MEASUREMENT, CHARACTERIZATION AND CLASSIFICATION FOR AUTOMATED LINE FEATURE GENERALIZATION

Corinne Plazanet

IGN - Service de la recherche - Laboratoire COGIT
2, Avenue Pasteur (B.P. 68) 94160 St Mandé Cedex FRANCE

plazanet@cogit.ign.fr

ABSTRACT

This paper proposes to describe geometrical characteristics of linear features by segmenting lines and qualifying sections detected according to different criteria at several levels ("hierarchical" process). Such a kind of line representation should provide basic and complementary information in order to guide global and local decisions. The ultimate aim is to choose and sequence adequate generalization tools and parameter values.

KEYWORDS: Automated linear feature generalization, segmentation, characteristic point detection, measurement, clustering

INTRODUCTION

Generalization is often described as an holistic process, due to the large number of constraints and interactions that must be taken into account. This paper addresses issues related to linear feature intrinsic generalization. At first, the main issue for such a generalization is that every linear feature seems to be a particular case (at least for a significant level of generalization). This may come out from the fact that a lot of complex, and difficult to formalize, cartographic constraints intervene in the generalization process, and *shapes at different levels of analysis* must be considered and processed (omitted vs retained, simplified vs enhanced ...). Therefore, an identification and analysis of these shapes, prior to the application of generalization operations, as well as a study of the relationships between shape configurations and generalization operations and tools, are crucial for proceeding towards more effective generalization solutions.

LINEAR FEATURE GENERALIZATION AND CARTOGRAPHIC CONSTRAINTS

A cartographer, when generalizing manually, has a global and continuous feeling of each line [Plazanet et al, 94]. He/she reacts as a human being who fully uses his/her artistic and aesthetic judgement and his/her experience. Most of the time, he applies cartographic knowledge that has not been formalized in rules form anywhere. For instance in mountainous areas, many roads need to be enhanced. To eliminate a bend from a series of laces on a given road, when a spatial conflict with a river or other feature occurs, French cartographers generally proceed by maintaining the first and last bends, choosing *the* bend to eliminate (often the smallest) and emphasizing the others.

Looking at the cartographers' own way to proceed should hopefully provide basic knowledge and should help to express and formalize cartographic constraints and rules for correct automated generalization.

What kind of constraints do cartographers take into account when generalizing ? What rules guide manual generalization ? And further *Can we produce objective rules in order to obtain consistent generalization ?* [McMaster, 89] So far, some cartographic rules which guide intrinsic linear feature generalization have been identified such as:

Respect geometrical shapes. Within a road section containing a series of hairpin bends for instance, at least one hairpin should be retained, even at small scales. More generally, it seems important to take into account and preserve as well as possible the geometrical characteristics of bends such as size, amplitude, type of curvature, bend main direction, etc, and to maintain the main directions of linear features.

Preserve intrinsic topology.

Take into account symbol width. Linear feature symbolization is an important factor for decisions and strength generalization operations. The larger the sign, the faster geometrical details on shapes are lost due to line intrinsic conflicts and the legibility limits on the map.

Take into account semantical nature of the features. French cartographers distinguish 2 types of linear lay-out: straight ones observing *soft sinuosity* (rail-roads, freeways), and sinuous ones (roads, foot-paths, rivers). The semantical nature of the features contains a first indication on geometrical properties. Not only this indicates the type of geometry which may be expected from semantics, but generalization results have to respect these relationships.

Focus on the message. Each geographical object contains a particular message in accordance with its semantical nature, the map resolution and the cartographic theme. When producing a road map for instance, roads intrinsic value becomes greater. If a globally straight road has a compact bend, it should be kept and emphasized for user's information purposes such as for instance: "be careful, the road at this point turns abnormally".

Shape importance has to be considered in each case according to its linear environnment. Thus the ϕ shape (see Fig. 1) has to be maintained in the second case, while in the first case it looks like another one and then should be smoothed.

Scale and shape environment notions are essential, introducing some kind of subjectivity into local generalization decisions.

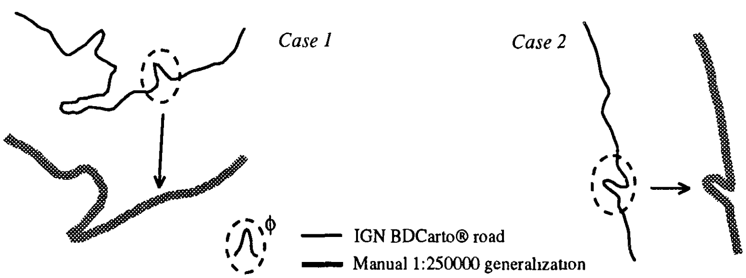


Fig. 1: The shape ϕ is the same, but in some case the generalization decision won't be the same

AUTOMATED GENERALIZATION

A lot of studies dealing with linear generalization have shown the lacks and limits of generalization algorithms, especially for simplification [Beard, 91] [Herbert et al, 92] [Jenks, 89] [McMaster, 89]. The quality of the result seems to vary according to the line characteristics [Buttenfield, 91]. These studies have clearly revealed the need to describe and qualify linear features. Checking on the significant geometric features of the line will hopefully guide the choice of tools and parameters.

More precisely the word "description" raises some questions such as:

- *What kind of information is important in order to generalize a line (Should we describe a line that we want to generalize as we perceive it visually ?) ?*
- *How to extract useful information from a set of points ?*
- *How to organize the extracted information ?*
- *How to take cartographic constraints into account ?*
- *How to take good decisions according to target scale, cartographic constraints, map theme and line description ?*

The automated intrinsic linear generalization process (first including simplification and enhancement) may be schematically divided into 4 tasks:

A - Linear features description

B - Analysis of the description

C - Decision of the generalization process according to analysis

D - Assessment of the generalization according to constraints and targeted scale

The present paper deals with tasks A and B proposing a way to build up a line feature description wich preceeds an approach of the generalization stage proper. The first part deals with linear feature description while the second addresses technical methods for description. The last section proposes first experimental results before we conclude and introduce next steps of the research.

LINEAR FEATURE DESCRIPTION

PRELIMINARY

When asked to describe a line, anybody will use the words *straight*, *sinuous*, *regular* (meaning homogeneous in some specific sense), etc. It is worth noting that such appreciations may be based on a more or less global (or local) inspection of the line. If the line is not regular, local bends are often described, for instance *"this straight road with 2-3 sharp U bends one mile away from the fork"*.

What are the criteria accounting for the global qualification of a line ? Is there nothing more than sinuosity vs. straightness ? According to Mc Master [McMaster, 93]: *"Indivuals seem to judge the shape of the line on two criteria: the directionality of the line and the basic sinuosity of the line"*. This observation induces him to use the sinuosity / straightness criterion in order to know where to cut the line.

Sinuosity may be qualified according to the semantical type of objects. For instance a river trace is quite often irregular, rough, while a road may follow "zigzags", hairpins, closed bends. Road features are human constructions, designed from regular mathematical curves. Modern roads are often composed of straight lines, circle arcs and clothoids [Affholder, 93].

One will instinctively cut the line of Fig. 2 into segments which look homogeneous. From the left runs first a hairpin asymmetric bend series followed by a more or less straight segment and a sinusoid series while the right part seems unhomogeneously softly sinuous. The sinuosity criterion itself is built up of several other criteria. In the meandering part on Fig. 2 the amplitude is much larger on the right. Bends in a homogeneous section can be described according to the kind of curvature (spirals, sinusoids...), and to amplitude and symmetry, etc. Local important details may also be noteworthy.

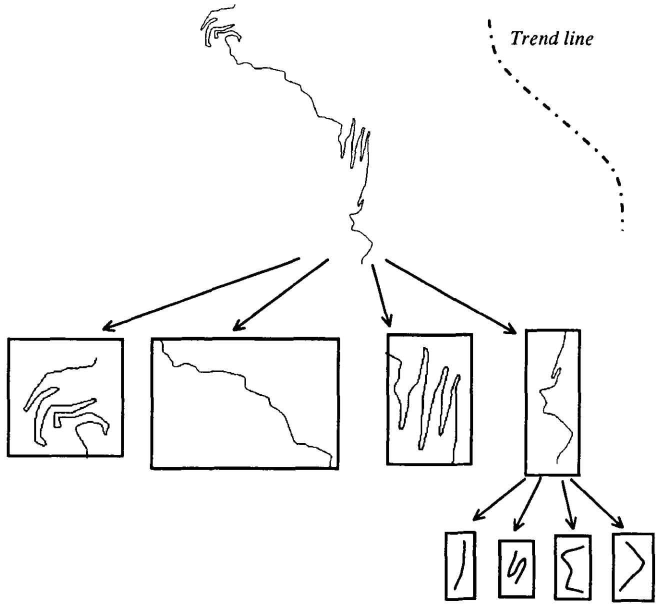


Fig. 2· An example of natural line segmentation (IGN BDCarto® road)

Our assumption is that, in fact, different levels of perception will consider different shapes of a line. Then the trend line can be seen as a first level of sinuosity (that might be called the global level). For instance on Fig. 2, it can be seen that the trend line has a convex shape. Clearly, the analysis of the trend line could provide us with additional indication of the line complexity.

SEGMENTATION AND ANALYSIS AT EACH LEVEL

At each level of perception, the line or a section of the line is segmented and analysed. If the line or section is unhomogeneous in some specific sense, it has to be segmented again.

A low level analysis is required to describe linear features at different levels. It consists in classifiying the considered line section and judging if it is homogenous or not, using some measures (see below in the "methods" section).

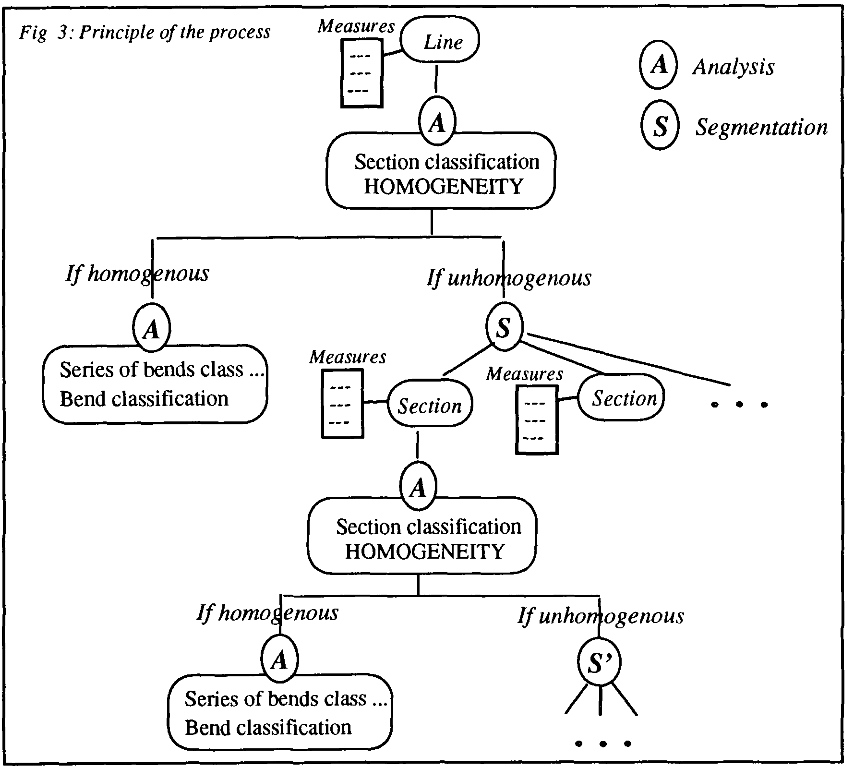
Further, besides a line description at several levels, a deepest analysis is required to extract complementary informations of prior importance for the generalization process itself such as:

- shape levels (relative levels of shapes within shapes),
- shape environment (relative positions of shapes within shapes),
- shapes or bends repetitions,
- intrinsic conflicts areas of the line.

ORGANIZATION OF THE PROCESS

Our approach is a hierarchical segmentation process analogous to *Ballard strip trees* and to the method proposed in [Buttenfield, 91] where series of measures are computed for each line. B. Buttenfield in [Buttenfield, 87] already proposed a very similar process for classification of cartographic lines also dealing with segmentation and measurement. Starting from the observation that *"a cartographic line is composed of a trend line, and features that bifurcate from it"*, she isolated several levels of details and elaborated measurements on the first level including the number of levels.

Principle of the process



First the line to be described is analysed using a first set of measures. If it is homogeneous, series of bends are qualified and classified if possible. If it is unhomogeneous, it has to be segmented. Then each section may in turn enter in the process recursively. (See Fig. 3)

Leaves of the tree

The first occurence of "hierarchical" description corresponds to a segmented section of the line with attributes and possibly a shape class code. The set of attributs is not always the same for every level of the tree; Most likely measures for segmentation won't be the same as for characterization.

At the finest analysis level, it is important to know two things about each of the final homogeneous sections: What are their geometrical characteristics ? And how do they fit into the global shape of the line ? The local analysis (between two successive inflection points) of curvature, amplitude and symmetry will decide the characterization of each section.

The lowest node may be a single bend defined as a constant sign curvature section delimited by two inflection points. It can be qualified by its width (large to small), amplitude (tight to open), bend direction (straigth to convex), symmetry and type of curvature (sinusoid, rectangle, spiral).

Potential outcomes

Together segmentation, analysis and classification will hopefully be useful for taking local decisions in order to generalize according to cartographic rules. For instance, for a particular bend or series of bends delimited by 2 inflection points, if the distance between these points is smaller than symbolization width, then a conflict area is detected. Or if a section of hairpin bends inside a straight section from the higher level node is considered as an accident then it should be preserved. If the two rules are verified then the compact series of bends has to be amplified. Then, when the generalization operation is chosen, simplification and enhancement algorithms corresponding to the lowest level class section are applied. Other potential outcome may be the quality assessment in terms of shape maintenance, still badly missing so far.

METHODS FOR SEGMENTATION AND ANALYSIS

CHARACTERISTIC POINTS DETECTED FOR SEGMENTATION AND ANALYSIS

The detection of the characteristic points is a fundamental operation for the process. Characteristic points are useful both for segmentation and analysis operations:

- Hoffman and Richards [Hoffman et al, 82] suggest that curves should be cut at the minima of curvature, which are particular characteristic points.
- The line geometry description is based on the characteristic points detection: Attneave in [Attneave, 54] has proved that "*information is further concentrated at points where a contour changes direction most rapidly*" i.e. vertices. As a rule, in literature, be it psychology or computer vision or more recently cartography, shape detection is based on the detection of characteristic points [Hoffman et al, 82] [Thapa, 89] [Muller et al, 93].

In Euclidean geometry, characteristic points are often confined to curvature extrema: Vertices and inflection points. As Freeman said, this definition is too restrictive: "*We shall expand the concept of critical points to include also discontinuities in curvature, end points, intersections (junctions) and points of tangency*" [Freeman, 77]. According to our needs, we define the following points as characteristic points:

- discontinuities in curvature,
- start or end points,
- maxima of curvature i.e. vertices,
- minima of curvature i.e. inflection points,
- critical points (chosen among of inflection points), which demarcate 2 line sections after a segmentation stage.

Detection of inflection points

The first step consists in detecting the inflection points. Because of the acquisition process, lines frequently contain spurious micro-inflections which are not characteristic. Moreover, according to the analysis level, only the main inflection points are of interest. A process has been implemented and tested at the IGN COGIT laboratory to detect characteristic inflection points [Affholder, 93]:

A smoothing using the convolution product of points with a Gaussian filter will delete details which are not significant for a given analysis level. According to McMaster, *smoothing routines relocate or shift coordinate pairs in an attempt to "plane" away small perturbations and capture only the more significant* [McMaster, 89]. Studying the vectorial product variation in each point of the smoothed line allows for the detection of characteristic inflection points: At these points, there is a significant change in the sign of the vectorial product. The smoothing may be more or less heavily applied, depending on the analysis level. The more global the analysis, the strongest the smoothing.

Detection of vertices

We approximate the actual vertex by computing the point which stands at the greatest perpendicular distance from the segment joiging the two consecutive inflection points (equivalent to the Douglas routine computation). We may go further to compute secondary pseudo vertices of a particular bend (running a more specific Douglas routine).

MEASUREMENTS FOR SEGMENTATION AND ANALYSIS

For a bend (Fig. 4), we may compute:

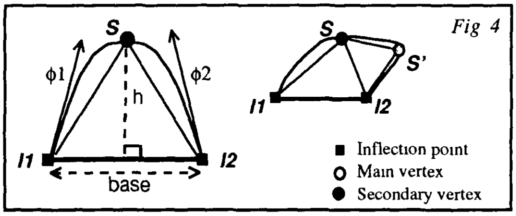


Fig 4

- the height *h* or surface of the triangle (S,I1,I2)
- the euclidean distance between the inflection points, *base*
- the curve length between inflection points I1 and I2, *l*
- the ratio between the curve length from I1 to I2 and the Euclidean distance, *l/base*
- the angle between I1 and I2
- the number of secondary vertices per bend
- the area of the main triangle (S,I1,I2) and secondary triangles (S,S',I2)

The number of secondary vertices or the sum of the area of triangles for instance gives an idea of the type of curvature fo the bend, while the variance of the curve length *l* between the inflection points indicates the regularity of line.

For a section, we may compute the mean (or median) value, variance, minimum and maximum values of each of these measures. Some more measures based on the line joining inflection points seem to be interesting:

- the total number of bends
- the total absolute angles of the inflection points line
- the ratio of the curve length of the line joining inflection points and the curve length of the original line

For instance the variance of distances between inflection points gives an appreciation of some kind of regularity of the line. The inflection points line in a sense may be seen as an approximation of the trend line.

Many of such measures may be easily computed. Also we need to choose and normalize a subset of significant and non-correlated measures that will account for each particular operation of segmentation and analysis of any line.

MEASUREMENTS FOR A FIRST LEVEL SEGMENTATION

With a heavy smoothing, it can be hoped that the strongest and most meaningful inflection points will be detected, and from them the critical points for this first analysis level.

The obvious fact that a very sinuous line has many inflection points in close succession (which is not the case for a straight line) will help us to divide lines into sinuous / straight segments. Looking at the curve representing the Euclidean distances between successive inflection points along the line, one can deduce sections on the line where these distances varie gently.

A FIRST EXPERIMENT OF LINE FEATURE DESCRIPTION

The description process has been attempted on a set of lines taken from the BDCarto® IGN data base, based on the measurements described above and using a classical cluster analysis software. The first objective is to split lines into straight / sinuous / strongly sinuous sections. The different parts of the experimentation are:

1 - INFLECTION POINTS AND VERTICES DETECTION

Hereafter is presented an example of inflection points detection on a 5 m resolution road, Fig. 5. The smoothing is rather strong as the neighbourhood involves a fifth of the total number of points ($\sigma = 40$).

2 - PARTITIONING INTO STRAIGHT / SINUOUS SECTIONS

Once the inflection points are detected as shown on Fig. 5, the line is cut into straight / sinuous parts and the critical points are selected.

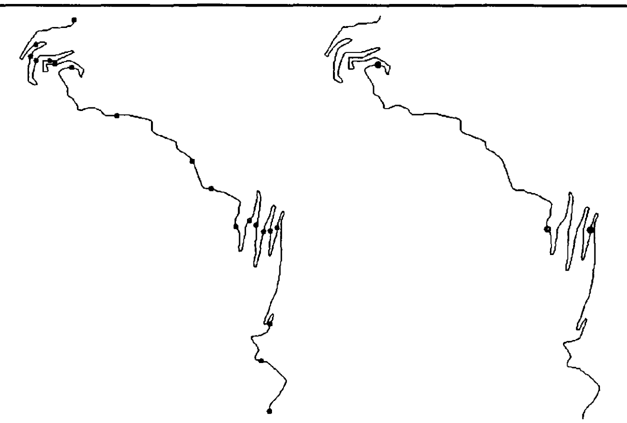


Fig 5 : Inflection points detected on the left side
Critical points retained on the right

These three critical points determine the four sections observed. Through closer analysis, the line will be cut recursively and even the sinuous sections, according to their characteristics: In the first sinuous section (top left), the bends are as sharp and close to each other as in the second sinuous section, but they are not symmetric.

3 - <u>CLUSTERING HOMOGENEOUS SECTIONS</u>

For this first experiment, chosen measures from the described set of measures are: the median ratio between the curve length *l* from I1 to I2 and the Euclidean distance *base* and median ratio between the *h* distance and the Euclidean distance *base*. A set of 40 lines have been classified using a S-PLUS cluster analysis package (See Fig. 6).

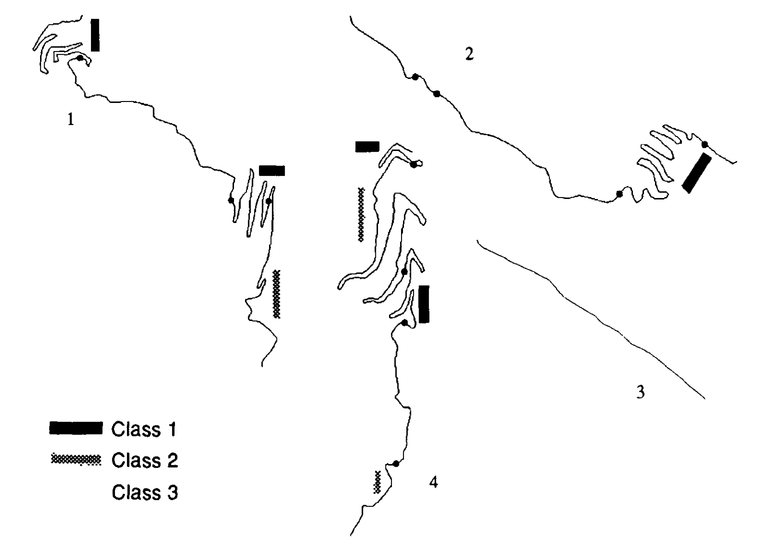


Fig. 6: An example of classification results

DISCUSSION

The results of the segmentation operation seem quite promising. In order to go further, the segmentation process has to be improved by adding complementary information especially about bend amplitudes. As to characterization and classification, the current results show the need for a better understanding of the measures and their interrelations. An interesting approach would consist in devising a large set of measures, and then trying to reduce this set to smaller sets of significant and non-correlated measures for the different purposes of our description tasks. The question which arises then is wether we will be able or not to choose among th initial measures and normalize the initial ones.

The aim of this study is to provide a method for linear shapes identification before generalization in order to choose adequate and effective generalization solutions. Other future work will focus first on the study of the effects of generalizing operations on line shapes, and on the selection of the best representation (cubic curve arcs for instance) for a given homogeneous section. Such kind of work could probably become useful in order to assess the quality of generalization in terms of shape maintenance.

It would be hopeful that a classification of measures, as well as a common terminology, be established and agreed by the research community. This would allow for a greater exchange of research results and could accelerate joint research in this crucial area of generalization.

ACKOWLEDGEMENTS

The author wishes to thank Robert B. McMaster and Jean-Philippe Lagrange for the numerous enriching discussions and their valuable suggestions.

REFERENCES

[Affholder, 93] J. G. Affholder. Road modelling for generalization. NCGIA Initiative 8. Spec. Meet. Buffalo 1993

[Attneave, 54] F. Attneave. Some informational aspects of visual perception. Psychological Review. Vol. 61, No. 3, 1954

[Buttenfield, 87] B. Buttenfield. Automating the Identification of Cartographic Lines. The American Cartographer Vol. 14, N.1, pp. 7-20, 1987

[Buttenfield, 91] B. Buttenfield. A rule for describing line feature geometry.In Map Generalization (B. Buttenfield & R. McMaster Eds) Part 3. P. 150-171 Ed. Longman Scientific & Technical. London 1991

[Herbert et al, 92] Herbert G., Joao E. M. et Rhind 1992 Use of an artificial intelligence approach to increase user control of automatic line generalization. EGIS 1992 p 554-563

[Hoffman et al, 82] D.D. Hoffman, W.A. Richards. Representing smooth plane curves for visual recognition AAAI Proc. p. 5-8. 1982

[McMaster, 89] R. B. Mc Master. The integration of simplification and smoothing algorithms in line generalization. Cartographica 26 p.101-121 1989

[McMaster, 93] R. B. Mc Master. Knowledge Acquisition for Cartographic Generalization: Experimental Methods. ESF GISDATA Workshop Compiègne France 1993. To appear in "GIS and Generalization: Methodological and Practical issues" Taylor & Francis, London. ESF GISDATA series

[Muller et al, 93] J. C. Muller, Z. Wang. Complex coast-line generalization. Cartographic and Geographic Information Systems. Vol 28-2,p. 96-106. 1993

[Plazanet et al, 94] C. Plazanet, J.P. Lagrange, A. Ruas, J.G. Affholder 1994 Représentation et analyse de formes pour l'automatisation de la généralisation cartographique. EGIS'94 Proc. Vol 2. p 1112-1121

[Ruas et al, 93] A. Ruas. JP Lagrange, L. Benders 1993 Survey on generalization. IGN Internal Report

[Rosin, 93] Rosin P.L. 1993 Non-parametric multi-scale curve smoothing SPIE Conference, XI: Machine Vision and Robotics, p. 66-77

[Thapa, 89] K. Thapa. 1989 Data compression and critical points detection using normalized symmetric scattered matrix. AUTOCARTO 9, P. 78-89