VISION-BASED MOVING VEHICLES DETECTION IN INTELLIGENT TRANSPORTATION SYSTEMS

WANG Chun-Bo ZHANG Wei-Dong XU Xiao-Ming
(Department of Automation, Shanghai Jiaotong Univ., Shanghai 200030, China)

Abstract A new moving vehicle detection method in complicated background was presented. First, moving regions are detected using the motion information of three frames by a statistical method, then the results of motion segmentation are regulated using the region information produced by static segmentation. And an image enhancement technique is proposed to improve the effects of segmentation. The results can be used as the basis of advanced vehicle control and traffic management in intelligent transportation systems (ITS).

Key words motion segmentation, region segmentation, image enhancement.

智能交通系统运动车辆的视觉检测

王春波 张卫东 许晓鸣
(上海交通大学自动化系,上海,200030)

摘要 提出了一种在复杂背景中检测行驶车辆的方法,首先用概率统计的方法通过三帧图像间的运动信息找出运动区域,然后用静止分割得到的区域信息修正运动分割的结果,并采用图像增强技术改进了分割效果。检测结果可作为智能交通系统(ITS)中高层交通管理和车辆控制的基础.

关键词 运动分割,区域分割,图像增强.

Introduction

In recent years many researchers have proposed computer vision techniques in intelligent transportation systems (ITS). Computer vision provides the possibility to extract complex, high-level road traffic information such as congestion, accident or fluid traffic, thus allowing efficient planning of a path through the road network, or to deviate the traffic. In order to extract this type of information it is first necessary to segment moving vehicles from the scene.

Methods based on optical flow or image difference are commonly used. Optical flow measurement is difficult in areas of little texture or object boundaries and computationally expensive. The real-time method is required for applications in ITS, so we choose image difference to deal with this problem. This method was primarily stated as simple pixel-based frame difference followed by thresholding[1]. In Ref. [2], an approach based upon subtracting the current image from the background image was presented. The background image is updated to account for changing external conditions by a Kalman filter. This method requires a certain number of frames until a reliable background is available. Recently statistical frameworks are adopted to improve the results of image difference. Mitiche and Bouthemy[3] applied Markov random fields theory to regulate the moving region labeling results based on spatiotemporal derivatives of three subsequent images. However, this technique is highly computationally intensive.

All the above methods only use motion infor-

Received 2000-01-31, revised 2000-03-27
稿件收到日期 2000-01-31,修改稿收到日期 2000-03-27

mation. There are also some methods using model-based techniques to extract vehicles from the image[4]. These are actually static segmentation methods. In this paper we propose a method for detecting multiple moving vehicles on highway. In our approach motion and static segmentation are integrated together to produce good results. An image enhancement technique is presented to improve segmentation effect. The algorithm is fast and easy to realize in practice.

1 Motion Segmentation

The first step of the algorithm is the motion segmentation. This module performs a segmentation between static and dynamic regions in the image sequence by providing a set of binary masks, which coarsely represent the shapes, and the positions of the moving objects. First we get the masks of moving regions by image difference of three frames. Dubuisson[5] proved that if all edges in an image are ramp edges, then the difference image $d(i,j)$ obtained by

$$d(i,j) = |f_1(i,j) - f_2(i,j)| \times |f_2(i,j) - f_3(i,j)| \qquad (1)$$

has non-zero value only at the locations of the edge in the second frame, where $f_1(i,j)$, $f_2(i,j)$, $f_3(i,j)$ are the gray value functions of three frames (not necessary to be consecutive), respectively. This difference image is then thresholded to produce the masks of moving regions.

This approach performs sufficiently well if moving objects in the sequence are well contrasted. But it is difficult to determine the thresholds to adapt to the changes along the sequence for different moving objects. Moreover, if the contrast of moving objects is not sufficiently high as compared to the camera noise, there might not exist the threshold able to get rid of noise and preserve the motion information at the same time. To obtain more robust results, we apply a method to dynamically change the threshold according to the sequence characteristics by taking into account the statistical behavior of each pixel's neighborhood. The technique we have chosen was proposed by Aach[6] first. It is based on that the noise that causes changes can be supposed to be uncorrelated Gaussian noise. Then we can discriminate the changes due to motion from noise by methods of hypothesis testing. Two hypotheses need to be defined:

H_0: there are not moving objects at image position $[i,j]$.

H_1: complementary of H_0.

Suppose $d_{[i,j]}$ is the gray value of the difference image at $[i,j]$. Given the hypothesis H_0, the probability $P(d_{[i,j]}/H_0)$ in which $d_{[i,j]}$ is different from zero obeys a zero-mean normal distribution $N(0,\sigma)$ with variance σ^2. In order to make the detection more reliable, we chose the test statistics as shown in Eq. (2).

$$\Delta^2_{[i,j]} = \sum_{w^n_{[i,j]}} \left(\frac{d_{[i,j]}}{\sigma} \right)^2, \qquad (2)$$

where $w^n_{[i,j]}$ is a window of width n centered at pixel coordinates $[i,j]$ (the window size can be 3×3 or 5×5).

From statistical theory[7] we know that $P(\Delta^2_{[i,j]}/H_0)$ obeys a χ^2-distribution with $n\times n$ degrees of freedom, then we can apply a significance test to determine the change at pixel $[i,j]$ because of motion or noise according to Eq. (3):

$$\alpha = P(\Delta^2_{[i,j]} > t_\alpha/H_0). \qquad (3)$$

If $\Delta^2_{[i,j]}$ at pixel position $[i,j]$ is over the threshold t_α, we mark the pixel as changed. Two parameters need to be determined beforehand. The variance σ^2 of the camera noise can be obtained off-line. The significance level α should be below 10^{-3} to increase the reliability of the results. With existing tables we can obtain the corresponding value of t_α.

Furthermore, in order to obtain a better result we add a preprocessing step to the above motion segmentation method. The following mask is proposed to enhance the edges and contrast of the image.

$$L \begin{pmatrix} 5 & 5 & -3 \\ 5 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The enhancement effect of this mask can be seen by the following analysis. The neighborhood of a step vertical edge MN is shown in Fig. 1. Suppose that the pixels on the two sides of the edge have the same gray levels, respectively (represented by "•" and "×"). The following results can be obtained from the convolution of the mask with the neighborhood:

1. The gray values of the pixels on the left side of the pixel "a" and the right side of the pixel "b" are magnified $12L$ times.

2. The step edge between "a" and "b" is changed into the ramp edge between "a" and "c".

3. The edge gradient is magnified $15L$ times.

Similar results can be obtained for the edge in the other directions. When $L>\frac{1}{12}$, the above mask has the effect of enhancing the contrast of the image and the larger value of L can produce better results, if only it doesn't make gray level saturation. This mask is applicable when most pixels in the image have gray level less than $255/12L$. In this paper L is chosen as $\frac{1}{6}$.

Finally, a 2-D median filter is applied on the difference image in order to smooth the masks boundaries, and small regions are eliminated. Figure 2(a) is one frame of a traffic image sequence. The obtained masks of moving regions without enhancement are shown in Fig. 2(b), and those obtained with enhancement shown in Fig. 2(c). Motion segmentation only provides us with coarse representations of the moving vehicles. Due to transparency and aperture effect, some parts of the moving objects are not detected as changed. A refinement process is thus required to extract a more accurate representation of the object shapes. We present a method integrating static segmentation to improve the detection results.

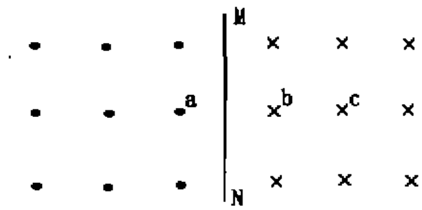


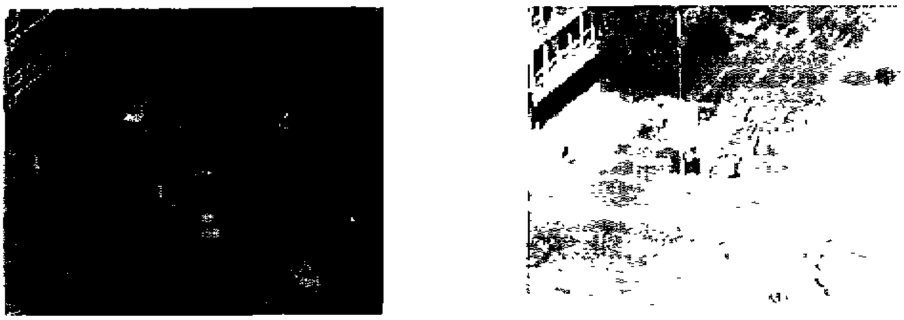
Fig. 1 The neighborhood of a vertical edge
图 1 垂直边缘邻域

2 Static Segmentation

The second step of the method is the static region segmentation. The purpose of this step is to divide the image into homogeneous regions. We choose Felzenszwalb's region segmentation approach as the basis for our algorithm[8]. Felzenszwalb treated image segmentation as a graph partitioning problem. In such approaches there is a graph whose nodes are in one-to-one correspondence with pixels of the image. Certain nodes are connected by edges whose weights meausre the distance between the corresponding pixels in some feature space. Central to Felzenszwalb's approach is a function that decides whether or not it makes sense for two regions to be distinct. A particular

(a) (b) (c)

Fig. 2 The motion segmentation results
(a) the original image (b) result without enhancement (c) result with enhancement
图 2 运动分割的结果
(a) 原图像 (b) 未经增强处理的分割结果 (c) 增强处理后的分割结果



(a) (b)

Fig. 3 The region segmentation results
(a) result without enhancement, (b) result with enhancement
图3 区域分割的结果
(a) 示经增强处理的分割结果, (b) 增强处理后的分割结果

function that measures the difference between two regions relative to some property of each of them is defined. The property is derived from the entire region, rather than from some local or fixed-size neighborhood. An important characteristic of this method is its ability to preserve details in low-variability image regions while ignoring details in high-variability regions. This assures that the segmentation results are neither too coarse nor too fine.

Our approach differs from Felzenszwalb's in the definition of the edge weight function, which in Ref. [8] is the absolute intensity difference between the pixels connected by an edge. A new item is added to measure the texture information, which is defined as the absolute difference of the response to a DOG filter between the same pair of pixels. We still need to apply the above enhancing method first. The above mask can enhance the edge gradient, thus the regions with small intensity difference can be separated correctly. By contrast, the region segmentation result without enhancement for the frame of Fig. 2(a) is shown in Fig. 3(a), and that with enhancement shown in Fig. 3(b).

3 Integration of Motion and Static Segmentation

Motion and static segmentation produce a list of regions, respectively. The result of motion segmentation is very rough. We integrate the above two steps to produce a result more close to the true shapes of the vehicles. Suppose C_i is the connected region produced by region segmentation, n_i is the total number of pixels in C_i, and n_i^m counts the number of pixels which belong to motion regions. If n_i^m/n_i is larger than the threshold, C_i is considered to be part of a moving vehicle. Thus a new group of connected regions C'_i is produced. By this means the undetected moving parts due to transparency and aperture effect can be found. The threshold n_i^m/n_i is chosen to be 0.5. This value can't be too large to compensate for transparency and aperture effect. Some background regions that have the gray level very close to the moving objects may be also taken as the motion regions. So C'_i should be refined further. From the gray-level histogram we can see that the gray level of the road is within a small range around one peak I^m_{hi}. Although there may exist several peaks, it's not difficult to find the peak corresponding to the road background. Using this prior knowledge we can regulate C'_i by the following method.

In each connected region C'_i:

(a) Is the current pixel a boundary one?

(b) Yes. Suppose I_i^c is the gray value of the current pixel. If the difference between I_i^c and I^m_{hi} is greater than the threshold (can be chosen as the distance between the peak corresponding to the road and the nearest dip in the histogram), the pixel is in the motion regions. Else it is removed. Search the next one in the connected region C'_i. Go

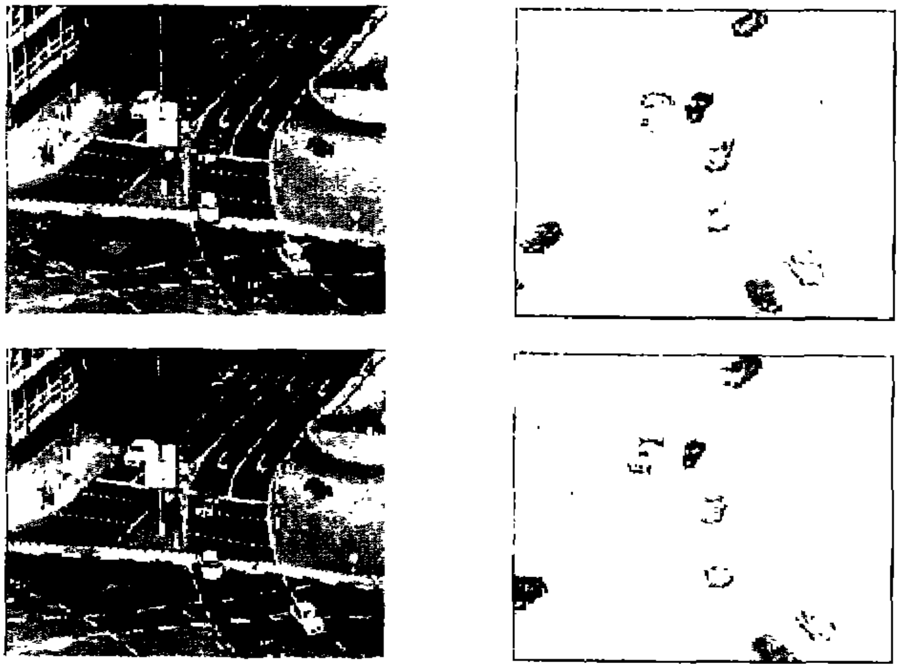


Fig. 4 The segmentation results of sequence 1
图 4　序列 (1) 的分割结果

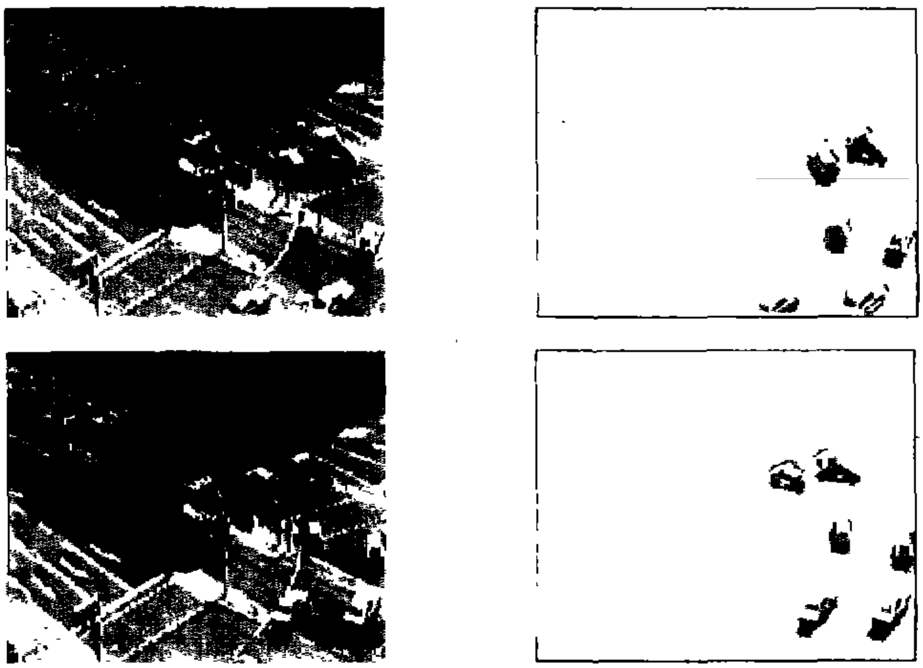


Fig. 5 The segmentation results of sequence 2
图 5　序列 (2) 的分割结果

to (a).

(c) No. Search the next pixel in the connected region C'_r. Go to (a).

The final segmentation results are shown in Figs. 4&5. Each shows the segmentation results of the two different frames in one sequence. In the left is the image after ecnhancement. In the right are the segmentation results. For most vehicles the results have been very close to their true shapes. All vehicles can be located correctly. But the result for the vehicle labeled as #1 in Fig. 5 is not perfect. There exists a large shadow region on one side of the vehicle, where edge characteristic is not distinct. This case will be studied further.

4 Conclusion

In this contribution a method integrating motion and static segmentation is presented for traffic surveillance. It is applicable to the case of multiple vehicles. The background can be arbitrarily complicated and contain many nonmoving objects. The result can be used to monitor the traffic situations and can be used as the basis of further classification and recognition.

REFERENCES

[1]Lipton A, Fujiyoshi H. Moving target classification and tracking from real-time video, *IEEE Workshop on Applications of Computer Vision*, Princeton NJ, Oct. 1998, 8—14

[2]Kilger M. A shadow handler in a video-based real-time traffic monitoring system, *IEEE Workshop Application Computer Vision*, Palm Spring, CA, 1992, 1060—1066

[3]Bouthemy P, Lalande P. Recovery of moving object masks in an image sequence using local spatiotemporal contextual information, *Optical Engineering*, 1993, **32** (6): 1205—1212

[4]Koller D, Daniilidis K, Nagel H H. Model-based object tracking in monocular image sequences of road traffic scenes, *International Journal of Computer Vision*, 1993, **10**(3): 257—281

[5]Dubuisson M P, Jain A K. Contour extraction of moving objects in complex outdoor scenes, *International Journal of Computer Vision*, 1995, **14**: 83—105

[6]Aach T, Kaup A, Master R. Statistical model-based change detection in moving video, *Signal Processing*, 1993, **31**: 165—180

[7]Korn G A, Korn T M. *Mathematical Handbook for Scientists and Engineers*. New York: McGrawHill, 1968

[8]Felzenszwalb P F, Huttenlocher D P. Image segmentation using local variation, *IEEE Conf. Comput. Vis. Pattern Recognition*, 1998, 98—104