WILEY | Hindawi

Research Article

Target Tracking and 3D Trajectory Reconstruction Based on Multicamera Calibration

Junfang Song [1,2] **Yao Fan,**[1,2] **Huansheng Song,**[3] **and Haili Zhao**[1]

[1]*School of Information Engineering, Xizang Minzu University, Xianyang, Shaanxi 712082, China*
[2]*Key Laboratory of Optical Information Processing and Visualization Technology of Tibet Autonomous Region, Xianyang, Shaanxi 712082, China*
[3]*School of Information Engineering, Chang'an University, Xi'an 710064, China*

Correspondence should be addressed to Junfang Song; 284786635@qq.com

In traffic scenarios, vehicle trajectories can provide almost all the dynamic information of moving vehicles. Analyzing the vehicle trajectory in the monitoring scene can grasp the dynamic road traffic information. Cross-camera association of vehicle trajectories in multiple cameras can break the isolation of target information between single cameras and obtain the overall road operation conditions in a large-scale video surveillance area, which helps road traffic managers to conduct traffic analysis, prediction, and control. Based on the framework of DBT automatic target detection, this paper proposes a cross-camera vehicle trajectory correlation matching method based on the Euclidean distance metric correlation of trajectory points. For the multitarget vehicle trajectory acquired in a single camera, we first perform 3D trajectory reconstruction based on the combined camera calibration in the overlapping area and then complete the similarity association between the cross-camera trajectories and the cross-camera trajectory update, and complete the trajectory transfer of the vehicle between adjacent cameras. Experiments show that the method in this paper can well solve the problem that the current tracking technology is difficult to match the vehicle trajectory under different cameras in complex traffic scenes and essentially achieves long-term and long-distance continuous tracking and trajectory acquisition of multiple targets across cameras.

1. Introduction

Target tracking is one of the research hot spots in computer vision, and it has been widely used in military, unmanned driving, video monitoring, and other fields. The current target tracking algorithm [1] can be divided into three categories from the observation model: the method based on the generated model, the method based on the discriminant model, and the method based on deep learning.

The method based on the generative model is also called the classical target tracking algorithm. This method extracts the features of the target in the current frame, constructs the target model, and searches the best matching region with the appearance model in the next frame as the prediction position of the target. Typical representative algorithms are as follows: particle filter algorithm, mean shift algorithm, and Kalman filter algorithm. The method based on the discriminant model regards the target tracking problem as a classification or regression problem. In this method, the target is separated from the background by combining the background information with the feature extraction. TLD (tracking-learning-detection) algorithm [2] is the representative of a long-time tracking algorithm in this kind of method. In view of the target deformation, scale change, and occlusion in the process of long-time target tracking, TLD combines tracking with a traditional detection algorithm and updates the model and parameters online to make the tracking more robust and reliable. The target tracking method based on correlation filtering also belongs to the discriminant model method. Based on the minimum output sum of squared error (MOSSE) algorithm [3], correlation filtering is applied to target tracking for the first time.

Through fast Fourier transform, the calculation is transferred from time domain to frequency domain, and the tracking speed is up to 615fps. The speed advantage of the target tracking algorithm based on correlation filtering shows its potential in target tracking. KCF [4] algorithm calculates the discriminant function by regression and introduces the cyclic shift method for approximate dense sampling. The kernel method is introduced to map the input to high-dimensional space, and hog feature is added to improve the tracking effect while maintaining fast calculation. SRDCF [5] introduces spatial regularization and weights the filter coefficients so that the filter coefficients are mainly concentrated in the central area, and the influence of boundary effects is alleviated.

In the method based on deep learning, C-COT [6] combines the shallow surface information and deep semantic information in-depth features, synthesizes the feature map information under multiple resolutions, interpolates the response map in the frequency domain, and then calculates the target position through iteration. SiamRPN [7] algorithm proposes a Siamese network structure based on RPN, which is composed of Siamese network and RPN network. Siamese network shares weights and maps the input to a new space to extract features. The RPN network generates candidate regions, which are used to distinguish the target background and fine-tune the candidate content to achieve end-to-end input and output. SiamMask [8] algorithm changes the previous rectangular box aligned with the coordinate axis to represent the target position, adds mask branch in Siamese network architecture, and generates a rotating rectangle through the target mask, which further improves the tracking accuracy.

Single object tracking (SOT) is the research content of the above target tracking methods. Different from single-object tracking, target tracking in practical application is more multiobject tracking [9] (MOT). The target is locked in the given video sequence, and each target is distinguished in the subsequent frame, and its motion trajectory is given. According to the initialization method of the target box, the multitarget tracking method is divided into two categories: DBT (detection-based tracking) and DFT (detection-free tracking). DFT needs to manually initialize the location box of the target, and it cannot deal with the new target problem in the video; DBT can detect new targets automatically and end the trajectory of the target leaving the visual field. In the multitarget tracking method, the key problem [10] is to detect the data association between nodes and existing trajectories and the correlation between trajectories. Xiang et al. [11] transformed the multitarget tracking problem into Markov decision process (MDP). The target trajectory is set to four different states, and the trajectory state and state transition process are described by MDP modeling and decision-making. Sort algorithm [12] uses Kalman filter algorithm to track the detected target, calculates the distance between IOU (intersection over union) measurement target frames, and performs optimal association matching through Hungarian algorithm. Deep sort algorithm [13] is improved on the basis of sort algorithm. Fast r-cnn is used to detect the target, and the Kalman filter is still used to track and predict the target. In distance measurement, Mahalanobis distance and the minimum cosine distance between the nearest depth feature set successfully tracked by the target and the feature vector of the detection result are integrated, and priority is assigned to the target through cascade matching. The problem of track association of target occlusion is solved. In the multitarget tracking method based on deep learning, Feng et al. [14] proposed a unified multitarget tracking framework. Siamrpn network is used for short-term target tracking, and the appearance characteristics of the long-term target are integrated. Reid network is used to improve the tracking stability when the target is occluded and deal with abnormal motion. Based on association matching, switch aware classification (SAC) is proposed to achieve a good multitarget tracking effect. However, due to the complexity of the model, the tracking speed is slow, which cannot meet the practical application.

It is still an important research task to track multitarget continuously and track accurately in complex traffic scenes. It is of great value to improve the utilization efficiency of traffic video monitoring data, timely and accurately to grasp road traffic information and regional road operation status. The cross-camera multitarget tracking can solve the problem that monocular camera cannot track accurately for a long time and a long distance, which lays an important foundation for the acquisition of wide-angle traffic information.

2. Principle of Multitarget Tracking

Traffic scene is a typical multitarget tracking application scene. This paper uses DBT detection target box to realize multitarget vehicle tracking in traffic scene. The process flow of multitarget tracking based on DBT is shown in Figure 1. The target detector will first detect the target in each frame of the video to obtain and identify multiple target positions. Multitarget tracking process is to associate the current detection result with the existing target track to extend the track.

Next, we need to solve the problem of effective association between trajectory and target. In cross-camera multitarget tracking, the first step is to obtain the multitarget vehicle trajectory in a single camera. Referring to the latest research results of the team [15], the similarity between the target frames is calculated based on IOU, and the Hungarian algorithm is used to complete the association between the new detection node and the existing vehicle trajectory. The definition and delimitation method of the stage and state of the trajectory are proposed to better classify the trajectory. Then, through cross-camera vehicle tracking, the problem of 3D trajectory reconstruction based on combined camera calibration in the overlapping area is solved, as well as the similarity association and cross-camera trajectory update between cross-camera trajectories, and the trajectory transfer between adjacent cameras is completed.

3. Data Association Based on Cross-Camera Calibration

For the multicamera monitoring scene with the overlapping area, as shown in Figure 2. In a long area, there are many cameras. From the end with the smaller camera number in the monitoring area, renumber the cameras from 0 in turn. Each camera is responsible for monitoring a section of the Road area. In Figure 2, different color blocks are used to mark the monitoring area of each camera. There is a view overlap between adjacent cameras, and the overlap area is indicated by yellow. On the premise of cross-camera calibration, the similarity association can be completed by calculating the similarity matrix of vehicle trajectories between adjacent cameras. The basic idea is through the joint calibration of multiple cameras, and the cameras are unified in a world coordinate system, and the similarity matrix is calculated according to the Euclidean distance of the track points in the adjacent cameras in the world coordinate system.

3.1. Cross-Camera Joint Calibration.

According to the imaging principle of the monocular camera and the description of the coordinate system in reference [16], the conversion relationship from pixel coordinate system to world coordinate system under the same camera can be obtained as follows:

$$P_W = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} K p, \tag{1}$$

where $P_W = \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}$ is the coordinates of the point in the

world coordinate system, $P_{uv} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ is the coordinates of the

point in the pixel coordinates,

$K = \begin{bmatrix} Z_C dx/f & 0 & -Z_C u_0 dx/f \\ 0 & Z_C dy/f & -Z_C v_0 dy/f \\ 0 & 0 & Z_C \\ 0 & 0 & 1 \end{bmatrix}$ is the camera's in-

ternal parameter matrix, and $\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$ is the camera's ex-

ternal parameter matrix.

The above formula is derived without considering distortion. If the distortion of the camera is considered, it can be divided into radial distortion and tangential distortion. For the image physical coordinate system, the corresponding radial distortion correction is shown in equation (2), and the corresponding tangential distortion correction is shown in equation (3). The corresponding formula can be introduced for parameter correction.

$$\begin{cases} x_{corrected} = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right), \\ y_{corrected} = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right), \end{cases} \tag{2}$$

$$\begin{cases} x_{corrected} = x + \left[2p_1 y + p_2\left(r^2 + 2x^2\right)\right], \\ y_{corrected} = y + \left[p_1\left(r^2 + 2y^2\right) + 2p_2 x\right]. \end{cases} \tag{3}$$

The conversion process of the world coordinate system between multiple cameras is as follows: first, a coordinate origin is selected, and the corresponding subworld coordinate system of each camera into the global unified world coordinate system is constructed. The schematic diagram of calibration conversion between adjacent cameras in a large area is shown in Figure 3. Taking two cameras as an example, the monitoring road is two lanes. Suppose that 3 (a) is the monitoring scene of camera i and 3 (b) is the monitoring scene of camera $i + 1$, through cross-camera calibration, the points under the field of view of each camera are converted to the same world coordinate system, as shown in 3 (c).

3.2. Calculation of Association Matrix.

The association matrix calculation of cross-camera vehicle trajectories is to calculate the Euclidean distance between adjacent camera trajectories to be matched after the trajectories are transformed from image coordinates to world coordinates. Suppose that the vehicle trajectories under each camera are divided into two sets: $T = \{RT \cup NT\}$, RT represents the real track set in the scene, NT represents the new track set that has just changed from the undetermined track to the real track, and the similarity matrix of vehicle trajectories between adjacent cameras is $D = (d_{ij}), d_{ij} \geq 0$, where d_{ij} is calculated as follows:

$$d_{ij} = \frac{\sum_{i=1}^m |P_k - P_k'|}{m}. \tag{4}$$

In formula (4), m is the number of trajectory nodes involved in the calculation. In this paper, $m \leq 15$ is determined by the number of nodes of the trajectory to be matched between adjacent cameras under the same frame number; p_k is the world coordinate of the track point in the real track RT of the current camera; p_k' is the world coordinates of the track points in the new track NT of the adjacent cameras, and the frame numbers of p_k and p_k' are the same, indicating the vehicle position at the same time. Taking $m = 5$ as an example, the calculation process of d_{ij} between vehicle trajectories across cameras is shown in Figure 4.

4. Multitarget Vehicle Tracking Algorithm across Cameras

Cross-camera vehicle tracking relies on the unified calibration between multiple cameras and single-camera multitarget vehicle tracking. Its main work is to associate the
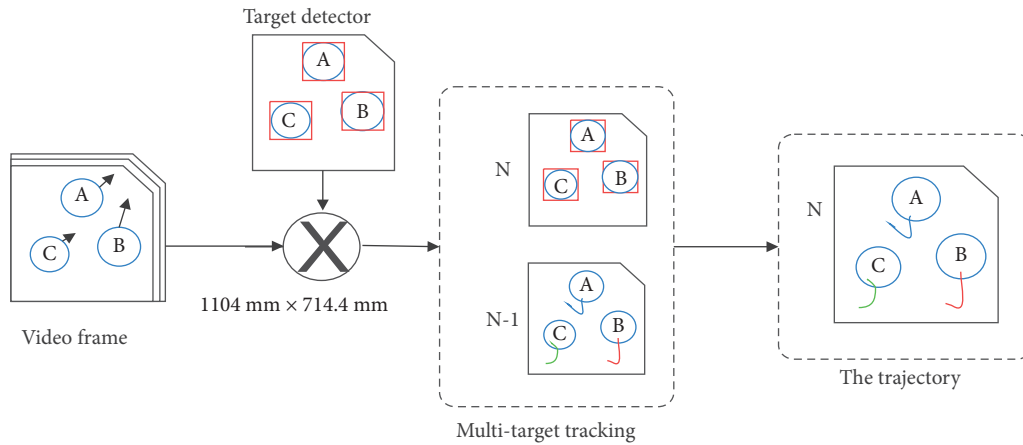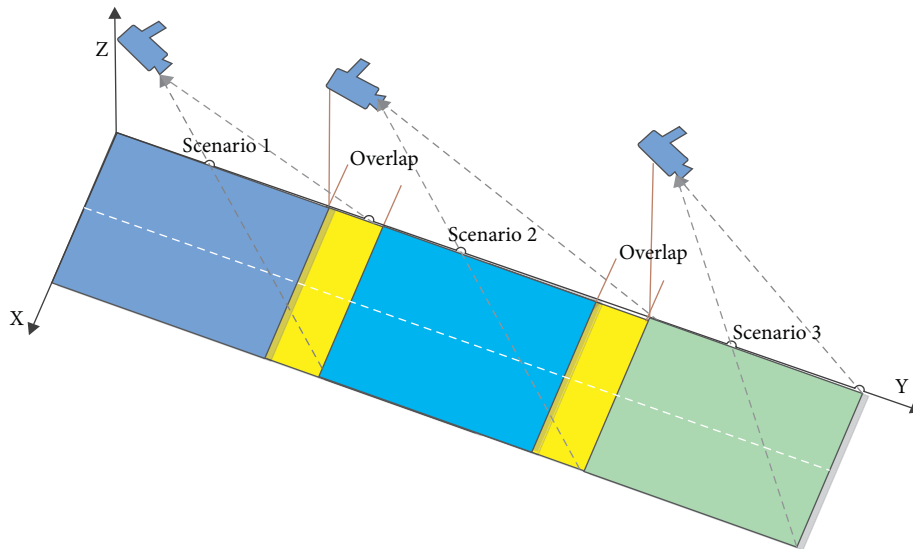
FIGURE 1: Multitarget tracking process.

FIGURE 2: Multicamera scene stitching.

tracking results of each camera. First, the global trajectory set *GT* is established to save the global trajectory information of the vehicle target from entering the monitoring area to leaving the monitoring area. When the target leaves the monitoring area, the corresponding vehicle target information is recorded in the file. After the cameras in the monitoring area are synchronized and the video frames are associated with each other, the new trajectory nodes will be updated into the global trajectory set *GT*.

Assuming that there are *n* cameras in a large monitoring area, the flowchart of the cross-camera vehicle tracking algorithm is shown in Figure 5, and the steps of the cross-camera vehicle tracking algorithm are as follows:

Step 1: the vehicle trajectory of *N* cameras is obtained at the same time. The multitarget vehicle trajectory in a single camera is obtained by the method in Section 2.

Step 2: association between adjacent cameras is tracked. The vehicle track in each camera is divided into two

sets: real track set *RT* and new track set *NT*. The track matching association method is set with the current camera number *i*. Considering the two-way driving of the vehicle, the similarity matrix is calculated between the real track set under *i* camera and the new track set in *i* − 1 and *i* + 1 of adjacent cameras. The final matching association results are as follows:

(i) The real trajectory does not match, indicating that the vehicle has not entered the overlapping area, and the trajectory attributes do not need to be changed.

(ii) The new trajectory does not match. As a new target entering the monitoring range, the target ID and trajectory color are assigned and recorded the starting frame number of the trajectory.

(iii) The real trajectory is successfully matched with the new trajectory, and the attributes of the vehicle trajectory are updated by migration. The matching trajectory information is migrated to the new

(a)

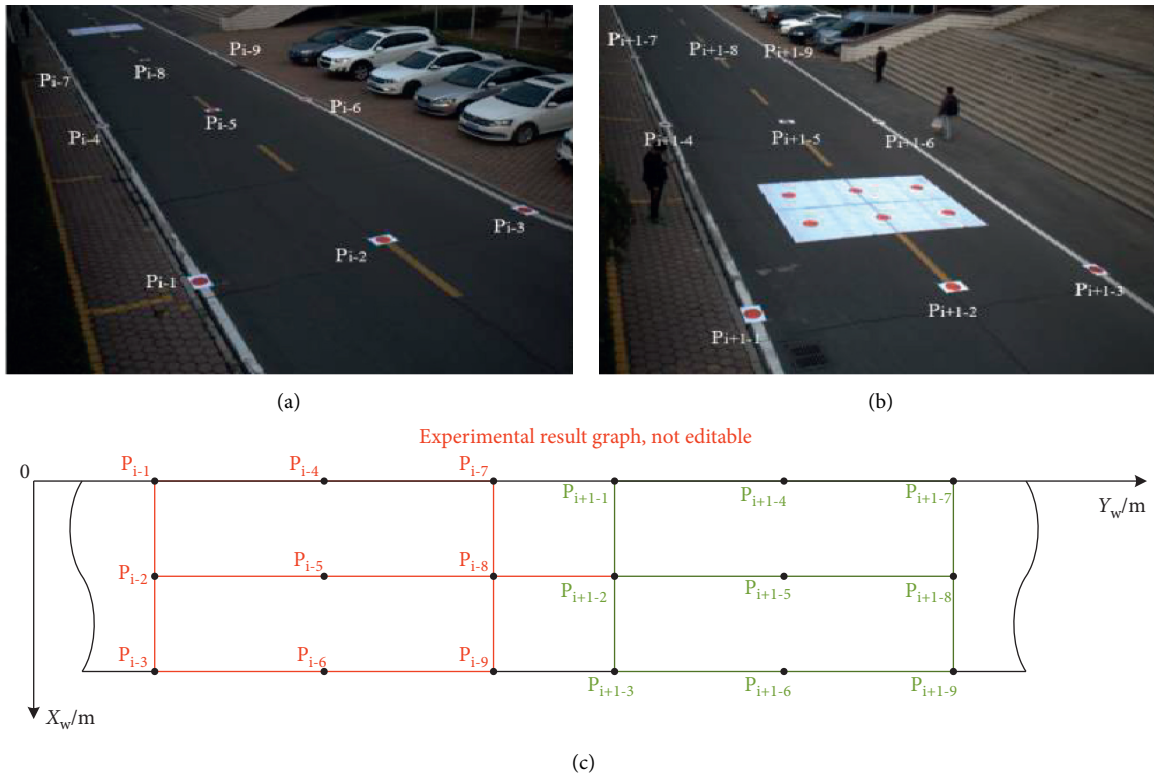(b)

Experimental result graph, not editable

(c)

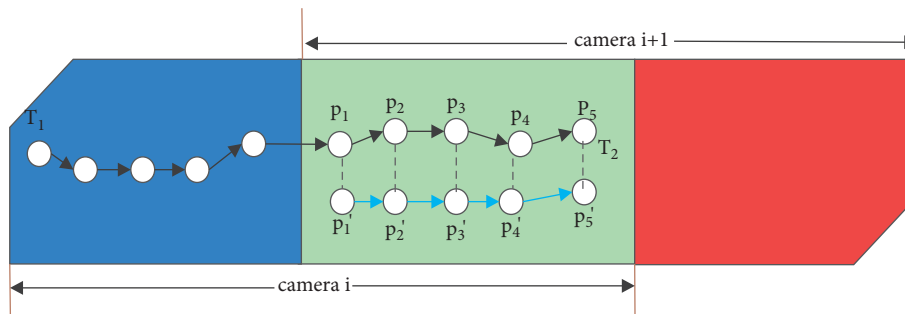FIGURE 3: The calibration conversion between adjacent cameras.

FIGURE 4: The principle of trajectory calculation.

trajectory, including target ID, trajectory color, and updated some trajectory attributes. Among them, the camera number and the starting frame number of the track under the camera are used to draw the target track under the camera.

Step 3: the global track is updated. Every frame needs to update the global trajectory:

(i) For unmatched real trajectories, the newly added trajectory nodes need to be updated into the global corresponding trajectories

(ii) The unmatched new trajectory is used as a new target, and its trajectory is newly added to the global trajectory

(iii) Between the successfully matched real trajectory and the new trajectory, in addition to the above-mentioned trajectory attribute changes, it is also necessary to fuse the trajectory nodes in the overlapping area of the two trajectories

Figure 6 shows the successful matching of vehicles between adjacent cameras. When the target vehicle moves from the current camera to the next camera, the vehicle will be in the overlapping area of the two cameras. The successfully matched vehicle target ID needs to be unified, and the vehicle trajectory color will follow the initial color attribute. In Figure 7, when a black car is driven from camera 0 field of view to camera 1, the black car can be detected in both camera fields of view in the overlapping area. The two cars connected by the yellow line are the position of the black car under the two cameras. The target vehicle is matched in the overlapping area, and the vehicle information is transferred to camera 1.
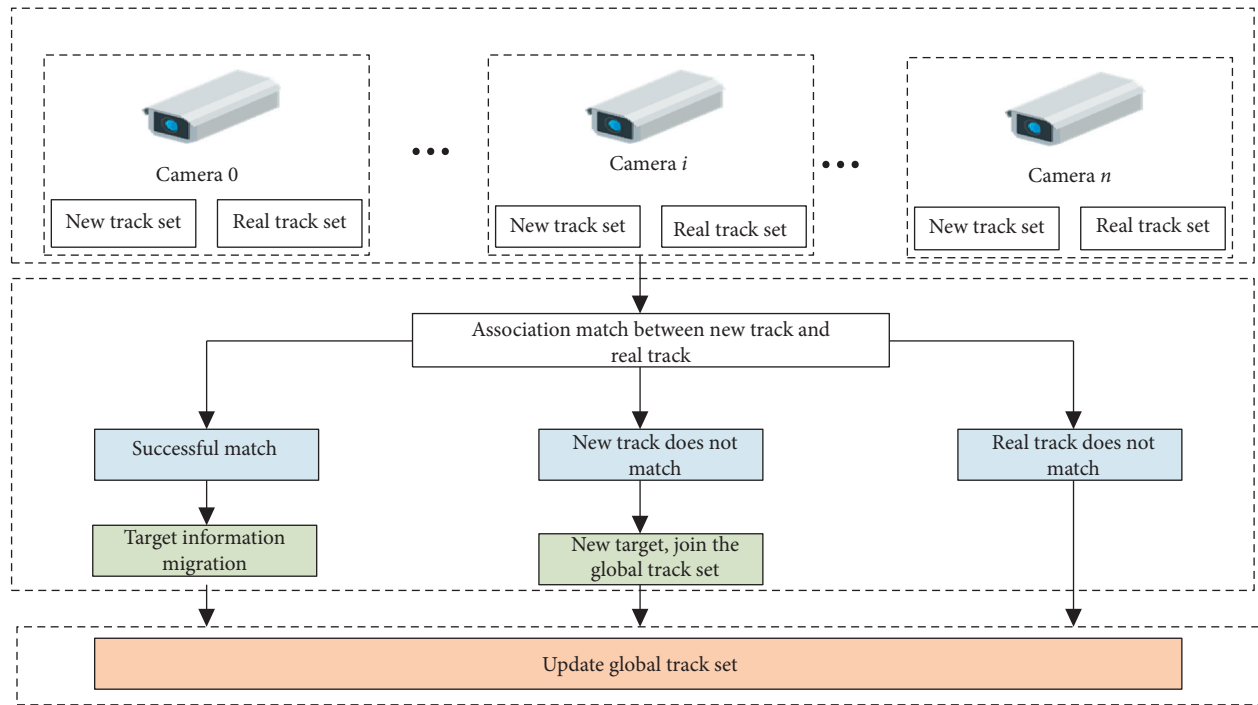
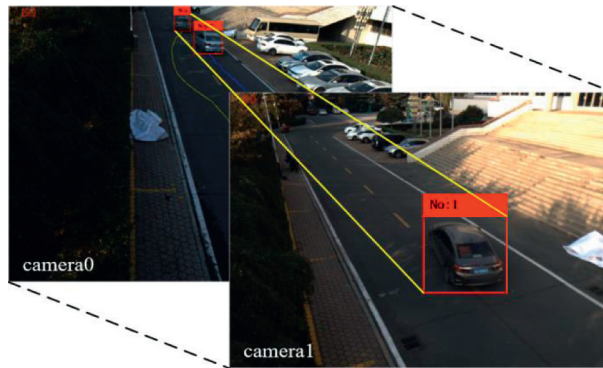Figure 5: Cross-camera vehicle tracking flowchart.

Figure 6: Matching results of adjacent camera scenes (experimental result graph, not editable).

5. Experiment and Analysis

Since this method is still in the simulation testing stage, there is no special scenario suitable for the experiment in the open data set. Cross-camera vehicle tracking takes the simulation test scene built-in campus as an example, in which two cameras collect images synchronously. After the detection results of the yolov3 detector are obtained, we load reference [15] and the algorithm in this paper to carry out the waiting tracking experiment and obtain the following experimental results. The following is a scene test of overtaking. The silver car first enters the surveillance area of camera 0, and the black car overtakes, as shown in Figure 7. In the collection of 58 frames of photos, two cars can be detected at the same time under camera 0. Since the silver car enters the field of view first, it will be detected first, with ID = 1. Enter after the black car, ID = 2. After overtaking the black car, it first enters

the camera 1 field of view. However, when the vehicle is driving across the camera, ID values are assigned in the order in which it first enters the entire monitoring area. After the cross-camera trajectory is matched, the trajectory information is migrated, so the ID of the silver car in camera 1 is still 1, and the trajectory color is blue, which is the same as the trajectory information of the vehicle under camera 0. The black car is the same as above for trajectory information migration. Figure 8 shows the tracking result of camera1 at frame 78.

After the two scenes are calibrated across cameras, the vehicle trajectory can be drawn in the panoramic view of the cross-camera reconstruction of the surveillance scene. Taking the 70th frame photo of the multitarget vehicle tracking panoramic reconstruction image as an example, you can intuitively see the entire overtaking process of the vehicle under the two cameras, as shown in Figure 9. The
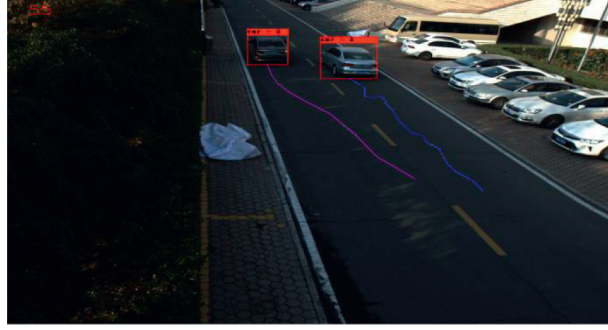
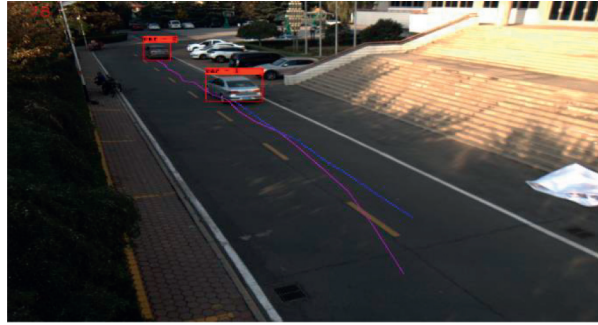FIGURE 7: Frame 58 of camera 0 experimental result graph, not editable.

FIGURE 8: Frame 78 of camera 1 experimental result graph, not editable.
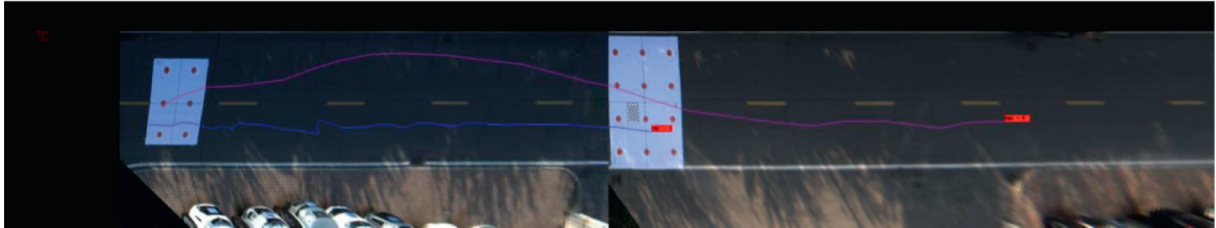

FIGURE 9: Reconstructed panorama (experimental result graph, not editable).

TABLE 1: Coincidence of trajectory under different cameras.

Trajectory coincidence degree	Camera 0 (%)	Camera 1 (%)	Restructure (%)
—	8	23	5

result of this panoramic reconstruction allows a real overview of the operating state of the vehicle from a macro perspective and is not affected by the loss of the occluded trajectory. The reliability of the data is a major technological breakthrough.

In order to further verify the effectiveness of the proposed method, the trajectory coincidence degree TC is used for description, and its definition formula is as follows:

$$TC = \frac{\sum_{0 \leq i < m, 0 \leq j < n} \left| P_{iA} - P_{jB} \right| < T}{m * n}. \tag{5}$$

Among them, m, n represents the number of discrete points on trajectory A and trajectory B, and P_{iA}, P_{jB} are points on trajectory A and trajectory B. By calculating the absolute distance between each point in trajectory A and each point in trajectory B, and then accumulating the number of distances less than the threshold T divided by the product of the number of discrete points on the two trajectory curves, the evaluation value of the coincidence degree of the two trajectories is obtained. The degree of trajectory coincidence obtained in the experiment is shown in Table 1.

It can be seen from Table 1 that the method in this paper unifies the cameras in a world coordinate system for target tracking and association matching. Results: the coincidence degree between trajectories was the lowest in camera0 and camera1, and the effect of trajectory-based target behavior analysis was the same as that of observation from high

altitude. So that the problem of occlusion overlap does not appear in the 2D image, and it can intuitively reflect the whole running state of the target in the large scene. Not only that, the proposed method also meets the real-time requirements.

6. Conclusion

Through the joint calibration between multiple cameras, the cameras are unified under a world coordinate system. The Euclidean distance between the trajectory nodes under the overlapping area at the same time is used to measure the similarity between the trajectories, and the trajectory association matrix is calculated to realize the matching between the real trajectory in the current camera and the new trajectory under the adjacent camera. Target tracking and association matching under single camera and cross-camera complete the trajectory transfer of the vehicle between adjacent cameras and realize the 3D bird's-eye view reconstruction of the vehicle trajectory. The result proves that the operating state of the vehicle can be viewed from a real macro perspective, and the data are reliable, which is a major breakthrough. It makes the long-term and long-distance continuous tracking of multiple targets across cameras reliable and accurate.

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.

Authors' Contributions

Junfang Song and Tao Fan mainly engaged in image processing and artificial intelligence research. Huansheng Song mainly engaged in image processing and recognition and intelligent transportation systems research. Haili Zhao mainly engaged in image processing and information security research.

Acknowledgments

References

[1] L. Xi, Y. Cha, T. Zhang et al., "Overview of deep learning target tracking algorithms," *Chinese Journal of image graphics*, vol. 24, no. 12, pp. 2057–2080, 2019.

[2] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking learning detection," *IEEE Transactions on Software Engineering*, vol. 34, no. 7, pp. 1409–1422, 2011.

[3] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of the The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition*, no. 6, pp. 13–18, CVPR 2010, San Francisco, CA, USA, 2010.

[4] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[5] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4310–4318, Santiago, Chile, December 2015.

[6] M. Danelljan, A. Robinson, F. Shahbaz Khan, and M. Felsberg, "Beyond correlation filters: learning continuous convolution operators for visual tracking," in *Proceedings of the 14th European Conference on Computer Vision*, pp. 472–488, Spring, Amsterdam, The Netherlands, October 2016.

[7] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8971–8980, Salt Lake City, UT, USA, June 2018.

[8] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. S. Torr, "Fast online object tracking and segmentation: a unifying approach," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1328–1338, Long Beach, CA, USA, June 2019.

[9] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: a literature review," *Artificial Intelligence*, vol. 293, no. 3, Article ID 103448, 2021.

[10] P. Emami, P. M. Pardalos, L. Elefteriadou, and S. Ranka, "Machine learning methods for solving assignment problems in multi-target tracking," pp. 1–33, 2020, arXiv https://arxiv.org/abs/1802.06897.

[11] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: online multi-object tracking by decision making," in *Proceedings of the IEEE international conference on computer vision*, pp. 4705–4713, Santiago, Chile, December 2015.

[12] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proceedings of the 2016 IEEE international conference on image processing (ICIP)*, pp. 3464–3468, IEEE, Phoenix, AZ, USA, September 2016.

[13] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proceedings of the 2017 IEEE international conference on image processing (ICIP)*, pp. 3645–3649, IEEE, Beijing, China, September 2017.

[14] W. Feng, Z. Hu, W. Wu, J. Yan, and W. Ouyang, "Multi-object tracking with multiple cues and switcher-aware classification," pp. 1–10, 2019, arXiv https://arxiv.org/abs/1901.06129.

[15] Li Ying, *Research on Traffic Video Intelligent Analysis System Based on Target Detection and Tracking*, Chang'an University, Xi'an, China, 2019.

[16] J. Song, H. Song, and S. Wang, "PTZ camera calibration based on improved DLT transformation model and vanishing Point constraints," *Optik*, vol. 225, no. 7, Article ID 165875, 2021.