An Algorithm for Composing Principal Road Network from Digital Road Map by Using Topological Information

Shigeki Toriumi

Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
E-mail address: toriumi@ise.chuo-u.ac.jp

(Received May 30, 2008; Accepted October 27, 2008)

In this paper, we propose an algorithm for composing a "practical" principal road network from the DRM Database which contains detailed road networks. When we study transportation problems between cities, it is difficult to take out "practical" principal road network from the DRM Database. For instance, there is a possibility of becoming the following "not-practical" principal road networks when we extract them based only on the attribute data of the links. That is, the principal road network (1) could become disjoint; (2) has nodes that are not intersections; (3) shows intersections expressed in more detail than is necessary; and (4) has links for roads with central dividers split into upper lines and lower lines. The algorithm solves these problems by using topological information in addition to attribute data. The proposed algorithm is not limited to the DRM Database, and can be applied to general digital road maps with a little modification.

Key words: Digital Road Map, Principal Road, Topological Information, Contraction

1. Introduction

Ordinary maps printed on paper are comprehensible for people. However, "digital road map", in which locations and other information are expressed in numeric form, is necessary so that the computer for the car navigation system and the road management systems may recognize roads and intersections. We propose an algorithm to reduce the network size of "digital road map" with macroscopic structure maintained. Here, we assume that the network size is defined by the number of links and the number of nodes. The smaller the network size is, the more preferable it is. If the network size is small, it becomes easy to calculate network flow.

In this paper, we treat "the DRM (Digital Road Map) Database" that is maintained and provided by Japan Digital Road Map Association. In the DRM Database, road networks are depicted as combinations of "nodes" and "links" as illustrated in Fig. 1. A unique ID number is assigned to each node and each link.

- Node: An intersection or other necessary nodal point of the road networks.
- Link: A road segment between two nodes.

The characteristics of the DRM Database are as follows:

(1) The data structure is suitable for representing networks and retrieving the route based on distance and time.
(2) Data on arterial roads of the prefectural level and higher is updated in advance of road openings and other changes.
(3) In addition to location information, the DRM Database also includes a great deal of other data such as each road's administrator, route number, width, and road structures such as bridges and tunnels.

Recently, the volume of the DRM Database has increased, and the number of links, which is an indicator of the size of network, is more than several million.

Researchers of the city planning and operations research use the DRM Database to their research. However, because they are not necessarily familiar with the DRM Database, it might be difficult to extract necessary information from the DRM Database. It is comparatively easy to extract information based on the region. However, when we study transportation problems between cities, it is difficult to extract "practical" principal road network. For instance, there is a possibility of becoming the following "not-practical" principal road networks when we extract them based only on the attribute data of the link (road type and width, etc.). That is, the principal road network

(1) could become disjoint;
(2) has nodes that are not intersections;
(3) shows intersections expressed in more detail than is necessary;
(4) has links for roads with central dividers split into upper lines and lower lines.

Problem (1) occurs when two regions are connected only on narrow side streets. When narrow side streets are omitted, those regions are divided into parts. Problem (2) occurs when each node represents an intersection of principal roads and narrow side streets. When narrow side streets are omitted, each node is not an intersection any longer. Problem (3) and (4) occur in a lot of places. Those information is necessary to guide the route by the car navigation system etc., however, when the transportation problem is solved macroscopically, it is excessive.

In this paper, we propose an algorithm for composing a "practical" principal road network from the DRM Database.

Table 1. Attribute data of link.

Contents	
Link id (node id numbers of each end)	Number of lanes
Administrator code	Carriageway width [unit: meter]
Road type [expressway, urban expressway, national road, principal prefectural road, general prefectural road, principal city road, general city road, municipal road, others]	Condition of use [in-service, under construction, under contemplation]
Route number	Median width
Overlapped route information	Median extension length
Municipal area code	12-hours traffic volume
Link length	Travel speed (at peak times)
Width classification	Speed limit and other traffic regulations
Width of narrowest carriageway portion	Coordinates of interpolation point locations
Types of weather-related traffic restriction zones	Link type
Vehicle weight	Toll road [true, false]
Height and width restrictions	etc.

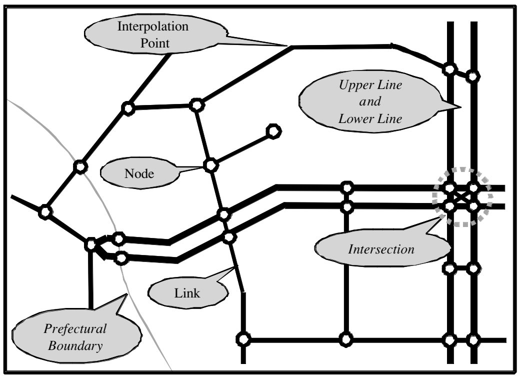


Fig. 1. Data model of the DRM Database.

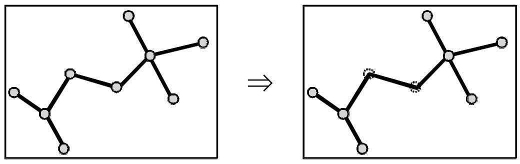


Fig. 2. Step1.

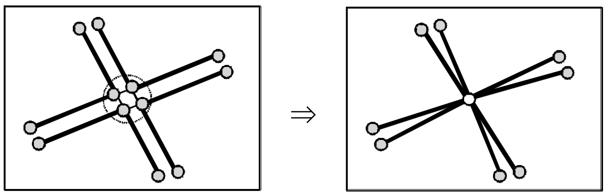


Fig. 3. Step2.

Of course, the composed network does not have the above problems and is suitable for solving transportation problems. The feature of the proposed algorithm is that it makes use of topological information of the road networks in addition to attribute data. The proposed algorithm is not limited to the DRM Database, and can be applied to general digital road maps with a little modification.

Various studies have been conducted on the automatic generalization of maps. For example, the Lang method (Lang, 1969) and the Douglas-Peucker method (Douglas and Peucker, 1973) are known as algorithms for simplifying broken-line sections. However, most of these methods focus on displaying maps, and they ignore the structure of networks.

2. Principal-Road Network Structural Algorithm

2.1 Digital road maps

The DRM Database is maintained and renewed under standards established by the Japan Digital Road Map Association (Ishida and Yaguchi, 2007). The DRM Database has the topological structure of the road networks and the attribute data of links. Table 1 shows a part of the attribute data that the DRM Database has. For example, “Road type” has the value such as an expressway, a national road, a principal prefectural road, a general prefectural road and a municipal road, etc. “Link type” has the value such as a main road, a junction road, an intersection road, a frontage road, etc.

2.2 Algorithm

The algorithm for composing a principal-road network from the DRM Database is described by the following five steps:

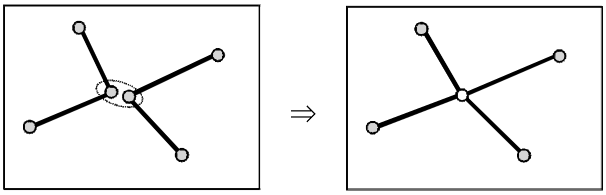
Step0 Composition of an initial network using attribute data,
Step1 Deletion of nodes of degree 2,
Step2 Contraction of intersection links and junction links,
Step3 Contraction of two adjacent nodes,
Step4 Contraction of adjacent nodes and links.

In the following, we explain each step.

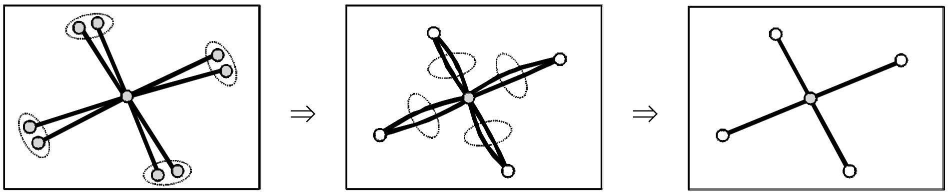
Step0

We compose the network consisting of the road links that meets the following three requirements:

- “Road type” value is “national road”, “principal prefectural road”, “general prefectural road” or “principal city road”. Or, “Carriageway width” value is “13 m”

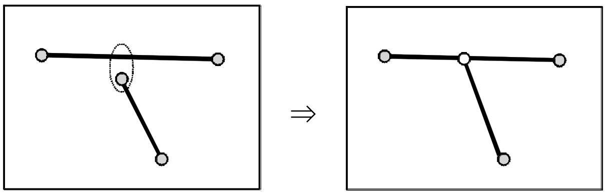


(a) Case without multiple links

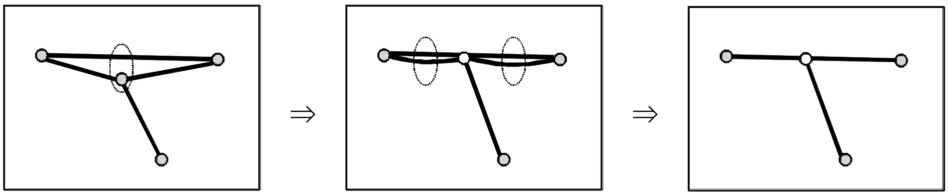


(b) Case with multiple links

Fig. 4. Step3.



(a) Case without multiple links



(b) Case with multiple links

Fig. 5. Step4.

or more.

- “Toll road” value is “false”.
- “Condition of use” value is “in-service”.

Furthermore, in the network, we find the maximal connected sub-network (i.e. connected component), and we use this sub-network as an initial network for composing a principal road network. When the network is disjoint, that is there are nodes that cannot be reached mutually, there is a possibility that the transportation problem cannot be solved.

Step1

In this step, we find the nodes of degree 2 and delete them. Here, the degree of the node is defined as the number of links connected with the node; for example, the node of degree 2 means the number of links connected with the node is two. Generally, when solving transportation problems, in the nodes of degree 2 the flow enters in through one link, and exits out through the other link (except for the source and the sink), the nodes are redundant. Thus, the nodes of degree 2 can be deleted without affecting the solution to the transportation problem. Therefore, because the number of nodes decreases, the network size can be reduced.

In the DRM Database, a link has interpolation points which represent shape in addition to two nodes which represent endpoints of a link. Nodes represent an intersection, a dead-end point or a change point of attribute data (for instance, the intersection of prefectural boundary). The degree of the node representing an intersection is three or more, the degree of the node representing a dead-end point is one, and the degree of the node representing a change point of attribute data is two. Therefore, all nodes of degree 2 are change point of attribute data. However, in the initial network (in Step0), some nodes of degree 2 are intersections of principal roads and narrow side streets. So, we find and delete them (see Fig. 2). When this is done, the two links connected to the node are joined. The attribute data, such as link length, is updated at the same time as the sum of the length of adjacent links. At this time, the shape of the joined link does not change because the interpolation points are maintained.

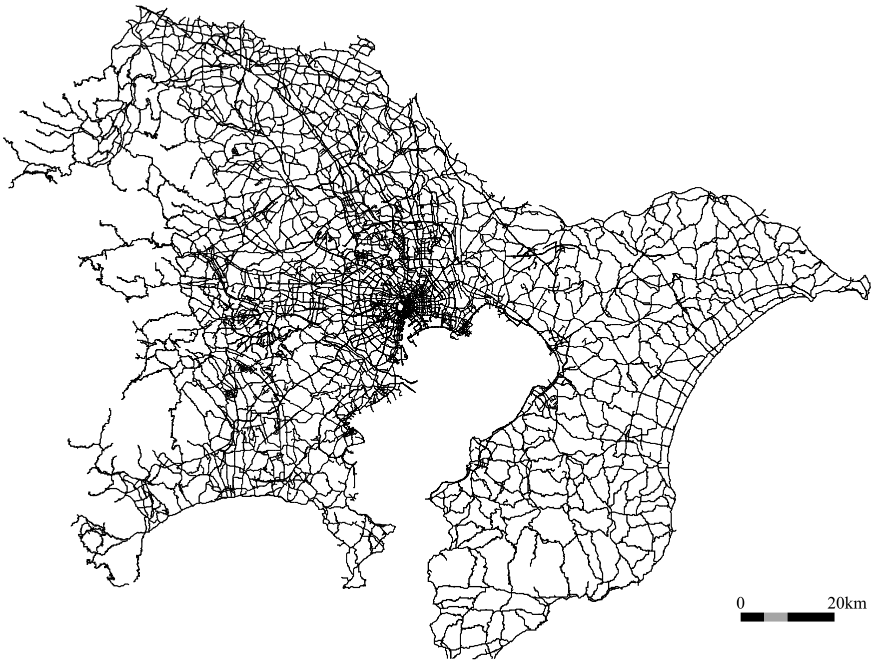


Fig. 6. Initial network (South Kanto).

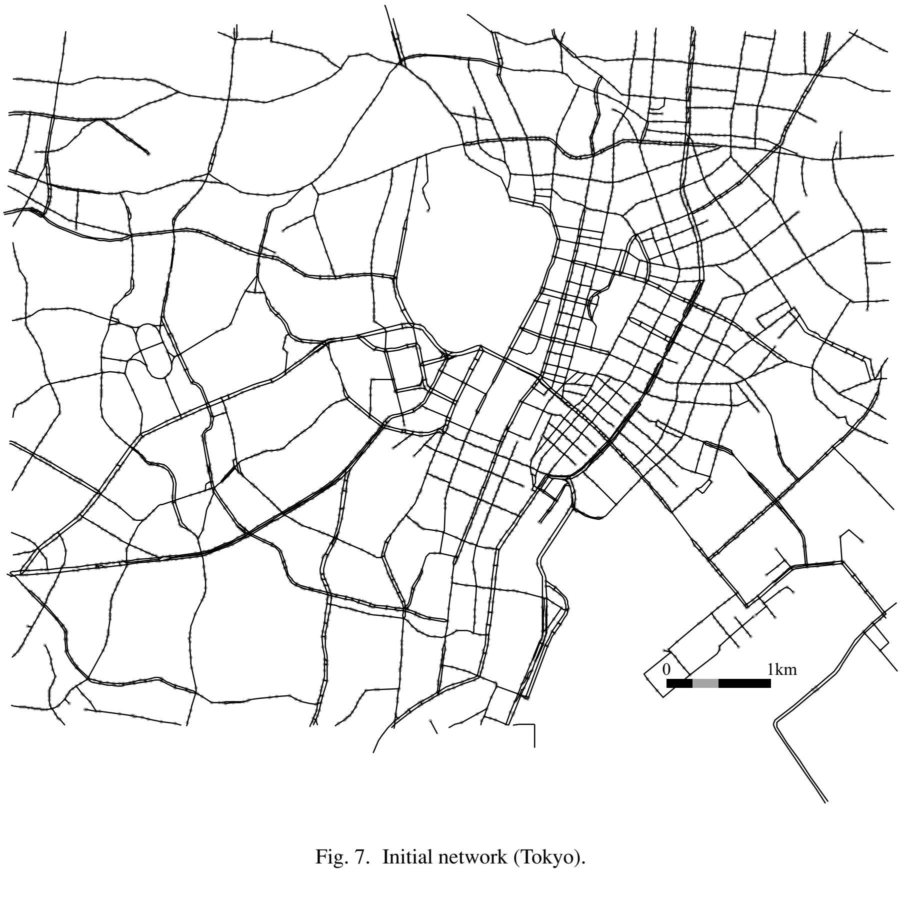


Fig. 7. Initial network (Tokyo).

Step2

In the DRM Database, to guide the route by the car navigation system, intersections and junctions might be represented by two or more links and nodes. However, when solving transportation problems macroscopically, those links and nodes are unnecessary. So, in this step, intersection links and junction links with length smaller than a given threshold value δ are contracted in the above network (Iri and Koshizuka, 1993), and the intersection is represented by one node (see Fig. 3). When the link is contracted in the above network, the number of nodes decreases at the same time. We assume the position of the consol-

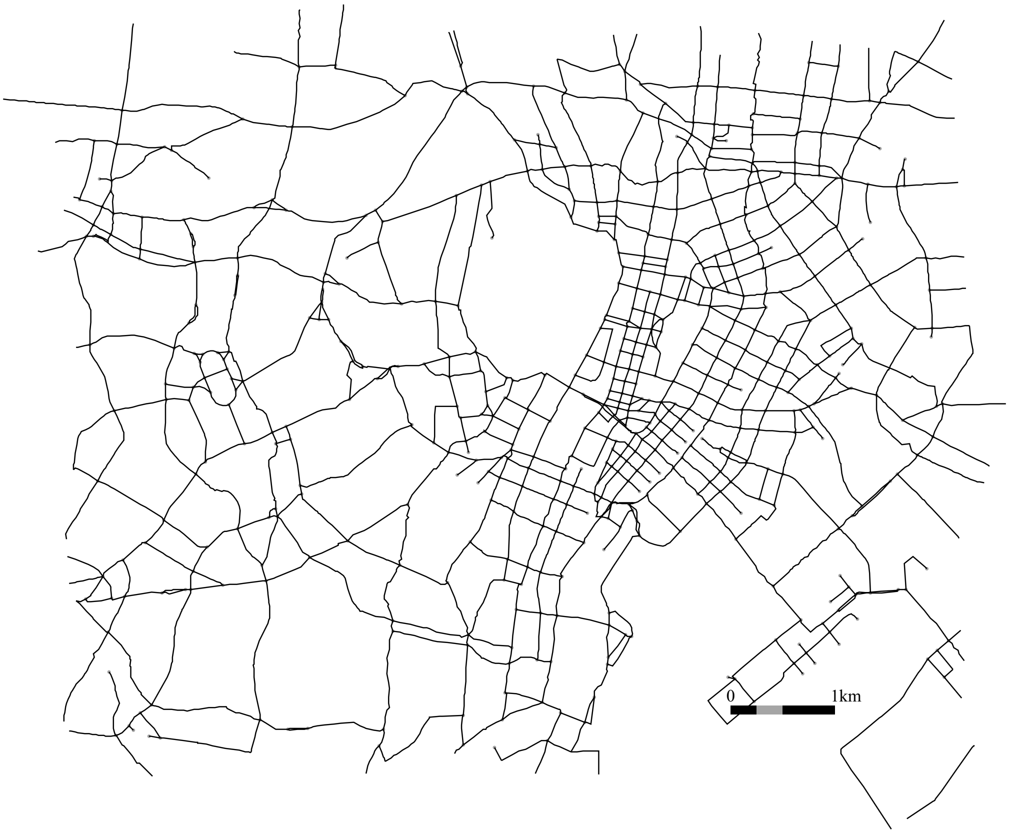


Fig. 8. Composed principal road network ($\delta = 50$ m).

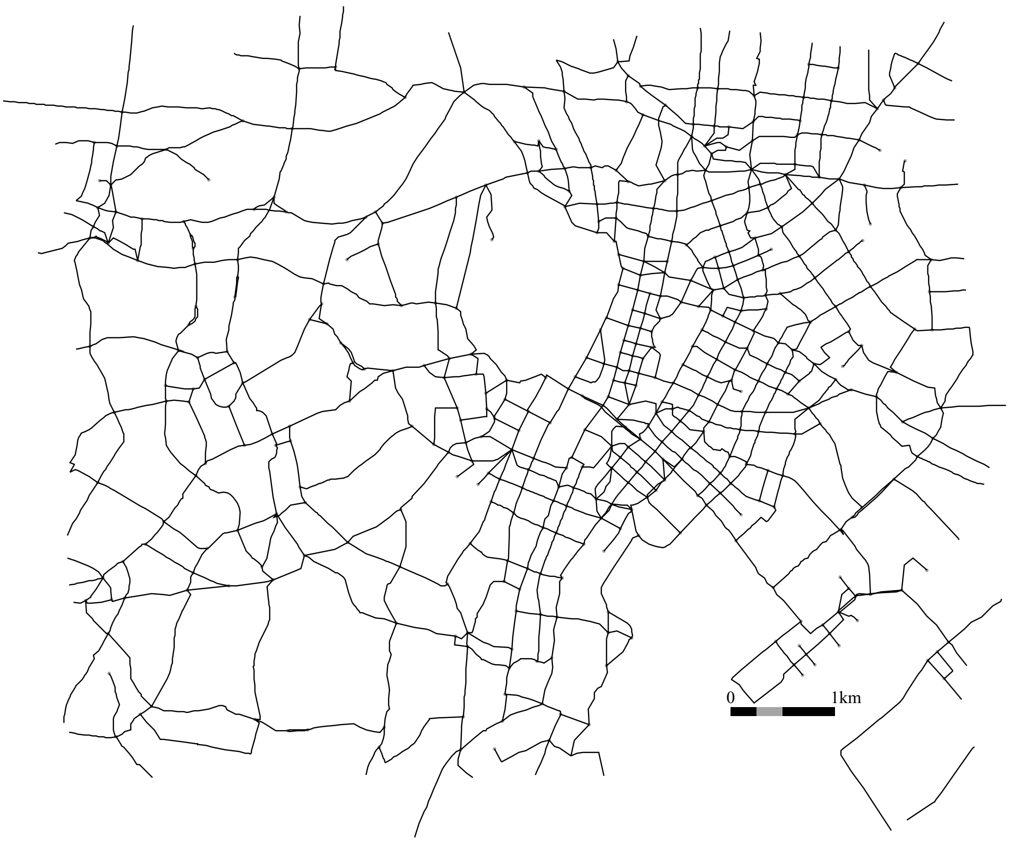


Fig. 9. Composed principal road network ($\delta = 100$ m).

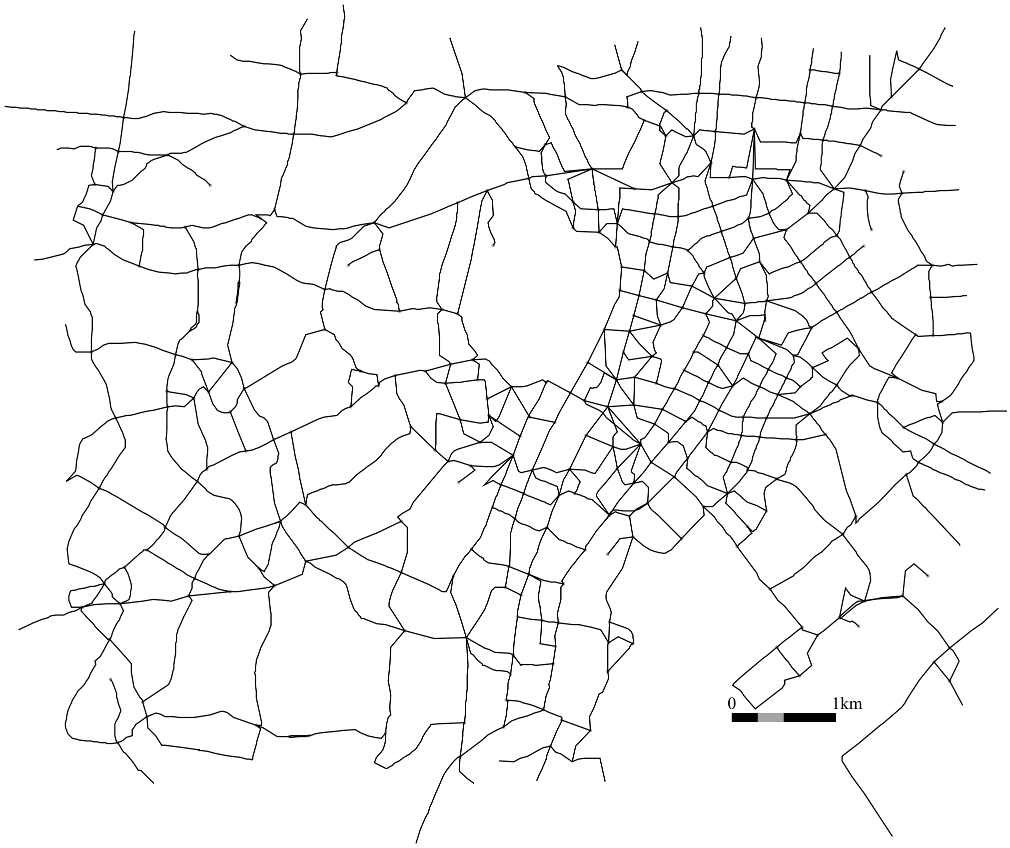


Fig. 10. Composed principal road network ($\delta = 150$ m).

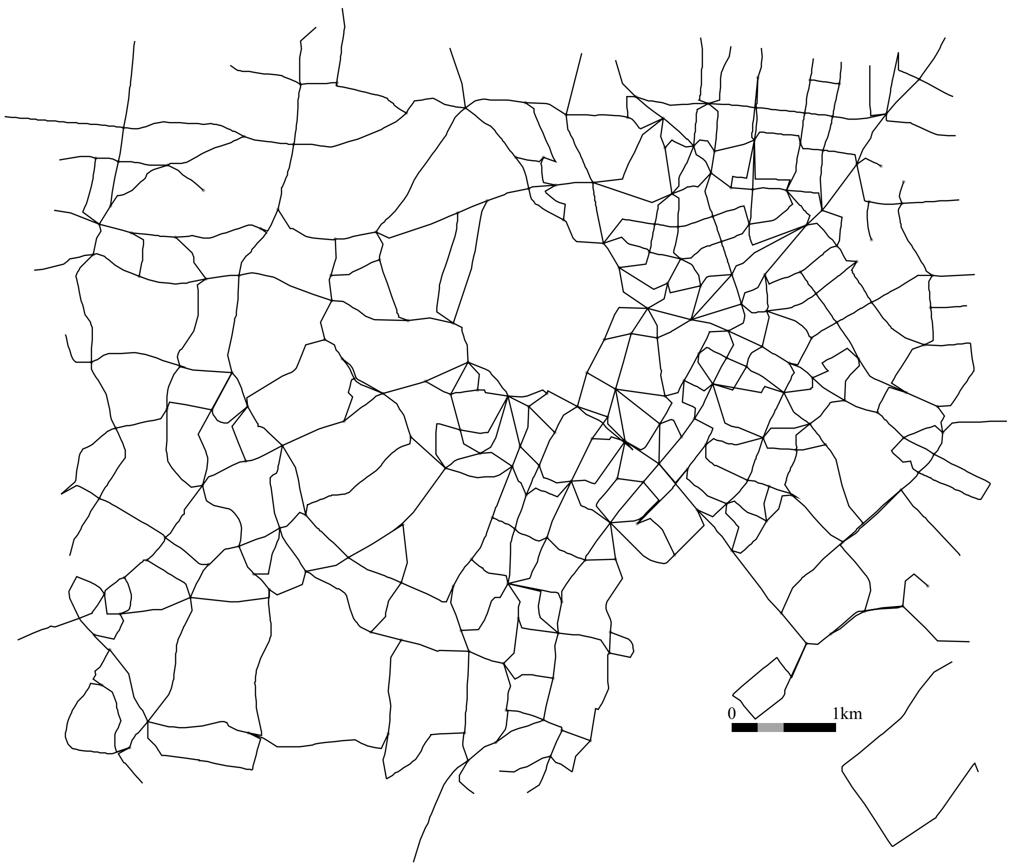


Fig. 11. Composed principal road network ($\delta = 200$ m).

Table 2. Number of nodes and number of links ($\delta = 100$ m).

	Initial network		Composed network ($\delta = 100$ m)	
	number of nodes	number of links	number of nodes	number of links
Saitama	26,875 (100%)	29,522 (100%)	1,851 (6.9%)	2,750 (9.3%)
Chiba	22,219 (100%)	23,890 (100%)	1,688 (7.6%)	2,407 (10.1%)
Tokyo	43,880 (100%)	48,651 (100%)	2,601 (5.9%)	4,024 (8.3%)
Kanagawa	25,063 (100%)	26,701 (100%)	1,318 (5.3%)	1,871 (7.0%)

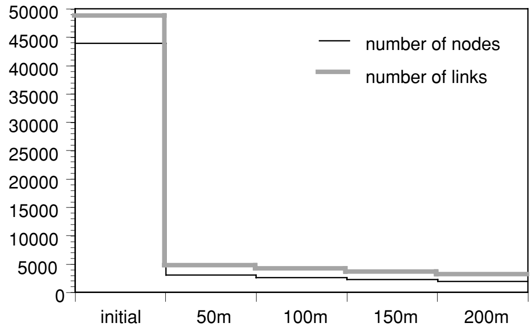


Fig. 12. Number of nodes and number of links (Tokyo).

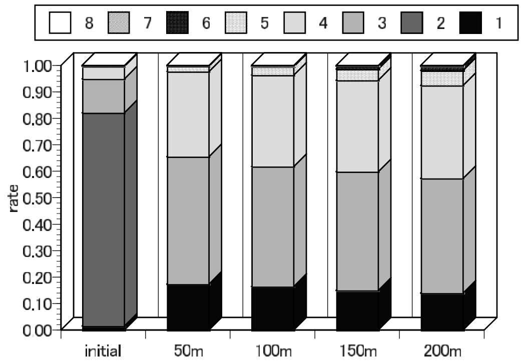


Fig. 13. Distribution of the degree of nodes (Tokyo).

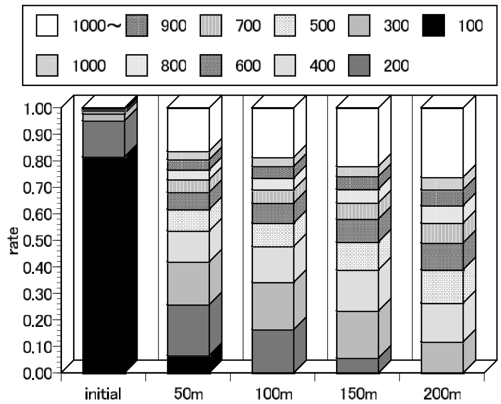


Fig. 14. Link-length distribution (Tokyo).

idated node is moved to the middle point of the link that is contracted, so the network shape changes minutely and locally by this operation.

Step3

Toriumi and Taguchi (2006) proposed an algorithm for obtaining arbitrarily precise spatial data from spatial data of a certain precision by contraction of nearby spatial objects. We herein make use of this algorithm to contract node pairs which are separated by a distance shorter than the given threshold value δ (see Fig. 4(a)). When the links connected with the contracted node have another common node, those links become multiple links (see Fig. 4(b)). In this case, we leave behind only one link of each pair and delete the other. As well as Step2, we assume the position of the contracted node is moved to the middle point of original two nodes, so the network shape changes minutely and locally by this operation.

Step4

Similarly with Step3, we contract node and link pairs which are separated by a distance shorter than the given threshold value δ (see Fig. 5(a)). At this time, the link is divided into two. If there are multiple links, we leave behind only one link of each pair and delete the other, as in Step3 (see Fig. 5(b)).

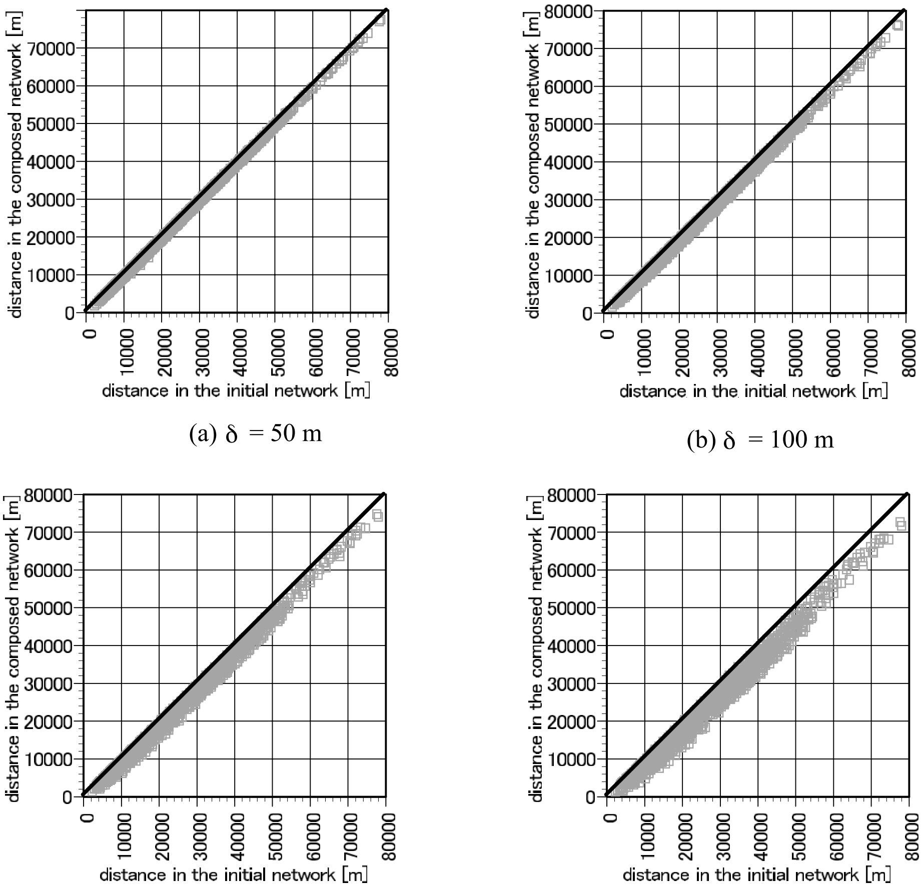
The number of contracted nodes and links becomes large when we use a larger threshold value δ. The more the number of contracted nodes and links increases, the more the number of nodes and links decreases. In addition, the amount of the movement of the node increases by enlarging of the threshold value δ. The structure (i.e. shape, topology, etc.) of original network is changed. Therefore, it is necessary to choose an appropriate threshold value δ according to the purpose of the analysis.

3. An Application Example using South Kanto as the Target Region

We apply the algorithm to South Kanto (Saitama, Chiba, Tokyo and Kanagawa) area after setting the threshold value δ to from 50 m to 200 m in 50 m increments. The entire target region is shown in Fig. 6, while an enlarged image of the central area of Tokyo is shown in Fig. 7. The composed principal road networks by applying the algorithm are shown from Figs. 8 to 11 (all of which show an enlarged image of the central area of Tokyo).

By comparing Fig. 7, which depicts the initial network, with from Figs. 8 to 11, which depict the composed principal road network, shapes of these networks are similar. However, as the threshold value increases, the difference from the initial network increases. Furthermore, it is also clear that many of the links which had been divided into upper and lower lines in the initial network have been consolidated into single links in the composed principal road

Fig. 15. City office (Tokyo).



(a) δ = 50 m

(b) δ = 100 m

(c) δ = 150 m

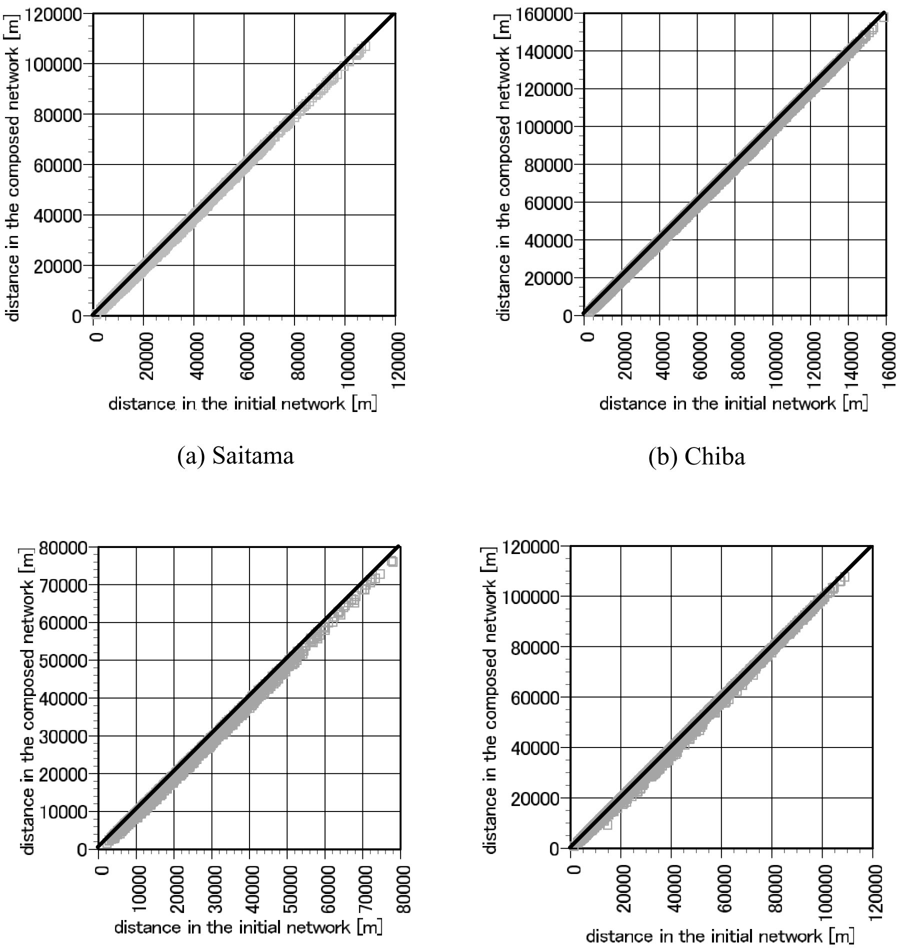
(d) δ = 200 m

Fig. 16. The shortest route distance between two city offices (Tokyo).

networks.

Next, Fig. 12 shows the number of nodes and the number of links in the Tokyo Metropolitan Area. It is clear that as the threshold value δ increases, the number of nodes and the number of links decrease. Table 2 shows the number of nodes and the number of links in the composed principal road network for each target region with a threshold value δ set at 100 m. By examining Table 2, it is understood that the number of nodes and the number of links have been greatly decreased in any region. The number of nodes is approximately 5% to 8% as that of the initial network, and the number of links is approximately 7% to 10% as that of the initial network.

Let us examine the nodes and the links in a bit more detail. First, regarding the nodes, we looked at the distribution of the degree of nodes in each network. Figure 13 shows the distribution of the degree of nodes. In the initial network, the ratio of the nodes of degree 2 is approximately 80%,



(a) Saitama

(b) Chiba

(c) Tokyo

(d) Kanagawa

Fig. 17. The shortest route distance between two city offices (=100 m).

Table 3. Coefficient of regression and coefficient of determination.

	Threshold δ [m]			
	50	100	150	200
Saitama	0.9988 (0.9997)	0.9922 (0.9993)	0.9766 (0.9983)	0.9574 (0.9967)
Chiba	1.0009 (1.0000)	0.9970 (0.9999)	0.9901 (0.9997)	0.9744 (0.9984)
Tokyo	0.9958 (0.9998)	0.9776 (0.9986)	0.9496 (0.9940)	0.9079 (0.9853)
Kanagawa	1.0028 (0.9999)	0.9939 (0.9998)	0.9795 (0.9991)	0.9401 (0.9955)

(Coefficient of determination.)

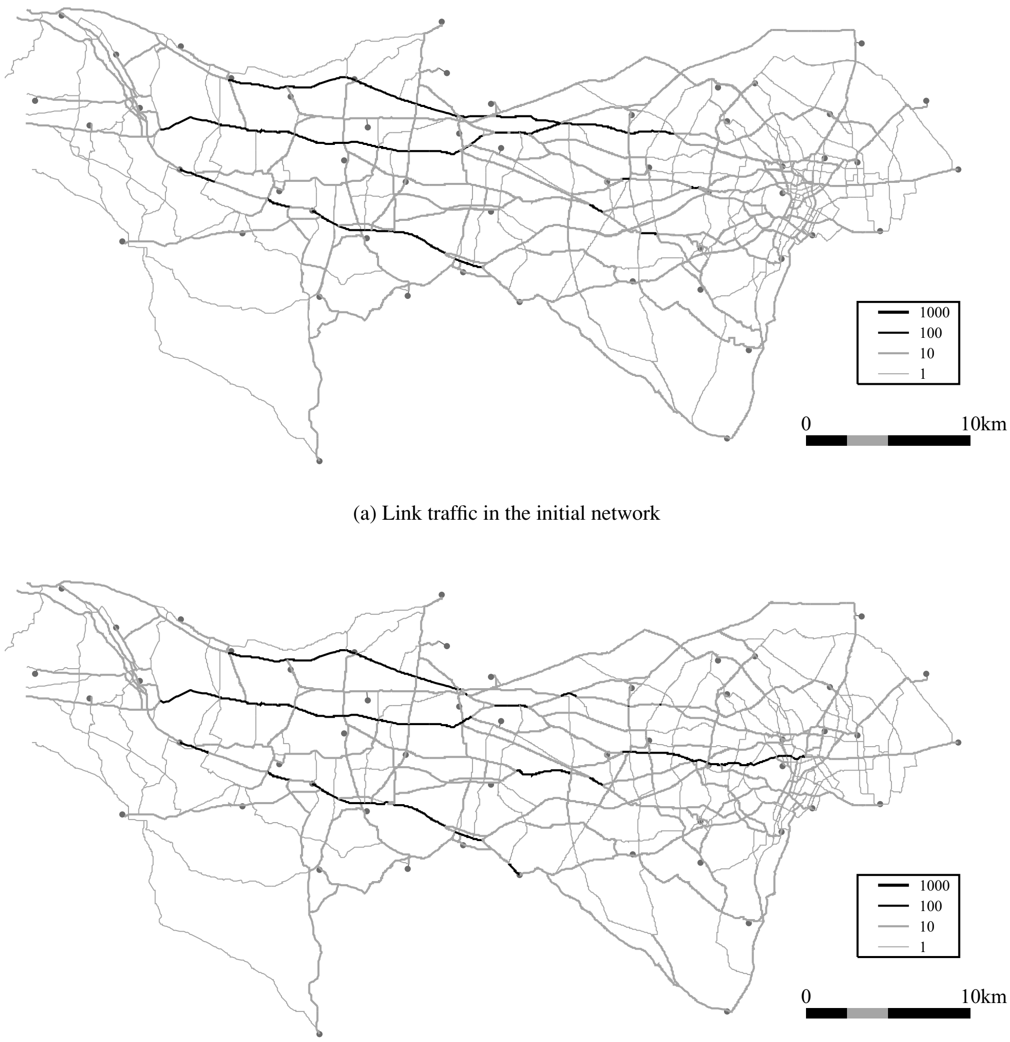
while the ratio of the nodes of degree 3 or above is 20% or less. In the composed principal road networks, the nodes of degree 2 are deleted by the algorithm, consequently, the ratio of the nodes of degree 3 or above becomes 80% or more. Among the obtain networks, the differences are not so large by the value of δ.

Next, regarding the links, we focus on the link-length distribution in each network as shown in Fig. 14. In the initial network, the ratio of links with length smaller than 100 m is approximately 80%, while the ratio of links with length longer than 100 m is 20% or less. In the composed principal road networks, the ratio of longer links increased, because short links were deleted by the algorithm. This tendency strengthens as the threshold value δ increases.

4. Evaluation of the Composed Principal Road Network

4.1 Shortest route distance

We compare the shortest route distance between two geographical points for each network to evaluate the extent to which the composed principal road network preserves the characteristics of the initial network. Naturally, it is preferable that the distance between two points in the composed principal road networks be equal to the distance in the initial network. Here, we took the municipal government building in each city, ward, town and village in each region as the representative point, and found the shortest route distance for all of these pairs (see Fig. 15). We solved the all-pairs shortest path problem by repeating the Dijkstra method.



(a) Link traffic in the initial network

(b) Link traffic in the composed network ($\delta = 100$ m)

Fig. 18. Link traffic.

Figure 16 shows scatter plot diagrams of the shortest route distance for each pair, with the horizontal axis displaying the shortest route distance in the initial network, the vertical axis showing the shortest route distance in the composed principal road network. In each scatter plot diagram, there are a strong positive correlation. However, as the threshold value δ increases, the difference from the initial network becomes large. There is a tendency that the shortest route distance in the composed principal road network becomes smaller than the shortest route distance in the initial network. As the threshold value δ increases, more intersection links are contracted. So the shortest route distance shrinks by exactly the same length as the contracted link.

Furthermore, to observe the difference among the target region, similar scatter plot diagrams for each region are created for $\delta = 100$ m as shown in Fig. 17. By examining Fig. 17, the difference among regions is not large. It is clear that there is a stable, strongly positive correlation between them.

To evaluate both sets quantitatively, we perform a regression analysis with the shortest route distance in the initial network as the explanatory variable and the shortest route distance in the composed principal road network as the explained variable. Table 3 shows the results of the regression analysis for each composed principal road network. Any coefficient of determination is over 0.98, the explaining variable is well suitable for the explained variable. Therefore, the composed principal road networks maintain overall characteristic of the shortest route distance of initial network.

4.2 Shortest route

Here, let us analyze results by using the shortest route. We assume that there is one unit of traffic between each pair. So, we count how many times each link was used as the shortest route in the initial network and in the composed principal road network (herein, this is called the volume of link traffic), and then compare them (Taguchi and Ohyama, 1993). Fig. 18 shows the results in which the target region was the Tokyo Metropolitan Area. Figure 18(a) shows the volume of link traffic in the initial network, while Fig. 18(b) shows the volume of link traffic in the composed principal road network. Links with a link traffic volume of zero, however, are not displayed. By comparing Fig. 18(a) with Fig. 18(b), they look similar. Therefore, the composed principal road networks maintain characteristic of the shortest route of initial network too.

5. Conclusions

In this paper, we have proposed an algorithm for composing a "practical" principal road network from the DRM Database which contains detailed road networks using network topological information. The composed principal road network preserves the characteristics of the original road network, and is sufficiently practical for solving transportation problems.

We conclude this paper by examining possible further work.

First, construction of an algorithm to compose "practical" principal road network using only geometrical shape without attribute data is important. This algorithm can be applied to the digital road maps which have no attribute data.

Second, it is desirable to devise a method to choose an appropriate threshold value. At present, we should experimentally choose the threshold value. It might be possible to use information of the shape of the original road network.

Finally, it is interesting to examine a method of evaluating the shape of the networks. It is difficult to compare the shape of the initial network and the shape of the composed principal road network quantitatively, because we cannot relate the link in the initial network to the link in the composed principal road network.

Acknowledgments. This work was partially supported by Grant-in-Aid for Young Scientists (B) No. 20710125 of the Ministry of Education, Culture, Sports, Science and Technology.

References

Douglas, D. H. and Peucker, T. K. (1973) Algorithm for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer*, **10**(2), 112–122.

Iri, M. (director) and Koshizuka, T. (editor) (1993) *Computational Geometry and Geographic Information Processing*, 2nd ed., Kyoritsu Shuppan, Tokyo (in Japanese).

Ishida, M. and Yaguchi, A. (2007) Review of digital road map, *Traffic Engineering*, **42(4)**, 60–66 (in Japanese).

Lang, T. (1969) Rules for the robot draughtsmen, *The Geographical Magazine*, **42**(1), 50–51.

Taguchi, A. and Ohyama, T. (1993) Importance degree evaluation of road based on network structure, *Communications of the Operations Research Society of Japan*, **38**(9), 465–470 (in Japanese).

Toriumi, S. and Taguchi, A. (2006) Development of a three-dimensional geographic database with a spatial precision, *Development of a Three-Dimensional Geographic Database with a Spatial Precision*, **14**(2), 19–30 (in Japanese).