1. Introduction

The video image is an important way for people to understand the objective world. With the rapid development of electronic technology and Internet technology, image processing technology has attracted the attention of researchers in various fields as a convenient and practical technology. In recent years, with the popularity of smartphones equipped with high-definition cameras and video surveillance systems, people's demand for intelligent video analysis is increasing [1, 2]. We hope to use computers instead of human brains to complete complex tasks such as video classification, intelligent monitoring, and so on, in order to improve the convenience of daily life. This requires the computer to be able to understand and process such visual information, and the tracking algorithm is the most basic and important part of it. In short, the goal of tracking is to analyze and predict the motion path of the object to be measured in the video and to determine its specific position in each frame [3, 4].

Effective detection and tracking of objects in video images is a very important branch of technology in the field of modern computer vision, which has a broad development prospect in many fields such as security monitoring, logistics management, automatic navigation, military investigation, criminal suspect locking, and so on. Especially in security monitoring, military reconnaissance, criminal suspects lock, etc., of multiobjective pedestrians on the purpose of effectively detecting and tracking is the target of pedestrian-related movement data obtained, such as the movement of the target location, target track, and target's movement speed and direction and goal of the activity area, so as to realize the analysis of the moving object behavior, understand, or in order to do more advanced tasks.

In order to improve the tracking accuracy of multitarget pedestrians, many researchers have carried out a lot of research on target tracking algorithms. At present, the main algorithms used in the field of target tracking include particle filtering algorithms, correlation filtering algorithms, and so on. In addition, a multitarget tracking algorithm based on the combination of adjacent frame matching and Kalman filtering is proposed in Reference [5]. In this algorithm, the pedestrian head is detected by the AdaBoost classifier with the direction gradient histogram (HOG) as the feature; then, the pedestrian trajectory is predicted by the Kalman filter and the adjacent frame matching algorithm is used to correct the target trajectory. In the adjacent frame matching
method, the target matching problem in the tracking process is transformed into an assignment problem based on the Hungarian algorithm. Reference [6] proposed a multitarget tracking algorithm based on YOLOV3 and Kalman filtering. The algorithm uses YOLOV3 to detect the target to be tracked in the current frame and uses the Kalman filter to predict its next position and the size of the bounding box according to the position of the current target. The improved Hungarian algorithm is used to carry out data correlation and matching according to the intersection ratio of the detection frame and prediction frame and color histogram, and the tracking is completed through continuous iteration of the system to obtain the target’s motion trajectory [7–9]. For the occluded target, a region-based quality assessment network is introduced, and multiframe high-quality detection images are combined to recover the occluded part and improve the tracking accuracy.

However, with the development of modern video technology and the increasing complexity of pedestrian movement background, the abovementioned traditional algorithm has the problems of a high error rate of pedestrian feature recognition and poor tracking effect. To solve this problem, a new multitarget pedestrian multiscale stable tracking algorithm is designed based on a contour template [10, 11]. The specific ideas of this paper are as follows:

First, the template voting strategy is adopted and according to the regional characteristics of pedestrian actions, the pedestrian template is divided into multiple contour areas.

Second, the feature similarity thresholds in different regions are divided by regional feature similarity classification to complete the classification of multitarget pedestrian features.

Then, the score function is selected as the matching criterion between the template and the candidate samples, and the multicontour template is matched according to the similarity of deformation diversity. Finally, the multitarget detection and tracking are realized based on the mean value.

Finally, the feature recognition effect of this method is verified by the accuracy of feature recognition, position tracking error, and tracking time.

2. Design and Matching of Contour Template

2.1. Contour Template Design. Multiscale feature areas of pedestrians mainly include the head, neck, shoulders, chest, back, abdomen, buttocks, upper limbs, and lower limbs. Different feature areas will be affected by different degrees of action.

Because the target object is different, the feature region affected by the action is also different, and the influence of different parts on the whole feature region is also different [12–15]. Therefore, a voting strategy based on multiple templates is applied in this article, according to the features of easily affected by the action area, the human body is divided into multiple outline template sections, calculating the similarity between the template area, and implementation of the match, again USES the voting statistics majority vote for final similarity of a large area of the matching results [16, 17]. The proposed method has good robustness of occlusion and tolerance of region division, so as to avoid the problem of too much dependence on feature points to complete region division and obtain several relatively reliable region classifiers at the same time. In this paper, the multiregion contour template is divided into 9, and the partition results of the multiregion contour template are shown in Figure 1.

2.2. Multiregion Profile Template Feature Similarity Classification. The feature similarity of the multiregion contour template is to classify the features in the divided multiregion contour template according to the similarity degree and compare the similarity between the well-classified regional features and the contours of different parts of the human body in the feature database. The higher the similarity value, the more similar the image features [18–22]. The similarity of feature information in nine multiregion contour templates is calculated, and the sum of Euclidean distance squares of coordinate points corresponding to regional semantics is taken as the similarity value of the two regions. The process is as follows:

\[
W = \sum_{j=1}^{N} \sqrt{\left( k(x_j, y_j, z_j) - h(x_j, y_j, z_j) \right)^2}. \tag{1}
\]

In formula (1), \( N \) is the number of templates; \( j \) is any point in the multi-region contour template; and \( k(x_j, y_j, z_j) \) and \( h(x_j, y_j, z_j) \) are points in the region to be tested and semantic corresponding points in the region in the feature library, respectively. The larger the similarity value is, the more similar the contour features in the two regions are. If the similarity value is greater than a certain threshold value, it is considered to be the same contour region, which is to complete the feature similarity classification of the multiregion contour template.

2.3. Contour Template Matching. The multiregion contour template was matched by feature similarity classification. Assuming that there is a point set \( P = \{p_i\}_{i=1}^{m} \) corresponding to the image template image and a point set \( Q = \{q_j\}_{j=1}^{n} \) corresponding to the image target, \( i, j, m, \) and \( n \) respectively, represent the point an ordinal number of the point set and the total number of the corresponding point set, the nearest neighbor formula of the multitarget pedestrian image target in the multiregion contour template is as follows:

\[
l_{mn}(q, P) = \arg \min P D(q, p). \tag{2}
\]

In formula (2), \( D(q, p) \) and \( l_{mn}(q, P) \), respectively, represent the distance function and the nearest neighbor of the point \( q \) in the point set \( P \) [22–27].

The purpose of calculating the feature similarity is to obtain the number of diversity of the nearest neighbor in the template image and the multi-objective pedestrian image. The calculation formula is as follows:
According to formula (3), the similarity between the multitarget pedestrian image and the multiregion contour template can be obtained, and the target deformation can be added to the above calculation to improve the matching performance of the contour template.

On this basis, $A$ and $k$ are used to represent the appearance model and spatial position coordinates, respectively, and the nearest neighbor formula of the appearance model in the point set $P$ can be obtained as follows:

$$p_i = R_m^A (q_j, P) \times k = \arg \min d(q^A_j, p^A) \times k.$$  

(4)

In formula (4), $d(q^A_j, p^A)$ is the model distance, and the number of nearest neighbors matching diversity of $q \in Q$ in point set $P$ is as follows:

$$\gamma_{p_i} = q \times R_m^A (q_j, P).$$  

(5)

Combined with the above equation, the formula can be obtained as follows:

$$\delta_{Q \rightarrow P} = \frac{|l_{mn}(q_j, P) \times p_i|}{\min[m,n]}.$$  

(3)

$$\delta = \sum_{q_j \in Q} \exp \left[ 1 - \xi \left| l_{nm}(q_j, P) \right| \right] / \gamma_{p_i} + 1.$$  

(6)

To sum up, contour template matching can be completed through the above process.

3. Multitarget Pedestrian Multiscale Stable Tracking Algorithm Design

On the basis of the above contour template matching, a multitarget pedestrian multiscale stable tracking algorithm is designed in this study. The process is as follows.

First, the candidate region where multiscale targets may exist in each frame of the multitarget pedestrian video image is determined. Then, the abovementioned steps are repeated to obtain the final histogram of the feature probability distribution of the candidate region at a multiscale, and the construction of the candidate model is completed [27–29]. Therefore, the multiscale candidate model is obtained as follows:

$$\omega = \frac{\delta (1 + g)^{-1}}{Z}.$$  

(7)

In formula (7), $Z$ and $g$ represent multiscale candidate regions and the corresponding label sets, respectively. Then, the similarity between the target model and the candidate model is calculated by using the similarity measurement method, and the maximum value of the similarity function Babbitt coefficient is selected as the mean vector of the target. This vector is the transition vector of the target from the initial position to the current position [30–33]. The search region moves continuously along the direction of the vector until the convergence condition is satisfied, and the real position of the multitarget pedestrian in the current video frame is determined, so as to achieve the goal of target tracking. The specific process is as follows:

Step 1: Read the video sequence of multiobjective pedestrians, obtain the video frame image to form the test set, and take it to the objective function $F$. The formula can be obtained as follows:

$$F = \omega H \frac{1}{\lambda}.$$  

(8)

In formula (8), $H$ represents the kernel function and $\lambda$ represents the target appearance model of the learned sample. Then, the inverse discrete Fourier transform is applied to $\hat{F}$, and the position of the pedestrian target is determined according to the result. Therefore, the test set and its weight are updated in real time to adapt to the change of pedestrian target in the tracking process

$$\tilde{\varphi}_t = \mu y + (1 - \mu) \alpha_{t-1}.$$  

(9)

In formula (9), $\mu$ and $\alpha_{t-1}$ represent the learning parameters and the weight matrix of the previous video frame, respectively; $\mu$ and $y$ represent the target template after the current frame is updated and the discrete
Fourier transform of the current frame target, respectively.

Step 2: Determine the initial frame image $I_0$, delimit the target area $S$, and determine the target center $y_0$.

Step 3: Build a multiscale target model with $y_0$ as the center, as follows:

$$q_u = \frac{I_0 \times S}{y_0}.$$  \hspace{2cm} (10)

Step 4: Initialize the candidate targets at the current frame center $y_0$ and then estimate the probability of the multiscale eigenvalues of the multitarget pedestrian contour template in the candidate region.

$$z[p(y_0), q] = \sum_{u=1}^{m} \sqrt{p_u(y_0)q_u}.$$  \hspace{2cm} (11)

Step 5: Calculate the weight $w_y$ of each pixel in the candidate target region [34, 35].

Step 6: Move $y_0$ to a new position, and record the position of the new target in the candidate region as $y_1$. Calculate the candidate target model $U$ with $y_1$ as the center and calculate as follows:

$$U_y = \sum_{u=1}^{m} \sqrt{w_y(y_1)q_u}.$$  \hspace{2cm} (12)

Step 7: Judge whether formula (13) is true as follows:

$$z[w_y(y_1), q_u] \leq z[w_y(y_0), q_u].$$  \hspace{2cm} (13)

If it is true, we set the target position to the appropriate area between the new position and the old position; otherwise, we proceed to the next step.

Step 8: Compare $z[w_y(y_1), q_u]$ with Kronecker delta function to obtain the following formula:

$$z[w_y(y_1), q_u] < \text{Kronecker Delta function}.$$  \hspace{2cm} (14)

If formula (14) is true, a new target location is found, and the algorithm stops to get the real pedestrian target location in the current frame; otherwise, return to step 6 and continue to find the candidate target location that meets the conditions. After several iterations, the real position of the target in the current frame is output.

Step 9: Repeat the above process until the real position of the pedestrian target in each frame is found, so as to achieve multiscale stable target tracking. The specific multiscale stable target tracking process is shown in Figure 2.

4. Experiment and Analysis

4.1. Experimental Environment Design. In order to verify the practical application performance of multitarget pedestrian multiscale stable tracking algorithm based on contour template, the following simulation experiments are designed to verify.

The experimental environment is as follows: the simulation experiment is carried out on the MATLAB platform, the operating system is windows 10, and the program development tool is Visual C++ 2008 express. The experimental database contains 500 pedestrian feature models with different postures and regions. According to the difficulty of facial feature tracking, the database can be divided into two categories as follows: the first category is facial feature videos captured from different angles, the second category is videos captured under completely arbitrary conditions, including severe occlusion, extreme lighting, and posture.
In order to avoid the singleness of the experimental results, the multitarget tracking algorithm based on the combination of adjacent frame matching and Kalman filter in reference [5] and the multitarget tracking algorithm based on yoloV3 and Kalman filter in reference [6] are used as comparison algorithms to complete the performance verification with the algorithm in this paper.

4.2. Performance Index

(1) Feature recognition accuracy: this indicator is used to measure the effect of different algorithms on multiscale multitarget pedestrian feature recognition. The higher the recognition error rate is, the more accurate the algorithm can capture pedestrian features.

(2) Pedestrian barycenter position tracking error: this index is used to measure the effect of different algorithms on multiscale multitarget pedestrian feature tracking. The lower the barycenter position tracking error is, the more accurate and timely the algorithm can track and capture the pedestrian position.

(3) Tracking takes time: This index is used to verify the timeliness of different algorithms. The less the tracking time is, the more efficient the algorithm is.

4.3. Result Analysis. First, the feature recognition accuracy of the multitarget pedestrian multiscale stable tracking algorithm based on contour template, the multitarget tracking algorithm based on the combination of adjacent frame matching and Kalman filter in reference [5], and the multitarget tracking algorithm based on yoloV3 and Kalman filter in reference [6] are verified. The results are shown in Figure 3.

Analysis of the results shown in Figure 3 shows that with the continuous change of frame scale, the accuracy of feature recognition of different algorithms also changes. The feature recognition accuracy of the algorithm in literature [5] is between 88% and 91%, the feature recognition accuracy of the algorithm in literature [6] is between 89.5% and 93%, and the feature recognition accuracy of the algorithm in literature [6] is slightly higher than that of the algorithm in literature [5]. The feature recognition accuracy of this algorithm is between 93% and 98%, which is significantly higher than the two contrast algorithms. With the change in video frame scale, the feature recognition accuracy of this algorithm gradually increases. Therefore, this algorithm has more application advantages, which shows that the algorithm can accurately capture the characteristics of multitarget pedestrians.

On this basis, the multitarget pedestrian multiscale stable tracking algorithm based on contour template, the multitarget tracking algorithm based on the combination of adjacent frame matching and Kalman filter in reference [5], and the application performance based on yoloV3 and Kalman filter in reference [6] are verified with the tracking error of pedestrian gravity center position as the index. The results are shown in Figure 4.

By analyzing the results shown in Figure 4, it can be seen that with the continuous change of frame scale, the tracking error of pedestrian gravity center position of different algorithms also increases. The tracking error of the pedestrian center of gravity position of reference [5] algorithm is always below 8%, the tracking error of the pedestrian center of gravity position of reference [6] algorithm is always below 10%, and the tracking error of pedestrian center of gravity position of reference [6] algorithm is slightly higher than that of reference [5]. However, the tracking error of the pedestrian gravity center is always below 5%, which is
significantly lower than the two comparison algorithms. Therefore, this algorithm has more application advantages, which shows that the algorithm can accurately and timely track and capture the position of pedestrians.

Finally, taking the tracking time as the index, the application performance of multitarget pedestrian multiscale stable tracking algorithm based on contour template, multitarget tracking algorithm based on the combination of adjacent frame matching and Kalman filter in reference [5], and the application performance based on yolov3 and Kalman filter in reference [6] are verified, and the results are shown in Figure 5.

By analyzing the results shown in Figure 5, it can be seen that with the continuous change of image data, the tracking time of different algorithms also decreases. The tracking time of the reference [5] algorithm is always between 11 s and 14 s, and that of the reference [6] algorithm is always between 13 s and 15 s. The tracking time of the reference [6] algorithm is more than that of the reference [5]. The tracking time of this algorithm is always less than 5 s, which is significantly less than the two comparison algorithms. Therefore, this algorithm has more application advantages, which shows that the algorithm has stronger timeliness.

In order to further verify the effectiveness of the method in this paper, considering the economic benefits of different algorithms, the results are shown in Table 1.

Analysis of Table 1 shows that when the image data volume is 100 GB, the tracking cost of the algorithm in reference [5] is 2223.7 yuan, the tracking cost of the algorithm in reference [6] is 2758.1 yuan, and the tracking cost of the method in this paper is 1086.2 yuan. When the image data amount is 600 GB, the tracking cost of the algorithm in reference [5] is 7995.3 yuan, the tracking cost of the algorithm in reference [6] is 7642.0 yuan, and the tracking cost of the method in this paper is 1339.7 yuan. The economic cost of this method is always lower than that of other methods, which indicates that this method has better economic benefits.

5. Concluding Remarks

In this study, a multiscale pedestrian tracking algorithm based on a contour template is designed. According to the regional characteristics of pedestrian action, the pedestrian template is divided into several contour regions, and the similarity threshold is obtained by using the regional feature similarity classification. The multicontour template matching is completed, and the multitarget object detection and tracking are realized. The results are as follows:

1. When the frame size is 56, the recognition accuracy of this algorithm can reach 96.5%, which shows that the recognition accuracy of this method is better.

2. When the frame size is 104, the tracking error of the pedestrian gravity center is only 5%, which shows...
that the tracking accuracy of the pedestrian gravity center is low.

(3) When the amount of image data is 80 GB, the tracking time of this algorithm is only 3 s, which shows that the tracking efficiency of this method is high.

(4) When the amount of image data is 600 GB, the tracking cost of this method is 1339.7 yuan, which shows that the method has good economic benefits.

To sum up, it can be proved that the algorithm has the advantages of high recognition accuracy, low tracking error, and less tracking time.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Key Project of Natural Science Research of Universities in Anhui Province: Research and Application of Video Pedestrian Detection and Tracking (no. KJ2018A0669); The Key Project of Natural Science Research of Universities in Anhui Province: Vehicle Trajectory Extraction and Behavior Analysis in Complicated Traffic Video Science (no. KJ2020A1216).

References


