Research on Recognition and Classification of Moving Objects in Mixed Traffic Based on Video Detection

Hongyu Hu (corresponding author)
College of Transportation, Jilin University, 130022, Changchun, P.R.China,
Tel: 86-431-85095857, e-mail: dayuhoo@gmail.com

Zhaowei Qu
College of Transportation, Jilin University, 130022, Changchun, P.R.China,
Tel: 86-431-85095857, e-mail: quzw@email.jlu.edu.cn

Zhihui Li
College of Transportation, Jilin University, 130022, Changchun, P.R.China,
Tel: 86-431-85095857, e-mail: lizhihui@email.jlu.edu.cn

Dianhai Wang
College of Transportation, Jilin University, 130022, Changchun, P.R.China,
Tel: 86-431-85095857, e-mail: wangdianhai@sohu.com

Total Words: 5576 (with 8 figures and 1 table)

ABSTRACT

At present, mixed traffic is the major property of the traffic in some developing countries, mixed traffic parameters can be obtained by video detection which is utilized more and more for the traffic information collection. In this paper, recognition and classification of moving objects is developed. In order to obtain moving objects from the video sequence efficiently, this paper presents a background initialization algorithm based on clustering classifier, all stable non-overlapping intervals in the training sequence are located for each pixel as possible backgrounds by slip window at first; then the background interval is obtained from the classified data set by unsupervised clustering. According to mixed traffic, a simple effective feature representation algorithm is proposed, the distance between the point of object's silhouette and the geometric center is defined as "centro-distance", all of centro-distances along the silhouette make up of a vector, and moving objects could be recognized with the vector of centro-distance. Minimum-distance method is used to classify moving objects in mixed traffic into three categories: vehicles, bikes and persons. Temporal consistency is combined with to further improve the accuracy of the classification. Experimental results show that the proposed method can be applied well in real-world.

KEY WORDS

Mixed Traffic, Video Detection, Recognition and Classification, Feature Representation, Minimum-Distance Method

INTRODUCTION

The motion status of moving objects can be obtained by video detection effectively in real-time, and traffic flow parameters can be utilized to control and manage the urban traffic. There have been not any appropriate video detectors applied for mixed traffic flow which includes vehicles, bicycles and persons by now. Therefore, the mixed traffic detection algorithm has become the urgent demand for traffic information collection.

Recognition and classification between vehicles and non-vehicles is a crucial part of video detection for mixed traffic. How to find the appropriate feature to represent objects is connected closely with recognition and classification. A better feature should enlarge the similarity of different category objects and reduce the similarity of same category objects. There have been lots of researches on the video detection system in the past, but these researches had limitations respectively. Collins (*1*) categorized moving objects with two steps. At first, the dispersedness and the area were taken as classification features and objects were classified by a three-layer neural network to be three categories: vehicles, humans and human groups; Secondly, the shape and color moment features and the linear discriminant analysis was used to divide vehicles into five categories: cars, vans, pickups, trucks and buses. Lapchev and Bogomolov (*2*) divided the moving objects into vehicles, persons, crowds, and others combined with motion status and shape information. Kuno and Watanabe (*3*) used silhouette parameters of the shape model to detect moving objects in video sequences. Toth and Aach (*4*) represented object shapes with Fourier description, and a feed-forward neural network was used to determine the object classes among human, vehicle and background clutters. Methods mentioned above were not adaptive enough to apply, because they could not get over the variability of the rotation and the scale. Objects may be classified by error in different traffic scenarios. There were also some methods relying on temporal motion characteristics for classification. These methods were usually suitable for the discrimination between rigid objects (vehicles) and non-rigid objects (persons). Javed and Shah (*5*) proposed feature representation based on “Recurrent Motion Image” to calculate repeated motion of objects and categorized the image blobs into three object classes: humans, vehicles, and human groups. Cutler and Davis (*6*) described objects’ self-similarities by tracking objects, applied Time-Frequency analysis to detect and characterize the periodic motion. Lipton (*7*) presented a reliable flow technique called dynamic region matching to allow flow-based analyses of moving entities in real-time. Generally， methods of detecting periodic motion based on temporal characteristic need two assumptions: firstly, every foreground object should be located in a fixed rectangle, the variation of the shape should be little; secondly, the frame rate should large enough to obtain periodic motion information, and the background should not vary obviously. These two suppositions limited algorithms’ application.

At present, mixed traffic still plays a leading role in some developing countries, thus the control and management of mixed traffic is not able to neglect, and the classification between vehicle and non-vehicle is a crucial part. In view of this, this paper researches on recognition and classification algorithm of moving objects in mixed traffic by video detection. A background initialization algorithm based on clustering classifier to obtain moving objects from the video sequence efficiently; according to mixed traffic, a simple effective feature representation algorithm which is fixed on rotation, translation and scale is proposed; and then minimum-distance method is used to classify moving objects in mixed traffic. The system includes three basic parts: background elimination, objects feature representation, objects recognition and classification. Fig.1 is the technical framework of the paper.

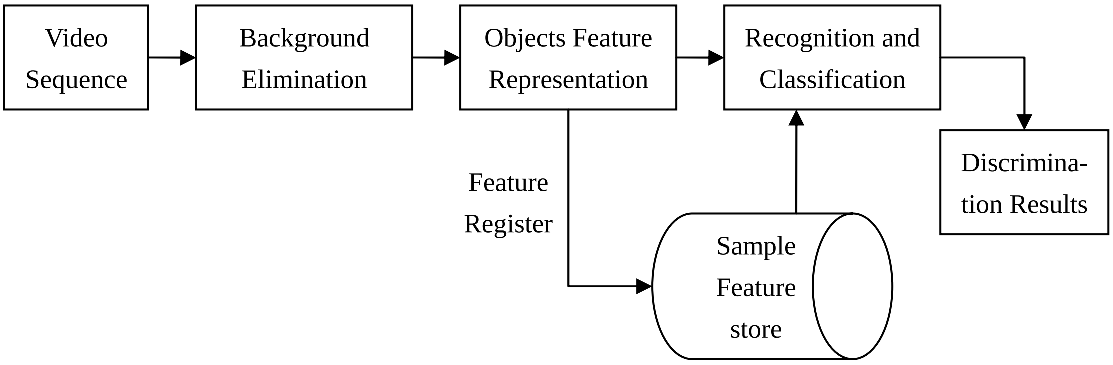

FIGURE1 The technical framework of recognition and classification

The paper structure is arranged as follows: Section 2 discusses the preprocess of video detection, background elimination algorithm is presented to extract foreground objects; moving objects are identified by feature representation in Section 3. Section 4 is focused on the approach to classify objects into different categories: vehicles, bicycles and persons. Section 5 illustrates the experimental evaluations of the algorithm. And finally, the paper is concluded in Section 6.

BACKGROUND ELIMINATION

Eliminating background, extracting moving objects effectively and adaptively is the basic of the object recognition and classification system. At present, background subtraction algorithm (*8*, *9*) is used popularly and efficiently. In the video sequence, the background is more stable to detect than foreground. Adaptive background models (Gaussian Model (*10*), Mixture Gaussian Model (*11*)) were usually used to obtain background, these methods were robust for the impact of the light changing or trees rocking in the dynamic environment, but when massive continual moving objects (for example: rush hour) or long-term stopping of the vehicle (for example: traffic jam) in the detection region, the methods were invalid. In order to detect moving targets effectively, a background initialization algorithm based on clustering classifier is proposed in this paper.

Generally, if the pixel is the background in the image, its value keeps long-term stable, and it would be changed when foreground objects passing. A set of values would be observed in the temporal training sequence of every pixel, and some stable non-overlapping intervals in the sequence could be located for each pixel as the possible background. Fig.2 is values at one pixel in a temporal sequence of 100 frames.

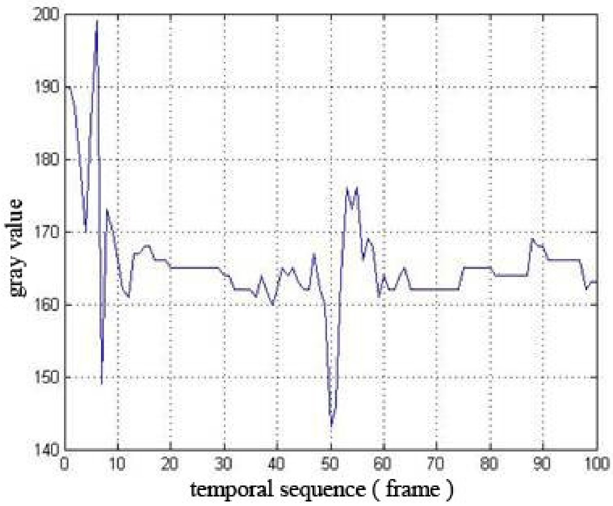

FIGURE2 Values over time at one pixel of a sequence

Because stable intervals may be made up of three parts: real backgrounds, still objects and big slow objects, a algorithm is utilized to eliminate intervals which include still objects and big slow objects, and it should satisfy assumptions followed:

(1) There is a period in the temporal training sequence not covered by the foreground object at least.

(2) The background interval is stable in the temporal training sequence.

(3) The amount of background intervals is more than others in the stable interval set.

The initialization background algorithm consists of two steps: the first step is to obtain all stable non-overlapping intervals in the temporal training sequence for each pixel as possible background; the second step is to get the background sub-set from the classified data set by unsupervised clustering to realize the background initialization.

Stable intervals Detection

Let $\{x_i \mid \mathrm{i}=1\cdots \mathrm{N}\}$ denote N observed values of every pixel. In order to obtain all stable non-overlapping intervals, a slip window with the initial length is defined in the temporal sequence. Continuous observed values are located at first, the number of which is equal to the initial window length. If observed values alter in a permissible range, the next observed value is pulled into the window, and the length of the window increase by 1; if the variation of observed values is out of the permissible range and the length of the window is larger than the initial length, observed values in the window is marked as a stable interval, and then renew the slip window beginning with the value after marked values of the stable interval. If the variation of observed values is out of the permissible range and the length of the window is not larger than the initial length, the length doesn't alter, and the window moves one value. A set of stable intervals can be obtained when the window slips all observed values of the pixel. The intervals detected are showed in Fig.3. Let $L=<l_1,\cdots,l_k>$ represent the stable interval set, and the interval $\{x_i,\cdots,x_j\}$ in it should satisfy Eq.1:

$$\begin{gathered}\omega \le j-i, \\ \forall(s,t) \mid x_s - x_t \mid \le \delta_{\max}\end{gathered} \quad (1)$$

Where ω is the initial length, $\delta_{\max}$ is the largest variation permitted.

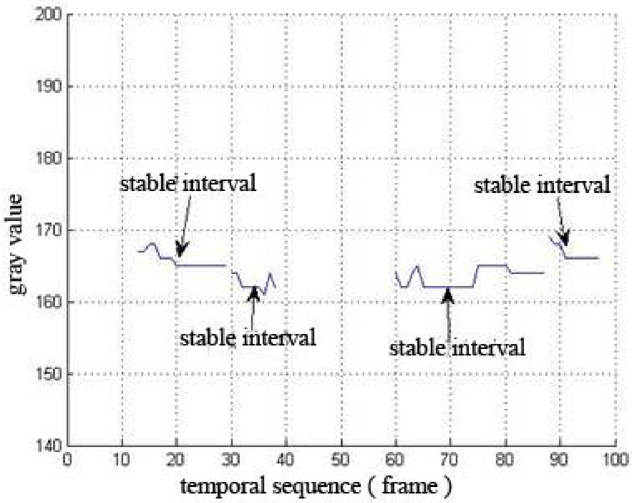

FIGURE3 Stable intervals candidates

Background Obtaining

Generally, the median of interval is robust, a classify set $< s_1, \cdots, s_k >$ is made up of medians of all stable intervals, and the median is expressed as:

$$s_j = median(l_j),\ 1 \le j \le k \quad (2)$$

The distance between medians relates to the similarity between intervals. The classify set could be classified into some sub-class with the similarity. Circular regions are constructed with s_j as the center, $\delta_{\max}$ as the radius, the number of samples in the region is denoted as "density". The interval which has the highest density is the background interval. And the sample which is nearest to the center of the background interval is the initial background pixel. When every pixel of the image is processed with the algorithm mentioned above, the background initialization would be realized.

After obtaining the initial background, the background update algorithm (*11*) is utilized to overcome the impact of the environmental variability. Ultimately, the foreground moving objects could be extracted by background subtraction. The motion detection result achieved (see fig.4 as an example) is reliable and satisfactory.

FIGURE4 Moving objects detection results
(a) Source frame; (b) Foreground moving objects

FEATURE REPRESENTATION

After eliminating the background, foreground objects extracted have integrated silhouettes. We denote the distance between the point of silhouette and the geometric center as centro-distance, all of centro-distances make up of a vector along the silhouette, and we recognize and classify moving objects with this vector.

The algorithm of the feature representation based on centro-distance vector is composed of following essential steps:

Geometric Center Calculation

A segmented $M * N$ binary image, the (p,q) moment of region R can be expressed as:

$$m_{pq} = \sum_{i,j \in R}^{M} \sum^{N} i^p j^q \quad (3)$$

m_{00} is denoted as the pixel amount of the region, m_{10}, m_{01} is denoted as the central moment. So the coordinate $C\ (x_c, y_c)$, the geometric center of the foreground region can be defined as:

$$x_c = \frac{m_{10}}{m_{00}} = \frac{\sum_1^n x_i}{n}\ , y_c = \frac{m_{01}}{m_{00}} = \frac{\sum_1^n y_i}{n} \quad (4)$$

Where n is the pixel amount of region R .

Representation of Centro-distance Vector

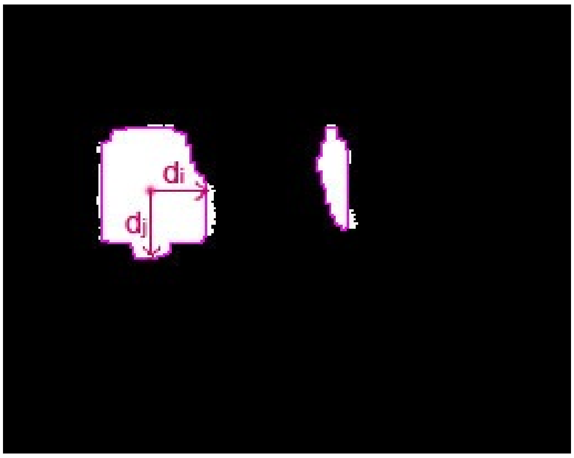

FIGURE5 Centro-distances of targets

At first, according to the connectivity of the segmented foreground object, points of silhouette are sorted as a sequence with anti-clockwise: $L(l_1, l_2, l_3, \cdots l_m)$. And then, we denote the distance between the point of silhouette l_i and the geometric center C as the centro-distance d_i, as illustrated in fig.5, and all centro-distances along the silhouette could be expressed as a sequence, in the paper we transform the sequence into a vector, which is defined as the centro-distance vector of the object:

$$D(d_1, d_2, d_3, \cdots d_m)^T \quad (5)$$

$$d_i = dist(C, l_i), \forall i \in (1...m) \quad (6)$$

Optimization of Centro-distance Vector

Different objects have different silhouette amounts. Accordingly, centro-distance vectors' dimensions of different objects are not same, too. Therefore, elements of the fixed amount should be picked up from the original sequence as a new vector to ensure centro-distance vectors of different objects have same dimensions, and these elements should be dispersible and well-proportioned to reflect the whole silhouette information of the moving object. So the centro-distance vector is optimized as:

$$D'(d_1, d_2, d_3, \cdots d_k)^T \tag{7}$$

$$D'[i] = D[i \cdot \frac{m}{k}], \forall i \in [1...k],\ k \in N \tag{8}$$

The k dimension centro-distance vector could be obtained from Eq.8. But owing to the variety of moving status, obviously different values would be taken on of one dimension in different situation, which is easy to classify by mistake. So more optimization should be carried on to overcome the impact from the scale varying, every dimension vector of different objects should have the same scale, and could be compared with each other. So the centro-distance vector further optimized is denoted as:

$$D''(d_1, d_2, d_3, \cdots d_k)^T \tag{9}$$

$$D''[i] = \frac{D'[\mathrm{i}]}{\sum_1^{\mathrm{k}} D'[\mathrm{i}]} \tag{10}$$

In Eq.10, $D''[i]$ denotes the ratio of the i dimension value to the sum of total dimensions', this optimization could get over the influence of scale variety. Because the centro-distance has the invariability of rotation and translation already, the feature D'' is fixed on rotation, translation and scale.

Representation Feature Based on Centro-distance Vector

Centro-distance vector curves of the vehicle and the bicycle are showed in Fig.6 and Fig.7, centro-distance vector features of moving objects in different classes have the relativity, the variety of one class feature curve is obviously different from others, accordingly, $D''(d_1, d_2, d_3, \cdots d_k)^T$ between the vehicle and the bicycle is observably different, too. From Eq.9, Eq.10, the mean value of the centro-distance can be obtained:

$$\bar{D} = \frac{1}{k}\sum_1^k D''[i] \tag{11}$$

The dispersedness of the centro-distance vector:

$$M_1 = \frac{1}{k}\sum_1^k (D''[i] - \bar{D})^2 \tag{12}$$

The ratio of the max centro-distance to the min:

$$M_2 = \frac{MAX(D''[i])}{MIN(D''[i])} \tag{13}$$

From Eq.12, Eq.13, The dispersedness of the centro-distance, the ratio of the max centro-distance to the min could be calculated, which are chosen as the representation feature for recognition and classification in the next step.

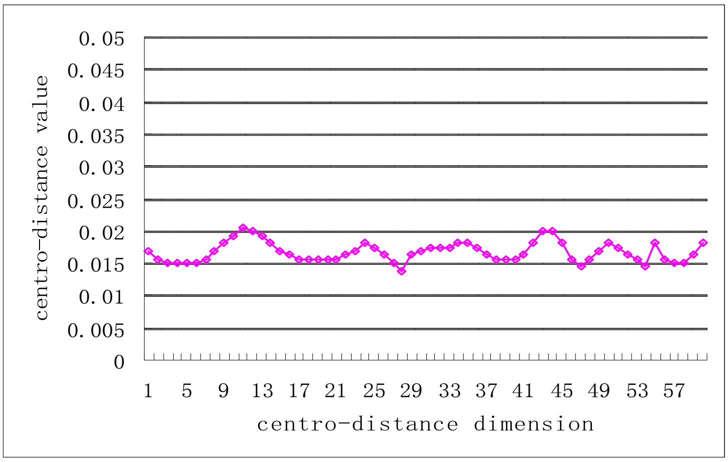

FIGURE6 Curve of vehicle's centro-distance vector

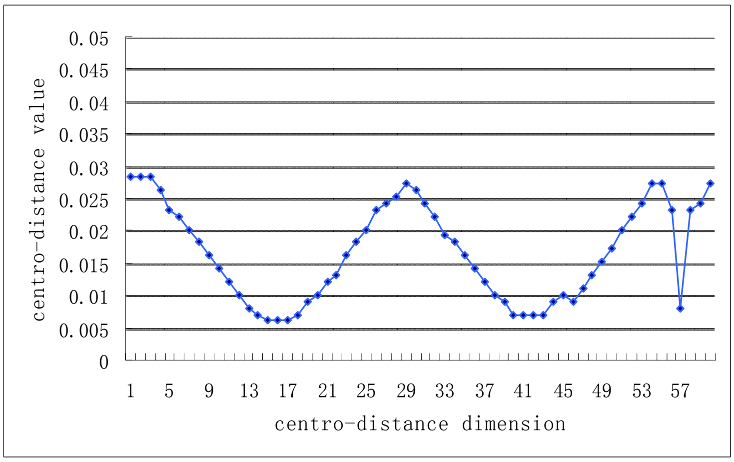

FIGURE7 Curve of bicycle's centro-distance vector

RECOGNITION AND CLASSIFICATION

Minimum-Distance Method

In the process of recognition and classification for moving objects, the dispersedness of the centro-distance vector and the ratio of the max centro-distance to the min are chosen to be distinguishing features. In the paper, objects of the mixed traffic flow are divided into three categories: vehicles, bicycles and persons. Firstly, some sample feature data of known category are collected to construct a sample storage, and the sample centers of different categories could be obtained by arithmetic mean based on minimum-distance method, which represent thresholds of the classification. And then when the new sample comes, the similarity between the new sample and the sample center of every class is calculated, and the new sample belongs to the class, the similarity of which is most.

The similarity between non-recognition object's eigenvector X and sample center's is calculated as Eq.14.

$$D(X, X_i) = \sqrt{(X - X_i)(X - X_i)^T} \quad (14)$$

If the similarity $D(X, X_i) < D(X, X_j)$, X is closer to X_i.If the similarity between X and X_i is the least of all, X belongs to the class whose sample center is X_i. When the new sample is put into the known class, the sample center is recalculated and updated. Minimum-distance method is advantage of to realize the classification quickly and efficiently due to its simple computation.

Temporal Consistency

In order to enhance the classification reliability, temporal consistency of the object feature should be considered. The distinguishing feature based on the centro-distance vector is only accounted for spatial information. And the spatial feature may be changed with the targets moving. So it's not enough to classify by the distinguishing feature in only one frame, the feature stability and consistency should be taken into account. In most video detection sequences, the status of moving targets will not change frequently within a few frames, which means in most cases the object class for a moving object should be the same in neighboring frames. In this paper, we classify the moving target detected in the process of motion tracking by Kalman filtering algorithm (5). In five continuous frames, if the object takes on the same type in three frames or more, it will be classified into the type. This method is applied to further improve the accuracy of the moving object classification.

EXPERIMENTAL RESULTS

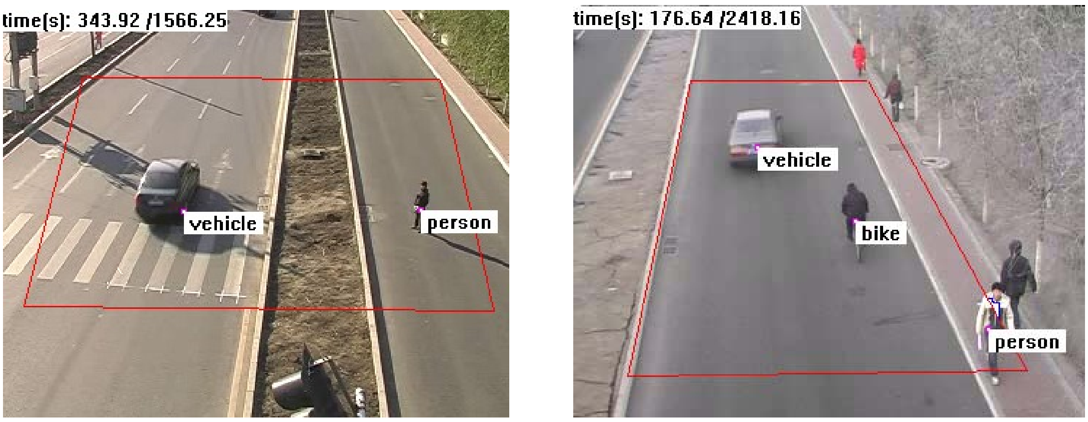

FIGURE8 Samples of classification in mixed traffic scenes

For testing the validity of the proposed algorithm, the video sequence is obtained from realistic scenarios by camera. Fig.8 is examples of classification results in different traffic scenes. In order to reduce the detection extent and increase the detection efficiency, the effectual detection region is constructed (In fig.8, lining quadrangles), and the algorithm presented is only used for the effectual detection region. Table1 is the result of classification for moving targets in mixed traffic.

TABLE 1 Classification Results of Moving Objects in Mixed Traffic

Type of Object	No. of Instances	Classified Correctly (One-Frame)	One-Frame Accuracy	Classified Correctly (Five-Frame)	Five-Frame Accuracy
Vehicle	52	47	90.38%	51	98.08%
Bicycle	45	31	68.89%	38	84.44%
Person	48	34	70.83%	42	87.50%

From table1 we can see that it's not good enough to classify by one frame, the recognition accuracy rate of bicycle or person is approximately 70%. When temporal consistency is taken into account, accurate classification results can be obtained. Especially the accuracy rate of vehicles, can achieve above 98%. However, because of the similarity of the representation feature between bicycles and persons, recognition accuracy rates of them are about 85%, new effectual features should be selected for the representation, and distinguishing between bicycles and persons should be further studied in future.

CONCLUSION

In this paper, recognition and classification of moving objects in mixed traffic by video detection is developed. A background extraction algorithm is utilized to detect moving targets effectively; a robust feature representation method based on centro-distance vector is presented, which has the invariability of rotation, translation and scale; Minimum-Distance method combined with temporal consistency is proposed to classify moving objects into three categories: vehicles, bikes and persons. Experiments have been conducted on real-world scenes under different conditions. Good classification performance and high accuracy have been confirmed.

ACKNOWLEDGEMENTS

This research was supported by The National Key Basic Research and Development Project of China (No. 2006CB705505).

REFERENCES

1．Collins R．T．，Lipton A．J．，Kanade T．et al．A System for Video Surveillance and Monitoring．*tech. report CMU-RI-TR-00-12，Carnegie Mellon University,* 2000．

2．Bogomolov Y．， Dror G．．，Lapchev S．et al．Classification of Moving Targets Based on Motion and Appearance．*British Machine Vision Conference*，2003，2，429~438．

3．Kuno Y．，Watanabe T．，Shimosakoda Y．． Automated detection of human for visual surveillance system．*Proc IEEE International Conference on Pattern Recognition*，1996，865~869．

4．Toth D．，Aach T．．Detection and Recognition of Moving Objects Using Statistical Motion Detection and Fourier Descriptors．*International Conference on Image Analysis and Processing*，2003，430~435．

5．O．Javed，M．Shah．Tracking and Object Classification for Automated Surveillance．*Proc of ECCV*，2002，Part IV，343~357．

6．Cutler R．，Davis L．． Robust Real-time periodic motion detection analysis and application．*IEEE Trans Pattern Analysis and Machine Intelligence*，2000，22(8)，781~796．

7．Lipton A．． Local application of optic flow to analyse rigid versus non-rigid motion．*Tech．report*，1999，13．

8．Mckenna S．． Tracking Groups of People．*Computer Vision and Image Understanding*，2000，80(1)：42~56．

9．B． Shoushtarian，H． E． Bez．A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking． *Pattern Recognition Letters*，26 (2005) ，5~26．
10． Wren Christopher Richard，Azarbayejani Ali，Darrell Trevor et al．Pfinder：real-time tracking of the human body． *IEEE Transactions on Pattern Analysis and Machine Intelligence*，1997，19(7)，780~785．
11． Stauffer Chris，Grimson W. E. L．Adaptive background mixture models for real-time tracking． *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*，1999，2，246~252．