PERGAMON

Engineering Applications of Artificial Intelligence 13 (2000) 165–178

Engineering Applications of
ARTIFICIAL INTELLIGENCE

www.elsevier.com/locate/engappai

Automated vectorization of cartographic maps by a knowledge-based system

Kyong-Ho Lee*, Sung-Bae Cho, Yoon-Chul Choy

Department of Computer Science, Yonsei University, 134 Shinchon-dong, Seodaemun-ku, Seoul, 120-749, South Korea

Received 1 September 1998; accepted 1 October 1999

Abstract

Developing an automated vectorizing system as an input method for a geographic information system (GIS) is of extreme importance due to the fact that an input process takes a lot of time and cost in constructing a GIS. Most vectorizing systems require users to set the parameters as appropriately as possible for a particular map image, but it is quite difficult for a novice to adjust the parameters appropriately. This paper proposes a knowledge-based system for automated vectorization, allowing an appropriate choice of the parameters. Since thinning of the input image to produce a skeleton of unit width is a prerequisite for the automated vectorization among several steps, the performance of representative thinning algorithms is systematically evaluated in various map images, and appropriate rules for the maps are devised. Each rule in the knowledge base is characterized by the type of map, and by the resolution, line width, slope and protrusions. Experimental results with various map images show that the proposed system is superior in terms of performance and convenience of use.

Keywords: Automated vectorization; Knowledge bases; Thinning; Geographic information systems; Cartographic maps

1. Introduction

In a variety of fields, there has recently been a growing interest in geographic information systems (GIS), which facilitate the efficient processing of geographic information. A GIS is a decision-support system that helps users to make appropriate decisions in manipulating given geographic information. These have been applied to a variety of fields, such as city planning, facility management, and resource development (Aronoff, 1989; Tayler, 1991).

Generally, a GIS consists of four basic components: input, management, manipulation and analysis, and output. An efficient input method of transforming concrete data into an accessible form in a GIS is very important, for this is the most time-consuming part. Thus, software that directly vectorizes scanned raster images has most commonly been developed. These can be categorized into automated and semi-automated systems, according to their different modes of operation (Hori and Tanigawa, 1993; Eikvil et al., 1995).

Most conventional automated vectorizing systems apply the same methods to all maps, and require users to set the most appropriate parameters for the map image manually (Hori and Tanigawa, 1993; Boatta et al., 1992; Kasturi et al., 1990; Pavlidis, 1986; Suzuki, 1990). They do not usually reflect the different characteristics of the various maps, and reinforce the need for users to have some preliminary information about the system.

* Corresponding author. Tel.: +82-2-361-2712; fax: +82-2-393-7663.

E-mail addresses: lkh@rainbow.yonsei.ac.kr (K.-H. Lee), sbcho@csai.yonsei.ac.kr (S.-B. Cho), ycchoy@rainbow.yonsei.ac.kr (Y.-C. Choy).

0952-1976/00/$ - see front matter
PII: S0952-1976(99)00049-4

In this paper, cartographic maps are classified into three types: contour, cadastral, and water and sewer maps. A knowledge base is constructed using their characteristic features, and an automated vectorizing system is implemented based on it. Furthermore, because thinning an input image to produce a skeleton of unit width is a prerequisite for the automated vectorization, the performance of various representative thinning algorithms has been systematically evaluated for various cartographic maps, to construct appropriate rules for the different types of maps.

The proposed system has produced vector data that is as exact as can be expected from very sophisticated manual digitization, because it is based upon the knowledge base. Moreover, it provides a convenient user interface that enables appropriate parameters to be chosen according to the visual information contained in the original map image.

The organization of this paper is as follows. Section 2 describes the results of the research aimed at discovering appropriate thinning algorithms for various cartographic maps. The development of the knowledge base is described in Section 3, and the design and implementation of the system are detailed in Section 4. Experimental results on a variety of cartographic maps are discussed in Section 5, and the conclusions are summarized in Section 6.

2. Thinning algorithms

Thinning is a prerequisite for an automated vectorizing system. Among the most recent publications (Chen and Hsu, 1990; Chu and Suen, 1986; Davies and Plummer, 1981; Lam and Suen, 1995; Lee et al., 1993; Naccache and Shinghal, 1984; Smith, 1987; Toriwaki and Yokoi, 1981) on the evaluation of thinning algorithms, (Lee et al., 1993) considers a comparison using performance characteristics for character data. Lam and Suen (1995) report the performance of 10 parallel thinning algorithms on an OCR system. However, there is no publication on the evaluation of thinning algorithms for the vectorization of cartographic maps. This section summarizes the results of the authors' recent research, which has attempted to discover the most appropriate thinning algorithm for automated vectorization.

In particular, the maps mostly used in GIS are contour maps, cadastral maps, and water and sewer maps, which have different characteristics. The performance of representative algorithms has been systematically evaluated on these three kinds of maps, and the appropriate algorithms for the maps are suggested. The following subsections describe in detail the criteria for the performance evaluation, the selection of algorithms and the evaluation results.

2.1. Thinning criteria

Previous publications (Kwok, 1988; Lam and Suen, 1995; Lee et al., 1993; Rosenfeld, 1975; Shin and Wong, 1994) have considered the important properties that the skeletons produced by thinning algorithms must possess. Based on these properties, this subsection describes the properties and criteria for a performance evaluation of thinning algorithms on cartographic maps. First, because cartographic maps represent spatial data, the skeleton that is produced must preserve the topology of a map, as well as its geometry, for accurate vectorization.

In order to preserve the topology, the skeleton should preserve the connectivity of an input image. Furthermore, because branches produced by boundary noise may change the topology, the thinning algorithm must not be sensitive to boundary noise. For the preservation of the geometry, the skeleton must exactly reflect the original pattern of the map image. In particular, when cartographic maps are reduced on a large scale in the real world, extensive erosion of end points must be prohibited, in order to create accurate vector data. A simple description on the criteria for the evaluation of algorithms follows.

1. Processing time: This is the amount of time taken for thinning. Ideally, sequential and parallel thinning might be implemented in serial and parallel computers, respectively. Because all the algorithms have been implemented on a serial computer, processing time is not that significant here.
2. Connectivity: For preserving the overall topology, the skeleton should preserve the connectivity of the original pattern. The algorithms were examined on eight-neighbor connectivity (Chen and Hsu, 1989).
3. Noisy branches: As described above, it is very important for the skeleton to maintain the topological properties of the original map image. Noisy branches in a skeleton image, which are caused by boundary noise, may change the topology of a cartographic map. Therefore, in order to choose an algorithm that is insensitive to boundary noise, the number of noisy branches in the skeleton image is computed.
4. Similarity (Plamondon and Suen, 1989): Generally, because cartographic maps are drawn on a reduced scale, careful consideration should be given to obtain accurate vector data. For this reason, the skeleton must run along the medial axes of the original pattern. Comparisons have been made between skeletons produced by the thinning algorithms and reference skeletons prepared by human experts.
5. Erosion of end points: For the preservation of geometrical properties, extensive erosion of the original

pattern should be prevented. Therefore, the extent of the erosion of end points is computed.

2.2. Algorithms for evaluation

Thinning algorithms can be divided into two categories: iterative and non-iterative algorithms. The iterative thinning algorithms, which produce a skeleton by repeatedly deleting pixels from the boundaries of patterns, are generally adequate for automated vectorization of cartographic maps. These iterative algorithms can be further divided into two categories: sequential and parallel thinning algorithms (Lam et al., 1992; Lam and Suen, 1995; Stefanelli and Rosenfeld, 1971). Sequential algorithms examine contour points for deletion by either raster scanning or contour-following algorithms. Parallel algorithms, however, have different modes of operation for preserving the connectivity of patterns: to preserve connectivity, they use either larger neighborhoods than 3×3, or use more than one pass over the pattern. As a result of these considerations, representative recent sequential and parallel algorithms have been selected from among those in the literature. The selected algorithms were implemented in C, and were tested on the same data as their papers for a fair evaluation.

1. SPTA (Naccache and Shinghal, 1984): The SPTA is a sequential algorithm that uses two raster scans per cycle, left-to-right and top-to-bottom, and decides on the deletion of pixels using a boolean expression in each scan.
2. CGT (Xu and Wang, 1987): This algorithm, proposed by Xu and Wang, is a sequential algorithm that uses a contour-generation method to save processing time. In this algorithm, the four types of contour pixels (east, north, west and south) are placed in buffers. Each type of buffer point is sequentially processed and checked by means of windows for connectivity and end-point preservation.
3. Arcelli (Arcelli and Baja, 1981): Arcelli's method preserves significant contour protrusions and prominences in the thinning process, in which prominences are first detected and labeled. These significant protrusions are defined as connected subsets of the contours that are beyond a threshold distance from the core. This is a sequential algorithm that uses a crossing number (Rutovitz, 1966) to determine whether the pixel is to be deleted.
4. Chen (Chen and Hsu, 1988): Irrespective of immunity to boundary noise, Zhang's method (Zhang and Suen, 1984) seriously erodes two-pixel wide diagonal lines and completely deletes 2×2 squares. Chen's method is a parallel algorithm with two subcycles which avoids the serious erosion found in Zhang's method. In this algorithm, the information about pixel deletion is stored in look-up tables.
5. Lü (Lü and Wang, 1986): Lü's method is a parallel thinning algorithm which operates in two subcycles over 3×3 windows, and improves on both the serious erosion of the Zhang method and the processing time of the Stefanelli method (Stefanelli and Rosenfeld, 1971).
6. Holt (Holt et al., 1987): Holt's method is a parallel thinning algorithm which considers information from a 4×4 window in order to determine pixel deletion. This algorithm considers the edge information on neighboring pixels in order to use only one subiteration per cycle.

2.3. Performance evaluation

The performance of the representative thinning algorithms on cartographic maps has been evaluated in terms of the processing time and the quality of the skeleton produced. In terms of processing time, CGT is the fastest, as shown in Tables 1–3. However, CGT is highly subject to boundary noise, and produces a skeleton with many noisy branches. In vectorizing cartographic maps, the thinning algorithm must be insen-

Table 1
Performance of thinning algorithms for contour maps

Algorithm	Criteria				
	Time	Connectivity	Number of noisy branches	Similarity ranking	Erosion of end points (pixels)
SPTA	15.2	Perfect 8	152	1	2.5
CGT	7.5	Perfect 8	229	3	1.5
Arcelli ($L^{a} = 3$)	12	Perfect 8	41	5	2
Arcelli ($L = 5$)	12	Perfect 8	23	5	2
Chen	11	Perfect 8	49	7	4.5
Lü	14.6	Imperfect 8	90	4	2.5
Holt	10.3	Perfect 8	79	2	3

[a] L is the distance from the core (Arcelli and Baja, 1981).

Table 2
Performance of thinning algorithms for cadastral maps

Algorithm	Criteria				
	Time	Connectivity	Number of noisy branches	Similarity ranking	Erosion of end points (pixels)
SPTA	30.6	Perfect 8	62	1	1
CGT	12.6	Perfect 8	141	6 (poor)	0.5
Arcelli (L = 3)	24	Perfect 8	35	3	1.5
Arcelli (L = 5)	24	Perfect 8	20	3	1.5
Chen	22	Perfect 8	20	5	1.5
Lü	31	Imperfect 8	19	7 (poor)	1.5
Holt	20	Perfect 8	22	2	1.5

Table 3
Performance of thinning algorithms for water and sewer maps

Algorithm	Criteria				
	Time	Connectivity	Number of noisy branches	Similarity ranking	Erosion of end points (pixels)
SPTA	31.5	Perfect 8	75	1	14
CGT	14.6	Perfect 8	187	3	14
Arcelli (L = 5)	26	Perfect 8	10	6 (poor)	49
Chen	19.1	Perfect 8	10	4	25
Lü	36	Imperfect 8	9	5	14
Holt	19	Perfect 8	13	1	15

sitive to boundary noise that is not necessary to preserve the topology and geometry of the map image. CGT is more appropriate for fields where the skeleton that is produced must contain more information about the original pattern.

In terms of the quality of the skeleton, in the case of a contour map, protrusions may exist in a scanned line image because a contour map consists of both dotted and solid lines. It is essential to select an algorithm that is insensitive to boundary noise, and to prevent extensive erosion. As shown in Table 1, the bigger the value of the threshold, the less sensitive Arcelli's method is to noise. In the case of a cadastral map, Holt's method reflects the characteristics of a cadastral map well at the corners of buildings, which are mostly used in such maps. In water and sewer maps, both Holt's method and Chen's method are superior in terms of processing time, connectivity, similarity, and noise insensitivity. In particular, in spite of boundary noise, Chen's method produces rectangular skeletons at the +-shaped or T-shaped intersections. Table 4 summarizes the appropriate thinning algorithm for each type of map.

3. The knowledge base

A knowledge-based system has been developed for automated vectorization, using a rule-based model. This system takes as input various kinds of map images, and produces their vector representations as output. Various image-processing operations are performed to extract the characteristics of the image, and these data are analyzed under the control of a rule-based system, which operates in conjunction with a global data structure to monitor the entire process (Nazif and Levine, 1984; Nii, 1986a, 1986b). A conventional inference mechanism has been used within the rule-based system; this explains about the maps using a knowledge base that contains rules describing all the characteristics of the map images (Nazif and Levine, 1984; Niyogi and Srihari, 1996).

This section describes the knowledge base that is utilized in the proposed system. There are many things to be considered while constructing a knowledge base. First, various types of maps are to be classified into contour, cadastral, and water and sewer maps, and then the characteristics of each type should be examined. Most automated vectorizing processes are currently applied to line segments in cartographic maps.

Table 4
Summary of the best algorithms for different maps

Contour map	Cadastral map	Water and sewer map
Arcelli	Holt	Chen

Consequently, they are classified into three categories: contour maps, mostly composed of curved lines, cadastral maps, consisting largely of rectangles, and water and sewer maps, in which straight lines cross each other. The rules in the knowledge base are expressed in terms of first-order predicates, based upon a careful examination of different characteristics such as resolution, scale, line width, slope, and protrusions in cartographic maps.

As described in the previous section, the varying characteristics of the various types of maps pose an obstacle in applying the same thinning algorithm to all the maps. Therefore, these characteristics must be reflected in the vectorization process. For example, most cadastral maps are made up of building blocks, which are polygons, so this characteristic of cadastral maps must be taken into account in the application of the vectorizing process. Fig. 1 shows the result of applying the knowledge base to a cadastral map: (a) is a skeletal image of the corner of building, (b) is the vector data before applying the knowledge base, and (c) is the result of applying it. It can be easily appreciated that (c) reflects the characteristics of the cadastral map more accurately than (b).

Moreover, because water and sewer maps are composed of cross-shaped or T-shaped intersections, where line segments cross each other, it is desirable to approximate multiple line segments with one segment as long as this remains within the boundary of the original pattern. Because both solid and dotted lines may overlap in the case of a contour map, protruding patterns can exist in the scanned maps. The characteristics of contour maps have been examined in terms of protrusions and the slopes of the patterns, since such maps are mostly composed of curved lines.

This paper considers contour maps whose scales are between 1 : 1000 and 1 : 5000, cadastral maps, and water and sewer maps between 1 : 3000 and 1 : 12,000. Due to the fact that cartographic maps are drawn on a reduced scale, close attention is required to obtain accurate vector data. Meanwhile, the pattern width of a scanned image is affected by the resolution of the scanner: in the case of maps with the same size, the larger the resolution is, the wider a pattern becomes. Some examples of knowledge rules are shown in Fig. 2.

In constructing the knowledge base, a number of thresholds were used, which were manually adjusted according to the output of the system, on a large number of test images. The detailed description of how to determine the value of threshold is as follows. In the case of Rule 9 in Fig. 2 where patterns are steep, thin, and excessively protrusive, Arcelli's method is employed with a threshold of 3 to preserve significant protrusions, as mentioned in Section 2. Taking the width of the pattern into consideration, a threshold of 1.5 is used so as to eliminate noisy branches.

In addition, it is essential to apply line approximation to the skeleton in order to produce vector data. For the purpose of determining suitable thresholds for line approximation, the accuracy of the vector data was examined in relation to a variation of the threshold at every step, and an example of the experimental results is shown in Table 5. The best results were obtained in the case where the maximum and mean values of the threshold were 0.3 and 0.15, respectively. Therefore, the "Method 9" in Rule 9 uses 0.3 and 0.15 as the maximum and mean values of thresholds for line approximation, respectively.

4. Implementation of the knowledge-based vectorizing system

This section describes the proposed system, which produces vector data by utilizing a knowledge base constructed as outlined in the previous section. The

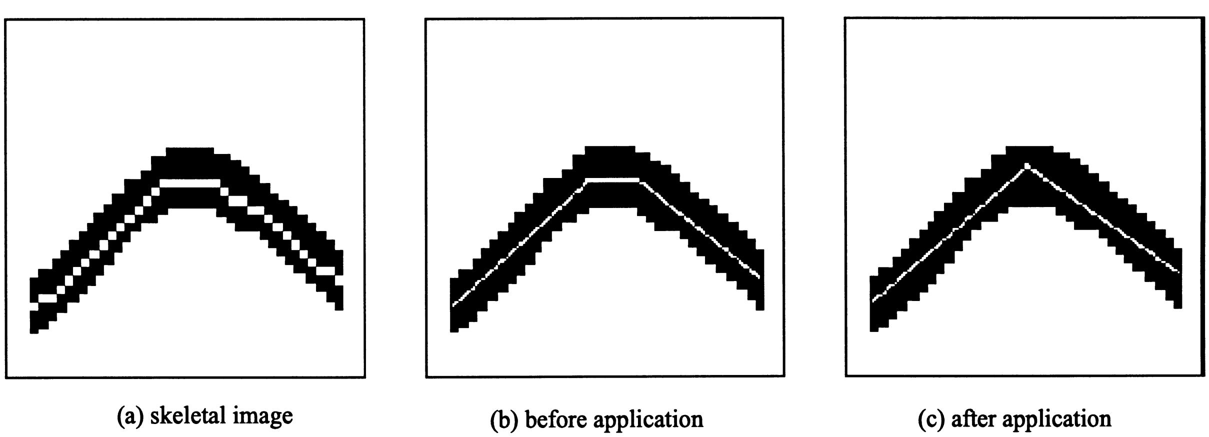


Fig. 1. Comparison of the skeletal images depending on the application of the knowledge base for a corner of a building block in a cadastral map: (a) skeletal image, (b) before application, (c) after application.

Rule 9:
IF (a map X is of type "contour")
AND (its line width is "thin")
AND (its slope is "steep")
AND (its protrusion is "high")
THEN apply method 9

Rule 10:
IF (a map X is of type "contour")
AND (its line width is "middle")
AND (its slope is "gentle")
AND (its protrusion is "low")
THEN apply method 10

Rule 36:
IF (a map X is of type "contour")
AND (its line width is "very thick")
AND (its slope is "steep")
AND (its protrusion is "high")
THEN apply method 36

Rule 39:
IF (a map X is of type "cadastral")
AND (its line width is "very thick")
THEN apply method 39

Rule 43:
IF (a map X is of type "water and sewer")
AND (its line width is "thick")
THEN apply method 43

Rule 45:
IF (the line approximation mode is on)
AND (an arc has been selected)
THEN the line approximation is performed through the application of the least-squares method

Fig. 2. Part of the knowledge rules constructed.

system is composed of three parts, as shown in Fig. 3. The first is the preprocessing part, where noises are eliminated from scanned raster images, and the map image is edited and modified to facilitate accurate line tracing. The second is the vectorizing part, where vector data are produced from the preprocessed map image by using the knowledge base. Finally, the last section is the postprocessing part, where users are able to correct errors manually through the vector editor. The system has been implemented in C on a SUN-compatible workstation.

4.1. Preprocessing

The system provides facilities that can remove noisy objects and edit a map image. In the process of noise

Table 5
Determination of suitable thresholds for line approximation

Threshold (pixels)		Total number of vector data (pixels)	Number of deviating pixels
Maximum	Mean		
0.8	0.4	4732	258
0.7	0.3	5027	211
0.6	0.25	5078	147
0.5	0.2	5132	143
0.4	0.2	5204	112
0.3	0.15	5400	81
0.25	0.15	5432	81

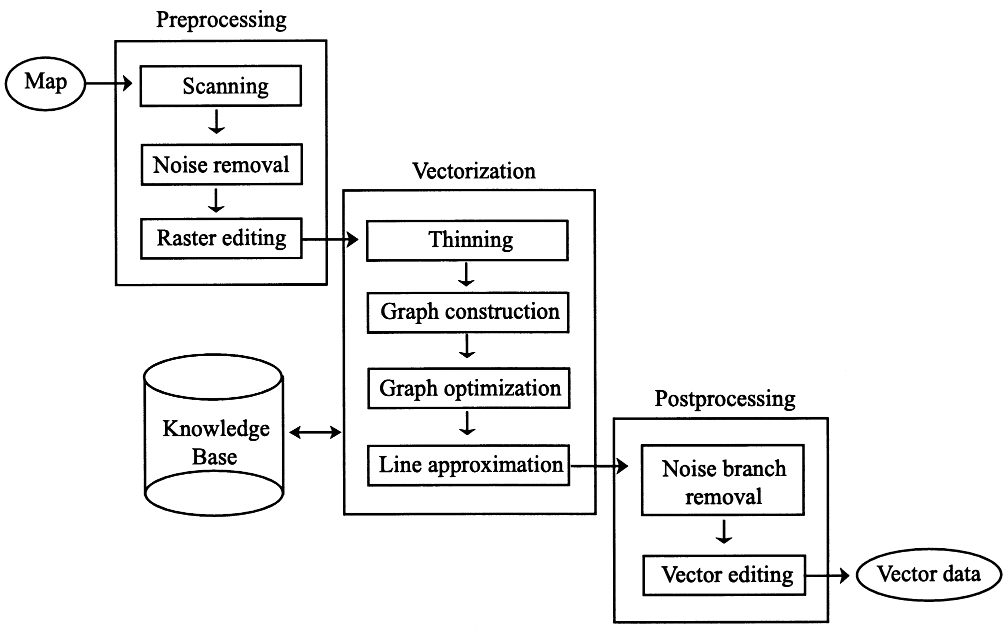


Fig. 3. The structure of the proposed system.

removal, holes and speckles can be eliminated. A speckle is removed in cases where both width and height are smaller than some threshold. A hole is filled in the same manner. However, it is necessary to set the thresholds with care according to the condition of the maps, because large thresholds may result in the removal of dotted lines, small building blocks, and letters.

4.2. Automated vectorization

This system provides a convenient user interface based on the knowledge base supplied by the system designer. As shown in Fig. 2, 36 different kinds of contour maps can be processed according to the thickness of patterns, their slopes, and any protrusions. In case of cadastral maps and water and sewer maps, four types of maps can be processed, according to the thickness of the pattern.

The vectorizing process consists of four steps. The first is the thinning step, where a skeletal image is generated from the cartographic map. The second is the transformation of this skeletal image into graph. The third step involves optimizing the graph, where extra edges unnecessary for preserving the topology of the map are deleted. The last step is the extraction of feature points and line approximation. A detailed description of each step follows.

1. Thinning: The purpose of thinning a map image into a skeleton with a unit width is to reduce and simplify the data needed to be processed. An appropriate thinning algorithm is applied to each map, as described in Section 2.
2. Graph representation: The skeletal image is transformed into a graph in order to produce vector data effectively (Suzuki, 1988). During raster scanning, black pixels in a skeletal image correspond to nodes and neighboring nodes are connected by an edge. Fig. 4 shows the result of transforming a skeletal image into a graph. (a) is a skeletal image, (b) is the graph representation of (a), and (c) is the data structure of representing the graph, respectively. Fig. 5 shows the tracing order of neighborhood pixels.
3. Graph optimization: The edges that are unnecessary for preserving the topology of a cartographic map in the graph representation may cause ambiguity in line-tracing at intersection points, and may create incorrect vector data. Therefore, the deletion of edges makes it possible to identify the feature points clearly, and to reduce the amount of vector data produced. Fig. 6 shows the result of applying the edge deletion to the graph representation.
4. Line approximation: First of all, feature points are extracted from the graph. There are two kinds of feature points. One is a node that has not yet been traced during raster scanning, and the other is a node where the degree determined by the number of connected nodes is not equal to two. Secondly, sets of arcs, which consist of nodes between features, are created. In particular, because any feature point whose degree is larger than two is an intersection

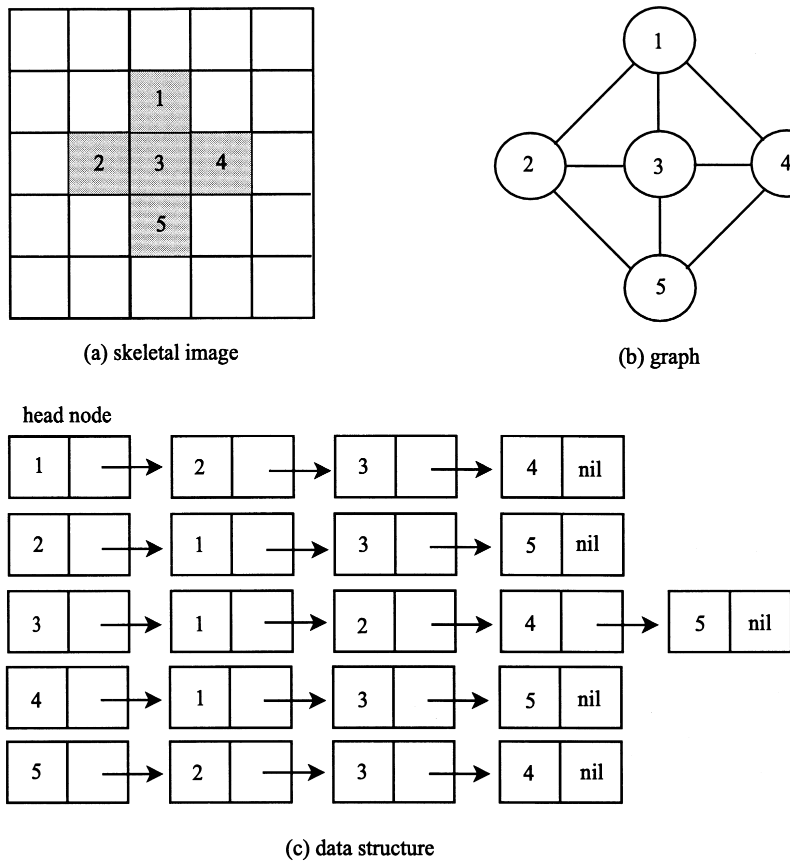


Fig. 4. Transformation of skeleton into graph: (a) skeletal image, (b) graph, (c) data structure.

point, nodes other than the current direction must be stored in the stack. This tracing will be repeated until there are no nodes remaining in the graph. Finally, line approximation is performed through the application of the least-squares method (Burden and Faires, 1993) to the resulting set of arcs.

The error in a line approximation is defined as the perpendicular distance from each node to the approximated line, and its maximum and mean values are used as the thresholds. If either of two errors is larger than its threshold, the approximated line is considered to be inadequate, and the range of approximation is reduced by half. On the other hand, if both errors are smaller than their thresholds in the reduced range, the range is expanded by half of its reduced size, and then the line approximation is repeated. Through these iterative rearrangements of the range, multiple nodes in an arc might be approximated into one straight line. This method is performed rapidly in the case where the slope of an arc varies gently. Fig. 7 shows the result of applying the knowledge base to a cadastral map: (a) is a map image, (b) is the preprocessed image of (a), (c) is the skeletal image, and (d) is the result of line approximation. In particular, the vector data produced is shown in Table 6.

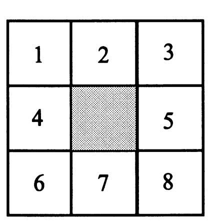


Fig. 5. Neighborhood pixels.

4.3. Postprocessing

Due to boundary noise in the original pattern, noisy branches may exist in a skeleton image, and an additional vector data can be generated. Therefore, any arc whose length is shorter than some threshold can be deleted. Also, the vector editor is provided for users to modify vector data manually.

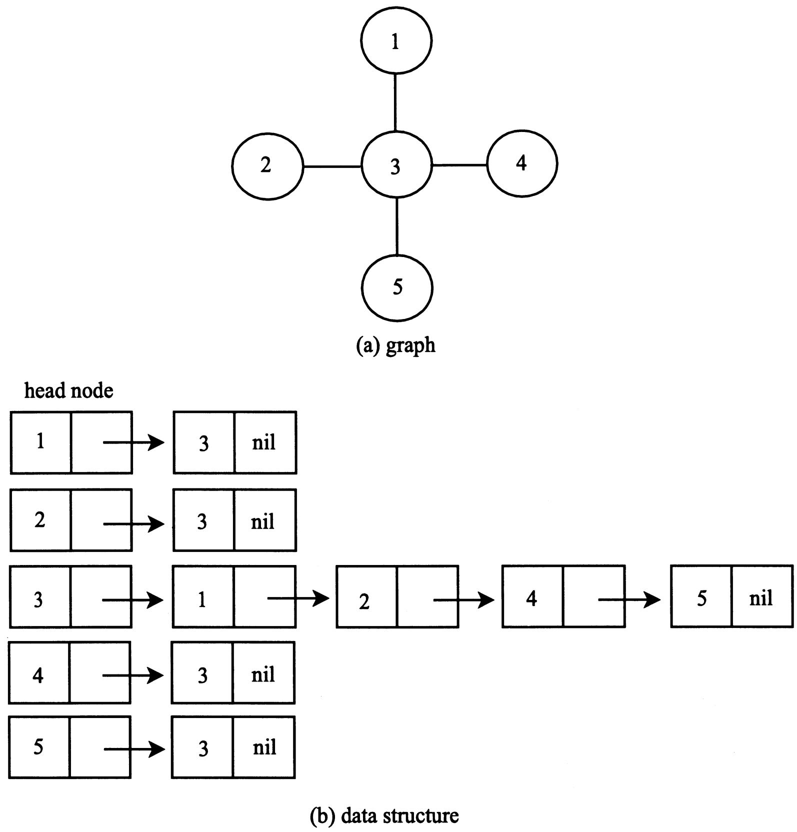


Fig. 6. Edge deletion: (a) graph, (b) data structure.

5. Experimental results

Tests have been conducted on various cartographic maps with scales of between 1 : 1000 and 1 : 12,000, as shown in Fig. 8. In particular, the proposed vectorizing system has been compared with existing ones in terms of performance and convenience of use. From the perspective of convenience, unlike existing systems which require users to set appropriate parameters for the map image manually, the proposed system enables appropriate parameters to be chosen according to the visual information in its original map image, and thus provides users with lucid guidelines.

However, in spite of the convenience, the proposed method would be useless if its performance were to be inferior to that of existing system. Therefore, its performance has been evaluated on a variety of cartographic maps in terms of the accuracy of the vector data and the number of line segments.

First, the error of the vector data was defined as the number of pixels deviating from the original map; based on that, the accuracy of vector data was calculated. Table 7 shows that, on an average, the total number of pixels is 6596 and the number of deviating pixels is 152, resulting in an accuracy of 97.7%. In the case of water and sewer 1, poorer result were obtained due to the bad quality of the map.

Second, in terms of the number of line segments, this has been significantly reduced because an appropriate thinning algorithm was applied to each map, and errors in the vector data were corrected by using the knowledge base. Fig. 9 shows the result of applying the knowledge base to a cadastral map: (a) is a map image from which noises are removed, (b) is the skeletal image of (a), (c) is the vector data before applying the knowledge base, and (d) is the result of applying it. In particular, (d) reflects the characteristics of the cadastral map more accurately than (c). Moreover, the number of edges in (c) is 170 and that in (d) is 155. Therefore, a significant reduction of the amount of vector data produced can be gained by applying the knowledge base.

6. Conclusions

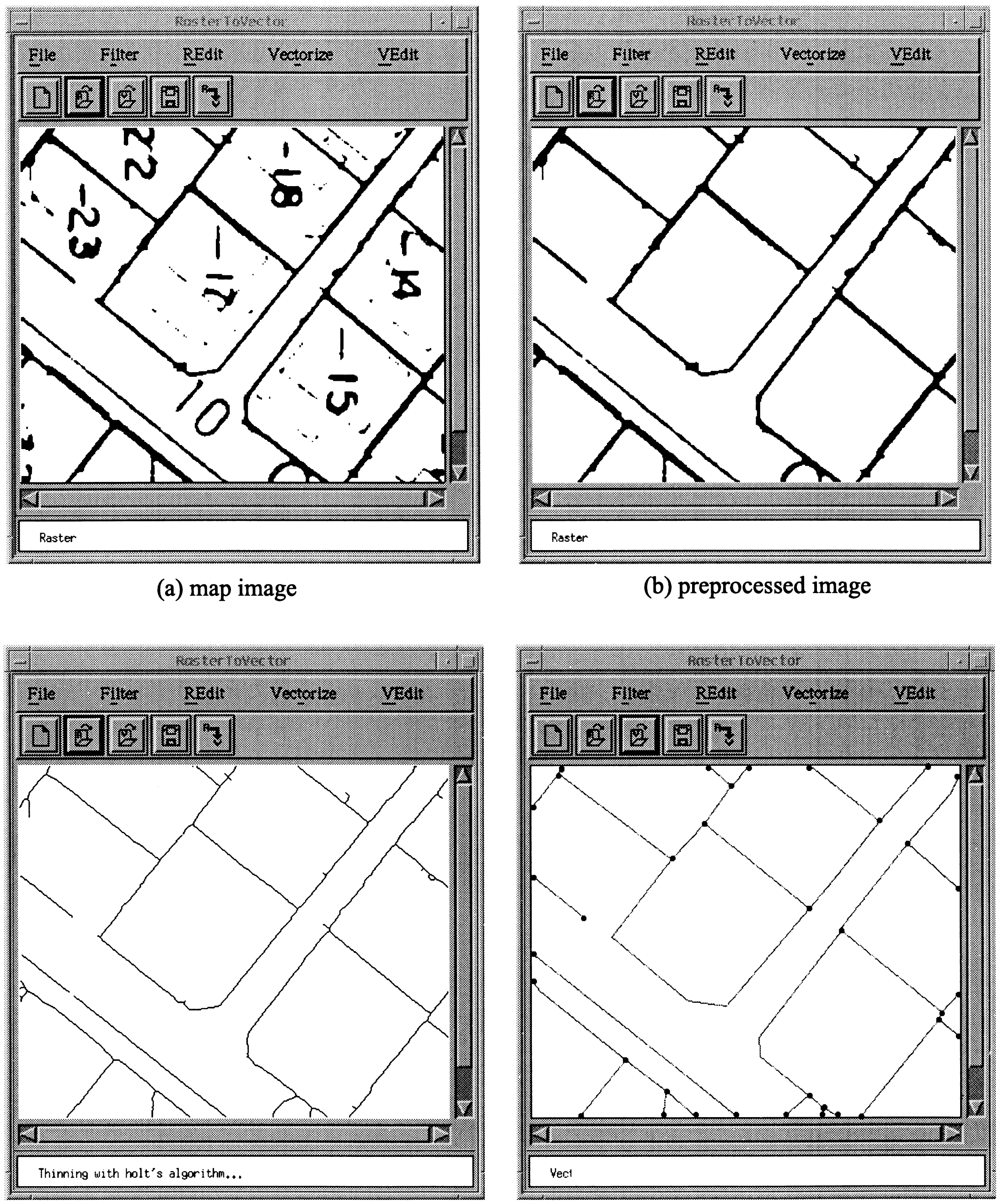
Since cartographic maps have different characteristics according to their types, it is very important to take this factor into consideration when designing a vectorizing system. Most vectorizing systems apply the same method to all maps, and require users to set appropriate parameters for the map image manually, leading to problems where only a uniform method is

Table 6
Vector data obtained in Fig. 7

Arc ID	Start point		End point		Number of intermediate points	Intermediate point					
	x	y	x	y		x	y	x	y	x	y
1	30	0	26	7	0						
2	26	7	0	39	0						
3	26	7	131	86	0						
4	131	86	166	55	0						
5	166	55	190	15	0						
6	190	15	170	0	0						
7	190	15	207	0	0						
8	166	55	266	133	0						
9	266	133	333	50	0						
10	333	50	267	0	0						
11	333	50	378	0	0						
12	266	133	131	86	3	189	229	149	222	77	163
13	414	9	362	75	1	403	26				
14	362	75	276	156	0						
15	276	156	263	312	2	218	262	219	277		
16	263	312	244	334	0						
17	263	312	279	325	0						
18	279	325	281	334	0						
19	279	325	296	334	0						
20	276	156	392	231	0						
21	392	231	413	217	0						
22	392	231	392	242	0						
23	392	242	311	340	0						
24	392	242	415	260	0						
25	362	75	414	118	0						
26	1	105	51	144	0						
27	3	180	198	334	0						
28	1	205	86	278	1	7	213				
29	86	278	47	334	0						
30	86	278	127	308	0						
31	127	308	128	334	0						
32	127	308	159	334	0						

Table 7
Accuracy of the vector data

Maps	Total number of pixels	Number of vector	Number of the deviated	Accuracy (%)
Contour 1	17413	5400	81	98.5
Contour 2	109626	32028	644	98
Contour 3	29724	6319	82	98.7
Contour 4	27779	7210	201	97.2
Cadastral 1	17194	1944	3	99.8
Cadastral 2	8397	2281	93	96
Cadastral 3	23967	2023	4	99.8
Cadastral 4	13829	3102	106	96.6
Water and sewer 1	10323	2802	265	90.6
Water and sewer 2	19681	2246	37	98.4
Average	27793	6596	152	97.7



(a) map image

(b) preprocessed image

(c) skeletal image

(d) vector data

Fig. 7. Results of vectorization: (a) map image, (b) preprocessed image, (c) skeletal image, (d) vector data.

available and the users must be knowledgeable about the system.

Therefore, this paper has proposed a new method as a solution to these problems. The proposed system has the following characteristics. First, it uses a knowledge base related to cartographic maps. It makes use both of appropriate thinning algorithms and of the knowledge about cartographic maps. The maps are categorized as contour, cadastral, or water and sewer maps, and the map images are subdivided in terms of the thickness of pattern, the slope, and protrusions, and then different criteria are applied to each individual map.

Second, a convenient user interface is provided for novice users, as an assistance to promoting the system's user-friendliness. That is, users are able to select

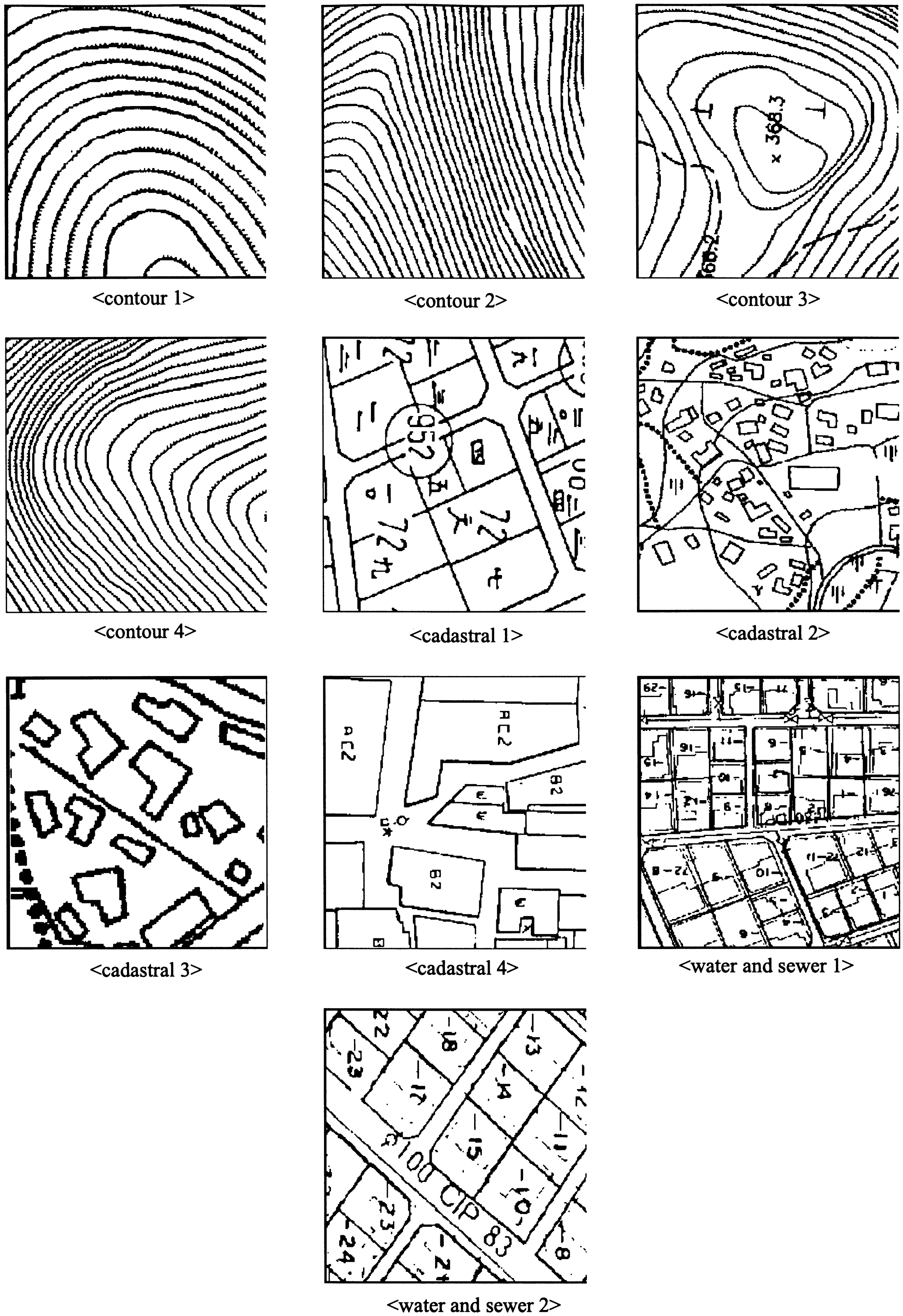
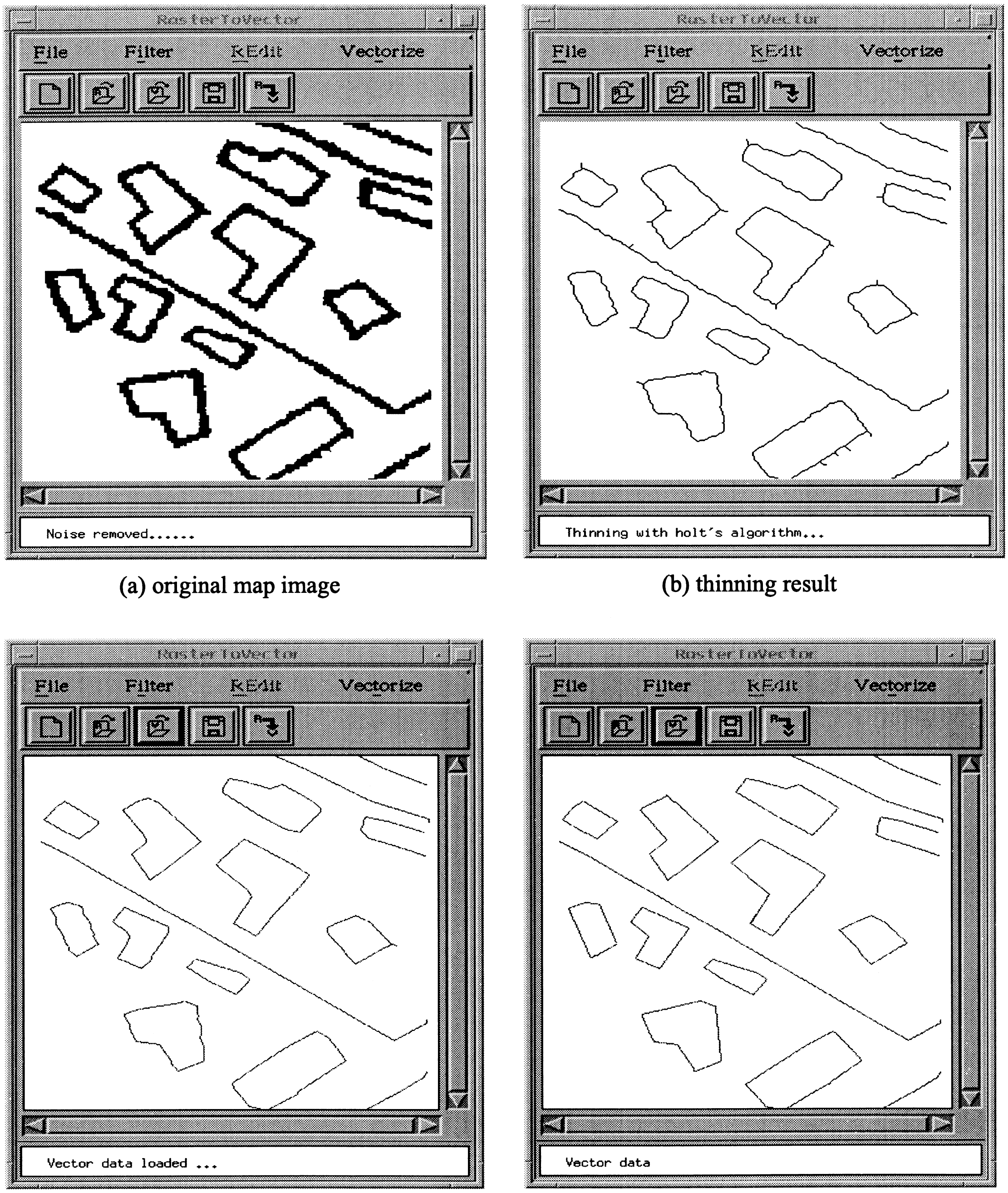


Fig. 8. Map images used in the experiment.



(a) original map image

(b) thinning result

(c) before application of the knowledge base

(d) after application of the knowledge base

Fig. 9. An application of knowledge base: (a) original map image, (b) thinning image, (c) before application of the knowledge base, (d) after application of the knowledge base.

from a menu on the basis of their visual recognition of the characteristics of maps, rendering the proposed system superior in terms of performance and convenience.

References

Arcelli, C., Baja, G.S.D., 1981. A thinning algorithm based on prominence detection. Pattern Recognition 13 (3), 225–235.

Aronoff, S., 1989. Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa.

Boatta, L., Consorti, L., Buono, M.D., Zenzo, S.D., Eramo, V., Esposito, A., Melcarne, F., Meucci, M., Morelli, A., Mosciatti, M., Scarci, S., Tucci, M., 1992. An interpretation system for land register maps. IEEE Computer 25 (7), 25–33.

Burden, R.L., Faires, J.D., 1993. Numerical Analysis. Prindle, Weber & Schmidt, Boston, MA.

Chen, Y.S., Hsu, W.H., 1993. A modified fast parallel algorithm for thinning digital patterns. Pattern Recognition Letters 7, 99–106.

Chen, Y.S., Hsu, W.H., 1989. A 1-subcycle parallel thinning algorithm for producing perfect 8-curves and obtaining isotropic skeleton of an L-shape pattern. In: International Conference on CVPR. San Diego, USA, 208–215.

Chen, Y.S., Hsu, W.H., 1990. A comparison of some one-pass parallel thinning. Pattern Recognition Letters 11 (1), 35–41.

Chu, Y.K., Suen, C.Y., 1986. An alternative smoothing and stripping algorithm for thinning digital binary patterns. Signal Processing 11 (3), 207–222.

Davies, E.R., Plummer, A.P., 1981. Thinning algorithms: a critique and a new methodology. Pattern Recognition 14 (1), 1150–1152.

Eikvil, L., Aas, K., Koren, H., 1995. Tools for interactive map conversion and vectorization. In: Third International Conference on Document Analysis and Recognition, 927–930.

Holt, C.M., Stewart, A., Clint, M., Perrott, R.H., 1987. An improved parallel thinning algorithm. Communication of the ACM 30 (2), 156–160.

Hori, O., Tanigawa, S., 1993. Raster-to-vector conversion by line fitting based on contours and skeletons. In: Second International Conference on Docoment Analysis and Recognition. Tsukuba, Japan, 353–358.

Kasturi, R., Bow, S.T., Masri, W.E., Shah, J., Gattiker, J.R., Mokate, U.B., 1990. A system for interpretation of line drawings. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (10), 973–992.

Kwok, P.C.K., 1988. A thinning algorithm by contour generation. Communication of the ACM 31 (11), 1314–1324.

Lam, L., Lee, S.W., Suen, C.Y., 1992. Thinning methodologies–a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (9), 869–885.

Lam, L., Suen, C.Y., 1995. An evaluation of parallel thinning algorithms for character recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (9), 914–919.

Lee, S.W., Lam, L., Suen, C.Y., 1993. A systematic evaluation of skeletonization algorithms. International Journal of Pattern Recognition and Artificial Intelligence 7 (5), 1203–1225.

Lü, H.E., Wang, P.S.P., 1986. A comment on a fast parallel algorithm for thinning digital patterns. Communication of the ACM 29 (3), 239–242.

Naccache, N.J., Shinghal, R., 1984. SPTA: a proposed algorithm for thinning binary patterns. IEEE Transactions on System, Man, and Cybernetics 14 (3), 409–418.

Nazif, A.M., Levine, M.D., 1984. Low level image segmentation: an expert system. IEEE Transaction on Pattern Analysis and Machine Intelligence 6 (5), 555–577.

Nii, H.P., 1986a. Blackboard systems. Part 1: The blackboard model of problem solving and the evolution of blackboard architectures. AI Magazine 7 (2), 38–53.

Nii, H.P., 1986b. Blackboard systems. Part 2: Blackboard application systems, blackboard systems from a knowledge engineering perspective. AI Magazine 7 (3), 82–106.

Niyogi, D., Srihari, S.N., 1996. An integrated Approach to document decomposition and structural analysis. International Journal of Imaging Systems and Technology 7, 330–342.

Pavlidis, T., 1986. A vectorizer and feature extractor for document recognition. Computer Vision Graphics and Image Processing 35, 111–127.

Plamondon, R., Suen, C.Y., 1989. Thinning of digitized characters from subjective experiments: a proposal for a systematic evaluation protocol of algorithms. In: Krzyzak, A., Kasvand, T., Suen, C.Y. (Eds.), Computer Vision and Shape Recognition. World Scientific Publishing, Singapore, pp. 261–272.

Rosenfeld, A., 1975. A characterization of parallel thinning algorithms. Information and Control 29, 286–291.

Rutovitz, D., 1966. Pattern recognition. Journal of Royal Statistical Society 1129 (A), 504–530.

Shin, F.Y., Wong, W.T., 1994. Fully parallel thinning with tolerance to boundary noise. Pattern Recognition 27 (12), 1677–1695.

Smith, R.W., 1987. Computer processing of line images: a survey. Pattern Recognition 20 (1), 7–15.

Stefanelli, S., Rosenfeld, A., 1971. Some parallel thinning algorithms for digital picture. Journal of ACM 18, 255–264.

Suzuki, S., 1988. Graph-based vectorization method for line patterns. In: IEEE Conference on CVPR, 616–621.

Suzuki, S., 1990. MARIS: Map recognition input system. Pattern Recognition 23 (8), 919–933.

Tayler, D.R.F., 1991. Geographic Information Systems: The Microcomputer and Modern Cartography. Pergamon Press, Oxford.

Toriwaki, J.I., Yokoi, S., 1981. Distance transformation and skeletons of digitized pictures with applications. In: Kanal, L.N., Rosenfeld, A. (Eds.), Process in Pattern Recognition. North-Holland, New York, pp. 189–264.

Xu, X., Wang, C., 1987. CGT: a fast thinning algorithm implemented on a sequential computer. IEEE Transactions on Systems, Man and Cybernectics 17 (5), 847–851.

Zhang, T.Y., Suen, C.Y., 1984. A fast parallel algorithm for thinning digital patterns. Communication of the ACM 27, 236–239.

Kyong-Ho Lee received his B.S. and M.S. degree in Computer Science from Yonsei University, Seoul, South Korea, in 1995 and 1997, respectively. Currently, he is working toward his Ph.D. degree in Computer Science at Yonsei University, Seoul, South Korea. His research interests include document image analysis, document management system, markup languages, and multimedia. He is a student member of the Korea Information Science Society, the Korea Information Processing Society, the ACM, and the IEEE Computer Society.

Sung-Bae Cho received the B.S. degree in Computer Science from Yonsei University, Seoul, South Korea, in 1988 and the M.S. and Ph.D. degrees in Computer Science from KAIST (Korea Advanced Institute of Science and Technology), Taejeon, South Korea, in 1990 and 1993, respectively. He worked as a Member of the Research Staff at the Center for Artificial Intelligence Research at KAIST from 1991 to 1993. He was an Invited Researcher of Human Information Processing Research Laboratories at ATR (Advanced Telecommunications Research) Institute, Kyoto, Japan, from 1993 to 1995. In 1998, he was invited as a Visiting Scholar at University of New South Wales, Canberra, Australia. Since 1995, he has been an Associate Professor in the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man–machine interfaces, evolutionary computation, and artificial life. Dr. Cho was awarded outstanding paper prizes from the IEEE Korea Section in 1989 and 1992, and another one from the Korea Information Science Society in 1990. He was also the recipient of the Richard E. Merwin prize from the IEEE Computer Society in 1993. He was listed in Who's Who in Pattern Recognition from the International Association for Pattern Recognition in 1994, and received the best paper awards at International Conference on Soft Computing in 1996 and 1998. He is a Member of the Korea Information Science Society, INNS, the IEEE Computer Society, and the IEEE Systems, Man, and Cybernetics Society.

Yoon-Chul Choy is a professor of Computer Science at Yonsei University in Korea. He received B.S. degree from Seoul National University, Korea in 1973, and M.S. and Ph.D. degree from the University of California, Berkeley in 1976 and 1979, respectively. After working as a research staff for Lockheed Corp. and Rockwell International, he joined Yonsei University in 1984. At present, he leads the Multimedia and Graphics Lab. at Yonsei University. His main areas of interests are multimedia, computer graphics, virtual reality, and geographic information system.