Video Scene Invariant Crowd Density Estimation Using Geographic Information Systems

SONG Hongquan¹, LIU Xuejun², LÜ Guonian², ZHANG Xingguo³, WANG Feng¹

¹College of Life Sciences, State Key Laboratory of Cotton Biology, Key Laboratory of Plant Stress Biology, Henan University, Kaifeng 475004 Henan Province, P. R. China

²Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023 Jiangsu Province, P. R. China

Nanjing 210025 Stangsu 110vince, 1. K. enina

³College of Urban and Environmental Science, Xinyang Normal University, Xinyang 464000 Henan Province, P. R. China

Abstract: Crowd density is an important factor of crowd stability. Previous crowd density estimation methods are highly dependent on the specific video scene. This paper presented a video scene invariant crowd density estimation method using Geographic Information Systems (GIS) to monitor crowd size for large areas. The proposed method mapped crowd images to GIS. Then we can estimate crowd density for each camera in GIS using an estimation model obtained by one camera. Test results show that one model obtained by one camera in GIS can be adaptively applied to other cameras in outdoor video scenes. A real-time monitoring system for crowd size in large areas based on scene invariant model has been successfully used in 'Jiangsu Qinhuai Lantern Festival, 2012'. It can provide early warning information and scientific basis for safety and security decision making

Keywords: crowd density estimation; video scene invariant; GIS; video spatial registration

I. INTRODUCTION

Excessive crowding will result in unexpected events, such as riots, fights or emergencies. Crowd density can serve as an important

descriptor of crowd stability, because it can quantitatively or qualitatively provide the amount of pedestrians in an area[1,2]. There have been many methods for crowd density estimation using computer vision techniques. These methods can be divided into three categories, 1) pixel-based analysis[3-6], 2) detection-based analysis [7-10], and 3) texture-based analysis[11-15]. These methods rely on specific camera training data which requires a system to be trained and tested on the same viewpoint, using potentially hundreds or thousands of annotated training frames. Even though large-scale CCTV networks are becoming increasingly common, automated crowd counting is not widely deployed. Cao and Huang[16] developed a video-based crowd density estimation and prediction system for wide-area surveillance. One of the largest barriers to full deployment of this technology is the requirement to train each camera independently, which is both time-consuming and expensive.

In recent years, some scene invariant methods have been proposed for crowd density estimation. Kong[17] *et al.* described a viewpoint invariant method that took into account feature normalization to deal with perspective projection and different camera orientation for This paper presented a video scene invariant crowd density estimation method using Geographic Information Systems (GIS) to monitor crowd size for large areas. counting people in crowds from a single camera. Ryan[18] et al. extended the method of Kong[17] et al. using a global scaling factor to relate crowd sizes from one scene to another. Then Ryan[19] et al. proposed another scene invariant crowd counting algorithm using camera calibration that used local features to monitor crowd size. Dong[20] et al. built a set of feature templates for different crowd density scenes, and calculated the similarity between templates and features that were extracted from surveillance video frames. This method can be deployed with minimal setup for other new cameras. Lin[21] et al. proposed a cross camera people counting model that can adapt itself to a new camera scene without the need of manual setup. In order to achieve video scene invariance, these methods should calculate the scaling factor for each camera scene using camera calibration, feature normalization or scales-covariant description, et al. However, these methods are also tedious, such as multi-scales patches generation and feature templates building. Sometimes it is impossible to obtain the scaling factor when we cannot get the camera parameters.

The major obstacle to developing scene invariant crowd density estimation model is the multi-scales of surveillance images because of the differences of camera parameters at different surveillance sites. Geographic information systems (GIS) integrate hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information. So GIS can manage, analyze and display all forms of geo-referenced data in a single scale by the geographical reference. In this paper, the multi-scales surveillance images were mapped to the unified geographic reference using GIS. We can make use of GIS to analyze the crowd image with the same scale. So the novel method we proposed based on GIS can estimate the crowd density without scene-specific learning. Scene-specific means that we should train crowd density estimation models for each camera.

Our contributions to the scene invariant crowd density estimation method are in two

aspects. Firstly, the crowd density estimation model learned with video data taken by one specific camera can be adaptively applied to other cameras. This make the automatic crowd density estimation possible in large-scale surveillance area. Secondly, we can make use of GIS to analyze the spatial-temporal distribution patterns for regional crowds. And this can provide the scientific basis for the decision making of security departments.

The remainder of this paper is organized as follows. Section II describes the details of the proposed approach. Section III presents the experimental results and gives some discussions. Section IV summarizes the approach and presents some clues for the future research work.

II. RESEARCH METHOD

In our method, video data were mapped to GIS. Then, we train the crowd density estimation model based on the crowd images captured by one camera in GIS. This model can be applied to other cameras in GIS. For example, let $A = \{S_1, S_2, ..., S_n\}$ be a set of cameras, it means that there are n cameras in surveillance area A, where S_i is one of the cameras in this visual surveillance system. And $M = \{f_1, f_2, ..., f_n\}$ denote the crowd density estimation model set corresponding to set A, whereas $f_i = \{f_{pixel}, f_{texture}, ...\}$ is the crowd density estimation model for the camera S_i . So one of the largest obstacles to the widespread use of the automatic crowd density estimation is the requirement of train estimation model for each camera independently, which is both time consuming and expensive. In this paper, we attempt to find a crowd density estimation model *f* that can be applied to each camera.

2.1 Overview of our framework

Now we provide the technical details of our approach and present how the crowded dynamic information is analyzed, and then visualized in GIS. Fig. 1 shows an overview of our framework to accomplish crowd density estimation in GIS. Firstly, live videos have to be registered onto GIS by three stages of crowd activities region of interest (ROI) segmentation, perspective distortion correction, and spatial registration. Then, the crowd foreground features have to be extracted from the image in GIS. Next, the foreground features act as the inputs to the crowd density estimation model that learned with video data in GIS taken by one particular camera. Finally, we can obtain the number of people or the crowd density for each camera in this surveillance area. These results will be displayed on the map, so we can get temporal and spatial patterns for crowds in the large-scale surveillance area.

2.2 Video spatial registration

GIS allows us to view, understand, question, interpret, and visualize the same scale data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts. In order to construct scene invariant crowd density estimation model and analyze estimation results in GIS, the original video frames have to be registered onto GIS and image pixels have to be resampled to make crowd images with the same scale and pixel size. Fig.2 presents the main geometric relationships between image plane and GIS map plane. A fixed camera is located at point C. The plane I represents the projection of the world onto this camera. The plane R is the image plane after perspective rectification. Point *P* is an arbitrary point on the ground plane and point p is its projection on the space of the camera, and point p' is the point on the rectified image plane corresponding to the point *p*. Our goal is to find the transformation relationship between point p and point P.

Fig.3 shows the process of video spatial registration. Perspective distortion is a fact that far objects in the scene appear smaller than near ones in the image. So perspective distortion causes each person or each group of people to occupy different amount of pixels in one image. Pixel-based methods and texture-based methods are substantially affected by this perspective distortion. Therefore, im-

age rectification is necessary to bring all the objects at different distances in a scene to the same scale. Over the years, several methods have been developed to rectify images[22-24]. In this paper, a method presented by Liebowitz and Zisserman[25] is successfully employed in the rectification of images obtained by uncalibrated cameras. This method requires the estimation of two vanishing points and the prior knowledge of two angles on the ground plane. Most of the public gathering places are ground plane scenes, such as station square, temple fairs, pedestrian streets, etc. Consequently, these parameters can be easily obtained by a large amount of parallel and perpendicular lines. In general, they estimate the projective transformation by establishing three matrices or transformations.







Fig.2 Geometry model of the video spatial registration



Fig.3 Flow of video spatial registration



Fig.4 Crowd images spatial registration example

where S represents the similarity transformation, A is the affine transformation, and P is the pure projective transformation. Each one of these transformations is responsible for the restoration of certain geometric and metric properties. They can be accomplished by known parameters on the image and ground planes. The pure projective transformation is responsible for restoring line parallelism and area ratios. This can be easily achieved by estimating the homogeneous representation of the vanishing line, which is determined by locations of two vanishing points. The affine transformation restores angle and length ratios of nonparallel lines. This transformation can be obtained using two known angles on the ground plane. The last transformation is similarity transformation, which performs rotation, translation and isotropic scaling of the affine transformation image.

Fig. 4a is one frame of crowd images, and we selected a shape on the image as region

of interest (ROI) as shown in Fig.4b. The image block is then extracted based on the ROI. Then we can get the rectified ROI image block by pure projective transformation and affine transformation (see Fig.4c). In order to recover the metric geometry and accomplish scale consistency for videos, we align the image with GIS plane X_gY_g by means of a linear transformation. Let (x_g, y_g) be a ground truth location point in the GIS coordinate space X_gY_g and let (x, y) be the corresponding location in the coordinate space XY. Then the linear transformation matrix which gives the transformation between these two points can be written as:

$$\begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & t_x \\ k_3 & k_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
(2)

where t_x and t_y are the translation parameters, k_1 , k_2 , k_3 , and k_4 are the affine parameters. In our case, the transformation matrix H is a homography from a patch of an image to the patch in the plane of GIS map. Then we can project the location of a point in the image onto the plane of GIS map environment. Since the transformation matrix has 6 degree of freedom (DOF), the 6 unknown parameters can be solved using more than three pairs of points. Fig.4d illustrates an example of the spatial registration results for a background image.

2.3 Scene invariant crowd density estimation

Crowd images that have been captured by different cameras will have the same scale through the spatial registration. In this paper, we estimate the crowd density for each camera based on the images with the same scale in GIS. Consequently, the crowd density estimation model only needs to be learned with video data taken by one specific camera. Methods based on pixels can only be applied when crowd density is low because of its decreased accuracy when occlusion becomes serious. Methods based on texture analysis could solve occlusions to some extent, but it cannot get desirable results when the crowd density is low. In order to test the effectiveness of the scene invariant crowd density estimation method that we have proposed in this paper, we adopt pixel based method and texture analysis based method for low crowd density estimation and high crowd density estimation respectively.

2.3.1 Pixel-based method

Compared with the previous pixel-based approaches, the present one accomplishes image feature extraction under the environment of GIS. Fig. 5 presents the process of pixel-based scene invariant crowd density estimation. Firstly, the crowd movement ROI has to be extracted from the original crowd image. Then, the ROI image has to undergo spatial registration to correct perspective distortion and register this image onto the plane space of the GIS. The foreground pixels can be extracted from the difference between adjacent frames' ROI in GIS. Next, we can get the total number of the extracted foreground pixels based on the binary image. Finally, the number of people within the crowd can be obtained by the functional relationship between the number of people and the number of foreground pixels, and we can get the crowd density by the number of people and the ROI area in GIS.

2.3.2 Texture-based method

In situations of very dense crowds, the performance of pixel-based approaches is highly affected due to occlusions. Thus, we adopt the texture-based method proposed by Marana et al[1] to deal with very dense crowds that may appear during peak hours. Fig. 6 illustrates the flow of texture-based scene invariant crowd density estimation method. The image preprocessing is the same as the pixel-based method, such as ROI segmentation, spatial registration. For the texture-based method, the foreground extraction based on the background images at different time stored in the database. Then texture features of foreground image extracted in GIS are calculated based on gray level dependence matrix (GLDM)[1]. Gray level dependence matrix is a statistical method which was put forward on the basis of second-order



Fig.5 Flow chart of the pixel-based method



Fig.6 Flow chart of the texture-based method

conditions joint probability density function $f(i, j|d, \theta)$. Each $f(i, j|d, \theta)$ represents a probability that gray level value (i, j) take place on the two pixels whose distance is d in the angle of θ . Next, support vector machine (SVM) is used to train and classify feature vectors for crowd density classification.

2.3.3 Adaptive crowd density estimation

Crowd density is classified into five distinct levels based on the five levels of service as shown in Table I [26]. Each level of service is defined based on the range of average area occupancy for a single pedestrian. The five density groups are very low (VL), low (L), moderate (M), high (H), and very high (VH). Pixel-based methods give us reliable results when crowd density is low. And texture-based methods have outstanding performance when the crowd density is high. In order to achieve good estimation results, we need to develop a solution to solve the adaptive crowd density estimation for crowds real-time monitoring.

We adopt the pixel-based method to cope with two density groups of VL and L. Texture-based method is used to deal with the rest of three high density groups. The very first thing to be done is threshold determination.

Table I Level of service		
Level of Service	Density Range(people/m ²)	Group
A: Free flow	<0.50	Very Low
B: Restricted flow	0.50-0.80	Low
C: Dense flow	0.81-1.26	Moderate
D: Very dense flow	1.27-2.00	High
E: Jammed	>2.00	Very High



Fig.7 Adaptive crowd density estimation

We determine the threshold according to the area of ROI and the crowd density value of L. We will select the texture-based method if the crowd density is higher than 0.8 people/ m^2 . So the threshold of people number is the product of 0.8 and the area of ROI. Also, we can transform the threshold of people into the threshold of foreground pixels through the function of pixel-based method. Fig. 7 shows the process of scene invariant adaptive crowd density estimation algorithm. First, we get the pixels number of foreground through the original crowd image through ROI segmentation, spatial registration, foreground extraction, edge detection and binary image generation. Then, we compute the people number using pixel-based method and compare the people number with the threshold value. We can use

the pixel-based method to accomplish the crowd density estimation if the pixels number is no more than the threshold value, and the crowd density is divided into VL or L according to the range of density. Otherwise we will adopt the texture-based method to estimate crowd density.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The Confucius Temple in Nanjing was originally constructed in the year of 1034 in the Song Dynasty. Over the years, the temple has become a favorite entertainment area, with clothing shops and restaurants spread around it. Therefore, this has become a busy area filled with people. In order to ensure security, the public security department has installed a large number of cameras at important junctions.

We chose three cameras with different setups at different sites to test the feasibility of the scene invariant crowd density estimation method proposed in this paper. Fig. 8 a, c, eshow sample images for camera a, b, c with different tilted angles respectively. Fig.8 b, d, f are rectified results of the ROI in sample images.

The crowd density estimation models was trained by crowd images acquired from camera *a*. These model was then applied in the other two cameras to verify the scene invariant of these crowd density estimation model. In our experiment, the size of each pixel is resampled in GIS and its size is 2cm. Based on the pixel-based method we have proposed, the linear function relationship between people number (N_p) and the number of pixels (N_{px}) can be obtained by using a least squares fit, as shown as follows:

$$N_p = aN_{px} + b \tag{3}$$

where a = 0.01973 and b = -5.3126 for "camera *a*". The ground truth area of any polygon can be easily obtained by GIS. The surveillance areas of these three cameras are $271.78m^2$, $150.62m^2$, and $135.53m^2$ respectively. The

crowd density is VL or L when the crowd density is less than 0.8 *people*/m²[26]. So pixel-based method should be adopted by these three cameras when crowd size is no more than 217, 120 and 108 respectively. We can set the threshold value for each camera according to the surveillance area and the crowd density value 0.8 *people*/m². We collected 100 crowd images as testing sample data for each camera. Fig.9 shows the results of pixel-based scene invariant estimation method.

In order to measure the accuracy of this pixel-based scene invariant crowd density estimation model, we calculated the error of each camera by formula 4.

$$err = (\sum_{i=1}^{n} (|y(i) - p(i)|/p(i)))/n \times 100\%$$
 (4)

where y(i) is the estimation result, p(i) is the manual counting result, and *n* is the total number of sample crowd images. Therefore, the accuracy of this estimation model for each camera is 87.96%, 87.67% and 87.60% respectively. If we apply the model learned with data acquired by camera a without spatial registration to camera *a*, *b* and *c*, the accuracy of this model for each camera is 83.21%, 60.37% and 51.68% respectively. So the scene invariant pixel-based crowd density estimation model learned with video data taken by camera *a* can be adaptively applied to other cameras.

Texture-based method is adopted when the crowd density is more than 0.8 $people/m^2$ [26]. It means that we should use texture-based method when the people number of these three cameras is more than 217, 120 and 108 respectively. In this paper, we calculated the GLDM whose gray level and distance are all 8, and select the angle of θ as 0° and 90° because the texture characteristics are similar when θ is 0°, 45° and 135°. After repeated experiments we extract 5 texture descriptors (that are energy, correlation, contrast, entropy, and homogeneity) according to GLDM. We trained the SVM classifier based on the video data collected by camera a. We chose 120 crowd images for each camera to test the scene invariant feasibility of the SVM classifier, and the number of sample images for each level of service (they



Fig.8 Test cameras' sample crowd images and their perspective rectified results



Fig.9 Pixel-based scene invariant estimation results

are M, H, VH) is 40. Based on Table II, we can conclude that it is capable of constructing the scene invariant SVM classifier to estimate

Table II Texture-based scene invariant estimation results				
Camera ID	Moderate	High	Very High	
а	33	35	36	
b	32	36	37	
с	31	34	38	



Fig.10 System interface

the crowd density.

In this paper, we mapped all the videos with diversity of scales to the GIS map environment. Then the crowd image can be processed in GIS with the same scale. The results show that our proposed method can be adaptively applied to different cameras in outdoor scenes. But we did not test the feasibility of this method in indoor scenes. Because there are many types of indoor scenes, such as subways, supermarkets, sports stadiums, railway stations, etc. It is difficult to locate and map these indoor scenes to GIS environment. In the future work, we will test the applicability of this method in the indoor scenes.

Based on the scene invariant crowd density estimation model, we designed and implemented a system for real-time monitoring crowd density by the integration of video surveillance system and GIS. We can acquire real-time data of crowd density levels and the dynamic number of people for each camera at different locations, and they can be displayed by the way of map and curves as shown in Fig. 10. Also, we can retrieve history data and analyze them by spatial analysis tools. This system has been successfully used in 'Jiangsu Qinhuai Lantern Festival, 2012'. It can provide early warning information and scientific basis for safety and security decision making.

IV. CONCLUSIONS

This paper proposed a novel scene invariant crowd density estimation algorithm based on GIS to estimate crowd size and its spatial distribution patterns in a large scale area. Scale consistency was taken into account to scale features between different scene viewpoints. The experimental results from different sites with different camera setup demonstrate the accuracy and reliability of our method. It does not require additional training when deployed for crowd density estimation on a new camera. We can analyze the crowd movement patterns by the spatial-temporal distribution of crowds using scene invariant crowd density estimation model and GIS for a large scale area with a large number of cameras. So many other descriptions of crowd status can be obtained from crowd density, such as the crowd distribution and the trend of abnormal changes of crowd. These factors make it easy to control the amount of pedestrian or make a quick response to the emergent situations. So this method has many potential applications, such as safety monitoring, abnormality detection and so on.

Even though this method works well for crowd density estimation, it is still untested on other crowd characteristics, such as velocities and directions of pedestrian movements. We can detect abnormal behaviors according to the characteristics of pedestrian movements. The spatial-temporal distribution patterns of abnormal hot spots would be able to be displayed on GIS maps. Also, once the characteristics of crowds are estimated for the surveillance areas, they can be used to estimate and simulate missing data for the unobserved regions based on the relationships between density, velocity and flow in closely packed crowds. In the future, the performance of the system may be improved by incorporating the characteristics of pedestrian movement.

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their detailed reviews and constructive comments. We are also grateful for Sophie Song's help on the improving English. This work was supported in part by the 'Fivetwelfh' National Science and Technology Support Program of the Ministry of Science and Technology of China (No. 2012BAH35B02), the National Natural Science Foundation of China (NSFC) (No. 41401107, No. 41201402, and No. 41201417).

Corresponding author

Xuejun Liu, E-mail: liuxuejun@njnu.edu. cn

References

- MARANA A, VELASTIN S, COSTA L, et al. Automatic Estimation of Crowd Density using Texture[J]. Safety Science 1998, 28(3): 165-175.
- [2] ZHAN B, MONEKOSSO D, REMAGNINO P, et al. Crowd Analysis: a Survey[J]. Machine Vision and Applications, 2008, 19(5): 345-357.
- [3] DAVIES A, YIN J, VELASTIN S. Crowd Monitoring using Image Processing[J]. Electronics & Communication Engineering Journal, 1995, 7(1): 37-47.
- [4] MA Ruihua., LI Liyuan., HUANG Weimin, et al. On Pixel Count Based Crowd Density Estimation for Visual Surveillance[C]// IEEE Conference on Cybernetics and Intelligent Systems. Singapore: IEEE Press, 2004: 170-173.
- [5] CHAN A, JOHN Z, VASCONCELOS L. Privacy Preserving Crowd Monitoring: Counting People Without People Models or Tracking[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Alaska, USA: IEEE Press, 2008: 1-7.
- [6] HUSSAIN N, YATIM H, HUSSAIN N L, et al. CDES: A Pixel-Based Crowd Density Estimation System for Masjid al-Haram[J]. Safety Science, 2011, 49(6): 824-833.
- [7] LIN S, CHEN J, CHAO H. Estimation of Number of People in Crowded Scenes using Perspective Transformation[J]. IEEE Transactions Systems, Man and Cybernetics, Part A: Systems and Humans, 2001, 31(6): 645-654.
- [8] ZHAO T, NEVATIA R. Bayesian Human Segmentation in Crowded Situations[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Wisconsin, USA: IEEE Press,

2003: 459-466.

- [9] RABAUD V, BELONGIE S. Counting Crowded Moving Objects[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2003: 705–711.
- [10] KILAMBI P, RIBNICK E, JOSHI A, et al. Estimating Pedestrian Counts in Groups[J]. Computer Vision and Image Understanding, 2008, 110(1): 43–59.
- [11] MARANA A, VERONA V. Wavelet Packet Analysis for Crowd Density Estimation[C]// Proceedings of the 19th IASTED International Symposia on Applied Informatics. Innsbruck, Austria: Acta Press, 2001: 535-540.
- [12] LI Xiaohua, Shen Lansun, Li Huanqin. Estimation of Crowd Density Based on Wavelet and Support Vector Machine[J]. Transactions of the Institute of Measurement and Control, 2006, 28(3): 299-308.
- [13] RAHMALAN H, NIXON M, CARTER J. On Crowd Density Estimation for Surveillance[C]// The Institution of Engineering and Technology Conference on Crime and Security. London, England: IEEE Press, 2006: 540-545.
- [14] WU X, LIANG G., LEE K, et al. Crowd Density Estimation using Texture Analysis and Learning[C]// IEEE Conference on Robotics and Biomimetics. Kunming, China: IEEE Press, 2006: 214-219.
- [15] GUO Sen, LIU Wei, YAN Heping. Counting People in Crowd Open Scene Based on Grey Level Dependence Matrix[C]// IEEE Conference on Information and Automation. Zhuhai/Macau, China: IEEE Press, 2009: 228-231.
- [16] CAO Lijun, HUANG Kaiqi. Video-Based Crowd Density Estimation and Prediction System for Wide-Area Surveillance[J]. China Communications, 2013, 10(5): 79-88.
- [17] KONG D, GRAY D, TAO H. A Viewpoint Invariant Approach for Crowd Counting[C]// 18th International Conference on Pattern Recognition. HongKong, China: IEEE Press, 2006: 1187-1190.
- [18] RYAN D, DENMAN S, FOOKES C, et al. Scene Invariant Crowd Counting for Real-Time Surveillance[C]// The 2nd International Conference on Signal Processing and Communication Systems. Gold Coast, Australia: IEEE Press, 2008: 1-7.
- [19] RYAN, D, DENMAN S, SRIDHARAN S,. et al. Scene Invariant Crowd Counting[C]// International Conference on Digital Image Computing: Techniques and Applications. Noosa, Queensland, Australia: IEEE Press, 2011: 237-242.
- [20] DONG N, LIUM F, LI Z. Crowd Density Estimation using Sparse Texture Features[J]. Journal of Convergence Information Technology, 2010, 5(6): 125-137.
- [21] LIN T, LIN Y, WENG M, WANG Y, et al. Cross Camera People Counting with Perspective Es-

timation and Occlusion Handling[C]// IEEE International Workshop on Information Forensics and Security. Foz do Iguaçu, Brazil: IEEE Press, 2011: 1-6.

- [22] BOSE B, GRIMSON E. Ground Plane Rectification by Tracking Moving Objects[C]// Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Nice, France: IEEE Press, 2003: 94-101.
- [23] DAILEY D, CATHEY F. A Novel Technique to Dynamically Measure Vehicle Speed using Uncalibrated Roadway Cameras[C]// IEEE Conference on Intelligent Vehicles Symposium. Las Vegas, France: IEEE Press, 2005: 777-782.
- [24] MADURO C, BATISTA K, PEIXOTO P, et al. Estimation of Vehicle Velocity and Traffic Intensity using Rectified Images[C]// IEEE Conference on Image Processing. San Diego, USA: IEEE Press, 2008: 777-780.
- [25] LIEBOWITZ D, ZISSERMAN A. Metric Rectification for Perspective Images of Planes[C]// IEEE Conference on Computer Vision and Pattern Recognition. Santa Barbara, USA: IEEE Press, 1998: 482-488.
- [26] POLUS A, SCHOFER J, USHPIZ A.. Pedestrian Flow and Level of Service[J]. Journal of Transportation Engineering, 1983, 109(1): 46-56.

Biographies

SONG Hongquan, received his Ph.D degree in Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University in 2013. He is currently a lecturer in College of Life Sciences, Henan University. His research interests include spatial analysis, GIS software design and development, GIS and remote sensing applications. Email: hqsong@henu.edu.cn

LIU Xuejun, is currently a professor in Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University. His current research interests includes Video-GIS, DEM and digital terrain analysis, and transportation GIS. Email: liuxuejun@njnu.edu.cn

LÜ Guonian, is currently a professor in Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University. His research interests covered a wide range of topics on GIS, including geographical information science, virtual geographical environment, and geographic modeling. Email: gnlu@njnu.edu.cn

Zhang Xingguo, received his Ph.D degree in Key Laboratory of Virtual Geographical Environment, Ministry of Education, Nanjing Normal University in 2014. He is currently a lecturer in College of Urban and Environmental Science, Xinyang Normal University. His research interest includes video GIS, GIS software design and development, digital terrain analysis. Email: zhangxingguo2012@163.com.

WANG Feng, received her B.S. degree in College of Real Estate, Beijing Normal University Zhuhai Campus in 2010. She received her M.S. degree in Geographical Information Science from University of Edinburgh in 2011. She is currently a research assistant in College of Life Sciences, Henan University. Her research interest includes spatial analysis and applications of GIS. Email: f_wang108@126.com