Development of Image Sequences Based Traffic Incident Detection System for Urban Intersection

Guangyi SHI[1], Yuexian ZOU[1*], Yiyan WANG[1], Hang SHI[1] and Qiang PENG[2]

[1]The Key Laboratory of Integrated Microsystems, Shenzhen Graduate School of Peking University
[2]School of Information Science & Technology, Southwest Jiaotong University, Chengdu, P.R. China

*Contact Author: zouyx@szpku.edu.cn

Abstract

***Abstract*—Traffic incident detection is one of the most important issues for intelligent transportation systems (ITS), especially in urban area which is full of signaled intersections. This paper presents the development of a novel traffic incident detection system based on image signal processing, feature extraction algorithms, and hidden Markov model (HMM) classifier. First, a traffic surveillance system was set up at a typical intersection of china, traffic videos were recorded and image sequences were extracted for image database forming. Second, several features extraction algorithms were used and compared. Finally, HMM was used for classification of traffic signal logics (East-West, West-East, South-North, North-South) and accident of crash. Feature generation with DCT-FFT process gives the best result with total correct rate of 91% and incident recognition rate of 95%.**

Keywords — ITS, Incident Detection, HMM, Intersection

I. INTRODUCTION

Intersections come to be common places for traffic incidents, which may be due to the fact that there are several conflicting movements, as well as different intersection design characteristics [1-2]. The conditions of breaking the traffic regulations, e.g. illegal stop vehicles, converse driving, etc. Intersections also tend to experience severe congestion due to the several types of injurious crashes, furthermore, it will increase the chance of a second accident because of the lane changing to avoid the stopped vehicles. Therefore, accurate and prompt detection of accidents at intersections offers tremendous benefits of saving properties and lives and minimizing congestion and delay.

Typically, an efficient automatic incident detection (AID) system is composed of two main components: a traffic detection system and an incident detection algorithm. The traffic detection system provides the traffic data necessary for detecting an incident while the incident detection algorithm analyses these data and makes the judgment of the presence or absence of incidents. The very first automatic traffic detection system is made of inductive loop detectors embedded in the road pavement [3], followed with other kinds of spot sensors such as microwave and magnetic sensors, are all capable of providing presence, occupancy, speed, in nonintrusive way. A number of event detection algorithms have been developed from a variety of theoretical foundations utilizing various kinds of sensors [4-7].

In recent years, much more image sensors has been used in traffic surveillance systems, result in many research works concerning with image processing and computer vision techniques as solutions for AID systems. Ikeda et al. present [8] outline an automatic abnormal incident detection system, which could detect four types of incidents including stopped vehicles, slow vehicles, fallen objects and successive changing vehicles. Ching-Po Lin etc. [9] presented a traffic monitoring system based on real-time visual tracking techniques. Active contour model based tracking algorithms is realized in their project, and the stand-along image tracker can extract traffic parameters and detect car accident from image sequences. However, these methods have rather limited capability to detect accidents at an intersection, which has a very complicated scene and background. Weimin Hu et al. [10] reported a method of traffic accident prediction using 3-D model-based vehicle tracking. A fuzzy self-organizing neural network algorithm is then applied to learn activity patterns from the sample trajectories. Vehicle activity is predicted by locating and matching each partial trajectory with the learned activity patterns, and the occurrence probability of a traffic accident is determined. Therefore, it requires detailed geometric data of vehicles, and the computing complexity is high.

Yong-Kul Ki etc. [11] made their efforts to traffic accident recording and reporting model at intersections. This model first extracts the vehicles from the video images, tracks the moving vehicles and extracts features such as variation rate of velocity, position, area and direction. Based on these features, the model can make decisions on the traffic accident. The work is remarkable with correct detection rat of 50% and a detection rate of 60%. Shunsuke Kamijo group in Tokyo Univ. [12] made a successful traffic incident detection system based on hidden Markov model (HMM). The system learns various event behavior patterns of each vehicle in the HMM chains and using the output from the tracking system to identify current event chains. The system can recognize bumping, passing, and jamming. These methods which utilize the individual vehicle features are relative effect and partially in real-time. However, in

This project is partially supported by the Chinese National 863 Program (2007AA11Z224) and Shenzhen Science and Technology Program (SZKJ-200716)

China, when a lot of vehicles and pedestrians appear in the detection zone, the tracking algorithm will consume much more computing resource, which will result in the computing collapse.

In our previous work [13], traffic image sequences in a whole were studied, image difference and FFT were used for the feature generation. HMM was used as the training model for the classification of traffic signal logics (East-West, West-East, South-North, North-South) and accident of crash, where the recognition result is 74%. In this paper, several figure extraction algorithms were used including principal component analysis (PCA), discrete cosine transform (DCT) and fast Fourier transform (FFT). With same HMM training process, DCT-FFT based features have best classification result with correct rate of 91%. This paper is organized as follows. In section 2, the vehicle detection system at intersection will be described. Section 3 will focus on the feature generation algorithms for incident detection as well as HMM training process. Experimental result will be presented in section 4. Conclusion will be drawn in section 5.

II. VEHICLE DETECTION SYSTEM AT INTERSECTION

A. *Camera Installation (CI)*

Currently, vehicle detection system, deployed at intersections, usually installed the cameras at roadside poles or traffic light poles. It can only view a part of intersection and collect part of video information. Therefore, it is required to use at least more than one camera to cover an entire intersection. Moreover, the video information, collected by this installation, was easily affected by vertical and lateral viewing angles, number of lanes observed and image quality, and had the serious problem of vehicle occlusion, directly affect the result of vehicle detection and segmentation. Some research group presented vehicle detection and tracking system, and utilized a single camera which installs a fisheye lens or wide view-angle lens, to cover an entire intersection. However, the video image information collected by this method existed geometric distortion and required to calibrate before vehicle detecting and tracking, and therefore, the real-time performance of system was poor. In addition, it also had the serious problem of vehicle occlusion.

In our vehicle detection system, the camera is installed at the roof of a high building beside the intersection, and it provides the full coverage of the intersection. The actual traffic image can be seen in Figure 1.

B. *Detection Region of Interest (DROI)*

In the original video image captured by camera, it includes centre region, crosswalk region and vehicle lane region. In this paper, our proposed system aims to detect vehicles and traffic status or accident in the centre region. Therefore, we define the central area as detection region of interest, and extract the interesting frame from original video image. The advantages of it are: 1) Achieving a real-time performance for the system; 2) Eliminating the effect of pedestrian on crosswalk and saving of the computational complexity.

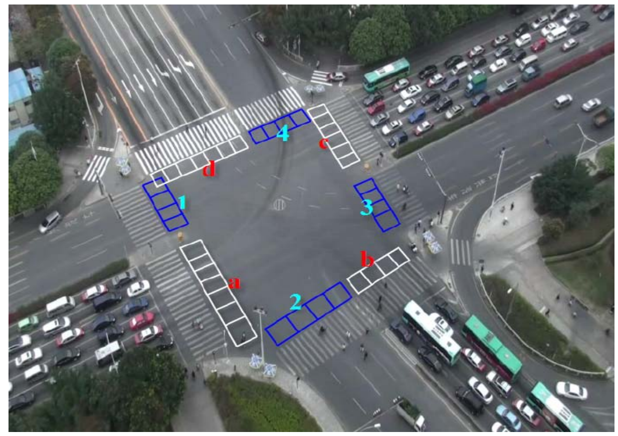

Figure 1. Actual traffic image gathered from top building around a typical urban intersection.

III. INCIDENT DETECTION BASED ON HMM

For the purpose of incident detecting at intersection, we proposed an algorithm based on hidden Markov model, which can realize multi-classification for different traffic scenes. The process mainly consists of three parts: ROI setup, image compression and feature extraction, and HMM training process and testing experiments. The region of interest is designed to generate the area that contains most information. It will also eliminate the zebra crossing and useless background like plants and irrelevant constructions, which will bring disturbance to the feature generation. The feature extraction from image sequences is very important since it will determine the classification result a lot. As good features will lead to higher incident detection rate, as well as fasten the feature generation speed for the implementation for real-time system. Finally, a good classifier should be chosen for the AID system, HMM was selected since it is a good multi-classifier for special-temporal input. The block diagram of our method is shown in Figure 2.

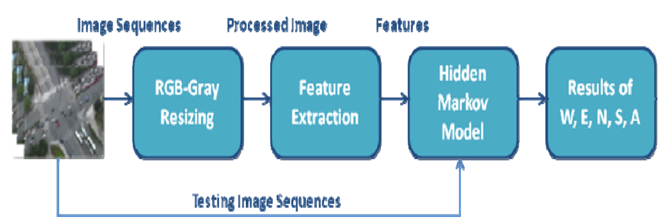

Figure 2. Block diagram of image processing and HMM training process.

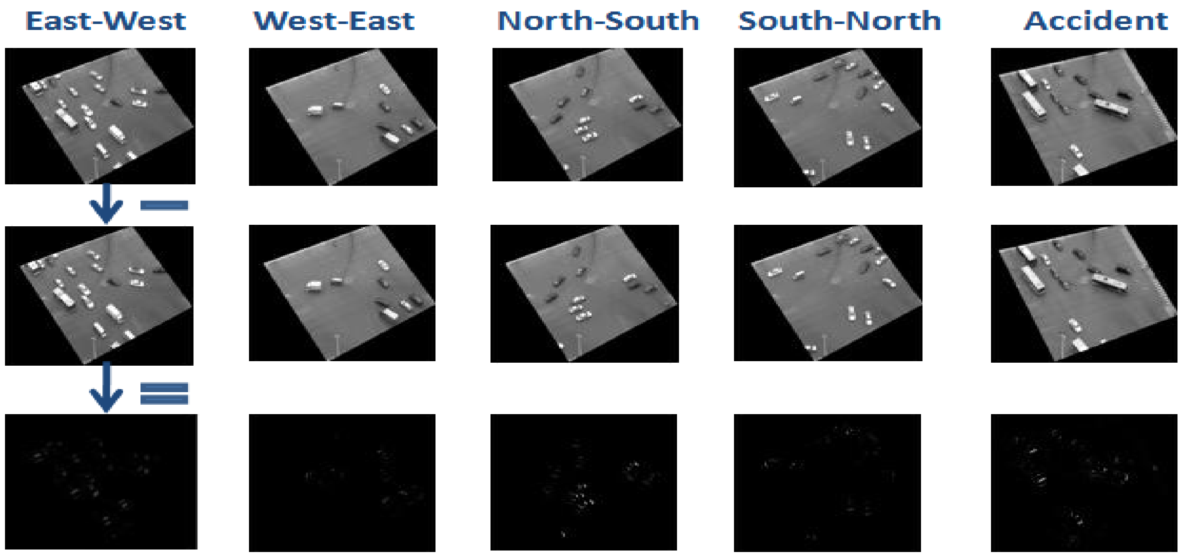

Figure 3. Image difference for the feature extraction with spatio-temporal information.

A. Database Construction and Feature Extraction

For a typical intersection, there are mainly four typical signaled movement, West-east, East-west, North-South, and South-North, which includes left-turn and U-turn for each moving condition. Accident images will be marked as the fifth condition, which should include crash, illegal adverse drive, illegal stopped vehicles and so on. Since there are only several accidents happened in the traffic videos we recorded, we only make use of these data for the classification in this paper, other conditions will be added in later research work. Therefore, five types of traffic signaled movements based on image sequences are constructed. We sampled 100 images for each condition for HMM training and 100 other images each for testing, totally 1,000 images.

In our project, the videos were recorded with HD mode, each image has the resolution better than D1 standard. If we input the data directly into HMM training process, it will cause a dimension disaster. We performed the image processing steps as: 1) Transfer the RGB images to grey ones; 2) Resize the images to pixel of 100*100; 3) Add one mask to the ROI and extract central area for further analysis; 4) Perform the image difference of two consecutive video sequences, which include the temporal information of moving vehicles as well as spatial information of the moving trajectory.

B. PCA for Feature Extraction

PCA can be used for dimensionality reduction in a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data [14]. It is found that PCA is closely related to singular value decomposition (SVD). In practice, we used SVD for the image decomposition.

Let $\mathbf{X}$ be an arbitrary $n \times m$ matrix and $\mathbf{X}^T\mathbf{X}$ be a rank r, square, symmetric $n \times n$ matrix. Let us define all of the quantities of interest.

$\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_r\}$ is the set of orthonormal $m \times 1$ eigenvectors with associated eigenvalues $\{\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_r\}$ for the symmetric matrix $\mathbf{X}^T\mathbf{X}$.

$$(\mathbf{X}^T\mathbf{X})\hat{\mathbf{v}}_i = \lambda_i \hat{\mathbf{v}}_i \tag{1}$$

$\sigma_i \equiv \sqrt{\lambda_i}$ are positive real and termed the singular values. $\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_r\}$ is the set of orthonormal $n \times 1$ vectors defined by $\hat{u}_i \equiv \frac{1}{\sigma_i}\mathbf{X}\hat{\mathbf{v}}_i$. We have the properties of $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j \equiv \sigma_{ij}$ and $\|\mathbf{X}\hat{\mathbf{v}}_i\| = \sigma_i$

We now have all of the pieces to construct the decomposition. The "value" version of singular value decomposition is just a restatement of the third definition.

$$\mathbf{X}\hat{\mathbf{v}}_i = \sigma_i \hat{\mathbf{u}}_i \tag{2}$$

We start by constructing a new diagonal matrix Σ.

$$\mathbf{\Sigma} \equiv \begin{bmatrix} \sigma_{\tilde{1}} & & & & \\ & \ddots & & \mathbf{0} & \\ & & \sigma_{\tilde{r}} & & \\ & \mathbf{0} & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \tag{3}$$

where $\sigma_{\tilde{1}} \geq \sigma_{\tilde{2}} \geq \ldots \sigma_{\tilde{r}}\}$ are the rank-ordered set of singular values. Likewise we construct accompanying orthogonal matrices $\mathbf{V}$ and $\mathbf{U}$.

$$\mathbf{V} = \{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, ..., \hat{\mathbf{v}}_m\} \qquad \mathbf{U} = \{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_m\}$$

where we have appended an additional (*m-r*) and (n-*r*) orthonormal vectors to "fill up" the matrices for **V** and **U** respectively. Because **V** is orthogonal, we arrive at the form of decomposition.

$$\mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T \tag{4}$$

C. DCT and FFT

The DCT is often used in signal and image processing, especially for lossy data compression, because it has a strong "energy compaction" property (Rao and Yip, 1990). We make use of DCT in our difference images for better classification.

Typically, the N real numbers $\mathbf{x_0, ..., x_{N-1}}$ are transformed into the N real numbers $\mathbf{X_0, ..., X_{N-1}}$ according to one of DCT formulas. The commonly used expression of DCT is:

$$X_k = \sum_{n=0}^{N-1} \mathbf{x}_n \cos[\frac{\pi}{N}(n + \frac{1}{2})k] \tag{5}$$

$$k = 0, ..., N-1$$

FFT is widely used for frequency domain processing and spectrum analysis. It is a computationally efficient discrete Fourier transform (DFT), which is defined as

$$X(k) = \sum_{N=0}^{N-1} X_n W_N^{kn}, \ k = 0, ..., \mathrm{N-1}, \tag{6}$$

where the twiddle factor is defined as

$$\mathrm{W}_{\mathrm{N}}^{\mathrm{kn}} = \mathrm{e}^{-2\mathrm{j}\,\pi n k / N} \tag{7}$$

D. Hidden Markov Model

A Hidden Markov Model derives from Markov Chains, which is a doubly embedded stochastic process with an underlying stochastic process that is not observable, and can only be observed through another set of stochastic processes that produce the sequence of observations. An N-state $(S_1, S_2, ... S_N)$ of HMM can be described as $\lambda = (\pi, A, B)$. The parameters are as follows.

The initial state distribution,

$$\pi = \pi_1, \pi_2, ... \pi_N \tag{8}$$

which is used to describe the distribution of the state at time $t = 0$, and $\sum_{i=1}^{N} \pi_i = 1$.

The state transition probability distribution is:

$$A = \{a_{i,j}\}_{N \times N}, \quad i, j = 1, 2, ... N \tag{9}$$

$$a_{i,j} = P(q_j, t+1 | q_i, t), \quad 1 \le i, j \le N \tag{10}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{11}$$

In the special case where any state can reach any other in a single step, $a_{ij} > 0$ for any *i, j*. For other types of HMMs, $a_{ij} = 0$ for one or more (i, j) pairs.

The observation symbol probability distribution in state j can be described as $B = b_j(k)$, where

$$B = \{b_{jk}\}_{N \times N}, \quad k = 1, 2, ..., M \tag{12}$$

N denotes the hidden states, and M denotes the length of the codebook (discrete observation density). For continuous observation densities, it is $B = b_i(x)$.

Associating with the distribution of *B*, HMMs are categorized into Discrete Hidden Markov Model (DHMM with discrete probability density) and Continuous Hidden Markov Model (CHMM, with continuous probability density). Using these parameters, testing vectors can be interpreted properly. A testing vector is classified into the Model with the highest Maximum *A* Posteriori.

IV. Experimental Result

A. Traffic Movement Modeling

We trained five 6-state left-right HMMs of the five traffic movements mentioned above. All HMM parameters including two model parameters (M and N) and three probability matrixes A, B, and π were initialized using a uniform segmentation for each training sequence. The Baum-Welch algorithm was used to iteratively re-estimate the parameters according to the forward and backward variables. 20 iterations were run for each training process. Fig. 4 shows all 5 HMMs forward scores of the results for the Forward-Backward procedure, denoted as the log-likelihood versus the learning iteration index. After the learning iterations, 5 HMMs representing the five traffic conditions are retrieved based on the learning data samples.

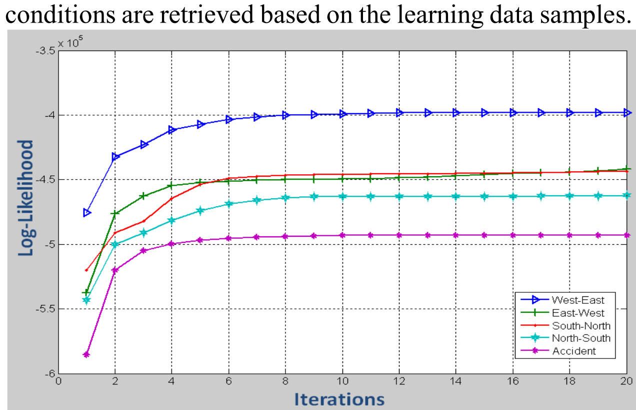

Figure 4. Log-likelihood versus learning iteration index in training process.

B. Model to Model Similarity Evaluation

In order to evaluate the 5 trained motion models (λ_1-λ_5), each model randomly generated 1 observation sequences which are as the same size as the ones for the training and testing targets. We applied those generated observation sequences O_i as the test data to all trained models λ_j ($j = 1, 2, 3, 4, 5$) one after another. The probability results $P(O_i | \lambda_j)$

are derived in the form of log-likelihood shown in table I. It is obviously to find that the log-likelihood reaches the maximum marked with shadow, when $i = j$. That is to say, each generated observation sequence has much more similarity with the model generating it than all the other models.

TABLE I.
Log-Likelihood of randomly generated sequences to the trained HMMs

Log-Likelihood	**W-E**	**E-W**	**W-N**	**N-S**	**Accident**
GS #1-1	-0.1501	-0.2327	-0.2181	-0.2102	-0.2359
GS #2-1	-0.3097	-0.2688	-0.3508	-0.6208	-0.2732
GS #3-1	-0.2546	-0.3096	-0.1783	-0.3570	-0.2716
GS #4-1	-0.2396	-0.2638	-0.2567	-0.2014	-0.2508
GS #5-1	-0.3427	-0.3281	-0.3255	-0.6362	-0.2767

C. *Traffic Incident Detection*

As introduced before, we have extracted data of five traffic signaled conditions, each have 200 samples. We divided the samples into two groups. One is for HMM training process, the other is for testing. Table shows the recognition result with different features that generated from PCA, FFT and DCT-FFT process. Therefore, DCT-FFT features have best total correct rates of 91%. As for the traffic incident detection, the correct rate is 95%.

TABLE II.
Recognition result with different generated features

Iterations	**W-E**	**E-W**	**W-N**	**N-S**	**Accident**
PCA	28/100	86/100	22/100	53/100	62/100
FFT	66/100	61/100	78/100	83/100	84/100
DCT-FFT	100/100	68/100	94/100	100/100	95/100

V. CONCLUSION

The development of a novel image sequence-based traffic incident detection system using feature extraction algorithms and HMM classifier is presented in this paper. A HD camera was installed at the roof of a high building beside the intersection as our traffic detection system, and it provides the full coverage of the intersection. Based on the image sequences, database was setup including four signaled logic movements at intersection and one condition of crash. Effective features were extracted from original videos using image processing and different compression method including PCA, FFT and DCT. Through HMM training process, the five traffic conditions can be recognized from each other with correct rates range from 68% to 100%, with incident detection correct rate of 95%.

REFERENCES

[1] The Ministry of Public Security of China: http://www.mps.gov.cn/English/index.htm.

[2] Yong-Kul Ki, and Dong-Yong Lee, "A traffic accident recording and reporting model at intersections," IEEE Transactions on Intelligent Transportation Systems, Vol. 9, NO. 2, June 2007.

[3] Dia H., and Rose G., "Development and evaluation of neural network freeway incident detection models using field data," Transp. Res. C, Emerg. Technol., 1996, 5, pp. 313–331.

[4] Ashkan Sharafsaleh, and Ching-Yao Chan, "Experienmtal evaluation of commercially-off-the-shelf sensors for intersection decision support systems," 12th World Congress on ITS, November 6-10, 2005, Paper No. 1925.

[5] Dougherty M., "A review of neural networks applied to transport," Transp. Res. C, Emerg. Technol., 1995, 3, pp. 247–260.

[6] Seung-Heon Lee, Jin-Woo Choi, and Nam-Kwan Hong, "Development of incident detection model using neuro-fuzzy algorithm," Proceedings of the Fouth Annual ACIS, International Conference on Computer and Infromation Science (ICIS'05).

[7] J. Owens, and A. Hunter, "Application of the self-organizing map to trajectory classification," in Proc. IEEE Workshop Visual Surveillance, Dublin, Ireland, 2000, pp. 77–83.

[8] H. Ikeda, T. Matsuo, Y. Kaneko, and K. Tsuji, "Abnormal incident detection system employing image processing technology," in Proc. IEEE Int. Conf. Intell. Transp. Syst., Tokyo, Japan, Oct. 1999, pp. 748–752.

[9] Ching-Po Lin, Jen-Chao Tai, and Kai-Tai Song, "Traffic Monitoring based on real-time image tracking," Proceedings of the 2003 IEEE International Conference on Robotics and Automation, Taipei, Taiwan, September 14-19, 2003.

[10] Weiming Hu, Xuejuan Xiao, Dan Xie, and Tieniu Tan, "Traffic accident prediction using 3-D model-based vehicle tracking," IEEE Transaction on Vehicular Technology, Vol. 53, NO. 3, May, 2004.

[11] Yong-Kul Ki, and Dong-Young Lee, "A traffic accident recording and reporting model at intersections," IEEE Transactions on Intelligent Transportation Systems, Vol. 8, NO. 2, June 2007.

[12] Shunsuke Kamijo, Yasuyuki Matsushita, Katsushi Ikeuchi, and Masao Sakauchi, "Traffic monitoring and accident detection at intersections," IEEE Transactions on Intelligent Transportation Systems, Vol. 1, NO. 2, June 2000.

[13] Yuexian ZOU, Guangyi SHI, Hang SHI, and Yiyan WANG, "Image sequences based traffic incident detection for signaled intersections using HMM," IEEE International Conference on Hybrid Intelligent Systems 2009, HIS 2009, August 12-14th 2009, Shenyang, China.

[14] www.wikipedia.org.