# **Combining Float Car Data and Multispectral Satellite Images to Extract Road Features and Networks**

Chun Liu, Zhiwei Jian and Xiaolin Meng

Abstract This chapter presents an automatic methodology for the extraction of spatial road features and networks from floating car data (FCD) that was integrated with multispectral remote sensing images in metropolitan areas. This methodology is divided into two basic steps. Firstly, a spatial local statistical examination is carried out to extract the nodes of each road segment. Based on the local Moran's I statistics, a new statistic method is developed to detect local clusters. Significance is assessed using a Monte Carlo approach to determine the probability through observing large samples under the null hypothesis of no pattern. When all the necessary nodes are detected, spatial road segments can then be organized by linking pairs of nodes, which are used as the candidate road segments for the next step. Secondly, pre-processed multispectral remote sensing images are utilised for testing those initial road segments. To prove the concept, a Metropolitan area is employed as a case study. Road segments with high significance values in the tests are selected to construct the spatial road network. The developed methodology could be adopted for the provision of high quality navigational road maps in a cost-effective manner and the experimental results are presented.

Keywords Local Moran's I statistics  $\cdot$  Floating car data (FCD)  $\cdot$  Monte Carlo approach

College of Surveying and Geoinfomatics, Tongji University, Shanghai 200092, China e-mail: Liuchun@tongji.edu.cn

C. Liu (⊠) · Z. Jian

X. Meng Nottingham Geospatial Institute, The University of Nottingham, Nottingham NG7 2TU, UK

## **1** Introduction

The spatial road networks, as a fundamental component of GIS, especially in the transportation GIS (GIS-T), play very important roles both in practical applications and dedicated theoretical studies. Topics related to generation of spatial road networks are therefore discussed by many researchers. Meanwhile, multispectral remote sensing images, as a new data source for the establishment of geographical information databases, have been used to provide more detailed landscape information in recent years. Due to their unique characteristics such as short acquisition periods and large land coverage, multispectral images are widely utilised in the extraction of spatial road networks, especially from high spatial resolution images. Actually, many studies on feature extraction from multispectral imagery have been carried out. For instance, a semi-automatic approach was developed by Yang and Zhu (2010) to extract main road centre lines from high-resolution satellite images. This method is based on the active window line segment matching and an improved sequential similarity detection algorithm. Mohammadzadeh et al. (2006) designed an approach using fuzzy logic and mathematical morphology to extract main road centre lines from pan-sharpened IKONOS images. Besides, multi-scale structural features and support vector machines are applied to extract road centre line (Huang and Zhang 2009). Mokhtarzade and Valadan Zoej (2007) investigated the possibility of using artificial neural networks for road feature detection from high-resolution satellite images on a part of RGB IKONOS and Quick-Bird images. Ravanbakhsh et al. (2008) proposed a zip lock-snake approach to the extraction of road junctions from Digital Mapping Camera (DMC) ortho-images. Hu et al. (2007) presented a two-step approach that includes detecting and pruning for automatic extraction of road networks from aerial images through tracking road footprints. An Artificial Intelligence (AI) and statistics oriented methodology was also developed to output road features and road networks. A hierarchical grouping strategy was also proposed to automatically extract main road centre lines from high-resolution satellite imagery (Hu and Tao 2007). A Kohonen-type selforganizing map was applied to detect the seed points in candidate road regions and presented road tracking to search for connected points in the direction and candidate domain of a road (Yun and Uchimura 2007). Tournaire and Paparoditis (2009) proposed a top-down approach for road marks such as dashed line detection based on stochastic geometry.

Furthermore, object-oriented methodology was introduced to model road objects and road networks, in which road extraction is generally based on the properties of roads and their linked networks (Peteri et al. 2003; Dal Poz et al. 2006). Shackelford and Davis (2003) combined pixel-based fuzzy and object-based methods to extract road networks from high-resolution multispectral satellite imagery. Skourikhine (2005) proposed an image vectorization approach to road network extraction from digital imagery, which is based on proximity graph analysis. In addition, knowledge-based methodologies are also very popular. Zhu et al. (2005) extracted road networks based on the binary and grey scale

mathematical morphology and a line segment match method. Miriam et al. (2006) presented a user-guided method based on the region competition algorithm to extract roads. Lacoste et al. (2010) extended previously marked point processes developed for line network extraction to a process of manipulating polylines of variable dimension rather than segments. However, no matter whether date-driven approaches or knowledge-driven approaches are used, they all, to some extent, largely depend on the inherent character of grey scale 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 the grey scale character of features and shapes of features. In this sense, these methods all have their own limitations.

Fortunately, Floating Car Data (FCD) concept, using a dynamic sensor such as moving vehicle to collect spatial information, is currently under a rapid development. The FCD technology requires a positioning system such as the Global Position System (GPS) and a wireless communication unit installed onboard the car. The real-time position of the car is transmitted at a regular interval to the server at a data centre, which collects and processes all the GPS data packages to facilitate the determination of the traffic pattern (Liu et al. 2008). Nowadays, research on FCD mainly focuses on the application of FCD in traffic state detection (Kerner and Rehborn 2001; Schafer et al. 2001; Kerner et al. 2005; Kwella and Lehmann 2000), FCD analysis (Fouladvand and Darooneh 2005), updating the road network in an existing GIS database (Smartt 2006) and traffic information publication (Liu et al. 2008).

In this chapter, we propose a new method to integrate the advantages of multispectral remote sensing imagery and FCD for the extraction of a spatial road network. The significance of this method lies in not only helping to extract the spatial road network using FCD and multispectral RS imagery but also assisting division of a road network automatically into reasonable road segments, which are compatible with FCD. In this chapter, discussion involved is mainly on the feasibility of integrating multispectral remote sensing imagery data with FCD and the geometric registration between them. Based on local Moran's I statistics, a new statistic is defined to carry out a spatial cluster analysis to detect nodes of the road network. The Monte Carlo simulation process is adopted to evaluate the significance. This chapter will also explain the strategy used to construct a spatial road network with nodes detected and pre-processed by multispectral imagery. Experimental results are presented and discussed. Preliminary conclusions are drawn from these above tests and analysis.

#### 2 Integration of FCD with Multispectral RS Imagery

#### 2.1 FCD Approaches

No matter whether in the countryside or in urban areas, the traffic status on a road network can surely be monitored by direct measurements (e.g., induction loops and radar devices). This traditional method is effective only when there are not so many vehicles and the demand on the monitoring sensors is relatively low, because those devices could not, to some degree, monitor the dynamics of the traffic flow. With the development of GPS and wireless communication technologies, real-time positions of vehicles can be transmitted at a regular interval to a central site when they are equipped with GPS devices. With sufficient vehicles being monitored, velocities on road segments can be directly estimated. This is called floating car data or FCD (Turksma 2000). FCD have become a very important data source for the establishment of traffic information systems.

Lorkowski et al. (2005) proposed two approaches for the collection of FCD, i.e., passive and active FCD approaches. The passive FCD extraction means "recognizing" a vehicle at one section of the road network and later on another one, e.g., by automatic vehicle identification or passive onboard transponders responding to roadside stationary beacons. The time interval between the events allows an estimation of the average travel time between the two sections. The active FCD extraction method, which is widely used nowadays, requires installing a positioning system such as GPS and a wireless communication unit onboard the car. The instantaneous positions of the car are transmitted at regular intervals to a data centre. The locations of the car are then superimposed on a digital road network through map matching, and the routing velocity is further obtained through the calculation of the road segment travel time. For the above implementations, a detailed digital road map database is developed and contains well organized features with strict topological relationship to support various ITS and LBS operations on its spatial and temporal features. In Fig. 1, the logical structure of the methodology discussed in this chapter is presented. In the experiments introduced in this chapter, the active method of collecting FCD is adopted. The data are used to establish the digital road map instead of operating them on the map.

# 2.2 Geometric Registration of FCD and Multispectral RS Imagery

The positioning information of FCD sent to the traffic centre is the geodetic coordinates in the WGS84 coordinate system. However, multispectral remote sensing imagery data are always in a projected coordinate system. Therefore, to integrate these two, they must be registered into a common coordinate system in advance. In Shenzhen City, China, the "Shenzhen Urban Transport Simulation System" (SUTSS) project was developed. In this project 5,000 taxis are equipped with GPS and wireless communication devices to obtain FCD, and millions of GPS data packages have been recorded since May 2006. These FCD and the corresponding high spatial resolution images are used to study the feasibility of integrating these two data sources. The coordinate systems of these two data types are listed in Table 1.

For convenience, FCD are projected with a simple cylindrical projection to match these two data types in the same coordinate system. The selected test area ranges between 22°31′48.00″N and 22°32′24.72″N in latitude and 113°54′34.86″E



Fig. 1 The relationship of each component of the proposed approach

Table 1 Coordinate systems of FCD and RS imagery				
Data type	FCD	RS image		
Coordinate system	WGS84	Simple cylindrical projection with a WGS84 datum		

and 113°55'13.56"E in longitude. Figure 2a, b show the area before and after overlaying FCD on a high resolution image, respectively.

From Fig. 2, it can be seen that after geometric registration, FCD are projected on the image and FCD almost covers all the roads. After measuring the distances from each FCD to its corresponding central road line and performing a statistical analysis, it was found that the accuracy of FCD positions is normally within 10 m and most of the FCD have an accuracy better than 4 m. The mean deviation of the arithmetic mean is 2.28 m, sample variance is 6.92 m and the standard deviation is 2.63 m (Fan 2007). The spatial resolution of the image used is around 1 m. Thus, taking into account the width of the road, it is evident that these two data sources match very well.

## **3** Local Cluster Detection from FCD

Local statistics are a useful indicator applied in many fields. Based on Moran's I statistics (Moran 1950), the local indicators of spatial association were derived by Anselin (1995) (also see Getis and Ord 1996), to resemble passing a moving



Fig. 2 Images before and after overlaying FCD on the high resolution image. a Before the overlay, b after the overlay

window across the data, and examining dependence within the chosen region for the site on which the window is centred. The specifications for the window can vary, using perhaps contiguity or a distance at some spatial lag from the considered zone or point. With this concept, a new statistic to detect clusters from FCD is proposed.

## 3.1 Fundamentals

The local Moran statistic can be expressed by Eq. 1:

$$I_{i} = \frac{n(y_{i} - \bar{y})}{\sum_{j} (y_{i} - \bar{y})^{2}} \sum_{j} w_{ij} (y_{j} - \bar{y})$$
(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 Eq. 1 and considering the character of FCD, a new statistic is given in as follows (Eq. 2):

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

where,  $y_i$  is the number of FCD within an area centred at point *i*,  $w_{ij}$  is the weight,  $\overline{k}$  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 simulation is employed to obtain the critical value. Monte Carlo simulation of the null hypothesis of no local clustering confirmed the actual value of  $L_t$  consistent with  $\alpha = 0.05$  each time. The simulations were carried out by firstly filling the study area with randomly distributed points. The local statistics were then estimated using Eq. 2 and then the critical value for  $L_t$  was found.

Because multiple testing was employed, in which each  $L_t$  was tested whether it is significant until no significant  $L_t$  is found, to keep the experiment error rate to a specified level (usually  $\alpha = 0.05$ ), the Bonferroni adjustment was implemented (David 1956). If k independent tests are made, instead of choosing a critical value of the test statistic using  $\alpha$  as the Type I error probability for each test, simply  $\alpha/k$  can be used for each test. Therefore, it is an iteration process because parameter k of the Bonferroni adjustment is involved with the times of multiple testing. An iterative procedure is needed to approximate the optimal solution.

#### 4 Strategy to Construct a Spatial Road Network

There are two assumptions made before extracting the spatial road network.

- (a) There is a traffic pattern in the centre of the local cluster, and
- (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 because of high FCD density. Local clusters mean that there is a significant change of speed in the clustered place. Between every two clusters and along a road, vehicles 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. After finding the nodes of road segments, each two nodes can be linked as the candidate road segments. A pre-processed multispectral image is then used to determine the final spatial road network. This can be explained by as follows (Fig. 3). In Fig. 3a, seven nodes are found with the method introduced in Sect. 3. They are a, b, c, d, e, f and g. Then each two nodes can be linked and 21 candidate road segments are formed. R1, R2 and R3 are the road areas determined from the pre-processed multispectral image. The strategy used for the selection of road segments can be described by the following steps:

- (a) Decompose the vector data of candidate road segments into raster data.
- (b) Overlay the decomposed raster data with the pre-processed multispectral image.
- (c) Calculate the ratio of the number of pixels overlaid to the total number of pixels of each road segment.
- (d) Select road segments whose overlay ratio is over a certain probability.



Fig. 3 Strategy to determine the final spatial road network.  $\mathbf{a}$  Candidate road segments and road area,  $\mathbf{b}$  detected road network

In Fig. 3b, an ideal result is shown, but actually the result from the preprocessed multispectral image may contain certain levels of noise and uncertainty, which are endurable. Therefore, if areas that 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 at a given significance level. Figure 4 gives more detailed data processing procedure that is employed in this study.

#### 5 Experiment and Case Study

The testing area is as described in Sect. 2.2 and a simple cylindrical projection is used. FCD were collected from 3 to 5 June 2007. There are in total 21,935 FCD (in WGS84) during this period in this region. The high spatial resolution image was obtained from Google Earth and the spatial resolution is around 1 m.

### 5.1 The Structure of the Experiment

Fig. 4, the structure of this experiment is displayed. There are three inputs for the multiple testing: (1) FCD, (2) weight matrix, and (3) critical value. The critical value is obviously from the Monte Carlo simulation process, which was introduced in Sect. 3. After the multiple testing, road segment nodes are obtained and then candidate road segments can be produced. The selection procedure is conducted according to the discussion in Sect. 4. Finally, the spatial road network is obtained.



#### 5.2 The Weight Matrix

In Fig. 4, the weight matrix is one of three inputs for the multiple testing. Actually, weight is essentially important for the multiple testing. To define the weight matrix, the character of FCD should be studied. In Fig. 5, a corner of FCD is shown.

From Fig. 5, it is apparent that at each road intersection FCD are not distributed evenly. In the centre of intersections, there are fewer FCD but more on the roads near the centre. This does make sense, because in the real traffic situation, due to the effect of traffic lights, vehicles must wait until they are permitted to go. When vehicles are allowed to go, they must go through intersections without stopping.

Based on the area from Fig. 5, a kernel density procedure with a search radius 15 m (mean accuracy of the FCD position) and output cell size of 4.5 m is made and the result is shown in Fig. 6, in which darker areas represent higher density of FCD. From the analysis of the kernel density of the whole study area, the weight matrix is defined as follows:

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

where dist(i, j) is the distance from point i-j.



## 5.3 Monte Carlo Simulation Process

As set rules have been applied to build the weight matrix, the Monte Carlo simulation process can be carried out. Firstly, 21,935 points are randomly arranged in the study area. Secondly, a weight matrix is built with Eq. 3. Finally, the local statistic  $L_t$  is calculated with Eq. 2. These three steps are repeated 1,000 times and a series of  $L_t$  can be found. The result is shown in Fig. 7. Figure 7 shows the histogram of the Monte Carlo simulation process results. The x axis stands for the value of  $L_t$  and the y axis records the corresponding occurrence times of  $L_t$  in the Monte Carlo simulation process. Taking into consideration of the Bonferroni adjustment, critical values at confidence level  $\alpha = 0.05$  are given in Table 2.

As illustrated in Table 2, more tests could lead to higher critical values. This means if more tests are carried out, a higher critical value is needed to avoid conservative estimation.





Table ? Critical values		
considering the Bonferroni adjustment	Times of multiple testing	Critical value
	10	6.0539
	20	6.1526
	30	6.2009
	40	6.2074
	50	6.2138

# 5.4 Node Detection

After 38 tests with a critical value of 6.2061, 38 significant local clusters are found. Their points are plotted along with all FCD points in Fig. 8.

In Fig. 8, it is evident that all local clusters have been found. Besides road intersections, some clusters are located along roads, where it can be assumed that there must be some traffic patterns. This kind of traffic pattern requires those roads to be divided into road segments to comply with FCD.

# 5.5 Determination of Final Spatial Road Network

Based on the strategy discussed in Sect. 4, the pre-processed imagery of the study area should be prepared. As the aim of this chapter is to introduce FCD for the road network extraction, here the road frame is roughly described by hand, and used to explain the candidate road segment selection procedure. Figure 9 shows the preprocessed imagery and the final spatial road network overlaid with high spatial resolution imagery.



Fig. 8 Distribution of centres of local clusters



Fig. 9 Candidate road segment selection. a Pre-processed imagery, b final spatial road network

For convenience, in Fig. 9a, the pre-processed imagery is given as a binary image, in which the white area roughly covers the road area. In Fig. 9b, 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. This process will provide very useful information for the analysis of actual traffic conditions on the local

road network for the improvement of traffic flow. Actually, by applying the method on the whole Hangzhou city, the time cost of entire network features extraction with a common PC is less than 8 h. So the computing efficiency of the proposed method is proved feasibility for a practical application on a large region.

#### 6 Conclusions

Aiming at provision of large high precision digital databases for wide adoption of ITS and LBS for managing transport system, the authors of this chapter present a cost-effective approach through taking the advantages of FCD and high-resolution images. For achieving the goal, 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, the Monte Carlo simulation process is employed. The Bonferroni adjustment is also utilised to keep the experimental error rate to a specified level. In the case study, kernel density analysis was carried out to acquire the key parameter for building the weight matrix. After all the road segment nodes were detected, candidate road segments were formed successfully. Assisted by the preprocessed high spatial resolution imagery, the spatial road network was finally decided. Besides the final spatial road network obtained with this methodology, it should be noted that all the road nodes are detected based on FCD. Therefore, on the one hand, these nodes are the most compatible with FCD and on the other hand, these nodes match multispectral remote sensing imagery very well. Thus the significance of this methodology is the provision of an automatic way to construct spatial road networks that is compatible with both FCD and multispectral imagery. Using real data sets gathered with FCD cars in Shenzhen City in China, this chapter demonstrated that FCD can help to construct spatial road networks with multispectral remote sensing imagery. The short cycle period and wide coverage of multispectral remote sensing imagery make the updating of spatial road networks quicker and the cost of the update less. In future work, the more convenient way to obtain the pre-processed imagery will be investigated and the robustness of this methodology will be further addressed.

Acknowledgments The work described in this chapter was supported by National Basic Research Program of China (2012CB957702).

## References

Anselin L (1995) Local indicators of spatial association-LISA. Geogr Anal 27:93-115

Dal Poz AP, Zanin RB, Do Vale GM (2006) Automated extraction of road network from medium to high-resolution images. Pattern Recognit Image Anal 2:239–248

David HA (1956) On the application to statistics of an elementary theorem in probability. Biometrica 43:85–91

- Fan YM (2007) Computation of spatial road network and transportation information publication based on WebGIS and floating car data. MS thesis, Department of Surveying and Geo-Informatics, University of Tongji, Shanghai
- Fouladvand ME, Darooneh AH (2005) Statistical analysis of floating-car data: an empirical study. Eur Phys J B 2:319–328
- Getis A, Ord JK (1996) Local spatial statistics: an overview. In: Longley P, Batty M (eds) Spatial analysis: modeling in a GIS environment. Geoinformation International, Cambridge, pp 261–277
- Hu XY, Tao V (2007) Automatic extraction of main road centerlines from high resolution satellite imagery using hierarchical grouping. Photogram Eng Remote Sens 9:1049–1056
- Hu J, Razdan A, Femiani JC, Cui M, Wonka P (2007) Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Trans Geosci Remote Sens 12:4144–4157
- Huang X, Zhang LP (2009) Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. Int J Remote Sens 8:1977–1987
- Kerner B, Rehborn H (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 BS, Demir C, Herrtwich RG, Klenov SL, Rehborn H, Aleksic M, Haug A (2005) Traffic state detection with floating car data in road networks. In: Proceedings of the 8th international IEEE conference on intelligent transportation systems, Vienna
- Kwella B, Lehmann H (2000) Floating car data analysis of urban road networks. In: Proceedings of the 7th international workshop on computer aided system theory, Vienna
- Lacoste C, Descombes X, Zerubia J (2010) Unsupervised line network extraction in remote sensing using a polyline process. Pattern Recogn 4:1631–1641
- Liu C, Meng X, Fan Y (2008) Determination of routing velocity with GPS floating car data and WebGIS based instantaneous traffic information dissemination. J Navig 2:337–353
- Lorkowski S, Mieth P, Schäfer R-P (2005) New ITS applications for metropolitan areas based on floating car data. ECTRI Young Researcher Seminar, Den Haag
- Miriam Amo, Fernando Martínez, Margarita Torre (2006) Road extraction from aerial images using a region competition algorithm. IEEE Trans Image Process 5:1192–1201
- Mohammadzadeh A, Tavakoli A, Zoej MJV (2006) Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images. Photogram Rec 113:44–60
- Mokhtarzade M, Valadan Zoej MJ (2007) Road detection from high-resolution satellite images using artificial neural networks. Int J Appl Earth Obs Geoinf 9:32–40
- Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17-23
- Peteri R, Celle J, Ranchin T (2003) Detection and extraction of road networks from high resolution satellite images. In: Proceedings 2003 international conference on image processing, Barcelona, Spain
- Ravanbakhsh M, Heipke C, Pakzad K (2008) Road junction extraction from high-resolution aerial imagery. Photogram Rec 124:405–423
- 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, Prague
- Shackelford AK, Davis CH (2003) Urban road network extraction from high-resolution multispectral data, 2nd GRSS/ISPRS joint workshop on remote sensing and data fusion over urban areas, Berlin
- Skourikhine AN (2005) Road network extraction from digital imagery. In: Proceedings of mathematical methods in pattern and image analysis, San Diego
- Smartt BE (2006) Methods and systems for deducing road geometry and connectivity. CircumNav Networks, Inc
- Tournaire O, Paparoditis N (2009) A geometric stochastic approach based on marked point processes for road mark detection from high resolution aerial images. ISPRS J Photogram Remote Sens 6:621–631

- Turksma S (2000) The various uses of floating car data. Road transport information and control, London
- Yang Y, Zhu CQ (2010) Extracting road centrelines from high-resolution satellite images using active window line segment matching and improved SSDA. Int J Remote Sens 9:2457–2469
- Yun L, Uchimura K (2007) Using self-organizing map for road network extraction from ikonos imagery. Int J Innovative Comput Inf Control 3:641–656
- 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. Int J Remote Sens 24:5493–5508