REAL-TIME TRAFFIC OBJECT DETECTION TECHNIQUE BASED ON IMPROVED BACKGROUND DIFFERENCING ALGORITHM

Yingying, Zhang, Department of Automation, Tsinghua University, Room 3-423, FIT Building, Tsinghua University, Beijing, China

Chenyao, Geng, Department of Automation, Tsinghua University, Room 3-423, FIT Building, Tsinghua University, Beijing, China

Danya, Yao, Department of Automation, Tsinghua University, Room 3-427, FIT Building, Tsinghua University, Beijing, China

Lihui, Peng, Department of Automation, Tsinghua University, Room 501A, Main Building, Tsinghua University, Beijing, China

ABSTRACT

Background differencing is the major algorithm for real-time object detection, which is widely used in traffic data collection and incident detection techniques. The speed of this method in object detection usually heavily depends on the size of the images and the number of the objects. This paper presents an improved background differencing technique, which draws several crossing lines in the interested region and detections are accordingly performed on these lines. A real-time object detection system built on this method significantly improves the speed of the system. An object detection experiment for real traffic scene is conducted to compare the speed and accuracy between traditional and improved background differencing techniques. Effects on the size of the crossing lines are discussed in this paper as well.

Keywords: grid processing, improved background differencing, motion detection

1. INTRODUCTION

Traffic incident detection technology has been widely studied over the world since the 1960s. As the computer technology develops, video based traffic incident detection comes to its intelligent times. Generally speaking, real-time traffic incidents include irregular stop in free flow, traffic congestion, scattered objects, vehicles going in a direction not allowed by traffic regulation, ultralimit queues, break-in pedestrians, etc. Real-time motion detection is the key issue of these incidents detection because it highly contributes to the following steps, which are object classification, tracking and understanding.

Motion detection is defined as interested objects detection from sequential frames. It involves detection in both dynamic and static scenes. The three major methods for motion detection in static scene are as follows:

1. Background differencing (JAIN, R., 1984), which detects motion by differencing the sequential frames and the reference background pictures.

2. Frame differencing (Barron, J. and Fleet, D., 1994), which detects motion by differencing the sequential frames by time.

3. Optical flow method (Bert, H. and Brian, S., 1981), which detects motion by merging similar motion vectors. (Wang, J. and Adelson, E. 1985)

Of all these methods, object detection based on background differencing algorithm can detect all the pixels of moving objects, but it is highly depended on the environment like daylight illumination etc. Object detection based on frame differencing can finely adapt to the environment but is only sensitive to the distinct changed pixels (Jain, R., 1981). For those slightly changed pixels, this algorithm cannot detect them well. Optical flow method is the best method for moving camera scenes, but most kinds of these algorithms need too much calculation and time, which limit their applications in real-time incident detection.

This paper focuses on camera fixed real-time traffic scenes, which require fast detection, change smoothly and is mainly affected by the weather and daylight illumination. An improved background differencing algorithm is accordingly developed as follows.

2. OBJECT DETECTION BASED ON BACKGROUND DIFFERENCING

2.1 Improved Background Differencing Algorithm

Background differencing algorithm differences the present frame and reference background image to get the foreground image. Foreground objects are collected by threshold segmentation. However, if the reference background image is good enough, the algorithm can obtain all the foreground objects needed by using only one single frame. Besides, comparing with the frame differencing algorithm, this algorithm works better on slow objects.

Commonly used incident detection system for real scene is inevitably affected by the environment, major including the weather and the daylight illumination. In this situation, the reference background image cannot be exactly suited to the real scene, which means too much noise for the foreground object extraction. Meanwhile, in traffic detection system, the foreground information is relatively complex because it includes shadows and overlapping objects. Further processing is definitely necessary since interested targets always accompany the incorrect foreground information by using the traditional background differencing algorithm independently.

Theory of the traditional background differencing algorithm can be described as:

$$\text{Video}(x,y,t) = \text{Foreground}(x,y,t) + \text{Background}(x,y,t) + \text{Noise}$$

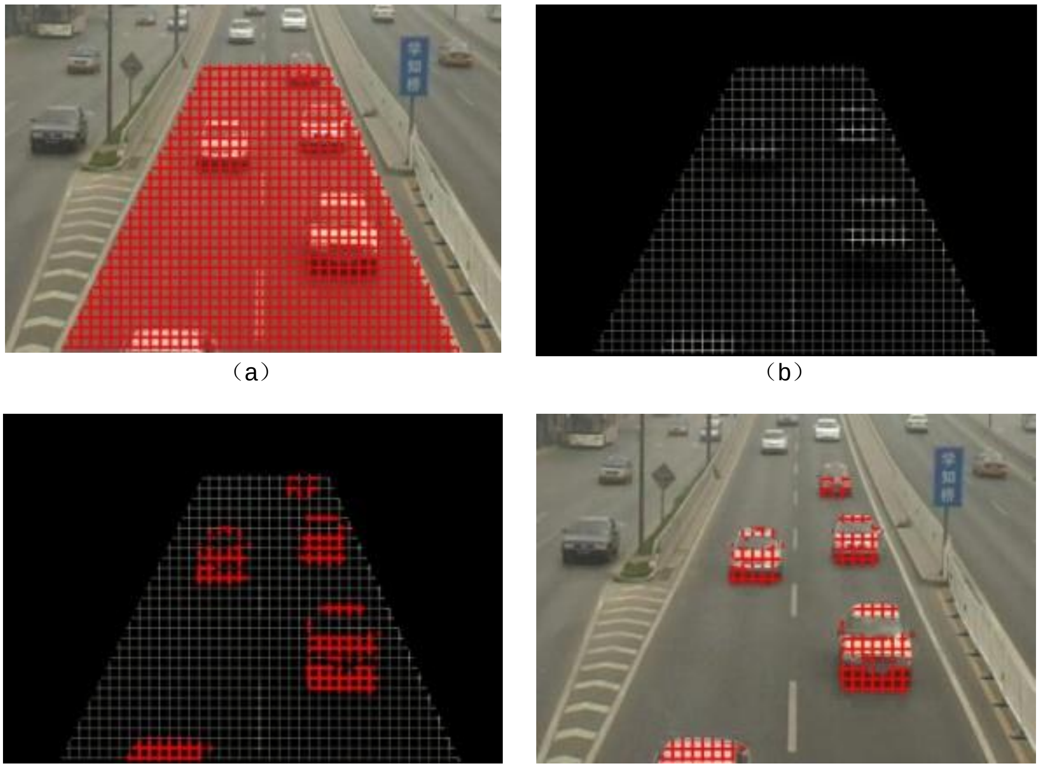
Where Video(x,y,t) is the video sequence for processing, Foreground(x,y,t) is the foreground moving object, Background(x,y,t) is the background information, and the Noise is random noise. The sum of Foreground(x,y,t) and the random noise can be obtained by background differencing with Background(x,y,t). The entire targets can be collected by filtering processing.

Every single pixel needs to be processed to get the foreground objects, which means much calculation and processing time depends on the size of the image. Worse still, too much noise also brings a good deal of extra calculation for the following corrosion expansion step. All these are really disadvantages for the real-time video processing.

In traffic scenes, the interested targets are known to have some video characters in common. For example, cars are usually bigger and rectangular while pedestrians are smaller and often with uncertain shapes. For some mixed traffic scenes, cyclists are to be detected as well, and the speed and location information are highly recommended while the exact outline information is not so much important. An improved background differencing algorithm is presented to accelerate the traditional algorithm and limit the impact of noise.

First, N×N crossing detection lines are drawn in the interested region, which divides the region into $(N+1)^2$ grids. Detection is done on the pixels located on the detection lines. It highly reduces the number of pixels that need to be calculated. For example, for a 100×100 pixels object, there is 10^4 times calculation by using traditional method, which can be reduced to 819 times (9×100-9×9=819) by drawing 10×10 crossing lines on the image. It means a reduction of about 90 percent. Besides, the size of the grids can be determined by the object size, which consequently gets rid of a mass of broken pieces.

A series of lines and points can be obtained because of the crossing line. In order to gain the initial interested objects, expansion processing is required. The expansion window is the same size as the grids. Figure 1 shows the results of gridding and background differencing. Processed foreground objects after morphological filtering and connection recognition are shown in figure 2.



（a） （b） （c） （d）

Figure 1 – Initial foreground objects segment processing

(a) Detection region after grid processing (b) Grey image after background differencing (c) Threshold segmentation on the grid (d) Obtained initial foreground objects

（a） （b）

（c）

Figure 2 – Foreground images after morphological filtering

(a) Foreground image after corrosion and expansion (b) Segmented foreground image with shadows and several broken regions (c) Object identification based on connection recognition

2.2 Shadow Detection Algorithm

The study shows that the shadow region can be considered as semi-transparent region. (Paul L. Rosin and T. Ellis. 1994) It has a photometric gain with respect to the background image, which is less than unity. And this gain is reasonably constant over the shadow region. According to daily experience, the hue of an object does not change whether it is covered by shadow or not. In RGB colour space, it means the R, G, B value reduces proportionally in the shadow region, while the value of R:G:B is more or less the same in both situations. In HSV colour space (Cucchiara, R., Grana, C., and Piccardi, M., 2001), H remains, while S and V reduce by the same ratio.

For the grey images in traffic scene, as the luminance source of the visible camera region is close, the gain is reasonably constant over all the shadow region, which appears as the product of relevant background point and the optical gain k (k<1). For the other regions, the grey value of the image is relative to the foreground image characteristic, gain k varies widely. Therefore, after calculating every k value on the detection lines, it is easy to find out that k for shadow regions rarely changes while k for the real foreground images changes frequently. The road surface as the main background is relatively smooth in traffic scenes, while the foreground objects (e.g. vehicles) are always with clear and definite visible outlines, which can be easily detected by gradient calculation.

The outlines can be detected by using Sobel operator. The edges between the objects and the shadows can be easily detected since the grey values are different. At the same time, the optical gain of the shadows does not change with the objects, so we can remove the shadows by threshold segmentation. Foreground objects without shadows are finally obtained.

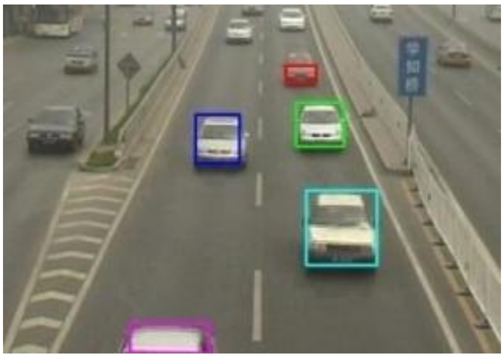
2.3 Merging Broken Pieces

Single objects may be divided into two or more adjacent parts because of the process described earlier in this paper. In traffic scenes, vehicles are always divided into roof parts and body parts because the grey values of car windows are very close to the road. It is then necessary to merge the broken pieces after shadow detection.

This paper merges the broken pieces by neighbourhood detection. Firstly, we classify the foreground objects by size: those bigger ones are recognized as main body, and the smaller ones as broken pieces. Then, we search the interested region until we find a broken piece and search its nearby neighbours that are less than D pixels far. If there is a main body, we merge the broken piece into it and continue searching until no broken pieces can be merged. Here, we define the distance between the objects as the distance between centres. Finally, the foreground targets are collected. Figure 3 shows the process of shadow detection and broken pieces merging.

(a)

(b)



(c)

Figure 3 – Further process for final foreground targets

(a) Foreground image after gradient calculation (b) Foreground image after removing the shadows (c) Finally merged foreground targets

2.4 Motion Detection

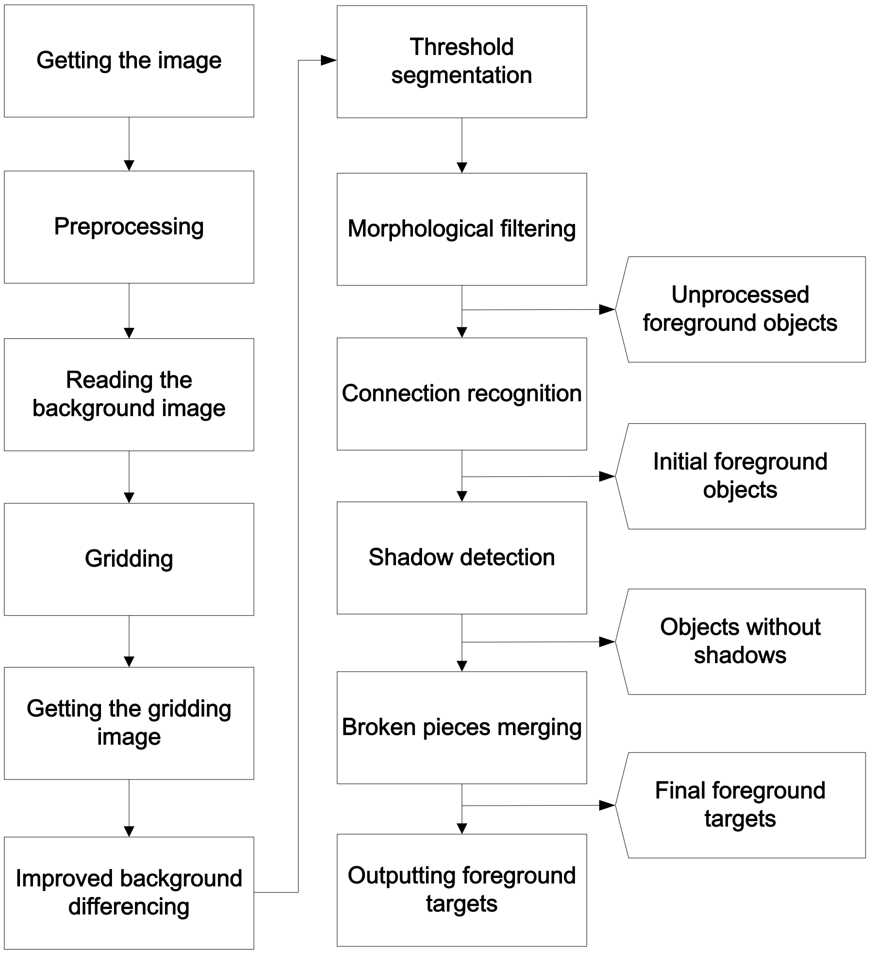


Figure 4 – Basic processing for motion detection

Motion detection processing can be described as:

1. Preprocessing: including median filtering, mean filtering and image enhancement in order to provide better primary images. It is also necessary to set the interested region for further detection.

2. Gridding: drawing N×N crossing detection lines in the interested region, N is relative to the size of the object in the image. Further calculations are only carried out on the lines.

3. Getting the grid image: taking the image pixels that covered by the detection lines as the grid image.

4. Background differencing: differencing the grid image and background image on the detection lines.

5. Morphological filtering: removing small noise by corrosion and expansion, merging some nearby broken pieces and obtain initial objects. In this part, only quite close pieces can be merged. For a preferable result, the threshold is generally set as less than 5 pixels.

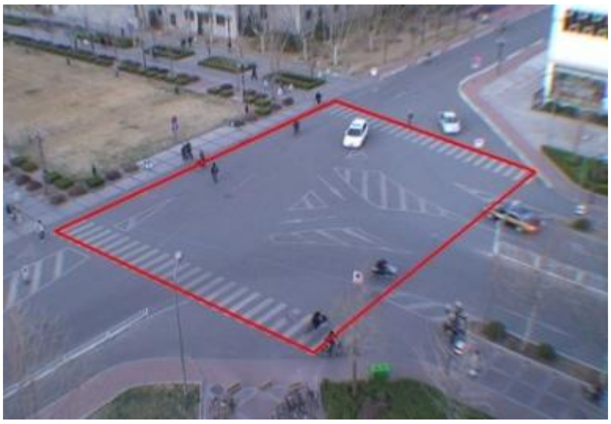
6. Connection recognition for indentifying different objects.

7. Shadow detection for getting objects without shadows.

8. Merging broken pieces and obtaining the final foreground targets.

3. EXPERIMENT AND ANALYSIS

Experiments are conducted on the video of the first T-junction inside the east gate of Tsinghua University and the San-Huan-Zhong-Lu Road in Beijing, China. The detection region is set as Figure 5.



（a）

（b）

Figure 5 – Detection region of express way scene and mixed traffic scene

(a) Mixed traffic scene (b) Express way scene

Figure 5(a) shows mixed traffic scene, which involves certain numbers of cyclists and pedestrians. Figure 5(b) presents express way traffic scene where only vehicles involved. The size of both frames is 1024 pixels × 576 pixels.

For each scene, we test the time of object detection on 500 sequential frames and the results of both traditional and improved background differencing processing are shown in Table I.

Table I – Time test of object detection

Processing time	Mixed traffic scene		Express way scene	
	Without grid	Grid=2	Without grid	Grid=5
Average time（ms）	94.0	86.4	193.7	79.4
Maxim time（ms）	438.0	99.9	1089.9	89.5
Minim time（ms）	44.7	84.1	42.4	73.3

The result shows, the average time of object detection has been limited to its 40% in express way scene by improved background differencing algorithm. For mixed traffic scene, the grid should be small enough to make sure the tiny objects would be detected, including cyclists and pedestrians. In this mixed traffic scene, the grid is set to 2 and the average time reduces about 9%. Generally speaking, it can accelerate the average speed of video processing by about 30%～40%. Comparing with the traditional method, the maxim processing time is highly reduced. The reason is that the traditional method operates much on the noise in some particular frames involving too much noise , while the improved method can left most of the noise out by grid processing.

Table II – Statistical result in mixed traffic scene

Frame No.	Cars			Tiny objects		
	Right report	False report	Misreport	Right report	False report	Misreport
0	1	0	0	5	2	0
1	1	0	0	5	1	0
2	1	0	0	5	1	0
3	1	0	0	4	0	0
4	1	0	0	4	1	0
5	1	0	0	4	1	0
6	2	0	0	4	1	0
7	2	0	0	3	0	0
8	2	0	0	3	0	0
9	2	0	0	3	0	0
Total	14	0	0	40	7	0
Rate of failure		0%	0%		15%	0%

Table III – Statistical result of express way scene

Frame No.	Right report	False report	Misreport
0	5	0	1
1	5	0	1
2	4	0	1
3	4	0	0
4	4	0	0
5	4	0	0
6	4	0	0
7	4	0	0
8	4	0	0
9	4	0	0
Total	42	0	3
Rate of false		0%	7%

Table II and Table III are the statistical results of object detection from the first 10 frames. Every segmentation mistake is counted. In mixed traffic scene, tiny objects, including cyclists and pedestrians, are also counted.

The results show that improved background differencing algorithm can achieve a preferable and efficient goal for real-time incident detection. Comparing with the mixed traffic scene, the targets for detection in the express way scene are relatively bigger and with larger areas, thus the grid can be bigger with relatively fine detection result. But for the mixed traffic scene, the grid should ensure both the cars and the tiny targets to be detected. However, the objects far away from the camera are much smaller than the near ones in the express way scene since the lack of projective transformation. It results in the major misreport.

4. DISCUSSION

The advantages of improved background differencing algorithm are:

1. Faster processing speed. Compared to the traditional methods which cost about 150ms to process a single frame that is 348 pixels × 288 pixels and may reach up to 180ms per frame for some particular scenes, this improved method limits the processing time to less than 100ms for a single frame that is 1024 pixels × 576 pixels with relatively high accuracy. It thus provides a way to process more frames per second and saves more time for motion tracking and other dynamic operations.

2. Preferable segmentation effect. In express way scene, the rate of misreport is 7%, which is relatively tolerable for real-time video detection and the segmentation effect is good enough for further operation. Projective transformation before gridding obtains relatively higher accuracy.

3. Better-adaptation. By setting reasonable grid size, this method fits both mixed traffic scene and express way scene very well.

However, this method has some disadvantages. The major interested targets and the size of the grids have to be set according to different traffic scenes with different kinds of targets require different accuracy. Thus, for some fixed scenes with well-established kinds of targets, this method may provide fault information if other kinds of targets break in.

On the whole, the average time of object detection has been averagely reduced to its 60%～70% due to the grid processing. Real-time object detection system built on this method can significantly improve the speed of the system.

REFERENCES

Barron, J. and Fleet, D. (1994). Beachem in S. Performance of optical flow techniques, International Journal of computer Vision. J., 12(1), 42-47.

Bert, H. and Brian, S. (1981) Determine Optical Flow. Artificial Intelligence, 185 - 203.

Cucchiara, R., Grana, C., and Piccardi, M. (2001). Improving shadow suppression in moving object detection with HSV color information. IEEE Intelligent Transportation System Conference, pp. 334 - 339. Oakland, CA, USA.

JAIN, R. (1981). Dynamic Scene Analysis Using Pixel-based Processes. Computer. J., 14(8), 12–18.

JAIN, R. (1984). Difference and accumulative difference pictures in dynamic scene analysis. Image and Vision Computing. J., 2(2), 99-108.

Paul L. Rosin and T. Ellis. (1994). Image Difference Threshold Strategies and Shadow Detection. Proc.of the 6th British Maehine Vision Conference, pp. 347–356.

Wang, J. and Adelson, E. (1985). Determining three-dimensional motion and structure form optical flow generated by several moving objects. IEEE Trans on Image Processing. J., 7(4), 384–401.

This work was supported in part by The National High Technology Research and Development Program of China (863 Program) 2009AA11Z206, National Basic Research Program of China (973 Project) 2006CB705506, National Natural Science Foundation of China 50708055, 60834001, Postdoctors in China Science Foundation 20090460326 and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.