International Journal of Intelligent Transportation Systems Research (2021) 19:312–323
https://doi.org/10.1007/s13177-020-00247-w

An Improved Method of Nonmotorized Traffic Tracking and Classification to Acquire Traffic Parameters at Intersections

Xiaohan Liu[1,2,3] · Hao Wang[1,2,3] · Changyin Dong[1,2,3]

Received: 27 July 2020 / Revised: 8 December 2020 / Accepted: 15 December 2020 / Published online: 13 February 2021

Abstract

High computational cost and low tracking stability make it still a challenging task to acquire nonmotorized traffic parameters at intersections via vision-based method. In order to address the above issues, our study improves a cooperative tracking and classification method, and proposes a vision-based data collection system to monitor nonmotorized traffic at intersections. The system utilizes the combination of two tracking algorithms, Kernelized Correlation Filter and Kalman filter, to ensure the continuous tracking. Based on multivariate feature, K-means clustering and Support Vector Machine are implemented to classify nonmotorized traffic according to the motion and appearance feature respectively. As a result, the proposed system can acquire trajectories of pedestrians and cyclists and extract traffic parameters, including flow and velocity. Our method performs well in both efficiency and accuracy by fusing simple but effective algorithms and is robust in the complex scenario especially at large-scale intersections with limited training samples. The experimental results show that it can extract more trajectories with low computational lost. Moreover, the error of flow and velocity result is controlled within acceptable limits, which directly proves it feasible to collect field data in project applications.

Keywords Nonmotorized traffic · Intersection · Computer vision · Tracking · Object classification

1 Introduction

Nonmotorized traffic plays an important role in the whole traffic system. More and more people choose to walk or bike to their destination, especially for a short distance travel or the connection to other traffic modes. Nonmotorized road users are more vulnerable to injuries than other due to their labile velocity and direction [1]. Some research has proved that higher nonmotorized traffic rates lead to safer road, but the pedestrian and bicycle fatalities increase, both in absolute numbers and the proportion of all traffic [2]. Furthermore, intersections are among the most dangerous locations of a roadway network due to complex traffic conflicting movements. In order to ensure an efficient and safe operation at intersections, nonmotorized traffic parameters are indispensable.

Computer vision provides a good way to obtain traffic parameters from video. There are also some other methods, such as infrared detectors, pneumatic tube, radio beam, inductive loop, piezoelectric strip, radar, thermal imaging sensors, etc. Compared with them, vision-based method is easy to operate and the cost is low. It can obtain a variety of traffic data in higher dimensions. Therefore, this method has received widespread attention in recent years. FHWA and the NHTSA initiated Bicycle-Pedestrian Count Technology Pilot Project to collect more accurate data on pedestrian and bicyclist behavior in 2015 [3].

In general, a successful vision-based system needs multiple technologies naturally incorporated, one of which is target tracking. Tracking models are mainly divided into two types, generative models and discriminative models. Generative models use the target information in current frame to model and find the most model-conforming region as the target in the next frame. The representative generative models include Kalman Filter (KF) [4], mean shift [5], etc. Discriminative models make full use of background information

✉ Hao Wang
haowang@seu.edu.cn

[1] Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing 210096, People's Republic of China

[2] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing 210096, People's Republic of China

[3] School of Transportation, Southeast University, Nanjing 210096, People's Republic of China

to train a detector, which detects the target in each frame. The representative discriminative models include correlation filter [6, 7], convolutional neural network (CNN) based models [8], etc. Another key technology is traffic objects classification. It is to determine the type of nonmotorized traffic. At present, deep learning has been used widely for vehicle classification [9, 10].

As the technology improved, some achievements have been made in the field of traffic data acquisition. Compared with nonmotorized traffic, technology of obtaining vehicle data is more mature. Currently, vehicles on the road can be detected and tracked to monitor their behavior precisely [11, 12]. Compared with vehicle, nonmotorized traffic objects are small and morphologically transformed, which increases tracking difficulty and has limitations on applicable scenarios. Many researches were devoted on the detection and tracking technologies for pedestrian [13–15]. Zhao et al. investigated deep learning to track pedestrians at construction sites [16]. For the application in intersections, Li et al. used KF to track nonmotorized targets and chose backpropagation neural network to identify pedestrians and bicycles [17]. Zhu et al. used the Kernelized Correlation Filter (KCF) to track and simulate pedestrian movements at intersections [18]. Guo et al. analysed the pedestrian walking behaviour and walking mechanism represented by gait parameters acquired from video [19]. Shirazi used contextual fusion of motion and appearance cues to more reliably track pedestrians during stop-and-go movements at intersections [20]. Although some progress has been made in the existing researches, there is still great potential in acquiring nonmotorized traffic data at intersections through video processing.

- Currently, nonmotorized traffic tracking performs well in small-scale scenarios , such as intersection approach. For large-scale scenarios such as a large intersection, nonmotorized traffic targets are small and blurred, causing tracking easy to lose.
- Some processing algorithms need high computational cost, which is limited by the hardware and hard to get the real-time data. It is unfeasible to apply into the project.
- Many researches concentrate on the pedestrian detection and tracking algorithms. Various traffic parameters need to be acquired accurately and data reliability needs to be verified. Additionally, few studies have verified the reliability of velocity and cyclist should be paid more attention.

Therefore, this study presents an effort to address the aforementioned issues. A vision-based system is organized to automatically complete nonmotorized traffic parameters collection. To improve the tracking continuity of nonmotorized traffic at large intersections, generative and discriminative algorithms are integrated, where KCF is the major tracking algorithm and KF is cooperated to predict the possible position of missing target. Prediction with KF relies on the trajectory before tracking is lost. This system also benefits from a simple but effective classification methods. Specifically, K-means clustering and Support Vector Machine (SVM) can take advantage of motion and appearance differences between pedestrians and cyclists to classify them. Ultimately, complete trajectories can be extracted within acceptable computational cost. More importantly, the extraction traffic parameters are based on trajectories, which ensures the accuracy and high-dimension of the data.

The remainder of this paper is organized as follows: The second part introduces the framework of our data collection system and the tracking and classification algorithms. The third part presents the experimental results and compares with other algorithms. The fourth part discusses the potential problems in the system and outlooks on future research. The last part concludes the paper.

2 Methodology

This section presents the framework of nonmotorized traffic data collection system and describes the improved tracking and classification methods.

2.1 Framework of Data Collection System

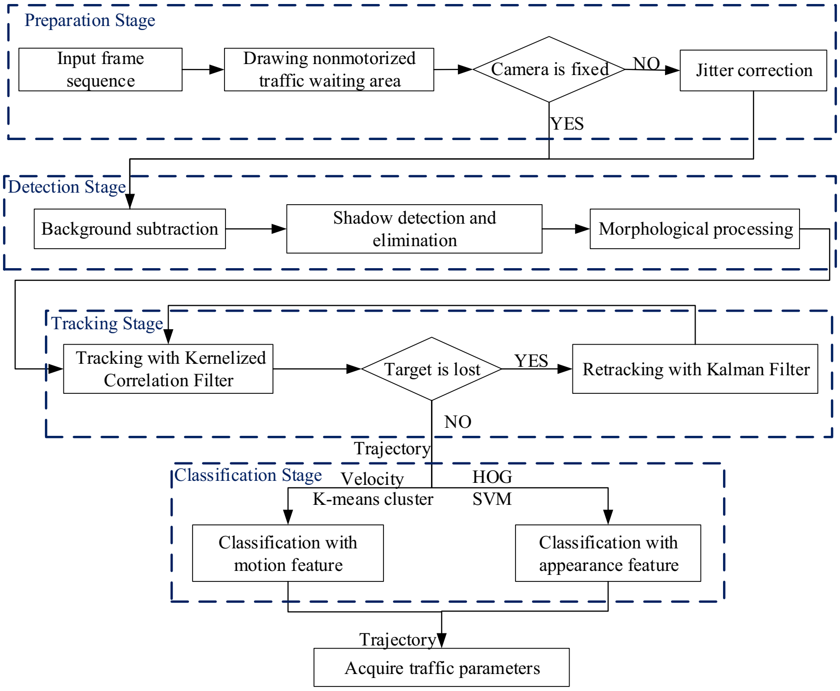
This system consists of several basic functional parts: target detection, target tracking, target classification and trajectory processing. Figure 1 presents the whole framework of the system. This subsection will introduce the preparation and detection stage and explain how to acquire traffic parameters from trajectories. Tracking and classification stage will be emphatically illustrated in the next subsections.

1) Preparation Stage

As the video pre-processing stage, preparation stage needs to provide the ideal state for video processing.

- Nonmotorized traffic waiting area is defined to account for regions where the nonmotorized traffic waits to pass at intersections. Drawing the waiting area can concentrate the processing area and prevent extra motor vehicle detection, which will significantly reduce the computational cost.
- Target detection is affected by shooting condition. Specifically, if the video is taken by a moving camera, slight jitter will produce many false positive detections. With the principle of affine transformation, slight jitter can be eliminated. Affine transformation used 2*3 matrix trans-

Fig. 1 Framework of nonmotorized traffic parameters collection system at intersections


Preparation Stage
Input frame sequence
Drawing nonmotorized traffic waiting area
Camera is fixed
NO
Jitter correction
YES
Detection Stage
Background subtraction
Shadow detection and elimination
Morphological processing
Tracking Stage
Tracking with Kernelized Correlation Filter
Target is lost
YES
Retracking with Kalman Filter
NO
Trajectory
Classification Stage
Velocity
K-means cluster
HOG
SVM
Classification with motion feature
Classification with appearance feature
Trajectory
Acquire traffic parameters

formation to transform an image from two-dimensional coordinates to another. Therefore, three key points were found in the video starting frame, and the corresponding points would be found in each subsequent frame by key point tracking. Jitter will be removed based on the parallelogram relationship constructed by the three points.

2) Detection Stage

Target detection aims to find out complete nonmotorized traffic target contours, so as to provide them for tracker in the follow-up process.

- Background subtraction was used to filter the foreground. This method uses a reference image as the background model and subtracts the pixel value of the current frame image from the corresponding point of the reference image, which can adapt to the videos from different shooting angles and heights. Gauss Mixture Model worked to construct background model for accurately segmenting foreground target [21]. Once the nonmotorized targets moved, they can be detected immediately.
- The detected foreground contains part of shadow, which causes some problems like adhesion. With the feature that shadow in Hue-Saturation-Value color space has lower value and almost no change in hue compared with the background, shadow can be well detected and eliminated.
- Morphological methods like corrosion and expansion were implemented in final to obtain the complete counter.

Figure 2 shows the effect change in detection stage. The final effect in Fig. 2(c) meets the detection request, where traffic objects are emerged with complete counter.

3) Traffic Parameters Acquisition

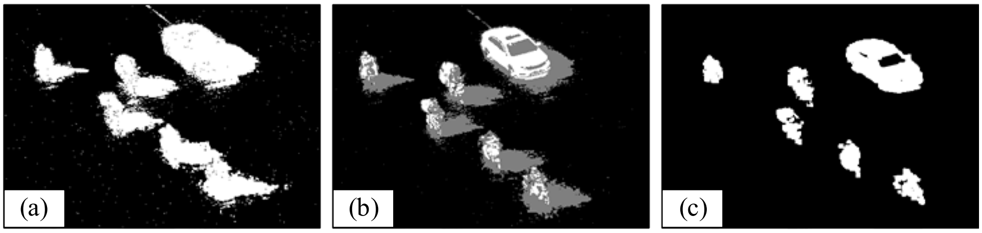


Fig. 2 Detection result **a** Background subtraction, **b** Shadow detection, **c** Morphological reconstruction

Traffic parameters acquisition requires attaching space-time dimension to trajectory. Time dimension can be easily established by recording frame sequence in video processing, while space dimension must be established by coordinate transformation. Eventually, a variety of traffic parameters comes from these proceeded trajectories.

2.2 Tracking Algorithm

It is necessary to track the detected nonmotorized traffic objects to get their position at each moment, and then draw their trajectories at intersections. KCF tracking algorithm worked as the major tracking algorithm, with KF as the auxiliary tracking algorithms to retrack lost targets. This combination is extremely synergistic, taking into account both efficiency and accuracy, and obtains as many complete trajectories as possible in acceptable time.

2.2.1 KCF Tracking Algorithm

KCF is a typical algorithm in correlation filter class [6]. The basic idea of KCF is to construct a large number of training and test samples by cyclic shifting, selecting the target region image as positive sample, and the surrounding environment images as negative samples. The sample is mapped to a linearly separable space by a kernel function, in which the target detector is trained by ridge regression. An outstanding contribution of KCF algorithm is that all the samples acquired by cyclic shift are diagonalized in the Fourier space using the discrete Fourier matrix, which reduces the amount of data stored and computation and greatly improvs the tracking speed.

For each frame of video, the response value of test samples is calculated by the trained detector, and the sample with the largest response value is selected as the new tracking target. Then, the training set is updated by the new detection results, and the target detector is updated correspondingly. Tracking is thus established and the tracking effect is shown in Fig. 3. The red box is the previously drawn nonmotorized traffic waiting area, and only the targets passing through this area have possibility to be identified as nonmotorized traffic targets. The blue boxes indicate that nonmotorized traffic targets are being tracked.

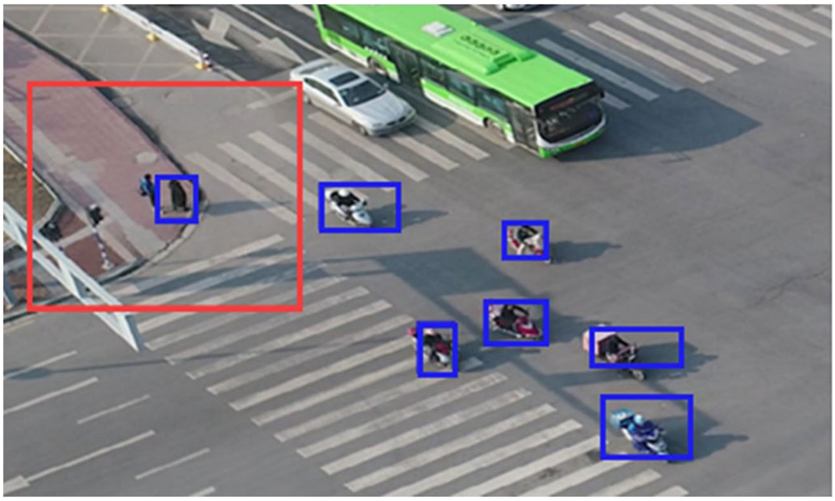
2.2.2 KF for Tracking

Although KCF algorithm is a well-behaved tracking algorithm, tracking lost will appear in some cases. Thus, KF is used to retrieve lost targets. Because the time interval between the adjacent trajectory points is very short, the state can be approximately represented by a linear system, which is suitable for KF to predict the possible location of missing targets and rebuild tracking.

KF is an optimized autoregressive data processing algorithm. It uses a linear stochastic differential equation to calculate the current optimal value based on the current measured value, previous predicted value and error, and can predict the value at the next moment.

The application of KF in tracking is to predict the possible position of the missing target according to its previous trajectory. If there is an object around the prediction point, the object is identified as the missing target and the tracking is rebuilt. The system state can be represented as a 4×1 matrix, which stores the target horizontal and vertical coordinates and the change in the horizontal and vertical coordinates. The measured value is a 2×1 matrix, which stores the target horizontal and vertical coordinates. Figure 4 shows the retracking effect. The blue box attached the man on the right insinuates that he will lose tracking due to the white car shadow, while the green box shows that tracking is retrieved.

Fig. 3 Tracking effect



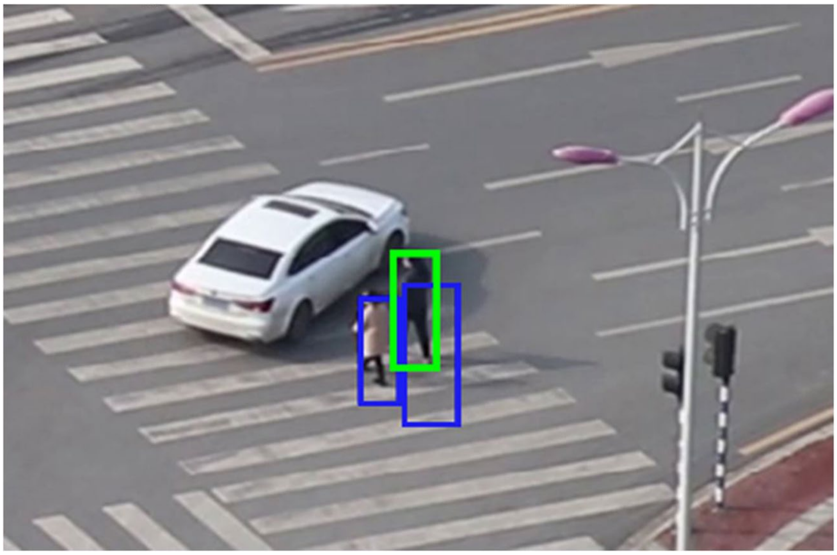


Fig. 4 Retracking effect of lost target

2.3 Classification Algorithm

The category of each target trajectory acquired by tracking can't be distinguished whether it belongs to cyclist or pedestrian. Multivariate feature can efficiently complete classification from multiple perspectives. The category of nonmotorized targets can be determined quickly based on motion and appearance feature.

2.3.1 Classification with Motion Feature

There are obvious differences in velocity between pedestrian and cyclist. Walking velocity is about one meter per second, cyclist velocity is about four to five meters per second, or even higher. We selected mean velocity as the motion feature delivering to K-means clustering algorithm for preliminary classification.

K-means clustering is an unsupervised learning method, which attempts to find the category of data. K-means clustering was used instead of delineating the velocity range for classification, because if the relevant information is not available, the pixel coordinates cannot be converted to actual coordinates, and the velocity range cannot be known. With the increase of computer computing speed, selecting different starting points as clustering centers and clustering for many times to select the result with the least variance can get good classification results in a few seconds.

Although velocity performs well in classification as a kind of motion feature, some cases which are hard to classify correctly also exist. Great differences in vehicle condition and driving habit result in the velocity changing in a wide range. Some pedestrians choose to run through intersections at the end of green light, whose velocity is close to biking, forming the velocity dispute area which is hard to classify. After multiple experiments, it can be proved that nonmotorized traffic can be divided into four categories with K-means clustering. The category with the lowest velocity can be identified as pedestrians, the category with the highest velocity and the second-highest velocity can be identified as cyclists, while the remaining sub-low-velocity category contains some pedestrians with fast velocity and cyclists with slow velocity. Nonmotorized traffic in this category can be regarded as outliers and needs to be further classified to obtain more accurate results.

2.3.2 Classification with Appearance Feature

For targets in the sub-low-velocity category, appearance feature can assist in more accurate judging. Appearance feature adopted the aspect ratio of the circumscribed target rectangle and Histogram of Oriented Gradient (HOG). HOG took the lead, combined with aspect ratio to make the final judgement.

HOG feature is proposed as an image feature descriptor based on gradient direction [22]. It calculates the histogram of the oriented gradient in local image patches and can describe the appearance and shape of the target well, which can be calculated as follows.

$$\begin{cases} G_x(x,y) = f(x+1,y) - f(x-1,y) \\ G_y(x,y) = f(x,y+1) - f(x,y-1) \end{cases} \quad (1)$$

$$\mathrm{M(x,y)} = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (2)$$

$$\theta(\mathrm{x,y}) = \arctan(\frac{G_x(x,y)}{G_y(x,y)}) \quad (3)$$

where f(x, y) is the pixel value. $G_x(x,y)$ and $G_y(x,y)$ denote gradients in the vertical and horizontal directions of (x, y).

M(x, y) and θ(x, y) are the gradient magnitude and direction of (x, y).

Gradient direction is divided into several intervals, which are called bins. The measured region is segmented into small regions, which are called blocks. Block is segmented into small regions, which are called cells. The gradients of all the cells in the block are concatenated to obtain the feature vector of the block.

In this study, the nonmotorized traffic sample size was uniformly normalized to 48×48 for HOG feature extraction. Block size was set to 16×16, block moving step size was set to 8×8, cell size was set to 8×8, and gradient direction was divided into nine bins in [0, π]. Through the size setting, the dimension of feature vector is 900.

HOG feature needs to be conveyed to SVM to achieve classification. SVM is a classification algorithm, which needs training based on a given set of known samples. SVM uses kernels to map the data points in particular dimensions into a space of much higher dimension, where a linear classifier can often be found to separate the two classes. When the number of samples is limited, SVM can achieve good results.

In this study, cyclists were taken as positive samples, pedestrians and other intersection environment were taken as negative samples. The training samples were captured from the field video. The number of positive samples in the training set is 1013, while the number of negative samples is 1181. The number of samples is slightly more than the feature dimension. It is suitable for classification with SVM classifier. The k-fold cross validation was used for training. The proportion of the incorrect classification in the training set is 9.66%, and the accuracy in the test set is 88.24%.

Because the training model has a 11.76% probability of false detection in the test set, so it is necessary to use the aspect ratio to assist the discrimination. The aspect ratio of cyclists is often larger than that of pedestrians. Once the predicted result is cyclist and the aspect ratio of tracking frame is larger than the set threshold, the tracking object is considered to be cyclist. Vice versa, if the predicted result is pedestrian and the aspect ratio is less than the set threshold, the tracking object is considered to be pedestrian. For the very few remaining targets that cannot be determined at last, a simple manual classification is needed.

3 Application and Results

3.1 Application Environment

This study chose a four-leg (Intersection 1) and a three-leg intersection (Intersection 2) as the experimental intersections. Figure 5 shows their location and surroundings. The reasons for choosing these intersections are as follows:

(1) The branches of the intersections are urban main roads, so the intersection size is large and the motor vehicle flow is high.
(2) The intersection is close to subway station and shopping mall, so the nonmotorized traffic flow is sufficient. The average pedestrian and cyclist flow rates were 343 persons/h and 600 vehicles/h at Intersection 1 and 465 persons/h and 180 vehicles/h at Intersection 2.

The experiment was conducted on a sunny day, when there were obvious shadows interfering with the detection. The intersection was shot by UAV from oblique angle. The frame rate of the video is 25 frames per second and the resolution is 1080P. In summary, it is not easy for these intersections to manage nonmotorized traffic detection and tracking.

3.2 Results

3.2.1 Flow Result

The hardware environment chose Core i7-8700 as the central processor unit, which is six cores and twelve threads with 3.2GHz basic frequency. RAM capacity is 16G. Video processing could be completed within 3.15 times as long as the actual video time.

The trajectory results showed that 67 trajectories were detected at Intersection 1, of which 59 were complete, and 8 were lost during the tracking process. And the number of trajectories detected at Intersection 2 was 37, of which 24 were complete. ‘Complete’ means that the nonmotorized traffic keeps in the tracking state in the process of moving from one waiting area to the next. The trajectories can be projected on the video image, as shown in Fig. 6. Green points are pedestrian trajectories and blue points are cyclist trajectories.

Flow data were obtained from trajectories. In order to better reflect the accuracy of flow data acquisition, three indicators are introduced to quantitatively evaluate the counting results, which are *Precision*, *Recall* and *Critical Success Index* (*CSI*). The indicators are defined as follows.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$CSI = \frac{TP}{TP + FP + FN} \tag{6}$$

where *TP* is the true positives value, which denotes the number of correctly detected objects. *FP* is the false positives value, which denotes the number of invalid detections. *FN*

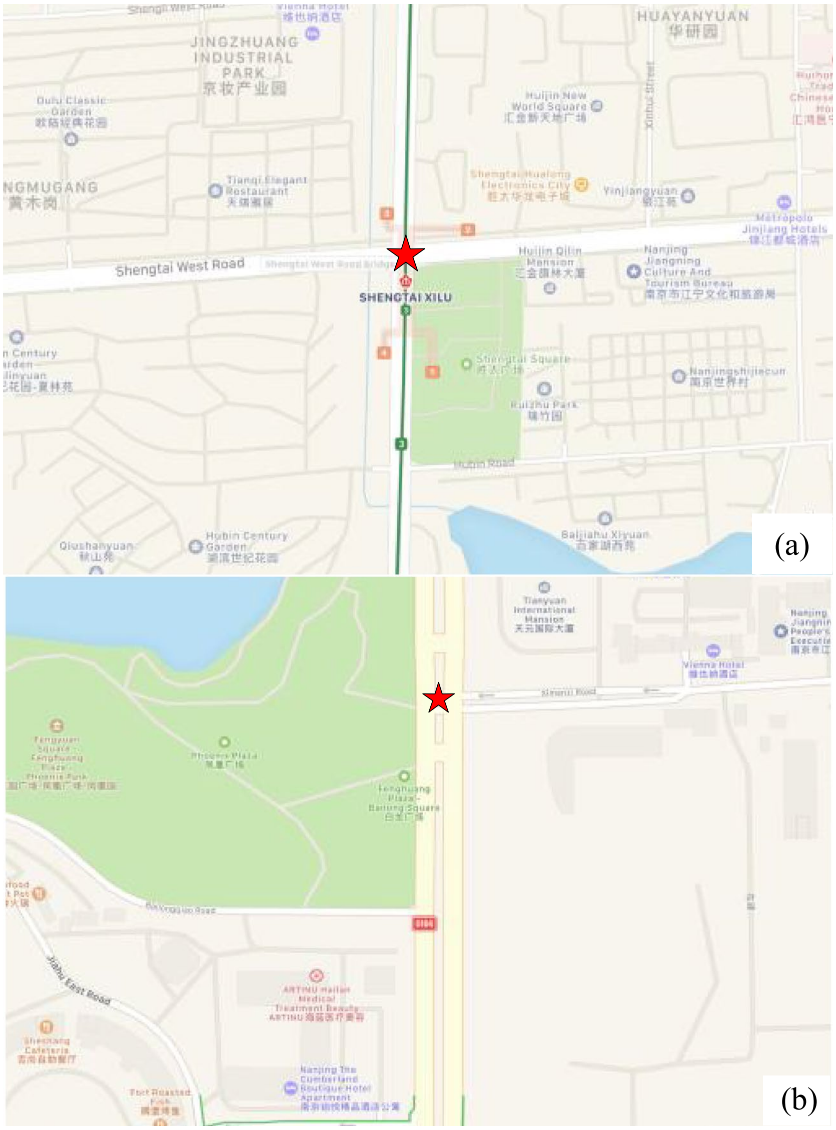


Fig. 5 Intersection location and surroundings **a** Intersection 1, **b** Intersection 2

is the false negatives value, which denotes the number of missed objects. Furthermore, quality is the most important indicator to reflect the final accuracy, since it considers both correctness and completeness.

The flow results of each direction are summarized in Table 1. The results of video detection are compared with the actual flow and mark the differences with underscores. The flow evaluation results are summarized in Table 2. We can see from the table that there is one target missing in the cyclist test and four targets missing in the pedestrian test for the detection of Intersection 1. Two cyclist targets are false detections. Qua is more than 80% for pedestrian and more than 90% for cyclist. The overall Qua is nearly 90%, which can be accepted in application. As for Intersection 2, a lot of occlusions occur due to the high density of non-motorized traffic. Qua drops to 70%. Therefore, our system needs to improve performance in high-density scenarios. By the way, the system saved the picture of tracked objects, finding that only 1 misclassification occurs in the whole 93 classifications.

3.2.2 Velocity Result

Velocity data were also extracted from the trajectories at Intersection 1. The minimum, maximum and average values of mean velocity for pedestrians were 1.28, 2.12 and 1.55 meters per second respectively. Similarly, the relative settings for cyclist were 1.91, 8.88 and 4.78. Standard deviation was 0.21 for pedestrian and 2.05 for cyclist. Cyclists had greater velocity fluctuations at intersections. And the instantaneous velocity results are shown in Fig. 7. The points in the figure are trajectory points, of which the size and color reflect the magnitude of velocity. The darker color and the

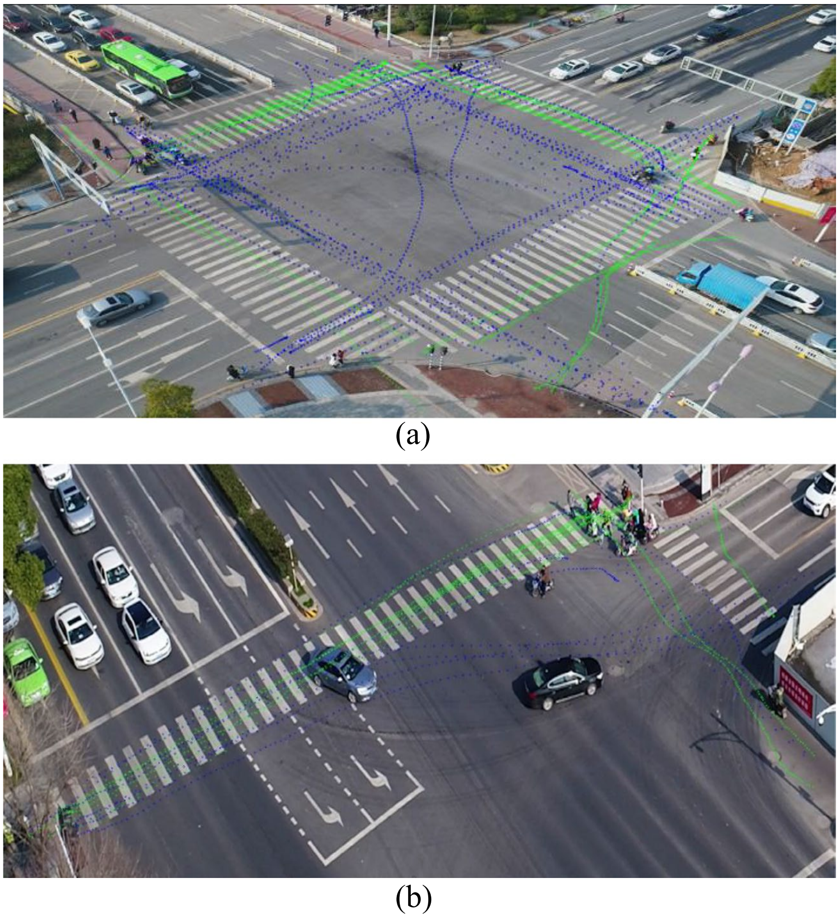


Fig. 6 Nonmotorized traffic trajectories distribution **a** Intersection 1, **b** Intersection 2

Table 1 Summary of flow results

		Video detection results			Ground truth		
Intersection	Direction	Nonmotorized	Cyclist	Pedestrian	Nonmotorized	Cyclist	Pedestrian
1	1 to 2	11	10	1	12	10	2
	1 to 3	0	0	0	0	0	0
	1 to 4	6	3	3	7	3	4
	2 to 1	3	0	3	3	0	3
	2 to 3	8	5	3	10	5	5
	2 to 4	1	1	0	1	1	0
	3 to 1	1	1	0	1	1	0
	3 to 2	1	1	0	2	2	0
	3 to 4	11	9	2	11	9	2
	4 to 1	13	8	5	13	8	5
	4 to 2	1	1	0	1	1	0
	4 to 3	5	2	3	5	2	3
	Overall	61	41	20	66	42	24
	Accuracy Rate	92.42%	97.62%	83.33%	—	—	—
2	1 to 2	16	2	14	21	1	20
	1 to 3	5	2	3	7	2	5
	2 to 1	9	3	6	7	2	5
	2 to 3	2	2	0	3	3	0
	3 to 1	5	4	1	5	4	1
	Overall	37	13	24	43	12	31
	Accuracy Rate	86.05%	108.33%	77.41%	—	—	—

larger size of the point is, the higher the instantaneous velocity at that point is.

3.2.3 Reliability Verification of Velocity

In order to verify the accuracy of velocity data, we compared the detected velocity with the velocity measured by GNSS-RTK (Global Navigation Satellite System Real-time Kinematic), whose error is within centimeter scale. Velocity was tested by a pedestrian walking separately at three different approach lanes. The comparison results are shown in Fig. 8. The orange line represents the velocity detected by video and the blue line represents the velocity measured by GNSS-RTK.

As shown in Fig. 8, due to the instability in establishing and releasing tracking state, the gap between the detected and the actual velocity is relatively large in the beginning and ending stages, so the velocity results in the first two seconds and the last two seconds were eliminated. The reliability verification results of velocity are shown in Table 3. Test mean velocity is the mean value of velocity detected from video, while validation mean velocity is measured by GNSS-RTK. Max difference is the maximum difference between the detected and validation instantaneous velocity, and the

Table 2 Summary of flow evaluation results

	Type	TP	FP	FN	Precision	Recall	CSI
Intersection 1	Pedestrian	20	0	4	100%	83.33%	83.33%
	Cyclist	39	2	1	95.12%	97.5%	92.86%
	Overall	59	2	5	96.72%	92.19%	89.39%
Intersection 2	Pedestrian	23	1	8	95.83%	74.19%	71.88%
	Cyclist	11	2	1	84.62%	91.67%	78.57%
	Overall	34	3	9	91.89%	79.07%	73.91%

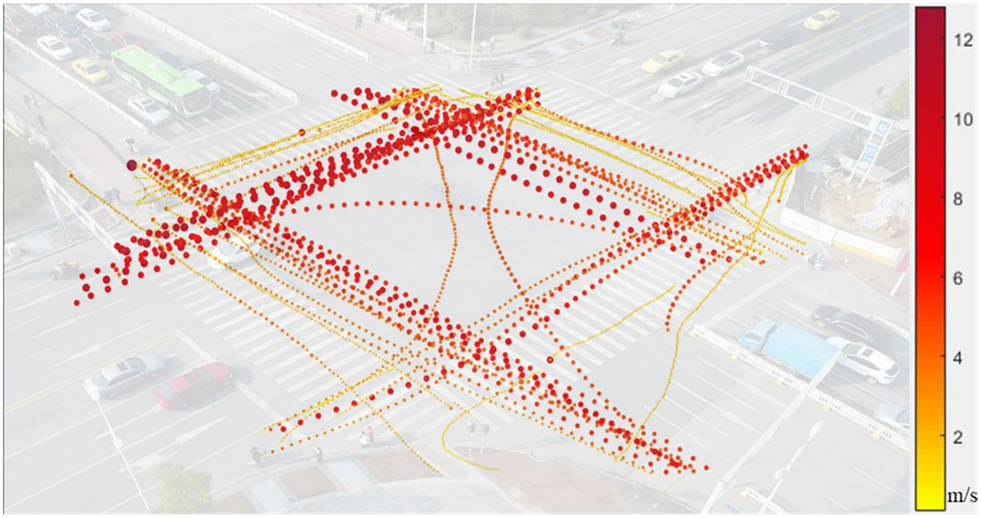


Fig. 7 Instantaneous velocity distribution.

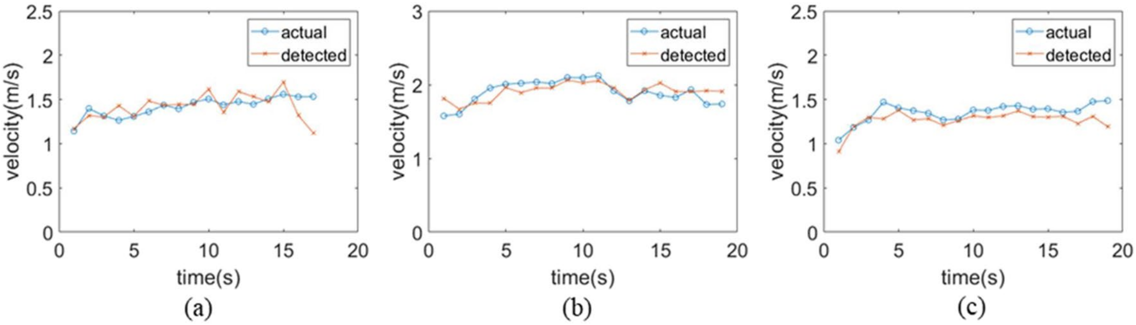


Fig. 8 Comparison results for velocity reliability verification.

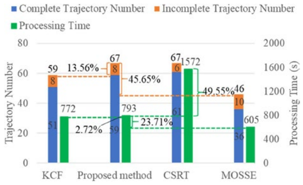
definition of min and average difference is the like. Error is the relative error of the detected and validation mean velocity. We can see from Table 3 that average difference of all approaches can be controlled around 7 to 8 centimeters per second and error is controlled within an acceptable range.

3.2.4 Tracking Efficiency Assessment

In order to assess the improvement of the tracking algorithm in this paper, we compared our method with KCF and other classic tracking algorithms, which includes Channel and Spatial Reliability Tracking (CSRT) [23], Minimum Output Sum of Squared Error filter (MOSSE) [24] The comparison results are shown in Fig. 9.

MOSSE, KCF and CSRT are three tracking algorithms of different computational levels. MOSSE is considered to be a relatively fast-tracking algorithm. Although it could reduce the processing time by 23.71% compared with our method, 21 trajectories couldn't be detected, which is of less confidence for the practical application. Compared with the single KCF algorithm, our method only took 2.72% more time to detect 13.56% more trajectories. With respect to CSRT, it could obtain as good results as our method, but it took twice as long, namely, more than 6 times as long as the actual video time to complete video processing. In consequence, our method could get the best results with the consideration of both efficiency and accuracy.

Fig. 9 Comparison results for tracking effect.



4 Discussion

Our method is suited for the videos taken at oblique angle. However, if the traffic flow is large and density is high, this shooting angle is prone to side-by-side and occlusion problems, affecting the accuracy of detection. Therefore, the accuracy of flow result at Intersection 2 dropped to 70%, compared with 90% at Intersection 1. In order to solve the problem, some attribute information such as the aspect ratio and height of tracking frame were adopted to judge the number of targets in tracking box. Logical judgment was added to determine whether separation behavior appears during movement. Although some measures have been taken to improve the system robustness as far as possible, shooting intersections from the top view can solve the problems better. However, this will result in a new issue, that when looking down straight at intersections, particularly a large-scale intersection, pedestrian almost appears as a point which is hard to be tracked. An efficient tracking algorithm for small targets needs to be found. Through the experiment, recommended camera angle is set to 50 to 60 degrees.

Weather and accuracy of the lenses are also factors affecting the system performance. Shadow detection mechanism makes the detection effect on sunny days almost consistent as on cloudy days, but it needs to be improved for night, rain, and snow conditions. Theoretically, the camera model should be calibrated due to lens distortions. However, as a robust system, it is impractical to calibrate the model for all the videos from different lens. The lens in our system is not calibrated, but velocity reliability verification results show that error is kept within acceptable limits.

As the proposed method combines KCF and KF for target tracking, it can meet the tracking requirements for nonmotorized traffic theoretically. Tracking effect depends partly on the effect of target detection. Although shadows are eliminated and relatively complete contours can be extracted by morphological methods, there are also a few incomplete or wrong detections. Some advanced detection algorithms may perform better, which are based on the deep learning. However, these methods have strict requirements on the sample and need to consider the impact of the change in shooting angle and height.

Table 3 Reliability verification results of velocity

Approach	Detected mean velocity (m/s)	Measured mean velocity (m/s)	Standard Deviation of detected velocity (m/s)	Standard Deviation of measured velocity (m/s)	Max difference(m/s)	Min difference(m/s)	Average difference(m/s)	Error (%)
Southbound	1.4699	1.4214	0.1160	0.0882	0.1673	0.0002	0.0722	3.41
Northbound	1.9279	1.9591	0.1004	0.1095	0.2054	0.0067	0.0723	1.59
Eastbound	1.2942	1.3676	0.0447	0.0587	0.1904	0.0176	0.0776	5.37

5 Conclusions

In this paper, we develop a lost target retracking mechanism for joint tracking method and propose a multivariate feature classification method, taking which as the core, a vision-based surveillance system for nonmotorized traffic at intersections is constructed. For tracking, since nonmotorized traffic tends to be small and blurred at large-scale intersections and tracking is easy to lose, we combine two different tracking strategies, using KCF as the major algorithm and KF to retrieve the lost targets. Consequently, long-term tracking is achieved and more complete trajectories can be found relying on the retracking mechanism. In terms of classification, we provide a precise classification method with low computational cost and limited training samples. Therefore, multivariate feature is associated to classifiers. Velocity is selected as motion feature plugging into K-means clustering for the first classification and HOG is used as the major shape feature for the further classification with SVM. Accordingly, the achievements and conclusions can be summarized:

(1) Although our method needs 2.72% more time than the single KCF, it detects 13.56% more trajectories. Furthermore, it can achieve as good results as more complex algorithm, CSRT, with only half the time cost.
(2) The accuracy of classification results reaches nearly 100%, which means that traffic parameters for pedestrians and cyclists can be obtained separately.
(3) Based on the trajectory, the system extracts flow and velocity data. The accuracy of flow results can reach nearly 90%. Meanwhile, the mean velocity error is within 6%. The average difference is around 7 to 8 centimeters per second compared with the actual value and can acquire all the data within nearly three times as long as the actual video time.

It is certain that the system works well in the application of collecting field data and our method provides a superior solution to monitor nonmotorized traffic and obtain their trajectories at intersections. However, some problems remain and need to be solved in the future. It is necessary to make the system more robust for practical application. In future studies, the system needs to include efficient tracking algorithm for small targets based on deep learning methods that provide users with a variety of options which can meet both accuracy and speed demand. The adaptability of the system to night, rain and snow condition also needs to be further developed.

Acknowledgments This study is supported by the National Key Research and Development Program of China (No. 2019YFB1600200), the national nature science foundation of China (No. 51878161) and the China Postdoctoral Science Foundation (No.2020M681466).

References

1. Transport Canada: Canadian Motor Vehicle Traffic Collision 2014. Publication Transport Canada (2016) Available online: https://tc.canada.ca/sites/default/files/migrated/cmvtcs2014_eng.pdf. Accessed 7 Jan 2021
2. U.S. Government Accountability Office: Pedestrians and cyclists—cities, states, and DOT are implementing actions to improve safety. Publication GAO-16-66. U.S. Government Accountability Office (2015) Available online: https://www.gao.gov/assets/680/673782.pdf. Accessed 7 Jan 2021
3. Baas, J., Galton, R., Biton, A.: FHWA bicycle-pedestrian count technology pilot project—summary report. FHWA, U.S. Department of Transportation (2016) Available online: https://www.fhwa.dot.gov/environment/bicycle_pedestrian/countpilot/summary_report/fhwahep17012.pdf. Accessed 7 Jan 2021
4. Shantaiya, S., Verma, K., Mehta, K.: Multiple object tracking using Kalman filter and optical flow. European Journal of Advances in Engineering and Technology. **2**(2), 34–39 (2015)
5. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. J. IEEE Trans. Pattern. Anal. Mach. Intell. **24**(5), 603–619 (2002)
6. Henriques, J.F., Caseiro, R., Martins, P.: High-Velocity Tracking with Kernelized Correlation Filters. J. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
7. Danelljan, M., Häger, G., Khan, F.S. (eds.): Discriminative Scale Space Tracking. J. IEEE Trans. Pattern Anal. Mach. Intell. **39**(8), 1561–1575 (2017)
8. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. 2017 IEEE International Conference on Image Processing, Beijing (2017)
9. Liu, W., Zhang, M., Luo, Z. (eds.): An Ensemble Deep Learning Method for Vehicle Type Classification on Visual Traffic Surveillance Sensors. IEEE Access. **5**, 24417–24425 (2017)
10. Gao, H., Cheng, B., Wang, J. (eds.): Object Classification Using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment. IEEE Trans. Ind. Inform. **14**(9), 4224–4231 (2018)
11. Khan, M.A., Ectors, W., Bellemans, T. (eds.): Unmanned Aerial Vehicle-Based Traffic Analysis: A Case Study for Shockwave Identification and Flow Parameters Estimation at Signalized Intersect. J. Remote Sens. **10**(3), 458 (2018)
12. Xin, J., Du, X., Shi, Y.: Classifier Adaptive Fusion: Deep Learning for Robust Outdoor Vehicle Visual Tracking. IEEE Access. **7**, 118519–118529 (2019)
13. Cui, Y., Zhang, J., He, Z. (eds.): Multiple pedestrian tracking by combining particle filter and network flow model. Neurocomputing. **351**, 217–227 (2019)
14. Yu, P., Zhao, Y., Zhang, J. (eds.): Pedestrian detection using multichannel visual feature fusion by learning deep quality model. J. Vis. Commun. Image Represent. **63**, 102579 (2019)
15. Wang, Z., Cui, J., Zha, H.e.: Foreground Object Detection by Motion-based Grouping of Object Parts. Int. J. Intell. Transp. Syst. Res. **12**, 70–82 (2014)
16. Zhao, Y., Chen, Q., Cao, W. (eds.): Deep Learning for Risk Detection and Trajectory Tracking at Construction Sites. J. IEEE Access. **7**, 30905–30912 (2019)
17. Li, J., Shao, C., Xu, W. (eds.): Real-Time System for Tracking and Classification of Pedestrians and Bicycles. Transport. Res. Rec. **2198**(1), 83–92 (2010)

18. Zhu, J., Chen, S., Tu, W. (eds.): Tracking and Simulating Pedestrian Movements at Intersections Using Unmanned Aerial Vehicles. Remote Sens. **11**(8), 925 (2019)
19. Guo, Y., Sayed, T., Zaki, M.H.: Automated analysis of pedestrian walking behaviour at a signalised intersection in China. IET Intell. Transp. Syst. **11**(1), 28–36 (2017)
20. Shirazi, M.S., Morris, B.T.: Vision-based pedestrian behavior analysis at intersections. J. Electron. Imaging. **25**(5), 051203 (2016)
21. Stauffer, C., Grimson, W.: Adaptive Background Mixture Models for Real-Time Tracking. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Ft. Collins (1999)
22. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego (2005)
23. Lukezic, A., Vojir, T., Zajc, L.C. (eds.): Discriminative Correlation Filter with Channel and Spatial Reliability. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu (2017)
24. Bolme, D.S., Beveridge, J.R., Draper, B.A. (eds.): Visual object tracking using adaptive correlation filters. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Providence, San Francisco (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Xiaohan Liu received the B.S. degree in transportation engineering from Southeast University, Nanjing, China, in 2019. He is currently a postgraduate with School of Transportation, Southeast University. His research interests include intelligent transportation systems and traffic signal control.

Hao Wang received the M.Eng. and Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2002 and 2008, respectively. He is currently a professor with the Jiangsu Key Laboratory of Urban Intelligent Transportation Systems at Southeast University. He is the vice director of Institute of Traffic Engineering at Southeast University and also serving as a member of National Transportation Modeling and Simulation Association. His research interests include intelligent transportation systems, traffic flow theory and traffic safety.

Changyin Dong received the M.Eng. and Ph.D. degree in transportation planning and management from Southeast University, Nanjing, China, in 2014 and 2020, respectively. He is currently a postdoctoral researcher with the Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University. His research interests include traffic flow theory, intelligent transportation systems and highway control.