ISE

ISE-TR-90-87

An Optimum Data Reduction Algorithm for General Plane Curves

by

Sumiko Hiyama, Takao Hanada and Hitoshi Imai

November 30, 1990

INSTITUTE
OF
INFORMATION SCIENCES AND ELECTRONICS
UNIVERSITY OF TSUKUBA

An Optimum Data Reduction Algorithm for General Plane Curves

Sumiko Hiyama, Takao Hanada* and Hitoshi Imai**

Earthquake Research Institute, the University of Tokyo
Bunkyo-ku, Tokyo, 113 Japan

*Faculty of Electro-Communications, the University of Electro-Communications
Chofu-shi, Tokyo, 182 Japan

**Institute of Information Sciences and Electronics, University of Tsukuba
Tsukuba City, Ibaraki, 305 Japan

Keywords: algorithm, coastline, data reduction, optimum, plane curve

Abstract.

We introduce a data reduction algorithm for general plane curves without losing visual acceptability.Plane curves generally include not only planar curves but also complex zig-zag lines such as coast-lines or frontier lines between countries.

Although several data reduction methods were already given to date, their application is limited to planar curves or small amount of data.

Our new algorithm is applied for a large quantity of data which constitute geometrical lines and coast-lines. The results show that the algorithm is practical and effective on both planar curves and complex curves.

1.Introduction

In usual graphic systems, plane curves are represented by sequences of connected line segments and the data of curves are given in the coordinates of knots. In some cases such as digital map, a great many data are provided in order to be adapted to any computer graphic system. Too many data for systems of low resolution, however, need a large amount of storage in a computer and require a long time to display. Then it is necessary to reduce the number of data without losing visual acceptability of the display image.

Many data reduction algorithms have been already developed. One of them is Agui's[1] algorithm in image processing. This uses the quasi-coding method for figures based on the fractal dimension. Another algorithm by Tomek[4] generates line segments in the polygon which enclosing original lines.In these algorithms , the reduced data are not subsets of original data. However,in Sato's algorithm[2] and Roberge's one[3] respectively, reduced data are included in their original data. The former is suitable for curves of small amount of data such as alphabets, and the latter is only for planar curves.

The curves obtained from the reduced data must resemble closely original curves, and the data reduction must be carried out correspondingly to a graphic system.

In the paper we propose a new data reduction method which satisfies the following four conditions.

1. To be reduced as a subset of original data.
2. To be useful for plane curves, not only for planar but also for complex zig-zag curves.
3. To be a pre-processing of original data instead of sending them directly to display packages(Fig.1), but to be practical in computing time.
4. To be easily applicable for the user who knows only a number of data.

2.New optimum data reduction algorithm

2.1 Definition of data reduction

In a digital map system, coast lines and frontier lines are given as an assemblage of unbranched curves. Each in this assemblage is called a chain curve.

Let $\vec{P}$ be a chain curve. The plane curve is represented by a sequence of points $(p_0, p_1, p_2 \ldots, p_n)$,

p_0: the first point of the chain curve $\vec{P}$,

p_n: the last point of the chain curve $\vec{P}$,

n: the number of line segments formed using points p_{i-1} and p_i.

For the original curve $\vec{P} = (p_0, p_1, p_2, .., p_i, .., p_n)$, 'Data reduction' means to create a new curve $\vec{Q} = (q_0, q_1, q_2, .., q_j, .., q_m)$ according to some criterions(Fig.2). Here, the set $Q = \{q_0, q_1, q_2, \ldots, q_m\}$ is a subset of $P = \{p_0, p_1, p_2, \ldots, p_n\}$;$m \leqq n$.

In usual case, certain scheme $f(p_i)$ is defined for a point p_i.(For example $f(p_i)$ expresses a degree of the curve bending at p_i.)

Data is selected as follows:

if $f(p_i) \geqq \tau$, p_i is preserved,and

if $f(p_i) < \tau$, p_i is omitted,

where a tolerance value τ is given beforehand. By this procedure for all points of P, a subset $Q = \{q_0, q_1, q_2, \ldots, q_m\}$ of P is selected, and data reduction is carried out.

2.2 Our algorithm

We present here a new optimum algorithm of data reduction. The algorithm consists of next two steps.

Step1.

First, take an appropriate tolerance value τ_1 and then select data points according to the following scheme in the order of $p_0, p_1, p_2, \ldots, p_n$.

The first point p_0: preserved.

For a point p_i $(0 < i < n)$, take the point p_b which is the last preserved, and define d_i as

$$d_i = S_i / \ell_i.$$

Where S_i is twice of the area of triangle $\Delta p_b p_i p_{i+1}$,

ℓ_i is the length of line segment $p_b p_{i+1}$.

If $d_i > \tau_1$, p_i is preserved, and if $d_i \leqq \tau_1$, p_i is omitted.

The last point p_n: preserved (Fig.3).

Preserved points are renamed as $q_0(=p_0), q_1, q_2, \ldots, q_m(=p_n)$ in serial order. By the procedure, the subset $Q = \{q_0, q_1, q_2, \ldots, q_m\}$ of P is selected.

Step2.

Take an appropriate value τ_2 as another tolerance and consider omitted points $p_{r+1}, p_{r+2}, p_{r+3}, \ldots, p_{r+t-1}$ between $q_i(=p_r)$ and $q_{i+1}(=p_{r+t})$: $(i=0,1,2,\ldots,m-1)$. There are two cases.

I The case that there are three or more successive points lie in one side of the line $q_i q_{i+1}$,

II the other case.

For the case II no procedure is applied. In the case I the following procedure is applied for all points p_{r+s} $(0 < s < t)$ (Fig.4).

a) Define h_s as the distance between the segment $q_i q_{i+1}$ and the point p_{r+s}, and let M_s be the foot of the perpendicular from p_{r+s} to the line $q_i q_{i+1}$. That is, for the case that M_s is between q_i and q_{i+1}: $h_s = \overrightarrow{p_{r+s}M_s}$,

for the case that M_s is on the q_i side of q_iq_{i+1}:

$h_s = \overrightarrow{p_{r+s}q_i}$.

For the case that M_s is on the q_{i+1} side of q_iq_{i+1}:

$h_s = \overrightarrow{p_{r+s}q_{i+1}}$.

b) Let h_c be the longest of $h_s(s=1,2,...,t-1)$.

If $\tau_c > \tau_2$, the point p_{r+c} is picked up.Then reset $q_{i+1} = p_{r+c}, q_{i+2} = q_{i+1},..,q_{i+m+1} = q_{i+m}, m = m+1$,and return to the step a).

c) Reset $i = i+1$,and if $i < m$ then return to the step a)

By carrying out this procedure, the final subset Q is obtained.

In the step 1, we intend to remove the point p_i from which the length of the perpendicular to the line p_bp_{i+1} is smaller than τ_1. Then,points on a strongly irregular line are surely choosen.

By the stepl only,too many points are reduced in a case in Fig.5. The step 2 is added to improve this defect, and some points are retaken appropriately.

4. Determination of the tolerance and evaluation of the degree of approximation

In this chapter, we consider the degree of approximation between original and reduced image, and then we treat how to determine the tolerances τ_1 and τ_2 introduced in the step 1 and the step 2 of section 2.2.

For two images generally the feelings 'being largely different' or 'being far' come from an areal difference or a gap between them.

First we consider the areal difference between an original and its reduced image. For example, two areal differences are the same for case i and ii in Fig.6. But the change of shape by data reduction seems more remarkable in case i . So,it is adequate to evaluate that the degree of approximation in case i is smaller than in case ii . The step 1 is devised so that points in case i may be picked up, and points in case ii may be omitted. So the smaller τ_1 is specified ,the better approximation is obtained.

For a gap between original image and reduced one, as shown in case iii , Fig.6, τ_2 is considered to be an upper bound. The smaller τ_2 is specified, the narrower gap is obtained. As mentioned above,the degree of approximation can be evaluated with two tolerances τ_1 and τ_2.

We show here how to determine two parameters τ_1 and τ_2. Finding the optimum τ_1 and τ_2 is generally very difficult.Even for cases that equal number of points are reduced,there are a lot of combinations of τ_1 and τ_2. Obtaining an image of good approximation depends greatly on the choice of tolerances τ_1 and τ_2.

In general case,n,the number of original points and m, an expected number of retained points, are known in advance.Then,it is desirable to determine τ_1 and τ_2 from n and m only.

At first we consider the best proportion of the points to be reduced or to be supplied through the step 1 and the step 2, respectively.

After many experimental computations, we reached the conclusion that in the step 1 points should be made to 0.6m and the number of points supplied in the step 2 should be 0.4m .

1) how to determine τ_1

The empirical formula for τ_1 is obtained as

$\tau_1 = \ell(-1.423 + 0.856 (2.775 - \log (0.6m/n))^{1/2})$,

where ℓ is the mean interval of successive two points in the original data.This formula is obtained from the relation between the number of omitted points and τ_1 , using 1971 data points of digital map in Tohoku district, Japan . The usefulness of the formula is confirmed by other applications.

2) how to determine τ_2

For retaining m points exactly , the determination of τ_2 by an empirical formula as τ_1 is impossible, τ_2 is therefore determined with the method of trial and error. First we choose an appropriate τ_2 (for example $\tau_2 \sim 0.1\tau_1$). After applying the procedure of the step 2 , fewer points are supplied,then τ_2 has to be increased and vice versa The procedure is repeated until the intended number of points is retained.

4.Numerical examples

4.1 Examples of application

We show here some applications of our algorithm.

<u>Example 1.</u>

As an example of application to irregular lines, our algorithm is applied to Rias coast-line in Sanriku,Japan(Fig.7). It consists of 351 data points.We carried out two cases,the first case that m is a half of n and the other case that m is a quater of n. In these cases,

$\tau_1 = 0.0042$,and $\tau_2 = 0.006$ ....for $m = 0.5n$

$\tau_1 = 0.0063$,and $\tau_2 = 0.018$ ....for $m = 0.25n$

,here the mean interval ℓ is 0.0148.

Example 2.

As an example of application to planar line,our algorithm is applied to a coastline of Aomori prefecture(Fig.8).It consists of 350 data points.We carried out two cases ,one that m is half of n and the other case that m is a quater of n. In these cases,

$\tau_1 = 0.0056$,and $\tau_2 = 0.0045$ ....for $m = 0.5n$

$\tau_1 = 0.0084$,and $\tau_2 = 0.012$ ....for $m = 0.25n$

,here the mean interval ℓ is 0.0197.

Example 3.

As an example for low precision data, the shoreline of Australia is chosen. This example shows that our algorithm is independent of data precisons by normarizing τ_1 using the mean interval of successive two points. It consists of 1337 data which are 1/10 lower precision than previous examples. Fig.9 shows one case that data reduction is done to a quarter of the original. Determinded values for τ_1 and τ_2 are

$\tau_1 = 0.0679$

$\tau_2 = 0.0106$.

And the mean interval is 0.159.

4.2 Comparison with other methods

In this section, we compare our method with the following three methods, that is Sato's[2] ,Roberge's[3] and ODYSSEY's[6] methods.

The Sato's method is an algorithm for small amount of data. It needs 10^5 times of distance calculations in the cases that 117 original points are reduced into 25.Our method needs calculations only 117 times. So,Sato's method is not practical when the number of data points is large.

The second is Roberge's method. This is usually called ESA. ESA is a little better than ours for spiral curve. Fig.10 shows the comparison of two methods. The upper image is by ESA and the lower by ours. Here 1000 points spiral are reduced to 26 points. However, for an unit circle two methods are equal level. In the case of shore line consist of planar and complexed curve, our method is better than ESA. Fig.11 shows the result by ESA. Ours are already given in Fig.8 in the same condition , that is 350 data are reduced into a quarter of original ones , where the tolerance d = 0.0168.

The third method is included in the generalization command in ODYSSEY map system.This is prepared to reduce the number of points to describe boundaries in ODYSSEY system . This system is georgraphic information processing system produced by "Laboratory for Computer Graphics and Spatial Analysis".The result by the generalization command is shown in Fig.12. In this case, the tolerance 0.0208 is used and 350 data points are reduced to 92. This method is the worst of three.

5.Conclusion

By many numerical experimentations, we come to a conclusion that our algorithm is practical and effective for complex curves, not to mention for planar curves.

Acknowledgment

We are very grateful to Prof.M.Natori(University of Tukuba) for his advice and useful discussions throughout this study.

References

1.Agui,T.,Miyata,K. and Nakajima,M. , Quasi-coding Method for Digital Figures Based on Fractal Dimension. J.Inst.Televi.Eng. of Japan, **39**(1985),979-987.

2.Sato,Y.,Optional Piecewise Linear Approximation of Plane Curves ,Trans.IEICE,**65**-D(1982),1145-1150.

3.Roberge,J.,A Data Reduction Algorithm for Planar Curves,Comput. Gr.Image Process.,**29**(1985),168-195.

4.Tomek,I.,Two Algorithms for Piecewise-Linear Continuous Approximation of Function of One Variable,IEEE Trans. Comput., C-23(1974),445-448.

5.ODYSSEY User's manual,Harvard Labo. for Comp.Gr.Spatial and Analysis(1982),8-32.

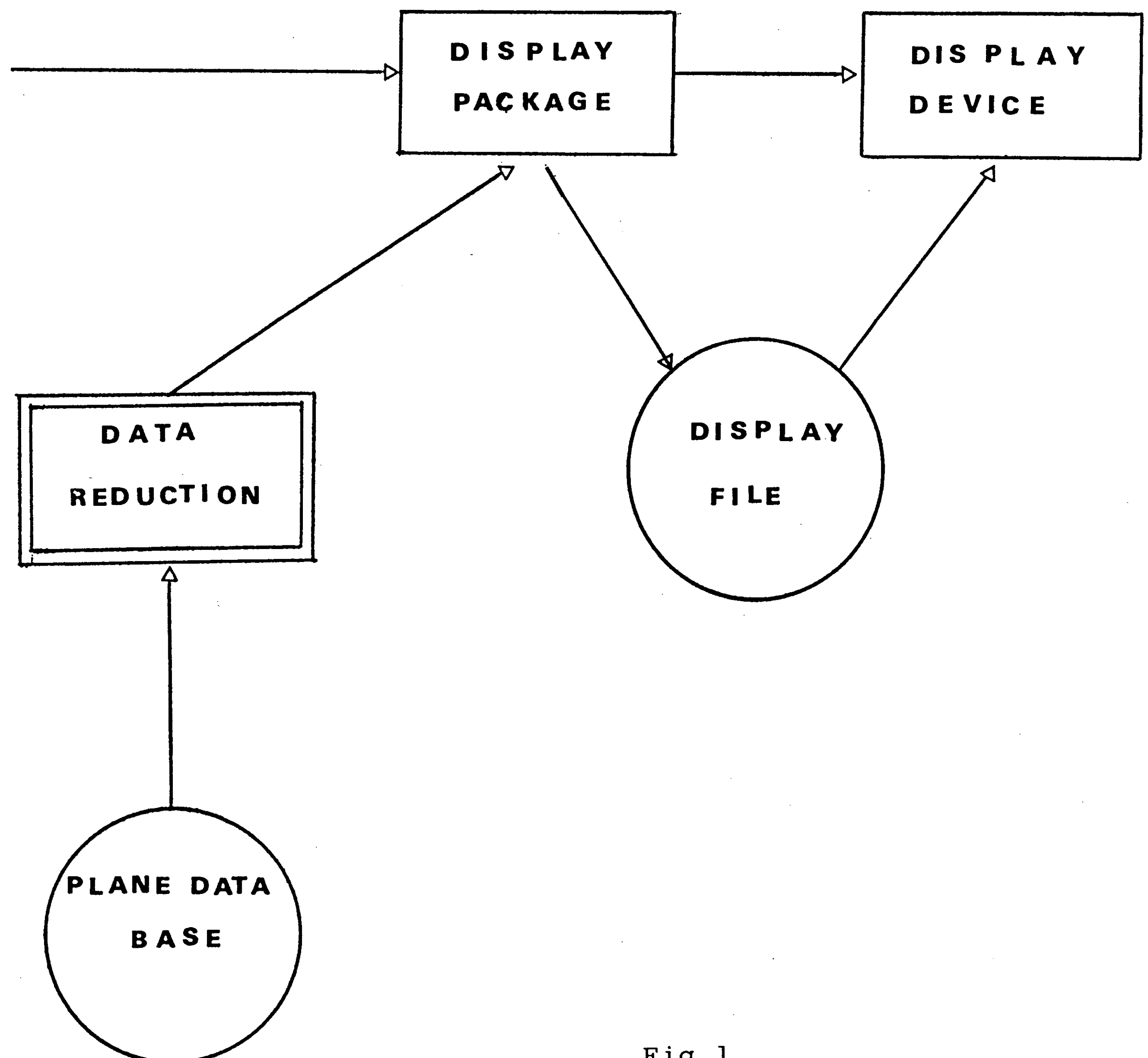


Fig.1.

Graphic system with data reduction of plane curves.

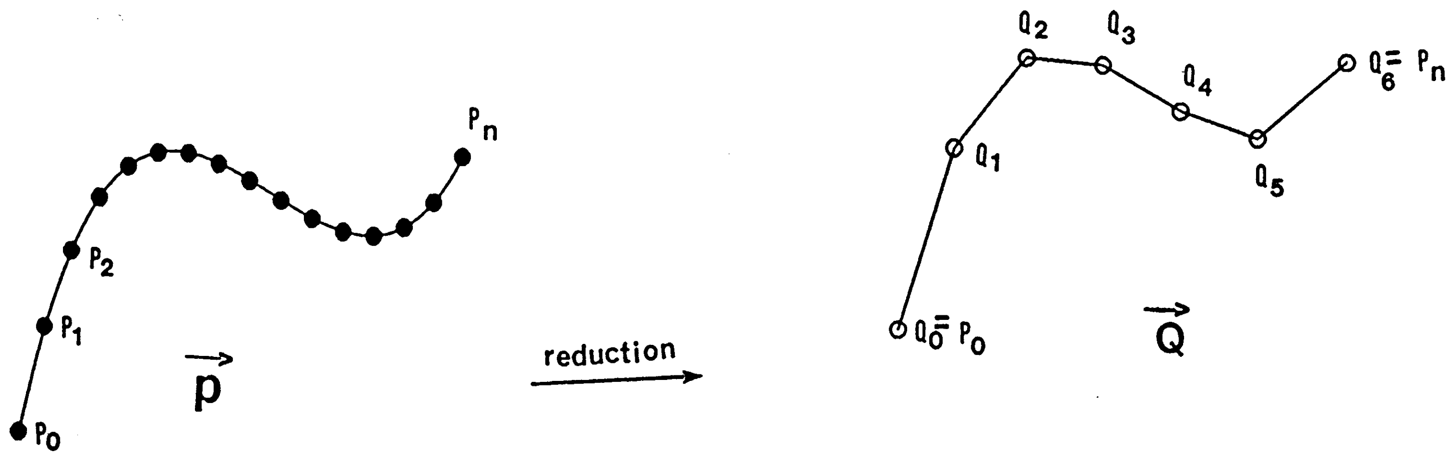


Fig.2.

Illustration of data reduction. $\vec{P}$ is the original curve and

$\vec{Q}$ is the reduced curve under some criterions.

The set $P \supseteq Q$.

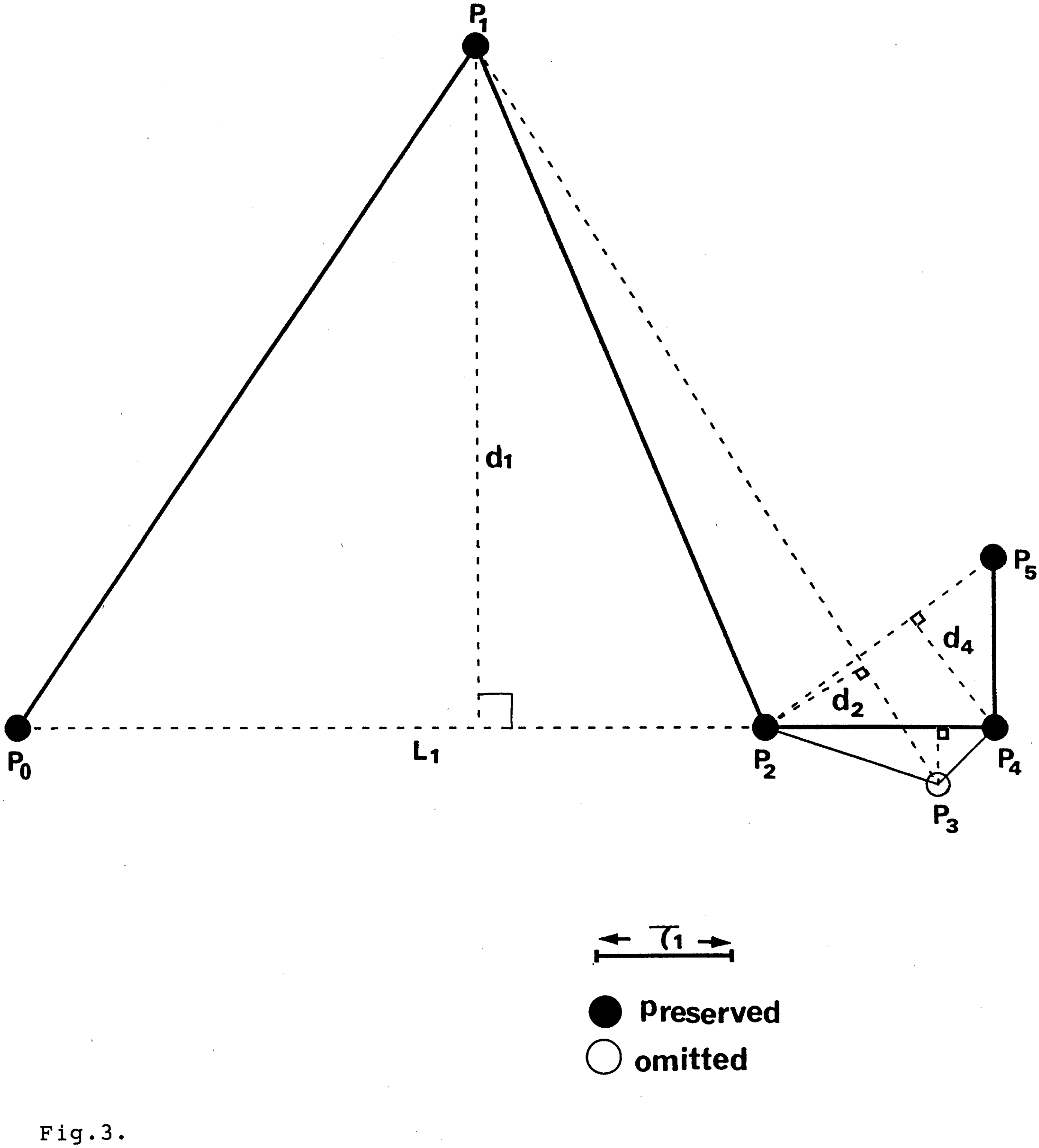


Fig.3.

Determination of q_i using the step 1 algorithm.

The reduced subset is $Q = \{p_0, p_1, p_2, p_4, p_5\}$.

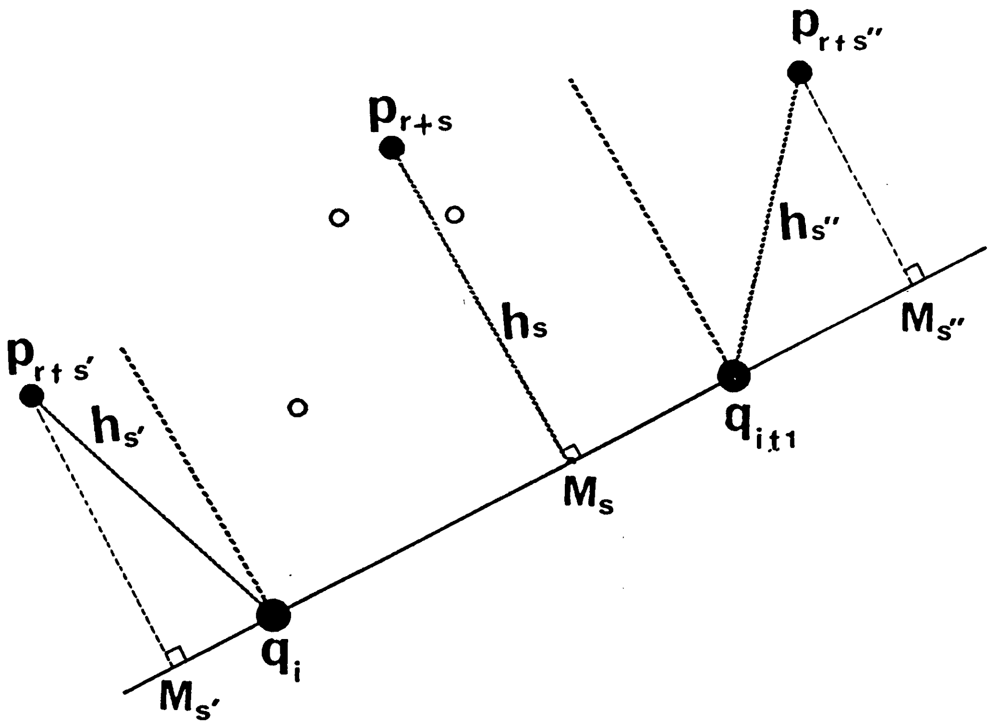


Fig.4.

Definition of the distance from p_{r+s} to the segment q_iq_{i+1}.

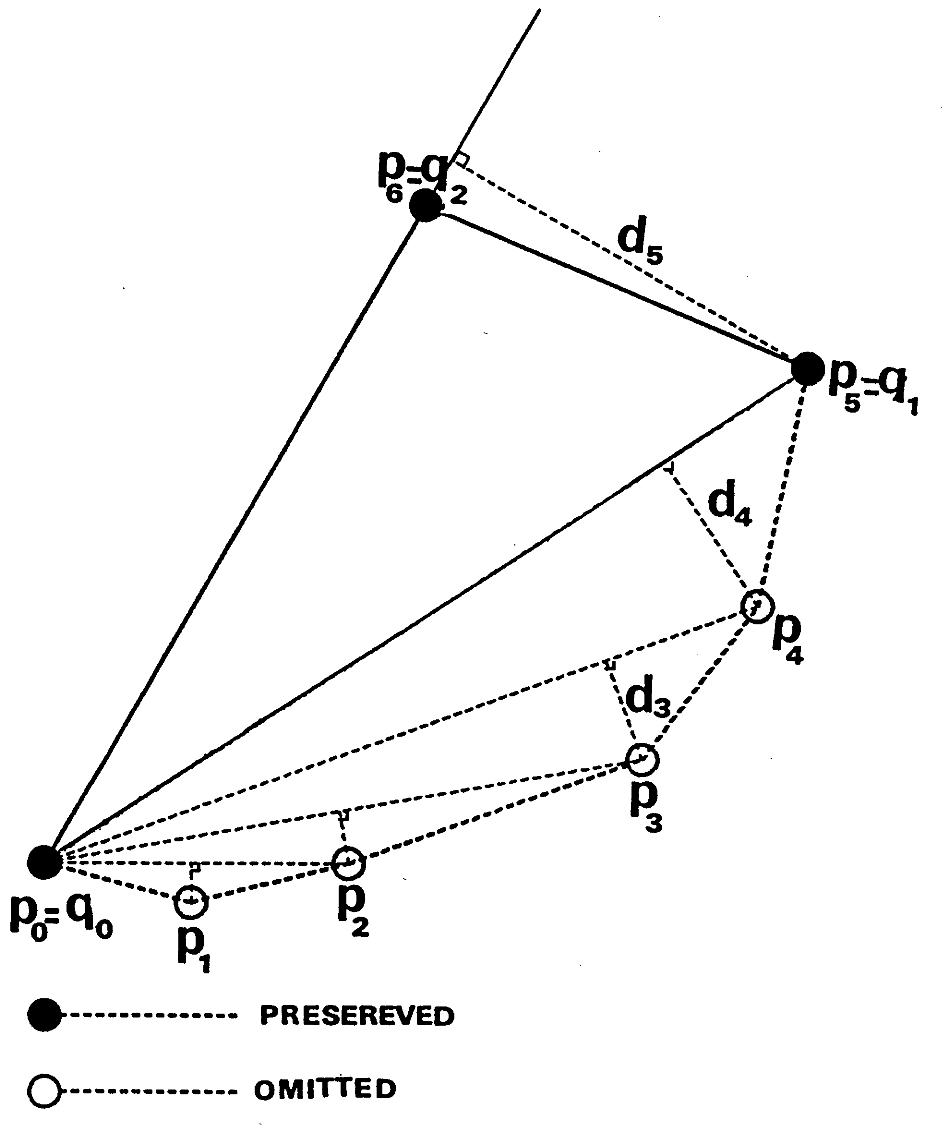
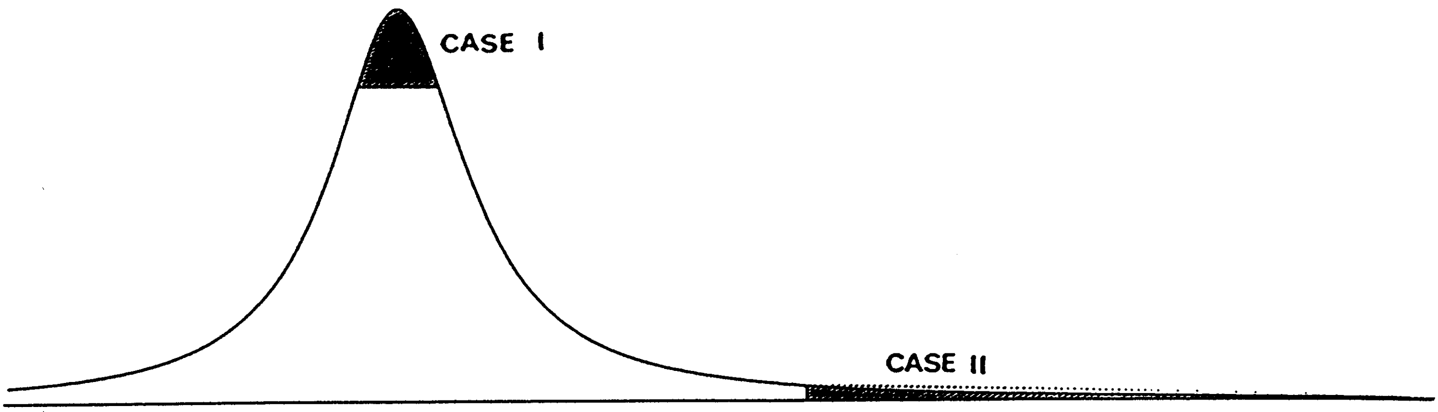


Fig.5.

Illustration of defects of the algorithm in step 1.

By the step 1 reduced curve $\vec{Q}$ = (p_0,p_5,p_6) is far from the original curve $\vec{P}$. By the step 2 , p_3 is selected again.



$$\int_{-0.1}^{0.1} \frac{dx}{25x^2+1} - 0.16 = \int_{1}^{2.848} \frac{dx}{25x^2+1} = 0.025459$$

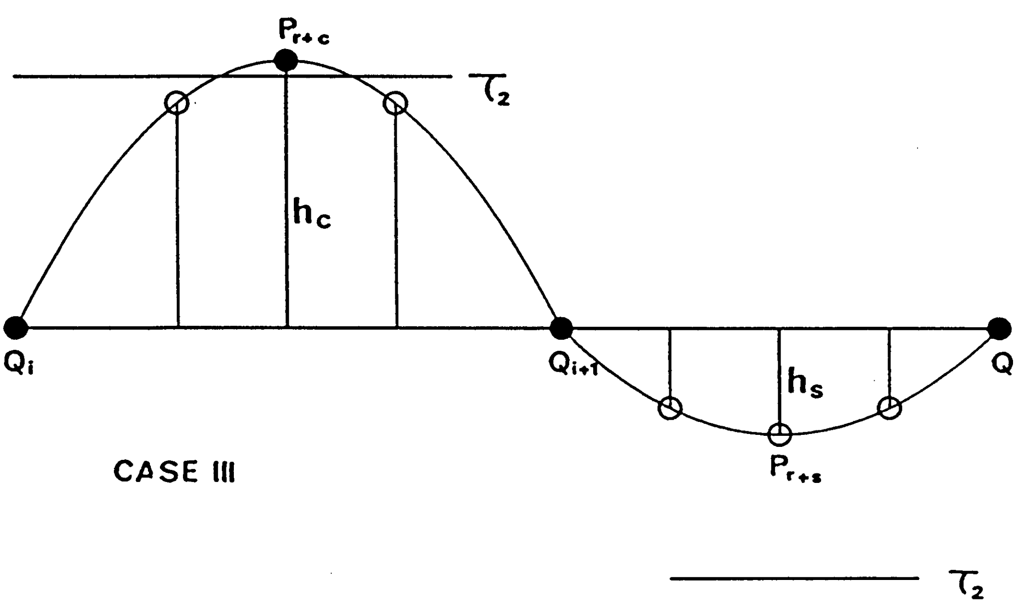
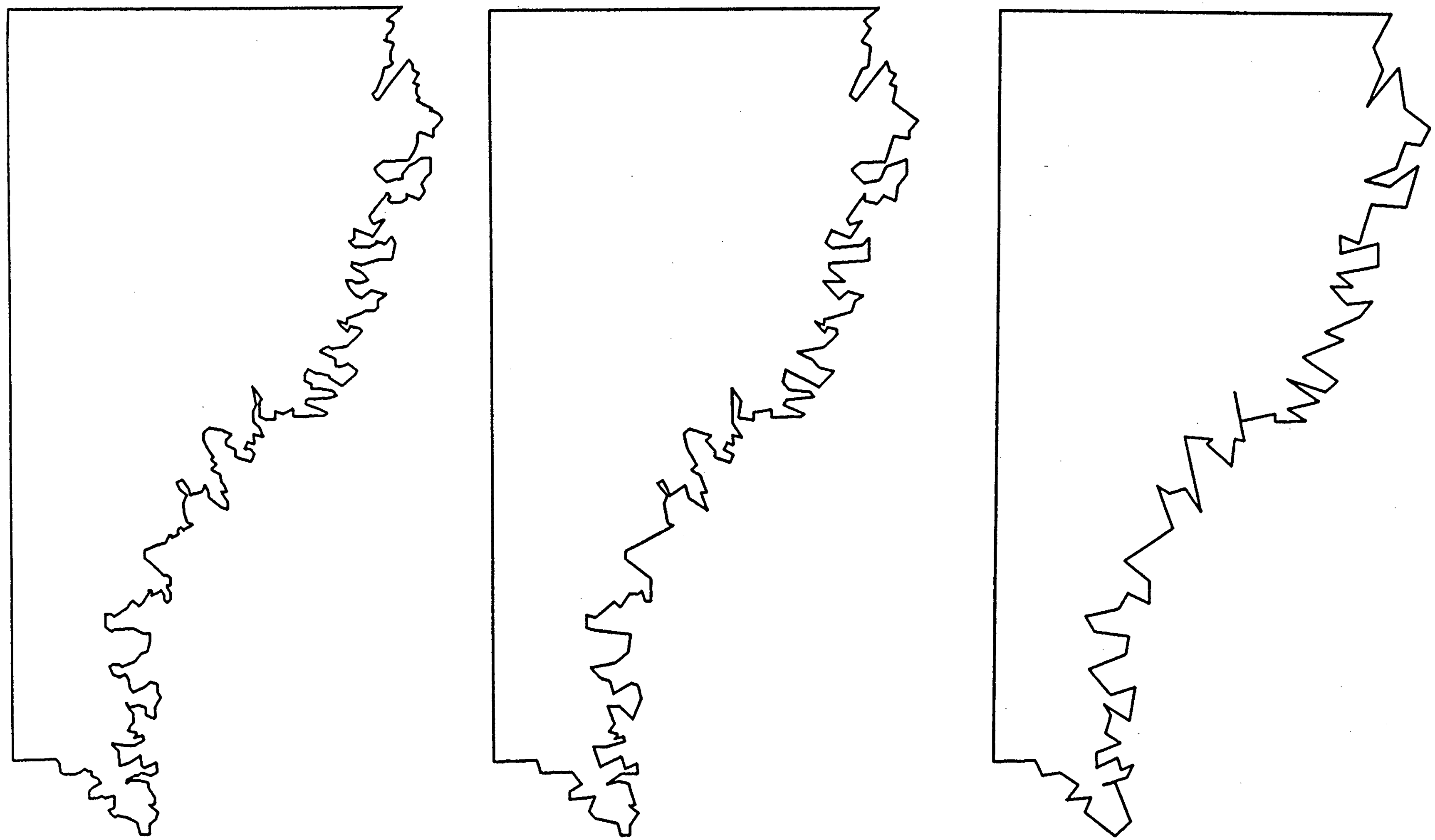


Fig.6.

Meaning of the tolerances τ_1 and τ_2.



ORIGINAL N= 351

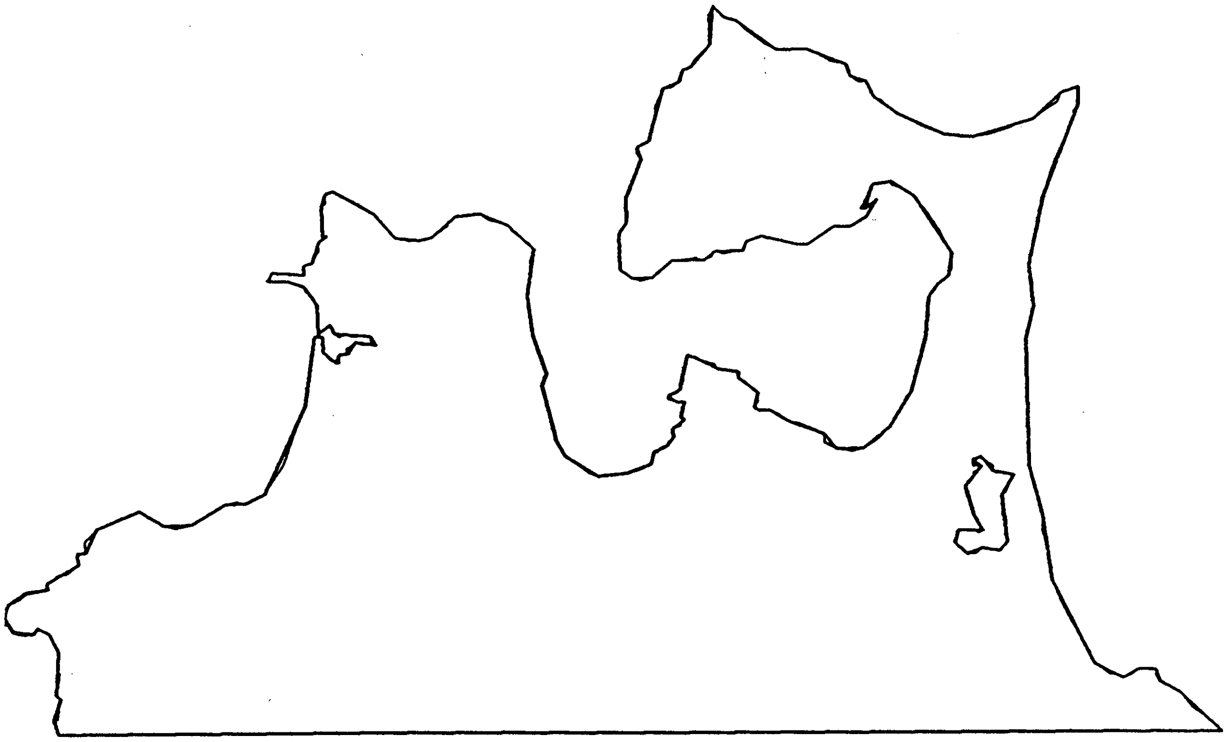
REDUCED TO 50 %

REDUCED TO 25 %

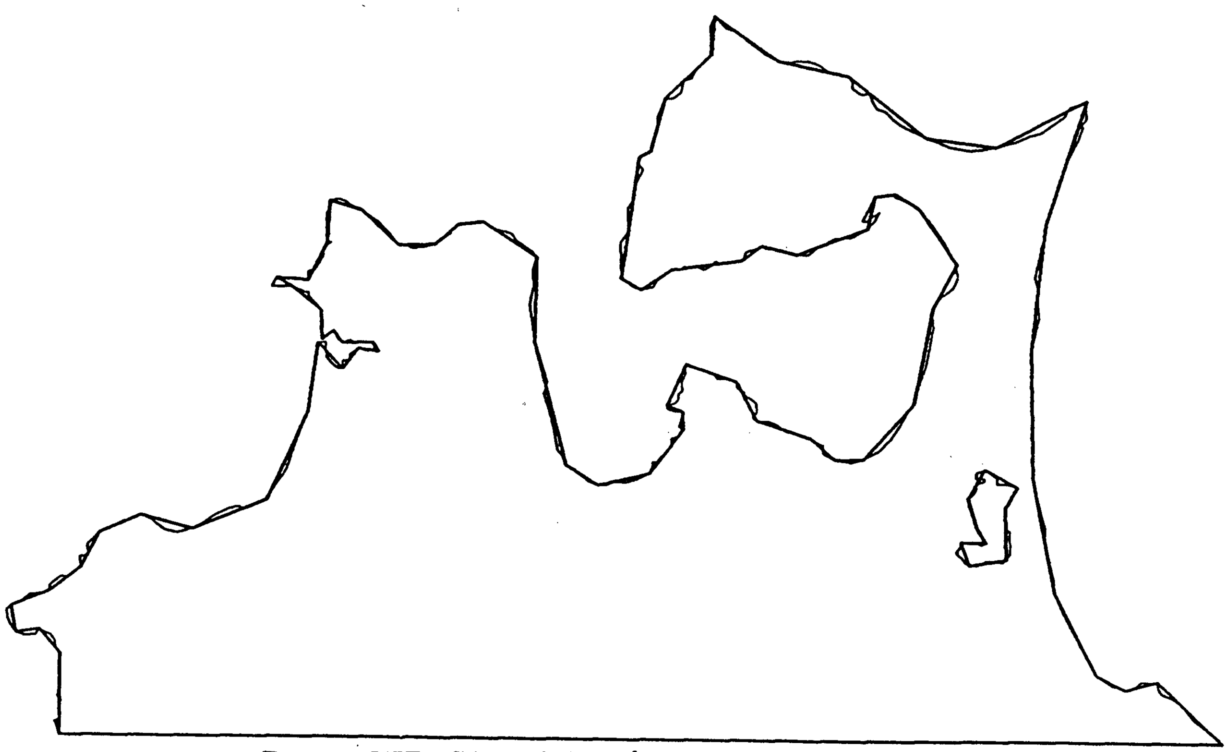
Fig.7.

The example of application to irregular Rias coast-lines.

The original data are consist of 351 points,the middle are 175 points,and the right are 87 points.



REDUCED TO 50 %   8 - 1



REDUCED TO 25 %   8 - 2

Fig.8.

The example of application to planar curves.

A fine line represents the original data and a heavy line represents the reduced data.

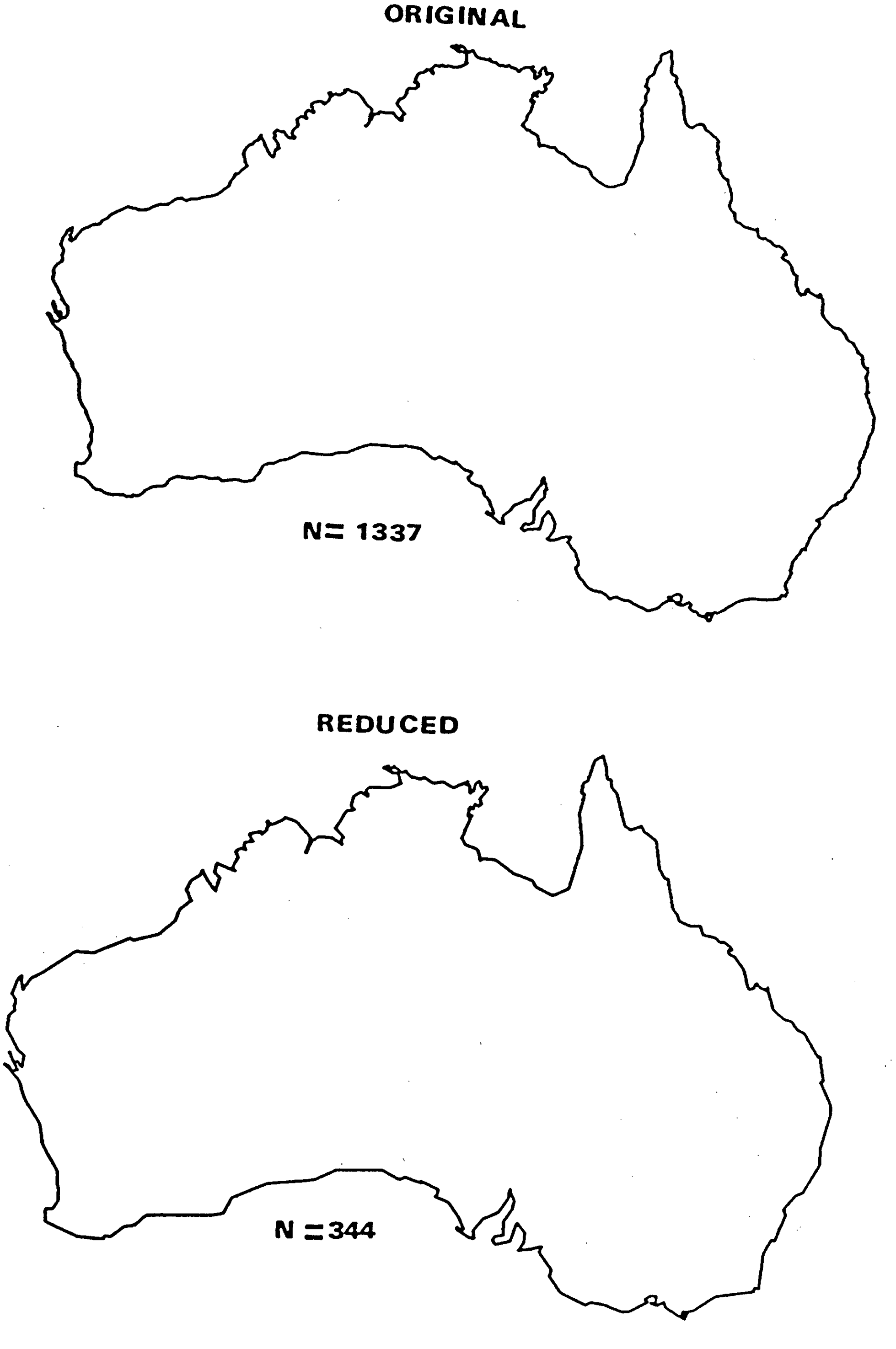
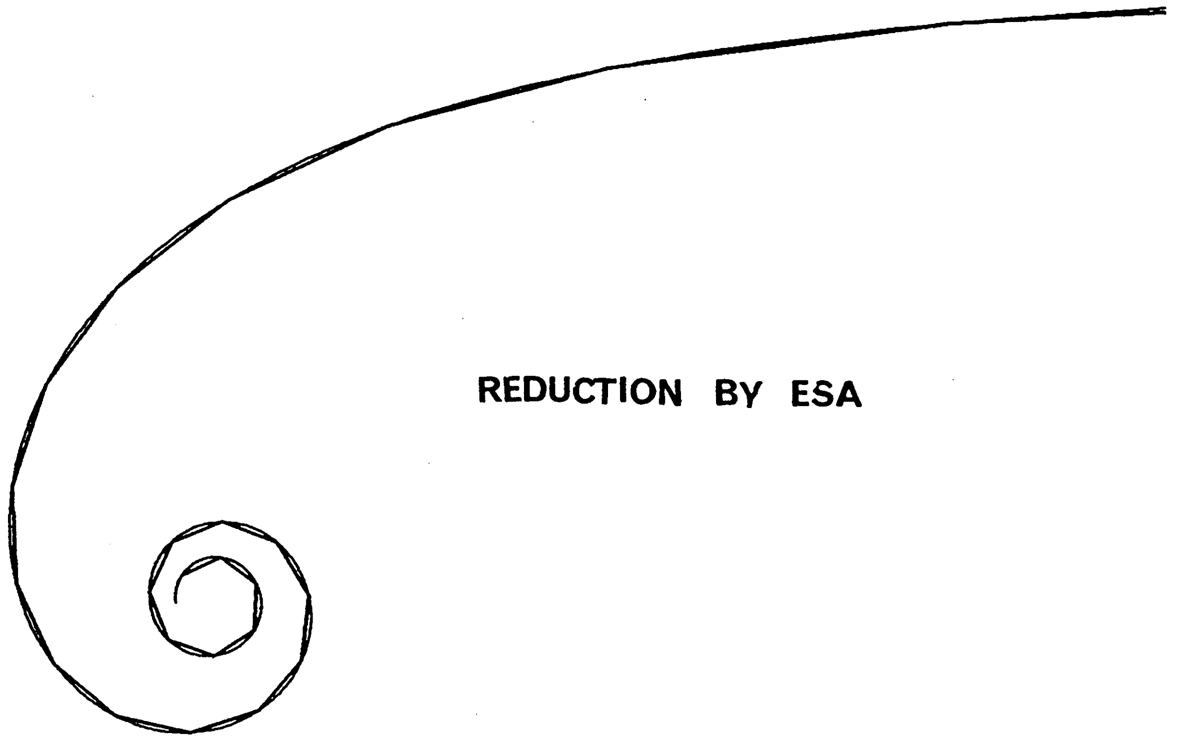


Fig.9.

The example for low precision data. This example means that our algorithm is independent of a precision of data.



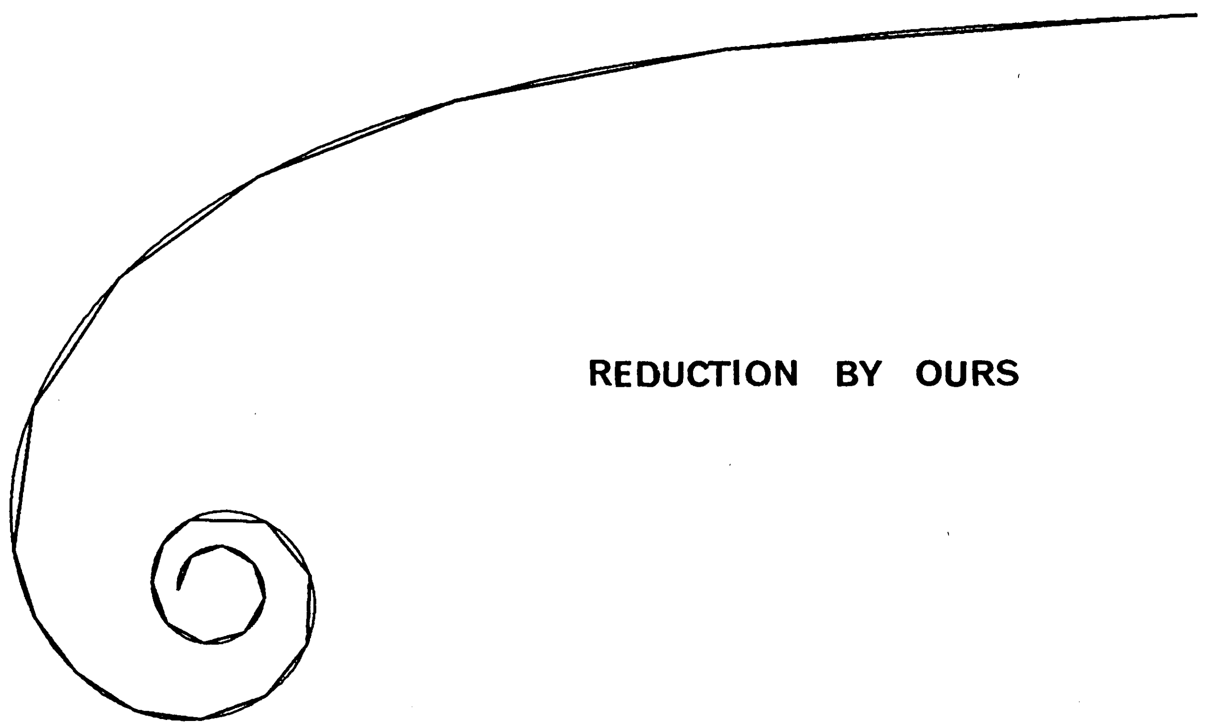


Fig.10.

Comparison between ESA algorithm and ours for spiral curve.

A fine line represents the original 1000 points and a heavy line represents the reduced to 26 points.

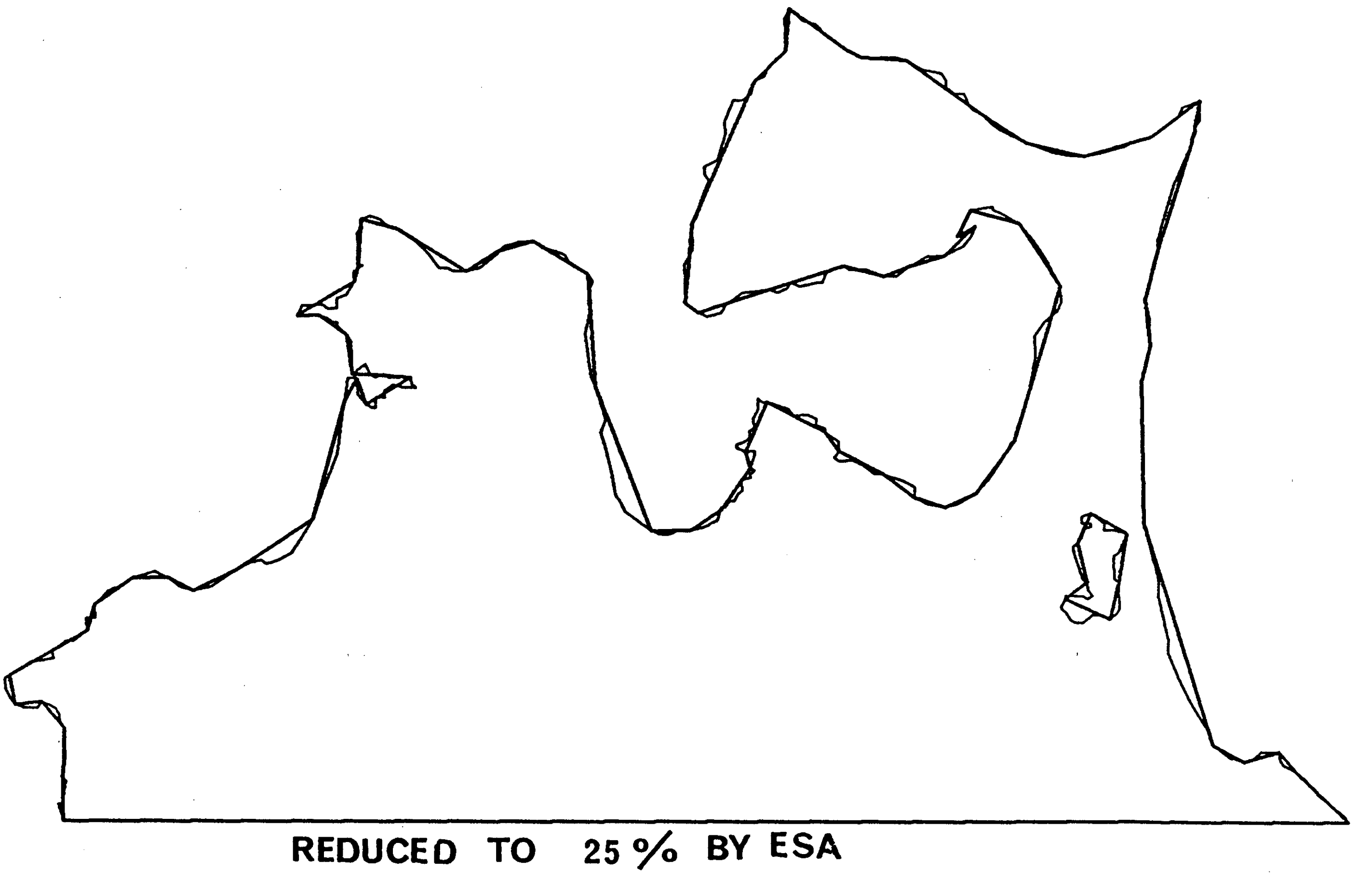
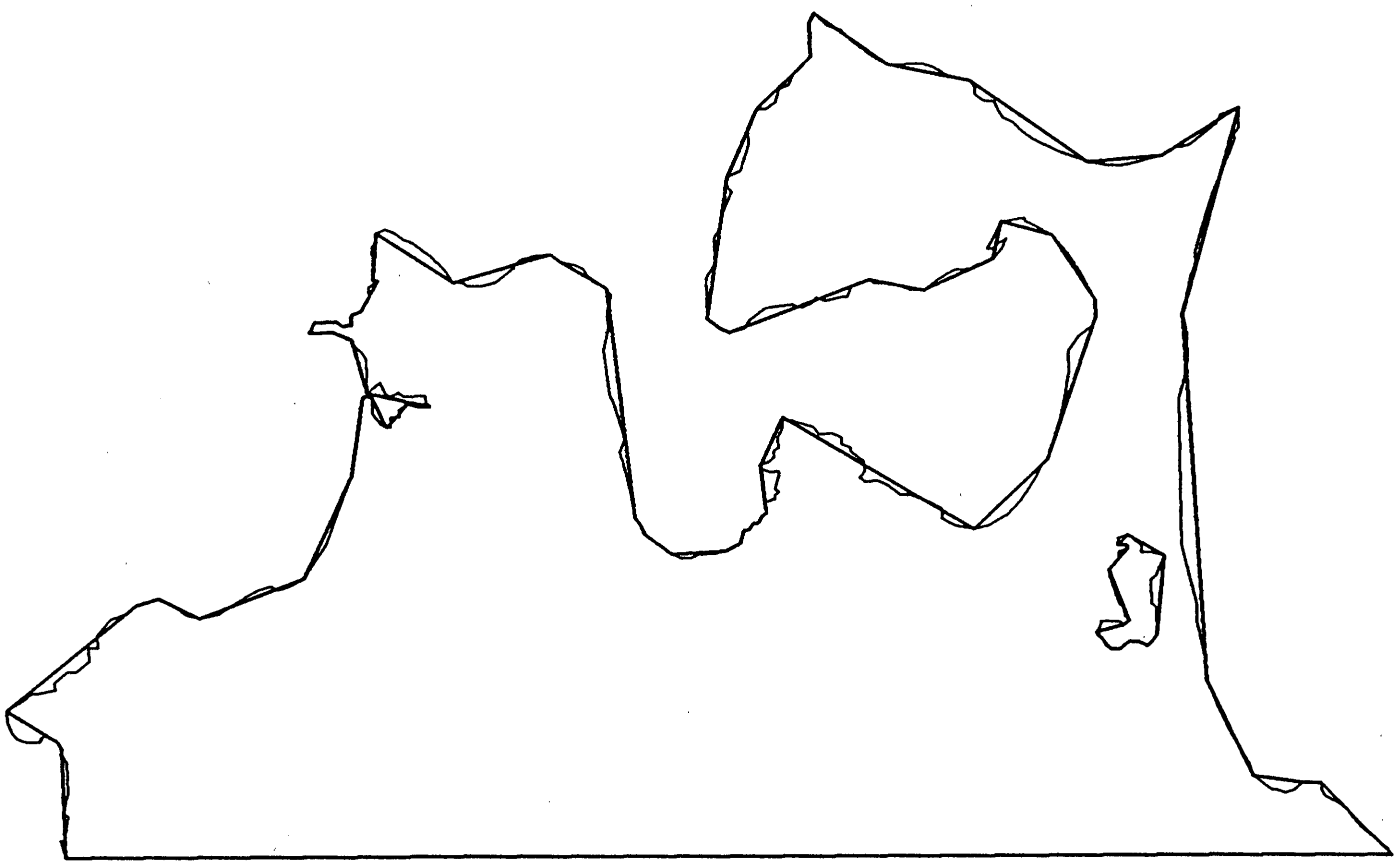


Fig.11.

Reduction of planar shore lines by ESA. Reduction 350 data to 88 points .



REDUCED TO 92 POINTS BY ODYSSEY

Fig.12.

Reduction 350 data to 92 points by the generalization command in ODYSSEY system.

INSTITUTE OF INFORMATION SCIENCES AND ELECTRONICS
UNIVERSITY OF TSUKUBA
TSUKUBA-SHI, IBARAKI 305 JAPAN

REPORT DOCUMENTATION PAGE	REPORT NUMBER ISE-TR-90-87

TITLE

An Optimum Data Reduction Algorithm for General Plane Curves

AUTHOR(S)

Sumiko Hiyama, Takao Hanada* and Hitoshi Imai**

Earthquake Research Institute, the University of Tokyo
Bunkyo-ku, Tokyo, 113 Japan

*Faculty of Electro-Communications, the University of Electro-Communications
Chofu-shi, Tokyo, 182 Japan

**Institute of Information Sciences and Electronics, University of Tsukuba
Tsukuba City, Ibaraki, 305 Japan

REPORT DATE	NUMBER OF PAGES
1990-11-30	23
MAIN CATEGORY	CR CATEGORIES
Numerical Analysis	G.1.0

KEY WORDS

algorithm, coastline, data reduction, optimum, plane curve

ABSTRACT

We introduce a data reduction algorithm for general plane curves without losing visual acceptability. Plane curves generally include not only planar curves but also complex zig-zag lines such as coastlines or frontier lines between countries.

Although several data reduction methods were already given to date, their application is limited to planar curves or small amount of data.

Our new algorithm is applied for a large quantity of data which constitute geometrical lines and coastlines. The results show that the algorithm is practical and effective on both planar curves and complex curves.

SUPPLEMENTARY NOTES