Moving Object Classification Method Based on SOM and K-means

Jian Wu[1,2]
1.The Institute of Intelligent Information Processing and Application;
2.Provincial Key Laboratory for Computer Information Processing Technology,
Soochow University, Suzhou 215006, China
Email: szjianwu@163.com

Jie Xia[1,2], Jian-ming Chen[1,2], Zhi-ming Cui[1,2]
1.The Institute of Intelligent Information Processing and Application;
2.Provincial Key Laboratory for Computer Information Processing Technology,
Soochow University, Suzhou 215006, China

***Abstract*—We do research on moving object classification in traffic video. Our aim is to classify the moving objects into pedestrians, bicycles and vehicles. Due to the advantage of self-organizing feature map (SOM), an unsupervised learning algorithm, which is simple and self organization, and the common usage of K-means clustering method, this paper combines SOM with K-means to do classification of moving objects in traffic video, constructs a system including four parts, and proposes a method based on bidirectional comparison of centroid to do tracking, and an improved method to obtain initial background when using background subtraction method to detect motion of moving objects. Experimental results show the effectiveness and robustness of the proposed approach.**

***Index Terms*—motion detection, object tracking, object classification, SOM, K-means**

I. INTRODUCTION

Intelligent video surveillance is an important component of the social security system. Automatic classification of moving objects in video is an integral part of intelligent video surveillance. Video surveillance requires to classify moving objects as quickly and detailed as possible. With object type information known, more specific and accurate methods can be developed to recognize high level actions of video objects. Classification of moving objects into predefined categories allows the operator to program the monitoring system by specifying events of interest, such as 'alarming when a vehicle is running at too high speed' or 'alarming when a pedestrian is coming into a forbidden area', which is very common in smart video surveillance. And there are some common applications in traffic video surveillance including the identification of moving object, traffic counting, vehicle classification and identification and etc, which are all based on the classification of moving objects.

Much work has been done in the field of moving object classification in traffic scene videos. Viola and etc. [1] give us a good framework for feature selection and object class recognition using boosting. However, it is a tremendous work to collect large samples of training data in all kinds of conditions and label all of them manually. In [2], foreground objects are detected using motion information and certain image features, like area, compactness, speed and bounding box aspect ratio are extracted for training and classification. However, most of these are 2D features so cannot avoid projective distortion, which is much more significant in far-field traffic scene videos. For example, nearby objects in images appear to be larger and move faster than those far away. Therefore, simply using these features for classification is unsuitable and limits the accuracy rate of the results. Zhaoxiang Zhang and etc. [3] extract several features and adopt K-means clustering along with decision level fusion for automatic labeling and a subregion strategy of traffic video scene is used to solve the distortion phenomenon in traffic scene.

In this paper, we propose an approach to classify moving objects of traffic scene videos into three categories: pedestrian, bicycle and vehicle. A framework is designed to achieve automatic classification. Firstly, by improved background subtraction method, moving objects are detected and a small number of 2D features are extracted; secondly, with bidirectional comparison of centroid which is proposed in this paper, motion information of moving objects are extracted; then by using SOM neural network, we cluster these feature vectors into three categories: pedestrian, bicycle and vehicle, and obtain their respective cluster centers. At last, after neural network clustering, we respectively calculate the distances between the new feature vector and the three centers to determine which category the moving object belong to.

The remainder of the paper is organized as follows. In Section 2, we introduce the method which we improved for motion detection. The tracking of moving objects is described in Section 3. Classification framework is described in Section 4. Experimental results and analysis are presented in Section 5. Finally, we draw our conclusions in Section 6.

doi:10.4304/jcp.6.8.1654-1661

II. MOVING OBJECT DETECTION

Moving object detection is a prerequisite of moving objects tracking, but also the basic issues in computer vision, using motion detection can obtain many kinds of information of foreground objects. The commonly used method of motion detection [4] is background subtraction and frame difference method. Background subtraction obtains foreground by subtracting the background model from the current video frame. The key lies on how to obtain the original static background model; and because the background often dynamically changes with the light, movement and objects in and out the scene, so it needs an efficient strategy to maintain and update the background model. The commonly used strategy is median filter method, mixed-Gaussian method and so on, but they all have some disadvantage of such as imprecise initial background, maintenance and updating difficulties of background model.

This paper first obtain initial background by using the approved double difference method, then use background subtraction method to extract moving objects.

A. Double Difference Method

In this paper, we adopt an improved frame-difference method: double difference method, also known as three-difference method [5] to obtain initial background, which can be summarized as follows:

Difference image is obtained by absolute value between corresponding pixels of two sequential frames. Suppose $\{F_n\}$ is input image frames, so the nth difference image is defined as:

$$D_n(i,j)=\left|F_{n-1}(i,j)-F_n(i,j)\right| \tag{1}$$

Where $F_{n-1}(i,j)$ denotes the pixel value at (i,j) in the $(n-1)$ th frame, $F_n(i,j)$ denotes pixel value at (i,j) in the nth frame, so the n th difference $D_n(i,j)$ is the absolute value between pixel values at (i,j) in the $(n-1)$ th frame and n th frame.

In the definition of double-difference image, a flag will be used, which is denoted as follow:

$$flag=\left(D_{n+1}(i,j)>T_0\right)\wedge\left(D_n(i,j)>T_0\right) \tag{2}$$

Where T_0 is a given threshold. $D_{n+1}(i,j)>T_0$ denotes that difference value of the $(n+1)$ th frame is larger than the threshold T_0. $D_n(i,j)>T_0$ denotes that difference value of the n th frame is larger than threshold T_0. Therefore, when both difference values at coordinate (i,j) in two consecutive difference images are larger than a given threshold T_0, the value of $flag$ is equal to 1.

Thus, the double-difference image DD_n is defined as follow:

$$DD_n(i,j)=\begin{cases}255, flag=1\\ 0, others\end{cases} \tag{3}$$

Where if $flag=1$, $DD_n(i,j)=1$ denotes that both difference value images has detected pixels whose difference is larger than threshold T_0, and these points compose of the moving target to be tracked. That is, double-difference image is obtained by logical-and between two consecutive difference image. Because noise is not easy to repeat in two consecutive difference images, double-difference method can suppress noise better than general frame-difference method.

B. Motion Detection

Based on the double difference method mentioned above, we propose an approach to obtain the initial background. $\{F_{i,j}^t\}$ represent the gray image of the t th frame, $F_{i,j}^t$ represents the pixel value at (i,j). $\{\mathrm{B}_{i,j}^t\}$, $\{O_{i,j}^t\}$ represent the background and foreground image of the t th frame respectively.

(1)Obtain $\{D_{i,j}^x\}$ by subtracting $(t-1)$ th frame from t th frame and obtain $\{D_{i,j}^y\}$ by subtracting t th frame from $t+1$ th frame with (1)

(2)Obtain $\{O_{i,j}^t\}$, $O_{i,j}^t=\begin{cases}255, if\left(D_{i,j}^x>T_0\right) and \left(D_{i,j}^y>T_0\right)\\ 0, other\end{cases}$

(3)Eliminate all the pixels whose $O_{i,j}^t$ is equal to 255 from the t th frame and make the reminder part as the background image of the t th frame.

(4)Repeat above operations on the first N frames, so obtain the corresponding $N-2$ background images. Calculate the statistical character pixel by pixel, find out the pixel value corresponding to the most frequent appearances, set it as the corresponding pixel of the background to construct the whole initial background B_0.In this paper, N is set as 80, T_0 is 120.

We obtain the initial background, but the background image is changing by time, so the background updating is defined as:

$$B_i(x,y)=(1-\alpha)B_{i-1}(x,y)+\alpha F_i(x,y) \tag{4}$$

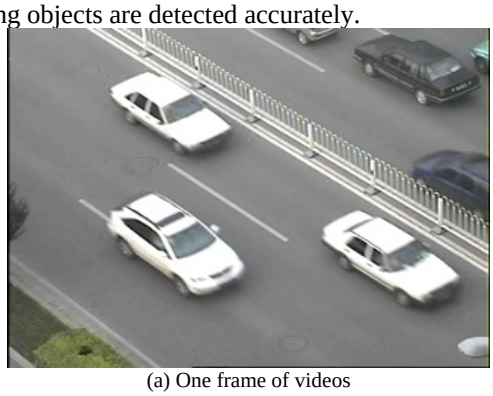
Where α is the coefficient of background updating. So in fact, the updating of background is obtained by the weighted average of video sequence, which is the process of eliminating noise.

Foreground object is obtained by background subtraction which is defined as follow:

$$Fore_i(x,y)=\left|F_i(x,y)-B_i(x,y)\right| \tag{5}$$

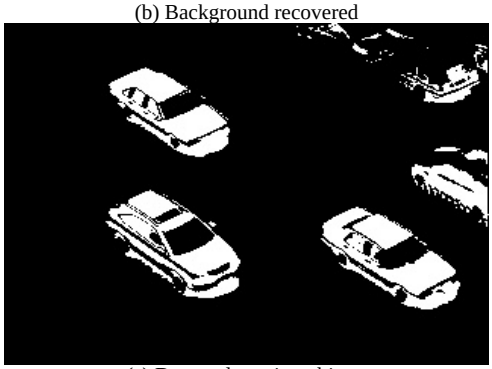
Where foreground object $Fore_i(x,y)$ is obtained by the difference between the i th frame $F_i(x,y)$ and the i th background image $B_i(x,y)$.

Experimental results of background modeling and motion detection are shown in Fig.1. As we can see, moving objects are detected accurately.



(a) One frame of videos

(b) Background recovered



(c) Detected moving objects

Figure 1. Motion detection results

(a) Original frame

(b) Motion detection by simple back ground subtraction

(c)Motion detection by our method

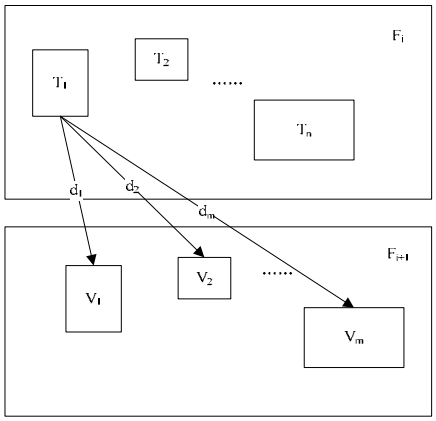
Figure 2. Detection results comparison

In Fig.2(b), We can find the ‘ghost’ of another vehicle appeared in last several frame, by simple background subtraction method, ‘ghost’ can’t be eliminate in time. But with the method proposed in this paper, perfect background is obtained and updated by time, so we can get perfect motion detection results.

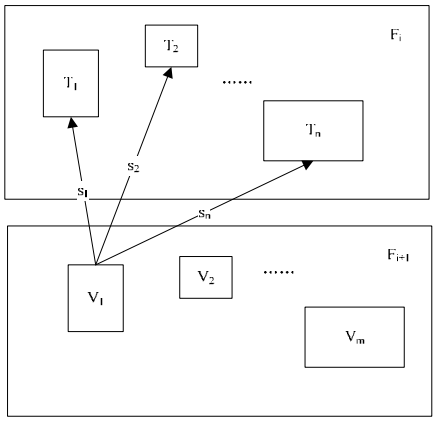
III. MOVING OBJECT TRACKING

The purpose of moving object tracking is to analyze video sequences to calculate the object location on each frame, and estimate object speed. Two commonly used indexes to evaluate tracking algorithm is reliability and real-time. Commonly used tracking methods [6] can be divided into two broad categories: the relevant tracking and optical flow tracking. The relevant tracking doesn’t require high quality of image scene, that is, it can work in low SNR conditions, but it requires very large computational complexity. The optical flow tracking method not only uses the motion information of object to avoid the impact of gray change, but also has a good anti-noise capability; and it only need to process very small number of feature points in image, thus computational complexity is quite small. But it is difficult to extract the precise shape of moving object, so it can not solve the feature matching problem. In this paper, tracking moving object is mainly in order to extract motion features of objects for object classification.

This paper proposes a bidirectional comparison method, which is shown in Fig.3:



(a) Forward comparison of centroid



(b) Reverse direction comparison of centroid

Figure 3. Bidirectional comparison of centroid for tracking

Where F_i is the i th frame, in which moving objects $T_1, T_2, \cdots, T_n$ are detected. F_{i+1} indicates the $(i+1)$th frame, in which moving objects $V_1, V_2, \cdots, V_m$ are detected.

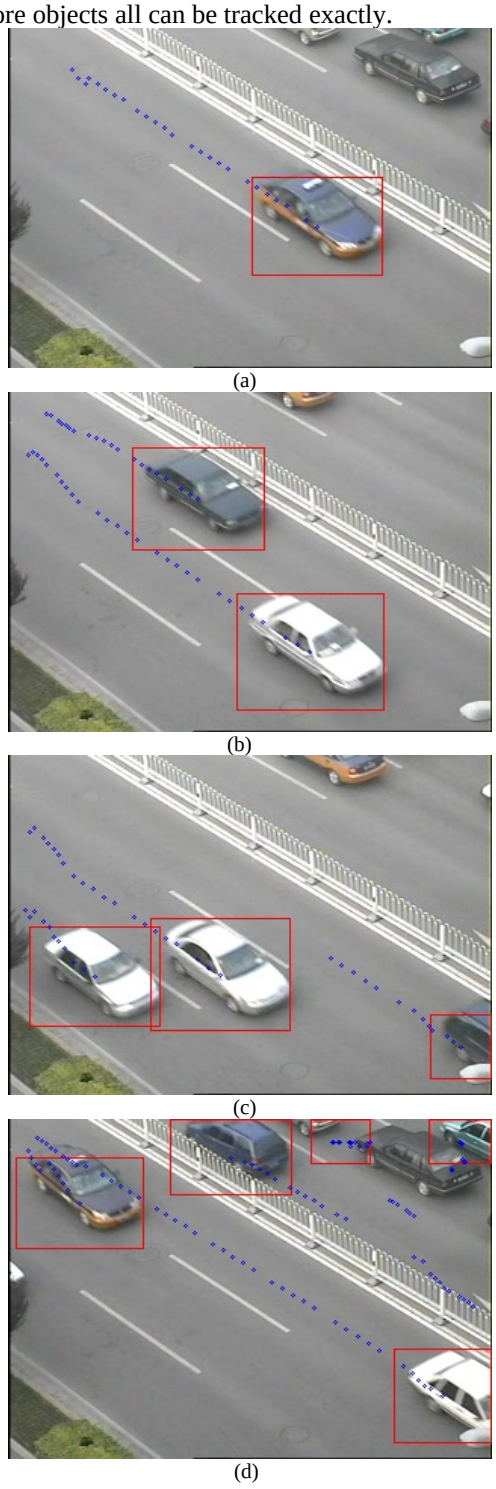
The bidirectional operation can be described as follows:

(1)Fig.3(a) is the forward comparison of centroid. Aim at every object T_i in F_i, we obtain a group of centroid distances $d_1, d_2, \cdots, d_m$ by calculating the centroid distance between T_i and every object $V_1, V_2, \cdots, V_m$ in F_{i+1}.And we will find the object V_i which is corresponding to the smallest distance, but we don't think the matching procedure is over but go ahead to step (b).

(2)Fig.3(b) represents the reverse direction comparison of centroid. Aiming at the object V_i which is found in step (a), we calculate the centroid distance between V_i and every object $T_1, T_2, \cdots, T_n$ in F_i to obtain a group of centroid distances $d_1, d_2, \cdots, d_m$. Then we find the object T_x corresponding to the smallest distance, only when T_x is T_i, we think the objects in pre and post frame match successfully.

Here we complete object tracking in two frames. Perform orderly, every moving object can be tracked as well as their motion information will be saved.

The results of this tracking method are shown in Fig.4. As we can see, in traffic scene, single moving object and even more objects all can be tracked exactly.



(a) (b) (c) (d)

Figure 4. Tracking results with bidirectional comparison of centroid

IV. CLASSIFICATION

A framework for classification of moving objects in traffic scenes is described in detail in this section.

A. System Diagram

The entire system consists of four components: object detection and tracking, feature extraction, network training, and object classification. In which, moving object extraction adopts an improved background substraction method which is mentioned in section 1 to ensure the rapid object extraction as well as the accuracy of object extraction. This paper proposes an improved centroid-comparison method to realize moving objects tracking. In the whole procedure of detection and tracking, extract the shape features and motion features of object, and save them as feature vector. Then put these samples into network training component to train by SOM neural network, and obtain the identified cluster centers. The object classification component will respectively calculate the distances between the new extracted feature vector and the three cluster centers, and thus output the result. The flowchart of the approach is shown in Fig.5.

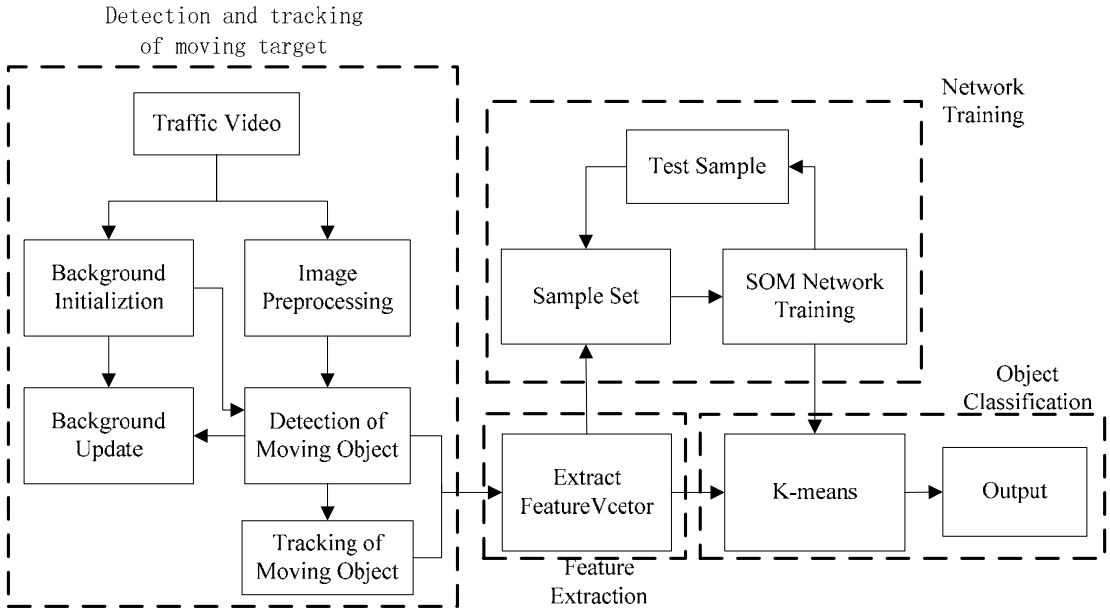


Figure 5. Flowchart of SOM-based classification of traffic moving object

B. SOM Neural Network

SOM (Self-orgnizing feature Map, SOM) is an unsupervised learning network [7] [8], which can change the network parameters and structure self-organizationly, adaptively by automatic finding the internal laws and properties of samples.

A typical topology is shown in Fig.6, which consists of input layer and competitive layer (output layer). In which, the input layer is single layer and one-dimensional neurons, while output layer are two-dimensional neurons, and neural inhibition exists between the neighborhood. Network is fully connected, that is, each input node is connected with all the output nodes.

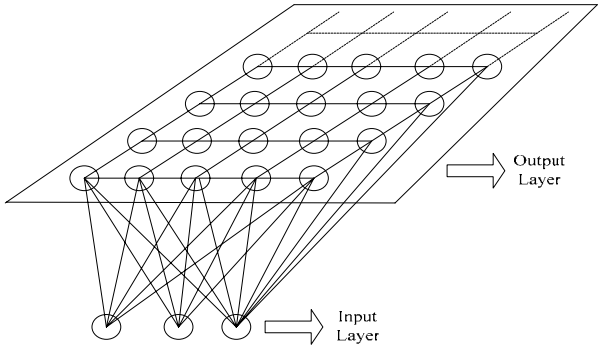


Figure 6. Network Topology Diagram of SOM

The SOM network in this paper adopts Kohonen weight adjustment rule [9], that is, not only adjust the weight of win neurons, but also relatively adjust the weight vectors of surrounding neurons. We use the Mexican hat function: give the winner greatest weight, and give smaller weight to neighboring nodes of winner; the distance away from the winner is larger, the value of adjust weight is smaller, until reach a certain distance, the adjust weight is zero; when the distance away from the winner is larger than the certain distance, the value of adjust weight is a negative; then if go on farther, it's back to zero. Mexican hat function has a very good localization property in the time domain and frequency domain, as shown in Fig.7:

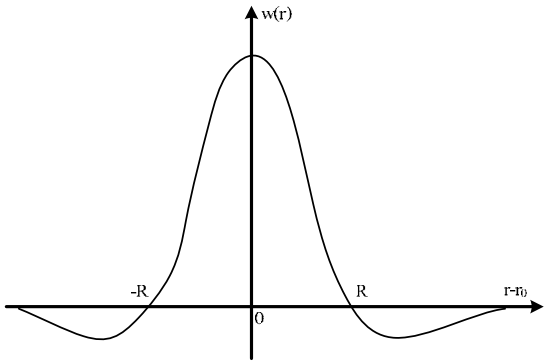


Figure 7. Mexican Hat Function

In this paper, SOM clustering function is used to cluster the extracted features of moving objects in traffic into three categories: vehicles, pedestrians and bicycles. This function is realized mainly through the two following steps:

(1) Aim at any input vector $X=[x_1,x_2,\cdots,x_n]$ supplied to network, the fan-in weight W_j of neuron j on output layer is represented as: $W_j=[w_{j1},w_{j2},\cdots,w_{jn}], j=1,2,\cdots,n$, through Euclidean distance method to determine the corresponding winner neuron on output layer q,

$$q=\arg\min\left\{\left\|X-W_i\right\|\right\}, i=1,2,\cdots,n \tag{6}$$

where the winner q is determined by the smallest Euclidean distance between X and vector W_j.

(2) Define a neighborhood range of winner neuron q. As the training is repeating, the neighborhood range is reduced gradually, so we don't adjust the connect weight vector of neuron which is out of the neighborhood range, but to adjust weight vector of neuron which is in the neighborhood range according to the following equation:

$$w_{ij}(t+1)=w_{ij}(t)+\alpha(t,N)\left[x_i^p-w_{ij}(t)\right] \tag{7}$$

This adjust procedure make the fan-in weight vector in the scope of $N_q(t)$ move to input vector. Through training, the winner in competitive layer and its fan-in weight vector of neuron in the scope of $N_q(t)$ neighborhood approaches to input vector. As the neighborhood is gradually reduced, all the weight vectors will be separated from each other to achieve clustering.

C. K-means

K-means algorithm [10] is a typical division method in clustering methods, whose basic idea is to divide the objective data into different clusters by iterating, so that similarity between objects inside one cluster is large while the similarity between clusters is smaller. The specific process can be described as follows:

(1) Select k objects as initial cluster centers;

(2) Reassign each object to the most similar cluster according to cluster centers;

(3) Recalculate the average value of each cluster, and use these average values as new cluster centers;

(4) Repeat steps (2) (3), until there is no change of cluster centers.

The choice of initial cluster centers has an intensive impact on the clustering results of K-means algorithm. If choose improper initial cluster centers, clustering results may fall into local optimum solution.

So we combine SOM and K-means algorithm. The method can be described as follow:

(1) Send the feature vectors into SOM to train, implement SOM algorithm, and obtain three cluster centers. At this stage, it is not needed complete convergence of SOM, so it can reduce time comsumption.

(2) Initialize the cluster centers of K-means with SOM clustering results, implement K-means clustering algorithm.

This combination algorithm of SOM networks and K-means algorithm not only maintain the self-organization characteristics of SOM but also absorb the high efficiency of K-means algorithm, and as well compensate the disadvantage of the too long converging time of SOM and the deficiency of bad clustering results caused by improper initial cluster centers when use K-means algorithm.

D. Feature Extraction of Moving Object

The ultimate purpose of video surveillance system is to analyze and judge objects in scene, therefore, accurate classification of moving objects in scene is needed necessarily. Video-based moving object classification methods are mainly based on shape information, motion features [10]. In order to classify moving objects into three categories, we adopt SOM neural network and K-means clustering. And considering that the moving direction of objects in traffic scene is uncertain, this paper combines shape features with motion information and there are totally five shape and motion features are used in our algorithm:

(1) area: size of objects in pixels.

(2) velocity: time derivative of centroid of the object.

(3) ratio: $\frac{length}{width}$.

(4) area': time derivative of area.

(5) dispersion: equals to $\frac{perimeter^2}{area}$.

As we all know, area and velocity have the most significant projective distortion among all the five features. So, we use the additional three features for SOM clustering and automatic labeling. After videos processed frame by frame for a period of time, SOM clustering is adopted to establish three clusters and obtain the three cluster centers. One cluster corresponds to one category, respectively.

We choose these features based on the following intuitive rules:

(1) area has the advantages of distinguishing pedestrians from vehicles and bicycles.

(2) velocity has the advantages of distinguishing vehicles from pedestrians and bicycles.

(3) ratio has the advantages of distinguishing vehicles from pedestrians and bicycles.

(4) area' has the advantages of distinguishing pedestrians from vehicles and bicycles.

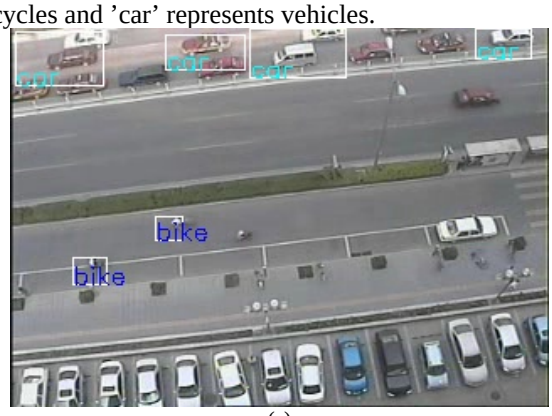
(5) dispersion has the advantages of distinguishing vehicles from pedestrians and bicycles.

As a result, we obtain five-dimensional feature vector which is composed of area, velocity, ratio, area' and dispersion. We send them into SOM neuron network, clustering them into three categories, and label them as pedestrian, bicycle, vehicle respectively, and save the three cluster centers; in the following videos, we extract feature vectors of new moving objects, and calculate the distance between the vector and the three cluster centers, and choose the category corresponding to the smallest distance as the category which the object is belong to.

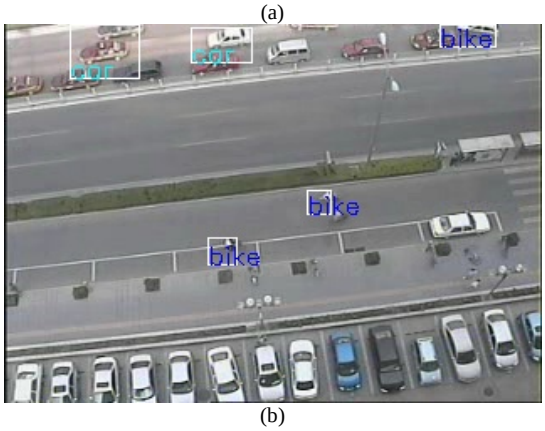
V. Experimental Results and Analysis

Experiments are conducted to demonstrate the performance of the proposed algorithm. All the experiments are carried out on computers of AMD 1.9 GHz CPU and 1G DDR, in Microsoft Visual C++ 6.0 programming environment, based on Intel's OpenCV library.

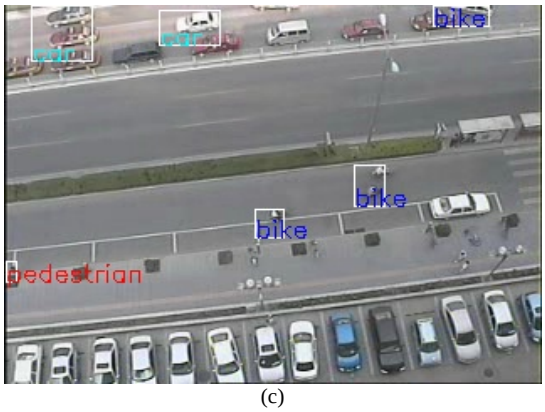
Several video scenes have been used in this experiment and two scenes are shown in Fig.8 and Fig.9 respectively. And classification results are given. Here, 'pedestrian' represents pedestrians, 'bike' represents bicycles and 'car' represents vehicles.



(a)



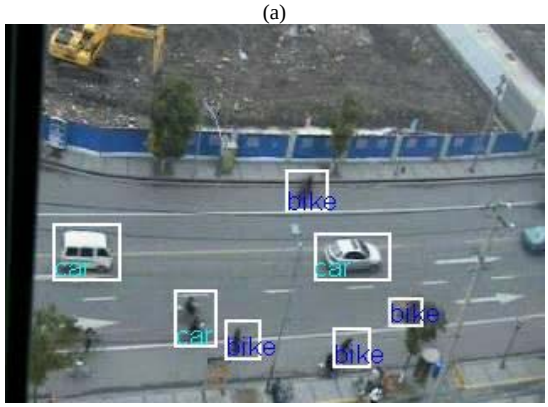
(b)



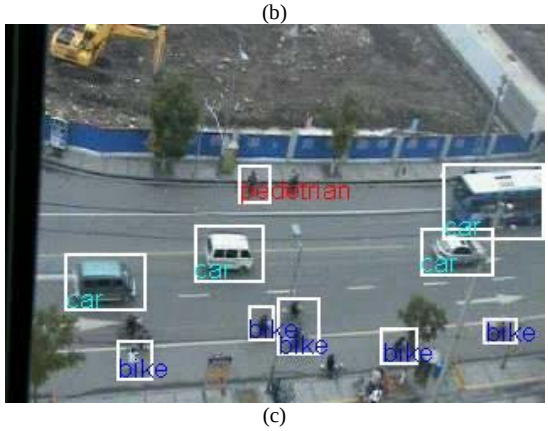
(c)

Figure 8. The results of classification in one scene

(a)



(b)



(c)

Figure 9. The results of classification in another scene

From above experimental results we can find that the moving objects classification can be basically realized. Especially vehicles can be recognized perfectly. The difficulty appears at distinguishing bicycle from

pedestrian because of their similar size, but due to the usage of velocity and dispersion features, we can distinguish objects quite good.

The defect appeared in these experimental results is that sometimes, two(or more) moving objects with similar velocity approach to each other and even intersect with each other, they will be recognized as single object, thus an unstable result happened, as shown in Fig.8(a) and Fig.9(b). In Fig.8(a), the 'car' in the top left corner in fact is two vehicles. While in Fig.9(b), the 'car' in the middle position actually is two bicycles.

In short, we simply realize the classification by combination of SOM and K-means algorithm. The results of the first video using are shown in Table 1.

Table 1 Classification accuracy using distance from cluster

	Pedestrians	Bicycles	Vehicles
Pedestrians	96.3%	3.7%	0.0%
Bicycles	3.7%	94.6%	1.7%
Vehicles	0.0%	1.7%	98.3%

From all of the above, we can see our method is effective.

VI. CONCLUSION

This paper proposes an approach for classification of moving objects in traffic scenes. We propose a motion detection method combining double difference method and background subtraction method, propose a tracking method based on bidirectional comparison of centroid, extract feature vectors of moving objects in the above process, and send them to SOM neural net work to get cluster centers, and finally use K-means approach to do classification. Using a novel classification framework, the experimental results show that the method is simple and effective, which can be applied to many systems. Solve the wrong classification caused by objects occlusion and overlap will be the future research direction.

ACKNOWLEDGEMENT

This research was partially supported by the Natural Science Foundation of China under grant No. 60970015, the 2008 Jiangsu Key Project of science support and self-innovation under grant No. BE2008044, the 2009 Special Guiding Fund Project of Jiangsu Modern Service Industry (Software Industry) under grant No. [2009]332-64, the Project of Jiangsu Key Laboratory for Computer Information Processing Technology grant No.KJS0924, the Applied Basic Research Project (Industry) of Suzhou City under grant No. SYJG0927 and the Beforehand Research Foundation of Soochow University.

REFERENCES

[1]. Paul A. Viola, Michael J. Jones, and Daniel Snow, "Detecting pedestrians using patterns of motion and appearance," in Proceedings of 9th IEEE International Conference of Computer Vision, 2003.

[2]. Lisa M Brown, "View independent vehicle/person classification," in Proc. of the ACM 2nd international workshop on Video Surveillance and Sensor Networks, 2004.

[3]. Zhaoxiang Zhang, Yinghao Cai, Kaiqi Huang and Tieniu Tan. Real-time moving object classification with automatic scene division.ICIP2007

[4]. DAI Ke-xue， LI Guo-hui， TU Dan， YUAN Jian. "Prospects and Current Studies on Background Subtraction Techniques for Moving 0bjects Detection from Surveillance Video". Journal of Image and Graphics. 2006, Vol.11, pp.919–927

[5]. Collins R, Lipton A J, Kanade T, et a1. "A system for video surveillance and monitoring: VSAM final report". Technical Report: CMU-RI-TR-00,Carnegie Melon University, Pittsburgh, Peen, America, 2000

[6]. YU Jing, YOU Zhi-sheng. Survey of Automatic Target Recognition and Tracking Method[J]. Application Research of Computers. 2005, 1:12~15

[7]. Kohonen T.Self-organizing Maps, 24dedition. Bedin: Springer. 1997

[8]. YA IG Zhanhua . YANFG Yan. Research and Development of Self-organizing Maps Algorithm[J]. Computer Engiineering. 2006, 32(16): 201~202-228

[9]. Ventura Dan, Tony Martinez. Quantum Associative Memory with Exponential Capacity. Proceedings of the International Joint Conference on Neural Networks, Anchorage, Alaska, 1998, pp.509~513

[10]. Richard O.Duda, Peter E.Hart, David G.Stock. Pattern Classification[M].2007.1:423-425

Jian Wu was born in Nantong on the 29th April, 1979, and got master degree in the field of computer application technology from Soochow university, Suzhou city, China in 2004. The main research direction is computer vision, image processing and pattern recognition. He works as a teacher in the same college after his master graduation. Now he is pursuing the doctoral degree. He was awarded the Third Prize of 2007 Suzhou City Science and Technology Progress and the 2008-2009 Soochow University Graduate Scholarship Model.

Jie Xia was born in Hefei on the 20th January, 1986, and got bachelor's degree in the field of computer application technology from Anhui Normal university, Wuhu city, China in 2007. And now she is studying in Soochow university to pursue the master degree. Her main research direction is image processing and video retrieving.

Jian-ming Chen was born in Suzhou on February, 1960. Associate professor, Master Supervisor. The main research direction is intelligent information processing and software engineering.

Zhi-ming Cui was born in Shanghai on the 4th July, 1961. Professor, PhD Candidate Supervisor. The main research direction is deep web and video mining.