



Research on Real Time Tracking Method of Multiple Moving Objects Based on Machine Vision

Yuan Wang^(✉)

Wuhan Institute of Design and Sciences, Wuhan 430205, China
yinwar822172@163.com

Abstract. In the process of real-time tracking of multiple moving targets, there is a big gap between the tracking effect and the ideal effect due to the influence of the objective environment state. Therefore, a real-time tracking method of multiple moving targets based on machine vision is proposed. A machine vision acquisition unit including lighting device, camera, image acquisition card, processing system and control actuator is constructed. At the same time, in order to avoid the impact of outdoor extreme weather on image acquisition effect, MF-DSC03 with waterproof function and automatic light compensation function is used as the camera head device. In the phase of real-time tracking of multiple moving targets, based on HSV color features and apparent features in real-time video image data of multiple moving targets captured by machine vision, iterative tracking of the state of multiple moving targets is realized by means of KM bipartite graph matching algorithm. In the test results, the design method outperforms the control group in multi-target tracking performance on different data sets, showing an ideal tracking effect.

Keywords: Machine Vision · Multiple Moving Targets · Real-Time Tracking · Machine Vision Acquisition Unit · Hsv Color Characteristics · Apparent Characteristics · Km Bipartite Graph Matching Algorithm

1 Introduction

Computer vision is an important research direction in the field of artificial intelligence, which aims to study how to make computers perceive, analyze and process the real world intelligently like human vision systems. The camera is easy to install and use as a relatively inexpensive environment sensing sensor [1]. Therefore, various computer vision algorithms based on images and videos have already penetrated into the daily life of the public, such as face recognition technology used in access control and security inspection, intelligent monitoring in traffic field, visual navigation in automatic driving field, etc. [2]. Video target tracking technology, as one of the basic and important research directions in the field of computer vision, has always been the focus of researchers [3]. Generalized target tracking usually includes single object tracking (SOT) and multiple object tracking (MOT) [4]. Video target tracking focuses on the ability to locate

and estimate the size of an object of interest in a frame for a long time in subsequent video frames [5]. Usually, multi-target tracking is limited to scenes with known target categories, such as visual tracking of multiple pedestrians and vehicles [6]. In order to distinguish multiple targets, each target is usually marked with a digital ID representing a unique identity. Therefore, in addition to tracking the location and scale information of the target [7], simultaneous movement of multiple targets should also ensure that the same target remains fixed unique digital identification. At present, the research results of multi-target tracking methods show good tracking performance in some simple tracking scenarios, and there is no big gap [8]. As the difficulty of the scene increases, various evaluation indicators of tracking performance begin to decline in varying degrees [9] due to the different focus of different research results. The difficulty of tracking the scene mainly refers to the irregular movement of the target, the complexity of the target and the background, the similarity of the target and the interaction between the targets, etc. These scene difficulties are mainly manifested as occlusion problems in the multi target tracking research. Occlusion is the main cause of target loss during tracking [10]. Object occlusion can be divided into two situations: being occluded by background objects and being occluded by other objects. The occluded area will affect the probability of target loss to varying degrees. The situation that the target no longer appears in the scene after being occluded is generally classified as the situation that the target disappears, which is not within the scope of the occlusion problem. The occlusion problem focuses on the situation that the target reappears after being occluded. The occlusion phenomenon does not happen instantaneously but has a process, which can be divided into three stages; The first stage is the process of the target being blocked by obstacles, and the target gradually disappears from the image; The second stage is that the target is completely blocked by obstacles, at this time, the target is completely invisible in the image; The third stage is the process of the target gradually moving out of the obstacle, and the target information reappears in the image, which is also called target reappearance. The longer the whole process takes in the second stage, the longer the effective information of the target is lost, and the harder it is to find the target when the target reappears. At present, Most multi-target tracking algorithms do not have good methods to deal with the situation that the target is occluded for a long time. How to effectively deal with the occlusion problem will be the purpose of scholars' continuous research.

Therefore, this paper proposes a real-time tracking method for multiple moving targets based on machine vision. By introducing machine vision, this method collects the data of multiple moving targets, constructs the video image data acquisition process of moving targets, and the hardware used in machine vision, etc., and realizes target tracking through the combination of HSV color histogram and KM binary graph matching algorithm. Through the comparison test, the practical application effect of the designed tracking method is analyzed and verified, and the method has better tracking effect.

2 Design of Real-Time Tracking Method for Multiple Moving Targets

2.1 Data Acquisition Based on Machine Vision

Machine vision [11, 12] simulates and searches for human vision rules through computers, so as to input images into the image processing process. In order to meet the different functional requirements in different fields, understanding the user's performance requirements for each component of data collection is the key to designing a good machine vision product. And the application of machine vision is because the method has higher automation, efficiency, stability, diversity and flexibility, which can process a large number of image data in a short time. The main structure of machine vision acquisition includes: lighting device, camera, image acquisition card, processing system and control actuator. For a machine vision system, there can be several input signals, which can be obtained from different perspectives, different times or different directions. The obtained 3D scene projection signal is input into the system. Usually, the input signal represents the spatial relationship of objects, such as surface spatial structure, surface shape characteristics, texture characteristics, color characteristics and other information. The cooperation between the modules of the machine vision system directly affects the function of the whole system. Any problem in any module will directly affect the performance of the next module or even the whole system, and eventually lead to deviation or error in the system results. Obtaining high-quality image information is the premise of machine vision system. However, in order to obtain ideal image quality, we must start from the selection of acquisition methods, lens selection and other aspects, then input the obtained image information into the processing platform, and cooperate with the module controller to achieve the higher-level functions of machine vision system.

In order to track pedestrians in real life, this paper provides the remote monitoring data adopts the camera for image shooting, and processes the image data information collected by the camera, so as to achieve the goal of target tracking. The flow chart of image acquisition used in this study is shown in Fig. 1.

The key step in the target tracking process is to use the camera to obtain the image, obtain the external environment information, and further process the acquired image. In order to achieve better real-time image acquisition, the main components needed include: light source, light source controller and lens. The display displays the image processing effect in real time, which is convenient for human observation. The image acquisition module usually uses CMOS or CCD cameras to convert the collected image information into digital signals and transmit them to the software platform with processing functions. According to the specific needs of different environments, different processing algorithms are used to achieve the analysis and processing of targets, extract the target area of interest, and make real-time judgments on the processing results, Then control relevant equipment to respond, achieve the purpose of image processing module, and provide basis for more advanced behavior.

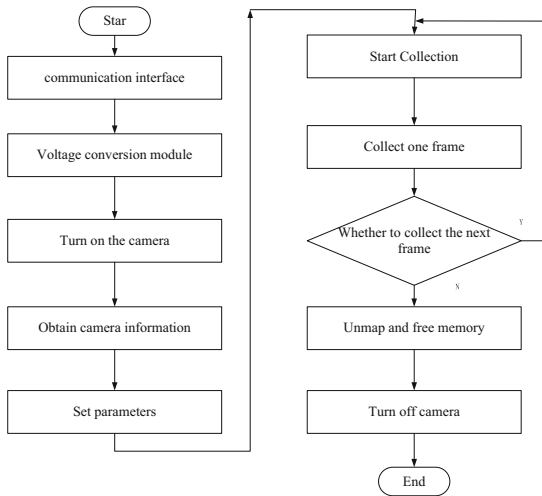


Fig. 1. Video image data acquisition process based on machine vision

The camera is the core of the image acquisition system. This paper selects a camera with waterproof function and automatic light compensation function, which can still achieve the purpose of image acquisition in extreme weather outdoor. The MF-DSC03 series serial camera used in this research is a highly integrated embedded digital camera. The camera has 2 million pixels and dedicated image compression technology. The DSP microprocessor is used for image compression processing. The image quality is clear, the color is lifelike, and it has the function of infrared night vision light compensation, Its image sensor adopts the world's best 0V series high-performance CMOS sensor, with progressive scanning function, which has high practical value in traffic supervision, intelligent agricultural production and other aspects. The output image format of MF-DSC03 series products can be directly converted to JPEG format. The data is transmitted through RS232/RS485/TL, and the electronic shutter is used for shooting. The MF-DSC03 series serial port camera heads are best used in places that require high image quality and need to transmit images. The main performance parameters of MF-DSC03 series serial camera are shown in Table 1.

On this basis, the real-time video image data of multiple moving objects is collected using machine vision, which provides the basis for the follow-up target tracking.

2.2 Real Time Tracking of Multiple Moving Targets

In most tracking scenes [13, 14], the color information of the target is the most recognizable appearance description, so this paper chooses to extract HSV color histogram statistical features of the target to represent its color information. The reason for choosing the HSV color histogram is that the HSV space can express the light and shade, tone, and vividness of the color very intuitively, so as to facilitate the contrast between colors. However, in a few tracking scenes, there will be target objects with highly similar appearance, whose color characteristics are not enough to distinguish each other's

Table 1. Parameters of MF-DSC03 Series Serial Camera

S/N	function	model
1	Image sensor type	CMOS 1/4 inch
2	Camera pixels	2 million
3	Pixel size	2.2umX22um
4	Output format	JPEG format
5	Balance mode	automatic
6	Exposure mode	automatic
7	Gain mode	automatic
8	Signal-to-noise ratio	40 dB
9	dynamic range	50 dB
10	Maximum analog gain	16 dB
11	Frame rate	UXGASXGA: 15fps SVGA: 30 fps CIF: 60fps
12	viewing angle	90 degrees (120 or other angles are optional)
13	Distance range	0-20m (adjust according to specific requirements)
14	Serial port speed	Between 15000 and 120000
15	Operating current value	≤ 100 mA (turn off the infrared device) ≤ 200 mA (turn on the infrared device)
16	Operating voltage value	DC 5V

identities, and sometimes the deformation of targets will lead to differences in color histogram statistics. Considering that in the Re ID field, the convolutional network can extract differentiated depth information from the appearance of targets, It can make similar targets show differences at the depth semantic feature level. Therefore, the fusion of depth apparent features enhances the feature representation of targets.

To sum up, the motion information generated by the extended Kalman filter for position prediction, the color information of the HSV color histogram [15] and the apparent information of the Re ID depth feature are fully considered, and the three feature information are fused to measure the similarity between the detection sequence and the tracking track sequence, so as to increase the accuracy of data association.

For HSV color feature measurement, HSV corresponds to hue (HueE (0, 360)), saturation (SaturationE (0, 1)), and brightness (ValueE (0, 1)) respectively, which is a color space constructed according to the intuitive features of image color, and can reflect the true color information of objects. Therefore, HSV color feature histogram features are used to describe the color information of detection and tracking tracks. The

three channel colors in the HSV space are uniformly quantized. The specific uniform quantization method can be expressed as

$$H \left\{ \begin{array}{l} 0, H \in [0, 22.5) \\ 1, H \in [22.5, 45) \\ 2, H \in [45, 67.5) \\ \dots \\ 14, H \in [315, 337.5) \\ 15, H \in [337.5, 360) \end{array} \right. \quad (1)$$

$$S \left\{ \begin{array}{l} 0, S \in [0, 0.25) \\ 1, S \in [0.25, 0.5) \\ 2, S \in [0.5, 0.75) \\ 3, S \in [0.75, 1) \end{array} \right. \quad (2)$$

$$V \left\{ \begin{array}{l} 0, V \in [0, 0.25) \\ 1, V \in [0.25, 0.5) \\ 2, V \in [0.5, 0.75) \\ 3, V \in [0.75, 1) \end{array} \right. \quad (3)$$

On this basis, the quantized histogram eigenvalues of real-time video image data of multiple moving targets are used V_{HSV} Represents, which can be expressed as

$$\begin{aligned} V_{HSV} &= 16H + 4S + V \\ V_{HSV} &\in (0, 255) \end{aligned} \quad (4)$$

For color characteristics f_{color} express, f_{color} . It represents the 256 dimensional color histogram feature vector of a target. The chi square distance is used to measure the color difference, which can be expressed as

$$d_{color}(i, j) = \sum \frac{[f_{color(i)}(n) - f_{color(j)}(n)]^2}{f_{color(i)}(n) + f_{color(j)}(n)} \quad (5)$$

where, $d_{color}(i, j)$ Represents the chi square distance of colors in real-time video image data of multiple moving targets, $f_{color(i)}(n)$ Record as detection target i Color characteristics, $f_{color(j)}(n)$ Record as track j Color characteristics of.

For the apparent feature measurement of real-time video image data of multiple moving objects, the specific implementation method is.

Initialize the target track with identity information Q Set to

$$Q = \{Q_1, Q_2, Q_3 \dots Q_n\} \quad (6)$$

The track information is transmitted frame by frame, and then the detected target sequence is measured in the next frame according to its apparent characteristics G Expressed as

$$G = \{G_1, G_2, G_3 \dots G_n\} \quad (7)$$

Use it with Q According to the measurement results, the target is re identified and confirmed. For this purpose, this article uses f_{reid} To describe the apparent characteristics of a target, f_{reid} Specifically, it refers to the 128 dimensional apparent feature vector extracted. Assuming that there are N detection objects in the current frame, their corresponding characteristic matrix F_D Can be expressed as

$$F_D = \{f_{reid(1)}, f_{reid(2)}, f_{reid(3)}, \dots, f_{reid(N)}\} \tag{8}$$

For M confirmed tracking tracks, each confirmed associated track $T(j)$ stores the composition set of its previous 50 frame feature vectors $F_{confirm}^{T(j)}$ Can be expressed as

$$F_{confirm}^{T(j)} = \{f_{reid(1)}^{T(j)}, f_{reid(2)}^{T(j)}, f_{reid(3)}^{T(j)}, \dots, f_{reid(N)}^{T(j)}\} \tag{9}$$

Here, the number of temporary features is adjusted by setting 50 super parameters. By calculating the number of i Apparent characteristics of detected objects $f_{reid(j)}$ And the j 50 frames of tracks $F_{confirm}^{T(j)}$ The minimum cosine distance between is used to measure the apparent similarity. The specific calculation method can be expressed as

$$d_{reid}(i, j) = \min\{1 - f_{reid}^{T(j)} F_{confirm}^{T(j)}\} \tag{10}$$

Among them, $d_{reid}(i, j)$ Represent each feature difference parameter of the tracking track, calculate the cosine distance between them and the apparent features of N detection objects in the current frame, and then take the minimum value as the calculation value between the track and the detection result, and the cosine cost matrix can be expressed as

$$\text{cos } t_{\text{cos}(M * N)} = \begin{bmatrix} d_{reid}(1, 1) \dots \dots d_{reid}(1, N) \\ d_{reid}(2, 1) \dots \dots d_{reid}(3, N) \\ \dots \dots \dots \dots \\ d_{reid}(M, 1) \dots \dots d_{reid}(M, N) \end{bmatrix} \tag{11}$$

Among them, $\text{cos } t_{\text{cos}(M * N)}$ The cosine cost matrix representing the real-time video image data of multiple moving targets.

On this basis, because the number of targets in the tracking scene is constantly changing, not all of the tracks in each frame can form a one-to-one correspondence relationship with the detection data after association, there must be unmatched tracks and detection targets [16–18], so it is necessary to divide the tracks and detection states. The KM bipartite graph matching algorithm is introduced, because the algorithm is more efficient, comparable to the cost flow, much simpler than the cost flow, and the idea is clearer, and solves the problem of maximum weight matching, which is suitable for the state division of trajectory and detection. After calculation by KM bipartite graph matching algorithm, set the two states of matched detection and unmatched detection, and set the two states of unconfirmed and confirmed for all tracks. The confirmed state refers to the track that has completed the correct matching. In the current frame, the confirmed state is divided into unmatched track and matched track, and the unconfirmed track is initialized by the unmatched detection target. The unconfirmed track is set to

reduce the chance of matching and prevent the detection target from not being associated with tracks of the same identity. The specific method is to set the matching threshold $Hits = 2$ frames (too many frames will affect real-time), that is, the KM algorithm completes the matching of two consecutive frames before it is allowed to initialize to the confirmed track. Because the target is blocked or the detector misses detection, the tracks in the confirmed state will appear unmatched tracks [19, 20]. In this case, set the lifetime Age and consecutive lost frame count Count for the confirmation state track. Where $Age = 100$ (if the value is too high, the garbage track will be cached too much) indicates the upper limit of the number of consecutive lost frames. When $OSCountSAge$, the track is valid. In the matching process, count + 1 iterative correlation is performed to match the tracks with fewer lost frames in priority, while ensuring the effectiveness of the lost tracks in the survival period. The specific implementation process of the update algorithm is divided into the following steps:

- (1) Set the variable $Count = 0$ for the number of consecutive lost frames and the constant $Age = 100$ for the lifetime;
- (2) At time t , n detection objects are output by machine vision and numbered $d \in \{0, 1, 2, 3, \dots, n\}$, d contains unmatched detection objects and matched detection objects in the previous frame;
- (3) At time t , the m tracks predicted by the extended Kalman filter in this paper are numbered $k \in \{0, 1, 2, 3, \dots, m\}$, and each track has a Count variable to record the number of consecutive lost frames of tracks, and k contains unmatched tracks and matched tracks whose Count is less than Age times;
- (4) Use the weighted multi feature fusion algorithm proposed in Sect. 4.23 of this paper to measure the similarity between d and k , and output the cost matrix $t_{\cos(M*N)}$. The KM algorithm is used to perform Count + 1 iteration matching association. This process starts from the track with Count = 0 and iterates to the maximum of Count, that is, the stable tracking track is first associated;
- (5) After iterative correlation, the unmatched track Count value is increased by 1, and the matched track Count is set to 0, unmatched

The detection target of is initialized as an unconfirmed track, and the track whose Count is greater than Age is deleted.

So far, the status update of all tracks in one frame has been completed, realizing real-time tracking of moving objects.

3 Test Analysis

3.1 Test Data Set Preparation

In order to verify the effectiveness of multi-target tracking algorithm based on attention mechanism in improving the accuracy of tracking algorithm, this chapter conducts experiments on MOT17, MOT16 and MOT15 datasets respectively, and uses MOTA, MOTP, IDF1, MT, ML, FP, FN and IDS to evaluate and analyze.

MOT Challenge is a public platform that can upload and publish multi-target tracking research results, and has the largest pedestrian tracking data set. The platform is designed to evaluate the detection performance and tracking performance of multi pedestrian in complex environments, mainly including MOT15, MOT16 and MOT17 data sets, which are introduced below. Table 2 shows the detailed attribute information of the dataset.

Table 2. Basic Information Statistics of Test Data Set

data set	Video source	length	Number of tracks	FPS	Camera condition	launch a pilot project
MOT15	TUD-Crossing	201	13	25	static	level
	PETS2009-S21.2	436	42	7	static	high
	ETH-C rossing	219	26	14	move	low
	ADL -R undle-1	500	32	30	move	level
	KITTI-16	209	17	10	static	level
MOT16	MOT1 6-01	450	23	30	static	level
	MOT1 6-03	1500	148	30	move	high
	MOT16-06	1194	221	14	move	high
	MOT1 6-12	900	86	30	static	level
MOT17	MOT1 7-01	450	24	30	static	level
	MOT1 7-03	1500	148	30	move	high
	MOT1 7-06	1194	222	14	move	level

The basic situation of three test data sets is analyzed.

- (1) MOT15: The data set disclosed by MOT Challenge in 2015, in which video sequences are shot by static or mobile cameras under unlimited conditions. The data set contains 22 video sequences, 11 of which are test sets and 11 are training sets. There are various scenes. The dataset has 11283 frames in total, including 1221 different pedestrians and 101345 pedestrian detection frames.
- (2) MOT16: The data set released by MOT Challenge in 2016 includes 7 training sets and 7 test sets, totaling 14 video sets. Compared with MOT15 dataset, its annotation is more accurate, the scene is more complex, and the number of pedestrians is more, including 11235 frames, including 1342 pedestrians and 292733 pedestrian detection frames. The official test results are the results detected by DPM detector.
- (3) MOT17: The dataset released by MOT Challenge in 2017 has the same video sequence as MOT16. The only difference is that the targets in the dataset are detected by three different detectors (Faster R-CNN, DPM, SDP), so the detection results are more accurate.

On this basis, multi moving target tracking test is carried out.

3.2 Multi Target Tracking Evaluation Indicator Setting

In order to evaluate the performance of multi-target tracking algorithm fairly and accurately, Table 3 lists the most commonly used multi-target tracking evaluation indicators.

Table 3. Multi target tracking evaluation indicators

S/N	Measure Name	Expected score	sketch
1	MOTA↑	100%	Multi target tracking accuracy
2	MOTP↑	100%	Multi target tracking accuracy
3	IDF1↑	100%	Accuracy of identity representation
4	MT↑	100%	Most tracked targets
5	ML↓	0%	The least lost goal
6	FP↓	0	Number of unmatched tracks
7	FN↓	0	No matched real target
8	IDS↓	0	Total number of identity switches

Carry out specific analysis on the multi-target tracking evaluation indicators shown in Table 3.

- (1) MOTA: Accuracy of multi-target tracking. This indicator is used to determine the number of targets and the cumulative error of statistical tracking. It is mainly related to three factors: FP (number of false detections), FN (number of missed detections), and IDSW (number of identity information handovers). The fewer false positives, false positives, and identity switches in the tracking process, the higher the tracking accuracy of the model. The calculation formula can be expressed as

$$MOTA = 1 - \frac{\sum (m_t + f_{pt} + mme_t)}{\sum g_t} \quad (12)$$

Among them, m_t Section t Number of missed frames, f_{pt} For t Number of false detections of frames, mme_t Is the number of mismatches IDSW, that is, the t . The number of targets with identity switching in the frame, g_t For t The number of real targets in the frame.

- (2) MOTP: Accuracy of multi-target tracking model. This indicator is used to evaluate the precision of the predicted target box position

Accuracy is generally defined by calculating the coincidence rate between the prediction box and the real dimension box. The calculation formula can be expressed as

$$MOTP = \frac{\sum d_i^t}{\sum c_t} \quad (13)$$

Among them. c_t Represents the t . The number of successful matches between the real target and the predicted target in the frame, d_i^t It indicates the Euclidean distance between target matching pairs, i.e. error.

- (3) MT: indicates the proportion of most tracked target tracks, that is, when the tracks tracked account for more than 80% of the real track length of the target, most of the target is tracked.
- (4) ML: indicates the proportion of most untracked target tracks, that is, when the tracks tracked by the target account for less than 20% of the real track length of the target, most of the target tracks are missing.
- (5) FP: indicates the number of false checks, that is, the number of tracks not matched in the current frame.
- (6) FN: indicates the number of missed checks, that is, the number of true annotations on the current frame that cannot be matched.
- (7) IDS: The number of times the trace changed the identity information.
- (8) IDE1: measures the accuracy of identity in tracking.

Based on the multiple target tracking evaluation indicators set above, the specific tracking effect is analyzed.

3.3 Systematic Testing

In order to ensure the accuracy of the experiment, it is necessary to ensure the normal operation of each module designed in this paper. Therefore, systematic testing is carried out to ensure the effectiveness of the module and the accuracy of the experiment. The test results are shown in Table 4.

Table 4. Systematic test results

Function name	Status
Moving target image video acquisition	Normal
Data transmission	Normal
Target image display	Normal
Data compression	Normal
Target feature extraction	Normal
Target matching	Normal
Target tracking	Normal

According to Table 4, the functions of each module of the system designed in this paper can run normally, and it has the basic feasibility of experimental analysis of target tracking.

3.4 Test Results and Analysis

In this section, on the open data sets of MOT17, MOT16 and MOT15, the attention mechanism based multi-target tracking method proposed in this chapter is compared with the reference [4] method and the reference [5] method. The reference [4] method is a moving target location method of moving image sequence based on the optimized particle filter hybrid tracking algorithm. The optimized particle filter hybrid tracking algorithm is introduced to achieve target tracking and positioning, while the reference [5] method is a fast moving and deformed target tracking method based on heterogeneous feature fusion, which realizes target tracking by integrating heterogeneous features. The experimental results of the three methods are shown in Tables 5, 6 and 7 below.

Table 5. Statistical Table of Comparison Experiment Results on MOT15 Data Set

evaluating indicator	Tracking method		
	Reference [4] method	Reference [5] method	Design method in this paper
MOTA	44.10	46.33	46.59
MOTP	77.50	53.40	79.10
IDF1	46.00	46.69	47.59
MT	17.20	19.00	18.31
ML	26.60	27.74	28.02
FP	6058	4627	4558
FN	26917	27065	26951
IDS	1347	1280	1304

It can be seen from Table 5 to Table 7 that on the MOT17 data set, the real-time tracking method for multiple moving targets proposed in this paper has increased by 2.58%, 1.61%, 1.24% and 1.53% on MOTA, MOTP, IDF1 and MT indicators respectively, and decreased by 1.23%, 3041 and 11471 on ML, FP and FN indicators respectively, compared with reference [4] method. On MOT16 dataset, the multi moving target real-time tracking method proposed in this paper has also greatly improved the evaluation indicators other than IDS. On the MOT15 dataset, the multi moving target real-time tracking method proposed in this paper is only higher than reference [4] method in ML and FN. This may be because MOT15 dataset is collected in an unconstrained environment, and the pedestrian in the video is too small, resulting in the loss of tracking targets. However, in general, the performance of the multi moving target real-time tracking method

Table 6. Statistical Table of Comparison Experiment Results on MOT16 Data Set

evaluating indicator	Tracking method		
	Reference [4] method	Reference [5] method	Design method in this paper
MOTA	54.80	56.86	56.95
MOTP	54.15	78.81	55.04
IDF1	53.40	55.14	55.51
MT	19.10	20.82	21.08
ML	37.00	34.91	34.91
FP	2955	2673	2625
FN	78765	75334	75217
IDS	645	647	644

Table 7. Statistical Table of Comparison Experiment Results on MOT15 Data Set

evaluating indicator	Tracking method		
	Reference [4] method	Reference [5] method	Design method in this paper
MOTA	53.70	56.11	56.28
MOTP	77.20	53.80	75.25
IDF1	53.80	54.15	55.04
MT	19.40	20.38	20.93
ML	36.60	35.50	35.37
FP	11731	8854	8690
FN	247447	237008	235976
IDS	1947	2065	1968

designed in this paper based on adaptive particle swarm optimization algorithm is significantly better than that of reference [4] method. In this paper, the tracking method is designed to reconstruct the features of adjacent frame targets through adaptive mechanism, so that the target features can ignore the similar features within the frame, enhance the feature expression ability of inter frame targets, and the correlation matrix obtained by relying on the reconstructed features is more reliable, thus improving the tracking accuracy of multi-target tracking.

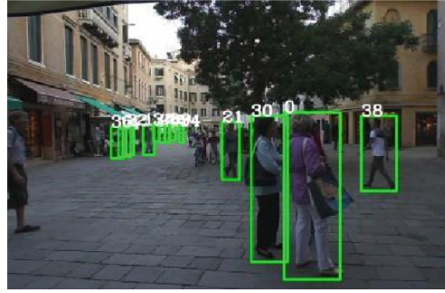
On MOT17 dataset, the real-time tracking method for multiple moving targets proposed in this paper is 0.17% higher, 3.56% higher, 0.89% higher and 0.55% lower in MOTA, MOTP, IDF1 and MT than in reference [5] method, and 0.13% lower, 1032 and 97 lower in ML, FN and IDS respectively. On MOT16 data set, the real-time tracking method of multiple moving targets proposed in this paper is superior to reference [5] method in all indicators. On the MOT15 dataset, the multi moving target real-time tracking method proposed in this paper is superior to the reference [5] method in MOTA, MOTP, IDF1, FP, FN indicators, but the MT indicators are low and the ML indicators are high, which may be due to the unstable tracking results caused by the MOT15 dataset's own attributes. In general, the multi moving target real-time tracking method based on adaptive particle swarm optimization algorithm is superior to the reference [5] method in many indicators, which may be because reference [5] method heavily relies on the target dependency set manually, is sensitive to the setting of many super parameters, and fails to fully exploit the correlation between targets, However, this adaptive framework is introduced into the multi moving target real-time tracking method proposed in this paper. The adaptive mechanism is used to learn the dependency between targets in the frame, and the cross attention mechanism is used to learn the correlation between targets in the frame, so that high-quality target dependency can be automatically learned, and the tracking performance of multi target tracking is improved.

In order to more intuitively compare the tracking performance of the reference [5] method and the multi moving target real-time tracking method proposed in this paper, Fig. 2 shows some visualization results of the two algorithms in the two video sequences of MOT17 dataset 02 and 09.

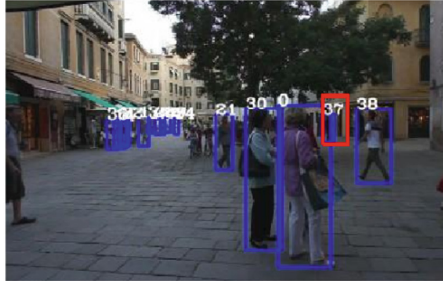
It can be seen from Fig. 2 that Fig. 2 - (a) compares the tracking effect under the complex background with dim light. From the figure, it can be seen that in the 308th frame of MOT17-02 dataset, the real-time tracking method of multiple moving targets proposed in this paper can track the smaller target 37 compared with the reference [5] method.

Figure 2 - (b) compares the tracking effect when the target is blocked by a static object. It can be seen from the figure that at frame 391 of MOT17-09 dataset, the real-time tracking method of multiple moving targets proposed in this paper can track 29 targets blocked by buildings very well.

Therefore, combined with the visual test results, it can be further concluded that the multi moving target real-time tracking method proposed in this paper further improves the tracking accuracy compared with the reference [5] method.

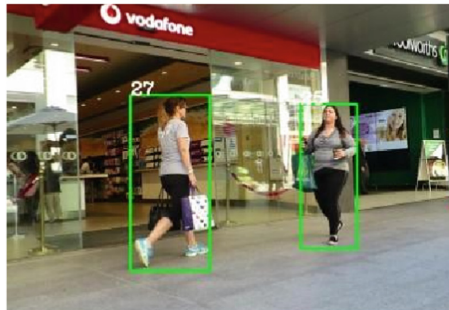


Reference [5] method

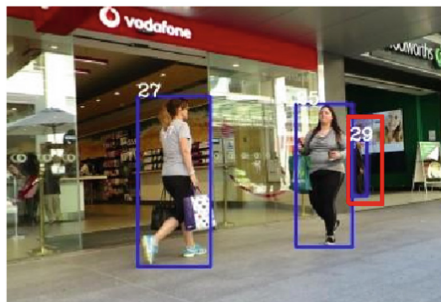


This article designs a tracking method

(a) Comparison Diagram of Tracking Effect under Complex Background with Dim Light



Reference [5] method



This article designs a tracking method

(b) Comparison of tracking effect when the target is statically occluded

Fig. 2. Comparison of real-time tracking effects of multiple moving targets

4 Conclusion

Real time online tracking of multiple moving objects in video has always been a key task in computer vision, and plays an important role in camera based environment perception. For example, the practical application of multi-target tracking technology can be seen in scenes such as automatic driving, security and traffic monitoring. Since the development of online multi-target tracking technology, the technical target has always been how to maintain the trajectory of multiple objects of interest from beginning to end without loss. To achieve this goal, it is necessary to solve the problems of object of interest detection, motion trajectory prediction, detection and trajectory correlation, and object occlusion. In this paper, a real-time tracking method for multiple moving objects based on machine vision is proposed, which can accurately locate multiple specific objects in the video during the multi-target tracking phase, maintain the identity information of the target in complex scenes, and output complete target motion tracks. Different from single target tracking, the multi-target tracking designed in this paper not only realizes the purpose of accurately outputting the position of each target in the video frame, but also realizes the effective target matching between frames.

Acknowledgement. 2022 Hubei Provincial Department of Education Science Research Program Guiding Project: Application of Gesture Interaction Technology in Public Art Design in Smart Cities (B2022422).

References

1. Subrahmanyam, K., Kavitha, L.M., Rao, S.K.: Shifted Rayleigh filter: a novel estimation filtering algorithm for pervasive underwater passive target tracking for computation in 3D by bearing and elevation measurements. *Int. J. Pervas. Comput. Commun.* **18**(3), 272–287 (2022)
2. Wang, X., Xie, W., Liangqun, L.I.: Labeled Multi-Bernoulli Maneuvering target tracking algorithm via TSK iterative regression model. *Chin. J. Electron.* **31**(2), 227–239 (2022)
3. Dong, J., Li, Y., Guo, Q., Liang, X.: Through-wall moving target tracking algorithm in multipath using UWB radar. *IEEE Geosci. Rem. Sens. Lett.* **19**, 1–5 (2022). <https://doi.org/10.1109/LGRS.2021.3050501>
4. Guo, Y.: Moving target localization in sports image sequence based on optimized particle filter hybrid tracking algorithm. *Complexity* **2021**(7), 1–11 (2021)
5. Li, B., Jing, Q.: Fast moving and deformational target tracking approach based on heterogeneous features fusion. *Trans. Inst. Meas. Control.* **43**(3), 612–622 (2021)
6. Yang, J., Liu, X., Sun, J., Li, S.: Sampled-data robust visual servoing control for moving target tracking of an inertially stabilized platform with a measurement delay. *Automatica* **137**, 110105 (2022). <https://doi.org/10.1016/j.automatica.2021.110105>
7. Guo, C., You, S., Luo, Y., et al.: A fast moving target detection tracking and trajectory prediction system for binocular vision. *Wuhan Univ. J. Nat. Sci.* **26**(1), 12 (2021)
8. Xu, W.: Adaptive probability hypothesis density filter for multi-target tracking with unknown measurement noise statistics. *Measure. Control London Inst. Measure. Control* **54**(3–4), 002029402199280 (2021)
9. Lin, C., Shi, J., Lyu, Y., et al.: Over-flight and standoff tracking of a ground target with a fixed-wing unmanned aerial vehicle based on a unified sliding mode guidance law. *Trans. Inst. Measure. Control.* **44**(2), 410–423 (2022)

10. Zhou, C., Wang, Y., Xiong, R.H., et al.: Research on improvement of efficiency of radio monitoring and positioning based on UAV. *Comput. Simul.* **003**, 039 (2022)
11. Zhang, Y., Zhao, J., Han, H.: A 3D machine vision-enabled intelligent robot architecture. *Mob. Inf. Syst.* **2021**, 1–11 (2021). <https://doi.org/10.1155/2021/6617286>
12. Ahn, J., Kim, S.: Automated textile circuit generation using machine vision and embroidery technique. *Textile Res. J.* **92**(11–12), 1977–1986 (2022). <https://doi.org/10.1177/00405175221075062>
13. Du, D.: Multiperson target dynamic tracking method for athlete training based on wireless body area network. *Adv. Math. Phys.* **2021**, 1–9 (2021). <https://doi.org/10.1155/2021/2287751>
14. Wang, X., Xie, W., Li, L.: Labeled multi-Bernoulli Maneuvering target tracking algorithm via TSK iterative regression model. *Chinese J. Electron.* **31**(2), 227–239 (2022). <https://doi.org/10.1049/cje.2020.00.156>
15. Lv, C., Li, J., Kou, Q., et al.: Stereo matching algorithm based on hsv color space and improved census transform. *Math. Probl. Eng.* **2021**(32), 1–17 (2021)
16. Yang, D.: Target tracking algorithm based on adaptive scale detection learning. *Complexity* **2021**, 1–11 (2021). <https://doi.org/10.1155/2021/9033912>
17. Zhang, H., Cheng, L., Zhang, J., Huang, W., Liu, X., Yu, J.: Structural pixel-wise target attention for robust object tracking. *Digital Signal Process.* **117**, 103139 (2021). <https://doi.org/10.1016/j.dsp.2021.103139>
18. Liu, Y., Wang, X.: Mean shift fusion color histogram algorithm for nonrigid complex target tracking in sports video. *Complexity* **2021**, 1–11 (2021). <https://doi.org/10.1155/2021/5569637>
19. Pirovano, L., Armellin, R., Siminski, J., Flohrer, T.: Differential algebra enabled multi-target tracking for too-short arcs. *Acta Astronautica* **182**, 310–324 (2021). <https://doi.org/10.1016/j.actaastro.2021.02.023>
20. Song, D., Gan, W., Yao, P., Zang, W., Zhang, Z., Qu, X.: Guidance and control of autonomous surface underwater vehicles for target tracking in ocean environment by deep reinforcement learning. *Ocean Eng.* **250**, 110947 (2022). <https://doi.org/10.1016/j.oceaneng.2022.110947>