Traffic Flow Detection for Complex Scene Based on Image Sequence

Li Na
1. Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
2. National Supercomputer Center in Jinan
Ji'nan, China
lina@sdas.org

Feng Jin-qiao
1. Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
2. National Supercomputer Center in Jinan
Ji'nan, China
fengjq@sdas.org

Pan Jing-shan
1. Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
2. National Supercomputer Center in Jinan
Ji'nan, China
panjs@sdas.org

Gu Wei-dong
1. Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
2. National Supercomputer Center in Jinan
Ji'nan, China
guwd@sdas.org

Zhao Yan-ling
Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
Ji'nan, China
zhaoyl@sdas.org

Liu Guang-qi
Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
Ji'nan, China
liugq@sdas.org

Zhong Hong-jun
1. Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
2. National Supercomputer Center in Jinan
Ji'nan, China
zhonghj@sdas.org

Abstract—A method of traffic flow detection for complex scenes based on image sequence is proposed, which is aiming at realizing intelligent vehicle detection and flow statistics for traffic videos shot by single camera located at the urban traffic intersection. For target vehicle detection, Gaussian-filtering mean method is utilized to create the dynamic real-time background; meanwhile it is combined with frame difference method to locate the motion vehicles in the foreground. Moreover, test-stripe detecting method is used for vehicle counting, in which data streams representing vehicle information are extracted with sliding window and are modified by Predictor-Corrector scheme to count the vehicles more accurately. Finally, a prototype system is designed to realize vehicle flow statistics for complex traffic scenes, combined with the judgment algorithm of vehicle traveling direction. Experimental simulation represents that target vehicle detection avoids the respective disadvantages of the two methods, that is background difference method and frame difference method, and obtains better detection effects. The prototype system captures good real-time performance and statistical data of the vehicle flow are comparatively accurate.

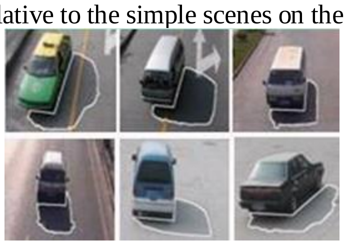
Keywords-vehicle detection; statistics of vehicle flow; complex scenes; background difference method; frame difference method

I. INTRODUCTION

Recognized as an advanced subject mainly studied by various countries nowadays, the intelligent transportation system (ITS) is an effective approach for the solution of conflicts between the rapid development of the automobile industry with urbanization and the limited road capacity, which plays an important role in the optimization of traffic layout, the prediction of traffic flow, the improvement of traffic efficiency and the ensurence of traffic safety. Study on ITS in China started in the late 1980s [1], and remarkable effects have been obtained in research, development and application until now. Among many provinces and cities, ITS has been constructed and applied in the aspects like traffic police management, road management, urban traffic planning and etc. [2].

With rapid development in recent years, intelligent traffic monitoring and control system combines technologies like computer vision and activity recognition with ITS, which is implemented effectively in vehicle recognition, vehicle speed-test, traffic flow measuring, vehicle license plate recognition, violation recognition and etc. As a result, it is recognized as an important branch of ITS. Although the traditional buried coil inductor has great accuracy in detection, the device is required to be installed in the civil structure of the road, so the life period is short and the road is damaged during construction and installation. However, traffic behaviors can be analyzed via image contents when the intelligent traffic monitoring and control techniques are applied. So the advantages are obvious that the cost is low and it is easy to use, to install, to adjust, and to change the monitor area.

For intelligent traffic monitoring and control, one of the significant problems is the video-based vehicle detection and tracking of which the accuracy can be greatly influenced by environmental factors. Difficulties in detection can be caused by resolution and vibration of the camera, change of the illumination, shaky movement of trees in the background, parallax and etc.. Among all the factors, the shadow of motion vehicle (shown in Fig. 1a) and the occlusion between vehicles (shown in Fig.1b) are of great significance in the influence of detection accuracy. Conglutination of adjacent vehicles during detection caused by the former one can lead to detection error, while the latter one may cause some superposed vehicles to be detected as a single object. In addition, problems mentioned above are more serious in the complex traffic scenes of which traffic intersections in urban areas (shown in Fig.1c), relative to the simple scenes on the highways,



(a) the motion shadow

(b) the vehicle occlusion

(c) Low resolution in urban intersection

Figure 1. Influencing factors of video vehicle detection

because unclear identification of excessively small moving target under low resolution image condition will also bring difficulties for detection.

II. LITERATURE REVIEW

The research of the video vehicle detection in complex scene mainly concentrated in two aspects, the target detection and accurate background model, focusing on solving the problem of the motion shadow, the illumination variation, the occlusion between vehicles.

Aiming at the complex monitoring scenes whose key monitoring area and non-key monitoring area have significant degrees of distinction, *Yong Luo et al.* [3] propose a fast motion targets detection algorithm using edge intra-frame difference to achieve background modeling rapidly and moving targets detection accurately. *Sixing Xiao et al.* [4] bring forward a tracking method for moving objects in complex scenes based on mean-shift template updated with a motion estimation method using Particle filter. *Zhihong Zhao et al.* [5] propose an improved Gaussian Mixture Model algorithm for real-time motion detection in complex scenes. Texture feature named scale invariant local ternary pattern was used to verify the candidate fore-ground points to get rid of the moving shadow and illumination variation.

Sen-Ching S et al. [6] compare various background subtraction algorithms for detecting moving vehicles and pedestrians in urban traffic video sequences, such as frame differencing, median filter, linear predictive filter, Kalman filter, Gaussians Mixture Model (GMM). The paper reveals GMM produces the best results, while it has its own drawbacks (computationally intensive & parameters sensitive) and adaptive median filtering offers a simple alternative with competitive performance. *Liyuan Li et al.* [7] present a Bayesian framework that incorporates spectral, spatial, and temporal features to characterize the background appearance and to detect foreground objects from complex environments, e.g., subway stations, campuses, and sidewalks. *A. Colombari et al.* [8] describe a technique that produces a content-based representation of a video shot for multiple moving objects, including object segmentation based on ego-motion compensation, background modeling using tools from robust statistics, and region matching based on the Mahalanobis distance.

For the complex scenes in the traffic video, we put forward the target vehicle detection and traffic flow statistics method, to solve the various problems appeared in actual application, e.g., background modeling is not reliable when the light changes, frame difference method is failure caused by the vehicle congestion, the optimal segmentation threshold is difficultly obtained on the low resolution images, vehicle counting is inaccurate when it changes traveling lane. Experimental results show that the traffic flow detection has the characteristics of accuracy and real time.

III. DETECTION OF THE TARGET VEHICLE

The moving object extraction methods which are used commonly include optical flow method, edge detection method, background difference method, frame difference method and etc. Background difference method and frame difference method can meet the real-time requirement of vehicle video detection, but have disadvantages respectively. For the background difference, the detection precision relies on the reliability of the background model to a great extent. However, for the frame difference method, the static or slowly-driven vehicles cannot be detected effectively. In this paper, the two methods are combined, which ensures both speed and accuracy of the target detection simultaneously.

Suppose $f(x,y,t_1)$ and $f(x,y,t_2)$ are the values of the coordinate point (x,y) from the gray image in the original video at time t1 and t2, so the binary frame-difference image satisfies the following equation,

$$f_{out} = \begin{cases} 255 & |f(x,y,t_1) - f(x,y,t_2)| > T_d \\ 0 & \text{others} \end{cases} \tag{1}$$

where T_d denotes the segmentation threshold. The frame difference method describes the inter-frame motion change of the target object, so there exits needless little-holes or isolated -region during the extraction of the target vehicle. It cannot get an ideal result just with morphological operator. This disadvantage is designed to be improved with the application of background difference method.

The widely concerned Gaussian Mixture Model better adapts to changes of illumination and the environment, but for the complex traffic scenes with low video resolution, the histogram of the gray image is relatively intensive, which makes it difficult for choosing the parameters of the algorithm. In this paper, background modeling is realized by a mean method based on 2D Gaussian low-pass filter. Since the Fourier transform of Gaussian function is still Gaussian function, the function itself forms a low-pass filter with smooth performance in the frequency domain, of which the expression of the transfer function is,

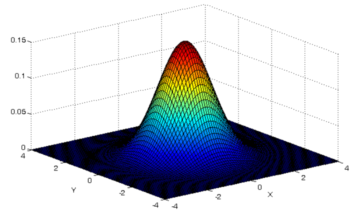
$$H(u,v) = \exp(-\frac{D^2(u,v)}{2\sigma^2}) \tag{2}$$

where $D(u,v)$ is the distance from the origin of the Fourier transform, σ denotes the expansion degree of the Gaussian curve. The basic idea realized by the filter is to construct a point spread function with a presentation of 2D Gaussian distribution. The function is called convolution kernel in discrete cases, of which the cutoff point is set at the value three times of the standard deviation(schematic diagram of 2D Gaussian distribution is shown in Fig.2a, while a 5×5 Gaussian convolution kernel function based on the distribution with the variance $\sigma = 1.0$ is presented in Fig.2b).

Suppose a video sequence of N continuous frames is $f_t(x,y), t = 1,2,\ldots,N$, the Gaussian convolution kernel function is $K(i,j), 1 \le i \le n, 1 \le j \le n$, where $[n \times n]$ denotes the kernel function size of, so the dynamic background of the video sequence is expressed by the following equation,

$$B(x,y) = \sum_N \frac{1}{N} f_t(x,y) \otimes K(i,j) \tag{3}$$

where $\otimes$ is the discrete convolution symbol. Physical meaning of the above equation is that the weighted average value of some pixel is taken as the corresponding pixel value on the static background, while the closer the pixel is to the center, the higher weight it gets. After subtracting the obtained background modeling from the video frame, we get binary background-difference image, which is taken into the logical "or" operation with the binary frame difference image, and detect the target moving vehicle.



(a) Schematic diagram of 2D Gaussian distribution

$$\frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

(b) Kernel funcion（5*5, $\sigma = 1.0$ ）

Figure 2. Principle of the Gaussian low-pass filter

IV. STATISTICS OF VEHICLE FLOW

The test-stripe detection method is designed for statistics of vehicles in complex scenes like urban traffic intersections. The vehicle information is transformed into vehicle data stream which will be modified through Predictor-Corrector scheme. Moving vehicles in different orientation are counted by the location and change information of the data stream in corresponding test-stripe zone. The method can reduce the erroneous inspection induced by vehicle shadow with high precision and low calculation cost.

The test-stripe is set up in the image according to actual situation and experience. Then we obtain the binary segmentation image using the method of section 3. Sequentially it is processed with opening operations and closing operations by a 3×3 diamond operator with a radius of 1. This morphological algorithm is propitious to remove noises produced by camera vibration. The vehicle information after morphological de-noising is quantified to generate data stream. "1" denotes that there is vehicle information at the corresponding location in the test-stripe, while "0" denotes that there is not. Then the vehicle information can be completely represented by data stream of the frames like (00011111100 00001111100000).

The target detection image of the test-stripe zones after morphological de-noising is denoted as $I(x,y), x = 1,\ldots,m; y = 1,\ldots,n$, here $m \gg n$. Suppose that sliding window width is w, so the gray value sum of each window can be calculated as follows,

$$\sum_{x=1}^{w}\sum_{y=1}^{n} I(x,y) \quad i = 1,\ldots,\frac{m}{w} \tag{4}$$

If the result exceeds the specified threshold, it is indicated there exists a passing vehicle on the location of the sliding window at present, so the location is assigned with the value "1". Otherwise, the value is "0". Then the data stream is generated from the current video frame representing the vehicle motion information. Each "0"or "1" is called an *information bit* in further detail below.

The preliminary extracted data stream may have strip breakages caused by partial information lost of the vehicles.

Here a scheme named Predictor-Corrector is designed to modify the data stream. As the minimum distance of two parallel vehicles is known, the breakages would be wrapped by the filling operator to make the data stream as continuous as possible. The filling operator is expressed as follows,

$$I_{0n} = \begin{cases} 1 & L_0 < D_v \\ 0 & L_0 \geq D_v \end{cases} \quad n = 1, \cdots, L_0 - 1 \tag{5}$$

where I_{0n} is the nth "0" information bit; L_0 is the length of the strip breakage; D_v denotes the minimum horizontal distance between vehicles. To reduce the undetected error rate caused by parallax or the headlight, filtering operator is applied on the few consecutive information bits that are not enough to represent vehicles,

$$I_{1n} = \begin{cases} 1 & L_1 < W_v \\ 0 & L_1 \geq W_v \end{cases} \quad n = 1, \cdots, L_1 - 1 \tag{6}$$

here I_{1n} is the nth information bit of the few consecutive "1"; L_1 is the length of the strip breakage; w_v denotes minimum width of the vehicle. After modification with Predictor-Corrector scheme, the data stream can be applied for detection of the normal running, slowly moving, stop-and-go vehicles of various sizes.

The error counting easily occurs when vehicle is changing lane or long bus is passing. So we calculate the distance of the vehicle centers between the adjacent frames. If the distance is greater than or equal to the width of one lane, it will be assessed that two vehicles passed. Otherwise, it will be regarded as the same vehicle; simultaneously we use the data stream of the frame with more information bits to modify that of the other frame.

Finally, we compare the data stream of the adjacent video, which will generate four situations: a) the data streams share the same value of "0" in both frames, which means that there are no vehicles passing; b) the value of the data streams changes from "0" of the former frame to "1" of the latter, which means that new vehicles are arriving; c) the data streams share the same value of "1", which means that vehicles are passing by; d) the value of the data streams changes from "1" to "0", which means that vehicles are leaving.

Total process of flow statistics in complex scenes is shown in Fig. 3. The judgment algorithm of vehicle traveling direction [9] is improved by utilizing Mahalanobis distance to region matching. The image region to be matched is denoted as **I**, while the target vehicle matrix is denoted as **X**, so we get the registration matrix **Y** which make the Mahalanobis distance minimized by the following expression:

$$\underset{\mathbf{Y} \in \mathbf{I}}{\arg\min} \sqrt{(\mathbf{X} - \mathbf{Y})^T \mathbf{S}^{-1} (\mathbf{X} - \mathbf{Y})} \tag{7}$$

where S is the covariance matrix. Mahalanobis distance describes the similarity of two unknown sample sets and is independent of the scale.

V. EXPERIMENTAL SIMULATIONS

Several videos shot by the traffic management department of Jinan are applied in the experiment. The videos are shot in daytime for normal weather and road conditions (represented in Fig. 1c) from the complex traffic intersections in prosperous and bustling downtown of Jinan. The videos are collected by a single fixed camera located at the intersection of Jingshi Road and East Youth Road in Jinan City with a frame rate of 25 per second and size of 560×474.

A. *Adaptive threshold selection*

Parameter T_d in target detection method denotes the segmentation threshold of the image. With appropriate selection of the threshold, frame difference noises can be effectively removed and target segmentation effect can be enhanced. The method of maximum classes square error (Otsu method) [10], iterative method [11] and least-half-samples method (LHS) [12] are applied in the experimental simulations (based on the video in Figure. 1c), of which the results are shown in TABLE I.

In TABLE I, the iterative error of the iterative method takes the value of 0.1, meanwhile in the LHS method T_d=3.4*σ, where σ is the standard deviation calculated from the frame difference image. As shown in the table, the method Otsu obtains the highest operation speed and it generates a similar threshold to that calculated by the iterative method. Actually, theories of the two methods are the same that the image is segmented into background and foreground objects according to the gray characteristics to obtain the optimal threshold. For the image sequences of

TABLE I. COMPARISON OF THREE ADAPTIVE THRESHOLD METHODS

Adaptive Threshold Methods	Threshold (0-255)	Running Time(ms)
Otsu	44	15.4
Iterative method	46	96.2
LHS	25	86.1

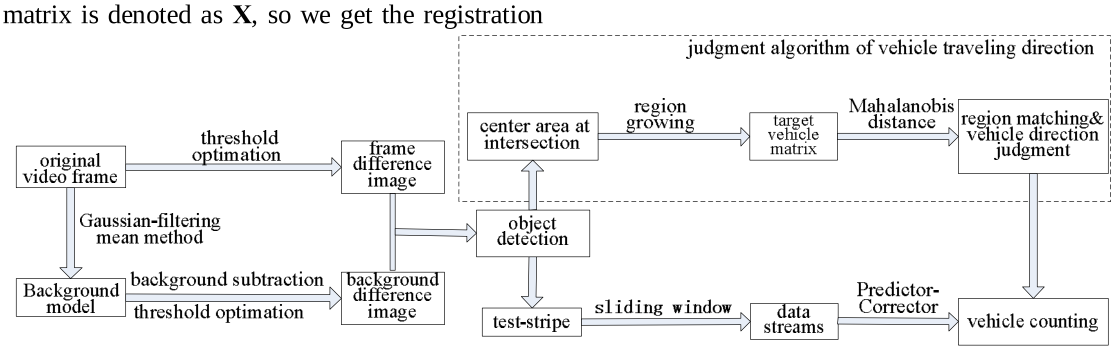


Figure 3. The flow chart of traffic flow statistics method in complex scenes for urban intersection

traffic scenes (Fig. 4a), the motion region is easy to be misjudged as the background (Fig. 4b-c) by these two methods for the reason that the difference induced by vehicle motion is very small. Results of experiments showthat LHS method keeps the vehicle information as much as possible, so that it is suitable for the threshold segmentation in traffic scenes. Simultaneously, the time cost of this method is relatively low. So LHS method is applied in the adaptive threshold selection and the value of T_d is modified according to the actual effect ($T_d=2.5*\sigma$ in the original paper).

(a) 317th frame of the video sequence

(b) frame difference image of Otsu

(c)frame difference image of the iterative method

(d)Frame difference image of LHS ($T_d=3.4*\sigma$)

Figure 4. Comparison of the segmentation threshold

B. Target vehicle detection

Experimental results of the target vehicle detection are shown in Fig. 5. The background model is shown in Fig. 5b created from the pre and post 10 frames around the 317th frame (Fig. 5a)in the video using the Gaussian-filtering mean method. Fig. 5c is the frame difference image. Fig. 5d is the background difference image. Fig. 5e is the final target detection image; Fig. 5f is the target detection image after morphological de-noising. It can be seen from the Fig. 5 that the frame difference image can represent contours of vehicles, but there are some "little-holes" at the positions of vehicles with low speeds. However, the target vehicles in the background difference image generated by the background subtraction method are clearer, but some vehicles (for example, the green in the figure) are completely not detected as their pixelsô values are too close to the background after transformation into the gray image. The final detection image is obtained with advantages of the two methods combined. And the failure of single method in the special situations is prevented.

(a) 317th frame of the video

(b) background modeling

(c) Frame difference image

(d) Background difference image

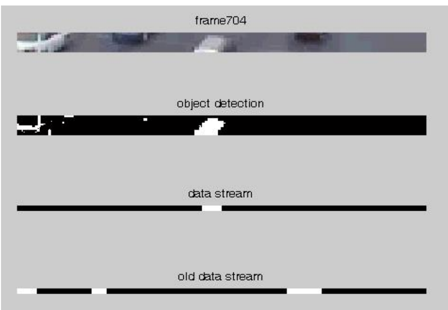
(e) the final target detection

(f) after morphological de-noising

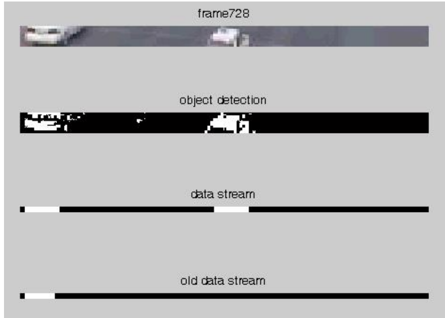
Figure 5. Results of the method of target vehicle detection

C. Vehicle counting

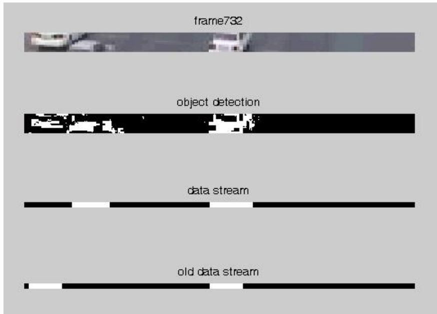
Some results of the vehicle counting method are shown in Fig. 6. Video frame numbers are marked on the pictures in which actual situation, object detection image in test-stripe, data stream image for the present frame and the previous frame are displayed successively from the top to the bottom. It can be seen from Fig. 6a that the data streams of their locations are turning from 1 to 0 when two vehicles of deep colour are driven out of the test-stripes; from Fig. 6a~b and Fig. 6d, the data streams are turning from 0 to 1 when five vehicles of different type and colour are driven into the test-stripes; from Fig. 6b~c, the data streams continues to be 1 when the vehicles are passing by the test-stripes. Experimental results represent that the vehicle counting is of high accuracy and good real-time performance.



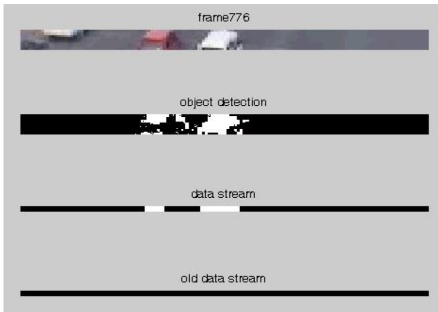
(a)



(b)



(c)



(d)

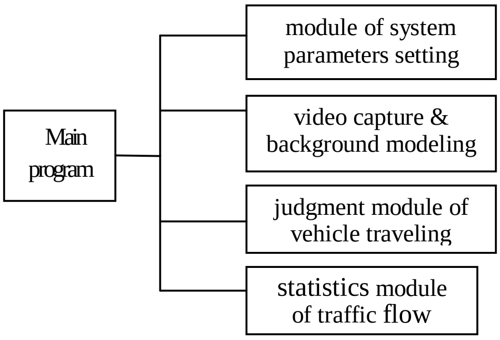
Figure 6. Experimental results of vehicle counting by test-stripe detection method

D. System simulation

A video vehicle detection system is also developed for complex scenes, which has been tested by a video file of 3554 frames including 3 periods of the traffic light. Hardware and software platforms for the system simulation tests are as follows:

- Hardware environment ： Pentinum4 CPU 3.20GHz, 1G memory, 80Ghard disk;
- Software environment ： Windows XP professional, Visual C++6.0, DirectShow9.0.

The development is based on the principle that the system should be convenient for algorithm improvement and modular design. Composition of function modules is shown in Fig. 7a, and the main interface of the system is shown in Fig. 7b. The system tests represent that the bottom test-stripe captures the highest detection accuracy rate with an average of 95%; next is the vertical two test-stripes with 89%; last one is the top test-strip with 86%.



(a) function modules of the prototype system

(b) main interface

Figure 7. Illustration of the prototype system

VI. CONCLUSIONS

To solve the problems in actual situations, the method of traffic flow statistics for complex scenes is proposed in this paper, which is aiming at the test videos shot by single camera located at the urban intersection. For target vehicle detection, the Gaussian-filtering mean method is utilized to create the dynamic real-time background, so background difference method and frame difference method are combined to locate the vehicles for detection. The test-stripe detection method is applied for vehicle counting, in which data streams representing vehicle information are extracted with sliding window after morphological de-noising and then modified through Predictor-Corrector scheme to count the vehicle number accurately. Combined with the judgment algorithm of vehicle traveling direction, the prototype system for complex traffic scenes is realized finally. Experimental simulations represent that target detection image avoids disadvantages of the two methods respectively and obtains good detection effects. The system test captures good real-time performance and statistical data of the vehicle flow are comparatively accurate.

ACKNOWLEDGMENT

This work is supported by Shandong Province Outstanding Research Award Fund for young scientists of China (Grant No. BS2012DX036), the National Key Technology R&D Program (Grant No. 2012BAH09B03), Major Project of Jinan City (Grant No. 201208003), Shandong Provincial Science and Technology Development Planning (Grant No. 2013GGX10133) and National Major Project of High-tech Service Research, Development and Industrialization (Grant No. [2013] 1383) which are gratefully acknowledged.

REFERENCES

[1] Yang Zhao-sheng. Theory and method for intelligent public traffic system in urban area[M]. Beijing: China railway publishing house, 2004.

[2] Pinghua Zhang, Adaptive control system of traffic signals in Qingdao, Chinese Transportation Information Industry, 2004, 10:50-51.

[3] Yong Luo, Jianjun Zhao, Linjiang Zheng. The research of key technologies on moving targets detection under complicated scenes [D]. Shanghai: Shanghai Jiao Tong University, 2012.

[4] Sixing Xiao, Jingxin Hong. Move object detect and tracking under complex scene [D]. Fujian: Xiamen University, 2008.

[5] Zhihong Zhao, Xuhong Tian, Yuefang Gao. Real-time motion detection in complex scenes[J]. Journal of Chinese Computer Systems, 2013(4):915-920.

[6] Sen-ching S. Cheung , Chandrika Kamath. Robust techniques for background subtraction in urban traffic video. Proceedings of Video Communications and Image Processing, SPIE Electronic Imaging (January 2004), pp. 881-892.

[7] Liyuan Li,Weimin Huang, Gu I.Y.-H. , Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. IEEE Transactions on Image Processing, 2004,13(11): 1459ï 1472.

[8] A. Colombari, A. Fusiello, V. Murino. Segmentation and tracking of multiple video objects. Pattern Recognition, 2007.4, 40(4) : 1307-1317.

[9] Li Na, Zheng Xiao-shi, Zhao Yan-ling, Yang Cheng-zhong. Video-based Vehicle Detection Scheme in Complex Traffic Scene at Urban Intersection[C]. International Conference on Information Engineering and Computer Science, 2009. ICIECS 2009, Page(s): 1009-1012.

[10] N.Otsu. A threshold selection method from gray-level histograms[J], IEEE Transactions on Systems, Man and Cybernetics,1979,9(1) : 62-66

[11] R.C. Gonzalez, R.E. Woods, "Digital Image Processing"[M], Prentice Hall, 2002. p.305

[12] Ong E P, Tye B J, Lin W S, et al. A Fast Video Object Segmentation Scheme for MPEG-4 Video Coding[J]. Nanyang Technological University, Innovation Centre.