Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

An Effective Background Reconstruction Method for Complicated Traffic Crossroads

Hong Liu
Key Laboratory of Machine Perception and Intelligence
Key Laboratory of Integrated Micro-system
Shenzhen Graduate School, Peking University, CHINA
E-mail: hongliu@pku.edu.cn

Wei Chen
Key Laboratory of Integrated Micro-system
Key Laboratory of Machine Perception and Intelligence
Shenzhen Graduate School, Peking University, CHINA
E-mail: chenw06206@szcie.pku.edu.cn

***Abstract*—Effective background reconstruction is the key for real time traffic flow monitoring. High traffic density and complexity of background scene make reconstruction more difficult. Background estimation based on the median method is imprecise under a complex traffic flow condition. In this paper, a new background estimation method based on the similarity of background using parameters of gray mean and variance is proposed. Therefore, a two-dimensional clustering and merging mechanism is introduced. At last, accurate decision about the category of the background is made by analyzing the distribution characteristic of the frame numbers in one category. Our algorithm works on the difficult condition of traffic congestion with higher reliability. The proposed method can be used in background reconstruction of the crossroads based on video sequences.**

***Keywords*—Intelligent Transport System (ITS), Crossroads, Two-dimensional clustering, Background reconstruction, Block matrix**

I. Introduction

A. Problem definition

Video image processing technology has been widely used in Intelligent Transportation Systems (ITS). In the metropolis, traffic congestion is serious in crossroads. Traditional flow detection methods have a requirement of huge operating cost. Therefore, the convenient real-time video processing technology has become more popular and important.

Many automatic video object segmentation algorithms are based on the assumption that the background image of video sequences is given. This assumption uses the difference between the video input frame and background as a first estimation. However, in many situations, recording a background image alone without foreground objects is difficult and unrealistic, because no one can control the scene change. It is also hard to adapt to the slow changes of the background. Therefore, there is a need to estimate the background images using automatic algorithms.

B. Previous works

Researchers have encountered a lot of background estimation problems, for which a good research can be found in [1], including changes of overall lighting and background appearance. Researchers have developed different approaches to monitor the traffic flow based on video sequences. Generally speaking, there are some traditional methods about back-ground reconstruction: median method [2], frequency method [3] and optical flow method [4] [5]. The median method was applied to scenes in which foreground objects are less complex. The method requires improvements in some specific situations. The frequency method was based on the assumption that background objects appear the most frequent. If the background throughout the video sequences only can be seen in minority, it is necessary to combine other methods. The optical flow method calculates changes of individual pixels, taking into account the motion of objects. It encounters difficulties at complex crossroads. First of all, the huge scene would make optical flow parameters inestimable. And stationary vehicles may be determined into background easily by mistake.

Literatures [6] proposed a differ matrix of frames. The author obtained the background image through calculating the degree difference of gray-value between sample frames in the same pixel block. The author only used gray-scale as a single feature, which should not get very stable results in complex and changeable scene like crossroads. Other method is proposed in [7]. It generated two background images. One recorded static information over the entire video sequence, while the other recorded gray-scale changes of pixel in a short period used to update the background information. *G.* Gordon and his team made a clustering based on two-dimensional, gray-scale amplitude and color feature in [8]. The algorithm might fail when the color of foreground objects is similar to background.

Mei Xiao and Lei Zhang set two threshold to the gray-value and the appear frequency of pixel in [9][10][11]. The value of threshold was determined by the clustering of similar elements. The method was applied to the scenes in which small amount of changes occurs, and the background objects must appear at most time.

Background estimation method based on the joint use of median-method and Support Vector Machine (SVM) method was suggested in [12]. It obtained the unknown surface parameters by training the known road and generated the unknown road image combining with median-method. The limitation of the method was that researchers must obtain the static calibration of the road. The authors of literature [13] also introduced texture feature to update the background. They held that the pixels of background dose have a different texture feature besides spatial and temporal characteristics. The authors calculated texture by circular Gabor Filter.

New method was presented in [14] in 2008. The author determined the candidate pixel value base on the gray-scale distribution, and calculated the optical flow of candidate to determine a most likely background value. The method may achieve very good result in not-too complex scenes. Besides [13], there are many new reconstruction algorithms.

In all experiments listed above, the background is generally visible at most time and the scene is not huge. Some of them can only be applied in specific occasions. Most methods are unstable when vehicles are static for a long time and even in congestion. In fact, no effective background estimation algorithm can apply to all scenes at present.

II. MOTIVATION

In this paper, we propose a clustering method based on the similarity of background using parameters of gray mean and variance to identify the background image of video frame sequences. The proposed method produces superior results on identifying the difference between stationary vehicles and road, with an accurate estimation on the whole background of crossroads.

Most background reconstruction techniques operate at the pixel-level, making decisions for each pixel. Fig.1 shows a typical change of intensity for one pixel in the traffic scene image over time. It can be seen that the middle part of curve is stable and the double ends are mutation. The true value of the background may be in the middle position or in double end if the road information can not be seen at most of time. Here, the middle position may be corresponding to the value of static traffics. Therefore, an accurate decision based on this factor alone is impossible.

The pixel-level data is useful in reducing the number of possible value of the background. However, it is insufficient. Dirk, Peter and Wolfgang in this paper processed video image in block-level [7]. They separated images into many blocks and used mean of intensity as a computing parameter. Their method worked well when all foreground objects are in motion during the sequences. However, many blocks may be in-correctly classified if foreground objects were stationary for a long period of time, or different objects may have the same gray scale mean (Fig.2). To overcome this problem, our method also considers the variance of gray scale in each block.

One significant step is clustering the sequences in the same block by the mean and variance of gray scale. Then extract the background and static foreground class. In most of the scenes, the static foreground gathers in a continuous period of time, and the background may disperse at any time of video sequences. At last, we make accurate decision on the true value of the background by analyzes the distribution characteristic of frame number in one class.

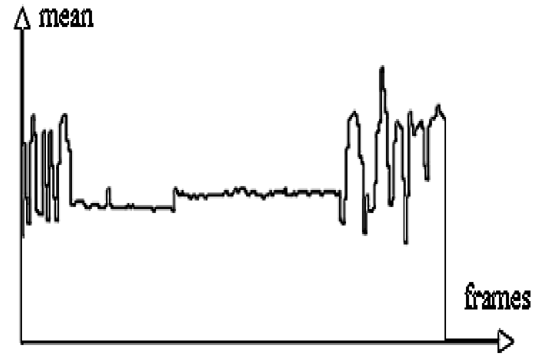

Fig.1 Mean histogram of pixel blocks.

Fig.2 Continuous vehicles flow at a crossroad.

Because of the specificity of vehicles scene, when vehicles move slowly (through a smaller margin between adjacent frames), one sample frame can be extracted at a long intervals. When the margin is comparatively large, we extract one frame at short intervals. Finally we get enough frames as a calculation sample, using mean and variance of gray-scale as parameters, which assure that enough background information can be obtained in video sequences.

III. CLUSTERING AND MERGER METHOD

A. Separate image into blocks

The time spent in calculating single pixel parameters is impressive, whether it is combined with gray scale, color or texture, even with multi-dimensional characteristics. Inspired by [4], this paper put block as a unit to deal with information. The computation can be reduced and noise interference can be eliminated with only considering the block as a unit. Compared to cluster computing of frame sequences, time of gaining parameters of one block can be neglected. Therefore, the computing time is inversely proportional to the size of each pixel block. However, a complete background block is scarce when the block size increases, which results the loss of eventual background unpredictable. Therefore, size of pixel block should be limited to an appropriate value.

B. The mean and variance of block matrix

Set the block size be $N \times N$ and let the length of sequences be L. In addition, let $f_i(x, y)$ be the intensity of pixel(x, y) in input frame I, assume that $f_i(x, y) \in [0, 255]$. For each block (u, v) with the top left pixel at position (u_n, v_n), we calculate the gray mean of block $E^{(u, v)}$ and variance $D^{(u, v)}$,

$$E_{(u,v)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f(u_N + i, v_N + j)}{N^2} \tag{1}$$

$$D_{(u,v)}^{\ 2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} (f(u_N + i, v_N + j) - E_{(u,v)})^2}{N^2} \tag{2}$$

here, each video frame is cut into pixel block equally. We train 200 frames in the same block.

Corresponding to the content of video changes when the time flow, $E^{(u, v)}$ and $D^{(u, v)}$ change in block (Fig.3). If the brightness of whole video scene does not change, the gray mean and variance of background block are approximate equivalent. It provides the basis for background extraction.

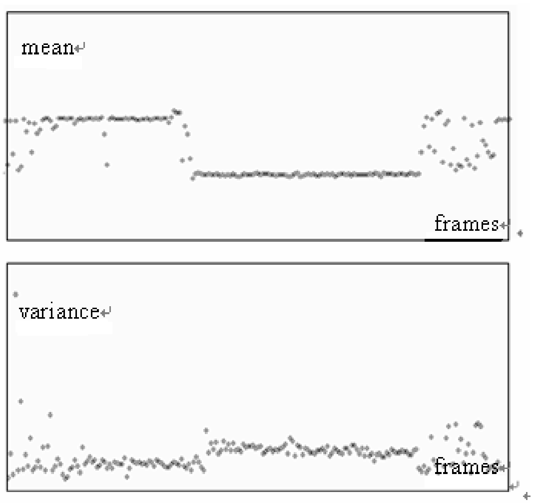

Figure 3: change chart of gray-scale mean (top) and variance (below) in 200 frames. Abscissa means frames, while ordinate corresponding to E (u, v) and D (u, v).

C. Character clustering

Considering a given block, the mean and variance of background change little in relatively long and stable video sequences. (The scene brightness does not have a mutation). The method classifies L frames in a block by comparing the mean and variance of gray scale with a clustering algorithm.

The distance $d_{(a,\ b)}$ between frame a and frame b is calculated using Formula (3). Considering that variance of overall brightness has less impact in changing environment, it is necessary to give greater weight for variance. The formula can be defined as following:

$$d_{(a,b)}^{(u,v)} = \sqrt{m * (E_b - E_a)^2 + n * (D_b - D_a)^2} \tag{3}$$

Here, $m + n = 2$. In the experiments, the value of m is 0.8 and n is 1.2. E_a, E_b, D_a, D_b are the mean and variance of frame a and b. Set the threshold d_k, if $d_{(a,\ b)} < d_k$, classify frame a and b as one category. Different thresholds make different clustering results. Fig 4 shows the graph of the clustering.

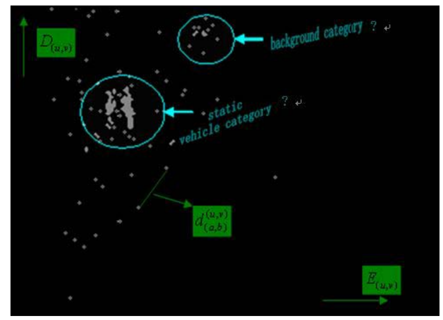

Figure 4 graph of two-dimensional clustering

D. Merger of the same category

Because of the little illumination changes in scene, there are two or more background categories existing. Maybe both are background categories. That should make a failure to classify them as the same category, however, just because the distance of them is larger than given threshold. The failure may lead to a lost of background information. Therefore, it is necessary to make the merger of category.

After clustering, elements of one category are in line with Gaussian distribution. Therefore, two background categories may contain some elements whose distance is less than given threshold d_k, which are called equal elements. We merge category A and B by searching the equal elements.

Assume that the experiment has generated category $A(x|\ x \in A)$, $B(x|\ x \in B)$, $C(x|\ x \in C)$, $D(x|\ x \in D)$,$\cdots$,in clustering steps. Let x_i (E, D) $(i = 1, 2, \cdots, L)$ be the elements of all categories in pixel block $(u,\ v)$. The algorithm of category merger is shown in Fig.5:

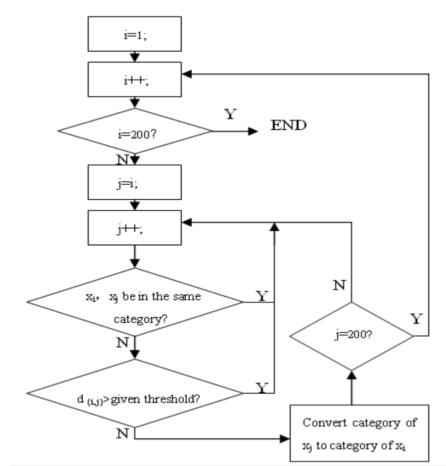

Figure 5: merger of category

The proposed method makes first rough clustering by calculating the distance between two frames. Then combine categories by comparing two elements in their category. Table Ⅰ shows that the content of video sequences is classified into 3 or 4 categories. In general, we conclude all non-static frames to a category, because each non-static frame category has few elements. The static frames included background frames and static foreground objects. (Waiting vehicles) The extraction of background was discussed in the following.

Table Ⅰ: different thresholds make different cluster results

d_k	Category numbers (before merger)	Category numbers (after merger)	mean	variance
80	6	3	133	22
			72	41
			Non-static frame	
50	7	3	132	24
			72	41
			Non-static frame	
30	11	4	132	24
			70	38
			73	41
			Non-static frame	

E. Recognition of background frame

At last, L frames must converge to some categories, one containing more elements will most likely be a background category. The parameters of the element are frame number with its gray mean and variance of pixel block. We note that categories containing the most elements do not correspond to background certainly. Because they may be represent foreground when vehicles wait for a long time.

One effective way to solve this question is calculating the dispersion of elements frame number. We know that the video sequences corresponding to static foreground category is continuous and unique, assuming that the vibration of camera is ignored or amended prior. That means they generally account for only major section of the sequence, while the background may exist in any section of video sequences. The dispersion is defined by calculating the variance of frame number of all elements in category. However, only computing the dispersion of element is unreliable. Because of the noise and accidental factors, some categories of non-background may generate a larger dispersion, even if they contain few elements. Therefore, the number of category must be considered.

Let M be the number of category A $\{a1, a2, a3, \cdots, a_M\}$, a_i be the frame number of element. Let E_A be the variance of A. then calculates category value with Formula (4):

$$Val_A = \sqrt{M} * E_A \tag{4}$$

here, Val_A presents category value. The greater it is, the more likely it is background category. In order to reduce some value of major static-foreground category, we select the square root of the value of M.

IV. EXPERIMENTS AND DISCUSSIONS

There are some tests in some traffic crossroads with our method. Some experiment samples are more complex than other video sequences. The experimental materials also include simple traffic junction. The tested video sequences were taken using digital camera. The video was sampled at a resolution of 640*480 and a rate of 30 frames per second. We extract one frame at intervals of 10 or 30, decided by the margin of adjacent frames. To select the size of each block, some work has been done. Table Ⅱ makes a description of relationship between block size and calculation time.

Table Ⅱ relationship between total calculation time and the block size

Block size	**5*5**	**7*7**	**9*9**	**11*11**
time(s)	281	164	106	79
accuracy	98%	96%	95%	92%

It can be seen from Table Ⅱ that the computing time is inversely proportional to the size of each pixel block. The table also lists the accuracy of generated background in complex scenarios of region. We have a 100 percent accuracy rate in simple region of scene. The accuracy means the ratio between correct blocks generated and all blocks in region. According to the table, block size about 7*7 is select with short computing time and high accuracy.

We compare our output of algorithm with median method used for background modeling. Tests are performed on several sequences representative of situations which might be commonly encountered in traffic crossroads. The performance statistics of two algorithms over a variety of sequences are compared in Table Ⅲ. One of the output of our algorithm for examples is shown in Fig.5.

To measure the accuracy of algorithms, we first test the reliability of background, then make a further detecting about flow of traffics in video sequences with our result. From table Ⅲ, error blocks is defined as pixel blocks whose mean value differs from the ideal value by more than some thresholds d_i.(we use a threshold of 20). The total blocks is all blocks in the region we mark off. When the scene is simple, median method has a high accuracy as well as our method in background reconstruction. However, if it becomes more complex, the propesed method get better results than median method with high accuracy.

The first example sequence is simple because there are few vehicle in the road and any area of road is visible in the most time. We can detect vehicle effectively by both methods. In the second sequence, some vehicles stop for a long time which make road information decreased. It result in errors using median method with a 81% correct rate in complex condition.

On the contrary, our method has a good result with 95.2 percentage.

Table Ⅲ: Comparing results of our method and median method in four sequences

Video sequences		**Reconstruction of background**						**Detection of vehicle numbers**						
num	note	Median(num of blocks)			Our(num of blocks)			Different period In sequences	Median(num of vehicles)			Our(num of vehicles)		
		error	total	rate	error	Total	rate		detected	real	Rate	detected	real	rate
1	simple	0	487	100%	0	487	100%	Simple	7	7	100%	7	7	100%
								complex	14	13	92.3%	14	13	92.3%
2	Normal	483	3124	84.5%	212	3124	93.2%	simple	10	9	88.9%	9	9	100%
								complex	24	21	81%	22	21	95.2%
3	Complex	476	1441	66.9%	87	1441	94.2%	simple	14	9	44.4%	10	9	88.9%
								complex	49	39	74.4%	41	39	94.9%

The third example in Table Ⅲ was difficult because some foreground vehicles is present in most of the frame of sequences. In this example, the results of our method are far better than median method. We have a good detection on vehicle flow. Results of effective background reconstruction of the third sequence with two methods are shown in Fig.5. Other tests under much more crowded conditions show that our method is robust to errors in the flow field caused by occlusions.

(a) The complex traffic crossroads

(b) Result of median method

(c) Result of our method

(d) Real background image

Figure 6: our method work well in complex traffic scene.

V. CONCLUSIONS AND FUTURE WORK

We have proposed a new background estimation algorithm in complex video sequences. Contrast to other iterative update algorithms, our approach determine the background frames in the extracting frames in every block using clustering method. Our method works well in general traffic scenes and has better results than some popular algorithms in complex crossroads.

However, if flow of vehicles and persons become huger and huger, the traffic scene also becomes more complex, which reduce the amount of information of background. The algorithm is unable to generate a good background and even makes a large mistake in the condition that many regions of background are not visible in the whole video sequences. Therefore, further improvement of our algorithm will be the next focus of the work.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China (NSFC, No.60675025, 60975050) and National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247, and No.2007AA11Z224), Shenzhen Bureau of Science Technology and Information.

REFERENCES

[1] Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: Principles and practice of background maintenance, Volume 1, 20-27 Sept. 1999 Page(s): 255-61 vol.1Digital Object Identifier 10.1109/ICCV.1999.

[2] Lipton A,Fujiyoshi H, Patil R, Moving target classification and tracking from real-time video, 19-21 Oct.1998 Page(s):8-14 Digital Object Identifier 10.1109/ACV 1998.732851

[3] Robert T. Collins, A system for video Surveillance and Monitoring, Technical Report CMU-RI-TR-00-12,Carnegie Mellon University 2000

[4] Barron J, Fleet D, Beauchem in S, 15-18 June 1992 Page(s): 236-242 Digital Object Identifier 10.1109/CVPR 1992.223269.

[5] Farin, D.; de With, P.H.N.; Effelsberg, W, robust background estimation for complex video sequences, Volume 1, 14-17 Sept. 2003 Page(s):I-145-8 vol.1 Digital Object Identifier 10.1109/ICIP 2003.1246919

[6] Pan Wei, Wu Kun, Chai Zhi, You Zhi-Sheng. A Background Reconstruction Method Based on Double-background, 22-24 Aug. 2007 Page(s):502-507 Digital Object Identifier 10.1109/ICIG 2007.14

[7] *G.* Gordon, T. Darrell, M. Harville, J. Woodfill. Background estimation and removal based on range and color, Volume 2, 23-25 June 1999 Page(s): Digital Object Identifier 10.1109/CVPR 1999.784721

[8] Mei Xiao, Lei Zhang. A Background Reconstruction Algorithm Based on Two-Threshold Sequential Clustering, Volume 1, 3-4 Aug. 2008 Page(s):389-393 Digital Object Identifier 10.1109/CCCM 2008.289

[9] Mei Xiao, Lei Zhang. A Background Reconstruction Algorithm Based on Modified Basic Sequential, Volume 1, 3-4 Aug. 2008 Page(s):47-51 Digital Object Identifier 10.1109/CCCM.2008.294

[10] Mei Xiao, Chongzhao Han, Xin Kang. A Background Reconstruction for Dynamic Scenes, 10-13 July 2006 Page(s):1–7 Digital Object Identifier 10.1109/ICIF.2006.

[11] Shuan Liu, Jun-Yu Dong, Sheng-Ke Wang , Guo-Jiang Chen. Road Junction Background Reconstruction Based on Median Estimation and Support Vector Machines, Aug. 2006 Page(s):4200-4205 Digital Object Identifier 10.1109/ICMLC.2006.

[12] Dongxiang Zhou, Hong Zhang and Nilanjan Ray, Texture Based Background Subtraction, 20-23 June 2008 Page(s):601-605 Digital Object Identifier 10.1109/ICINFA.2008.

[13] Xiaodong Cai, F.H.Ali, E.Stipidis. Robust Online Video Background Reconstruction Using Optical Flow and Pixel Intensity Distribution, 19-23 May 2008 Page(s):526-530 Digital Object Identifier 10.1109/ICC.2008.