Send Orders for Reprints to reprints@benthamscience.ae

1286 *The Open Cybernetics & Systemics Journal,* 2014, *8,* 1286-1292

Open Access

Road Network Matching Method Based on Particle Swarm Optimization Algorithm

Lin Yang*, Fang Fang, Songling Dai, Bo Wan and Zejun Zuo

Faculty of Information Engineering, China University of Geosciences (Wuhan), Wuhan, 430074, China

Abstract: Combined the global optimization ability of particle swarm algorithm and memory capacity of tabu algorithm, this paper proposed an automatic vector road network matching method based on the combination of particle swarm optimization and tabu strategy. Firstly, the similarity between node entities is evaluated by means of geometric and topological characteristics. Then, the basic principle of global optimization of particle swarm optimization is introduced and road matching model based on particle swarm optimization algorithm is designed. Meanwhile, the tabu search algorithm is joined, by using the ability of tabu strategy which expanded the search of the neighborhood. The algorithm fully reflects the "climbing" feature of tabu strategy, in order to find the global optimal solution of the matching relationship of road network entities. Three different forms of road network data of Wuhan are selected to test our method, the result indicates that the matching method based on the combination of particle swarm optimization and tabu strategy is effective and feasible, which can provide a new idea to solve the matching problem.

Keywords: Global optimization, particle swarm optimization, road network matching, tabu strategy.

1. INTRODUCTION

Nowadays, more and more geographic data are captured, updated and manipulated, besides, with the rapid development of Internet, an increasing need is thus appearing: integrating possibly heterogeneous databases and sharing this information on the Internet. Devogele points out that one of the major problems that the development of GIS applications is facing today is the integration of data from different sources [1].

Road network spatial target matching is a core and key technology of application such as integrate, update vector road map, navigation based on location service and so on. The matching method not only can reduce the high cost of data duplication, improve the quality of spatial data, but also can satisfy the strong demand of the rapid development of vehicle navigation products for the latest navigation electronic map. Furthermore it greatly improves the update efficiency and reduces maintain costs. Currently, matching problem is a hot-spot issue researched by domestic and foreign scholars and has very important scientific and applied value. Related scholars have realized that spatial data update will replace the data collection to become the bottleneck of GIS construction [2].

Early approaches of matching spatial data from different data models is the work of Bureau of the Census in Washington DC [3]. They developed the first automatic road matching system which started from dealing with simple 1:1 matching relationship. Later A star algorithm was used to solve the optimal solution by Walter and Fritsch which can deal with complex M:N matching relationship [4], however, it was very time-consuming. Since then, extensive studies such as road network matching and conflation have been performed.

Most of the existing matching strategies are focused on the optimization point and evaluation function of similarity measurement of line segments. Concerning point similarity evaluation, Zhao proposed the similarity evaluation function of the point of the similarity and the shape similarity of the adjacent arc segments [5]. Safraa calculated the similarity of points by the position of the endpoint of the line [6], Yang proposed the local structure of a point, which takes a point as the center, and extends some of the arc, using the similarity of local structure indicates the similarity of point [7].

From the perspective of line similarity evaluation, Huttenlocher proposed Hausdorff distance matching method and indicated that Hausdorff distance is a measure of the distance between the two line objects [8]. The Hausdorff distance is suitable for measure the similarity between line segments. Deng Min extended the traditional Hausdorff distance in order to make the Hausdorff distance more adaptive to the change of shape of the matching objects [9]. Zhang proposed a matching method based on Delimited-Stroke-Oriented method [10], the similarity of stroke is used to express the similarity between lines, in order to obtain the similarity evaluation function of the line, Zhang analyzed the difference in the distance, length and shape of the arc [11]. Tong proposed a method called SM_HD distance to express the similarity of the roads [12], SM_HD distance has better robustness and accuracy than Hausdorff distance.

In summary, scholars mainly concerned with the improvement of the similarity evaluation function on road network matching algorithm. About matching strategy, the local

1874-110X/14 2014 Bentham Open

optimal solution is found mainly in the local near space in the early stage, the stroke technology, the sub network and other local spatial structure were used to improve the situation later. Because of the complexity of the relationship between the elements of the road network, one difficulty in the matching of road network is that it is very difficult to effectively balance the contradiction between multiple entities and candidate option, thus unable to find a global approximate optimal matching solution. However, the road network matching is essentially a global optimization problem. Few people use artificial intelligent algorithms to study the problem of road network matching at home and abroad. Therefore, combined the global optimization ability of PSO algorithm with the memory ability of the tabu search algorithm, this paper proposed an automatic vector road network matching method based on the combination of particle swarm optimization and tabu strategy. The particle swarm optimization algorithm acts as an iterative strategy, and iterates towards optimal solution through the flying of particle. This algorithm can quickly converge to the global optimal solution. By using the ability of tabu strategy which expanded the search of the neighborhood, this algorithm fully reflects the "climbing" feature of tabu strategy, to find the global optimal solution of the matching relationship of road network entities.

2. ALGORITHM BASIC IDEA

Particle swarm optimization algorithm (PSO) was first proposed by Kennedy and Eberhart [13], which was got inspiration from bird foraging. PSO is used to solve optimization problem. A particle represents a solution of optimization problem. Particles constantly search for the current best particle's neighborhood. Particles would reach the global best solution through continuous iteration.

Tabu Search (TS) was first proposed by Glover [14]. Suboptimal solution could be accepted in order to avoid local optimal solution in Tabu Search. It pardons some excellent condition in the tabu list in order to increase search efficiency and find the global best solution.

This paper proposes a road network automatic matching method based on particle swarm optimization algorithm. Road network matching is an optimization problem in N (the number of nodes in road network) dimensional space. A particle's position in N dimensional space is a solution of optimization problem. One particle in iterations, there is the particle which has maximum fitness called personal best solution (P_{ld}). All particles there is a particle which have maximum fitness called group best solution (P_{gd}) after an iteration. PSO is an iterative strategy that all particles search the neighborhood of P_{ld} and P_{gd}. TS strategy is used to avoid falling into searching the same particle. It pardons some excellent condition in the tabu list in order to increase improve search efficiency and prevent the loss of the optimal solution.

Road network automatic matching algorithm based on particle swarm optimization includes six core parts.

2.1. Particle

Particle is the solution of the optimization problem. In the road network matching, particle represents a collection of all the matching relationships of all the nodes in the original data and target data. In other words, particle represents a result of road network matching. Particle is denoted as formula (1). M_{ij} represents a matching pair. i denotes the feature identity of original data and j denotes the feature identity of target data.

$$Particle = \{ M_{ij} \mid i \in O, j \in D \} \tag{1}$$

2.2. Similarity Measurement Function

The similarity measurement function expresses the similarity between nodes, it is very essential to evaluate the entities of the same name. The calculation formula is as follows,

$$Agl(S,T) = \frac{1}{n} * \sum_{j=1}^{n} (aglsim(SL,TL)) \tag{2}$$

$$aglsim(SL,TL) = \frac{1}{k} \sum_{i=1}^{k} agl(S_i S_{i+1}, T_i T_{i+1}) \tag{3}$$

$$tp(S,T) = \frac{Min\_TP(S,T)}{Max\_TP(S,T)} \tag{4}$$

$$F(S,T) = \frac{a}{k} * \sum_{i=1}^{k} Agl(S,T) + b * tp(S,T) \tag{5}$$

S denotes a node from original data, T denotes a node from destination data. Formula (2) expresses the vector similarity of the adjacent edges of the nodes. Formula (3) calculates the vector similarity of the edges. N is the number of adjacent edges of nodes. The following work should be carried out before the computation of the vector similarity. Suppose SL and TL are two edges needed to calculate. $S_0,S_1...S_n$ are the break point of SL, $T_0,T_1...T_m$ are the break points of TL. Firstly, cut the length of the long side and short side of the same length and get two new edges. Then, take S_0 as the starting point and find a virtual point S1 on SL which makes $|S_0S1| = |T_0\ T_1|$, and find a point S2 that makes |S1S2| = |T1 T2|, continuely find virtual points until the end of the line segment. Do the same for TL. Formula (4) represents the topological similarity degree of two nodes. Min_TP(S,T) represents the smaller topological degree of S,T. Max_TP(S,T) represents the larger topological degree of S,T. Formula (5) represents the initial similarity degree of two nodes, k is the number of adjacent edges, a and b are proportional factors, according to the experience value, we set a=0.5, b=0.5,

$$sim_{ab} = m * F(a,b) + n * \sum_{i=1}^{k} F(a_i, b_j) * \frac{1}{k} \tag{6}$$

The final similarity evaluation function for a and b, is shown in the Formula (6), k is the number of adjacent edges, a_i indicates the adjacency point of the node a, b_i indicates the adjacency point of the node b, m and n is proportional factors, according to the experience value, we set m=0.5, n=0.5.

2.3. The Fitness of Particle

The fitness of particle represents the sum of the similarity of each pair of nodes in a particle. The greater the fitness

value, the better the matching results. The calculation formula is as follows,

$$fitness = \sum_{i=0}^{N_a} Sim_{ij} \left(i \in \{0,1,2...N_a\}, b_j \in List_i^k \right) \quad ..(7)$$

$List_i^k$ represents the set of nodes that can be found in the candidate data set of node i, N_a represents the number of nodes.

2.4. Iteration Process

Particles in any dimension of the position are discrete. The particle position in any one dimension is represented by x_i, and the velocity is expressed by v_i, i represents the ith node (that is the ith dimension). The value of x_i indicates the similarity value between the current node and its matching node. Then the iterative process of the particle in the N dimension is the change of the position of each dimension, which is to change the matching nodes. The iterative process is that all particles fly close to the current optimal particle. The direction of the velocity is from the particle to the current optimal particle. The value of v_i indicates the distance from current state to the result state (that is change of similarity degree). The iterative formula is as follows [15].

$$v_i^T = wv_i^{T-1} + c_1 r_1 (P_{ld}.x_i - x_i) + c_2 r_2 (P_{gd}.x_i - x_i) \quad (8)$$

$$x_i^T = v_i^T + x_i^{T-1} (i = 0,1,2,....N_a) \quad (9)$$

T indicates the number of iterations, learning factor c_1 and c_2 are non-negative constants, r_1 and r_2 are a random number between [0,1], x_i indicates similarity value of the ith node pairs, P_{ld} is the optimal solution of history, P_{gd} is the group optimal solution, *w* is dynamic weight, *w* will change with the evolution of particles and the degree of aggregation. If the particle swarm evolution speed is relatively fast, the algorithm can continue to search in the larger search space, the particle can maintain a large range of optimization. When the evolution speed of the particle swarm is reduced, the value of *w* can be reduced, so that the particle swarm can be searched in a small space to find the optimal solution faster. If the particles are dispersed, the particle swarm is not easy to fall into local optimal solution. With the increase of the degree of aggregation of particles, the algorithm is easy to fall into local optimum. We should increase the search space and improve the global optimization ability of the particle swarm.

In summary, *w* has an inverse relationship with the evolution speed factor. *w* is proportional to the degree of aggregation factor. *w* calculation method is as follows [15].

$$w = w_{ini} - h * w_h + s * w_s \quad (10)$$

w_{ini} is the initial value of *w*, usually w_{ini} equals 1, h is an evolutionary factor, s is the aggregation factor. w_h and w_s are the ratio of h and s, according to the experience w_h is between [0.4,0.6], w_s is between [0.05,0.2].

$$h = \frac{\min(F(g_{best_{T-1}}), F(g_{best_T}))}{\max(F(g_{best_{T-1}}), F(g_{best_T}))} \quad (11)$$

$$s = \frac{\min(F(g_{best_T}), \overline{F_T})}{\max(F(g_{best_T}), \overline{F_T})} \quad (12)$$

F(gbest) represents the group optimal solution of fitness, T indicates the number of iterations, $\overline{F_T}$ is the average number of fitness of all particles. After one iteration, in order to get a new x_i, a matching point b_j is found which have closest value x_i from $List_i^k$. Assign Sim_{ij} to x_i as a new result. After all the x_i in the particle are calculated and the fitness value of the particles could be calculated, then the particle's one iteration is completed.

2.5. Tabu Search

PSO algorithm could fast find optimal solution, but it is easy to fall into local optimal. Therefore, tabu search algorithm is using to prevent premature or fall into local optimum. There are k optimal solutions should be recorded in the iterations. K is set 3 in this paper. Respectively, the particle history optimal solution P_{ld}, it records the maximum fitness of a particle in iterative processes. The group optimal solution P_{gd}, it records a particle which has maximum fitness of all particles in an iterative process. The global optimal solution called bestptr. It records a particle which has maximum fitness of all particles in iterative processes.

In the iterative process, P_{ld} and P_{gd} are part of iterative formula. The particle will be put into tabu list which is selected as P_{gd}. The time the particle stay in tabu list is $\sqrt{N}$, N is the number of particles. Particles in the tabu list will not be selected until the end of the time. The particles could be selected which meet the contempt criteria in order to prevent the loss of the global optimal solution. The contempt criteria is ruled that fitness of particle is greater than bestptr.

2.6. End Condition

There are two end conditions one is the algorithm reaches the preset number of iterations, and the other is the optimal solution is stable continuously in a fixed period. The calculation will be terminated if any of the conditions has been met.

3. ALGORITHM STEPS

This is the flow chart of the algorithm as Fig. (**1**).

The algorithm steps are as follows:

(1) Traverse all nodes of the original data, then use formula (6) to calculate the similarity of the node in original and all destination nodes in the node buffer, meanwhile sort by sim.

(2) Initialize particles in a random manner, select P_{gd} and bestptr, and put the particle into the tabu list.

(3) To determine whether or not meeting the end conditions, if satisfied, enter the step (6), otherwise enter the step (4).

(4) Formula (8) and (9) is used to update the particles.

(5) Get the fitness of the current particle, If there is a value bigger than bestptr, select the particle to be P_{gd} and update bestptr, otherwise choose the maximum particle which is not in the tabu list. Put the particle which is se-

lected as P_{gd} into the tabu list, update the term of all particles in the tabu list. Enter step (3).

(6) Output all the nodes pairs in the bestptr.

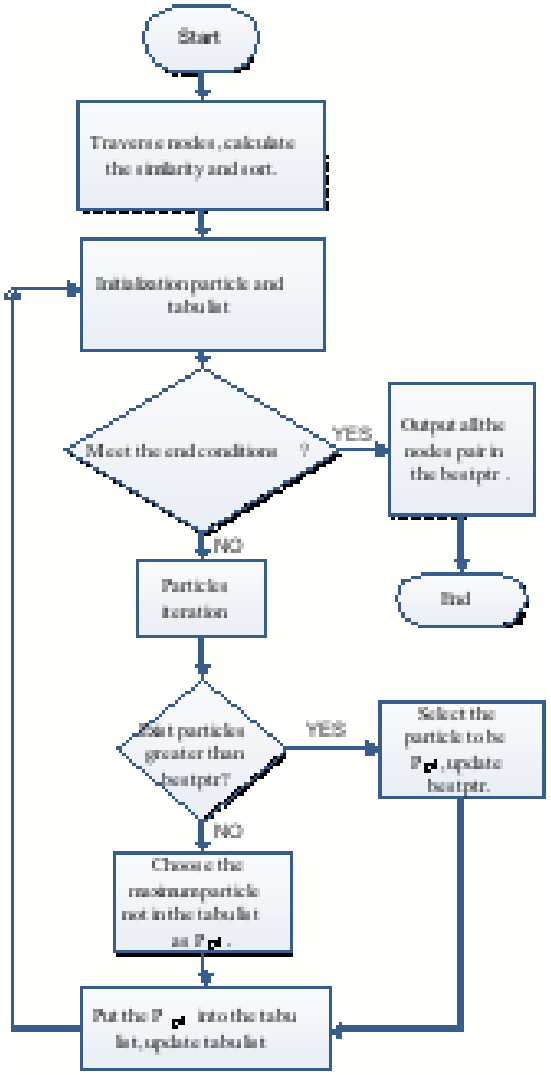

Fig. (1). Flow chart of the algorithm.

4. EXPERIMENT AND DISCUSSION

4.1. Experimental Environment

In order to verify the validity of the model of road network matching algorithm, C++ programming language is adopted in this paper, Based on MapGIS K9 in Visual Studio2010 environment, the design and development of the road network matching experiment system is developed. Three different forms of road network data of Wuhan are selected as experimental data. The acquisition time of the two data sources are 2011 and 2006 respectively. Two data sources come from different units. Experimental data are selected from two typical characteristics as shown in Figs. (**2**-**4**). The area in Fig. (**2**) has obvious characteristics of the grid. The area in Fig. (**3**) has obvious characteristics of the radiation modes. The area in Fig. (**4**) has a mix of the two characteristics above. Table **1** shows the node number, line number and total road length of three different experiment areas.

4.2. Algorithm Evaluation Index

The matching result is divided into matched objects and non-matched objects, the matched objects can be divided into accurate matches, mismatches and false positive matches; non-matched objects can be divided into the proper non-matches and negative matches.

In this paper, we use the recall rate (recall) and the correct rate (matchingCorrect) to evaluate the matching results, the calculation formula is as follows;

$$matchingCorrect = \frac{N_{am}}{N_{am} + N_{mm} + N_{pnm}} \times 100\% \quad (12)$$

$$recall = \frac{N_{am}}{N_{am} + N_{mm}} \times 100\% \quad (13)$$

N_{am}, N_{mm}, N_{pnm} representative accurate matches, mismatches and proper non-matches.

4.3. Algorithm Parameter Settings

The optimal parameters of the algorithm according to the experiments are listed in Table **2**.

Fig. (2). Grid area.

Table 1. Statistical information table.

Study Area	Data Source	Nodes Num	Lines Num	Length
1	Original	123	184	100148
	Destination	629	809	182075
2	Original	259	404	9257
	Destination	842	1152	137712
3	Original	91	111	50218
	Destination	438	536	98377

Table 2. Parameter setting.

Parameter Name	Mark	Reference Value
Particle number	P	50
Acceleration factor	c_1, c_2	2
Evolution factor	w_h	0.5
Aggregation factor	w_s	0.1
Maximum iteration number	K	45
Long tabu	T	$\sqrt{P}$

Fig. (3). Radiation area.

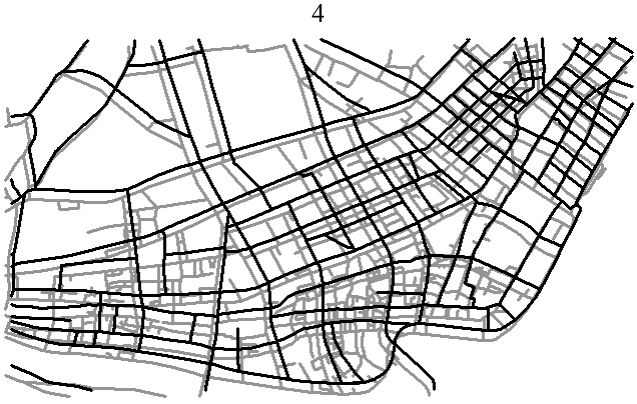

Fig. (4). Mixing area.

4.4. Algorithm Achieve And Result Analysis

According to the experimental analysis, through the method of matching result by human eyes, experimental result statistics that the model's average correct matching rate and recall rate were 75.8% and 73.2% respectively. Among them, the road network grid feature correct matching rate is 82.6%, the recall rate is 83.1%. The correct rate of the radiant road network is 62.2%, the recall rate is 59.8%, the correct rate of the mixed road is 74.4% and the recall rate is 76.7%. Results are shown in Fig. (**5**).

In this paper, we take the Fig. (**2**) as the experimental data and test the result under condition of different iteration and different particle number. Results as shown in Figs. (**6**) and (**7**).

The result shows that, when the number of iterations is relatively small, with the increase of the number of iterations, the recall rate and the accuracy of the matching have increased. However when the number of particles exceeds a certain number, recall rate and accuracy are stable. So it is important to select the appropriate number of iterations, less iteration number is not conducive to find the optimal solution. The iteration number will increase the experiment time and reduce the operating efficiency. According to the experiment the number of iterations is better in [40, 50].

4.4.1. Effect of Iteration Number and Particle Number on Results of Particle Swarm Optimization

The relationship between the number of particles and the results of the matching is basically the same as the number of iterations. More particles can produce more solutions and

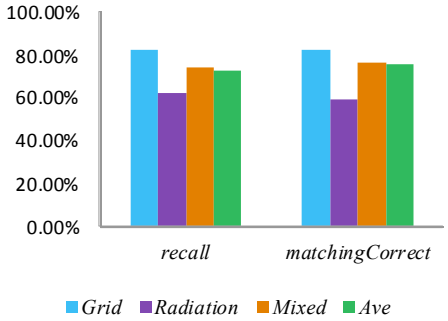

Fig. (5). Matching result.

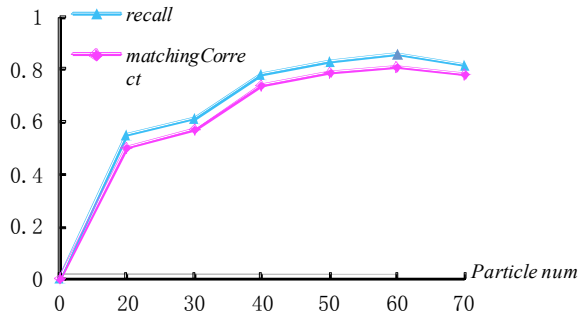

Fig. (6). The relationship between the number of particles and the matching results.

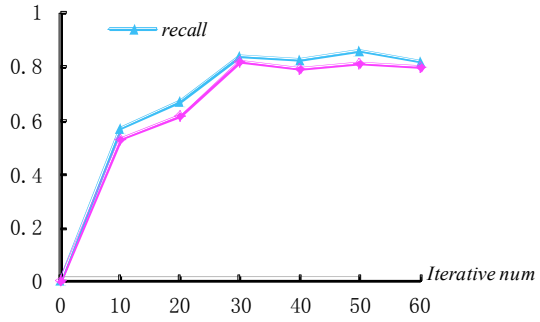

Fig. (7). The relationship between the number of iterations and the results of the matching.

are conducive to global optimization. However at the same time, it increases the running time of the experiment. So under the smaller experimental data, fewer particles should be selected. According to the experiment the number of particles is between [50, 60].

4.4.2. Comparison Between Particle Swarm Algorithm and Traditional Method

The traditional method is to calculate the similarity of each node with its candidate set of nodes and select the highest similarity as the matching node. Using Fig. (**2**) as the experimental data, the experimental results are as follows

From Figs. (**8**) and (**9**), we can find that the particle swarm optimization algorithm has a significant improvement in the matching accuracy and recall rate. Particle swarm algorithm can significantly reduce the duplication of the matching, and the particle swarm algorithm has a better effect on the identification of 1:0.

Fig. (8). Comparison between particle swarm algorithm and traditional method.

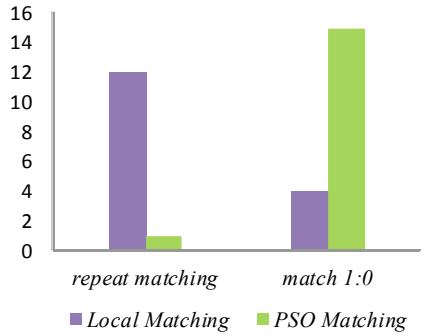

Fig. (9). Comparison between particle swarm algorithm and traditional method.

CONCLUSION

The paper proposed a road network automatic matching method based on particle swarm optimization algorithm and tabu search strategy. Particle' speed function could make particles to move to the global best solution quickly. The algorithm is robustness because of particle's parallelism. Moreover, the PSO has global optimization which increases the correct matching rate of road matching. The tabu search optimizes the iterative process and raises the efficiency of the iteration. In the after work, we will further study the problem of particle swarm optimization in dealing with discrete data and let the particles have more practical significance. Also we will discuss the iterative process of parallel computing in order to reduce the time complexity.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research has been supported by National Natural Science Foundation of China (No. 41201385, 41301426, 41301427, 4137142) and supported by Open fund of State Key Laboratory of Geographic Information Engineering (SKLGIE 2013-M-4-5).

REFERENCES

[1] S. Mustiere, T. Devogele, "Matching networks with different levels of detail", *GeoInformatica,* 12, 435-453, 2008.

[2] R. J. Zhai, "Research on Automatic Matching Methods for Multi-Scale Vector Spatial Data Based on Globla Consistency Evalua-

tion", *Zhengzhou: The PLA Information Engineering University*, 2011.
[3] A. Saalfeld, "Conflation: automated map compilation, International Journal of GIS, vol. 2, no. 3, pp. 217-228, 1988.
[4] V. Walter and D. Fritsch Matching spatial data sets: a statistical approach", *International Journal of Geographical Information Science*, vol. 13, pp. 445-473, 1999.
[5] D. B. Zhao, "Research on Automatic Matching of Vector Road Networks Based on Global Optimization", *Nanjing: Nanjing Normal University*, 2010.
[6] E. Safraa, Y. Kanza, Y. Sagiv, "Ad hoc matching of vectorial road networks", *International Journal of Geographical Information Science*, 2013, vol. 27, no. 1, pp. 114-153.
[7] B. S. Yang, "A Pattern-Based Approach for Matching Nodes in Heterogeneous Urban Road Networks", *Transactions in GIS* vol. 18, no. 5, pp. 718-739, 2014.
[8] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850-863, 1993.
[9] M. Zheng, X. Z. Feng, X. Y. Chen, "A generic model describing topological relations among area objects", *Acta Geodaetica et Cartographica Sinica*, vol. 34, no. 2, pp. 142-147, 2005.
[10] M. Zhang, L. Meng, "Delimited stroke oriented algorithm-working principle and implement for the matching of road networks", *Annals of GIS*, vol. 14, pp. 44-53, 2008.
[11] Y. F. Zhang, B. S. Yang, X. C. Luan, "Automated matching urban road networks using probabilistic relaxation", *Acta Geodaetica et Cartographica Sinica*, vol. 41, no. 6, pp. 933-939, 2012.
[12] X. H. Tonga, D. Lianga, & Y. Jina, "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.
[13] J. Kennedy, R. Eberhart, "Particle sawrm optimizatioin", *In: Proc of the IEEE Intertnational Conference on Neural Networks. Perth. Australia,* pp. 1942-1948, 1995.
[14] F. Glover, "Tabu search: Part I", *ORSA Joumal on Computing* vol. 1, pp. 190-206, 1989.
[15] X. P. Zhang, "Adaptive particle swarm optimization algorithm with dynamically changing inertia weight", *Journal of Xi'an Jiaotong University*, vol. 39, no, 10, pp. 1039-1042, 2005.

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014