A Flow Analysis Framework for Traffic Video

Lu-Shuang Bai* Ying Xia** Sang-Chul Lee***

Abstract The fast progress on multimedia data acquisition technologies has enabled collecting vast amount of videos in real time. Although the amount of information gathered from these videos could be high in terms of quantity and quality, the use of the collected data is very limited typically by human-centric monitoring systems. In this paper, we propose a framework for analyzing long traffic video using series of content-based analyses tools. Our framework suggests a method to integrate theses analyses tools to extract highly informative features specific to a traffic video analysis. Our analytical framework provides (1) re-sampling tools for efficient and precise analysis, (2) foreground extraction methods for unbiased traffic flow analysis, (3) frame property analyses tools using variety of frame characteristics including brightness, entropy, Harris corners, and variance of traffic flow, and (4) a visualization tool that summarizes the entire video sequence and automatically highlight a collection of frames based on some metrics defined by semi-automated or fully automated techniques. Based on the proposed framework, we developed an automated traffic flow analysis system, and in our experiments, we show results from two example traffic videos taken from different monitoring angles.

Keywords : Content-based Analysis, Flow Analysis Framework, Traffic Video

1. Introduction

During recent years, rapid advancement of visual material application urgently need a corresponding growth in the need for automatically extracting key information from images and videos, which is the basis of scene analysis, indexing and content based retrieval. Video is the image sequence which has strict order in succession and there is strong pertinence between neighbor frames. Shot, the important element of video, is defined as series of interrelated consecutive frames representing a continuous action in time and space, which is contiguously taken by a single camera[1].

The automatic partitioning of video into shots involves the detection of shot transitions. A transition is the boundary between two shots. There are two types of transitions: abrupt and gradual. An abrupt transition is an instantaneous change from one shot to another, with no transitional frames between two shots, which is also known as a cut, which is an abrupt shot change occurs in a single frame. For gradual transition, fade, dissolve and wipe are three common types of it. Fade is a gradual diminishing or heightening of visual intensity, which always named as fade out and fade in. Dissolve is a combination of fade out and fade in with some overlap. Wipe is a transition from one scene to another wherein the new scene is revealed by a moving boundary. The wipe transition can be either horizontal, vertical, diagonal or some other fancy types[2]. The detection of gradual transition is more difficult than that of abrupt cut, which need distinguish the operation of camera or objects movement.

Different with the ordinary videos, traffic videos have their own marked characteristics. Traffic data collection is an important issue for road traffic control, which has some requirements for real time information and traffic parameter estimation, such as road traffic intensity, lane occupancy, congestion level, estimation of journey time, etc[3].

Much effort has been made to meet the need of

† This research was supported by a grant(07KLSGC05) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Construction & Transportation of Korean Government.

* MS Student, Sino-Korea GIS Research Center, Chongqing University of Posts and Telecommunications, bails105@163.com
** Professor, Sino-Korea GIS Research Center, Chongqing University of Posts and Telecommunications, xiaying@163.com
*** Professor, Department of Computer and Information Engineering, Inha University, sclee@inha.ac.kr(corresponding author)

Manuscript received : 2009.05.12 Revised : 2009.06.19 Accepted : 2009.06.20

traffic video analysis, and many traffic surveillance systems have been built for application, some of which have good results[15,16]. In this paper, we propose a framework for analyzing long traffic video using series of content based analyses tools. Based on the proposed framework, we also develop an automatic traffic flow analysis system and make experiments on two example traffic videos taken from different monitoring angles.

The rest of our paper is organized as follows. Section 2 introduces some methods for video shot detection, and presents the related work about traffic video surveillance system. In Section 3, we describe the flow analysis framework for traffic video. In Section 4, the experimental results are detailed. Finally, the conclusion is presented in Section 5.

2. Related work

Shot detection, which also known as shot boundary detection or transition detection, is fundamental to kinds of video analysis and application, since it is a critical factor for segmenting a video into its basic components: the shots.

The detecting of shot changes depends on the suitable choice of the decision function between two frames. The functions are mainly based on pixel based methods or histogram based methods. In detail, some researchers used intensity histogram, edge change ratio (ECR), hue, saturation, or other features as the dissimilarity function. In [1], an adaptive threshold which applied histogram differences of a sliding window is used for abrupt shot change detection. In addition, dividing a selected frame into a fix number of blocks and performing an analysis of block wise histogram difference are used to detect gradual shot changes.

As for the video segmentation, the selection of threshold is always critical to the final result. In some video shot detection methods, the detection decision is based on a hard threshold of some dissimilarity measures, which is determined by many experiments. But in some situation, one fixed threshold may be not applicable enough for kinds of videos. In [4], an adaptive threshold which is applied for histogram differences of a sliding window is used.

The vehicle detection and tracking system is a very important part of Intelligent Traffic System (ITS), which always consists of three modules: region of interesting (ROI), vehicle detection, and vehicle tracking modules. The purpose of ROI extraction is to do rude segmentation on the vehicle in the traffic video, extracting the possible area of vehicle from the complex traffic scene for subsequent detection and tracking operation. The methods of this include analysis on the time domain, space domain, and spatio-temporal domain, among which the time domain analysis is the most important, owing to the special characteristics of video images. In other words, the detection of a pixel mainly depends on the pixels at the same location in the video sequence, instead of the surrounding pixels of the same frame. The Frame difference method, background difference method, and unsupervised segmentation method are mainly used for ROI extraction[5].

Furthermore, in [3][6][7], robust techniques for background subtraction of urban traffic video are recounted. Other analysis methods for traffic video, like classification and tracking of vehicles[8][9], on line anomaly detection[10], recognition and online route tracing[11] are also described. Additionally, P.Kumar, et al. had proposed a framework for real time behavior interpretation from traffic video[12].

3. Algorithm description

The generic shot detection method always focuses on distinguishing normal shot frames from cut frames, and aims at finding the abrupt shot change or gradual shot changes. However, the traffic videos gathered by road monitors, are always continuous without obvious transition of shots, therefore, the shot detection methods for ordinary videos may be impractical under certain circumstance. Motivated by this, we propose a novel flow analysis framework for traffic video based on their content. Re sampling is first performed for efficient and precise analysis, and employing foreground extraction on the original video for unbiased traffic flow analysis, after that, variety of frame characteristics including brightness, entropy, Harris corners, and variance of traffic flow are calculated for frame property analyses. All of these establish the basis of video segmentation and traffic flow analysis.

3.1. Re-sampling

In signal processing, re sampling is the process of reducing the sampling rate of a signal. This is usually done to reduce the data rate or the size of the data. For visual material, resampling is to make a digital image smaller by removing pixels. Scaling down always produces a sharper result than scaling to a larger size. With a high quality image or photo editor, re sampling produces an excellent image.

Camera and object motions always introduce large variation. Although the histogram difference is insensitive to object motion, it remains somewhat sensitive to camera motion, such as panning, or zooming. On the other hand, during the variations of video stream, the noise may cause errors in the detection of gradual transitions, which extends over several adjacent frames. To eliminate this effect, we take the process of interpolation to reduce the influence of noise.

The brightness value for the transformed pixel can be calculated by interpolating between the pixels surrounding the calculated address. In our proposed method, we adopt interpolation on the original video frames to get DC image of each frame, by calculating the average value of each 8*8 blocks, which is defined as:

$$C_x(m,n) = \frac{1}{8}\sum_{i=0}^{7}\sum_{j=0}^{7} f_x(8m+i, 8n+j) \tag{1}$$

where $f_x(i,j)$ is the original frame, and the comparison between original image and DC image is shown in Figure 1.

(a) Frame size: 320*240

(b) Frame size: 80*60

Fig. 1. The Original Image (a) of One Frame, compared with the DC image (b), whose size is the 1/16 of the original image.

3.2. Foreground extraction

For unbiased traffic flow analysis, foreground extraction is employed on the original video in order to get foreground video. First and foremost, the background is extracted by averaging the total frame images of each pixel location.

$$B_x(i,j) = \frac{1}{N}\sum_{x=0}^{N-1} f_x(i,j) \tag{2}$$

where N is the total frame number, and $f_x(i,j)$ is the pixel value at f_x in frame f_x.

In order to reduce the influence of background scene, we take the foreground extraction by subtracting the background image from each original frame, as equation (3) defined,

$$Fg_x(i,j) = |O_x(i,j) - B_x(i,j)| \tag{3}$$

Actually, although the traffic videos are captured by fixed monitors, the background maybe not unchanging all the way, due to the tiny movement of the monitor, and other influencing factor like the environment or weather, which might cause some minor mistakes on the foreground video, or bring some deviations on the final result, as shows in the Figure 2(b). For the purpose of removing these deviations on the foreground video, setting a threshold is much necessary. If the difference of pixel values between the original frame and the background image is less than the threshold T, it will be set to 0.

$$Fg_x(i,j) = \begin{cases} 0 & \mathrm{Fg}_x(\mathrm{i,j}) < \mathrm{T} \\ \textit{no change} & \mathrm{Fg}_x(\mathrm{i,j}) < \mathrm{T} \end{cases} \tag{4}$$

After this operation, the foreground video without the background will be obtained, as an example shown in the Figure 2(c).

3.3. Frame property analyses

For the generic video, the most common shot boundary detection techniques always based on the pixel differences, statistical differences, histogram, or motion vectors. As for traffic video analysis, we often pay more attention to the car flow in traffic, so it is necessary to select the features which may reflect the content of video well and truly. In our implementation, we employed brightness, entropy, and Harris corners as the features for analysis, and define a dissimilarity function by combining these features, which may reflect the complexity of the content very well.

3.3.1. Brightness analysis

Brightness is the perception elicited by the luminance of a visual target, whose range is from 0 to 255. In the RGB color space, brightness can be thought of as the arithmetic mean of the red, green, and blue color coordinates. As sometimes the three components may make uneven contribution to the brightness value, we defined the brightness by averaging all the pixel value in one image and employing normalization to decrease the range to [0,1], as the equation shows below,

$$B = \frac{\sum_{i=0}^{width}\sum_{j=0}^{height} f(i,j)}{255 * width * height} \quad (5)$$

Besides the normalization, the calculation is carried out on each band for improving the precision.

3.3.2. Entropy

Entropy is an important concept measuring the average information in an image, and shows how many bits need to code the image data. The entropy of an N*N image can be calculated by the equation [13]:

$$Entropy = -\sum_{g=0}^{L-1} P_i \log_2(P_i) \quad (6)$$

where

Pi= the probability of the ith gray level = nk/N2,

nk = the total number of pixels of gray value k,

L = the total number of gray levels.

This measure provides us a theoretical minimum for the average number of bits per pixel that could be used to code the image, and tends to vary inversely with the energy. As the pixel values in the image are distributed among more gray levels, the entropy increases. Entropy reflects the information quantity and complexity of an image, therefore, a complex image has higher entropy than a simple one.

3.3.3. Harris corners

The Harris corner detector is a popular interest point detector due to its strong invariance to rotation, scale, illumination and noise. The definition of it is based on the local auto correlation function of a signal, which measures the local changes of the signal with patches shifted by a small amount in different directions [14].

Given a shift (△x, △y) and a point (x,y), the auto correlation function is defined as:

$$c(x,y) = \sum_{W}[I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad (7)$$

where I(x,y) denotes the image function, and (xi,yi) is the points in the window W(Gaussian) centered on (x,y).

Let λ_1, λ_2 be the eigenvalues of matrix c(x,y). If both of the eigenvalues are high, the local auto correlation function is sharply peaked, and shifts in any direction may result in a significant increase, which indicates a corner.

3.4. Video segmentation

In particular, we explore the utilization of a feature function, combining of the three kinds of properties we talked above, to depict the variation of content for each frame along the video, as defined as:

$$f(n) = \frac{entropy + corner + brightness}{3} \quad (8)$$

For the convenience of operation, a visualization tool is also built, which summarizes the entire video sequence and automatically highlight a collection of frames based on some metrics defined by fully automated or semi automated techniques. Among the visualization tool for video segmentation, Function dif(n 1,n) is used to measure the discrepancy between two adjacent frames, which is defined as:

$$dif(n-1, n) = | f(n-1) - f(n) | \quad (9)$$

Based on the dissimilarity function (Eq.(9)) and threshold techniques, shot cuts can be easily detected. In our work, it was implicitly assumed that the distributions were indeed stationary, and thus that a single decision threshold could be used. To find this threshold, we experiment with a collection of values until the best one is found.

4. Experimental results

Our experiment is carried out on two traffic videos, and both of them are gathered by road monitors. One (video NO.01) contains 1426 frames, with the size of 320*240, and the other (video NO.02) contains 9538 frames, with the size of 644*160. The processing of each video follows the four steps recounted above. They are re sampling, foreground extraction, feature analysis, and video segmentation in sequence.

4.1. Foreground extraction

As we talked before, foreground extraction processing subtracts the background image from the original video, and identifies candidate foreground pixels on each frame. After extracting the background, shown as Figure 2(b) and Figure 3(b), foreground extraction is done based on Equation (3) and (4), in order to reduce the influence of the external influence, such as

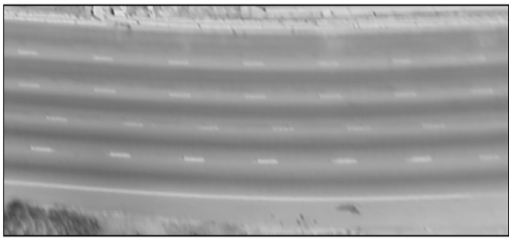
(a)

(b)

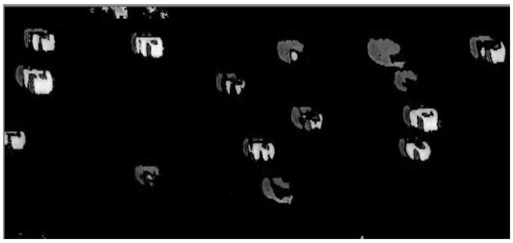
(c)

Fig. 2. Background (a), Foreground (b), and Foreground after Threshold Operation (c) of Video NO.01.

(a)



(b)



(c)

Figure 3. Background (a), Foreground (b), and Foreground after Threshold Operation (c) of Video NO.02

monitor moves or weather variations. The result of foreground extraction is shown in Figure 2 and Figure 3.

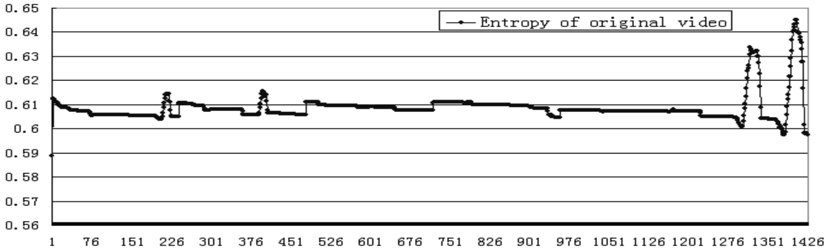
4.2. Comparison with the foreground video

Take Video NO.01 as example, after foreground extraction and features calculation, compare the feature curves between the original video and foreground video, we can draw the conclusion that foreground extraction is very useful and effectual for subsequent analysis, because this may make it easy to understand the content of each frame avoid being influenced by the background changes. Figure 4 shows the entropy curves of original video and foreground video, obviously, the latter describes the variation of frames more clearly and legibly.

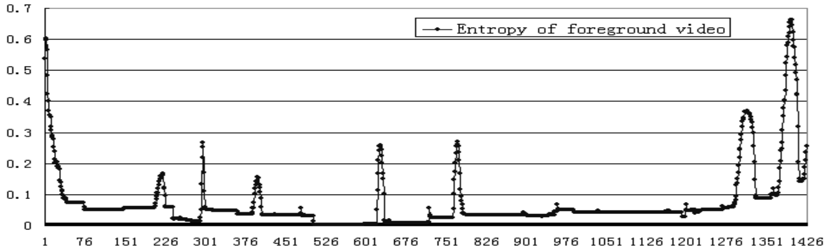
4.3. Dissimilarity measurement and video subsection

Based on the equations (5)~(9), we calculate the frame properties and dissimilarity values between adjacent frames along the video, and the three kinds of characteristic curves and dissimilarity curve of the two video segments are drawn in Figure 5 and Figure 6 respectively. From Figure 5, it is easy to see that although brightness, entropy and Harris corners represent different features of the frames, the general trends of them along the video are approximately similar. Moreover, the dissimilarity curve is more legible, as it is the combination of the total features.

In Figure 5 (b), without regard to the first several frames of the dissimilarity curve, from the start of the video to Frame 1255, most the values are lower than 0.2, although there are a few small peak values. So we set the threshold value as 0.2, and cut up the video into two subsections, and the results are shown in Figure 6.

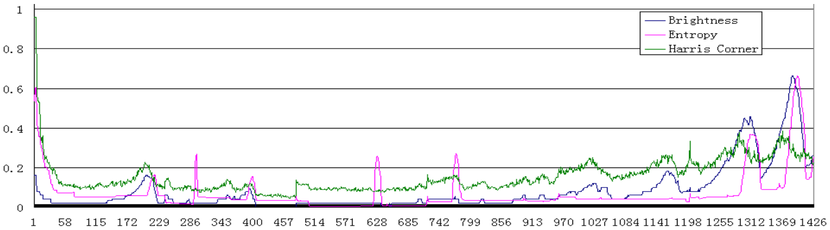


(a)

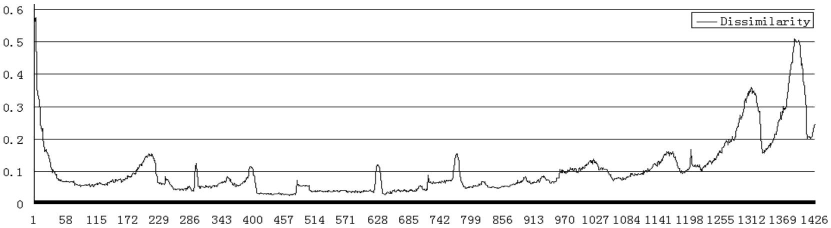


(b)

Fig. 4. Comparison of Entropy Curves between Original Video (a) and Foreground Video (b) of Video NO.01.



(a)



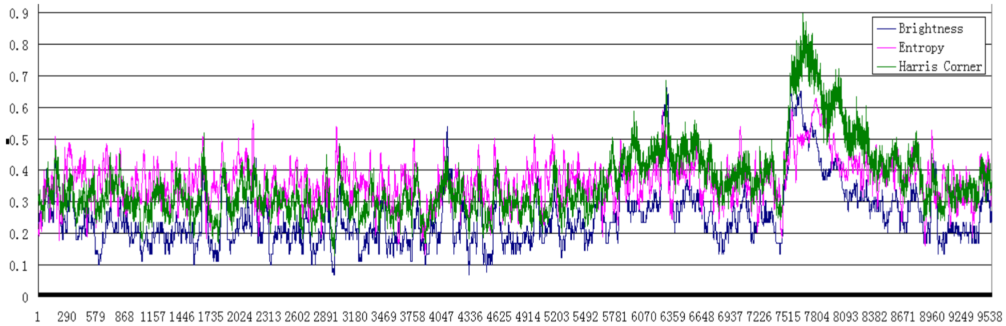
(b)

Fig. 5. Characteristic Curves (a) of Brightness, Entropy, and Harris Corners; and the Dissimilarity Curve (b) of Video NO.01

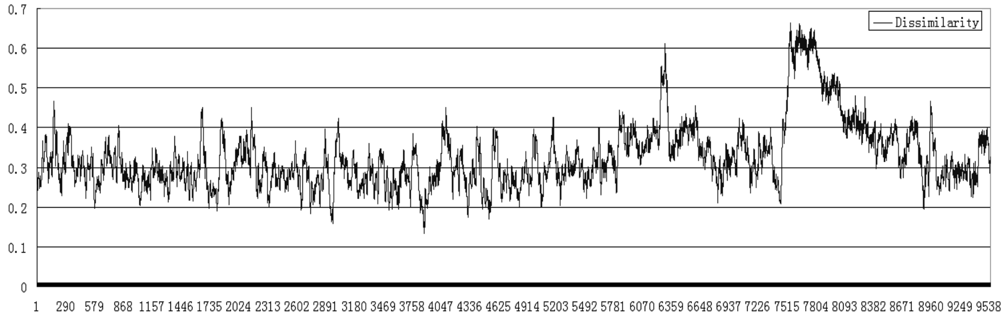
(a)

(b)

Fig. 6. Segmentation Result of Video NO.01, shows the Frame 40 (a) and Frame 1294(b).



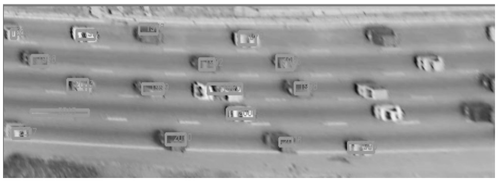
(a)



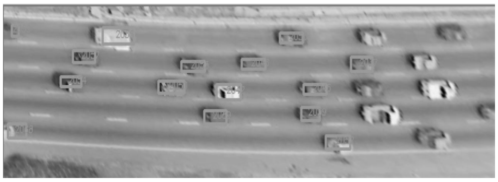
(b)

Fig. 7. Characteristic Curves (a) of Brightness, Entropy, and Harris Corners; and the Dissimilarity Curve (b) of Video NO.02

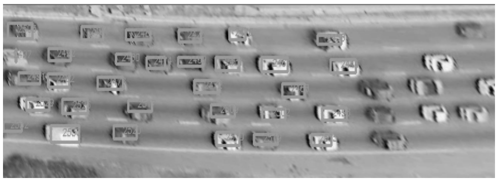
Frame 1



Frame 6229



Frame 6348



Frame 7749

Frame 8888

Fig. 8. Segmentation Result of Video NO.02

Video NO.02 is taken from different monitoring angles, in the same way, if we set the threshold value as 0.5, it will be segmented into five parts based on the dissimilarity curve of Figure 7(b), and parts of the segmentation result are showed in Figure 8.

Based on the dissimilarity curves and the segmentation results of the videos, further analysis can be carried out on the traffic condition. By professional analysis, available information may be obtained, like the precise position where there is traffic jam, or between which time slice there is a rush hour. All these information are useful for the traffic control and improving the road construction.

5. Conclusion

In view of the special characteristics of traffic videos, the ordinary shot detection methods cannot be used on them, hence, we propose a flow analysis framework for traffic video, based on which an automated traffic flow analysis system also has been developed.

Among the framework, a dissimilarity function which defined by three different features is elaborated. All the features in the function have close relationship with the content of the traffic videos. The experimental results on two example videos which taken from different monitoring angles illustrate the efficiency of our framework for traffic video analysis and practical application. In addition, our proposed framework is helpful for strengthening road traffic safety, estimating local infrastructure needs and making an optimal use of the existing traffic facilities.

References

[1] T.H. Tsai, Y.C. Chen, "A robust shot change detection method for content based retrieval", IEEE International Symposium on Circuits and Systems, 2005, pp. 4590-4593.

[2] A.Sundaresan, V.Raykar, "Video Codec Implementation, DC Image Extraction and Shot Segmentation", ENEE631 Fall 2001 Course Project Report. 2001.

[3] M.Vargas, et al. "An Enhanced Background Estimation Algorithm for Vehicle Detection in Urban Traffic Video", Proceeding of the 11th International IEEE Conference on Intelligent Transportation Systems. Beijing, China, 2008, pp. 784-790.

[4] Y.Yusoff, W.Christmas, J.Kittler, "Video Shot Cut Detection Using Adaptive Thresholding", British Machine Vision Conference, 2000.

[5] S.N.Wang, M.Yu, G.Y. Jiang, "Review on Vehicle Detection and Tracking Techniques Based on Video Processing in Intelligent Transportation Systems", Application Research of Computers, 2005, pp. 9-14.

[6] S.C.Cheung, C.Kamath, "Robust Techniques for Background subtraction in urban traffic video", Visual Communications and Image Processing 2004, Proceedings of the SPIE, Volume 5308, pp. 881-892.

[7] S.Cheung, C.Kamath, "Robust Background Subtraction with Foreground Validation for Urban Traffic Video". EURASIP Journal on Applied Signal Processing, 2005, pp. 2330-2340.

[8] B.Morris, M.Trivedi, "Robust Classification and Tracking of Vehicles in Traffic Video Streams", Proceeding of the IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September 17-20, 2006, pp. 1078-1083.

[9] B.Li, Q.M.Chen, F.Guo, "Freeway Auto surveillance From Traffic Video". 2006 6th International Conference on ITS Telecommunications Proceedings, pp. 167-170.

[10] F.Archetti, C.E.Manfredotti, et.al. "Parallel FirstMarkov Chain for on line Anomaly Detection in Traffic Video Surveillance". IET Conference on Crime and Security , London, UK, 13 14 June 2006, pp. 582-587.

[11] C.Busch, R.Dorner, C.Freytag, H.Ziegler. "Feature Based Recognition of Traffic Video Streams for Online Route Tracing", Vehicular Technology Conference, 48th IEEE, 1998, pp. 1790-1794.

[12] P.Kumar, S.Ranganath, W.M.Huang, K.Sengupta. "Framework for Real Time Behavior Interpretation From Traffic Video". IEEE Transactions on Intelligent Transportation Systems. VOL.6, NO.1, 2005, pp. 43-53.

[13] S.E.Umbaugh, Computer Imaging: Digital Image Analysis and Processing, CRC Press, the United States, 2005, pp. 489-492.

[14] K.G. Derpanis. "The Harris Corner Detector". 2004.

[15] Chang Rak Yoon, et al., "Realtime Video Visualization based on 3D GIS," Journal of Korea

Spatial Information System Society, Vol.11, No.1, 2009.

[16] Seong Hoon Kim, et al., "Understanding Lane Number for Video-based Car Navigation Systems," Journal of Korea Spatial Information System Society, Vol.11, No.1, 2009.

Lu Shuang Bai

2007 Dept. of Computer Science, Chongqing University of Posts and Telecommunications(B.S.)

2007~Present Master Student, Dept. of Computer Science, Chongqing University of Posts and Telecommunications

Research Interests : Database and Data Mining, Computer Vision.

Ying Xia

1993 Dept. of Computer Communication, Chongqing University of Posts and Telecommunications(B.S.)

2001 Dept. of Computer Science, Inha University(M.S.)

2006~Present, Doctoral Student, Computer Application, Southwest Jiaotong University

Research Interests : Database and Data Mining, GIS.

Sang-Chul Lee

1998 Dept. of Computer Science and Engineering, Inha University(B.S.)

2003 Dept. of Computer Science, University of Illinois at Urbana Champaign(M.S.)

2006 Dept. of Computer Science, University of Illinois at Urbana Champaign(Ph. D.)

2006~2008 National Center for Supercomputing Applications, USA

2008~Present Full-time Lecturer, Dept. of Computer and Information Engineering, Inha University

Research Interests : Computer Vision, Image Processing, GIS