Automatic Road Detection for Highway Surveillance Using Frequency-Domain Information

Qing-Jie Kong
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
qjkong.cas@gmail.com

Lucidus Zhou, Gang Xiong, Fenghua Zhu
Dongguan Research Institute of CASIA
Cloud Computing Center, Chinese Academy of Sciences
Dongguan, China
{fenghua.zhu; gang.xiong}@ia.ac.cn

***Abstract* —Video detection is one of the primary collection means of traffic states in Parallel traffic Management Systems (PtMS). In order to accurately and automatically obtain road areas in highway surveillance videos, this paper presents an automatic detection algorithm based on the frequency-domain information of video images. This algorithm uses the frequency-domain feature that is produced by the vehicles passing through road areas in videos, to realize automatic segmentation and recognition of the road areas. The experiment comparing with the traditional vehicle-tracking-based method, which uses the information in the time-space domain, illustrates the advantages of the proposed algorithm.**

Keywords—road detection; visual surveillance; frequency information

I. Introduction

Recently, Parallel traffic Management Systems (PtMS) [1], [2] has been verified to be an effective and better solution for management and control of complex urban transportation systems, by its successful applications in various cities in China, such as Suzhou and Guangzhou [3], [4]. As an essential component, traffic sensing, control and management systems (aDAPTS) are in charge of providing real-time traffic states for the PtMS. In aDAPTS, video detection is one of the primary means.

As the traffic video detection bears characteristics such as a huge number of video detectors and long collection time, intelligent and automatic detection is the necessary function. Currently, in the studies on traffic visual surveillance systems, automatic vehicle detection, tracking and recognition have always been popular research topics, and the related monitoring products gradually emerge as well [5]. However, most of the existing surveillance products are employed for specific monitoring locations, where cameras are specifically installed and calibrated. This process greatly reduces the efficiency of the installation of surveillance cameras. Furthermore, once the camera is damaged and replaced or the camera angle changes, it needs to be re-calibrated, which also increases the cost of maintenance. Therefore, if the computer can automatically identify the road area from the surveillance video, the manual calibration in advance can be eliminated, thereby reducing the cost of manpower and time to install monitoring equipments, and also reducing the workload of human-computer interaction in the monitoring process.

For the problem of automatic road detection, Morris and Trivedi [6] used the vehicle tracking method to automatically obtain the vehicle moving area. However, this method processes the image signal in the spatio-temporal domain, which is susceptible to interference of the noise signal, so that it is difficult to obtain accurate and complete road areas.

Compared to the methods based on the information in the time-space domain, the methods using the frequency-domain information have better accuracy and robustness in certain respects, and thereby have been widely used in object segmentation [7], [8] and pedestrian detection [9]. Therefore, this paper presents the method based on the frequency-domain information to automatically detect the road area in the highway surveillance videos. This method uses the frequency-domain characteristics generated by the vehicles passing through road areas in videos, to achieve the automatic segmentation and recognition of the road areas.

Research supported in part by National Natural Science Foundation of China grants 61104160, 61174172, 61233001, and 71232006.

II. Automatic Road Detection Method

In this section, the automatic road detection method will be presented in 4 steps in accordance with the order of execution. They are Gaussian smoothing, pixel matrix generation, frequency image generation, and automatic road area recognition. In this method, the pixel value of point (x, y) in frame t is recorded as $I(x, y, t)$.

A. Gaussian Smoothing

As the video images shot by surveillance cameras would exit the high-frequency interference, which can have big impact on the following processing steps, each frame, after read, is first smoothed by the Gaussian function as follows:

$$\mathrm{G}(x,y)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{1}$$

which means the intensity of pixels for later use will be:

$$I'(x, y, t) = \sum_{a=-1}^{1} \sum_{b=-1}^{1} \mathrm{G}(a, b) \cdot I(x-a, y-b, t) \,. \qquad (2)$$

Gaussian smoothing is an operation to convolute the video image with the Gaussian function.

B. Pixel Matrix Generation

Logically, the frames are stored in memory as a 3-dimensional matrix, and the vector with fixed x, y pair $p(x_0, y_0)$ holds the historic intensity values of that pixel. It represents the temporal information of the pixel. For instance, for a surveillance video shown as Figure 1, Figure 2 shows the temporal values of a column of pixels through the road area in the video image.

Fig. 1. An image frame of the surveillance video shot from a California Highway in USA.

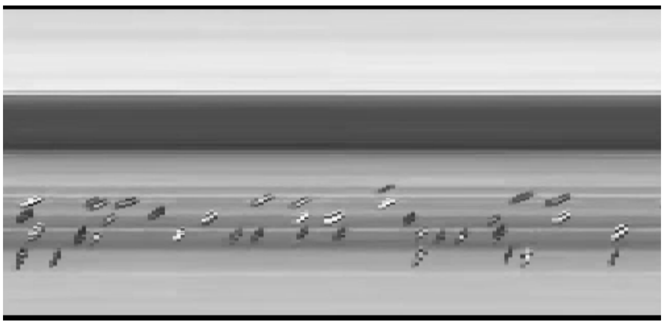

Fig. 2. Temporal image of a column of pixels.

C. Frequency Image Generation

First, given $I_p[t] = I'(x_0, y_0, t)$, where p is the point (x_0, y_0), we perform a discrete Fourier transform (DFT) on it and get:

$$F[u] = \frac{1}{N} \sum_{t=0}^{N-1} I_p[t] \cdot e^{-j2\pi ut/N}, \; u = 0, 1, 2, 3, \cdots, N\text{-}1. \qquad (3)$$

which is the spectrum of the intensity values on the respective pixel. The brackets mean it's a series. From this series, the power of the components can be computed by:

$$P[u] = \left|F[u]\right|^2, \; u = 0, 1, 2, 3, \cdots, N\text{-}1. \qquad (4)$$

Figure 3 is the result of performing the DFT on the pixel series, the same column of pixels with Fig. 2. In Fig. 3, the pixel intensities of the highest to lowest correspond to the frequencies of the high to low, respectively.

From this result, we can clearly see that the series of pixel values with no significant fluctuation, which correspond to non-road area, tend to have few high frequency components. On the contrary, the pixel series in the road area bear lots of middle and high frequency components.

Fig. 3. DFT results of the same pixel column with Fig. 1.

Therefore, all the low-frequency components of a pixel series are ignored. From the remaining spectrum, the highest single-frequency power (SFP) is selected by:

$$m(x_0, y_0) = \max\{P[u] \mid u > u_{thr}\}, \; u = 0, 1, 2, 3, \ldots, N\text{-}1. \; (5)$$

where u_{thr} is the low threshold to the eligible frequencies.

Then, each highest SFP is mapped into a visible intensity range and written back to the position of the pixel in the original video image. Finally, a transformed max SFP image is produced, as Figure 4 shows.

Fig. 4. Transformed max SFP image.

D. Automatic Road Area Recognition

From Figure 4, the road area can already be seen with the eyes. In this subsection, we will realize the automatic recognition of road areas, i.e. find the boundaries of road areas automatically. The recognition algorithm includes the following 4 steps, and the recognition result on the road example in Fig. 4 is shown in Figure 5.

1) Preprocessing:

Before the recognition, the SFP image is first binarized, and then a round of erosion and dilation is gone through the binarized image to eliminate outliers.

2) Definition of the detection line

Since the pixel values in the road area are 1 and those outside of it are -1, the detection line is defined by equation (6), just as Figure 6 shows.

Fig. 5. Recognition result on Fig. 4.

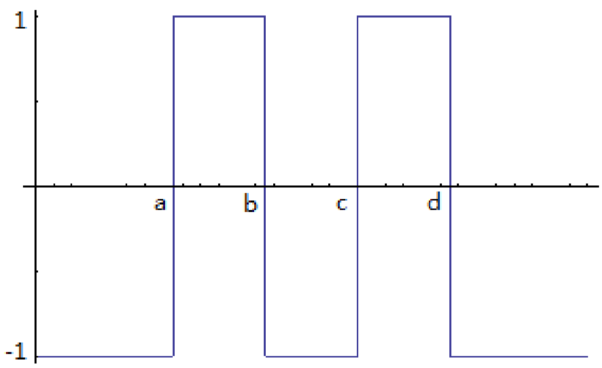

Fig. 6. Function of the detection line.

$$D[x] = \begin{cases} -1, & x < a \text{ or } b < x < c \text{ or } x > d \\ 1, & a \le x \le b \text{ or } c \le x \le d \end{cases} . \quad (6)$$

where a, b, c, and d are parameters to be determined.

3) Definition of the correlation

Denoting the line signal from the binarized image as $L[x]$, the correlation of the line from the image and the detection line is calculated by the following equation.

$$r_{L,D} = \sum_{x=1}^{w} L[x] \cdot D[x] . \quad (7)$$

where w is the width of the image.

4) Function fitting

The purpose of the function fitting is to determine the parameters of the detection function (i.e., a, b, c, and d) determined by searching the maximum correlation, so as to find the boundaries of road areas. The fitting procedure includes:

- First, line by line fitting upwards from the lowermost row of the image, to find the parameters of the best fitting;
- Second, line by line fitting starts from the middle column of the image to the two sides;

Finally, smooth the boundaries.

III. EXPERIMENTS

The data for the experiment are the video footages collected from the California Highways of the United States. Every footage contains at least 320 effective frames, and no accident or obvious violation of traffic rules can be observed. The programming tools are MinGW and OpenCV, and all of them are performed on a Windows PC with stock hardware.

For the sake of comparison, the traditional method based on the Gaussian mixture model (MOG) [10] is devised using the function provided by OpenCV. In this method, the road detection is realized by the vehicle tracking in the spatial and temporal domain.

The results of the comparison experiment are shown in Figure 7. From Fig. 7, it can be observed that the proposed method is hard to be disturbed by the other moving objects in the background region, such as the cloud in the sky (e.g., the situation in the 1st row). Furthermore, it can be robust to different road forms.

IV. CONCLUSION

This paper has proposed an automatic method to detect road for highway surveillances. This method takes advantages of the video information in the frequency domain to obtain better detection performance than that of the traditional MOG-based method. The proposed method bears high efficiency and good robustness, so that it is promising to be applied into the engineering practice.

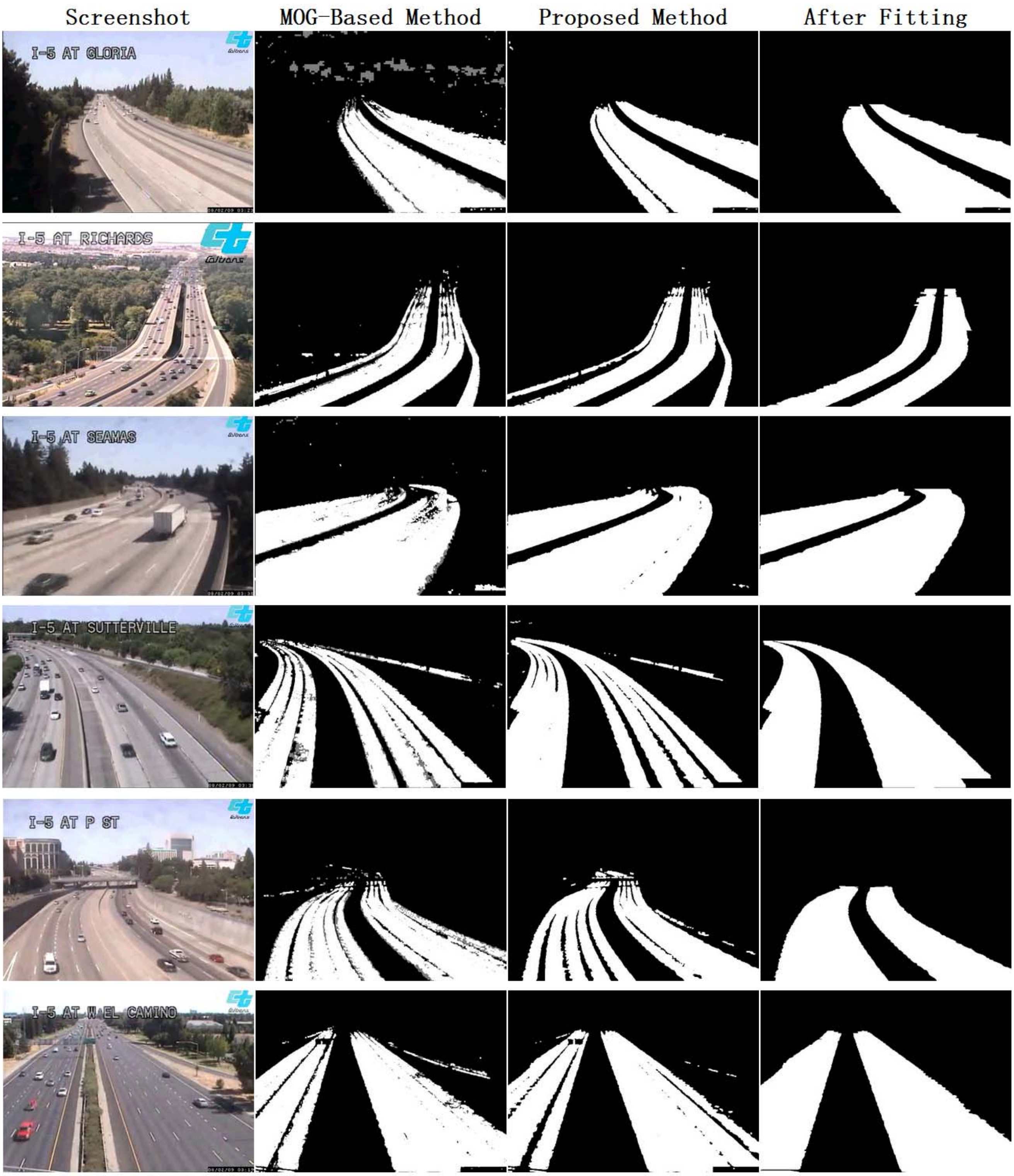

Fig. 7. Results of the comparison experiment.

REFERENCES

[1] F.-Y. Wang, "Toward a revolution in transportation operations: AI for complex systems," *IEEE Intell. Syst.*, vol. 23, no. 6, 2008, pp. 8–13.

[2] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, 2010, pp. 630–638.

[3] G. Xiong, K. Wang, F. Zhu, C. Chen, X. An, and Z. Xie, "Parallel traffic management for the 2010 Asian games," *IEEE Intell. Syst.*, vol. 25, no. 3, 2010, pp. 81–85.

[4] G. Xiong, S. Liu, X. Dong, F. Zhu, B. Hu, D. Fan, and Z. Zhang, "Parallel traffic management system helps 16th Asian games," *IEEE Intell. Syst.*, vol. 27, no. 3, 2012, pp. 2–6.

[5] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920-939, 2011.

[6] B. T. Morris and M. M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425-437, 2008.

[7] O. Azy and N. Ahuja, "Segmentation of periodically moving objects," in *Proc. 19th Int. Conf. Pattern Recog.*, Tampa, Florida, 2008.

[8] Q.-J. Kong, A. Kumar, N. Ahuja, and Y. Liu, "Robust segmentation of containers for intelligent train monitoring systems," in *Proc. IEEE Workshop Appl. Comput. Vision*, Snowbird, Utah, 2009, pp. 1-6.

[9] Y. Ran, I. Weiss, Q. Zheng, *et al*., "Pedestrian detection via periodic motion analysis," *Int. J. Comput. Vision*, vol. 71, no. 2, pp. 143-160, 2007.

[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Fort Collins, Colorado, 1999, pp. 246-252.