Research on Visual Detection of Intruding Foreign Objects in Rail Transit

Hu Bo, Liu Peiwen
College of Automation, Guangxi University of Science and Technology, Liuzhou, Guangxi, China

Abstract: Foreign object intrusion detection is one of the important means to ensure the safe operation of rail transit. The distance measurement of the intruding foreign object is conducive to reminding the train crew to take corresponding measures in time. This paper proposes a distance detection algorithm for intruding foreign objects based on monocular vision. Firstly, corner detection and threshold segmentation methods are used to obtain the image coordinates of a black-and-white checkerboard. Secondly, regression analysis is employed to establish a conversion model between image coordinates and world coordinates. Finally, an object detection algorithm based on the YOLOv10 network is used for image recognition to obtain the coordinates of the intruding foreign objects in the image coordinate system. The distance information of the intruding objects is detected. Experimental validation shows that the distance of the intruding objects has an error range of -5 to +6 cm in the X-axis direction and -9 to +16 cm in the Y-axis direction, demonstrating high accuracy.

Keywords: Coordinate Transformation; Distance Measurement; Regression Analysis; Threshold Segmentation.

1. Introduction

In rail transit, foreign object intrusion is a significant factor that can lead to train accidents. Detecting these intruding objects can effectively prevent such accidents, and determining the distance to the intrusion target can help train operators adjust their response plans in a timely manner [1-3]. Distance measurement between objects can generally be achieved using methods such as laser, radar, infrared, GPS, ultrasonic, and visual measurement. Among these, GPS-based distance measurement requires the object to be equipped with a GPS receiver, while infrared and ultrasonic measurements are greatly affected by environmental factors such as temperature, humidity, and light intensity, and their measurement range is relatively short, making them unsuitable for determining the distance to intruding objects. Xu Shixiong and colleagues[4] utilized an MCU as the main control system to collect CMOS image data and implemented distance measurement using laser triangulation. Liu Xin and colleagues[5] analyzed the timing of working signals to propose a method for addressing main echo overlap, thereby enhancing the detection probability of laser ranging systems. Ju Meiyu and colleagues[6] proposed a minimum divergence radar ranging method based on relative entropy, which demonstrated good performance in experimental validation. Although distance measurement methods based on laser and radar offer high precision, the equipment costs are also relatively high.

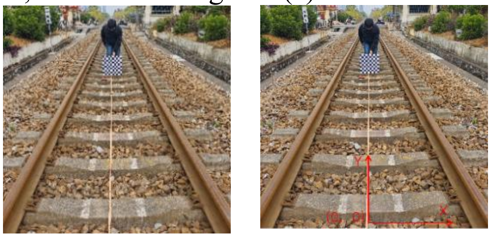
With the development of visual technology, distance measurement methods based on visual technology have also shown excellent performance. Among these, binocular vision measurement methods obtain disparity data through binocular matching, making it relatively easy to calculate the depth information of objects in the scene[7-8]. While binocular vision can accurately measure the distance to objects, the pixel matching mechanism results in a high computational load and requires advanced hardware. Additionally, occlusions in images can lead to matching failures, affecting distance calculations. Monocular vision, on the other hand, obtains the mapping relationship between image coordinates and world coordinates through coordinate transformation, using image coordinates to acquire the object's coordinates in the world coordinate system, thus measuring the distance to the object[9]. This method not only has lower costs but also features lower algorithm complexity and better real-time performance.

2. Establishing the Mapping Relationship of Coordinate Systems

To detect the position of foreign objects, it is essential to establish a mapping relationship between image coordinates and world coordinates. Essentially, this involves correlating the coordinates of the same point in both the image coordinate system and the world coordinate system. Using a black-and-white checkerboard as a marker is a common method for coordinate transformation.

2.1 Obtaining Image Coordinates

Since the black-and-white checkerboard is composed of alternating black and white squares, it has a high contrast and distinct corner points, making it easy for image processing algorithms to detect and recognize. The black-and-white checkerboard is used as a marker in the track scene, as shown in Figure 1(a).



(a) (b)

Figure 1. Using a Black-And-White Checkerboard as a Marker

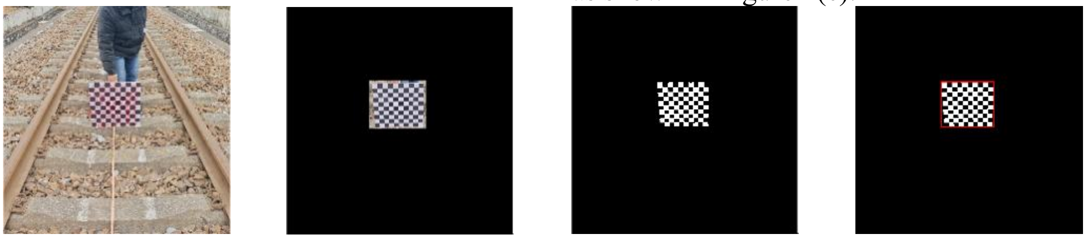
As shown in Figure 1(a), to obtain the distance information of foreign object targets, it is only necessary to know the positional information of the orbital plane where the foreign object is located, without needing to know the spatial information in the third dimension. Taking the camera position at the center of the two rails in the orbital area as the origin of the world coordinate system, an X-Y coordinate system is established as shown in Figure 1(b). The real coordinate values in the world coordinate system for the bottom-left, bottom-right, and center points of the lower edge of the black-and-white checkerboard are measured using a tape measure. There are two ways to obtain coordinate points in the image world coordinate system: one is through manual observation, which involves directly reading the corresponding coordinate data using a data cursor (u, v), and the other is to automatically acquire coordinate data by processing the image data.

Due to the complexity of the track area environment, it is relatively difficult to extract the black-and-white checkerboard area directly using threshold segmentation methods. Therefore, corner detection is first performed on the image. However, as shown in Figure 2(a), the corner detection results are incomplete due to the unclear corner points at the contact surface between the checkerboard and the track area. To obtain more accurate detection results, the eight continuous corner point coordinates in both the horizontal and vertical directions are extracted based on the detected corner points. The pixel distances occupied by each black-and-white square in the image are then calculated, with the formulas shown in Equations (1) and (2). Subsequently, using the obtained pixel distances, the region of interest where the black-and-white checkerboard is located is roughly determined, as shown in Figure 2(b).

$$\delta_u = \frac{1}{n}\sum_{i=1}^{n} u_{i+1} - u_i \tag{1}$$

$$\delta_v = \frac{1}{n}\sum_{i=1}^{n} v_{i+1} - v_i \tag{2}$$

After obtaining the approximate location of the black and white checkerboard, the region of interest is first converted to grayscale. Since the white areas of the checkerboard are very distinct, a higher threshold is used to binarize the grayscale image. At this stage, the resulting binary image contains a significant amount of noise. To reduce the impact of noise on the results, an erosion operation is first applied to the binary image to remove weaker noise. Subsequently, the connected components of the binary image are calculated, and by eliminating smaller connected components, the influence of noise is effectively mitigated. Finally, a dilation operation is performed to complete the binarization of the black and white checkerboard, as shown in Figure 2(c).

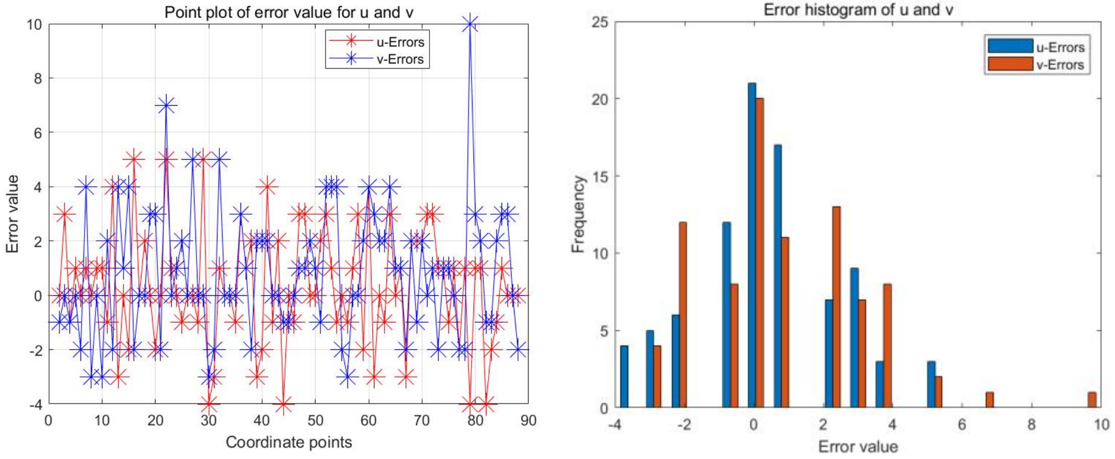


(a) (b) (c) (d)

Figure 2. Schematic Diagram of the Extracted Feature Point Image Processing Results

(a) Schematic diagram of corner point extraction from a black-and-white checkerboard, (b) Schematic diagram of the region of interest extraction from a black-and-white, (c) Schematic diagram of the binarization result of a black-and-white checkerboard, (d) Schematic diagram of the minimum enclosing rectangle drawing result.

After completing the binarized image of the checkerboard, its minimum enclosing rectangle is drawn to extract the checkerboard area, as illustrated in Figure 2(d). This results in the coordinates of the bottom-left and bottom-right points of the checkerboard in the image coordinate system, and the midpoint coordinates of the bottom edge are calculated. Obtaining the coordinates of the markers in the image coordinate system through image processing aids in reconstructing the mapping relationship between the image coordinate system and the world coordinate system when the visual sensor is displaced.



(a) (b)

Figure 3. Statistical Chart of Errors Between Coordinates Obtained through Image Processing and Manually Obtained Coordinates.

A total of 88 marker coordinate points were obtained in the experiment, with the differences between the manually obtained coordinates in the image coordinate system and those obtained through image processing shown in Figure 3. From the figure, it can be seen that the error values of the two sets of coordinates are mostly distributed between -2 and 2 pixels, with only a few differences exceeding 6 pixels.

2.2 Establishing the Coordinate Transformation Model

Establishing a coordinate transformation model is key to understanding world coordinates from image coordinates, which involves exploring the direct relationship between independent and dependent variables. Methods for establishing transformation models include polynomial fitting, nonlinear fitting, and machine learning fitting. While machine learning methods are suitable for complex relationships, they require a large amount of data for training. Since the landmark points are linearly correlated in both the world coordinate system and the image coordinate system, linear regression analysis can be used to establish the coordinate transformation model.

To establish the coordinate transformation model using linear regression analysis, it is first necessary to set up the linear regression equation. The higher the degree of the linear regression equation, the better the data fitting effect; however, a degree that is too high may lead to overfitting. The regression equation for a first-degree equation is given as shown in equation (3), and to reflect the coupling of coordinates, the equation includes u, v multiplicative terms.

$$\begin{cases} x = a_3 \times u + a_2 \times v + a_1 \times u \times v + a_0 \\ y = b_3 \times u + b_2 \times v + b_1 \times u \times v + b_0 \end{cases} \tag{3}$$

where (u, v) represents the coordinates in the image coordinate system, and (x, y) represents the coordinates in the world coordinate system. By using the least squares method, the coefficients $a_k(k = 0，1 \ldots)$, $b_k(k = 0，1 \ldots)$ can be determined, allowing the establishment of the coordinate transformation model. The corresponding world coordinates (x, y) can then be solved from the pixel coordinates (u, v).

2.3 Calibration Experiment of the Track Area

and Result Analysis

The black-and-white checkerboard landmark points in the image coordinate system are used as input, while the corresponding landmark points in the world coordinate system serve as output to establish the coordinate transformation model. The input matrix and output matrix are substituted into the regression equation to solve for the equation coefficients. Subsequently, the coordinates of the landmark points in the image coordinate system are used as input to predict the world coordinates using the linear equation model, with the prediction results shown in Table 1.

Table 1. Prediction Results and Errors of Regression Model

	Pixel coordinates/px		Corresponding world coordinates/cm		Regression to the predicted value of world coordinates/cm		Prediction error/cm	
	u	v	x	y	x	y	Δx	Δy
1	565	363	120	270	118.82	278.64	-1.18	8.64
2	310	156	0	839	2.27	827.88	2.27	-11.12
3	250	112	-120	1293	-117.43	1315.81	2.57	22.81
4	327	172	10	729	13.76	718.78	3.76	-10.22
5	281	196	-10	611	-12.57	598.92	-2.57	-12.08
6	236	160	-110	783	-107.45	785.38	2.55	2.38
7	378	148	110	839	104.93	850.76	-5.07	11.76
8	401	130	130	1011	136.77	1024.35	6.77	13.35
9	320	102	10	1520	16.43	1498.92	6.43	-21.08
10	364	104	120	1410	122.76	1427.42	2.76	17.42
11	310	89	0	1927	-2.97	1910.83	-2.97	-16.17
12	283	172	110	729	106.42	735.54	-3.58	6.54
13	125	228	-130	498	-132.24	503.62	-2.24	5.62
14	250	381	-10	270	-7.88	266.96	2.12	-3.04
15	323	115	10	1293	8.94	1312.59	-1.06	19.59

As shown in Table 1, the coordinate transformation model established using the regression equation predicts the world coordinates for 15 validation landmark points. The maximum error in the x-axis direction is 6.77 cm, and the maximum error in the y-axis direction is 22.81 cm. This error is due to the distance from the world coordinate origin being 1293 cm, which meets the actual requirements.

3. Identification of Intruding Objects

After establishing the coordinate transformation model between the image coordinates and the world coordinates, it is necessary to extract the coordinate points of the intruding objects within the image coordinate system. First, it is essential to identify the pixel regions in the image that represent the intruding objects, which allows the problem to be transformed into an object detection task for detecting intruding objects in the image data. Traditional methods such as threshold segmentation, color features, and morphological features struggle to complete detection tasks in complex track environments and with varying categories of intruding objects. However, with the development of deep learning technology, the YOLO series of object detection methods based on deep learning have demonstrated excellent performance in the field of object detection. Although deep learning-based methods require a large number of data samples for training, common intruding objects on tracks, such as pedestrians, vehicles, falling rocks, and livestock, are relatively easy to obtain. By simply adding a small amount of data on such objects in the track area to the dataset, the algorithm's adaptability to the track environment can be enhanced, enabling the detection of intruding foreign objects.

This paper uses YOLOv10 [10] as the object detection algorithm. As the latest version in the YOLO series, it improves upon the advantages of previous models by enhancing the network structure and training strategy, further increasing detection accuracy, especially in complex scenes. Additionally, while maintaining detection accuracy, the model parameters and computational load have been optimized to improve computational efficiency, making it more suitable for real-time scenarios and deployment on resource-limited devices. This study created a dataset using 4,650 images, divided into training, validation, and testing sets in a ratio of 7:2:1. After iterative training, the average precision mAP50 of the algorithm network detection reached 88.7%. The results of using the trained YOLOv10 algorithm network for intruding object detection on image data are shown in Figure 4.

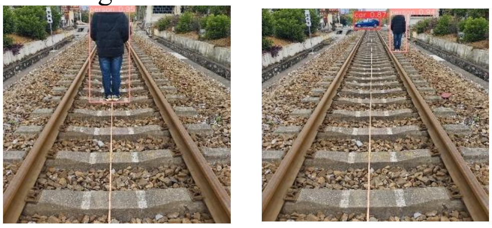

(a) (b)

Figure 4. Schematic Diagram of Object Detection Results Using YOLOv10

4. Calculation of the Position of Intruding Objects

The distance of an intruding object refers to measuring the coordinates of the intruding object in the world coordinate system, using the position of the visual sensor as the origin of the world coordinate system. The distance is the interval between the coordinate point and the origin. As shown in Figure 4, the algorithm detects intruding pedestrians and vehicles, marking the detection results with an anchor box. By obtaining the positions of the four corner points of the anchor box, the coordinates of the center of the lower edge of the anchor box can be found, which represent the coordinates of the intruding object in the image coordinate system.

By performing object detection on the image data, the coordinates of the intruding objects in the image coordinate system are obtained. The experiment detects 10 images of object intrusions, resulting in 10 coordinate points for intruding pedestrians in the image coordinate system. These coordinate points are used as input to a coordinate transformation model, which outputs the coordinates of the intruding pedestrians in the world coordinate system. The predicted results, true values of pedestrian positions, and measurement errors are shown in Table 2.

Table 2. Measurement Results of Encroaching Object Distance

	Pedestrian coordinates value/cm		Pedestrian coordinates predicted value/cm		Prediction error/cm	
	x	y	x	y	Δx	Δy
1	0	384	2.43	387.24	2.43	3.24
2	112	550	108.49	543.98	-3.51	-6.02
3	0	833	3.94	836.47	3.94	3.47
4	0	943	4.53	951.69	4.53	8.69
5	-105	1011	-99.74	1019.61	5.26	8.61
6	112	1180	107.67	1173.01	-4.33	-6.99
7	-105	1180	-102.39	1191.24	2.61	11.24
8	0	1350	5.12	1341.35	5.12	-8.65
9	-105	1469	-103.88	1481.92	1.12	12.92
10	0	1469	4.27	1483.42	4.27	14.42

As shown in Table 2, the measurement results of the distance to the encroaching objects have an error of no more than ±6 cm in the x-axis direction and no more than ±15 cm in the y-axis direction. The closer the measurement is to the world coordinate origin where the visual sensor is located, the smaller the measurement error becomes, with the minimum error being only 3.24 cm. This meets the task requirements for measuring the distance of encroaching foreign objects during rail transit inspections.

3. Conclusion

A visual measurement method for the distance of encroaching foreign objects in rail transit has been proposed. First, a coordinate transformation model was established. Then, the YOLOv10 object detection algorithm was utilized to select the area of the encroaching objects in the image. By determining the positions of the anchor box corners, the image coordinate points representing the locations of the encroaching objects were identified. Finally, the distance measurement of the encroaching objects was achieved through the coordinate transformation model.

The final measurement results indicate that the error in the x-axis direction is less than ±6 cm, and the error in the y-axis direction is less than ±15 cm, which meets the task requirements for determining the distance of encroaching objects during visual inspections in rail transit.

References

[1] Wang Tianyu, Peng Zaiyun, Liang Xiuhui, et al. Design of automatic monitoring System

for Urban Rail Transit based on multi-eye Machine vision [J]. Automation and Instrumentation. 2023, 38(12):1-5.

[2] Wang Hui, Jiang Zhufeng, Wu Yujie, et al. Rapid Detection of foreign body penetration in Railway based on Deep Learning [J]. Journal of Railway Science and Engineering. 2024, 21(5):2086-2098.

[3] Li Zhengzhong, Liu Shuang, PeiHuiran, et al. Research on Prevention and control measures of safety risks in protected areas of urban rail transit lines [J]. Modern Urban Rail Transit. 2024(3):113-117.

[4] Xu Shixiong, Yuan Zhenhuan, Cao Qiuting, et al. Design and implementation of CMOS single-line laser ranging system based on MCU [J]. Modern Information Technology, 2024, 8(21):1-5+10.

[5] Liu Xin, Yang Haima, Zhang Liang, et al. Main echo overlap and coping method of single-photon laser ranging system [J]. Applied Laser, 2024, 44(08):82-92.

[6] Ju Meiyu, Xu Junior College, Xu Huan. A method of radar range estimation based on relative entropy [J]. Data Acquisition and Processing, 2024, 39(06):1326-1332.

[7] He Junyao, Wang Wensheng, Han Yihang. Design of floating garbage detection and positioning system based on YOLOv8 and binocular ranging algorithm [J]. Modern Electronic Technology, 2024, 47 (20): 1-7.

[8] Liu Zhen, Dong Shaojiang, Luo Jiayuan, et al. Fire detection and ranging method based on binocular vision and improved YOLOv8n [J]. Journal of Shaanxi University of Science and Technology, 2025, 43(01):152-160.

[9] Weizhu Zhu, Zurong Cui, Lei Chen, et al. Robust monocular vision-based monitoring system for multi-target displacement measurement of bridges under complex backgrounds, Mechanical Systems and Signal Processing, 2025, 225:112242.

[10] Ao Wang, Hui Chen, Lihao Liu , et al. YOLOv10: Real-Time End-to-End Object Detection[J]. 2024.