SPATIAL ROAD NETWORK EXTRACTION FROM MULTI SPECTRAL REMOTE SENSING IMAGERY WITH FCD

Yemin Fan

Department of Geography, University at Buffalo, State University of New York, Amherst, NY 14261, USA; e-mail: yfan2@buffalo.edu

KEYWORDS: Local Moran's I statistics, Floating Car Data (FCD), Monte Carlo Approach, Multi spectral Remote Sensing Imagery.

ABSTRACT:

This paper presents an automatic methodology for spatial road network extraction from multi spectral remote sensing images with floating car data. Basically, it can be divided into two steps. In the first step, a spatial local statistics is carried out to extract nodes of road segments. Based on local Moran's I statistics, a new statistic is defined to detect local clusters. Significance is assessed by using a Monte Carlo approach to determine the probability of observing that many data under the null hypothesis of no pattern. When all necessary nodes are detected, spatial road segments then can be organized by linking each two nodes, which are used as the candidate road segments in the second step. In the second step, a pre-processed multi spectral remote sensing image is prepared for testing those candidate road segments. Finally, road segments which are significant in the test are selected to construct the spatial road network. Methodology and experiments are respectively given in this paper.

1. INTRODUCTION.

Spatial road network, as the fundamental component of GIS especially in GIS-T (GIS for Transportation), is very important both in practical applications and theoretical studies. Topics related with spatial road network therefore are discussed by many researchers. In recent years, multi spectral remote sensing imagery, as a new data source to build the geographical information database, is used to provide more detail landscape information. Due to its characters of short cycle period and wide land coverage, many studies are carried out to extract spatial road network from multi spectral images, especially from high spatial resolution images. Object oriented methodology is introduced to model road objects and road network, in which road extraction is generally based on properties of roads and road net work (Peteri, Celle and Ranchin, 2003, Dal Poz, Zanin, do Vale, 2006). Shackelford and Davis (2003) combined pixel-based fuzzy and object-based together to extract road network from high-resolution multispectral satellite imagery. SkourikHine and Alexei (2005) proposed an image vectorization approach to the road network extraction from digital imagery which is based on proximity graph analysis. Knowledge-based methodologies are also very popular. Zhu, et al. (2005) extracts road network based on the binary and greyscale mathematical morphology and a line segment match method. However, no matter date-driven approaches or knowledge-driven approaches are used, they all, to some degree, largely depend on inherent character of greyscale images. That means radiometric information plays a vital role in feature extraction. Therefore, prior knowledge, even when some intellectual computation methods are employed, is always composed of greyscale character of features and shapes of features. In this sense, these methods all have their own limitations.

Nowadays, floating car data (FCD), using dynamic sensor to collect spatial information, is developed rapidly. Each car is equipped with a GPS device and a wireless communication device. The instantaneous position of the car is transmitted at regular intervals to a data server centre. Data server centre collects and processes all the GPS data sets to facilitate decision making on traffic pattern. Nowadays, researches on FCD mainly focus on the application of FCD in traffic state detection (Kerner and Rehborn, 2001, Schafer, Strauch and Kelpin, 2001, Kerner et al. 2005, Kwella and Lehmann, 2000), FCD analysis (Fouladvand and Darooneh, 2005), updating road network in existing GIS database (Smartt, 2006) and traffic information publication (Fan and Liu, 2006).

In this paper, taking the advantages of multi spectral remote sensing imagery and FCD, a new method to extract spatial road network is proposed. The significance of this method lies in not only helping us to extract the spatial road network using FCD and multi spectral RS imagery but also assisting us to divide the road network automatically into reasonable road segments which are compatible with FCD. Therefore, spatial road networks, especially those in the urban area where road networks are changing rapidly and difficult to construct, can be upgraded both automatically and dynamically. In section 2, the feasibility of integrating multi spectral remote sensing imagery data with FCD and the coordinate transformation between them are discussed. In section 3, based on local Moran's I statistics, a new statistic to carry out a spatial cluster analysis is defined to detect nodes of road network. Monte Carlo simulation process is adopted to evaluate significance. In section 4, the strategy to construct spatial road network with nodes detected and pre-processed multi spectral imagery is given out. Experiment results are presented and discussed in section 5. Conclusions are provided in section 6. In figure 1, the relationship among topics discussed in this paper is given out.

As described in figure 1, FCD database is the data source. Monte Carlo simulation process is employed to obtain critical values used to detect local clusters. In the candidate link selection process, multi spectral remote sensing imagery plays an important role to filter candidate road segments. Finally, spatial road network is decided by road segments filtered.

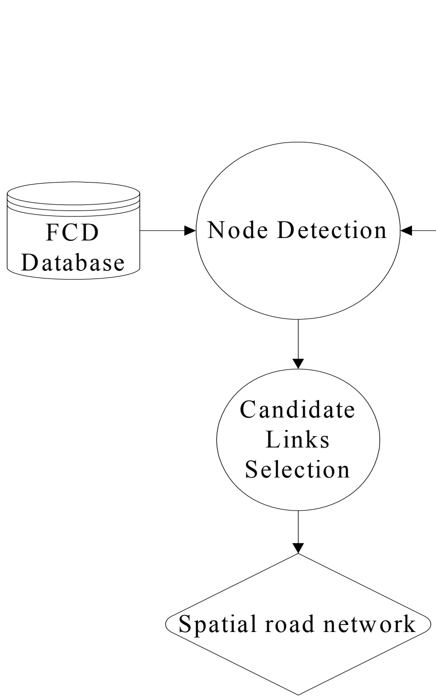


Figure 1. The relationship and structure of topics in this paper

2. INTEGRATION OF MULTI SPECTRAL RS IMAGERY DATA WITH FCD.

No matter in the countryside or in urban areas, the traffic state on the road network can be surely monitored by direct measurements (e.g. induction loops and radar device). This traditional method is effective only when there are not so many vehicles and the demand on the monitor is low, because those devices could not, to some degree, monitor the traffic state dynamically. As the development of GPS technology and wireless communication technology, vehicles can be equipped with GPS devices and their instantaneous positions can be transmitted at regular intervals to a central site. When the positions of a sufficient number of vehicles can be frequently communicated to a central site, travel times can be directly measured. This is called Floating Car Data or FCD for short (S. Turksma, 2000). FCD have become a very important data source for the establishment of a traffic information system.

The positioning information of FCD sent to the traffic centre is geodetic coordinates in the WGS84 coordinate system. However, multi spectral remote sensing imagery data is always with a projected coordinate system. Therefore, to integrate these two together, they must be registered into a same coordinate system in advance. There are about 5000 taxis already equipped with GPS receivers and wireless communication devices in “Shenzhen Urban Transport Simulation System” (SUTSS) project. Since May 2006, millions of GPS data sets from these 5000 taxis have been recorded. This FCD source and high spatial resolution image are used to explain the feasibility to integrate these two data together. Coordinate systems of these two data are listed in table 1.

In this example (Table 1), the coordinate system of FCD is WGS84 geodetic coordinate system, while the remote sensing image is projected by simple cylindrical projection method with a WGS84 datum. For convenience, FCD can be projected with simple cylindrical projection method and then these two data are registered in the same spatial reference. The area between 22°31'48.00"N and 22°32'24.72"N in latitude, 113°54'34.86"E and 113°55'13.56"E in longitude is selected as the study area. FCD is collected from June 3rd to June 5th, 2007. Figure 2(a) and Figure 2 (b) respectively shows the images before and after the overlay of FCD and high resolution remote sensing image.

From figure 2, it is obvious that after the coordinate transformation, FCD matches the remote sensing image very well and FCD almost covers all the roads. By measuring the distances from each FCD to its corresponding central line of road and do a statistics analysis, it is found that accuracy of FCD position is commonly within 10M and a large number of FCD has accuracy higher than 4M. The mean deviation of arithmetic mean is 2.28M, sample variance is 6.92 and standard deviation is 2.63M (Fan, 2007). In the mean while, the spatial resolution of this remote sensing image is around 1M. Thus, taking into account the width of road, it is convincible that these two data can be matched very well.

Data Type	FCD	RS image
Coordinate System	WGS84 Geodetic Coordinate System	Simple Cylindrical projection with a WGS84 datum

Table1 Coordinate systems of FCD and RS imagery

(a) Before the overlay	(b) After the overlay

Figure 2. Images before and after the overlay of FCD and RS image

3. LOCAL CLUSTER DETECTION FROM FCD

Local statistics is a hot topic in many fields. Based on Moran's *I* statistic (Moran, 1950), the local indicators of spatial association was derived by Anselin (1995, see also Getis and Ord, 1996) to resemble passing a moving window across the data, and examining dependence within the chosen region for the site on which the window is centered. The specifications for the window can vary, using perhaps contiguity or distance at some spatial lag from the considered zone or point. With this concept, a new statistic is proposed to detect clusters from FCD.

3.1 Fundamentals

The local Moran statistic is

$$I_i = \frac{n(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2} \sum_j w_{ij}(y_j - \bar{y}) \tag{1}$$

Where, I_i is the statistic of local Moran's I at region i, y_j is the attributes of region i, $\bar{y}$ is the expected value and w_{ij} is the weight. Based on the concept of equation (1) and considering the character of FCD, a new statistic is defined in equation (2)

$$\begin{cases} L_t = \max(F_i) \\ F_i = \dfrac{k_i - \bar{k}}{\sqrt{\bar{k}}} \\ k_i = \sum_j w_{ij} y_j \end{cases} \tag{2}$$

Where, y_i is the number of FCD within an area centred at point i, w_{ij} is the weight, $\bar{k}$ is median value of k_i and L_t is the new local statistic at test time t.

3.2 Significance test of

L_t Due to the uncertainty of FCD, it is hard to tell what kind of distribution L_t has. Therefore, Monte Carlo simulations are employed to get the critical value. Monte Carlo simulations of the null hypothesis of no local clustering confirmed the actual value of L_t consistent with $\alpha = 0.05$ for each time. The simulations were carried out by first filling the study area with randomly distributed points. The local statistics were then found using equation (2). And then the critical value for L_t is found. Because of making a multiple testing, in which each time L_t is found to test whether it is significant till no significant L_t is found, to keep the experimentwise error rate to a specified level (usually $\alpha = .05$), the Bonferroni adjustment (David, 1956) is implemented. If k separate and independent tests are made, then instead of choosing a critical value of the test statistic using α as the Type I error probability for each test, simply, α / k can be used for each test. Therefore, it is an iteration process because parameter k of Bonferroni adjustment is decided by the times of multiple testing. An iterative procedure is needed to approximate the optimal solution. In the case study in section 5, the detail procedure will be discussed.

4. Strategy to construct spatial road network

4.1 Assumptions

There are two assumptions to construct the spatial road network.

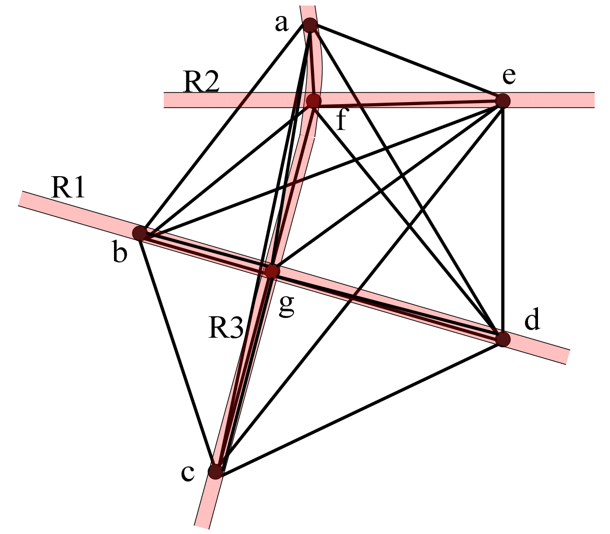
a) There is a traffic pattern in the centre of local cluster
b) There is no significant change of speed on each road segment linked by nodes

In the real traffic system, these two assumptions are always easy to realize. Local clusters mean a significant change of speed, because of high FCD density. Between every two clusters and along a road, vehicle can run at a relatively stable speed. Otherwise, there must be another cluster which should be detected. Therefore, local clusters can be regarded as nodes of road segments.

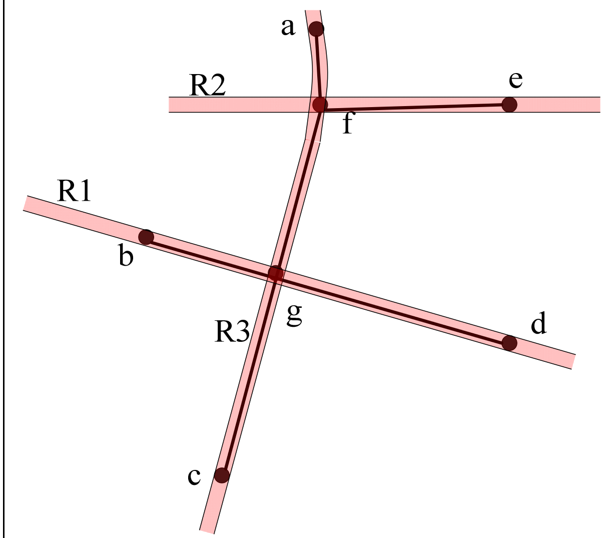
4.2 Strategy for candidate road segment selection

After finding nodes of road segments and projecting these nodes by simple cylindrical projection method with a WGS84 datum, each two nodes can be linked as our candidate road segments and in the meantime, these road segments and multi spectral remote sensing imagery are registered into the same coordinate system. After a pre-processed procedure for multi spectral remote sensing image, the result is used to determine the final spatial road network. This can be explained by figure 3.

In figure 3, 7 nodes are first found with method introduced in section 3. They are a, b, c, d, e, f and g. Then each two nodes can be linked and 21 candidate road segments are produced. R1, R2 and R3 are the road areas which are extracted from multi spectral image. In figure 3, this is an ideal result, but actually the result from pre-processed multi spectral image may contain noise and uncertainty, which are endurable because, from figure 3, if the areas cover all roads in the image can be roughly extracted, the spatial road network can be decided by calculating the probability that each candidate road segment falls in road areas and with a given significance level.



(a) Candidate road segments and road area



(b) Decision result

Figure 3. Strategy to determine the final spatial road network

5. A CASE STUDY

To illustrate the methodology proposed in this paper, a case study is carried out. In the area between 22°31'48.00"N and 22°32'24.72"N in latitude, 113°54'34.86"E and 113°55'13.56"E in longitude (Simple Cylindrical projection with a WGS84 datum), FCD is collected from June 3rd to June 5th, 2007. There are 21935 FCD (data is in WGS84 Geodetic Coordinate System) during this period in this region. The spatial resolution of the high spatial resolution image is around 1M.

5.1 The structure of experiment

In the following figure 4, the structure to carry out this experiment is given out.

In figure 4, there are three inputs for the multiple testing. They are FCD, weight matrix and critical value. The critical value is obviously from Monte Carlo simulation process, which is introduced in section 3. After the multiple testing, road segment nodes are detected and then candidate road segments can be produced. The selection procedure is discussed in section 4. Finally, the spatial road network is constructed.

5.2 The weight matrix

Weight matrix is one of three inputs for multiple testing. Actually, weight is essentially important for the multiple testing. Figure 5 shows a corner of FCD.

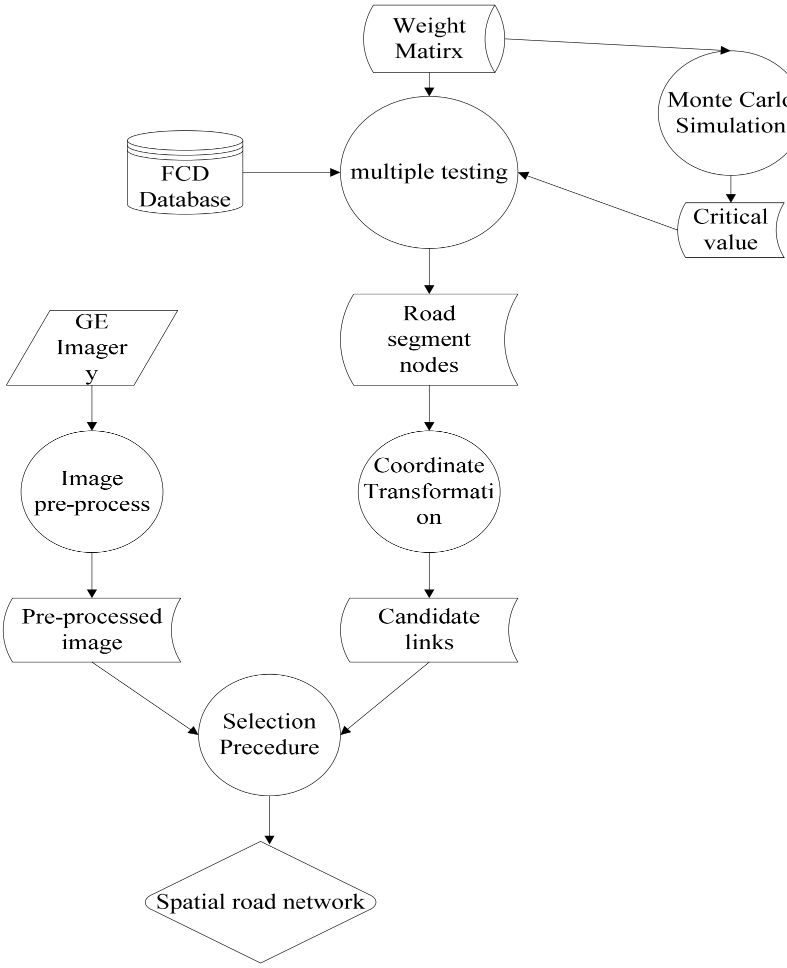


Figure 4. Structure of the experiment

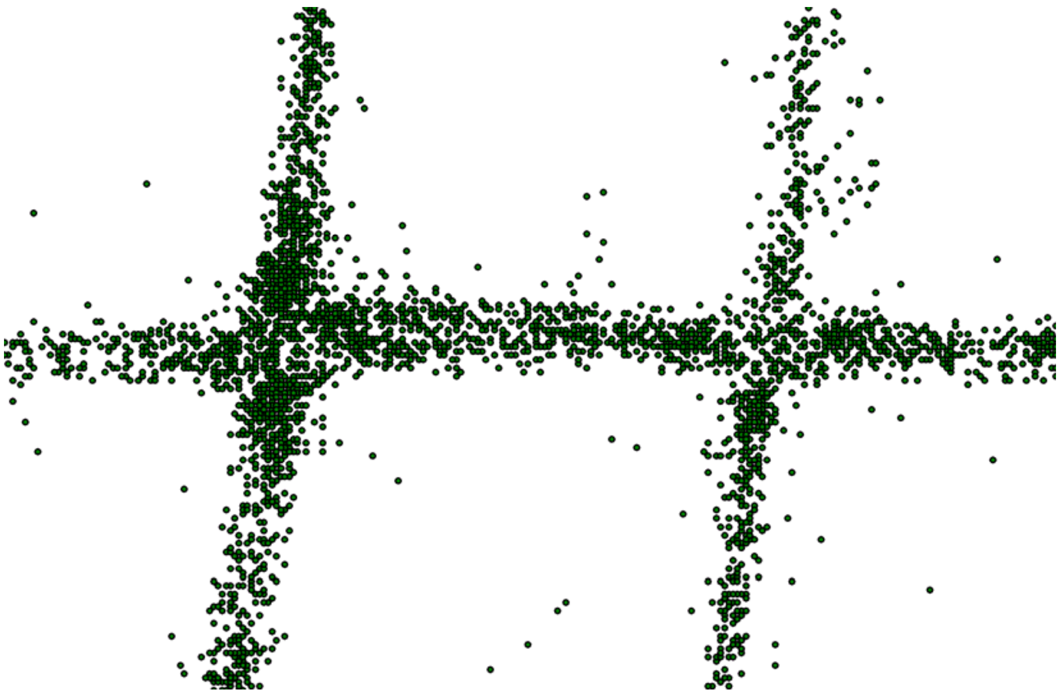


Figure 5. Distribution of FCD at road intersections

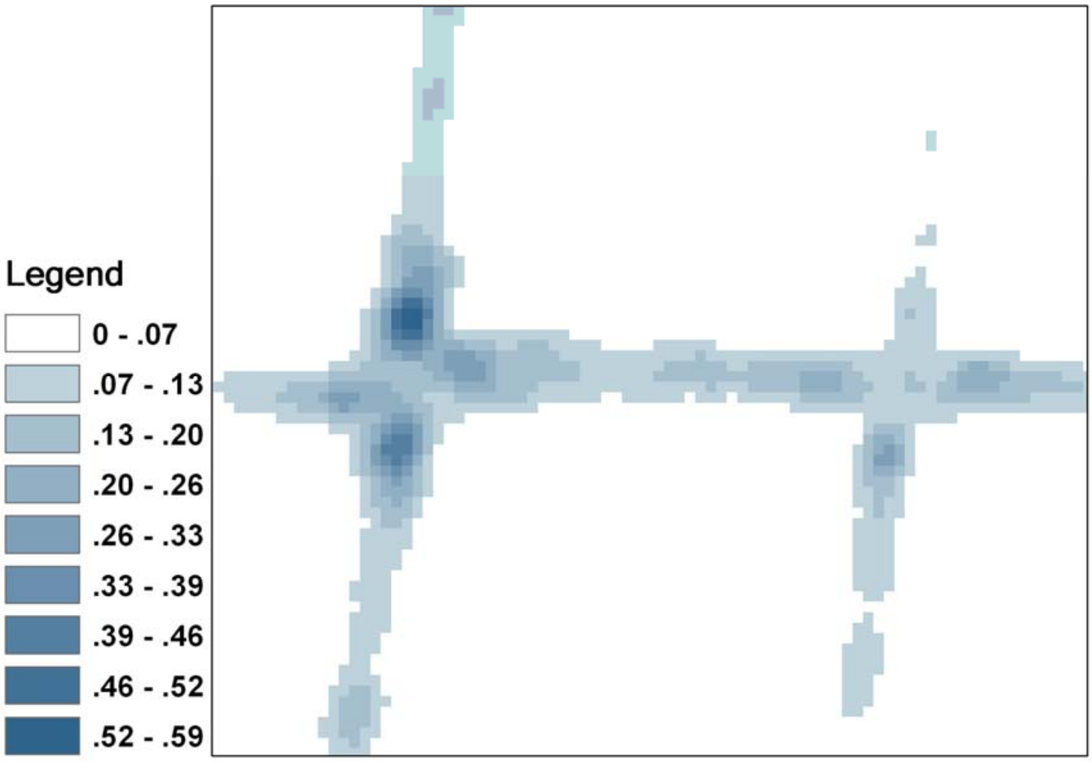


Figure 6. Kernel density of FCD

From figure 5, it is obvious that at each road intersection FCD are not distributed evenly. In the centre of road intersections, there are less FCD but more on the road near the centre. This does make sense because in the really traffic situation, due to the effect of traffic lights, vehicles must wait till they are notified to go which will cause the effect that more vehicles are on the road near the centre. When vehicles are allowed to go, they must go through road intersections without stop, in which situation there should be less vehicles in the centres of road intersections. Based on the interest area in figure 5, a kernel density analysis with a search radius 15M and output cell size 4.5M is carried out and the result is shown in figure 6.

The unit of the classification in figure 6 is number of points per square meters. From the analysis of the kernel density of the whole study area, the weight matrix is defined in equation (3),

$$W_{ij} = \begin{cases} -0.4 & dist(i,j) < 10m \\ 1 & 10 \le dist(i,j) < 30m \end{cases} \quad (3)$$

Where, $dist(i,j)$ is the distance from point i to point j. To be more intuitive, it is easy to explain equation (3) in the real traffic situation. In Shenzhen city, centred at centres of each road intersection, the average radius of all circles which can cover each intersection is 10 meter. Therefore, according to the analysis of kernel density shown in figure 6, to detect road intersections, a negative weight is preferred to give to points within 10 meter distance around each interest point. Along roads near to centres of each road intersection, the average waiting line of vehicle is 20 meter. Thus, according to the kernel density analysis, a positive weight needs to be assigned to points within 30 meter distance but farther than 10 meter distance around each interest point. Points farther than 30 meter distance around each interest point have little influence on detecting road intersections and they are not taken into account.

5.3 Monte Carlo simulation process

Since rules to build the weight matrix is set, the Monte Carlo simulation process can be carried out. There are three steps for each time of the simulation process. They are,

a) 21935 points are randomly distributed in the study area.
b) weight matrix is build with equation (3)
c) the local statistic L_t is calculated with equation (2)

Repeat these three steps 1000 times and a series of L_t can be found. In figure 7, results are arranged in a histogram.

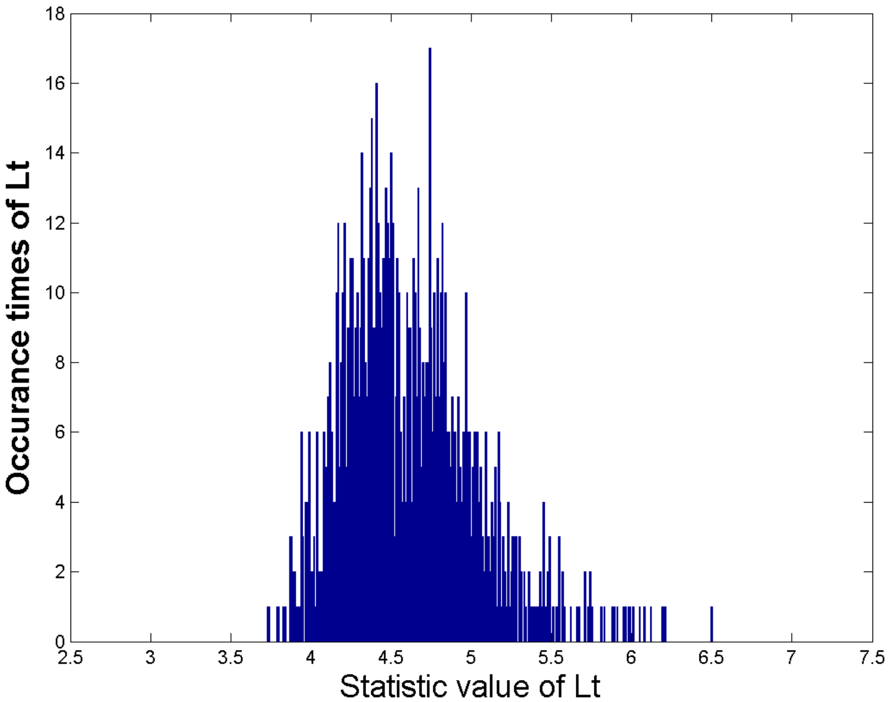


Figure 7. Histogram of Monte Carlo simulation process result

In figure 7, it can give out the critical value for single testing. When multiple testing is carried out, Bonferroni adjustment needs taking into consideration. According to the criteria introduced in section 3.2, critical values at confidence level $\alpha = 0.05$ can be easily calculated and the results are shown in table 2.

Times of multiple testing	Critical value
10	6.0539
20	6.1526
30	6.2009
40	6.2074
50	6.2138

Table2 Critical value considering Bonferroni adjustment

From table 2, it is obvious that if more tests are carried out, a higher critical value is needed to avoid conservative estimation of L_t .

5.4 Node detection result

After 38 times' test with a critical value of 6.2061, which is calculated from table 2 with a linear interpolation method, 38 significant local clusters are found. These points are plotted along with all FCD points in figure 8.

In figure 8, all local clusters are found. Besides road intersections, some clusters are located along the road, in which case it can be assumed that there must be some traffic patterns there. This sort of traffic patterns should be paid more attention and they should be regarded as nodes of road segments to comply with FCD.

5.5 Determination of final spatial road network

Based on the strategy discussed in section 4, a pre-processed imagery of the study area is needed. Due to the aim of this paper is to introduce FCD to road extraction, here the road frame is roughly described by hand, which is used to explain the candidate road segment selection procedure. Figure 9 shows the pre-processed imagery and the final spatial road network overlaid with high spatial resolution imagery.

For convenience, in figure 9(a), the pre-processed imagery is given as a binary image, in which white area is the road area roughly. In figure 9(b), when the spatial road network is overlaid with the high resolution imagery, they match each other very well. Nodes of all road segments are highlighted. Each node represents a traffic pattern, such as road intersection, traffic jam, etc.

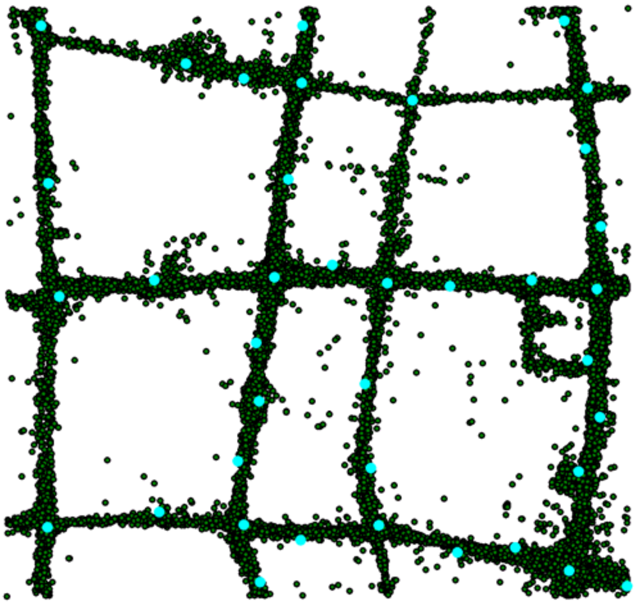
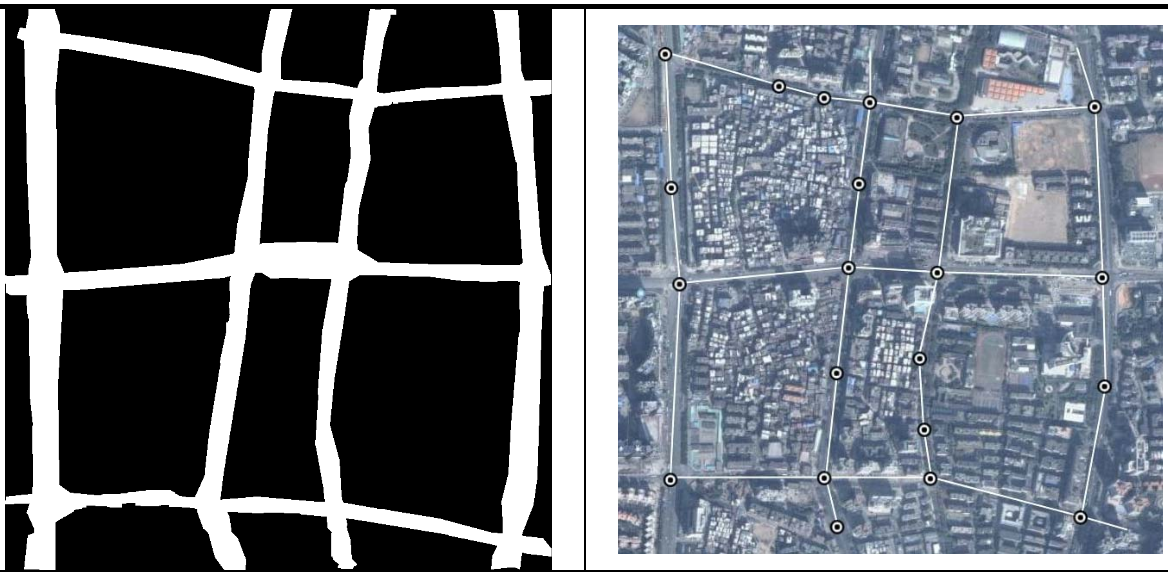


Figure 8. Distribution of centers of local clusters



(a) Pre-Processed imagery	(b) Final spatial road network

Figure 9. Candidate road segment selection

6. CONCLUSION

In the latest decade, GIS, especially WebGIS has been developing prospectively. GIS-T, as an important part of GIS, is discussed more and more. Spatial road network, the fundamental element, has become one of key topics. As the introduction of new form data, FCD, into GIS-T, there comes a new trend which leads to build a more dynamical and more intelligent traffic information system. In this paper, integrated the advantages of FCD and high spatial resolution imagery, a methodology to construct spatial road network automatically is proposed.

In this paper, a new statistic is defined to describe the local cluster based on the character of road intersections. To obtain the critical value of the statistic, Monte Carlo simulation process is employed. Bonferroni adjustment is suggested to keep the experimentwise error rate to a specified level. In the case study, kernel density analysis is carried out to get the key parameter for building the weight matrix. After all the road segment nodes are detected, candidate road segments are produced. Assisted with the pre-processed high spatial resolution imagery, the spatial road network is finally decided.

Besides the final spatial road network obtained with this methodology, it should be noticed that all the road nodes are detected based on FCD and candidate road segments are filtered based on multi spectral remote sensing imagery. Therefore, on the one hand, these nodes are the most compatible with FCD and on the other hand, these nodes match multi spectral remote sensing imagery very well. Thus the significance of this methodology is to provide an automatic way to construct spatial road network which can compatible with both FCD and multi spectral imagery.

In conclusion, FCD can help to construct spatial road network with multi spectral remote sensing imagery. The short cycle period and wide coverage of multi spectral remote sensing imagery make the update of spatial road network more rapidly and the cost of update less, which is vital to build the intelligent transportation system. Especially in urban areas where spatial road network is changing rapidly over time and the traffic situation is always complicated, this methodology can offer an automatic way to maintain the spatial road network database. In the future work, the convenient way to get the pre-processed imagery will be studied and the robustness of this methodology will be analyzed.

ACKNOWLEDGEMENTS

The work described in this paper was supported by National Natural Science Foundation of China (Project No.40501061) and Hi-tech Research and Development Program of China (Project No. 2007AA12Z178). The author would like to thank the data provided by SUTSS (Shenzhen Urban Transport Simulation System) project in China. The author would like to thank Prof. C. Liu from Tongji University of China for his comment and suggestion which helped to improve this paper.

REFERENCES

Anselin, L., 1995, Local Indicators of Spatial Association — LISA. *Geographical Analysis* 27: 93–115.

Dal Poz, A.P., Zanin, R.B., do Vale, G.M., 2006, Automated extraction of road network from medium- and high-resolution images, *Pattern Recognition and Image Analysis*, v 16, n 2, April-June 2006, p 239-48

David HA., 1956, On the application of statistics to an elementarytheorem in probability. Biometrica 43:85-91

Fan Y.M, Liu C., 2006, Computation and Matching of Road Segment Velocity With Floating Car Data Based on Road Network, *The 14th International Conference on GEOINFORMATICS, 2006, Proceedings of SPIE - The International Society for Optical Engineering*, v 6418, Geoinformatics 2006: GNSS and Integrated Geospatial Applications, 2006, p 641811

Fan Y.M., 2007, Computation of spatial road network and transportation information publication based on WebGIS and Floating Car Data, Tongji University Press, 2007

Fouladvand, M.E., Darooneh, A.H., 2005, Statistical analysis of floating-car data: An empirical study, *European Physical Journal B*, v 47, n 2, September, 2005, p 319-328

Getis, A., Ord, JK., 1996, Local Spatial Statistics: An Overview. In: Longley P, Batty M. , editors. Spatial analysis: modeling in a GIS environment. *Geoinformation International*; Cambridge: 1996. pp. 261–277.

Kerner, Boris, Prof. Dr., Rehborn, Hubert, Dr., 2001, Method to determine the traffic situation in a road network Assignee: DaimlerChrysler AG Publication Number: EP1154389 Publication date: 11/14/2001 Kind: APPLICATION WITH SEARCH REPORT

Kerner, B.S., Demir, C., Herrtwich, R.G., Klenov, S.L., Rehborn, H., Aleksic, M., Haug, A., 2005, Traffic state detection with floating car data in road networks, *IEEE Conference on Intelligent Transportation Systems, Proceedings*, ITSC, v 2005, ITSC`05: 2005 IEEE Intelligent Conference on Transportation Systems, Proceedings, 2005, p 700-705

Kwella, B., Lehmann, H., 2000, Floating car data analysis of urban road networks, Computer Aided Systems Theory - EUROCAST'99. Selection of Papers from *the 7th International Workshop on Computer Aided Systems Theory*. Proceedings (Lecture Notes in Computer Science Vol.1798), 2000, p 357-67

Moran, P.A.P., 1950, Notes on continuous stochastic phenomena, Biometrika **37**, 1950 , pp. 17–23.

Peteri, R., Celle, J., Ranchin, T., 2003, Detection and extraction of road networks from high resolution satellite images, *Proceedings 2003 International Conference on Image Processing* (Cat. No.03CH37429), 2003, p I-301-4 vol.1

Schafer, R.-P., Strauch, D., Kelpin, R., 2001, A floating-car data detection approach of traffic congestions, *Modelling and Simulation 2001. 15th European Simulation Multiconference 2001*. ESM'2001, 2001, p 958-60

Shackelford, A.K., Davis, C.H., 2003, Urban road network extraction from high-resolution multispectral data, *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*. URBAN 2003 (Cat. No.03EX646), 2003, p 142-6

Skourikhine, Alexei N., 2005, Road network extraction from digital imagery, *Proceedings of SPIE - The International Society for Optical Engineering*, v 5916, Mathematical Methods in Pattern and Image Analysis, 2005, p 1-9

Smartt, Brian E., 2006, Methods and systems for deducing road geometry and connectivity, Assignee: CircumNav Networks, Inc. Publication Number: US20060155464 Publication date: 07/13/2006 Kind: Utility Patent Application

Turksma S., 2000, The Various Uses of Floating Car Data, *Road Transport Information and Control*, Conference Publication No.472.

Zhu, C., Shi, W., Pesaresi, M., Liu, L., Chen, X., King, B., 2005, The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics, *International Journal of Remote Sensing*, v 26, n 24, Dec 1, 2005, p 5493-5508