RURAL ROAD NETWORKS MATCHING VIA EXTENDING LINE

Xiaofang Wang[a], Yu Zang[a,], Yiping Chen[a], Cheng Wang[a], Jonathan Li[a,b]*

[a]Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, FJ 361005, China;
[b]Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ABSTRACT

Road network matching has played an important role in road network extraction and update, yet has got extensive researching during the recent decades. Differ from previous road matching methods focus mainly on the city areas, which have accurate and regular road networks, this paper aim to address the matching between incomplete ground survey road network and extracted road network from remote sensing images. Specifically, we propose an extending line based matching scheme to calculate the road primitive similarity by taking into account the surrounding connections and contextual information. The experimental results show that the proposed method is able to provide high quality matching results, even the ground survey data are very different from the extracted road network of the satellite. Thus makes it possible to implement the road network update for the wide rural regions without interested ground survey road network data.

***Index Terms*—** rural road matching, similarity, line extending

1. INTRODUCTION

As the rapid improvement of informatization, road information is becoming an important part of Geospatial Information System (GIS), and it is increasingly applied to people's life, such as rural planning. Therefore, how to effectively use existing data sets to realize road data update has become a hot research problem. Road matching is one fundamental measure of road updating. Matching different road networks is essentially a process to identify the corresponding objects that represent the same real-world road in distinct data sets, which may differ in geometry, scale and coordinate.

Thus for, various methods have been proposed for matching roads in different data sets. The most commonly used data matching methods are based on geometric attributes. Geometric properties include position [1], length [2], distance [2, 3], orientation [2], shape [4, 5], topology [6] and so on. In addition, semantic properties such as the feature name and class type [7, 8] have been taken into account in a number of matching methods.

As for matching algorithms, the two most popular and well-known are Buffer Growing [9] and Iterative Closest Point [10]. In Buffer Growing, a buffer is created around an object from the reference date set. Then objects from the target data set which are inside this buffer will be considered as a potential matching candidate. The Buffer Growing is regarded as a line-based matching while the Iterative Closest Point algorithm bases on the points. Iterative Closest Point was initially developed to align two-dimensional or three-dimensional objects using a rigid transformation [10].

Most existing methods are able to provide satisfied matching results for urban area, where the roads are orderly and standardized, and the extracted road networks are well organized. But for large rural areas, road networks are messy, and even the official ground survey data can be incomplete. To address road network matching problem for rural areas, this paper proposes a novel matching scheme, which separates the road into segments. Then for each segment, we considers its surrounding sections by extending segments iteratively. The experiments verify that the method is effective.

2. METHOD

Our matching process consists of three key stages: 1) Data preprocessing. 2) Selecting potential candidates for the reference objects by utilizing **Buffer Growing** algorithm. 3) Getting the exact matching result by extending the candidate lines iteratively. Date preprocessing mainly includes topology checking, coordinate transformation, format conversion and graph construction. The aim is to reduce the noise of irrelevant details and systematic errors contained in data sets from difference sources. After these steps, the road networks become a series of line objects. Each line can be represented by two endpoints, and each endpoint in road network is actually a corner.

*Corresponding author.

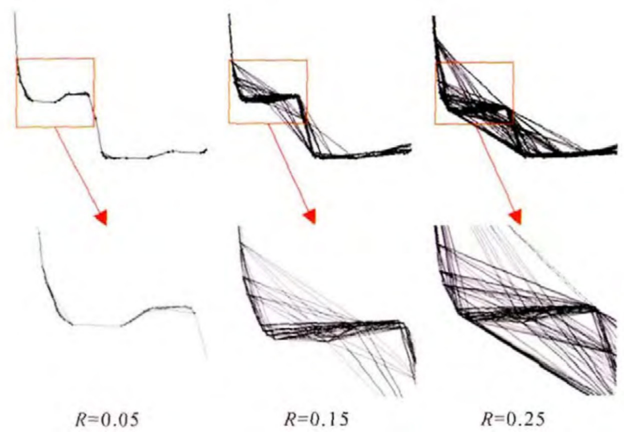


Fig. 1. Complex network model under different threshold values.

2.1. Selecting candidate segments

Buffer Growing (BG) [9] is a widely used algorithm for line matching. The process of **BG** is firstly to built a buffer around the reference line object. Then, all the line objects from the target data set that fall inside the buffer are selected. After that, similarities between the selected and reference objects are calculated. Finally, we could pick candidates from these selected objects if and only if their similarities are greater than the threshold value, which is obtained empirically.

The calculation of similarity between line objects is mainly based on geometric properties. There are many criteria related to geometric properties, but in this paper, we only take care about length, orientation, distance and shape.

Since we have the object type of linear vector, the length of the line is the Euclidean distance between two endpoints. And the orientation of the line is defined as the angle between the straight line and the horizontal axis. As for distance, we adopt **the short-line median Hausdorff distance** proposed by Tong et al [3].

Now that we have obtained candidates, we need to determine the exact matching pairs. To achieve this goal, an iterative extension line algorithm is introduced in section 2.2, in which shape becomes a primary attribute of similarities. We use the approach proposed by Wang et al [5] to calculate the shape similarities. Wang et al [5] gave a new definition of **complex network** : if the connection length between two arbitrary vertices of the curve is less than or equal to a threshold R_l, the connection is retained, and all of these retained connections form a complex network.

With different thresholds, we could obtain different complex networks. Fig. 1 shows three complex networks of the same curve with thresholds of 0.05, 0.15 and 0.25, respectively. It can be clearly seen that the number of connections in the curved and complex area is more than that in the smooth area, and the threshold larger one holds more connections.

Based on the complex network model, a new degree descriptor σ is proposed. It's a vector that contains pairs of mean degree K_a and maximum degree K_m of normalization.

$$\sigma = [K_a(1), K_m(1), K_a(2), K_m(2), ..., K_a(M), K_m(M)] \tag{1}$$

where K_a is the average value of all nodes' degrees in the network; K_m is the maximum degree of all nodes; $1, 2, ..., M$ correspond to different thresholds $R_1, R_2, ..., R_M$ and $0 \leq R_1 \leq R_M \leq 1$, $R_{j+1} - R_j = C$, and C is a constant.

So the shape similarity of two curves is calculated as follows:

$$Sim_{shape} = 1 - \frac{\sqrt{\sum_{h=1}^{M}[(K_{ia}(h) - K_{ja}(h))^2 + (K_{im}(h) - K_{jm}(h))^2]}}{\sqrt{2M}} \tag{2}$$

where M is the number of thresholds; i, j stand for different curves.

Finally, a series of appropriate weights were gave to calculate the total similarity:

$$Similarity = \frac{w_{len}Sim_{len} + w_{dis}Sim_{dis} + w_{ang}Sim_{ang} + w_{shape}Sim_{shape}}{w_{len} + w_{dis} + w_{ang} + w_{shape}} \tag{3}$$

2.2. Exactness selection of matching pairs

After **BG** process, more than one candidate could be chosen. In order to get exact matching results, an iterative extension line algorithm is introduced in this section.

Extending line (polyline): Take the existing line as center and extend it to the surround. The step size of extending is a line segment. For example, as shown in Fig. 2, the gray line $R_{A0}R_{B0}$ is the original line of reference date set, the red lines are increased parts after the first extension and the green lines represent the second extension result. To differentiate, we name the extended line as polyline. If we only concern similarity between lines, it's obvious that line $T_{C0}T_{D0}$ is more similarity to line $R_{A0}R_{B0}$ than line $T_{A0}T_{B0}$. But in practice, line $T_{A0}T_{B0}$ is the correct match for $R_{A0}R_{B0}$. This demonstrate that polyline has a better description for the overall structure of road.

This algorithm starts with an extended operation for each candidate line and the reference line object. Then the new similarity between the reference polyline and the candidate polyline is calculated based on length, distance and shape. Polyline's length is the sum of the length of the line segments that it contains. After that, top k polylines are selected as new candidates according to the similarities, where k

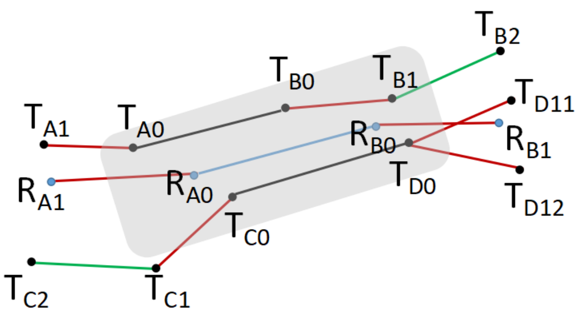


Fig. 2. Extending lines.

is an adaptive variable. For example, we have a sequence of similarities $\{Sim_{i,1}, Sim_{i,2}, ..., Sim_{i,m}\}$ that have been sorted in descending order, i.e. $Sim_{i,1} \geq Sim_{i,2} \geq ... \geq Sim_{i,m}$. Set j to start at 1, and when the first time meet $Sim_{i,j} - Sim_{i,j+1} \geq T_s$, T_s is an empirical threshold, then the value of k equals j. This algorithm ends when k equals to 1, otherwise, performs iterations until the most similar object is found. The more detailed overview is presented in Algorithm 1.

Algorithm 1 Exactness selection of matching pairs.

```
for each reference object l_i do
    The initial value of k is the sum of candidates.
    while k is greater than 1 do
        Extend the reference object l_i.
        for each candidate object l_j do
            Extend the candidate object l_j.
            Calculate the similarity Sim_{i,j} between the can-
            didate object l_j and the reference object l_i.
        end for
        Sort the new similarities
        {Sim_{i,1}, Sim_{i,2}, ..., Sim_{i,m}} in descending order.
        Update the value of k.
    end while
    Get the exact matching result for reference object l_i.
end for
```

3. EXPERIMENTS AND ANALYSIS

In order to verify the proposed matching algorithm, this paper takes experiments in Shaoshan city, Hunan province. Shaoshan is a typical mountainous city covering about 250 km^2. The road network data sets of Shaoshan we adopt are from two different sources. One is road vector extracted from the remote sensing image of Shaoshan according to the method introduced in literature [11]. The image is recorded by Pleiades-1A satellite with resolution 0.5 m and it's size is 28648 $*$ 37929 pixels. The other is the ground survey data set from China Transportation & Telecommunication Center. Road in this data set is continuous but incomplete, and it is regarded as reference data set in this paper. Fig. 3 shows a part of the used data sets of ShaoShan City. The pink lines are the target data set and the green lines represent the reference data set. It is clear that the target data set has more details than the reference one, beside this, significant location differences are existing between them.

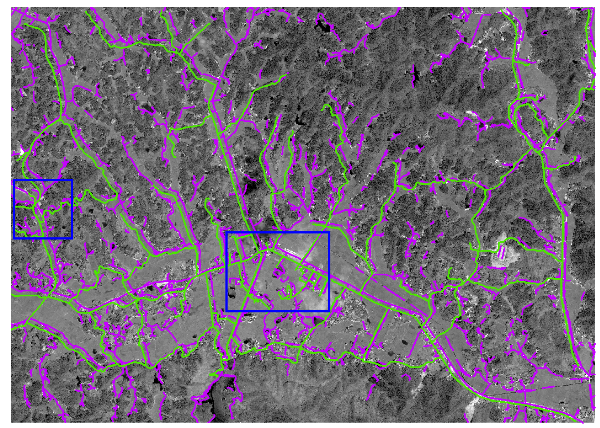


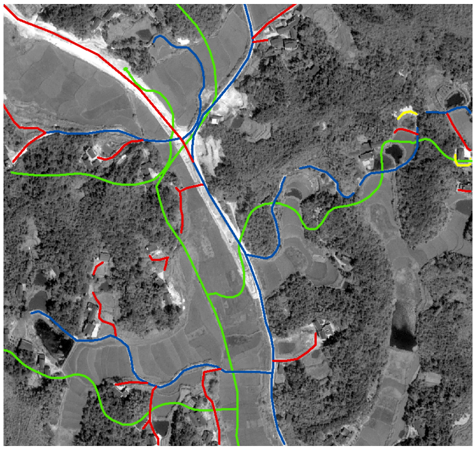
Fig. 3. Data sets utilized in this paper, where green lines are the reference data set and pink lines are target data set.

Fig. 4 are two partial enlarged view of blue rectangles in Fig. 3. In Fig. 4, the blue lines mean correct matched road objects of target data set, the yellow lines mean false matched road objects on the contrary. Beside that, the red lines represent unmatched road of target data set. The reference data set is showed in green lines. Specially, at the top left corner of Fig. 4(a), the red line indicates the missing part of the reference data set, but it can be seen from the remote sensing image that this is actually a real road section. In Fig. 3, there are totally 13640 line segments in the target data set and 6285 segments in the reference data set. 9042 of them are correctly matched to the reference data set and 1251 lines are mismatched. 3347 lines cannot matched to reference data set.

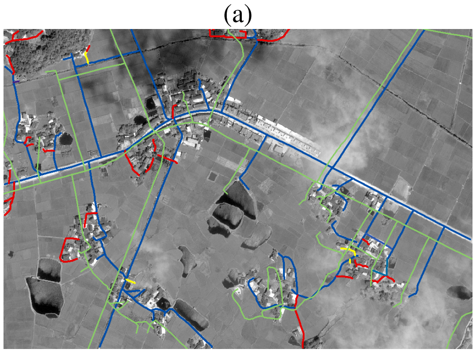
The statistical matching results of whole Shaoshan area are showed in Table 1. The total length of the reference road network is 908.45 km. Since the used two data sets have greatly difference in detail, the total matching rate is only 74.73%, but the rate of correctly matched pairs reaches 87.24%, which is far beyond literature [2]'s 83.41% accuracy. That means our new matching method is more accurate.

Table 1. Statistical results of the road matching.

Data sets to be matched	The target data set	The reference data set
Total amount of line objects	168453	77527
Matching rate	74.73%	
Accuracy	87.24%	



(a)



(b)

Fig. 4. Matching results of two randomly selected area.

4. CONCLUSION

This paper proposed a new road matching method for rural road networks, which mainly focused on extending lines iteratively. By this way, we could consider the similarities between potential matching pairs in the overall network structure rather than single line segments, that guaranteed high accurate matching rate than other common methods. Meanwhile, the experimental result of 87.24% accuracy also proves the feasibility of this method.

5. REFERENCES

[1] Ehsan Abdolmajidi, Ali Mansourian, Julian Will, and Lars Harrie, "Matching authority and vgi road networks using an extended node-based matching algorithm," *地球空间信息科学学报(英文版)*, vol. 18, no. 2-3, pp. 65–80, 2015.

[2] Meng Zhang and Liqiu Meng, "Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks," *Geographic Information Sciences*, vol. 14, no. 1, pp. 44–53, 2008.

[3] Xiaohua Tong, Dan Liang, and Yanmin Jin, "A linear road object matching method for conflation based on optimization and logistic regression," *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 824–846, 2014.

[4] Remco C. Veltkamp, "Shape matching: Similarity measures and algorithms," in *International Conference on Shape Modeling & Applications*, 2001, p. 188.

[5] Hao Wang, Renjian Zhai, Minghui Zhou, and Li Zhu, "A road matching method based on complex networks," *Journal of Geomatics Science & Technology*, 2016.

[6] Guillaume Touya, Adeline Coupé, Jérémie Le Jollec, Olivier Dorie, and Frank Fuchs, "Conflation optimized by least squares to maintain geographic ? shapes," *ISPRS International Journal of Geo-Information*, vol. 2, no. 3, pp. 621–644, 2013.

[7] William W. Cohen, Pradeep Ravikumar, and Stephen E. Fienberg, "A comparison of string distance metrics for name-matching tasks," in *International Conference on Information Integration on the Web*, 2003, pp. 73–78.

[8] Max J. Egenhofer, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure," *International Journal of Geographical Information Science*, vol. 18, no. 3, pp. 229–256, 2004.

[9] Volker Walter and Dieter Fritsch, "Matching spatial data sets: a statistical approach," *International Journal of Geographical Information Systems*, vol. 13, no. 5, pp. 445–473, 1999.

[10] P. J. Besl and Neil D. Mckay, "A method for registration of 3-d shapes," in *Robotics - DL tentative*, 1992, pp. 239–256.

[11] Yu Zang, Cheng Wang, Liujuan Cao, Yao Yu, and Jonathan Li, "Road network extraction via aperiodic directional structure measurement," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 54, no. 6, pp. 3322–3335, 2016.