



# A new Voronoi diagram-based approach for matching multi-scale road networks

Jianhua Wu<sup>1</sup> · Yu Zhao<sup>1</sup> · Mengjuan Yu<sup>2</sup> · Xiaoxiang Zou<sup>3</sup> · Jiaqi Xiong<sup>1</sup> · Xiang Hu<sup>1</sup>

Received: 18 March 2022 / Accepted: 22 March 2023 / Published online: 26 April 2023  
© The Author(s) 2023

## Abstract

Object matching is a key technology for map conflation, data updating, and data quality assessment. This article proposed a new Voronoi diagram-based approach for matching multi-scale road networks (VAMRN). Using this method, we first created Voronoi diagrams of the road network using the strategy of discretizing road lines into points and adding dense points to special road intersection segments. Then, we used the Voronoi diagram of road segment to find matching candidates. Finally, we obtained matching results by judging the geometric similarity metrics we designed and a heuristic combination optimization strategy. The experimental results demonstrated that the VAMRN outperformed two existing methods in generality and matching quality. The F-measures of VAMRN were 18.4, 29.6, 3.8, and 7.6% higher than the buffer growing method, and 4.5, 2.8, 1.8, and 6.1% higher than the probabilistic relaxation method. And the time performance is improved by more than 90% over the probabilistic relaxation method.

**Keywords** Voronoi diagram · Identical roads · Road update · Road network matching · Data integration

**JEL Classification** C60 · C61

## 1 Introduction

The road network is the connecting framework of cities and regions, and it plays a key strategic role in promoting regional economic development and improving people's living standards. At the same time, road network data are also the indispensable core data for emerging fields such as Location-based Services (LBS), Smart Navigation, and Social Networking Services (SNS) (Luan 2013). With

---

✉ Yu Zhao  
zy1352678154@163.com

Extended author information available on the last page of the article

the rapid advancement of the technologies like Geographic Information System (GIS), Remote Sensing (RS), Global Navigation Satellite System (GNSS), and social media, it is much easier to produce, distribute, and utilize digital geospatial information.

However, the spatial data collected by different departments have different application purposes, leading to duplicate collection of spatial data in the same area. And these data possibly exist large geometric difference due to map scale, image resolution, time, compilation standards, data accuracy, alignment, sensor characteristics, or error (Guo 2008). In order to keep the spatial data current and eliminate the differences between spatial data, it is necessary to update and fuse the spatial data in a reasonable way. While object matching is a basic and key step in the process of data updating and fusion. Object matching means that, through a series of similarity measures, to distinguish identical objects from different data sources, and then built the corresponding relations for related spatial objects (Fu et al. 2008). It usually includes match types of 1:1, 1:N, M:1, M:N, 1:0, and 0:1. M:N match type means that M objects in one dataset match N objects in another dataset. Table 1 shows the examples of the matching pair categories.

Therefore, it is of great research value and application significance to study road network matching. For example, Fig. 1 shows that two overlaid road datasets covering the same area with different map scales have obvious positional discrepancies. If you want to conduct the conflation of geometry or attribute information

**Table 1** The examples of the matching pair categories

Matching type	point	polyline	polygon
1 : 1			
1 : N			
M : 1			
M : N			
1 : *(0/Null)			
*(0/Null) : 1			



**Fig. 1** The non-systematic discrepancies between road dataset at 1:10,000 map scale and road dataset at 1:50,000 map scale

of the identical roads from two datasets, you must use object matching technology to recognize the identical objects.

However, due to the multi-source, multi-scale, multi-temporal of road network data, as well as relatively complex structure itself, it is relatively difficult to match multi-scale road network. The existing road network matching algorithms still have the problems of low matching automation or low computational efficiency. For example, the existing buffer-based method for searching matching candidates needs manually setting the buffer radius and can cause missing matches or computation workload. Therefore, there still exist challenges to implement a generic, automatic, and efficient matching of multi-source and multi-scale road networks.

In this paper, we present a novel multi-scale road network matching method based on Voronoi diagram (VAMRN). Our matching method can handle geographic datasets without attributes or having significant attribute differences (e.g., the difference in schemas, naming, or coding conventions), even the road networks that have large non-consistent positional discrepancies, and have higher matching accuracy. The VAMRN method innovatively uses Voronoi diagrams of road segments to find candidate matching road segments. Voronoi diagram has more advantages in spatial analysis and partitioning. It can establish the correlation between multi-source spatial data based on spatial location. When applied to object matching, it can effectively avoid manually setting the buffer radius used for searching candidates and prune the matching searching space. By combining geometric similarity metrics, such as length, shape, and direction, as well as a heuristic combination optimization strategy, which can effectively achieve identical road matching and improve the algorithm generality and matching quality. This method is conducive to solving the problem of integrating and updating multi-source road data.

## 2 Related work

The purpose of road network matching is to find out corresponding objects from different source datasets and to establish their association relationships. In general, two important factors that determine the merits of road network matching algorithms are similarity characteristic and matching model. Both of them have a significant impact on the matching accuracy of the overall matching algorithm. Many researchers have conducted related studies on road network matching.

In the aspect of study on similarity metric, the geometric similarity metric of line objects can be roughly judged and filtered in terms of characteristics such as distance, direction, length, and shape. Distance similarity characteristic is often used to measure correlation between corresponding objects. Gabay and Doytsher (1994) proposed a method for matching linear objects by using the distance between the points (vertices or nodes) and the angle between the line segments to be matched as a spatial constraint. Zhang (2002) proposed a method for calculating the distance using the middle area of two line objects, and designed a shape similarity of line objects using the change of direction of each line segment in each line object. By extending the Hausdorff distance algorithm, Deng et al. (2007) proposed a line object matching method based on the extended Hausdorff distance. Yang (2016) proposed a mixed-median Hausdorff distance to achieve a reasonable distance measure between line entities of varying length and shape. The directional similarity characteristic is usually used to determine the difference in the overall orientation of the objects. Walter and Fritsch (1999) proposed a probabilistic statistical matching method based on “buffer growing” to obtain the candidate matching set, specifically it uses the statistical properties of angles, lengths, shapes, and topological characteristics between line objects to determine the threshold value of confirming corresponding objects, and uses the dominance function in information theory to calculate the optimal matching results. The length similarity metric of line objects is mainly used to detect the size difference between two objects by calculating the length ratio of the two objects. The shape characteristics of line objects mainly include curvature, tightness, orthogonality, etc. Shape similarity metric of two objects is used to determine the deformation difference between two objects. Zhang et al. (2002) proposed a method for determining object shape difference by the comprehensive change in the directional variation of each line object for each road segment. Topological similarity metric is a mathematical method used to determine the spatial structure relationship between neighboring line objects (Wu 2008). The semantic similarity characteristics of line objects refer to the attribute information of each spatial object, such as naming, encoding, usage, and type. It is the most accurate and efficient matching similarity metric in matching, mainly by analyzing the difference of attribute information of line objects to measure the semantic difference.

In terms of automatic matching algorithm models of road networks, many scholars have tried to introduce models from other domains into the matching algorithm. The most widely used buffer growth method was proposed by Walter et al. (Walter and Fritsch 1999). By setting a reasonable buffer radius for

the dataset, the line objects that meet the distance, length, and direction similarity indexes are marked as candidate matching roads as the buffer continues to grow, and then the optimal matching pair is obtained by a probabilistic calculation method. Volz (2006) extended the Walter's algorithm by proposing an iterative node and edge matching algorithm applying relaxed constraints based on detection of similarity measures. Tong et al. (2007) proposed a probabilistic theory-based spatial data matching algorithm. The algorithm combined multiple similarity characteristics and determined whether two objects match by the matching probability. Li and Goodchild (2011) developed a new optimization model to improve linear feature matching. The model takes into account all potentially matched pairs simultaneously by maximizing the total similarity of all matched features, nevertheless, the matching efficiency is low. Zhang et al. (2012) proposed an automatic matching method for urban road networks based on a probabilistic relaxation method, which first estimates the initial probability of candidate road sections by the geometric difference among road segments; then, it continuously updates the original probability matrix until it converges to a certain minimal value based on the compatibility of neighboring candidate matching road segments; finally, it calculates the structural similarity of each candidate road segment based on the converged probability matrix, and selects and refines matching pairs by setting the corresponding rules. Chehreghan and Abbaspour (2017) used real coding genetic algorithm (RCGA) and sensitivity analysis for target identification based on consideration of geometric criteria. Their method eliminates the initial dependency on empirical parameters such as buffer distance, spatial similarity threshold, and weights of criteria; instead, the optimal values of these parameters are calculated based on input dataset. However, the method is time-consuming and there is biased in the training data. Zou et al. (2020) proposed a hierarchical matching method based on Delaunay triangulation for solving road network matching under different or unknown coordinate systems. Lei (2021) proposed a feature matching framework based on optimization and divide-and-conquer. His research was mainly conducive to improving the computational efficiency; however, he paid little attention to the matching accuracy, and the matching datasets used were also less complex.

To sum up the above methods, we find that (1) many methods require an empirically setting buffer distance for searching candidate matching elements. However, a small buffer distance causes missing the identical objects, and a large buffer distance increases computation time while reducing the matching efficiency; (2) many heuristic methods, such as probabilistic relaxation methods and genetic algorithms, have complex and massive calculation process that results in much lower efficiency in complex road network matching. Therefore, the practicability of these algorithms needs to be improved; (3) many of the previous studies are efficient in datasets with small geometric differences or consistent positional discrepancies, while they cannot be applied to matching datasets with significant inconsistent positional discrepancies. And they rarely simultaneously meet the requirements of urban road matching and mountain road matching.

Because Voronoi diagram has the unique properties in spatial proximity and spatial segmentation, it is widely used to describe spatial proximity, closest operation,

and scope of spatial influence (Hao 2010), and build the spatial corresponding relationship for multi-scale objects (Wu et al. 2018). So we can utilize the Voronoi diagram to query the candidate matching roads of different scales. In the study of the spatial analysis of Voronoi diagrams, Chen et al. (2003) noted that digital map synthesis and update is a direction for further research on Voronoi diagrams. In recent years, Voronoi diagrams have been applied preliminarily in point matching (Wu and Wan 2015; Ma 2020), line matching (Hu and Mao 2011; Yu 2017), and polygon matching (Wu et al. 2018). However, there is relatively little literature on road network matching using Voronoi diagrams. Compared with the buffer searching method used in many studies like Walter and Fritsch (1999) and Yang et al. (2013), Voronoi diagram has the advantages of establishing the relationship of corresponding objects from multi-source and multi-scale spatial data based on spatial location and controlling the selection range of the candidate matching set within a reasonable neighborhood area, thereby avoiding the matching error caused by the manually unreasonable selection of the buffer radius threshold. Therefore, we proposed a new multi-scale road network matching method based on Voronoi diagram (VAMRN).

### 3 Methodology

VAMRN consists of three processes: Voronoi creation, candidates acquisition, and matching pair selection. Figure 2 preliminarily illustrates the VAMRN process.

- *Voronoi creation*: Includes discretizing each road segment of the road network into a point set, creating Voronoi diagrams of the point set, and merging Voronoi diagrams by point set of each road segment.

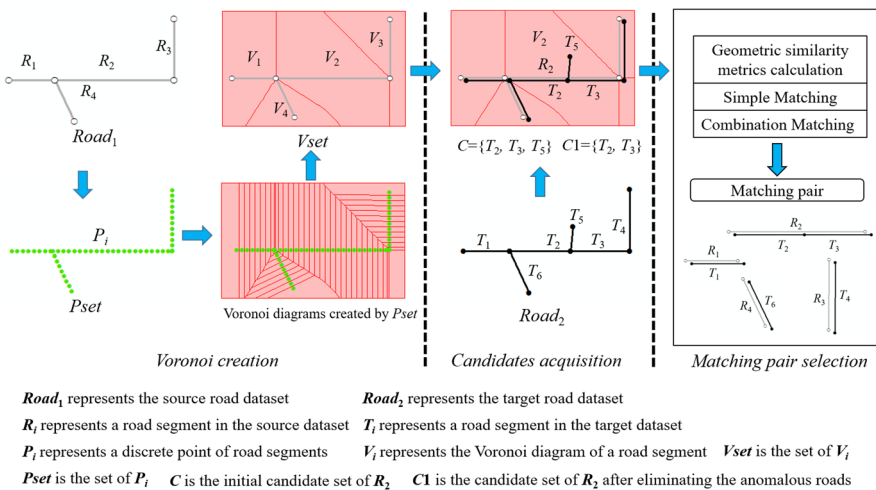


Fig. 2 Schematic diagram of VAMRN process

- *Candidates acquisition*: Includes obtaining potential candidate matching road segments and obtaining the candidate matching set after removing the anomalous road segments.
- *Matching pair selection*: Includes geometric similarity metrics calculation, simple matching, and combinatorial matching.

As shown in Fig. 2, let the two datasets to be matched be  $Road_1 = \{R_i \mid i=1, 2, \dots, m\}$  and  $Road_2 = \{T_j \mid j=1, 2, \dots, n\}$ , in which  $R_i$  represents each road segment in the source dataset and  $T_j$  represents each road segment in the target dataset. Based on the idea of occurrence element discretization, the Voronoi diagrams of points are firstly constructed by the point family  $Pset = \{P_i \mid i=1, 2, \dots, m\}$  into which VAMRN discretizes the road network, where  $P_i$  is a point of  $R_i$  discretization. Then, the Voronoi diagrams of  $Pset$  are merged into the Voronoi diagrams of  $Vset = \{V_i \mid i=1, 2, \dots, m\}$ , where  $V_i$  is the Voronoi diagram of  $R_i$  (Sects. 3.1). The initial candidate matching road set ( $C$ ) of  $R_i$  is then obtained by the intersection relationship between  $V_i$  and  $Road_2$ , and the candidate matching road set ( $C1$ ) of  $R_i$  is obtained after the operation of eliminating the anomalous roads (Sect. 3.2). Finally, depending on the number of elements in  $C1$ , simple matching and combinatorial matching are performed, and the geometric similarity metrics are calculated to determine whether a pair is a match or not (Sects. 3.3).

### 3.1 Voronoi creation

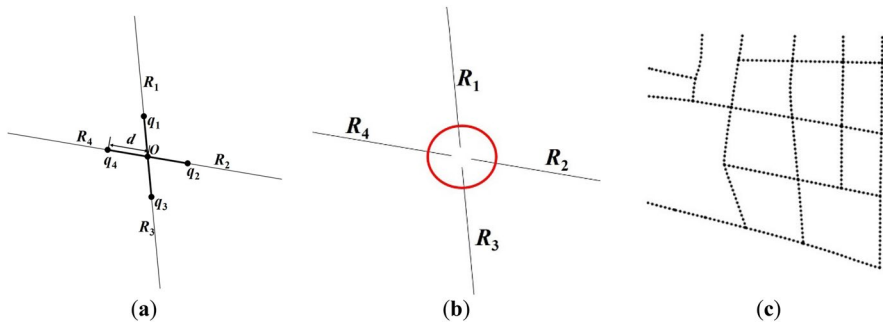
Because VAMRN requires the use of Voronoi diagrams of road segments to obtain the candidate matching set, the first step is to create Voronoi diagrams of the road network. We designed an algorithm for creating Voronoi diagrams of road network based on the idea of occurrence element discretization. This involved first breaking road network data at intersections; discretizing each road segment into a set of points that can replace the road segment; creating Voronoi diagrams of point sets for all road segments; and finally obtaining a Voronoi diagram for each road segment by merging the Voronoi diagrams generated by the corresponding point set of each road segment.

#### 3.1.1 Discrete point set construction method for road network

Discretizing road network into point set means that using a finite number of discrete points with equally spaced distance to replace each road segment in road network. Its main steps are as follows.

First, interrupting road polylines at intersections. VAMRN automatically interrupts all road polylines in two road network datasets at road intersections to obtain  $Road_1 = \{R_i \mid i=1, 2, \dots, m\}$  and  $Road_2 = \{T_j \mid j=1, 2, \dots, n\}$  ( $m$  and  $n$  denote the FID (unique identification number of the feature that is an object including geometry and attribute information) of each feature in  $Road_1$  and  $Road_2$ , respectively). Meanwhile, using the attribute field  $FID\_O$  of  $R_i$  (or  $T_j$ ) to record the FID value of its original road segment.





**Fig. 3** Discrete processing of road network ( $R_i$  is a road segment;  $O$  is the intersection;  $q_i$  is the offset point;  $d$  is the offset distance): **a** Insert the offset point; **b** Intersection point processing; and **c** The discrete point set  $P$

Second, endpoint offset processing. To avoid an error in creating the Voronoi diagram of road segment caused by the overlap of adjacent road endpoints, the endpoints need to undergo slightly offset processing. As shown in Fig. 3(a),  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$  are intersect at the point  $O$ , and then, the intersection  $O$  of each road is offset at a distance  $d$  (such as 1 m (meter)) along the  $Oq_n$  ( $n=1,2,3,\dots$ ) direction of each road. The schematic effect of the road segments after the endpoint offset is shown in Fig. 3(b). The red circle shows the effect of the intersection after the offset of the end points of each road segment.

Last, creating discrete points. First, creating a new discrete point layer for storing the discrete points of each road segment. Second, using any endpoint of each road segment in  $Road_1$  as the starting point, then adding new points along road segment  $R_i$  at intervals of distance  $g(\lambda)$  which is obtained by the calculation of Eq. (1) to the discrete point layer. In Eq. (1), MapScale ( $S$ ) is the denominator value of the map scale of the dataset  $S$ , MapScale ( $S$ ) divided by 1000 indicates the actual length of the surface corresponding to 1 mm on the map,  $1/10$  indicates the ratio at which 1 mm distance on the drawing can be recognized by the human eye (Pan 2004), and  $\lambda$  ( $1 \leq \lambda \leq 10$ ) indicates the distance tolerance factor.

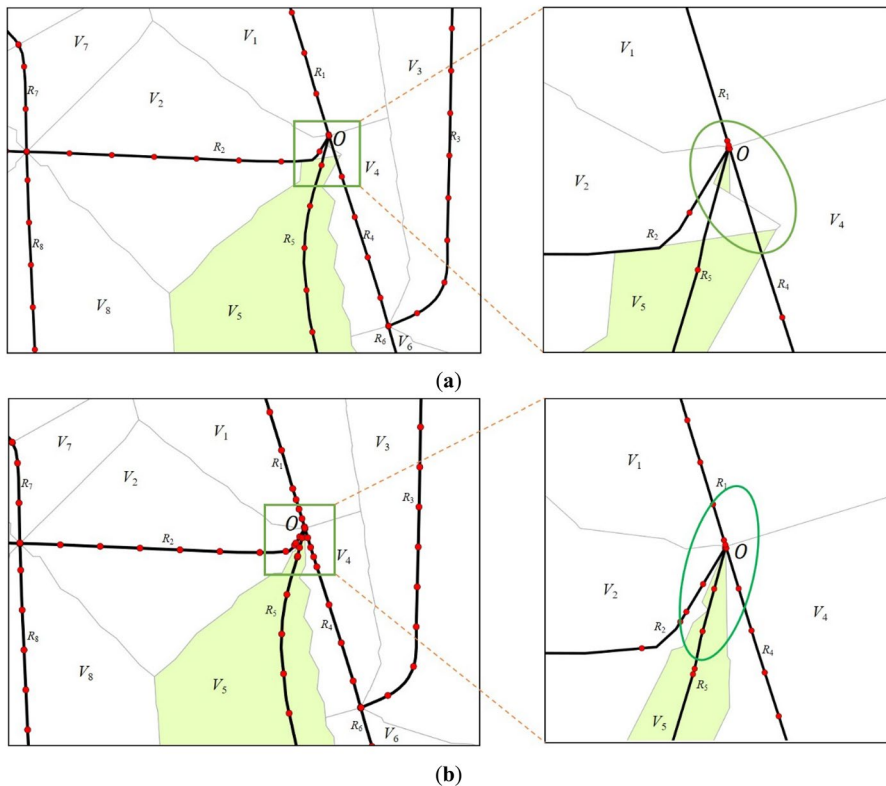
$$g(\lambda) = \lambda * \frac{\text{MapScale}(S)}{1000} * \frac{1}{10} \quad (1)$$

Meanwhile, the two endpoints and turning points of road segment  $R_i$  are added to the discrete point layer. At the same time, using attribute field  $FID\_O$  of each discrete points to store the FID of the road segment where the point is located in preparation for Voronoi diagram merging. Referring to the triple standard deviation principle, normally,  $\lambda$  takes the value of  $1 \leq \lambda \leq 3$ . Theoretically, the smaller the value of  $\lambda$ , the smaller the spacing of the discrete points, and the more discrete points are generated. We take into account the data error limit and the efficiency of creating Voronoi diagram,  $\lambda$  is set as 4 in this paper. Now we can obtain the point sets  $P_i$  of each road segment and the number of inserted points  $N_i$  of each road segment, where  $i$  denotes the FID of each road segment. Finally, all  $P_i$  are combined into  $Pset$ , as shown in Fig. 3(c).



### 3.1.2 Special intersection identification and addition of dense point processing

In a preliminary study of creating Voronoi diagrams of a road network, we found that when the angle size between two road segments was too small, the Voronoi diagrams of road segments generated at the intersection points were easy to cross and deform. As shown in Fig. 4(a),  $R_1, R_2, R_3$  and  $R_4, \dots, R_7, R_8$  denote several road segments to be matched in the source road set,  $V_1, V_2, V_3, \dots, V_7, V_8$  denote the Voronoi diagrams created by each road segment to be matched, the intersection  $O$  selected by the green box is a special intersection, and  $V_5$  is the Voronoi diagram with crossing and deformation. To solve the noted problems, we proposed a method for the identification of special intersections and the addition of dense points in the nearby road segments. When the angle between the two line segments, respectively, of two road segments that intersect at a point was less than  $45^\circ$  (this angle threshold is independent of map scale), the Voronoi diagrams of road segments generated at the angle were likely to experience these problems. Therefore, such intersections



**Fig. 4** Identification and processing of special intersections ( $R_i$  represents a road segment in the dataset;  $V_i$  represents the Voronoi diagram of a road segment;  $O$  is the intersection): **a** Error Voronoi diagram of road segment before adding dense points; and **b** Voronoi diagram of road segment after adding dense points

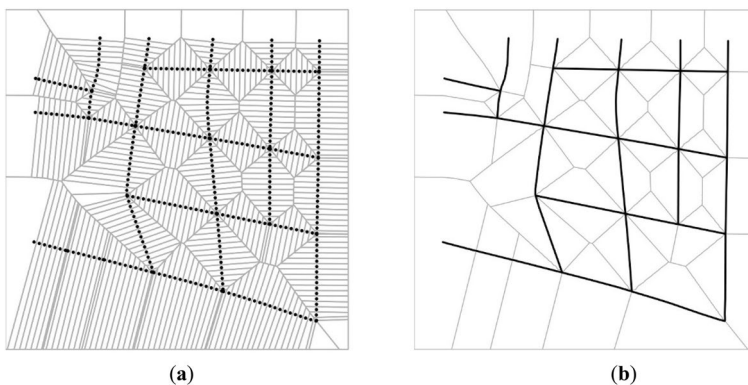
were marked as special intersections, and the road segments connected to such intersections were further processed by adding dense points to the original discrete point set. The distance of increasing the dense points is  $g(\lambda)$  ( $\lambda = 1$ ), and the effect of the Voronoi diagram of road segment generated after adding the dense points processing is shown in Fig. 4(b).

### 3.1.3 Building Voronoi diagrams of road networks

First, we created Voronoi diagrams of points using discrete point set  $P$ , as shown in Fig. 5(a). Next, we transformed Voronoi diagrams of the point set into Voronoi diagrams of road segments, that is, merging Voronoi diagrams generated by points belonging to the same point set into a Voronoi diagram of road segment. The result of the merging is the Voronoi diagram of all road segments in the road network is shown in Fig. 5(b).

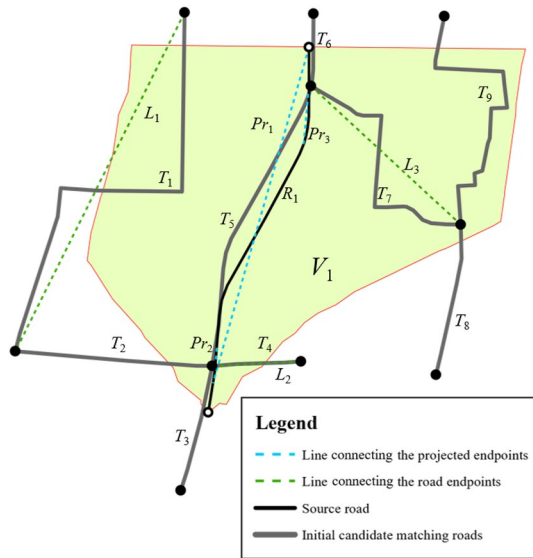
## 3.2 Candidates acquisition

Initially, we obtained the candidate matching road segments set  $C$  by traversing  $V_i$  in  $Vset$  and finding the road segments intersecting with  $V_i$  in  $Road_2$ . Then, some of the anomalous candidate matching road segments were eliminated by judging the size of  $\theta_{Proj}$ , ( $0 < \theta_{Proj} < 90$ ) (Zhang 2002), and  $dis_H$ , where  $\theta_{Proj}$  is the angular difference between the line connecting the first and last endpoints of each road segment in  $C$  and the line connecting the projection points of its two endpoints on  $R_i$ , and  $dis_H$  is the mixed-median Hausdorff distance between each road segment in  $C$  and  $R_i$ , which are marked as anomalous candidate matching road segments when  $\theta_{Proj}$  is greater than  $35^\circ$  or  $dis_H > g(\lambda)(\lambda = 10)$ . However,  $35^\circ$  and 10 are empirical values in spatial cognition that do not depend on map scale. As shown in Fig. 6,  $V_1$  is the Voronoi diagram generated by  $R_1$ ,  $C = \{T_1, T_2, T_3, \dots, T_9\}$ ;  $L_1, L_2$ , and  $L_3$  are the lines connecting the first and last endpoints of  $T_1, T_4$ , and  $T_7$  segments, respectively; and  $Pr_1, Pr_2$ , and  $Pr_3$  are the lines connecting the endpoints of  $T_1, T_4$ , and  $T_7$  on  $R_1$ ,



**Fig. 5** Process of creating a Voronoi diagram of a road network: **a** Voronoi diagram of the point set; and **b** Voronoi diagram of the road network

**Fig. 6** Process of obtaining candidate matching set ( $R_i$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset;  $V_i$  represents the Voronoi diagram of  $R_i$ ;  $L_i$  represents the line connecting the start point and end point of  $T_i$ ;  $Pr_i$  is the line connecting the projection points on  $R_i$  of the start point and end point of  $L_i$ )



respectively. The projection point is the point on source road closest to the target road endpoint. As shown in Fig. 6, although  $L_1$  and  $Pr_1$  satisfy  $\theta_{proj} < 35$ , the  $dis_H$  between them is greater than  $g(10)$ ,  $T_1$  is marked as an abnormal candidate matching road segment.  $\theta_{proj} > 35$  between  $L_2$  and  $Pr_2$ , as well as  $L_3$  and  $Pr_3$ , respectively, so  $T_4$  and  $T_7$  are also marked as abnormal matching road segments. After eliminating all the abnormal road segments,  $C = \{T_3, T_5, T_6\}$  significantly reduced the amount of computation in the matching process, improved matching efficiency, and reduced the number of false matches. Finally, we put the remaining road segments in  $C$  into the set  $C1$ , and  $C1 = \{T_3, T_5, T_6\}$  was the candidate matching set of  $R_i$ .

### 3.3 Matching pair selection

After obtaining the candidate matching set  $C1$  of  $R_i$ , we selected the corresponding objects of  $R_i$  from the candidate matching set, that is, the process of building road matching pairs. To select the final matching pairs, such as 1:0 (or 0:1), 1:1, 1: N, and M:N from the entire dataset, we divided the matching process into two cases (simple matching and combination matching) according to the number of road segments contained in  $C1$ , and determined whether they were identical objects by calculating the geometric similarity metrics between two road segments or combinators.

#### 3.3.1 Geometric similarity calculation and identification method of corresponding objects

Whether or not two road segments or road combinations are corresponding objects can be determined after calculating the geometric similarity metrics between the two road segments or road combinations. The metrics for evaluating the difference of two linear features are typically distance, length, shape, direction, semantics, and

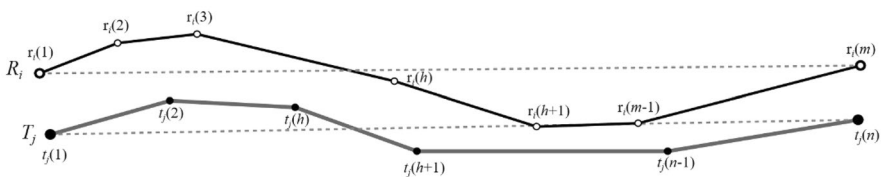
topology. Distance metrics are usually not suitable for matching data from road networks that are too dense and have large geometric discrepancies. Since road network datasets from multiple sources, especially for the VGI (Volunteered Geographic Information) data, usually have incomplete attributes or have different descriptions for the same attribute of corresponding objects, the semantic information is also difficult to be adopted. For multi-scale road network data matching, the topological characteristics of road with the same name may change due to the different levels of detail of data at different scales, so the topological similarity metric is not chosen in this paper. Therefore, considering the adaptability of similarity metrics, we designed length (Yang 2016), shape, and direction (Tong, et al. 2007) similarity metrics to measure the geometric difference characteristics among road segments. As shown in Fig. 7,  $T_j$  is a road segment in the  $R_i$  candidate matching set,  $R_i$  and  $T_j$  can be represented as a series of sequential nodes  $R_i = \{r_i(1), r_i(2), r_i(3), \dots, r_i(m)\}$ ,  $T_j = \{t_j(1), t_j(2), t_j(3), \dots, t_j(n)\}$ , the two endpoints of  $R_i$  are  $r_i(1), r_i(m)$ , and the two endpoints of  $T_j$  are  $t_j(1), t_j(n)$ . The length, shape, and direction similarity metrics of the candidate matching roads  $R_i$  and  $T_j$  are calculated according to Eqs. (2), (3), and (4), respectively.

$$\text{sim}_{\text{len}}(R_i, T_j) = \frac{\min \left( \sum_{h=1}^m |r_i(h)r_i(h+1)|, \sum_{h=1}^n |t_j(h)t_j(h+1)| \right)}{\max \left( \sum_{h=1}^m |r_i(h)r_i(h+1)|, \sum_{h=1}^n |t_j(h)t_j(h+1)| \right)}, \tag{2}$$

$$\text{sim}_{\text{shp}}(R_i, T_j) = \frac{\min \left( |r_i(1)r_i(m)|, |t_j(1)t_j(n)| \right)}{\max \left( |r_i(1)r_i(m)|, |t_j(1)t_j(n)| \right)}, \tag{3}$$

$$\text{sim}_{\text{dir}}(R_i, T_j) = dA(r_i(1)r_i(m), t_j(1)t_j(n)), \tag{4}$$

where  $|r_i(h)r_i(h+1)|$ ,  $|t_j(h)t_j(h+1)|$ ,  $|r_i(1)r_i(m)|$ , and  $|t_j(1)t_j(n)|$  all denote the Euclidean distance between two nodes;  $dA(r_i(1)r_i(m), t_j(1)t_j(n))$  denotes the angle ( $0^\circ$  to  $90^\circ$ ) size of the line connecting  $R_i$  and  $T_j$  endpoints of the corresponding road segment; and  $\text{sim}_{\text{len}}$ ,  $\text{sim}_{\text{shp}}$ , and  $\text{sim}_{\text{dir}}$  are the geometric similarity metrics of length,



**Fig. 7** Geometric similarity between candidate matching road segments ( $R_i$  represents a road segment in the source dataset;  $T_j$  represents a road segment in the target dataset;  $r_i(m)$  is the  $m$ th node on  $R_i$ ;  $t_j(n)$  is the  $n$ th node on  $T_j$ )

shape, and direction between candidate matching roads, respectively. Suppose their thresholds are  $\rho_{\text{len}}$ ,  $\rho_{\text{shp}}$ , and  $\rho_{\text{dir}}$ . When  $\text{sim}_{\text{len}} = \rho_{\text{len}}$ ,  $\text{sim}_{\text{shp}} = \rho_{\text{shp}}$ , and  $\text{sim}_{\text{dir}} = \rho_{\text{dir}}$  are satisfied at the same time,  $R_i$  and  $T_j$  are the identical roads.

### 3.3.2 Simple matching

When  $0 \leq |C1| \leq 1$  ( $|*|$  denotes the number of features in the set  $*$ ), the  $\text{sim}_{\text{len}}$ ,  $\text{sim}_{\text{shp}}$ , and  $\text{sim}_{\text{dir}}$  of the candidate matching road segments in  $C1$  and  $R_i$  are calculated, and if the threshold of each geometric similarity metric is satisfied at the same time, obtain a 1:1 matching pair; otherwise, obtain a 1:0 matching pair, and insert the matching result into the queue *Final*.

### 3.3.3 Combination matching

When the candidate matching set  $C1$  of  $R_i$  contains multiple road segments, screening the road segments for combination is required. By determining the constraints of  $U_T$  and  $R_i$ , where  $U_T$  denotes the combination of road segments in  $C1$ , if the constraints are met and the threshold of each geometric similarity metric is satisfied, the corresponding object of  $R_i$  is obtained. When combining candidate matching segments, the best combination is found under multiple constraints, based on the mixed-median Hausdorff distance between the candidate matching road segment and the source road segment, in order from near to far. Hausdorff distance is used for measuring the distance between two point sets, which is widely applied to describe the distance between various objects such as vector points, polylines, and polygons. However, it may have outliers when the shapes or lengths of two polylines that are used for calculating Hausdorff distance are different greatly. Nevertheless, the mixed-median Hausdorff distance can avoid the above problem. Since multi-scale road network datasets possibly have large differences in shape and length, we choose the mixed-median Hausdorff distance to calculate the distance between two road segments.

According to the first law of geography, everything is related to everything else, but near things are more related to each other, the VAMRN algorithm prioritizes the combination of candidate elements with small distances in combination matching. Table 2 shows the basic procedure of combining candidate matching elements in pseudocode.

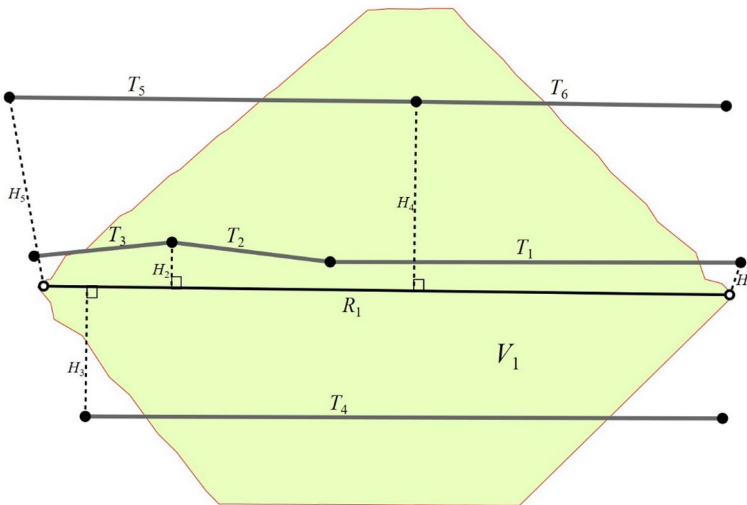
We calculated the  $\text{dis}_H(T_j, R_i)$  of each candidate matching road segment  $T_j$  in  $C1$  with  $R_i$  using the mixed-median Hausdorff distance, and combined results in the order of  $\text{dis}_H$  from the smallest to the largest. As shown in Fig. 8,  $R1$  is a road segment in the source dataset;  $V_1$  is the Voronoi diagram created by  $R_1$ ;  $T_1, T_2, T_3, T_4, T_5, T_6$  are the candidate matching set of  $R_1$ ;  $H_1$  is the Hausdorff distance between  $T_1$  and  $R_1$ ;  $H_2$  is the Hausdorff distance between  $T_2, T_3$ , and  $R_1$ ;  $H_3$  is the Hausdorff distance between  $T_4$  and  $R_1$ ;  $H_4$  is the Hausdorff distance between  $T_6$  and  $R_1$ ; and  $H_5$  is the Hausdorff distance between  $T_5$  and  $R_1$ . Where  $H_1 < H_2 < H_3 < H_4 < H_5$ ,  $T_1$  is the starting road segment of  $U_T$ . According to Hausdorff distance in order of smallest to largest, query the candidate matching road segment intersecting with  $U_T$ , and the intersecting road segment shown in Fig. 8

**Table 2** The basic steps of the combination of candidates

```

Procedure: CombinationOfCandidates
Inputs:  $R_i$  and its candidate matching set  $C1$ 
01: Set  $j=1$ , Dictionary<int, double>  $pDic = \text{new Dictionary}<\text{int}, \text{double}>()$ , and Set  $U_T$  as set of combinations
02: foreach ( $T$  in  $C1$ )
03: {  $H_j = \text{dist}(T, R_i)$ ,  $pDic.Add(T.OID, H_j)$  and  $j++$  } //records object ID of  $T$  and the  $\text{dist}(T, R_i)$ 
04:  $pDic = pDic.OrderBy(f=>f.Value).ToDictionary(f=>f.Key, f=>f.Value)$  //sort by distance
05: foreach (KeyValuePair<int, double>  $item$  in  $pDic$ )
06: {
07:   Set  $intOID = item.Key$ 
08:   Set  $dblH = item.Value$ 
09:    $T_j = \text{GetFeatureByIDFromC1}(intOID)$ 
10:   If( $j==1$ )  $U_T.add(T_j)$ 
11:    $U_{TB} = U_T.copy()$ 
12:   If( $U_T.Count > 1$  and  $U_T$  intersect with  $T_j == \text{Ture}$ )
13:   {
14:      $U_T.add(T_j)$ 
15:     If( $\text{sim}_{\text{top}}(R_1, U_T) < \text{sim}_{\text{top}}(R_1, U_{TB})$  or  $\text{sim}_{\text{shp}}(R_1, U_T) < \text{sim}_{\text{shp}}(R_1, U_{TB})$  or  $\theta_{\text{proj}}(R_1, U_T) \geq 10$ )
16:     {  $U_T = U_{TB}.copy()$ , Break }
17:   }
18:    $j++$ 
19: }
20: Return  $U_T$ 

```



**Fig. 8** Combined order based on the mixed-median Hausdorff distance from smallest to largest ( $R_i$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset;  $V_1$  represents the Voronoi diagram of  $R_1$ ;  $H_i$  is the Hausdorff distance between  $T_j$  and  $R_1$ )

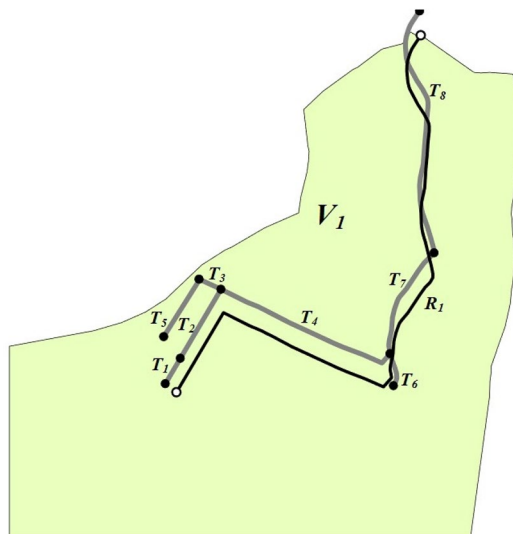
is  $T_2$ . Therefore,  $U_T$  and  $T_2$  will be merged to obtain the new  $U_T$ . The next candidate matching road segment intersecting with  $U_T$  is  $T_3$ , which is merged with  $U_T$  in the order of Hausdorff distance from smallest to largest. The next queries are  $T_4$ ,  $T_6$ , and  $T_5$ , which do not intersect with  $U_T$ , that is,  $\{T_1, T_2, T_3\}$  may match with  $R_1$ .

To ensure the correct growth of the combination of road segments, the following three constraints must be satisfied for the combination of road segments  $U_T$ .

1. The length similarity metric value of  $U_T$  and  $R_i$  is greater than the length similarity metric value of  $U_{TB}$  and  $R_i$ , where  $U_{TB}$  denotes the previous combination of  $U_T$ .
2. The value of the shape similarity metric between  $U_T$  and  $R_i$  is greater than the value of the shape similarity metric between  $U_{TB}$  and  $R_i$ .
3.  $\theta_{Proj} < 10$  of  $U_T$  and  $R_i$  to ensure that the directions of  $U_T$  and  $R_i$  are similar.

If these three constraints are satisfied simultaneously, it is proved that this is a set of further optimized combinations. The combined matching process is as follows: First, select the road segment with the smallest  $dis_H$  as the starting road segment of  $U_T$  and find its neighboring road segments in order, and combine them to obtain  $U_T$ . Then, determine whether it meets the constraints, and if it does, add the combined road segments of  $U_T$  to the temporary queue *Temp*, and repeat the combination and determination process until  $U_T$  and  $R_i$  meet the geometric similarity metrics at the same time, and the values of  $sim_{len}$  and  $sim_{shp}$  reach the maximum at the same time. Finally, add the matched pairs in *Temp* to the queue *Final*. If the selected  $U_T$  starting road segment fails to find a matching pair that satisfies the geometric similarity metric with  $R_i$  after an attempted combination, then remove the starting road segment from the candidate set. The road segment with the smallest  $dis_H$  in the candidate set is continued so as to be selected as the starting road segment for an attempted combination until the best combination is found. As shown in Fig. 9,  $V_1$  is the Voronoi diagram corresponding to  $R_1$ , and  $T_1, T_2, T_3, \dots, T_7$ , and  $T_8$  are the candidate matching road segments after  $R_1$  eliminates the anomalous road segments, where the smallest  $dis_H$  is  $T_6$ . The result of the first attempted combination is  $\{T_6, T_7, T_8\}$ . Because  $U_T$  cannot meet

**Fig. 9** 1:N road segments combination matching ( $R_i$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset;  $V_1$  represents the Voronoi diagram of  $R_1$ )

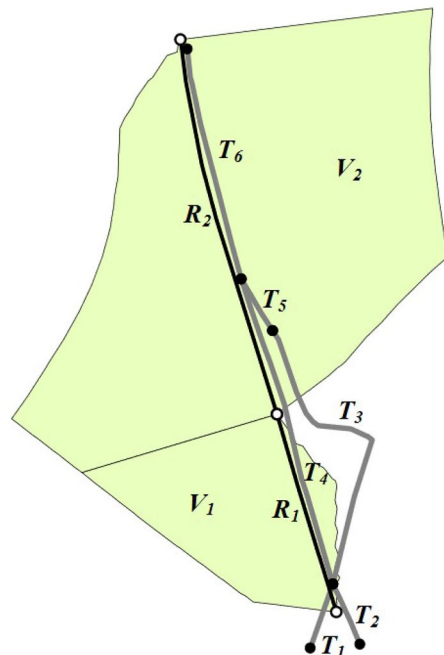




the geometric similarity metric threshold with  $R_1$ ,  $T_6$  is excluded and the second attempted combination is continued, with the result as  $\{T_1, T_2, T_4, T_7, T_8\}$ ,  $U_T$  meets the geometric similarity metric threshold with  $R_1$ , and the result is inserted into the queue *Final*. If no matching pair is obtained after eliminating all road segments in the candidate matching set, the selection of the next set of matching pairs is continued. By traversing all  $V_i$ , the obtained result is inserted into the queue *Final*.

So far, all the identical roads in 1:0, 1:1, and 1:N cases have been selected, and the identical roads in M:N cases still need to be selected. First, put the  $R_i$  segments that failed to find a matching pair into the set  $C2$ , assume that the combination of the neighboring road segments in  $C2$  is  $U_R$ , select the first  $R_i$  in the set  $C2$  as the starting road segment of  $U_R$ , query the road segments adjacent to  $U_R$ , and combine them by traversing the road segments in  $C2$  until there are no road segments adjacent to  $U_R$ . The  $U_V$  is obtained by combining the  $V_i$  of all the  $R_i$  segments that make up the  $U_R$ . Then, the candidate matching set  $C3$  of  $U_R$  is obtained by querying the road segments intersecting with  $U_V$  in  $Road_2$ , and M:N matching pairs are obtained through the process of  $dis_H$  calculation, elimination of outliers, and screening road segments of combination. Finally, the obtained result is inserted into the queue *Final*. As shown in Fig. 10,  $V_1$  and  $V_2$  are the Voronoi diagrams corresponding to  $R_1$  and  $R_2$ , respectively,  $R_1$  and  $R_2$  are combined to form  $U_R$ , and  $T_1, T_2, T_3, T_4, T_5$ , and  $T_6$  are the candidate matching road segments after  $U_R$  eliminates the abnormal road segments. The identical roads of  $R_1$  and  $R_2$  are obtained as  $T_4$  and  $T_6$  by attempted combinatorial matching. Once the matching result is obtained, all the road segments  $R_i$  that form  $U_R$  need to be

**Fig. 10** M:N road segment combination matching ( $R_i$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset;  $V_i$  represents the Voronoi diagram of  $R_i$ )



eliminated from  $C2$ , and then the next set of identical roads is selected. If a road segment  $R_i$  fails to find a neighboring road segment, it needs to be added to the 1:0 matching pair and eliminated from  $C2$  to avoid affecting the selection of the remaining matching pairs. When  $|C2|=0$ , the matching of all identical roads is completed, and the *final* queue is the final matching result.

## 4 Experiment and analysis

### 4.1 Experimental data

To verify the effectiveness and practicality of the VAMRN algorithm, we selected the experimental data in this study from the road networks in Nanchang, China; Zhejiang, China; and Buffalo, New York, in the USA. For Nanchang, we selected the road network data of 1:10,000 and 1:50,000, 1:50,000, and 1:250,000 map scales (hereafter referred to as 1WNCRoad, 5WNCRoad1, 5WNCRoad2, and 25WNCRoad, respectively) for the urban area. For Zhejiang, we selected 1:250,000 mixed urban and mountainous road network data and OpenStreetMap road network data (hereafter referred to as 25WZJRoad and ZJOSMRoad, respectively). For the city of Buffalo, we selected road network data from the city's official website (<https://data.buffalony.gov/Infrastructure/Roads/33ss-qmvk>) on August 13, 2021) and OpenStreetMap road network data (referred to as BuffaloRoad and BuffaloOSMRoad, respectively). The map scales of two datasets of Buffalo are unknown but they are similar, and their map scales are estimated approximately in the range of 1:2,000 to 1:10,000 according to the content details of road data. Four groups of experimental data are shown in Fig. 11a, b, c, and d. The road segment count, node count, total length, and coverage area statistics of each group of experimental data are given in Table 3.

### 4.2 Experiments and analysis of results

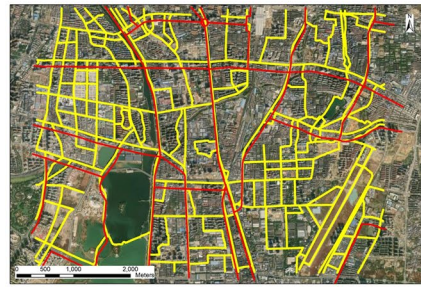
We conducted four sets of experiments using these four sets of road network experimental data. The hardware configuration of the experiment's computer was an Intel Core i7-9750H CPU, 16 GB RAM, and an Nvidia GeForce GTX 1660 Ti graphics card. The software environment consisted of 64-bit Windows 10, Microsoft Visual Studio 2019 (C#), and the ArcGIS Engine 10.2. VAMRN, the classic buffer growing method (BG) (Walter and Fritsch 1999) and the classic probabilistic relaxation method (PR) (Yang et al. 2013) that were published in high quality journals were used for the relevant experiments and algorithm comparison analysis.

In the VAMRN experiments,  $\rho_{len}=0.85$ ,  $\rho_{shp}=0.85$ , and  $\rho_{dir}=10$ . These parameters are obtained by analyzing the results of preliminary matching experiments and experts' cognitive experiences of graphic similarity and experiences in cartographic synthesis. If the length or shape threshold is greater than threshold 0.85, missing matches may occur, while if the threshold is less than 0.85, matching errors may occur. The same applies to direction threshold. In the BG

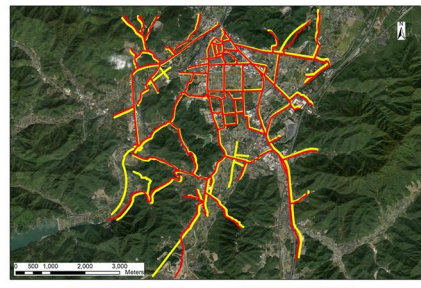
**Fig. 11** Experimental data of matching road network: **a** Group 1 dataset; **b** Group 2 dataset; **c** Group 3 dataset; and **d** Group 4 dataset



**(a)**



**(b)**



**(c)**



**(d)**

**Table 3** Road network statistics

Group	Road network name	Source (S) and target (T)	Number of segments	Number of nodes	Total road length (km)	Coverage area (km <sup>2</sup> )
1	1WNCRoad	T	402	3,924	90.00	16.86
	5WNCRoad1	S	171	1,091	59.25	
2	5WNCRoad2	T	405	2,583	136.64	30.11
	25WNCRoad	S	56	691	45.64	
3	25WZJRoad	T	122	1,402	84.80	40.97
	ZJOSMRoad	S	114	1,209	78.17	
4	BuffaloRoad	T	572	1,610	62.48	4.90
	BuffaloOSMRoad	S	206	1,371	64.47	

experiments, the buffer radius values were 38, 134, 90, and 14 m in four experiments. The method of setting buffer radius refers to the literature (Walter and Fritsch 1999). Namely the maximum value of Euclidean distances between corresponding objects recognized manually from two road datasets is selected as the buffer radius. In the PR experiments, the probability matrix iteration termination threshold was set to 0.0005, which is same to the threshold in the literature (Yang et al. 2013).

To quantitatively evaluate the matching quality and efficiency of each algorithm, we calculated and compared the *Precision*, *Recall*, *F-measure*, and the running time of the algorithms. Equations (5), (6), and (7) calculate the *Precision*, *Recall*, and *F-measure* of the matching results, respectively:

$$Precision = \frac{TP}{TP + FP} \times 100\%, \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100\%, \quad (6)$$

$$F\text{-measure} = \frac{(\alpha^2 + 1) \times Precision \times Recall}{\alpha^2(Precision + Recall)}, \quad (7)$$

where *TP* denotes the number of correct matching pairs; *FP* denotes the number of wrong matching pairs; *FN* denotes the number of wrong matches or missed matches, and the sum of *TP* and *FN* is the actual number of matching pairs of manual matching results;  $\alpha$  is the relative weight of precision; and when  $\alpha=1$ , then *Precision* in *F-measure* is considered equally important as *Recall*, and when  $\alpha>1$ , then *Recall* in *F-measure* is considered more important, and, vice versa, *Precision* is considered more important. We used the most commonly used value of *F-measure* at  $\alpha=1$ .

The results are shown in Table 4. All four groups of experimental results of VAMRN showed the largest values of *Precision*, *Recall*, and *F-measure*, with maximum values of 99.1, 97.4, and 98.2%, respectively. Compared with the BG method,

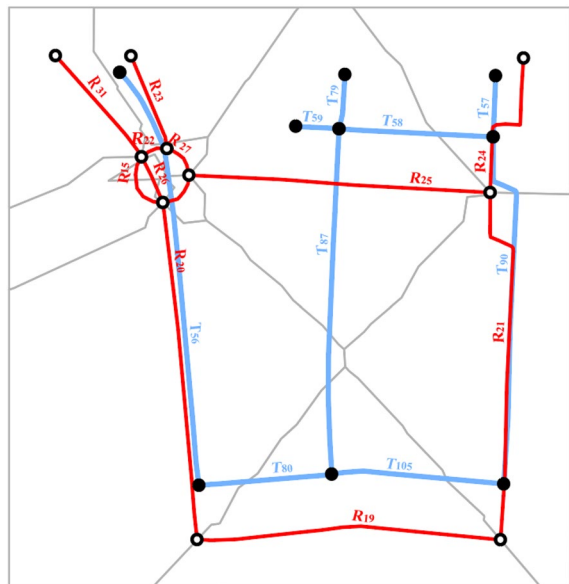
**Table 4** Matching result statistics of the three algorithms (*TP* denotes the number of correct matching pairs; *FP* denotes the number of wrong matching pairs; and *FN* denotes the number of wrong matches or missed matches)

Group	Algorithm name	<i>TP</i>	<i>FP</i>	<i>FN</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F-measure (%)</i>
1	VAMRN	147	3	4	98	97.4	97.7
	BG	117	27	34	81.3	77.5	79.3
	PR	138	7	13	95.2	91.4	93.2
2	VAMRN	49	1	4	98	92.5	95.1
	BG	27	5	26	84.4	50.9	65.5
	PR	48	3	5	94.1	90.6	92.3
3	VAMRN	111	1	3	99.1	97.4	98.2
	BG	103	1	11	99.0	90.4	94.4
	PR	108	2	6	98.2	94.7	96.4
4	VAMRN	159	15	30	91.4	84.1	87.6
	BG	138	18	51	88.5	73.0	80.0
	PR	150	29	39	83.8	79.4	81.5

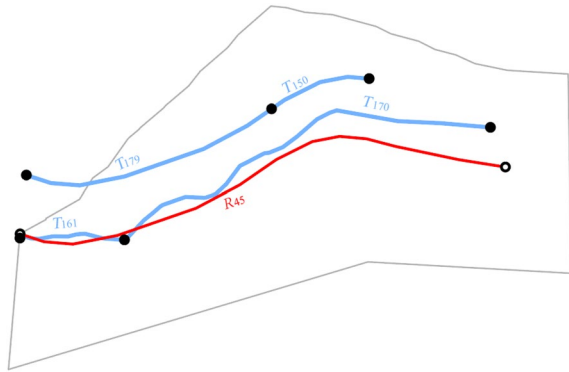
the *F-measure* of VAMRN improved by 18.4, 29.6, 3.8, and 7.6%, respectively; compared with the PR method, the *F-measure* of VAMRN improved by 4.5, 2.8, 1.8, and 6.1%, respectively.

The VAMRN algorithm also demonstrated good matching results for road networks with large positional discrepancies. As shown in Fig. 12, this part of the road network data came from the Group 2 dataset, the geometric structure of 25WNCRoad differed greatly from that of 5WNCRoad2, and the overall offset

**Fig. 12** Road network data with large positional discrepancies ( $R_i$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset)



**Fig. 13** Mismatching ( $R_{45}$  represents a road segment in the source dataset;  $T_i$  represents a road segment in the target dataset)



**Table 5** Time comparison of the three algorithms

Dataset size		Time consumption (seconds)		
Number of source road segments	Number of target road segments	VAMRN	BG	PR
171	402	9	6	783
56	405	5	1	412
122	114	3	1	187
206	572	217	3	1068

distance of 25WNCRoad was too great. The distance between  $R_{19}$  and  $T_{80}$  and  $R_{19}$  and  $T_{105}$  was too far to query the matching pairs by the buffer with an unreasonable radius in the buffer-based data matching. In the road network matching based on the PR method, the structural similarity of road segments at  $R_{26}$  was significantly different from that of  $T_{56}$ , which could easily cause the false matching. In the VAMRN-based road network matching, the Voronoi diagram corresponding to each road segment in the source dataset successfully selected the correct matching road segments in the target dataset as the matching candidates, and the correct and complete matching pairs were finally obtained through the selection of matching pairs.

VAMRN, however, also had limitations similar to the BG and PR methods, which may have resulted in false matches or missed matches in cases in which the source dataset was too offset from the target dataset and the roads were relatively dense. As shown in Fig. 13,  $R_{45}$  was a segment of the source dataset,  $T_{179}$ ,  $T_{150}$ ,  $T_{161}$ , and  $T_{170}$  were the candidate matching sets of  $R_{45}$ , and the correct matching pairs of corresponding objects were  $R_{45}$  with  $T_{150}$  and  $T_{179}$ . However, the offset of the dataset resulted in a smaller Hausdorff distance between  $T_{161}$  and  $T_{170}$ , and the combination satisfied the constraints and geometric similarity metrics, resulting in a match between  $T_{161}$  and  $T_{170}$ .

To analyze the efficiency of VAMRN, we counted the times of matching four datasets of different quantitative sizes with VAMRN, BG, and PR.

As shown in Table 5, the computation time of the VAMRN algorithm was slightly longer than that of the BG method, but much smaller than that of the PR method. The reason for the steep increase in time of the fourth group of VAMRN experiment is that there are too many intersections in the fourth dataset. Due to much intersections in the fourth dataset, the excessive number of road segments is obtained after interrupting the roads at the intersections, which leads to an increase in the number of combinations in the combination matching phase, thus it takes more time. Compared with PR, VAMRN has significant performance advantages. There are two main reasons for the long computation time of the PR method. The first is that it used a large buffer to obtain the candidate matching set, which resulted in a large amount of computation during the iteration and selection of matching results. The second is the complexity of the algorithmic process, which involved the steps of initial probability matrix acquisition, probability matrix iteration (including the calculation of compatibility coefficients and support coefficients of neighboring candidate matching pairs), and matching pair selection (including the calculation of structural similarity of matching pairs, matching growth process, and selection of robust matching pairs).

## 5 Conclusions

Object matching is the crucial technology of road data conflation. We proposed an innovative Voronoi diagram-based approach for matching multi-scale road networks (VAMRN). (1) VAMRN innovatively constructs Voronoi diagrams of road segments that can effectively avoid the intersection of Voronoi polygon and several road segments by the strategy of adding dense points at special intersections. Using Voronoi diagram to filter the candidate matching set of each road segment can effectively avoid manually setting the buffer radius used for searching candidate matching set, and prune the matching searching space, which avoids searching a large number of irrelevant candidates. It improved the universality and efficiency of object matching method. (2) Meanwhile, we designed the geometric similarity metrics including length, shape and direction, and a heuristic combinatorial growth strategy that are helpful for accurately matching multi-scale road data by using Voronoi diagrams. Our method only uses the geometric similarity metrics, which reduces the dependency on object attributes. However, the matching accuracy is still high. Moreover, the proposed approach has the ability to detect all matching pair types including 1:1, 1:N, M:1, M:N, 1:0, and 0:1.

In general, VAMRN demonstrated higher quality and good performance in overall matching results compared with the BG and PR methods, and offered a more universal and practical approach for matching multi-scale and complex road network data. However, VAMRN also had two shortcomings: (1) the threshold values for determining corresponding objects were set by expert experience, while when the map scales of identical objects span a large scale, some corresponding objects with large differences in shape may fail to match because of the threshold selection problem. The next work will investigate adaptively thresholds setting of geometric similarity metrics to reduce manual intervention; (2) when the positional discrepancies



between the two road networks are great or the local regional roads are relatively dense, the Voronoi diagram of the source road segment may not query all matching candidates, which leads to false matches or missed matches. These two shortcomings need to be further investigated in depth. In addition, in future research, more multi-source and multi-scale road network data will be acquired for validation and optimization of our method.

**Acknowledgements** We thank the anonymous reviewers for their constructive comments.

**Author contributions** JW, YZ, and MY proposed the matching method; YZ and JW implemented the algorithm and performed the experiments; JW, XZ, and XH collected and processed data; JX discussed the idea and analyzed the data; JW and YZ wrote the paper.

**Funding** This research was funded by National Natural Science Foundation of China [grant number 41201409 and 41561084] and Graduate Student Innovation Foundation of Jiangxi Normal University [grant number YJS2020012].

**Data availability statement** Not applicable.

## Declarations

**Conflict of Interest** The authors declare no conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


## References

- Chehreghan A, Abbaspour RA (2017) A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm. *Cartogr Geogr Inf Sci* 45:255–269
- Chen J, Zhao R, Qiao C (2003) Voronoi diagram-based GIS spatial analysis. *Geomat Inf Sci Wuhan Univ* S1:32–37
- Deng M, Li Z, Chen X (2007) Extended Hausdorff distance for spatial objects in GIS. *Int J Geogr Inf Sci* 21(4):459–475
- Fu Z, Wu J (2008) Entity matching in vector spatial data. The 21st ISPRS (International Society for Photogrammetry and Remote Sensing) Conference, Beijing, Vol. XXXVII, Part B4, 1467–1472
- Gabay Y, Doytsher Y (1994) Automatic adjustment of line maps. *Proc GIS/LIS 94*:332–340
- Guo L (2008) Theory and method research on multi-sources geospatial vector data fusion. Ph.D. Dissertation, PLA Information Engineering University, Zhengzhou, China.
- Hao Z (2010) Spatial-temporal database query and reasoning. China Science Publishing & Media Ltd.
- Hu T, Mao Z (2011) Methodological research on optimal matching candidates of line objects. *Sci Survey Mapp* 36(02):132–135
- Lei TL (2021) Large scale geospatial data conflation: a feature matching framework based on optimization and divide-and-conquer. *Comput Environ Urban Syst* 87(5):101618
- Li L, Goodchild MF (2011) An optimization model for linear feature matching in geographical data conflation. *Int J Image Data Fusion* 2:309–328

- Luan X (2013) Methods for modeling levels-of-detail road networks with the maintenance of structural patterns. Ph.D. Dissertation, Wuhan University, Wuhan, China.
- Ma J (2020) Research on theory and methodology of road and settlement fusion considering scale variation. Ph.D. Dissertation, PLA Strategic Support Force Information Engineering University, Zhengzhou, China.
- Pan Z (2004) Digital mapping principles and methods. Wuhan University Press
- Tong X, Deng S, Shi W (2007) A probabilistic theory-based matching method. *Acta Geodaetica Et Cartographica Sinica* 2:210–217
- Volz, S. (2006) An Iterative Approach for Matching Multiple Representations of Street Data. In M. Hampe, M. Sester, L. Harrie (Eds.): ISPRS Vol. XXXVI., ISPRS workshop – multiple representation and interoperability of spatial data; Hannover, Germany, 22–24.
- Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. *Int J Geogr Inf Sci* 13:445–473
- Wu J (2008) Researches on Methods of Entity Matching and Its Applications in Vector Spatial Data. Ph.D. Dissertation, Wuhan University, Wuhan, China.
- Wu J, Wan Y (2015) Research on point cloud data management based on spatial index and database. *Sci Survey Mapp* 40:97–100
- Wu J, Wan Y, Chiang Y, Fu Z, Deng M (2018) A matching algorithm based on Voronoi diagram for multi-scale polygonal residential areas. *IEEE Access* 6:4904–4915
- Yang Y (2016) Researches on Methods of Multi-characteristics Road Network Matching and Data Updating Applications. Ph.D. Dissertation, Wuhan University, Wuhan, China.
- Yang B, Zhang Y, Luan X (2013) A probabilistic relaxation approach for matching road networks. *Int J Geogr Inf Sci* 27:445–473
- Yu, M. (2017) Research on Voronoi-diagrams-based Matching Method of Multi-scale Road Networks. MA.Sc. Dissertation, Jiangxi Normal University, Nanchang, China.
- Zhang, Q. (2002) The study of map database entity matching and merging technology. Ph.D. Dissertation, Wuhan University, Wuhan, China.
- Zhang Q, Li D, Gong J (2002) Shape similarity measures of linear entities. *Geo-Spatial Inf Sci* 5(2):62–67
- Zhang Y, Yang B, Luan X (2012) Automated matching urban road networks using probabilistic relaxation. *Acta Geodaetica Et Cartographica Sinica* 41:933–939
- Zou Z, Yang L, An X et al (2020) A hierarchical matching method for vectorial road networks using delaunay triangulation. *ISPRS Int J Geo Inf* 9(9):509

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Jianhua Wu<sup>1</sup>  · Yu Zhao<sup>1</sup>  · Mengjuan Yu<sup>2</sup> · Xiaoxiang Zou<sup>3</sup> · Jiaqi Xiong<sup>1</sup> · Xiang Hu<sup>1</sup>

Jianhua Wu  
wjhgis@126.com

Mengjuan Yu  
yumengjuan@pudong.gov.cn

Xiaoxiang Zou  
zouxiaoxiang9803@126.com

Jiaqi Xiong  
201940100061@jxnu.edu.cn

Xiang Hu  
202040100120@jxnu.edu.cn

- <sup>1</sup> School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China
- <sup>2</sup> Pudong New Area Big Data Management Center, Shanghai 200135, China
- <sup>3</sup> Jiangxi Institute of Natural Resources Surveying, Mapping and Monitoring, Nanchang 330002, China