Research Article

Wenbo Fu, Zhihui Ban*, Xing Li, and Yingao Han

Optimization of multi-objective recognition based on video tracking technology

https://doi.org/10.1515/comp-2025-0040
received January 22, 2025; accepted June 20, 2025

Abstract: Considering the shortcomings of traditional video multi-target recognition technology in rapidly identifying criminal suspects in complex scenes, a multi-target recognition optimization method based on video tracking technology is proposed. This method constructs a multi-target recognition algorithm based on video feature matching, and introduces the Kalman filter algorithm to improve the accuracy and real-time recognition of criminal suspects through the definition of feature vector and similarity function. Experiments showed that the model proposed in the study performed exceptionally well in terms of tracking error; the highest precision was 94.75%, the recall rate was 96.59%, the tracking error of the horizontal axis was only 3.75%, and the tracking error of the vertical axis was 3.27%. In the crime detection video application, the accuracy–recall curve of the model was 0.94, and the feature recall rate was 94.83%, verifying the effectiveness and robustness of the model in complex and fast scenes. The results show that the proposed model has good feasibility and robustness in rapidly identifying criminal suspects. In addition, the work offered new technical concepts for improving target tracking precision and adapting to real-time scene changes, opening new research avenues in the field of multi-target recognition.

Keywords: video tracking, Kalman filtering, multi-target, feature matching, dynamic prediction

1 Introduction

In the fields of modern security management, traffic monitoring, and human-computer interaction, the application of multi-target recognition and tracking technology for rapid identification of criminal suspects in complex scenes has become increasingly significant [1,2]. With the rapid development of video surveillance technology, efficiently and accurately identifying multiple targets in complex environments, especially fast-moving criminal suspects, has become an important research topic [3]. This technology is of great significance to public security and criminal investigation and can effectively help prevent security risks and solve crimes. In view of this, a multi-target recognition model based on video tracking technology and the Kalman filter (KF) algorithm has been proposed. However, existing multi-target recognition and tracking technologies still face many challenges in complex scenarios. For example, in situations such as rapid target movement, mutual occlusion, lighting changes, and dynamic background changes, the tracking accuracy and stability of existing algorithms often fail to meet the requirements of practical applications. These problems not only limit the effectiveness of multi-target tracking technology in complex environments but also pose serious challenges to the intelligent development of fields such as public safety, traffic management, and industrial production. This algorithm extracts features, such as shape, color, and texture of the target, constructs feature vectors, and defines similarity functions to improve recognition accuracy and real-time performance. Concurrently, the KF algorithm was introduced to address the challenges posed by fast-moving targets and occlusion, thereby enhancing the efficacy of the recognition model. The integration of historical frame information with current frame data facilitates the prediction of the target's motion trajectory, thereby enhancing the tracking stability and accuracy of the model in complex dynamic environments. The innovation of the research lies in the construction of a video tracking method that integrates multiple feature information, such as shape, color, and texture, of the target. This method dynamically adjusts the weights of each feature by constructing feature vectors, thereby significantly improving the accuracy and real-time performance of target recognition in complex scenes. Second, the introduction of the KF algorithm combined with historical frame information and current frame data effectively solves the problems of rapid target movement and occlusion,

*** Corresponding author: Zhihui Ban,** College of Penology, The National Police University for Criminal Justice, Baoding, 071000, P. R. China, e-mail: 17726019076@163.com
Wenbo Fu: College of Penology, The National Police University for Criminal Justice, Baoding, 071000, P. R. China
Xing Li, Yingao Han: Department of Investigation, Hebei Public Security Police Vocational College, Shijiazhuang, 050091, P. R. China

thereby improving the tracking stability and accuracy of the model in dynamic environments. This research has the potential to contribute to the development of more reliable suspect identification technology for the public security field. Furthermore, it may offer novel technical concepts and solutions for multi-target tracking applications in intelligent transportation, industrial automation, medical health, and other domains. These advancements could stimulate technological progress and social development in related fields.

2 Related works

In response to the ever-increasing demand for multi-target recognition, it has been the subject of extensive research by many academics. Hu et al. proposed a novel dual-camera system that could quickly switch camera capture modes and achieve fast image processing through synchronized high frame rate zoom. The study was based on a convolutional neural network for recognition to achieve simultaneous recognition of camera and model. Experimental results showed that this new shooting mode effectively overcame the previous problem that a single fixed camera could not capture images clearly due to its wide field of view [4]. Zheng et al. proposed a method to manipulate the similarity between suspects and surrogates through a multi-dimensional scalingmodel to achieve the best recognition performance. The outcomes revealed that increasing the similarity between suspects and surrogates could reduce misidentification, but too much similarity could increase the difficulty of cases and reduce the correct identification rate of criminals [5]. Hou et al. proposed a faceted model based on similarity adjustment to focus on data association, which aimed to solve the case of mismatch of tracking association data for multi-targets and multi-cameras. After applying this model to the dataset, it could achieve more competitive performance and improve the accuracy and reliability of multi-target tracking and re-recognition [6]. Lakshmi and Arakeri proposed a sketch-based accurate face recognition technology to address the difficulty of identifying suspects in facial recognition in video surveillance, to solve the problem of using sketches for recognition without photos. The performance analysis of the proposed method on the Chokepoint dataset showed that the system had an accuracy of 89.02%, a recall rate of 91.25%, and an *F*-measure of 90.13%, proving its effectiveness [7]. As demonstrated in the preceding study, the majority of the methods may exhibit substantial variability in recognition across multiple frames when dealing with fast-moving targets. This can impede the efficacy of image segmentation and feature matching techniques in accurately capturing target states. Moreover, the algorithmic feature matching approach is susceptible to failure when multiple targets become occluded from each other, leading to instability in recognition and tracking performance. Therefore, to achieve accurate and stable multi-target video tracking, a multi-objective detection optimization model combining video tracking technology and a multi-target algorithm is proposed. This model integrates various feature information, including shape, color, and texture, of the target, and constructs feature vectors to dynamically adjust the weights of each feature. This process significantly improves the accuracy and real-time performance of target recognition in complex scenes.

3 Methods and materials

3.1 Algorithm for multi-target recognition with video feature matching

Rapid identification of criminal suspects in complex scenes is an important application of multi-target recognition algorithms, which covers many fields such as public safety, traffic management, and investigation. By improving the accuracy and real-time performance of target recognition, these algorithms can help the security monitoring system to respond in time, so as to effectively prevent security risks and solve crime problems [8,9]. Multi-target recognition algorithms first extract the unique features of the target through image processing techniques, which are multifaceted, such as shape, color, texture, and so on. Second, based on the extracted features and the target recognition results, the target is classified or tracked. Finally, the model outputs the prediction results or recognition results [10,11]. The multi-target recognition algorithm of criminal suspects based on video feature matching is constructed, as shown in Figure 1.

In Figure 1, the multi-target recognition algorithm of fast criminal suspects first recognizes the image in the target video and determines whether the recognized pixel is the key pixel of the video image. If it is not the key pixel, it re-recognizes it. If it is a key pixel, all pixels in the identification box are identified. Next, it determines whether the identification pixel is a domain pixel of the identification frame, and if it is a domain pixel, it expands the region of the identification frame to cover the pixel. If it is not a domain pixel, it adds a new identification frame to cover the pixel. Next, it determines whether the image is processed, and if it is not processed, the pixel recognition is performed again. If it is processed, the number of bounding recognition frames is returned, and finally, the

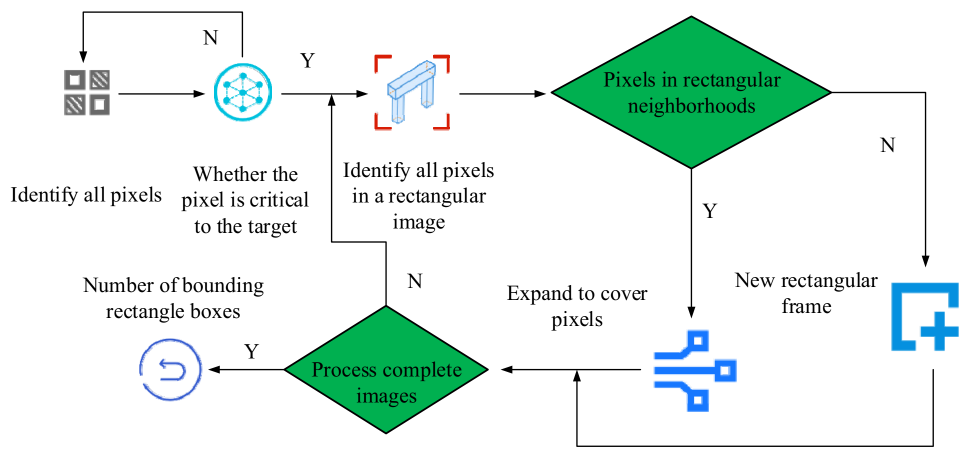

Figure 1: Multi-target recognition algorithm based on video feature matching.

outcomes are output. To utilize the above target feature matching metrics, the study defines a feature vector, which is utilized to correspond with the desired attributes. This feature vector consists of the size, color features, position information, shape ratio, and size information of the target region. The vector expression is as equation (1) [12].

$$a_{n,i} = (s_{n,i},\ R_{n,i},\ G_{n,i},\ B_{n,i},\ x_{n,i},\ y_{n,i},\ \mathrm{rate}_{n,i}). \tag{1}$$

In equation (1), $a_{n,i}$ represents the ith feature vector of image n. By effectively using the feature vector, the model can achieve higher accuracy and real-time performance in the process of rapid identification of criminal suspects. $s_{n,i}$ denotes the region in the chosen image that the moving target occupies. Among them, a large area may indicate that the target is a primary target, while a small area may indicate a secondary target or non-target area. $R_{n,i}$ denotes the average value of the red pixel point, which reflects the intensity of the target's red component in the image. $G_{n,i}$ denotes the green pixel point's average value, which reflects the intensity of the target's green component in the image mean value. It reflects the target's green component's intensity in the picture. $B_{n,i}$ represents the blue pixel point's mean value, which reflects the intensity of the blue component of the target in the image. $x_{n,i}$ denotes the horizontal coordinate in the matrix, and $y_{n,i}$ denotes the vertical coordinate in the matrix, which is used to characterize the position or contour of the target in the above image. The symbol $\mathrm{rate}_{n,i}$ stands for the proportion of the matrix's height to breadth, which is used to give details about the target's width and length. The moving target is used as a characteristic flow when defining a similarity attribute of the target image and is used in the feature matching job because the slight change in the target's movement between two frames of the graphic gives the image an apparent continuity. Equation (2) represents the expression of this resemblance function.

$$\Delta s_{i,j} = \frac{|s_{n,i} - s_{n-1,j}|}{s_{n,j}}. \tag{2}$$

In equation (2), $\Delta s_{i,j}$ denotes the similarity function. This formula defines a similarity function for measuring the degree of similarity between target images. The similarity function measures the similarity of the target in the time series by calculating the degree of change in the target feature values between two consecutive frames. This similarity function can quantify the changes of the target in a time series to evaluate the consistency and stability of the target in consecutive frames. To identify the same target in consecutive frames, it also allows feature matching between the current frame image and the previous one. The study identifies the similarity operations of the colors in the image and the previous frame image, which is expressed in equation (3). This allows one to ascertain the color mean value of the target.

$$\begin{cases} \Delta R_{i,j} = \dfrac{|R_{n,i} - R_{n-1,j}|}{R_{n,i}}, \\ \Delta G_{i,j} = \dfrac{|G_{n,i} - G_{n-1,j}|}{G_{n,i}}, \\ \Delta B_{i,j} = \dfrac{|B_{n,i} - B_{n-1,j}|}{B_{n,i}}. \end{cases} \tag{3}$$

In visual tasks, color is a key feature for classification and target tracking tasks. Therefore, equation (3) can help the algorithm to evaluate whether the color information in the currently displayed image and the prior frame image is consistent. The analogous function between the current i and j in the prior frame image is as equation (4).

$$\begin{cases} \Delta x_{i,j} = \dfrac{|x_{n,i} - x_{n-1,j}|}{x_{n,i}}, \\ \Delta y_{i,j} = \dfrac{|y_{n,i} - y_{n-1,j}|}{y_{n,i}}. \end{cases} \tag{4}$$

In equation (4), $(\Delta x_{i,j}, \ \Delta y_{i,j})$ represents the coordinates of the center position of the image recognition frame. The function of similarity is shown in equation (5) with respect to the ratio of height to width, which is characteristic of the outer rectangular box of the moving target in the movie.

$$\Delta \mathrm{rate}_{i,j} = \frac{|\mathrm{rate}_{n,j} - \mathrm{rate}_{n-1,j}|}{\mathrm{rate}_{n,i}}. \tag{5}$$

Equation (5) mainly analyzes the similarity of shape ratios. The algorithm can distinguish between different types of targets, such as long objects and round objects, and can help identify and classify different targets by the ratio of height to width of the outer rectangular box. Shape features can help analyze the behavioral characteristics of moving targets when detecting complex actions or for specific behaviors. The degree function is incorporated into the element fusion process to fuse the four elements used for target matching. The degree function's definition is provided in equation (6).

$$\Delta a_{i,j} = a(\Delta s_{i,j})^2 + \beta(\Delta r_{i,j})^2 + (\Delta g_{i,j})^2 + \chi(\Delta x_{i,j})^2 + (\Delta y_{i,j})^2 + \gamma(\Delta \mathrm{rate}_{i,j})^2. \tag{6}$$

The goal weighting of the characteristics coefficient in equation (6) is indicated by a, which can be adjusted to emphasize or mitigate the impact of a feature in the final metric results. β is the color weight coefficient. It helps the algorithm to determine the importance of color for target recognition and matching when features are fused. χ denotes the x-axis area weighting coefficient, which can influence the algorithm's emphasis on the target width. γ indicates the y-axis aimed feature weight coefficient, which reflects the influence of the target's features in the vertical direction. The study proposes a multi-target recognition process, as illustrated in Figure 2.

In Figure 2, first, the current frame image is subjected to background detection, while the background is continuously updated. The background updating method used in the study is the optical flow algorithm, and its updating principle is as follows: If the optical flow value of a pixel is small (i.e., the motion is not obvious), it is considered a background pixel and the background model is updated. If the optical flow value is large, it is considered a foreground target and the background model is not updated. Second, the difference between the two frames is calculated to determine if the amount of variation is greater than a set threshold. If it is smaller than the threshold, it indicates that there is no motion target in the video frame, and the result is output directly. If it is larger than the threshold, target recognition is performed. Meanwhile, the detected target is marked. Finally, the result is output. In the process of target recognition, the system will mark the suspicious moving target and compare the previously identified box with the real-time captured information to quickly identify the criminal suspect. Finally, the identification results are output to ensure the speed and accuracy of re-identification of criminal suspects in complex scenes, so that the security monitoring system can respond in time and take appropriate action.

3.2 Improved multi-target recognition model based on the KF algorithm

The multi-target recognition algorithm constructed above has a good recognition effect in stationary or slow-moving

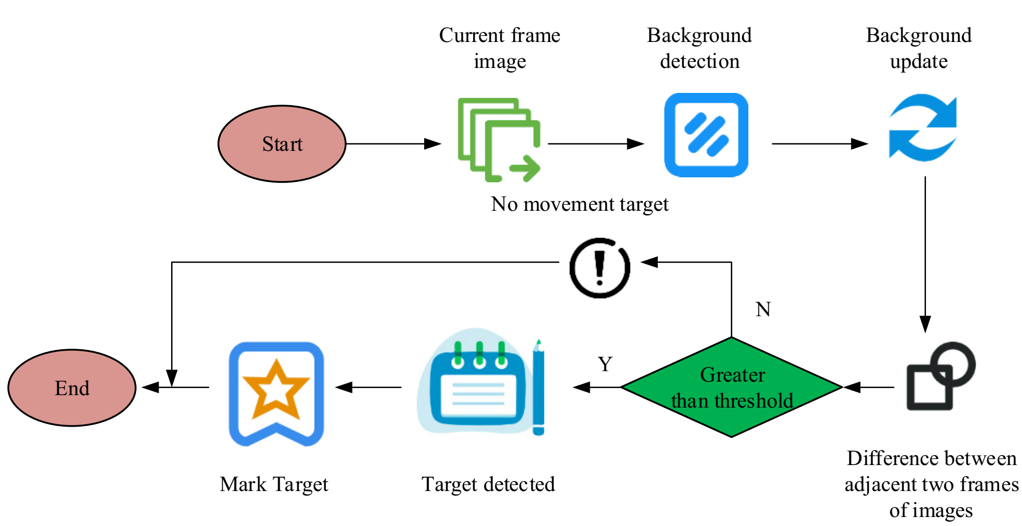

Figure 2: Feature matching-based multi-target tracking algorithm flow chart.

targets. However, when dealing with fast-moving suspects, the method based on inter-frame difference may lose the target, resulting in tracking instability. When multiple targets block each other, feature matching can fail, making it impossible to properly identify or track the blocked suspect. Therefore, to solve the above problems, the KF algorithm is introduced in this study to improve the recognition model. The KF algorithm is an algorithm for estimating the state of a linear dynamic system that adopts a recursive form to deal with noise and uncertainty [13,14]. In fast-motion target recognition, the KF algorithm can maintain accurate estimation of the target state under fast motion or instantaneous transformation of the target, thus improving the robustness and accuracy of tracking [15,16]. In dealing with the occlusion problem, the KF algorithm reduces the possibility of target loss by predicting the motion trajectory of the target. The flow of the recognition model introduced by the study into the KF algorithm is illustrated in Figure 3.

In Figure 3, the target recognition model first determines the detection target position and calculates the detection target position. Then, the detection target features are fused, and trajectory prediction and target tracking are realized by the KF algorithm. Finally, the target recognition position is obtained and the detection result is outputted. By introducing the KF algorithm, the model can not only quickly recognize and track suspects in complex scenes but also effectively deal with challenges such as fast movement and occlusion. This provides more reliable technical support for public safety. The tracking of the motion target state by KF is affected by the random noise, so it is necessary to determine the tracking state of the target first. The expression for the tracking state of KF is as equation (7) [17].

$$x_k = A\sigma_{k-1} + Bu_{k-1} + w_{k-1}. \tag{7}$$

In equation (7), A denotes the transfer matrix. σ_{k-1} reflects the destination state's noise value of A. B denotes the matrix of status control. u_{k-1} denotes the target state's noise value in the control matrix. w_{k-1} denotes the random noise value. Following the determination of the tracking, the feature extraction approach can be used to find the storage of the model tracking. The observation equation is displayed as equation (8).

$$z_k = H\sigma_k + v_k. \tag{8}$$

In equation (8), H denotes the matrix of observations. v_k denotes the observation disturbance. Due to the lengthy process involved in determining the multi-target state, the study uses the covariance of the disturbance and the observation noise to streamline the process. The tracking effect is then taken into consideration by evaluating the inaccuracy at step k of the tracking procedure. The estimation error's expression is provided in equation (9).

$$\begin{cases} \bar{x} = A\bar{x}_{k-1} + Bu_{k-1} \\ p_{\bar{k}} = AP_{k-1} + Q. \end{cases} \tag{9}$$

In equation (9), $\bar{x}_{k-1}$ denotes the outcome of the last observation in tracking. $\bar{x}$ denotes the observation result of the corresponding moment. P_{k-1} denotes the prediction of the last time. Q denotes the state noise covariance value. In the process of rapid identification of criminal suspects, after obtaining the tracking state of multiple targets, it is necessary to use the observation results to judge the error between the tracking state and the actual observation value. After obtaining the tracking state of multiple targets, it is also necessary to use the observation result pairs to determine whether the tracking state and the actual observation value differ in any way. Therefore, it can get the

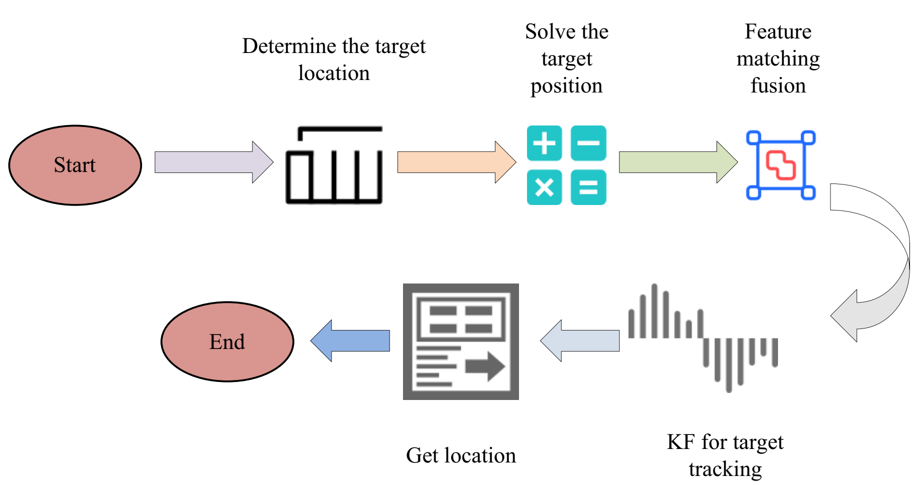

Figure 3: Flowchart of the improved recognition model based on KF.

corrected state estimate and noise value. This process is of filtering the noise using the KF algorithm, as displayed in Figure 4.

In Figure 4, the noise filtering of the KF algorithm is mainly divided into two parts. The first part is prediction, including target state value prediction and estimation error prediction. The second part is the correction part, including the calculation of Kalman gain, correction gain, and correction of estimation error prediction. The motion state of the crime target in the investigation video is high speed or irregular. Therefore, the common first-order motion model cannot complete the observation of the whole state. This work presents a second-order motion model. Moreover, the KF algorithm is used to predict the target in the second-order type motion model to obtain the relevant motion features. Then, it is successfully combined with feature extraction techniques [18]. The study assumes that the tracked motion target is in a motion rectangle at a specific tracking instant. At this point, the tracked motion target should move uniformly in both the horizontal and vertical directions, so that the motion state satisfies equation (10).

$$\begin{cases} V_x(t) = V_x(t-1), \\ V_y(t) = V_y(t-1). \end{cases} \tag{10}$$

In equation (10), $v_x(t)$ denotes the speed of motion at moment t. The state transfer of KF can be ascertained through gazing at the consistent motion during tracking, which can be expressed in equation (11).

$$x_z = Ax_{k-1} + w_{k-1}. \tag{11}$$

In equation (11), x_z denotes the KF transfer state value, when the corresponding transfer matrix can be expressed by equation (12).

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

If the state transfer equations satisfy this matrix at the same time, they can only be used as direct measurements. The measurement equation is shown in equation (13).

$$z_c = Hx_k + v_k. \tag{13}$$

In equation (13), z_c denotes the value of the obtained measurement equation, when the observation matrix can be expressed in equation (14).

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{14}$$

Through the analysis of criminal investigation video, it is found that most of the cases in the video have obvious object occlusion. It includes crowd occlusion, obstacle occlusion, and so on. This kind of occlusion can have a great impact on multi-target tracking. Through the above research, it is found that when the tracking target state is analyzed, if the occlusion phenomenon occurs, the image will disappear. After some time, if the obscured picture is blended with the target, it can be concluded that the image has vanished entirely. The occluded image is tracked, identified as a new tracking target, and matched with a new set of features when it separates from other targets and reappears in the video rectangle box [19,20]. Aiming to support real-time target tracking, the study presents the KF method based on image feature matching and uses the properties of both to anticipate the position information of the

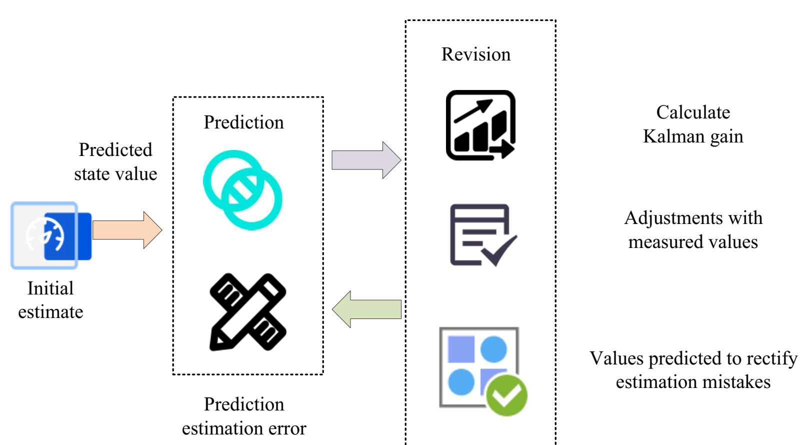

Figure 4: Noise filtering with KF algorithm.

occluded target. Figure 5 shows the flowchart for multi-target tracking of the model.

In Figure 5, the model analyzes the motion region-target relations for motion target recognition in video images. The motion region target relation is mainly divided into five kinds of targets. The first one is that the target already exists, and the model updates the tracking region. The model creates a new tracking region when the new target is detected, which is the second. The third one is that the region split is detected, and the model performs the segmentation processing. The fourth one is that the region merge is detected, and the model carries out the merge processing. The fifth is that the region disappearance is detected, and the model performs the disappearance processing. Finally, the model aggregates the five processing methods and outputs the target tracking results.

4 Results

4.1 Evaluation of multi-target recognition models' performance

The study is to verify the feasibility of the multi-target recognition model based on video tracking technology and the KF algorithm, using simulation experiments to conduct relevant performance experiments and comparison experiments. The specific environment of this experiment is as follows: operating system (Windows 10), GPU (RTX1080Ti), CPU (Inter i3), memory stick (2*8GB), and hard disk (SSD 500GB). The experimental data are collected using the crime and criminal tracking dataset, which contains a variety of videos related to criminal behavior. Behavior in the video can include theft, assault, burglary, and other typical crimes. Each video contains detailed annotations and labels. This is used to identify the particular crime, the participants in the crime, when the crime occurred, and the context in which the crime occurred. This provides comprehensive background information for algorithm training. Meanwhile, the study uses the dense trajectory (DT) algorithm and the multi-scale spatio-temporal graph convolution (MSSGC) algorithm for comparative analysis. Figure 6 shows the results of the comparison of the three approaches in terms of the error in the horizontal and vertical axes of the image.

Figure 6 (a) shows the tracking error of different algorithms on the horizontal axis. The results show that the tracking inaccuracy of the research proposed method is 3.75% on average, the inaccuracy of the DT algorithm is 7.94%, and the tracking inaccuracy of the MSSGC is 5.83%. Figure 6 (b) represents the inaccuracy of different algorithms on the vertical axis. The results show that the inaccuracy of the research proposed method is 3.27% on average, the tracking error of the DT algorithm is 6.12%, and the tracking error of the MSSGC is 5.06%. The results show that the model can maintain a good track tracking ability in the face of fast-moving suspects, such as fleeing in public places, which helps to timely capture more relevant information and make effective responses. This is because the DT algorithm has a relatively single feature description for the target, focusing mainly on motion information and not fully considering other features such as target color

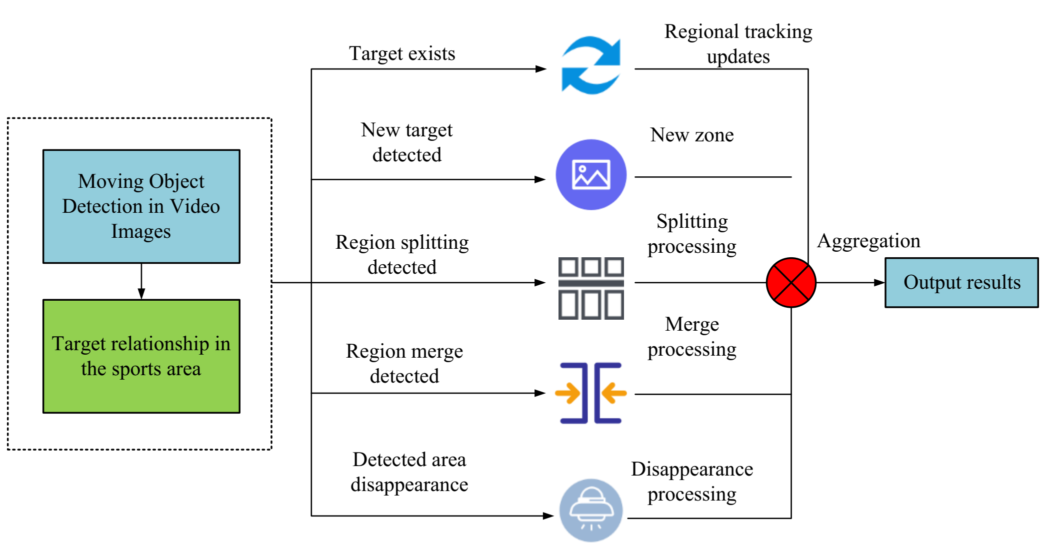

Figure 5: Multi-target recognition model based on video tracking with the KF algorithm.

and texture. The MSSGC algorithm has relatively insufficient continuity and accuracy in tracking fast-moving targets. The proposed algorithm integrates various feature information, such as shape, color, and texture, of the target and can dynamically adjust the weight of each feature. This multi-feature fusion method enables the model to comprehensively capture the characteristics of the target from multiple dimensions, thereby improving the accuracy of detection and tracking. The results of precision and recall during the tracking of the target by the model are shown in Figure 7.

Figure 7 (a) represents the accuracy results of different methods for target tracking in video. The outcomes demonstrate that the reliability of video tracking of the three models increases gradually, given the rise in the quantity of training. The highest accuracy of the proposed model is 94.75%, the highest accuracy of the DT algorithm is 79.27%, and the highest accuracy of the MSSGC algorithm is 82.20%. The high accuracy means that the model is highly reliable in identifying and confirming the location of criminal suspects. This is essential for the rapid detection of criminal incidents, reducing false targeting and ensuring that suspects are alerted and captured quickly and effectively in complex situations. Figure 7 (b) shows the recall results of different methods for target tracking in video, and the results show that as the number of model training increases, the recall of the model becomes better. The recall of the proposed model is 96.59%, the highest precision of the DT algorithm is 84.58%, and the highest precision of the MSSGC algorithm is 90.18%. The high recall rate demonstrates the model's sensitivity to criminal suspects in practical applications. For example, the model can not only identify currently visible suspects but also effectively track individuals moving in obscured or complex backgrounds, ensuring more comprehensive detection and prevention efforts. The dataset is used to validate the performance of several models. Figure 8 displays the details.

Figure 8 (a) represents the performance results of the DT algorithm in the dataset. Among them, the DT algorithm's accuracy is 84.57% inside the examine set, and the loss rate is 8.27%. The correctness in the validation set is 82.86% and the loss rate is 8.91%. Figure 8(b) represents the MSSGC algorithm's performance outcomes in the dataset. Among them, the accuracy in the training set is 92.74% with a loss rate of 5.27% and the accuracy in the validation set is 90.38% with a loss rate of 5.83%. Figure 8(c) represents the performance results of the research algorithm in the dataset. Among them, the accuracy in the training set is 95.28% with a loss rate of 2.44%, and the accuracy in the validation set is 93.35% with a loss rate of 2.64%. The results showed that the proposed model performed well in both precision and recall, demonstrating its effectiveness in the target tracking task. The reason why the proposed method of the study has significant advantages may be due to the fact that the model improves target recognition by recognizing multiple targets and fusing multiple target features. It also incorporates state prediction techniques such as KF, which enables the model to maintain efficient tracking capabilities in fast dynamic environments. These results show that when tracking criminal suspects, the research model can deal with interference such as occlusion and cross in the face of complex

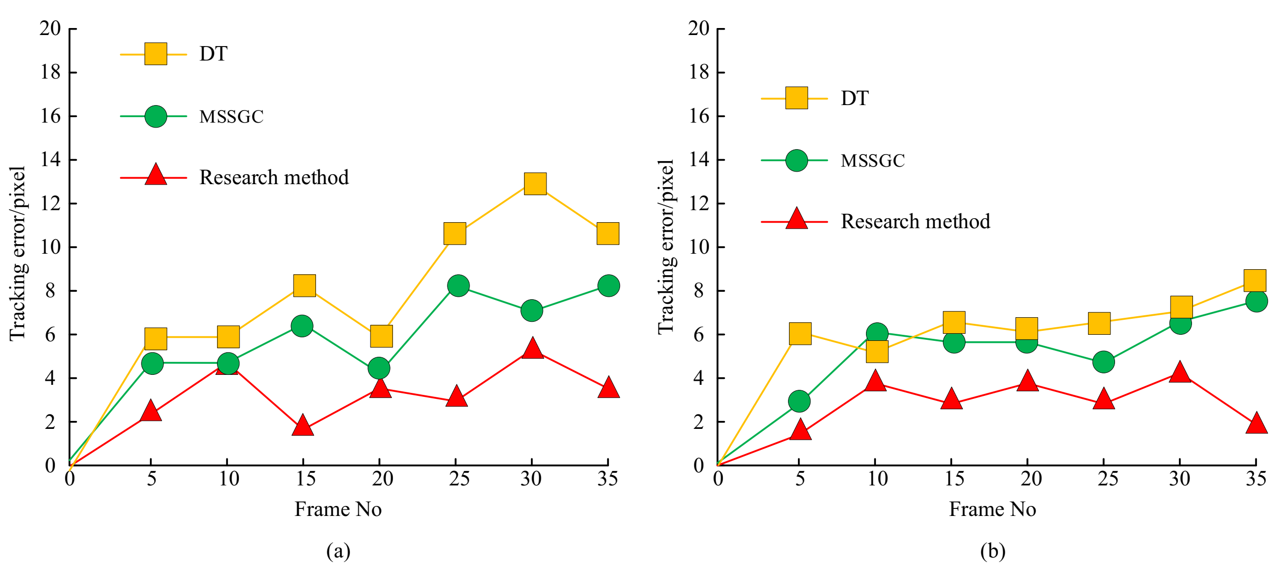

Figure 6: Comparison of the photos' vertical and horizontal axis inaccuracies. (a) *X*-axis. (b) *Y*-axis.

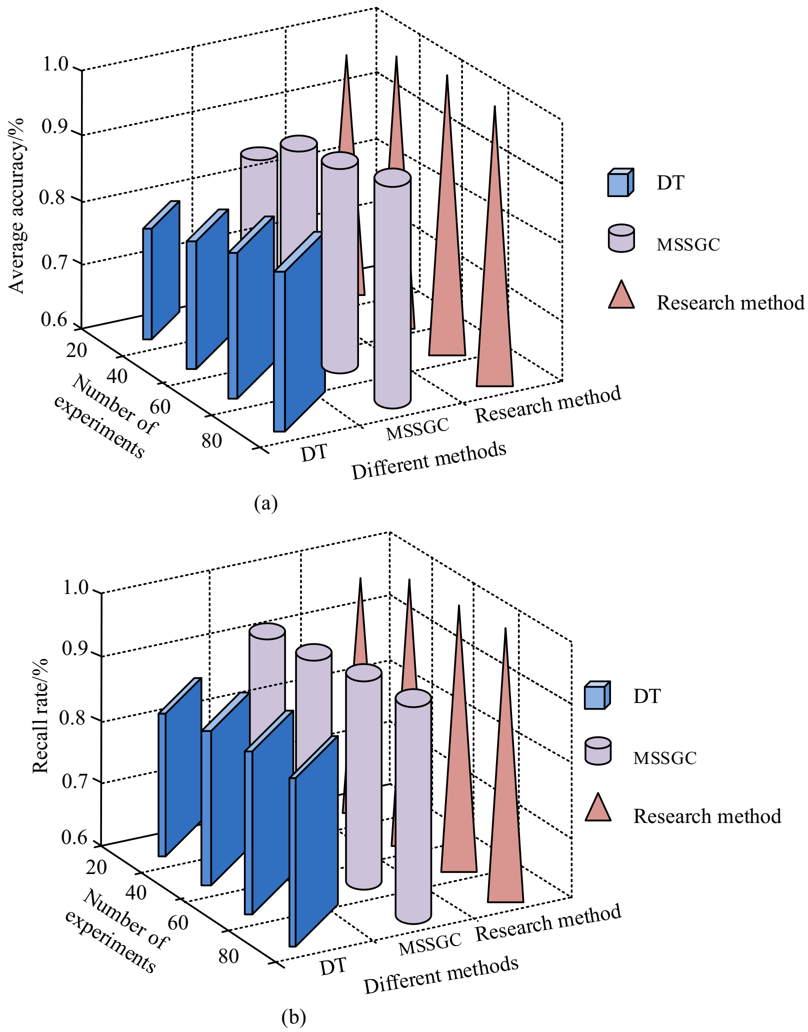

Figure 7: Accuracy and recall of target tracking in a video. (a) Average accuracy. (b) Recall rates.

scenes and continuously maintain effective tracking of the target. It greatly improves the professional ability of the security monitoring system in dealing with criminal behavior.

4.2 The application effect of multi-target recognition model improved with the KF algorithm

To verify the effect of the model method in practical application, the identification and tracking performance of the model is verified by a criminal investigation video. The criminal investigation video often has the characteristics of fast moving and multi-cover and involves the movement and interaction of many characters. This makes the crime investigation video an effective test of the performance of the model in complex scenes. It can also verify the robustness and adaptability of the model. At the same time, it can also verify the application effect of the model in crime investigation. The study is comparatively analyzed by the MSSGC algorithm. The results of the model's accuracy–recall (AR) curve are shown in Figure 9.

Figure 9 (a) represents the result of the comparison of the AR curve of the two models. The AR is between 0 and 1, and the larger the AR, the more discriminative the model is. Moreover, it can better separate the positive and negative samples. The value of the AR curve of the MSSGC algorithm is 0.87, and the value of the AR curve of the proposed model of the study is 0.94. The conclusion suggests that the proposed model of the study is more discriminative in comparison with the MSSGC algorithm. Figure 9 (b) represents the results of the feature checking rate of the two models. The results show that the feature checking rate of the method put forward is 94.83% and the feature checking rate of the MSSGC algorithm is 87.46%.

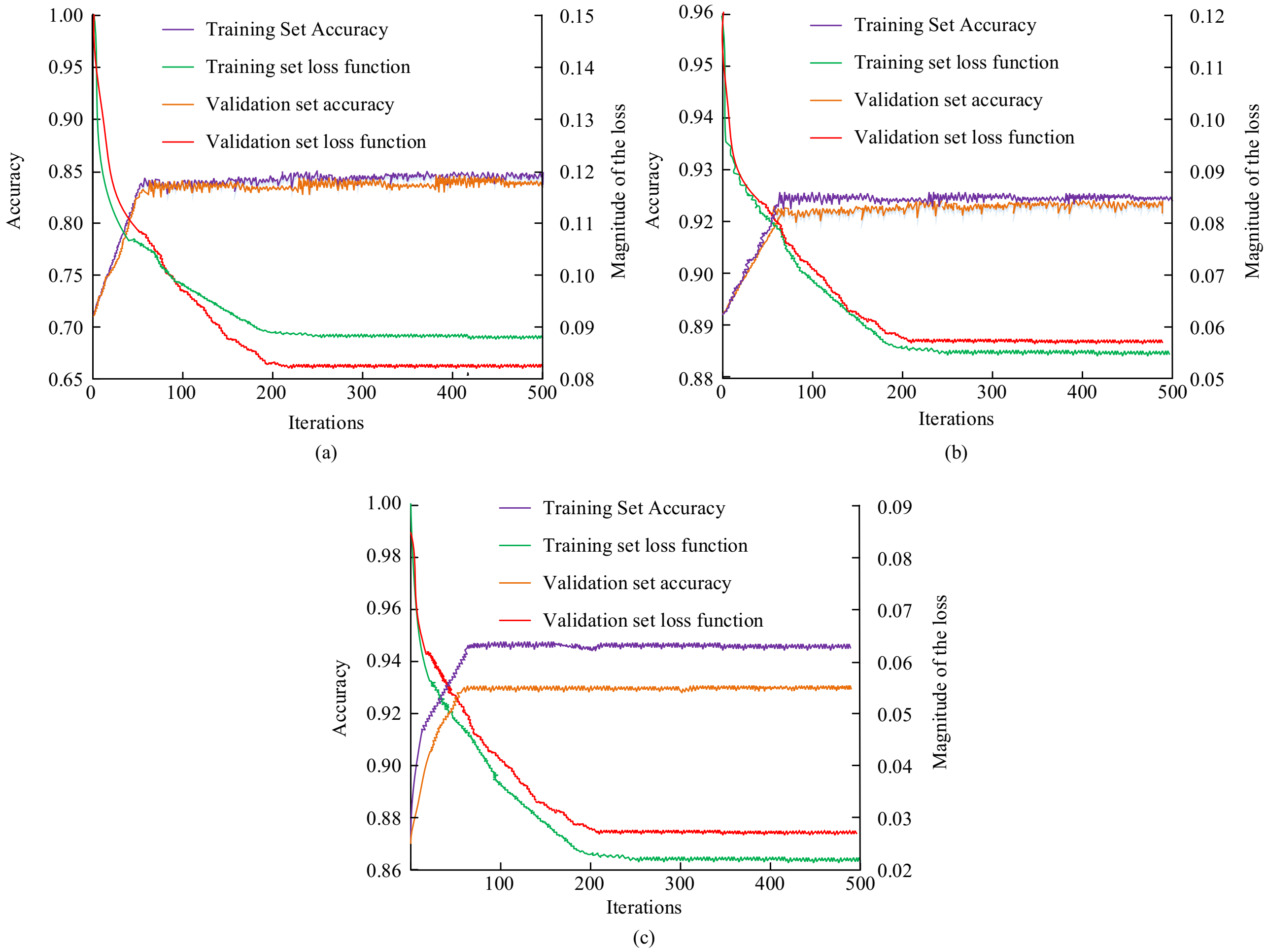

Figure 8: Performance validation of different algorithms in datasets. (a) DT algorithm performance results. (b) MSSGC algorithm performance results. (c) The performance results of the method were studied.

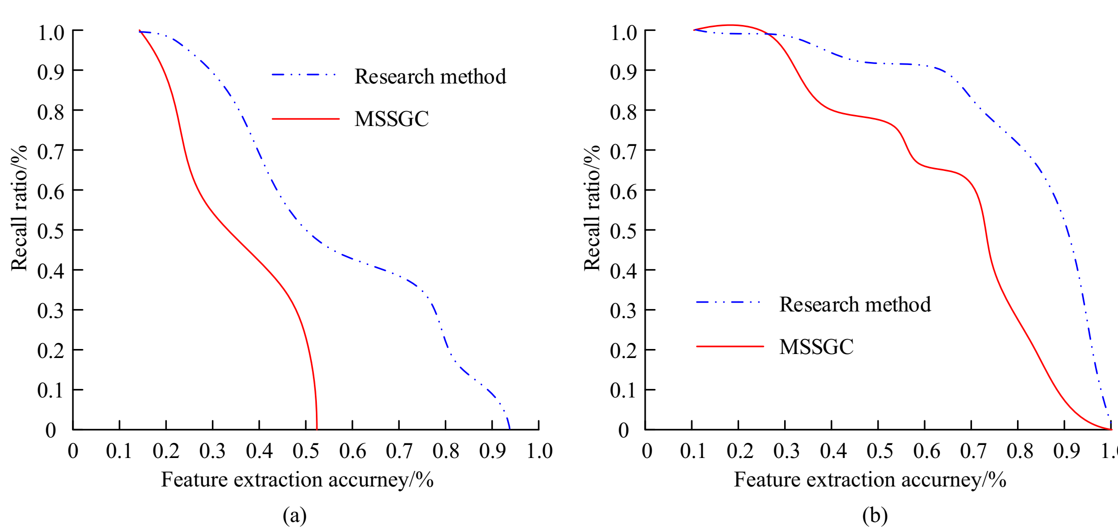

Figure 9: AR curve results. (a) AR curves for feature extraction of both techniques. (b) Comparison of the two characteristics.

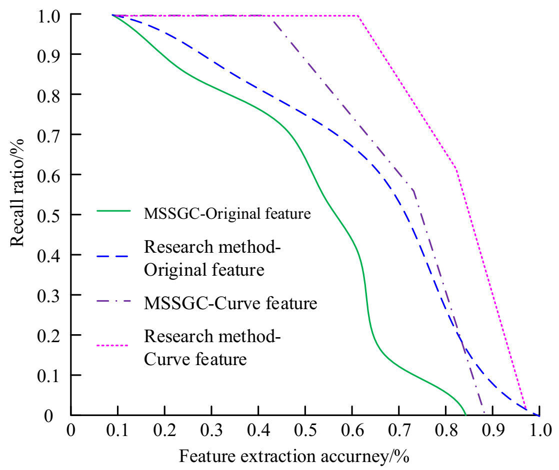

Figure 10: Target person tracking performance results.

The results show that the study has better feature extraction and good classification under multi-target recognition. The tracking performance test results of the model in the crime investigation video are shown in Figure 10.

Figure 10 shows the tracking performance results of the algorithm in the crime investigation video, which mainly includes the feature extraction of the original data of the target character and the feature extraction of the figure shape curve. The raw data feature extraction ability of the MSSGC algorithm is 85.31%, and that of the proposed algorithm is 89.86%. In curve feature extraction, the extraction ability of the MSSGC algorithm is 91.36%, and the extraction ability of the proposed algorithm is 95.12%. The proposed algorithm introduces the KF algorithm and combines historical frame information and current frame data to dynamically predict the motion state of the target. The KF algorithm can effectively handle noise and uncertainty, which improves the tracking stability of the model under fast target motion and occlusion. By predicting the motion trajectory of the target, the KF algorithm can reduce the possibility of target loss, thereby maintaining high tracking accuracy in complex scenes. The research applies the model method to four videos and verifies the tracking effect of the model method on the target person's actions by comparing the results of the number of annotations, correct detection, incorrect detection, missing detection, accuracy rate, and recall rate, as shown in Figure 11.

Figure 11 (a) shows the tracking and recognition performance of the model on people's actions in crime investigation videos. Among them, the number of wrong actions detected by the model in the four videos is 40, and the average number of missed detections is 38. Figure 11(b) represents the results of the model's check-accuracy and check-completeness rates in the four videos, in which the average check-accuracy rate is 82.57% and the average check-completeness rate is 84.46%. The results show that the research approach provides a superior feature extraction method that combines the target motion characteristics with the shooting action characteristics, thus improving the accuracy of action recognition. The model combines KF, which enhances the prediction precision and stability of the model in a fast dynamic environment, and improves the reliability of the detection results. In the target person action detection index, the detection effect of the research model is shown in Table 1.

The findings in Table 1 indicate that the detection effect is significantly improved in all cases where the

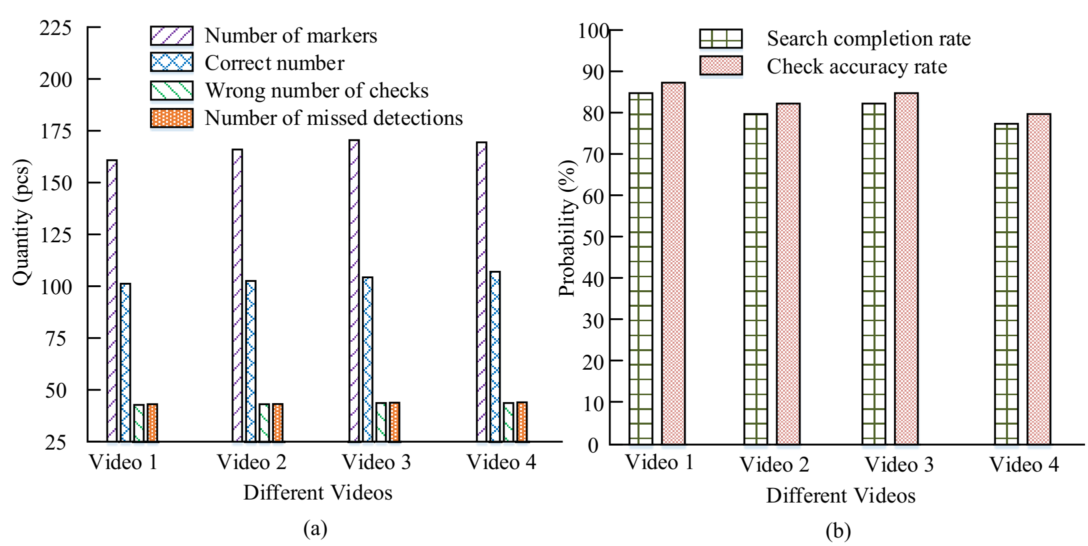

Figure 11: Results of character action tracking in crime detection video. (a) Target person motion detection. (b) Model accuracy and recall results.

Table 1: Abnormal behavior characterization test results

Type of crime	Algorithm	Sample size			
		40	60	80	100
Steal	MSSGC	0.799	0.818	0.855	0.917
	Research method	0.821	0.865	0.911	0.933
Tussle	MSSGC	0.728	0.835	0.864	0.900
	Research method	0.843	0.860	0.925	0.942
Escape gait	MSSGC	0.705	0.826	0.860	0.905
	Research method	0.802	0.870	0.925	0.933
Shoot	MSSGC	0.769	0.808	0.863	0.904
	Research method	0.835	0.873	0.930	0.938

number of samples is increased. Meanwhile, the research proposes that the recognition accuracy of the emanation is better than the MSSGC algorithm in each type of foul play. The highest feature detection accuracy of the MSSGC algorithm is 0.917, and the highest detection accuracy of the research method is 0.942. The results show that the research method's ability to learn and generalize can help the model identify and distinguish between different forms of foul play. The results show that the learning ability and generalization capability of the research method can help the model identify and discriminate different types of foul behavior and enhance the ability to capture various criminal behaviors. This makes it possible to effectively track and identify suspects in complex scenarios, thus helping the police to intervene and deal with potential security threats in a timely manner.

5 Discussion

The proposed multi-target identification technique was in accordance with matching features in videos and the KF algorithm, and through the fusion of multiple features, the model's average tracking inaccuracy in the horizontal and vertical axes was only 3.75 and 3.27%, respectively. The results demonstrated that the model could gather target information from many dimensions by fusing the target's color, shape, and motion state properties. With the introduction of the KF algorithm, the model not only relied on the image data of the current frame but also combined the information of the historical frames to predict the target position. In the application of crime detection video, the research showed that the feature recall rate of the proposed model was 94.83%, and the AR curve was 0.94. The results showed that the research model made full use of multiple information sources in the dynamic scene, which made it possible to maintain high recognition performance in fast-moving environments. Zhao et al. proposed a motion recognition model based on the DT algorithm, which relied on the motion path and the dense samples, with a maximum accuracy of 79.27% in the study [19]. However, compared to the proposed model, the DT algorithm has limitations in handling multi-feature fusion and dynamic environment adaptation, which may lead to insufficient tracking accuracy in complex scenes. Chen et al. proposed a multi-target recognition model based on the MSSGC algorithm, which had certain advantages in spatio-temporal map convolution. However, the tracking continuity and accuracy were relatively insufficient when dealing with highly dynamic targets [20]. This may be due to the algorithm's difficulty in effectively integrating historical and current information when dealing with rapidly changing target states. In general, the multi-target recognition method proposed by the research, based on video feature matching and the KF algorithm, fully considered the fast-moving characteristics of criminal suspects in complex scenes and demonstrated strong real-time tracking and accurate recognition capabilities. By deploying surveillance cameras in public places, such as shopping malls, train stations, and airports, the model can monitor suspicious behaviors such as stalking and wandering in real time and issue timely alerts. Although the proposed model has performed well in experiments, there are still a number of challenges to be overcome in practical applications. For example, when deploying models on resource-constrained devices, there may be a problem of insufficient processing power. In addition, it can be difficult to seamlessly integrate the model with existing monitoring systems and user interfaces. Therefore, to ameliorate the above problems, the use of a lightweight model architecture to accommodate resource-constrained environments is considered for the future. Meanwhile, a modular architecture that is easy to integrate and develop an intuitive user interface is considered to be designed.

6 Conclusion

With the increasing demand for rapid identification of criminal suspects in areas such as urban security, crime detection, and dynamic surveillance, traditional methods often face difficulties in complex environments. The study proposed a multi-target recognition approach that combined the KF algorithm with video tracking technologies, which combined the video feature matching algorithm to further improve the performance of multi-target recognition and tracking. Through comparative experiments and practical application validation, the study indicated that the model effectively improved the recognition ability of fast-moving suspects in the feature fusion strategy. It could better cope with complex situations such as fast movement and occlusion, enabling it to better cope with high-speed movement and occlusion. The dynamic prediction function of the KF algorithm improved the tracking stability and accuracy in high noise and uncertainty environments. Despite the good results of the study, there are still some limitations. For example, in extreme occlusion situations, the recognition rate of the model may be affected. Despite the introduction of the KF state prediction algorithm, it is difficult to achieve persistent tracking in cases of prolonged occlusion or complete disappearance of the target. Therefore, the future research direction can consider introducing more efficient image segmentation technology in order to identify and recover obscured criminal suspects, so as to further improve the model's recognition and tracking ability in complex scenes.

Funding information: The research received no external funding.

Author contributions: Wenbo Fu and Zhihui Ban provided the concept and wrote the draft and Xing Li and Yingao Han revised this paper critically. All authors reviewed this article carefully and approved this submission.

Conflict of interest: The authors declare that there is no conflict of interest.

Data availability statement: All data generated or analyzed during this study are included in this article. Further enquiries can be directed to the corresponding author.

References

[1] D. Cheng, J. Qian, X. Guo, Q. Kou, F. Xu, J. Gu, et al., "Review on key technologies of AI recognition for videos in coal mine," *Coal Sci. Technol.*, vol. 51, no. 2, pp. 349–365, 2023.

[2] C. Hebbi and H. Mamatha, "Comprehensive dataset building and recognition of isolated handwritten kannada characters using machine learning models," *Artif. Intell. Appl.*, vol. 1, no. 3, pp. 179–190, 2023.

[3] L. Yan, Y. Shi, M. Wei, and Y. Wu, "Multi-feature fusing local directional ternary pattern for facial expressions signal recognition based on video communication system," *Alex. Eng. J.*, vol. 63, no. 3, pp. 307–320, 2023.

[4] S. Hu, K. Shimasaki, M. Jiang, T. Senoo, and I. Ishii, "A simultaneous multi-object zooming system using an ultrafast pan-tilt camera," *IEEE Sens. J.*, vol. 21, no. 7, pp. 9436–9448, 2021.

[5] H. Zheng, Y. Bai, and Y. Tian, "A multi moving target recognition algorithm based on remote sensing video," *Comput. Model. Eng. Sci.*, vol. 134, no. 1, pp. 585–597, 2023.

[6] Y. Hou, Z. Wang, S. Wang, and L. Zheng, "Adaptive affinity for associations in multi-target multi-camera tracking," *IEEE Trans. Image Process.*, vol. 31, no. 10, pp. 612–622, 2021.

[7] N. Lakshmi and M. P. Arakeri, "A novel sketch based face recognition in unconstrained video for criminal investigation," *Int. J. Elect. Comput. Eng.*, vol. 13, no. 2, pp. 1499–1509, 2023.

[8] J. Li, Y. Wang, G. Fang, and Z. Zeng, "Real-time detection tracking and recognition algorithm based on multi-target faces," *Multimed. Tools Appl.*, vol. 80, no. 1, pp. 17223–17238, 2021.

[9] E. Kurtoğlu, S. Biswas, A. C. Gurbuz, and S. Z. Gurbuz, "Boosting multi-target recognition performance with multi-input multi-output radar-based angular subspace projection and multi-view deep neural network," *IET Radar, Sonar Navig.*, vol. 17, no. 7, pp. 1115–1128, 2023.

[10] X. Wang, X. Sun, and Z. Wang, "Construction of visual evaluation system for building block night scene lighting based on multi-target recognition and data processing," *IET Circ., Devices Syst.*, vol. 17, no. 3, pp. 149–159, 2023.

[11] Z. Zhang, "Based on YOLO v3 target recognition algorithm as vehicle tracking algorithm analysis," *Front. Comput. Intell. Syst.*, vol. 5, no. 2, pp. 140–144, 2023.

[12] J. Zhang, K. Chen, and Y. Ma, "Multi-target recognition utilizing micro-doppler signatures with limited supervision," *IEICE Trans. Electron.*, vol. 106, no. 8, pp. 454–457, 2023.

[13] M. Á. Naya, E. Sanjurjo, A. J. Rodríguez, and J. Cuadrado, "Kalman filters based on multibody models: Linking simulation and real world. A comprehensive review," *Multibody Syst. Dyn.*, vol. 58, no. 3, pp. 479–521, 2023.

[14] A. Murakami, K. Fujisawa, and T. Shuku, "Developments of inverse analysis by Kalman filters and Bayesian methods applied to geotechnical engineering," *Proc. Jpn. Acad., Ser. B*, vol. 99, no. 9, pp. 352–388, 2023.

[15] O. Al-Ghattas, J. Bao, and D. Sanz-Alonso, "Ensemble Kalman filters with resampling," *SIAM/ASA J. Uncertain. Quantif.*, vol. 12, no. 2, pp. 411–441, 2024.

[16] F. Bakhshi Ostadkalayeh, S. Moradi, A. Asadi, A. M. Nia, and S. Taheri, "Performance improvement of LSTM-based deep learning model for streamflow forecasting using Kalman filtering," *Water Resour. Manage.*, vol. 37, no. 8, pp. 3111–3127, 2023.

[17] L. Winiwarter, K. Anders, D. Czerwonka-Schröder, and B. Höfle, "Full four-dimensional change analysis of topographic point cloud time series using Kalman filtering," *Earth Surf. Dyn.*, vol. 11, no. 4, pp. 593–613, 2023.

[18] B. Shen, X. Wang, and L. Zou, "Maximum correntropy Kalman filtering for non-Gaussian systems with state saturations and stochastic nonlinearities," *IEEE/CAA J. Autom. Sin.*, vol. 10, no. 5, pp. 1223–1233, 2023.

[19] G. Zhao, Z. Chen, and W. Liao, "Reinforcement-tracking: An end-to-end trajectory tracking method based on self-attention mechanism," *Int. J. Automot. Technol.*, vol. 25, no. 3, pp. 541–551, 2024.

[20] G. Chen, X. Chen, C. Zheng, J. Wang, X. Liu, and Y. Han, "Spatiotemporal smoothing aggregation enhanced multi-scale residual deep graph convolutional networks for skeleton-based gait recognition," *Appl. Intell.*, vol. 54, no. 8, pp. 6154–6174, 2024.