaneously acquire color images and depth information, thereby realizing 3D scene perception. The position coordinates of the target object in space can be obtained by using depth data, which provides a basis for robot grasping.

Depth data can be filtered, segmented and feature extracted through point cloud processing technology, and area information suitable for robot grasping can be generated, thereby improving the grasping success rate.

The RGB-D camera provides both color texture and depth distance, enabling 2D detection results to be directly mapped to 3D space. In the camera coordinate system, the projection relationship between the pixel point (u,v) and the depth Z is shown in Equations (2-1) to (2-4):

$$X = (u - c_x)Z/f_x \tag{2-1}$$

$$Y = (v - c_y)Z/f_y \tag{2-2}$$

$$Z = D(u,v) \tag{2-3}$$

$$P = (X, Y, Z) \tag{2-4}$$

where f_x,f_y are the focal lengths, and c_x,c_y are the principal points.[6] To improve engineering usability, depth maps usually need denoising and completion: median/bilateral filtering can suppress random noise, and neighborhood interpolation or multi-frame fusion is used to alleviate holes caused by reflection and dark materials. On the point cloud side, voxel downsampling is commonly used to reduce the amount of calculation, outlier points are eliminated through statistical filtering, and RANSAC plane segmentation can remove the desktop/conveyor belt background. In the grasping candidate generation stage, geometric quantities such as local normal, curvature and main direction can provide hard constraints for "graspable/non-graspable", thereby maintaining an available candidate set even in weak texture or partial occlusion.

2.2 Force Feedback Control

In the actual grasping process, it is often difficult to accurately judge whether a stable contact is formed between the end effector and the parcel only by visual information. Especially in the case of flexible packaging or reflective surfaces,

even if the visual positioning is accurate, grasping may still fail due to unstable contact state. Force sensors provide physical information that directly reflects the contact state, enabling the robot to adjust the clamping strategy according to real-time force conditions, thereby improving the grasping success rate.

When the grasping force is insufficient, the parcel is prone to slip; while excessive grasping force may cause packaging deformation or even damage. By introducing force information into the control loop, the clamping force can be dynamically adjusted during the grasping process, so that the robot can avoid excessive extrusion of the parcel while ensuring stable grasping. The closed-loop control mechanism formed in this way makes the grasping process no longer rely on a single visual positioning result, but can continuously correct actions according to the actual contact state.

In system implementation, the end 6-axis force/torque sensor or the built-in force sensor of the gripper can provide contact information. These signals can reflect whether the contact is established, whether the friction conditions are sufficient, and whether there is abnormal collision during the grasping process. However, in logistics field tests, it can be observed that the original force signal is usually accompanied by obvious fluctuations, especially at the moment when the robot starts or stops, high-frequency vibration will affect the stability of slip judgment. Therefore, before using the force signal, it is necessary to filter the original measurement data. This paper uses a first-order low-pass filtering method to smooth the force signal:

$$F_t = \alpha F_{raw} + (1 - \alpha)F_{t - 1} \tag{2 - 5}$$

where F_{raw} is the current sampling value, F_t is the filtered force signal, and α is the filtering coefficient.

Slip judgment uses the Coulomb friction model as the basic basis. This model can describe the boundary conditions of contact stability from a physical perspective. When the tangential force is less than the friction limit, the contact usually remains stable:

$$|F_{tan}| \leq \mu F_n \tag{2 - 6}$$

If the tangential force exceeds the friction limit, relative movement is more likely to occur in the contact area:

$$|F_{tan}| > \mu F_n \tag{2 - 7}$$

where F_n represents the normal force, F_{tan} represents the tangential force, and μ is the friction coefficient.

In the actual logistics environment, the friction coefficient is not a stable constant. The surface roughness of cartons in different batches is significantly different, and the surface of plastic packaging is often attached with dust or label glue marks, which will cause changes in friction conditions. It can be found in the test process that even under the same clamping force conditions, the slip behavior of different parcels is significantly different. Therefore, this paper does not use a fixed threshold as the discrimination standard, but converts the slip detection result into an event-based score to trigger force increase or attitude adjustment actions. This processing method shows better stability in experiments.

3 System Design and Algorithm Implementation

3.1 Overall System Architecture

The overall structure of the system is shown in Figure1. The visual perception module and the force perception module run independently in the ROS environment, and complete data acquisition and preprocessing tasks respectively. The fusion node receives data from different sensors according to the control cycle, and completes time alignment and information integration. The fusion result is sent to the planning and control module to generate grasping and handling action commands.

This distributed structure shows good flexibility in the system debugging stage. When the visual algorithm needs to adjust parameters, it will not affect the operation of the force node; vice versa. In actual operation, the control cycle is kept within a stable range, which can meet the sorting cycle requirements.

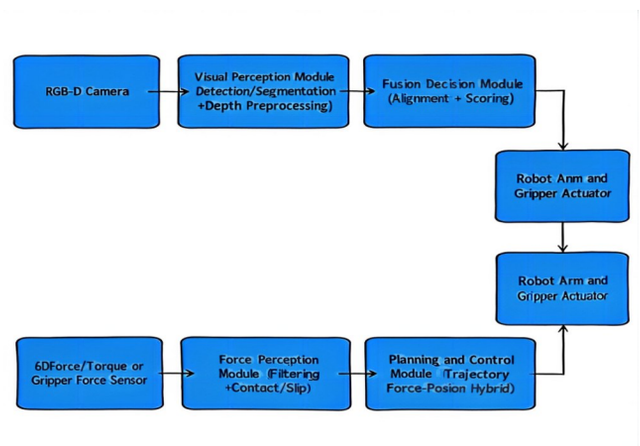

Figure 1 Overall system architecture diagram

3.2 Multi-modal Algorithm Flow

The algorithm flow is shown in Figure2. The system updates visual candidates, geometric scores and force event scores in each control cycle. When S is in the gray area, the system triggers a micro-probe action (such as a small close or light touch) to supplement evidence; if a force alarm is triggered or S is significantly low, it enters the repositioning/replanning/recovery process.

Supplementary note: In the visual side implementation, for targets with weak texture and clear boundaries such as cartons, a lightweight contour correction process can be added around the detection results. After the ROI is median filtered and thresholded, morphological opening operation is used to suppress noise; then Canny is used to extract edges, the largest connected domain is selected from all edge contours, and its minimum circumscribed rectangle and center point are solved, thereby stabilizing the target center and deflection angle estimation, and providing a more reliable initial value for the grasping attitude.

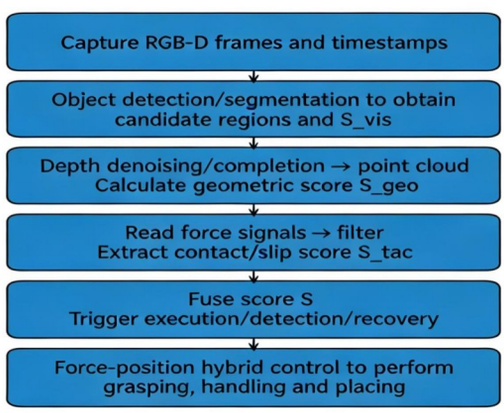

Figure 2 Flow chart of multi-modal fusion and control

3.3 Pseudo-code of Key Modules

Pseudo-code 1 Multi-modal Confidence Decision Logic

```
Plain Text
Input: S_vis, S_geo, S_tac, thresholds {T_low, T_high}, weights w
S ← w1·S_vis + w2·S_geo + w3·S_tac
2. if (S ⩾ T_high) and (not tactile_alarm) then
3.   EXECUTE(best_grasp_candidate)
4. else if (S < T_low) or (tactile_alarm) then
5.   TRIGGER(re_localize or re_plan or recovery)
6. else7.   MICRO_PROBE()   # small exploratory motion to gather evidence
8. end if
```

Pseudo-code 2 Slip-triggered Online Correction Strategy

```
Plain Text
Input: slip_event, F_n, F_max, transport_state
1. if slip_event then
2.   if F_n < F_max then
3.     F_n ← min(F_n + ΔF, F_max) # increase grip
4.   end if
5.   adjust_pose(Δθ)              # micro-rotation
6.   limit_motion(v_max_new, a_max_new)
7.   if slip_persists(N_cycles) then
8.     retreat(); regrasp()       # recovery
9.   end if
10. end if
```

4 Multi-modal Grasping and Sorting Method

The multi-modal grasping and sorting method proposed in this paper mainly includes three parts: perception layer, decision layer and control layer. The perception layer is responsible for acquiring visual and force information; the decision layer selects grasping strategies based on multi-modal data; the control layer is responsible for executing specific robot motion control.

In the visual detection stage, first obtain the scene point cloud data through the depth camera, and extract the parcel target using the point cloud segmentation algorithm. Then, generate multiple grasping candidate points according to the geometric characteristics of the object. The system selects the optimal grasping candidate through a multi-modal confidence scoring mechanism.

In the grasping execution stage, the force sensor monitors the contact force during the grasping process in real time. When a slip trend is detected, the system automatically increases the clamping force or adjusts the grasping attitude, thereby ensuring grasping stability.

To verify the effectiveness of the proposed method, this paper builds an experimental platform based on an RGB-D camera and conducts grasping experiments in a simulated warehousing environment. Experimental results show that under complex environmental conditions, the multi-modal perception system can significantly improve the grasping success rate compared with a single vision system. At the same time, in grasping tests of parcels of different shapes, the system shows good stability.

Experimental data show that the grasping success rate of the multi-modal grasping system in complex environments is increased by about 15%, and the positioning error is significantly reduced. It can be seen that fusing visual and force information can effectively improve the reliability of the robot sorting system.

5 Experimental Design and Result Analysis

Three methods are selected in the experiment: RGB only, RGB-D, RGB-D + force closed loop (proposed in this paper). It can be seen from Figure 3 that the multi-modal closed loop can still maintain a higher success rate under complex conditions. It is worth noting that the improvement in success rate does not come from simply "clamping tighter", but from the reduction of instability events in the handling stage through contact confirmation and slip correction. Figure 4 shows that the 3D positioning error decreases with the introduction of depth and force sensing; on reflective packaging materials, depth holes will cause fluctuations in geometric scores, and force micro-probing can provide supplementary evidence for candidate screening at this time. Figure 5 shows that the fusion strategy brings a slight computational overhead, but overall it is still within an acceptable range, meeting the station cycle requirements.

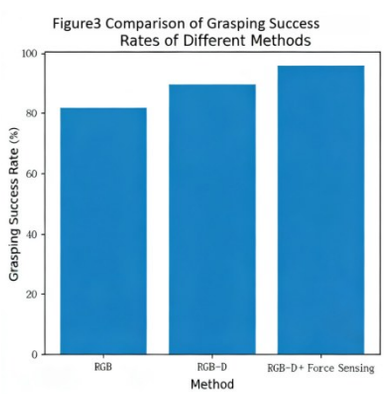

Figure 3 Bar chart of grasping success rate

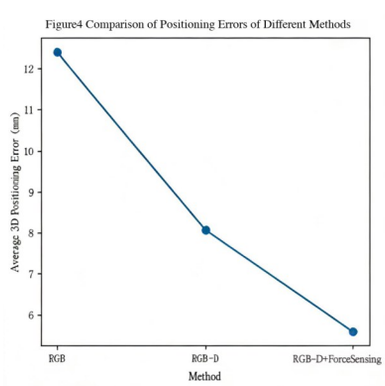

Figure 4 Comparison chart of average 3D positioning error

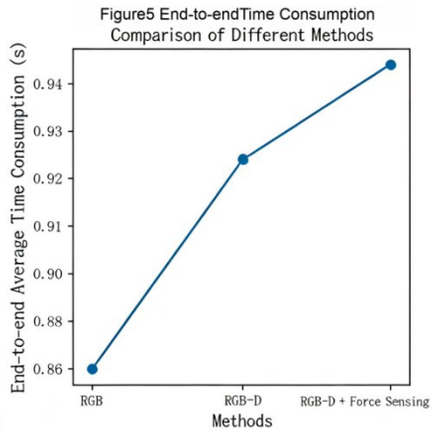

Figure 5 Comparison chart of end-to-end average time consumption

Compared with the optimization method proposed by Tan et al. (2021) for vertical sorting systems[7] and the sorting method for heterogeneous parcel stacking proposed by Grambo et al. (2019)[8], the multi-modal grasping method in this paper achieves better grasping stability in complex scenes and can also meet the requirements of sorting cycle .

Table 3 Comparison of comprehensive indicators of different methods

Method	Grasping Success Rate	Average Positioning Error (mm)	End-to-end Time Consumption (s)
RGB	82%	12.4	0.85
RGB-D	90%	8.1	0.92
RGB-D + Force Sensing	96%	5.6	0.95

6 Conclusion

Aiming at the robot grasping and sorting problem in complex warehousing environments, this paper proposes a multi-modal perception method fusing RGB-D vision and force feedback. By constructing a visual positioning model and a force feedback control mechanism, closed-loop control of the grasping process is realized. Experimental results show that this method can significantly improve the grasping success rate and reduce positioning errors in complex environments. Future research can further combine deep learning algorithms and intelligent control technology to improve the intelligence level of robot sorting systems.

References

[1] E, L.B., Huang, Y.W. (2014) New Mode of International Trade: The Latest Research on Cross-border E-commerce. Journal of Dongbei University of Finance and Economics, (02): 22-31.

[2] Li, C., Wei, X., Zhou, Y., Li, H.Y. (2022) Research on Control of Intelligent Logistics Sorting Robot Based on Laser Vision Guidance. Laser Journal, 43(08): 217-222.

[3] Liu, Y.X., Chen, S.X., Luo, J.J., Fu, B., Xu, Z.Y. (2025) System Solution of Intelligent Sorting Robot for Logistics Terminal. Information & Computer, 37(03): 18-20.

[4] Jiao, Z. Y., Cui, H. Y. (2025) Design and Research of Industrial Robot Sorting System Based on Machine Vision. Technology Innovation and Application, 15(22): 95-98.

[5] He, Z.Q. (2016) Design of Industrial Robot Sorting System Based on Machine Vision. Master's Thesis, Harbin Institute of Technology, Harbin.

[6] Liu, Z.Y., Zhao, B., Zou, F.S. (2012) Application of Machine Vision Technology in Workpiece Sorting. Computer Applications and Software, 29 (11): 87-91.

[7] Tan, Z., Li, H., He, X. (2021) Optimizing parcel sorting process of vertical sorting system in e-commerce warehouse. Advanced Engineering Informatics, 48: 101279.

[8] Grambo, P., Mullick, T., Furukawa, T., Matoba, M., Nasu, Y. (2019) Automatic sorting-and-holding for stacking heterogeneous packages in logistic hubs. IFAC-PapersOnLine, 52(10): 109-114.

[9] Xu, X., Chen, Y., Zou, B., Gong, Y. (2022) Assignment of parcels to loading stations in robotic sorting systems. Transportation Research Part E: Logistics and Transportation Review, 164: 102808.