Project Number	**ESPRIT / LTR /24 939**
Project Title	**AGENT**
Deliverable Type	**Report**

Deliverable Number	**D C1**
Internal ID	**AG-99-02**
Contractual Date of Delivery	**01/03/99 (Technical Annex)**
Actual Date of Delivery<	**23.02.99**
Version	**2.0**
Workpackage/Task contributing	**C1**
Authors	**UNI-ZH**
Confidentiality	**Public**

Title of Deliverable	**Selection of Basic Measures**

Abstract

This report provides a state of the art for measures for cartographic generalisation. It offers a classification and points out measures commonly in use, as well as new approaches. After an assessment, a justification of the selection of measures for the project prototype is presented.

Keyword List

Generalisation measures, feature characterisation

Executive summary

This report provides a state of the art on measures used during generalisation processing. In order to define a set of measure used by the micro-agents, we:

- Evaluate existing algorithms for characterisations
- Define missing algorithms and propose solutions
- Provide a set of measure.

This report engages the topic by defining what attributes a good measure should possess. It then proposes a suitable classification of internal measures (i.e. measures for micro-agents) as a clear distinction, between internal measures and other kinds, which enhance the complexity of internal measure. It also provides a list of major measures in each class.

However, this state of the art highlights most existing generic measure, are too indiscriminate to be used for characterising efficiently specific properties. The lack of discrimination led us to two conclusions. The measures should be:

1. Feature-oriented to characterise as closely as possible a property to be analysed for a generalisation process;
2. Applied, as far as the shape extent allows, on more local parts of each micro-agent because large variations within the object are frequent. Measures then return an incorrect description whether they are not correctly split in more homogeneous parts.

Furthermore, justifications of a set of measures for internal measures on micro-agents cannot be defended without investigating the generalisation algorithm purpose. These justifications are then made in the D-A3 report, based on recommendations of both the D-C1 (see below) and D-D2 reports (internal algorithms). The D-C2 and D-D1 reports then provide specifications of the retained measures and algorithms.

Table of Contents

1 Introduction

1.1 Purpose

The AGENT Technical Annex (IGN, 1996) describes the objectives of Task C1 as "to define a set of measures to qualify the geometry of an agent and to assess agent status according to constraints (local and global goals)".

While the identification of measures for characterising and assessing spatial attributes of an agent is described in the "DC2-Report on measure specification", this report is a state of the art in the field of vector-based spatial attribute analysis. As such the aims of this report are:

- to build a common basis for the choice of measures to be used in the prototype
- to define a common classification
- to present a list of the most useful measuring and characterisation algorithms available to help partners stay current with generalisation research

1.2 Preliminary Remarks

According to the Technical Annex, the aim of the C task is "to define and link measures to a specific task (quality control, algorithm choice,…). The objectives for C1 are "in defining sets of measures to qualify the geometry of an agent and to assess agent status according to constraints (local and global goals)".

This text focuses on measures used to assess the status of an agent regarding its cartographic and geographical natures only. Measures to assess pure agent-relevant status (e.g. is an agent active, is it waiting for an order, does it has plans to resolve…) are not treated here.

This document proposes a classification for measures that reflects the work decomposition used in our AGENT-project. Using this classification, a state of the art on measures is presented. While this document deals mainly with workpackage C1 some measures concerning C2 are also considered in order to illustrate the classification and allow continuity when approaching C2.

The reminder of this chapter is used to justify the choice of measures and templates and to define the most significant terms used throughout this text. These terms are discussed in greater depth later. Chapter 2 deals with some general remarks about measures, and ends with a short specification of 'a good measure'. Chapter 3 presents the proposed classification used in this document. In Chapters 4 through 7 we focus on the main categories used in this classification. The measures and methods referenced throughout Chapter 4 through 7 are described at the end of this report.

1.3 On the Choice of Measures

This report does of course not claim to be complete. We tried to respect the most common measures presented in the literature of digital cartography. We are convinced that many other interesting measures might be found in the literature not central to our topic (even if they are mostly raster oriented) e.g. Pattern Analysis and Image Processing, Pattern Analysis and Machine Intelligence, Computer Graphics and Image Recognition, and Computer Vision and Image Recognition. Though we reviewed several journals, we did not have the time resources to study and integrate all this knowledge.

Additional measures can be found in the field of statistics. We expect even simple measures, such as averaging or tests for distributions, valuable; though such methods are not described here. Another interesting source of knowledge is software in the cartographic and statistical domain. A powerful tool providing dozens of measures for analysing area patches and polygonal subdivisions is FRAGSTATS (Forest Science Department, Oregon State University).

These fields should be more explored whether missing and non-predicted characterisations will appear in the further developments but doesn't need any more attention for internal measures.

General Remark: *The methods described in the Appendix with a template are marked throughout this paper with an asterisk.*

1.4 Definitions

Measure: A measure is a procedure for computing measurements, which are the basis for evaluating characteristics of a geographical entity and assessing the need for and the success of generalisation. A measure might consist of a mere formula or it may involve a complex algorithm including, computation of auxiliary data structures or representations.

Wideness of shape definition contributes to generalisation process complexity so that usual measures do not supply all discriminative information needed for generalisation. We will extend our notion of a measure to each algorithm defining a characterisation by one or several quantified properties of an object, a set of objects, or a variation of a state of these, in order to allow a comparison or significant analysis, which leads to a decision.

Measurement: A numerical value assigned to an observation, which reflects the magnitude or amount of a characteristic (Davis, 1986).

Shape: Shape describes the geometric form of individual spatial objects (Wentz, 1997).

Shape analysis: Shape analysis is the process of building fundamental units for identifying and describing patterns in the landscape (Wentz, 1997).

Auxiliary data structure: A complex data structure applied to objects in order to reach concepts of higher geometric or topologic ordering (such as proximity or connectivity) of an object.

Representation for shape: A formal scheme for describing shape or some aspects of shape together with rules that specify how the schema is applied to any particular shape (Marr, 1982). In our case, a representation is an alternative formal scheme but which keeps all the shape properties. (Describe the shape in polar coordinates rather than cartesian).

Description: The result of using a representation to describe a given shape is a description of that shape in that representation (Marr, 1982).

Pattern: The organisation of phenomena in geographic space that has taken on specific regularity, which in turn is taken as a sign of the working of a regular process (Wentz, 1997).

2 Measures

2.1 Preliminary remarks

We will not deal with measuring throughout the whole generalisation process. Refer to A2 and E1 for initial comments on the significance and position of measuring in our project.

2.2 Measuring in other research-areas

The field of cartography is only one of a range of areas where shape is important. Computer science (esp. computer vision), mathematics, computational geometry, statistics, geology (see esp. Davis, 1986), geography and cognitive science participate in the search for measuring the nature of entities (or group of entities) and in building new representations. Nevertheless, the application, and therefore the goals and definitions in each discipline, are different. Also the demands on good measures vary accordingly. In the field of computer vision, for example, techniques are based on the need to recognise and represent shapes graphically. The rough shape of objects and group of objects have to be detected and the result must be stable if noise was added. For our purpose however, we already have the objects in a usable form and are interested in measuring differences. Besides that, computer vision mainly deals with Raster-graphics, while we mainly focus on vector-representations, which are more descriptive.

A main commonality between most research areas is that most effort was investigated to find new shape measures, while measures on topology or semantic have been neglected. See Section 4.1. for a discussion of this topic in more detail.

Due to different approaches taken by each field, the demands placed on measures vary. To see if other measures meet our needs, we first clarify what we consider a good measure to be for cartographic generalisation. Therefore we first propose a specification of “a good measure”.

2.3 A good measure

A theoretical definition of a good measure could be: “A good measure is discriminative of the observed characteristic of an object AND invariant under any other characteristics”.

However this would imply that each of the characteristics do not correlate with the others. Thus, this last assumption and the exact enunciation of the needed characteristics make it harder to design an adequate measure.

This is why we try to synthesise from a more practical viewpoint a specification for a good measure (see also Lee, 1970; Pavlidis, 1978; Wentz, 1997). As this catalogue should be applicable for all type of measures in our project, we aim to remain general. Additional requirements are mentioned when talking about specific measures (e.g. measures of shape).

From the theoretical point of view:

1. **Robustness**: Small variations should not lead to greatly differing measurements. This is a prerequisite for the next point.
2. **Separability**: A measure should capture as narrowly as possible the aspect it designed to measure.

From the process point of view :

3. **Invariance of person**: A good measure should result in the same measurement, independently of the person who applies it. This is only achievable if the measure is invariant under the choice of application options, such as the start-point where the measure is applied.
4. **Independence of point-representation**: The representation of a cartographic object must not influence the measurement. The same facts (objects) differently represented should result in the same measurement.

From the user point of view :

5. **Ease of calculation**: A good measure should be easy to calculate.
6. **Ease of use**: A good measure has a limited, documented and easy-to-use set of parameters.
7. **Ease of interpretation**: A good measure should be easy to interpret. Therefore, for quite similar measures a similar measurement needs to be calculated. And the result should be in interval/ratio scale.

For measures in computer science it is usual to demand **invariance under geometric transformations** (scale change, translation, rotation). In cartography however, this requirement is not generally applicable, as it is sometimes our intention to detect positional changes or deviations of orientation. This topic is further discussed in chapter 4.1.1.

At last, measures are not supposed to handle any possible data error.

One important step is data conditioning (which can be geometric or topologic or even attribute-based, such as weeding to the scale of digitisation), the use of which will help to standardise (if it can't eliminate) data artefacts. Though some measures may work fairly robustly on dirty data, the extent of object definition must not be enlarged any further because of additional possibilities of data error (two same vertices) or redundant information (for instance, collinear adjacent segments from a polygon). It is then assumed that data are pre-processed to be as consistent as possible and the most common data artefacts weeded out from each object field, before the generalisation process starts.

2.4 Aims of measurements

We have divided our aims of measuring in the generalisation process into three types (cf. Technical Annex, p.44):

To detect particular characteristic of a feature in order to be retained or to be removed by the generalisation process. This kind of measure should be used to select the action for a feature.

To evaluate characteristics in order to optimise the action of an algorithm. In that case, measurements may for instance influence the parameter values of algorithms but also to select distinct algorithms.

To assess actions: because most of the algorithms provide side effects, measures can be applied in order to evaluate efficiency. In that case, actions can be evaluated by comparing object characteristics modified by the action. But it can also measure the variation provided by a generalisation step.

These three classes are studied but can not be developed without algorithms relying on our measures. These correlations will be described in further reports (see task D).

On the contrary, characterisations provided by spatial measurements can be applied to different natures of the objects, for instance, a part of a feature, a set of features and also on different states of the same features. Furthermore, to provide these characterisations, it appears that transformations can be set up on the initial data set in order to facilitate the measure to reach its goal. These transformations can be needed by distinct measures. It is important to bring up these common points to separate them from exact measures. The following classification includes both of these aspects in order to extract a set of measures working on single objects.

3 Classification

3.1 Existing classifications

While a lot of researchers turned their attention to a classification of generalisation algorithms, only limited work was carried out to formalise a classification of measures.

McMaster (1983, 1986) was one of the first researchers to address measures in digital cartography. In a subsequent publication, (McMaster and Shea, 1992) proposed an incomplete summary; a classification using 7 types of measures: density measures, distribution measures, length and sinuosity measures, shape measures, distance measures, Gestalt measures and abstract measures. This classification is also used by some IGN-templates. Nevertheless, IGN adds a more detailed classification using 18 categories, which seem to be built mainly for classifying road-measures.

As the process of measurement is an act of quantifying the character/property of geographical objects, it is clear that measures can also be classified using any classes for the categorisation of character. For example, Clarke (1995) uses the class size, distribution, pattern, contiguity, neighbourhood, shape, scale and orientation.

Ruas and Plazanet (1996) made a meaningful difference between position and shape. These components are not only fundamental for describing the quality of geographic data, but crucial in assessing the initial situation and semantic aspects essential to a particular distribution of cartographic entities.

Another common distinction is made between shape and pattern. While shape focuses on individual objects, pattern describes the geographic distribution of a group of objects. For example, Campbell (1993) makes the distinction between shape, pattern and arrangement.

However, none of the above classifications fulfils our requirements completely. While adopting some ideas from them, we tried to set up another classification that:

1. integrates representations;
2. illustrates the difference between C1 and C2;
3. presents a more process and agent related classification;
4. adopts a hierarchical decomposition as already used for the classification of algorithms.

3.2 Framework for the classification of measures

Some preliminary remarks to the presented classification:

Besides the ‘pure’ measures, we think it is important to list additional functionality that supports, and may be a prerequisite for, the computation of measures. We identified (descriptions see below)

1. Structuring functions;
2. Representations and auxiliary data structures;
3. Support functions;

We are aware that it is not subject of this report to go into depth in auxiliary data structures and alternative representations. Nevertheless we think it is useful to list these structures if a measure is related to them.

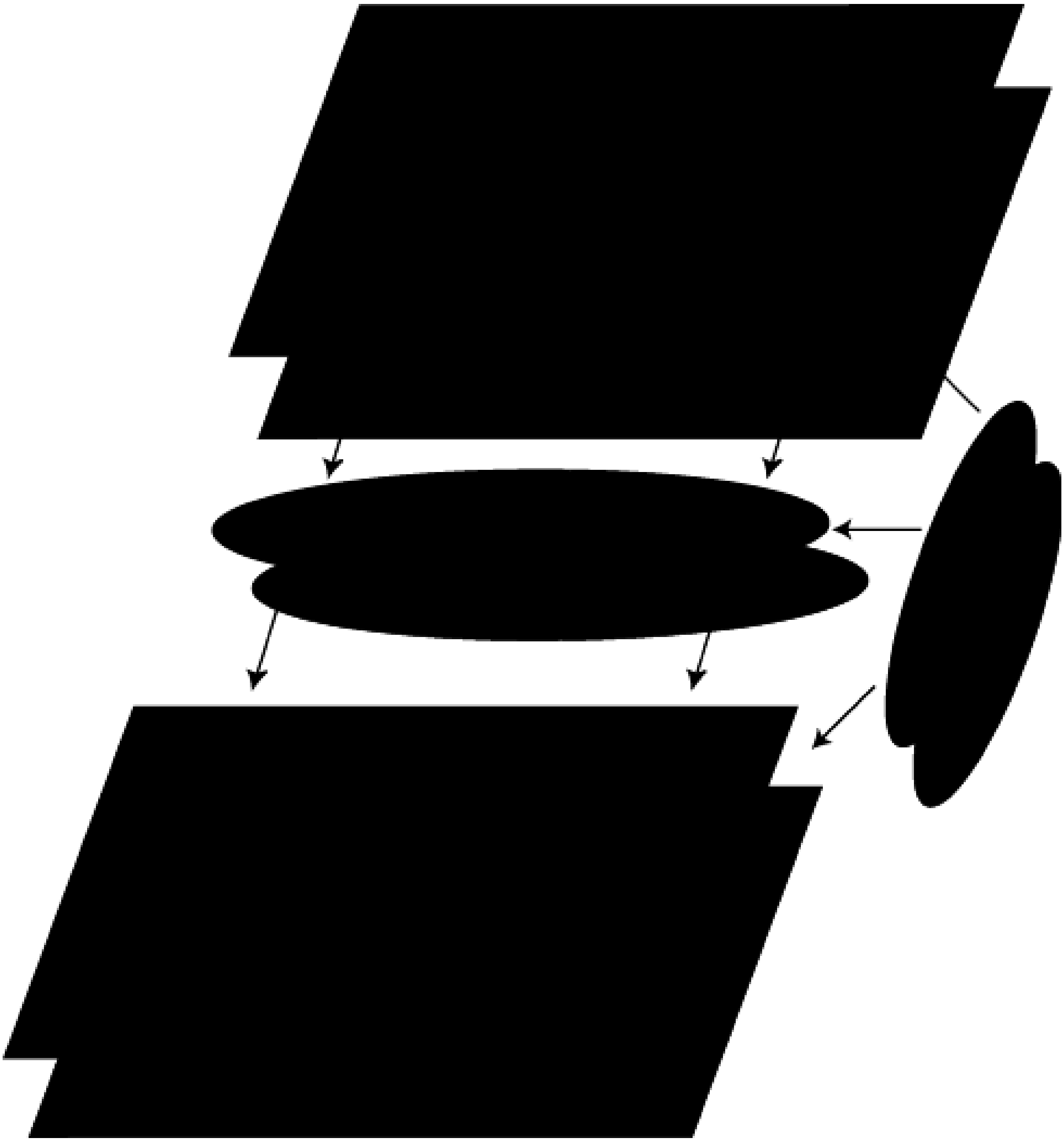

Figure 1 : Classification for measures.

The classification uses a distinction between measures on objects and measures on organisations which reflects the work decomposition for C1 resp. C2.

Measures on objects define ways for characterisations that can be logically attached to an object (entity). For example the sinuosity, length and orientation of a river might be computed. Another important information is its relation to neighbours, such as orthogonality to incoming streams or parallelism to roads. Even when such measures describe inter-object geometry, the resulting value is a description of a character that will be attached to the object. Furthermore, this value is computed without clustering groups of objects to analyse but only based on some local topological relationships.

In contrary, **measures on organisations** need always a specification of the set of objects that will be analysed. For example, the average distance between houses is calculated with respect to a predefined set of houses, like a block. Measures on organisations can use the attributes that arise through measures on individual objects to determine statistics for the whole group, but can also determine independent measurements. We suppose the final measurement will not be attached to every object, but managed in a higher structure.

It is obvious that this decomposition into measures for objects and measures for organisations reflects the proposed use of agents as meso- and micro-agents.

An important hinge between these two main categories is the **structuring functions class**. These functions are needed to set up appropriate groups for further processing (what is a prerequisite for some measures on groups) or to subdivide an object into smaller sub-object (which itself should be treated as individual objects). This allows a more adequate treatment of such objects (for measures as well as for generalisation algorithms). The keywords here are partitioning, clustering and segmentation. The result of such structuring is sometimes itself already the final result of a query, as it provides important information of the nature of objects or organisations. See Chapter 5 for more information.

Alternative **representations** of the geometry **and auxiliary data structures** serve as a foundation for several measures. Even if we could treat such representations as integral parts of such measures, we believe it is helpful to split up the measure and its underlying structure. Examples for representations are parametric curve representations such as the psi-s-plot or t-alpha-plot. An example for an auxiliary data structure is the Delaunay triangulation.

Methods are called **Support functions**, if they ease and support the handling of spatial data. On the one hand side, such functions are used to improve data for further processing, e.g. to interpolate curves at sharp angles. Furthermore, support functions are also data conditioning, providing “pure” data, decreasing risks of wrong interpretations due to artefacts (e.g. remove the vertex from two co-linear adjacent edges).

On the other hand, the functionality originally may have been built into a complex method, but then is separated, because it implements an important operation, such as detecting inflection points or bends. We are aware that all methods of this class might be described in another context. However, such functionality must be specifically identified in order to minimise redundant implementations in distinct algorithms.

A more detailed view of measures and their classification is presented in Chapter 4.

3.3 Extrinsic vs. Intrinsic Measures

Measures can be applied only once to detect an important characteristic, which leads to the best generalisation algorithm to apply. But more commonly, measures are used to compare or evaluate variations between objects in the same state and between same object in different states. This is why even if this report focus on internal measures, measures in generalisation cannot be presented without describing of variation measurements.

The distinction between extrinsic and intrinsic measures was proposed by the IGN. As these terms will be used further in the project we now discuss them in context of the proposed classification for sake of clarity.

An **intrinsic measure** is a measure that can be calculated using one state of an object or organisation only. Examples are the minimum width or area of a polygon, its Hausdorff distance to the closest neighbours, the pattern of a set of lakes or measures of connectivity in a road-network.

Extrinsic measures are computed by comparing different states of one object or one organisation. There are two different classes of extrinsic measures. On the one hand, an **implicit extrinsic measure** might be computed by comparing two intrinsic measurements calculated for both states in isolation (e.g. change in size by subtracting the two areas). On the other hand, there are **explicit extrinsic measures**. These measures really compare the state of the object concurrently (e.g. the vector displacement measures proposed by McMaster, 1986). How is the presented classification related to extrinsic and intrinsic measures? It is important to see that the presented classification builds classes at the concept level, while the distinction of extrinsic and intrinsic measures relates to the algorithm level. Therefore we suggest the proposed classification to build categories at a deeper level, while the terms extrinsic and intrinsic measures might be used to classify measures at the algorithm level. The relation between conceptual level and algorithmic level is normally 1:n, as the same concept can be implemented by more than one algorithm. Nevertheless an algorithm might also be used to implement different concepts.

For example, a measure of distortion of lines resulting through the application of a weeding algorithm is vector-displacement, as presented by McMaster (1986). Even though this measure does use inter-object computation, this algorithm must be classified as a Position-computing measure for extrinsic internal measure. Nevertheless the same algorithm might be used to detect the correspondence of a river and a border. Used in this context, the same type of measure is an inter-object measure.

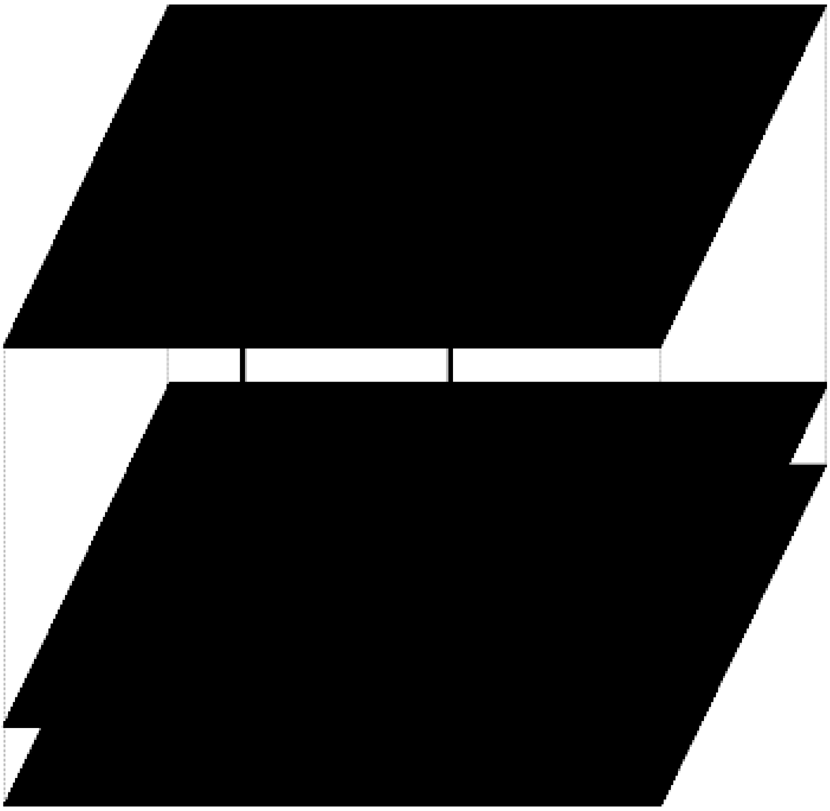

Figure 2 ; Extrinsic and Intrinsic Measures

3.4 The Role of the Data Model: Vector vs. Raster

The presented classification is independent of the underlying data model. For implementation the data model is, of course, relevant. However, the requirements for measures in the raster domain are the same as for vector, even if the formulation might differ.

Characterisation of shapes on raster data started far earlier, meaning a lot of measures already exist. But this project is only dealing with vectors data and because it appears that raster is an auxiliary data structure whose use implies a significant geometrical information loss, we decided to not go into depth in these measures.

4 A Selection of Measures

4.1 Internal measures

4.1.1 Internal Geometry: Position and Orientation, and Shape

Most research efforts were spent so far for the development of measures describing the shape of objects (geometry of objects). The objectives for shape analysis from several sciences are different. Nevertheless there are overlapping agendas that contribute to meeting goals for shape analysis for our cartographic research.

Geographers want to analyse objects on the earth's surface according to their shape, because this is one of their most significant properties. The goal of such measures was the development of shape indices. Shape indices try to assign each shape a unique number, while no two different shapes should be assigned the same number and similar shapes result in number that are close together. There are approaches to express shape by an individual number and to compare unknown shapes to standard ones (templates).

Research in computer science (see Ballard and Brown (1982), Marr, (1982)) is based on the need to recognise and represent shapes graphically rather than forming a unique descriptor that can be used to describe them. The objectives are shape identification and modelling. Shape indices originating from computer science need to remain invariant under rotation, translation and scaling. Such a prerequisite is not germane to our project, as objects need to be treated under rotation. That is why we propose to distinguish shape on the one hand and position and orientation on the other hand. While shape can be discussed without respect to underlying geographic reference (scale, origin, orientation), position & orientation deals with this missing component.

Desirable as it might be to have an assignment function described above as shape indices, no such function can possibly exist. As Lee (1970) proves: There exist no continuous one-to-one function from S, the set of all plane shapes, into R, the set of real numbers. To solve this problem Wentz (1997) suggests to deconstruct shape into different components, where each component can be represented with a number. This method is similar to the way color is separated into hue, saturation and value. Wentz (1997) further proposes using existing measures to identify distinct properties of shape rather than expecting any single measure to capture all aspects of shape. There are several ways to structure shape measures (see for example Pavlidis (1978) or Wentz (1997)).

A main distinguishing characteristic, coming from computer vision, is the observation whether the original shape can be regenerated using the shape measurement only or not (information preserving vs. information non-preserving). Information preserving measures themselves are not of use for cartographic evaluation, as they are too complex to interpret. But they give another viewpoint/insight into data, which can be the basis for further analysis. As such they represent the same nature in a different way (usually by some form of parameterisation). We therefore call such shape-preserving descriptions 'representations'. We did separate such representations from the measure-classification, as they serve rather as underlying structure and might be helpful in different contexts.

Also relevant to our work is whether a measure traces the boundary only to determine its shape measure, or whether the measure examines the points of the interior as well. While boundary measures can also be applied to linear feature, the use of compactness and component measures is restricted to polygons (area features).

4.1.2 Internal Topology

Measures to evaluate the correctness of topology of an object can be applied to connected line and area objects. The computation is limited to test for self-intersection of a line (respectively the outline of areal object) and closure of polygons as inclusions for complex polygons. Of course, topological evaluations will be very more important in inter-objects and organisations analysis.

4.1.3 Semantic

In order to avoid debate, we will for now call any object property, which can be neither fully understood by its stored spatial attributes nor extracted from them, a semantic property.

In the very wide area of map generalisation, categorical maps more often deal with semantic and less geometric accuracy than topographic maps. These kinds of map necessitate some semantic measures to control arising semantic modifications. But, because we are dealing with topographic maps, semantics can be fully handled by the database schema. Measures are then useless because semantic coherence controls are only database-queries, dealing then with the decision field and no more with the evaluation domain, the purpose of this report.

Nevertheless, semantic attributes can be highly correlated to spatial attributes of an object or its class, which could imply some "semantic" measures on single objects. But, literature did not provide any of these so far. This is also why this class of measure won't be mentioned in the following classification. Conversely, this kind of measure appears with inter-object and organisation measures where objects from different natures are compared.

Internal Measures		
Internal Geometry		**Topology**
POSITION & GEOMETRY	**SHAPE**	
Position	**Size**	
Absolute Geometric Position	Length	Self-intersection
Center of Gravity (e.g. Bader, 1997)	Area	
Center of Largest Inscribed Circle (Cromley, 1992)	Perimeter	Closure of Polygon
	Bend height	
A Maxmin-Center (Cromley, 1992)(*)	Maximal Bend height	Inclusion
Vector Displacement (McMaster, 1986)(*)	Minimum Width of Polygon	
Areal Displacement (McMaster, 1986)(*)	Minimum Bounding rectangle	
	Coalescence of Line (Mustiere, 1998)(*)	
	Coalescence Conflict Detection (Mustiere, 1998)(*)	
	Epsilon-Band (Perkal, 1966)	
	Turning Distance (Arkin, 1991)(*)	
	Radial Distance (Bel Hadj Ali, 1997)(*)	
Orientation	**Sinuosity/Complexity**	
Axis of Inertia (Boesch, 1993)(*)	Measures on Angularity (McMaster, 1986)(*)	
Absolute Orientation of a Building (Hangouet, 1998)(*)	Curvilinearity Measures (McMaster, 1986)(*)	
Orientation of a bend (Fritsch,97)(*)	Max Bend height	
	Slope-density Function (Ballard and Brown, 1982)(*)	
	Richardson-Plot (Buttenfield, 1989)(*)	
	Entropy (Bjorke, 1993)(*)	
	Sinuosity (Dutton, 1998)(*)	
	Fourier Descriptors	
	Density of Coordinates (McMaster, 1986)	
	Ratio of Maximum Chord (Ballard and Brown, 1982)(*)	
	Fractal Dimension (see templates)(*)	
	Number of Bends (Plazanet, 97)(*)	
	Elongation / Eccentricity	
	Brown-Eccentricity (Ballard and Brown, 1982)(*)	
	Elongation (Boesch, 1993)(*)	
	Regnauld-Elongation (Regnauld, 1998)(*)	
	Spreadness (Boesch, 1993)(*)	
	Circularity (Davis, 1986)(*)	
	Ellipticity (Davis, 1986)(*)	
	Moments (Boesch, 1993)	
	Compactness	
	Convex Deficiencies (Boesch, 1993)(*)	
	Bending Energy (Young et al., 1974)(*)	
	Miller's Measure (Campbell, 1993)(*)	
	Boyce-Clark radial shape index (Campbell, 1993)(*)	
	Compactness-Measures (Davis, 1986)(*)	
	Squareness (Regnauld, 1998)(*)	
	Wall Squareness (Regnauld, 1998)(*)	
	Important Aspects	
	Minimal Width Parts of a Building	
	Neck Searching (Wang and Mueller, 1993)(*)	
	Bend Shape (Plazanet, 1996)(*)	
	Bend Description (Wang and Mueller, 1998)(*)	
	Characterisation using a Distance-Direction Matrix (*)	
	Number of Points (Regnauld, 1998)(*)	
	Shortest Edge	

Table 1 : Classified internal measures

This state of the art provides a large set of internal measures, but many of their characterising output overlaps. The first set of needed measures for micro-agents is described in D C2 report, selected by the characterising descriptions needed by generalisation algorithms for micro-agents.

4.2 Other measures

In order to be complete in our measure overview, we briefly describe below additional measures applied on several objects.

4.2.1 Inter-object measures

As problems in generalisation arise through conflicts between objects, measurements for detecting and evaluating conflicts between objects are fundamental. Such measures are normally computed for any objects neighboring a given one. The results are stored as one or more attributes on the respective objects. They can reflect minimum, maximum, characteristic or dispersion values and utilise Euclidean, Hausdorff, topologic and other metrics.

The Delaunay triangulation was detected as a powerful auxiliary data structure for inter-object geometry, as well as topology. Nearest neighbours and topological inconsistencies due to displacement are easily detected.

Internal Measures			
Inter-object Geometry		**Inter-object Topology**	
Distance	Minimum Euclidean Distance (Mueller, 1990; Nickerson, 1988) Relative Object Position Coalescence (see Internal Geometry) Hausdorff distance (Hangouet, 1995)(*) Frechet distance (Missing Ref., IGN)(*) Mean Distance between Lines (Matos and Goncalves, 1998) Algebraic Area (IGN)(*) Absolute Area (McMaster, 1986)(*) Conflict Indicator using Voronoi Diagrams (Hangouet, 1998)(*) Distance to Neighbours using Delaunay Triangulation	**Containment** **Connectivity** **Contiguity** **Intersection** **Sideness**	Point-in-Polygon test Shortest Path Intersection of lines Left-of Test
Alignment	Parallelism of lines (Ip and Wang, 1997)(*) Centering (Mustiere, 1995)(*) Relative Orientation of a Building (Hangouet, 1998)(*)		

Table 2: Inter-object classified measures.

4.2.2 Measures on organisations

The measures listed in this section represent an initial list and proposed structure only. Measures for organisations are the subject of C2, and will be presented under that task.

Measures on the **geometry** of organisations are divided similar to measures on objects. The position of organisations can be described by central points and bounding boxes as well as their boundary. The orientation of a group of objects can be computed using geostatistics. While we call the form of individual objects *shape*, we say that organisations of objects have a *pattern*.

Of main interest for measures of **topology** are questions related to networks (road-network, river-network). Many auxiliary data structures for network analysis utilise graph-theoretical approaches.

MEASURES on ORGANIZATIONS			
Geometry of organizations			
POSITION & GEOMETRY		SHAPE	
Position	Location of bivariate centers (Cromley, 1992)	**Dispersion / Distribution**	Mean values on a group of buildings (Hangouet, 1998; Regnauld, 1998)(*)
	Point of central tendency (Cromley, 1992)		Buildings gathering (Regnauld, 1998) (see Structuring Function)
	Weighted bivariate center (Cromley, 1992)		Qualification of homogenity in a group (Regnauld, 1998)(*)
			Nearest Neighbour Analysis (Campbell, 1993; Clarke, 1995)(*)
			Quadrat Analysis (Campbell, 1993)(*)
Orientation	Regression-analysis	**Alignement**	Regression-analysis
		Density	Individual Density (Hangouet, 1998)(*)
		Important Aspects	Remarkable Aspects (Hangouet, Regnauld; Ref IGN)(*)

MEASURES on ORGANIZATIONS		
Topology of organization	**Semantic**	**Fullfilment measures & Measures on maps**
Complex Coastline Hierarchization (Wang and Mueller, 1993)	Importance of Roads (Morisset and Ruas, 1997)(*)	Abs. and Rel. Error Diagrams (Matos and Goncalves, 1998)
Bifuri cation Ration (Campbell, 1993)		Complexity of map (Mackaness, 1995)
		Consistency of generalisation (Jansen, 1998)

Table 3: Classified measures on organisations.

5 Representations and Auxiliary Data Structures

For our purpose we distinguish representations and auxiliary data structures.

Representations describe the geometry of *single primitives* using a transformation from 2-dim Euclidean space to another reference-system. To reconstruct the original Euclidean coordinate representation, specific rules must be applied (as such, they are a shape preserving description). Naturally, representations are only of use if they provide a more suitable starting point for algorithms to make them more efficient or effective. Therefore they should provide implicit geometrical knowledge that was not accessible using a standard Euclidean representation.

In **auxiliary data structures,** relations among a collection of objects are normally captured, in order to represent a geometric or topologic concept of higher order (such as proximity or connectivity). Therefore a complex data structure is maintained parallel to the original one(s).

Representations are used to facilitate computation of internal measures (on single objects). Auxiliary data structures are mainly used to support inter-object measures and measures on organisations. Nevertheless, an auxiliary data structure can also be of help for single object analysis (such as Delaunay triangulation for coalescence conflict detection). Therefore we did not split up representations and auxiliary data structures in our classification, but listed both approaches as sockets for measures. This might cause people to think that all measures are built on representations, which is logical if we regard the standard Euclidean representation as just another representation.

We are aware that data conditioning can be classified in auxiliary data structures but because it deals with data coherence and has to be applied before the generalisation process, we intentionally neglect this part in order to stay focused on our topic.

Representations, as well as auxiliary data structures, might be integrated in complex measures. For sake of clarity we separated measures and their underlying structure.

5.1.1.1 Representations

Plazanet, Affholder and Fritsch (1995), as well as Werschlein (1996) discuss the requirement concerning representations.

5.1.1.1.1 Line-representations

1. Parameterisation of x- and y- coordinates (Buttenfield, 1985)
2. psi-s-plot (O'Neil and Mark, 1987), as a vector analogon to the Freeman Code.
3. t-alpha-plot (Werschlein, 1996)
4. Fourier series (Moellering and Raynar, 1982; Fritsch, 1995)
5. Wavelet transformations (Plazanet, Affholder and Fritsch, 1995; Werschlein, 1996; Fritsch and Lagrange, 1995)
6. Turning Fuction (IGN, *)
7. Radial Function (IGN, *)
8. Curvature Scale Space Image (Mokhtarian and Mackworth 92)

5.1.1.1.2 Representation of polygons

1. Skeleton (Lee, 1982; Shapiro, 1981; Chithambaram, 1991)

2. Raster data

5.1.1.2 Auxiliary data structures

5.1.1.2.1 Subdivisions of geographical space

1. Delaunay-Triangulations (IGN; Bundy, Lee and Jones, 1995; Jones, Bundy and Ware, 1995) (*)
2. Voronoï-Diagrams
3. Voronoï Diagram on segments (IGN, *)
4. Hierarchical Coordinate Systems (Dutton, 1998, 1999)

5.1.1.2.2 Graph-theoretical techniques

1. Graph theoretical approach
2. Minimum Spanning Tree

6 Support Functions

As described above, support functions either improve the data for further processing or add important information that is used by other methods (or is itself already an important characterisation).

We believe that support functions become more important when implementing an extensive generalisation system. It is therefore no surprise that most of this functionality was described by the IGN, who strive for a rather extensible system (such as *Plage*).

6.1.1.1.1 Geometric Improvement of data

1. Cubic Spline Interpolation (IGN, *)
2. Polynomial Interpolation (IGN, *)
3. Weighted Parabola Interpolation (IGN, *)

6.1.1.1.2 Description of important information

1. Detection of Reversal (Nickerson, 1988)
2. Detection of Bends (Wang and Mueller, 1998)
3. Critical Points (Thapa, 1989)
4. Inflection points (IGN, 97)
5. Curvature (IGN, 97)
6. Vertices (IGN, 96)

7 Structuring Functions

We distinguish two different classes of structuring functions.

The first group deals with the process of building groups by selecting a set of objects from the map. There are several ways to accomplish this:

1. **Partitioning**: Partitioning is the process of decomposing a map into different parts (functionally or spatially);
2. **Grouping**: Objects can be grouped by any criterion and combine any sort of features. A special way of grouping is clustering. Also the whole map as a matter of research is a result of grouping all objects. Building groups can also consist of a simple operation such as extracting all roads for topologic measures. As such, grouping is mainly based on database queries.
3. **Clustering**: Clustering is the task of identifying groups in a data set by some criteria of similarity and/or proximity. As such, they focus mainly on a single feature class;

There is abundant literature on clustering. For example note Jain and Dubes (1988) and Kaufman and Rousseeuw (1990). A good introduction to clustering algorithms using graph-theoretical methods is presented by Zahn (1971), using single linkage clustering by Hartigan (1975). Other clustering methods are described recently by Castro and Murray (1998) and dynamic clustering of maps in autonomous agents by (Maio, Rizzi, 1996). The topic of clustering is not described in depth in this report.

The second group of structuring functions decomposes an object into subparts. So far this process is mainly used to segment a line in parts of similar properties. After segmentation, the line parts can be treated as individual objects. The measures described for objects can be applied to them separately. Work in this area was carried out by the IGN (see Plazanet and Affholder (1995), Plazanet (1995)) as well as UNI-ZH (Dutton, 1999).

Regnauld (1998) presented a method to partition a set of buildings into groups having some similar perceptual characteristics. An algorithm for the segmentation of roads and the gathering of buildings is described in the Appendix.

8 Conclusions and Recommendations

Based on the state of the art and in light of the purposes of this report on measures for individual features, we can conclude two points: 1) Effective characterisation and discrimination of features can only be achieved for very specific aspects of a feature influenced by the object semantics. 2) Extended (i.e. large and heterogeneous) features cannot be adequately characterised and generalised as a whole but may need to be segmented and the segments treated separately.

Feature-oriented measures

Measures and methods for generalisation can hardly be de-coupled from object semantics without loosing their discriminative value. Common simple measures such as length, area, elongation etc. can be used to detect basic geometric properties and conflicts. They also provide important elements of the object characterisation that is needed by support functions (e.g., clustering). However, they cannot directly be used to the cartographic nature and necessary consequences of a geometric conflict. In comparison, most effective generalisation algorithms are feature-oriented, meaning they have been specifically designed to enhance and remove feature characteristics of a particular feature class (or group of feature classes). Likewise, measures must also be oriented towards the nature of specific feature classes in order to more accurately detect the cartographic nature of conflicts and select the appropriate generalisation algorithms and agent behaviour, responding as precisely as possible to the specificities of the target generalisation algorithms.

Two complementary approaches are needed to deal with this specialisation. On the one hand, generic geometric measures can be studied in order to calibrate their parameterisation with respect to different feature classes and scales. The state of the art, however, has shown that this strategy is only partly useful. Beyond that, a specialisation of the measures themselves (rather than their conditions of use) is needed. Hence, specialised versions of generic measures must be developed for different feature classes, depending on their prevailing characteristics. This second strategy is illustrated by the measures for buildings specified and developed for E2 (e.g., minimal bounding rectangle, squareness, and compactness) which are all specialised variants of the generic versions of these measures, based on the fact that building sides are commonly angularised. Such an approach will also speed up micro-agent behaviour by aggregating object characteristics rather than applying a whole set of individual measures which later need to be analysed for their discriminating value.

Within the time frame available for C1, we were able to document a wide variety of potentially useful measures, including a general assessment of their usage and usefulness for certain generalisation problems (cf. Appendix). Based on this review, we were then able to select appropriate measures for selected feature classes and develop specialised measures that were identified as missing. It was decided to focus on two feature classes, buildings and roads. The reason for this selection is that these two classes are a) of overriding importance on topographic maps, and b) representative of man-made small area features and natural linear features, respectively. It is hoped that once all tools are available for roads, for instance, it is possible to derive rather directly additional tools and conditions of use for similarly behaving feature classes such as watercourses.

Besides the general assessment of the validity of measures, it was not yet possible to obtain accurate and detailed knowledge on the conditions of use (i.e. the procedural knowledge) of the various measures. This was partly due to the fact that we intended to address the issue of measures on a broad scale rather than narrowing in on buildings and roads from the very start. Furthermore, there was a sequencing problem that implementations of measures and generalisation algorithms had to be available before any empirical testing could be done.

Recommendations:

- Focus on buildings and roads, at least initially.

- Develop specifications and algorithms for the measures as specified in the SPECIFICATIONS for C1 (see deliverable D C2).
- Continue the work on the detailed conditions of use of measures for agents in parallel to C2 (measures for organisations) until Milestone 2. Coordinate with research conducted concurrently with AGENT at IGN, UNI-ZH, and UNI-ED.
- After Milestone 2, adapt the methods and knowledge found for buildings and roads to other feature classes.

Local internal measures

As soon as an object's extent becomes large and its internal structure heterogeneous (e.g., a road extends from a plane into a mountain range) it must be locally analysed in order to segment it into smaller but more homogeneous parts, which can support a generalisation that is adaptive to the particular characteristics of each segment. Thus, measures for differentiating and segmenting parts of objects are one of the most necessary pre-conditions for a successful generalisation. IGN has the longest experience on segmentation of cartographic lines, particularly roads (which form one of the key feature classes of this project). It therefore seems natural to first use IGN's existing algorithms for segmentation and localised generalisation of roads: Coalescence detection, noise detection, detection of inflection points, and characterisation of bends (height, width, orientation) for segmentation and localised generalisation. Further work, however, is needed to ensure appropriate automatic calibration of the measures. To date, manual interaction is still needed for some of these methods (e.g., detection of inflection points). Also, methods need to be extended to deal with artificial shapes on roads that need to be maintained. On the other hand, roads represent a good test case (i.e., the 'worst' case) for linear features; any other linear feature class will be at most as complex. Hence, once the tools for roads have been adequately developed, they can be quite directly specialised for other feature classes. Work is currently under way at the IGN in addition to AGENT purpose (it is to refine the existing knowledge on road segmentation).

While segmentation is an important preprocessing step for linear features it is less significant for areal features. For small areas such as buildings segmentation is only infrequently necessary. The shape character of most buildings is quite homogeneous and they are not large. However, there are cases where that may happen, such as when a building with two large wings connected by a narrow hallway needs to be generalised. In that case, tools for detecting narrow parts (such as the one specified in D C2) can be used. Large areas, such as large forest parcels or lakes are usually dealt with by generalising their outline rather than the area itself, simply because they are too large to visually perceive the entire shape at once; that is also true for conventional cartography. Hence, if segmentation is necessary the same methods can be applied to large areas as for lines (for instance, a forest bordered by a squared pasture and a lake on the other side).

The restriction to roads and buildings seems useful in several ways, as explained above. Furthermore, when more feature classes are added, most conflicts that are going to happen can be expected to be of topological nature, that is, relating to inter-object conflicts. This problem will be studied in more detail in task C2.

Recommendations:

- Focus on roads for segmentation methods for linear features.
- Focus on buildings for simple methods for detecting narrow parts.
- Develop specifications and algorithms for local internal measures for roads and buildings as specified in the SPECIFICATIONS for C1 (see deliverable D C2).
- Continue the work on knowledge acquisition with respect to road segmentation until Milestone 2. This research will be carried out in parallel to AGENT at the IGN.

Auxiliary data structures

Representations and auxiliary data structures must be clearly identified for each measure. We have therefore included a brief survey of auxiliary data structures in this report and in report D D1. This task has only dealt with measures for (micro) agents. Auxiliary data structures, however, are more relevant for measures for organisations (e.g., proximity measures and topological measures). Hence, task C2 will need to study more specifically the role of auxiliary data structures, in particular with respect to having all the important data structures available in the kernel of the prototype system to support all relevant measures. Many measures use the same auxiliary data structures, but there are subtle variations. If these exist, it must be ensured that the data structures or the measures are modified appropriately so they can make use of as few data structures as possible.

Recommendations:

- Work out specifications for auxiliary data structures as early as possible in C2.
- Avoid redundancies between variations of data structures; aim for the common denominator.

9 Bibliography

Affholder, JG, 1997, Généralisation du Linéaire : une approche nouvelle - Rapport interne IGN

Affholder, JG, 1998, Points d'inflexion et Sommets d'une polygonale - Rapport interne IGN

Alt, Behrends, Blömer , 1991, Approximate Matching of Polygonal Shapes - Proceedings of the 7th ACM Symposium on Computational Geometry, pp.186-193.

Alt, Godau, 1995, Computing the Fréchet distance between two polygonal curves - International Journal of Computational Geometry & Applications, vol.5 pp.75-91

Arkin, EM, Chew, Huttenlocher, Kedem, Mitchell, 1991, An efficient computable metric for comparing polygonal shapes - IEEE transactions on pattern analysis and machine intelligence, 13(3), pp 209-216.

Attali, D. and A. Montanvert (1997): Computing and Simplifying 2D and 3D Continuous Skeletons; Computer Vision and Image Understanding, 67 (3); 261-273

Ballard, D.H. and Ch.M. Brown (1982): Computer Vision; Prentice-Hall Inc., Englewood Cliffs (New Jersey); 523 S.

Barrault, M, 1995, Etude de la courbure d'un objet linéaire - Rapport interne IGN

Barrault, M, 1998, PhD Thesis, Le placement des écritures : résolution d'un problème à forte combinatoire et présentant un grand nombre de contraintes variées, Marne la Vallée University.

Beard, M.K. (1991): Constraints on rule formulation; In: Buttenfield, B.P. und McMaster, R.B. (Hrsg). Map Generalization: Making Rules for Knowledge Representation. London, Longman; 121-135

Bel Hadj Ali, A, 1997, Appariement géométrique des objets géographiques et étude des indicateurs de qualité (Mémoire de fin de stage, DEA SIG / GIS Masterthesis, ed. Vauglin) COGIT, IGN, Paris, France.

Bjørke, JT & Midtbø, 1993, Generalisation of digital surface model - 9th Int Carto Conf - Köln. pp 363-371

Boesch, R. (1993): Modellierung von Niederschlagsverlagerungen mit Radardaten; Geo-Processing Reihe, 21, Geographisches Institut der Universität Zürich; 143 S.

Brassel, K., M. Heller and P.L. Jones (1984): The construction of bisector skeletons for polygonal networks; Proceedings of the 4th International Symposium on Spatial Data Handling, 1, Zürich; 117-128

Brazile, F.L. (1998): A generalization machine design that incorporates quality assessment; Proceedings of the 8th Int. Symposium on Spatial Data Handling, Canada;

Brophy, D.N. (1973): An Automated Methodology for Linear Generalization in Thematic Cartography; Proceeding of the ACSM Conference, 33rd Annual Meeting, Washington D.C.; 300-314

Bundy, G.L., C.B. Jones and E. Furse (1995): Holistic approach of large-scale cartographic data; In: Muller, J-C., Lagrange, J.P. and Weibel, R. (Hrsg). GIS and Generalization, GISDATA 1, London, Taylor & Francis; 106-119

Buttenfield, B. (1985): Treatment of the cartographic line; Cartographica, 22 (2); 1-26

Buttenfield, B. (1989): Scale dependance and self-similarity in cartographic lines; Cartographica, 26 (1); 79-99

Buttenfield, B. (1991): A rule for describing line feature geometry; In: Buttenfield, B.P. und McMaster, R.B. (Hrsg). Map Generalization: Making Rules for Knowledge Representation. London, Longman; 150-171

Campbell, J. (1993): Map use and analysis; Wm.C.Brown Publishers, 2nd edition, Dubuque (Iowa); 428 S.

Castro, M. and A.T. Murray (1998): Mining Spatial Data via Clustering; Proceedings of the 8th Int. Symposium on Spatial Data Handling, Canada; 522-532

Chithambaram, R., K. Beard and R. Barrera (1991): Skeletonizing polygons for map generalization; Cartography and GIS/LIS, ACSM-ASPRS, 2, Baltimore; 61-72

Cromley, R.G. (1992): Digital cartography; Prentice Hall, Engelwood Cliffs (New Jersey); 317 S.

Danielsson, P.-E. (1978): A New Shape Factor; Computer Graphics and Image Processing, 7 (2); 292-299

Davis, J.C. (1986): Statistics and Data Analysis in Geology; 2nd Edition, John Wiley & Sons, USA; 646 S.

DeLucia, A.A. and R.T. Black (1987): A Comprehensive Approach to Automatic Feature Generalization; Proceedings of the International Cartographic Conference, Morelia, Mexico; 169-192

Dougenik, J. (1980): Whirlpool: A geometric processor for polygon coverage data; Proceedings of AUTO-CARTO 4; 304-311

Dutton, G. (1999): Scale, Sinuosity and Point Selection in Digital Generalization; Cartography and Geographic Information Systems, 26:1, 33-53.

Dutton, G.H. (1998): A Hierarchical Coordinate System for Geoprocessing and Cartography. Lecture Notes in Earth Sciences 79. Springer-Verlag, 230 pp.

Dutton, G.H. (1981): Fractal enhancement of cartographic line detail; The American Cartographer, 8 (1); 23-40

Ehrliholzer, R. (1995): Quality assessment in generalization: Integrating quantitative and qualitative methods; Proceeding ICA/ACI, 2, Barcelona; 2241-2250

Franklin, A. (1992): Faster line segment intersection; Graphics Gems III, edited by D. Kirk, Axademic Press Inc., San Diego; 199-203

Fritsch, E. (1997): Utilisation de la coubure pour la generalisation du lineaire routier; Bulletin d'Information n.66, IGN, France;

Fritsch, E, 1997, Représentation de la Géométrie et des Contraintes Cartographiques pour la Généralisation du Linéaire Routier - PhD thesis, Univ. Marne-la-Vallée / IGN

Fritsch, E. and J.-P. Lagrange (1995): Spectral Representations of Linear Features for Generalisation; In: Frank, A.U. und W. Kuhn (Hrsg): Spatial Information Theory, Lecture Notes in Computer Science 988, Springer; 157-171

Gouyet, JF, 1992 « Géométrie Fractales » dans Physique et structures fractales, Ed. Masson pp 1-38.

Hanguet, J.-F. (1996): Automated Generalization Fed on latent Geographical Phenomena; Proceedings of InterCarto 2, Irkutsk; 125-128

Hanguet, J.F. (1995): Computation of the Hausdorff distance between plane vector polylines; Proceedings of AUTO-CARTO 12, 4; 1-10

Hangouët, JF, Compte rendu de réunion Généralisation, IGN-DT/970290

Hangouët, JF, 1995, Comput. of the Hausdorff distance between plane vector polylines - AutoCarto12, pp.1-10

Hangouët, JF, 1996, OEEPE Measure Grid

Hangouët, JF, 1996, proceedings of InterCarto2, Irkutsk

Hangouët, JF, 1996, Hausdorff's distance between polylines - OEEPE report

Hangouët, JF, 1998, Voronoï Diagrams on Segments - Properties and Tractability for Generalization Purposes- Technical Report for Agent - Cogit

Hangouët, JF, 1998, Approche et méthodes pour l'automatisation de la généralisation cartographique ; application en bord de ville - PhD Thesis, Univ. De Marne La Vallée / IGN

Hartigan, J.A. (1975): Cluster Algorithms; John Wiley and Sons, New York;

Horton, R.E. (1945): Erosional Development of Streams and their Drainage Basins - Hydrophysical Approach to Quantitative Morphology; Geological Society of America, Bullein 56; 275-370

Ip, H.H.S. and W.H. Wong (1997): Detecting Perceptually Parallel Curves: Criteria and Force-Driven Optimization; Computer Vision and Image Understanding, 68 (2); 190-208

Jain, A.k. and R.c. Dubes (1988): Algorithms for Clustering Data; Prentice Hall, Englewood Cliffs, N.J.;

Jansen, M. and M. vanKreveld (1998): Evaluating the consistency of cartographic generalization; Proceedings of the 8th Int. Symposium on Spatial Data Handling, Canada; 668-678

Jasinsky, MJ 1990 The comparison of Complexity Measures for Cartographic lines, Rapport NCGIA 90.

Joao, E.M. (1995): The importance of quantifying the effects of generalization; In: Muller, J-C., Lagrange, J.P. and Weibel, R. (Hrsg). GIS and Generalization, GISDATA 1, London, Taylor & Francis; 183-193

Jones, C.B., G.L. Bundy and J.M. Ware (1995): Map generalization with a trinagulated data structure; Cartography and Geographic Information Systems, 22 (4); 317-331

Kaufman, L. and P.J. Rousseeuw (1990): Finding Groups in Data: An Introduction to Cluster Analysis; Wiley, New York;

Lam, N.S.-N, and L. DeCola (1993): Fractal simulation and interpolation; In: Lam, N.S.-N. und L. DeCola (Hrsg): Fractals in Geography, Prentice Hall, Englewood Cliffs, New Jersey; 56-74

Lam, N.S.-N. and L. DeCola (1993): Fractals in Geography; Prentice Hall, Engelwood Cliffs (New Jersey); 308 S.

Lecordix, F. and C. Plazanet (1996): Plate-Forme de généralisation PlaGe et application a la caricature; Bulletin d'Information n.65, IGN, France;

Lee, D.R. and G.T. Sallee (1970): A method of measuring shape; Geographical Review, Vol.60 (2); 555-563

Lee, D.T. (1982): Medial axis transformation of a planar shape; IEEE transactions on Pattern Analysis and Machine Intelligence, 4 (4); 363-369

Li, Z. and S. Openshaw (1992): Algorithms for automated line generalization based on a natural principle of objective generalization; IJGIS, 6 (5); 373-389

Longley, P.A. and M. Batty (1989): On the fractal measurement of geographical boundaries; Geographical Analysis, 21 (1); 47-67

Mackaness, W.A. (1994): An Algorithm for Conflict Identification and Feature Displacement in Automated Map Generalization; Cartography and GIS, 21 (4); 219-232

Maio, D., D. Maltoni and St. Rizzi (1996): Dynamic Clustering of Maps in Autonomous Agents; IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (11); 1080-1091

Marr, D. (1982): Vision; W.H. Freeman and Company, USA; 397 S.

Matos, L. and A. Goncalves (1998): Measurement and Analysis of Positional Error; Proceedings of the 8th Int. Symposium on Spatial Data Handling, Canada; 151-160

McMaster, R.B. (1986): A statistical analysis of mathematical measures of linear simplification; The American Geographer, 13 (2); 103-116

McMaster, R.B. (1987): Automated line generalization; Cartographica, 24 (2); 74-111

McMaster, R.B. and K.S. Shea (1992): Generalization in digital cartography; Association of American Geographers, Washington D.C.; 134 S.

Moktharian, F. and Macworth, A.K.(1992) : A theory of multi-scale, curved base shape for planar curves. IEEE Transactions of Pattern Analysis and Machine Intelligence, vol. 14, pp.789-805.

Moellering, H. and J.N. Rayner (1982): The dual axis Fourier shape analysis of closed cartographic forms; The Cartographic Journal, 19 (1); 53-59

Morisset, B. and A. Ruas (1997): Simulation and agent modelling for road selection in generalization; Proceedings 18th ICA/ACI, 3, Stockholm; 1376-1380

Muller, J.C. (1986): Fractal dimension and inconsistencies in cartographic line representations; The Cartographic Journal, 23 (4); 123-130

Muller, J.C. (1987): Optimum point density and compaction rates for the representation of geographic lines; Proceedings of AUTO-CARTO 8; 221-230

Muller, J.C. (1987): Fractal and automated line generalization; The Cartographic Journal, 24 (1); 27-34

Muller, J.C. (1990): The removal of spatial conflicts in line generalization; Cartography and GIS, 17 (2); 141-149

Müller, J.C. and Z. Wang (1992): Area-patch generalization; The Cartographic Journal, 29 (2); 137-144

Mustière, S, 1995, Mesures de la qualité de la gén. du linéaire, Rapport de stage DESS, Univ. Paris I - ENSG

Mustière, S, 1998, An algo for legibility evaluation of Symbolised Lines, Intercarto 4 proc., Barnaul, pp 43-47

Mustiere, S. (1998): Détection et résolution des conflits d'empâtement: une approche adaptative de la généralisation du linéaire routier; Bulletin d'information n.67, IGN, France;

O'Neil, M.P. and D.M. Mark (1987): The psi-s Plot: A useful representation for digital cartographic lines; Proceedings of AUTO-CARTO 8; 231-240

Pavlidis, T. (1978): A Review of Algorithms for Shape Analysis; Computer Graphics and Image Processing, 7 (2); 243-258

Perkal, J. (1966): An attempt at Objective Generalization; In: Michigan Inter-University Community of Mathematical Geographers, Discussion Paper 10, University of Michigan;

Peter, B. (1997): Ableitung von generalisierten Bodennutzungskarten aus der Arealstatistik 1979-1985; Diplomarbeit, Geographisches Institut der Universität Zürich; 167 S.

Plazanet, C. (1995): Measurement, characterization and classification for automated line feature generalization; Proceedings of AUTO-CARTO 12, 4; 59-68

Plazanet, C. (1996): Enrichissement des bases de données géographiques: analyse de la géometrie des objets linéaires pour la généralisation cartographique (application au routes); PHd Thesis of Marne la Vallée University;

Plazanet, C., J.-G. Affholder and E. Fritsch (1995): The importance of geometric modeling in linear feature generalization; Cartography and Geographic Information Systems, 22 (4); 291-305

Pullar, D. (1990): Comparative study of algorithms for reporting geometrical intersections; Proceedings of the 4th International Symposium on Spatial Data Handling, 2, Zürich; 66-76

Regnauld, N. (1996): Recognition of building clusters for generalization; SDH 96, pp.185-198.

Regnauld, N. (1997): Structuration du bati pour la generalisation; Proceedings 18th ICA/ACI, 3, Stockholm;

Regnauld, N. (1998): Generalisation du bati: structure spatiale de type graph et representation cartographique; PHd Thesis of Marne la Vallée University;

Ruas, A, 1995, Multiple paradigms for automating map generalization. Autocarto 12 Vol 4 pp. 69-78

Ruas, A. and C. Plazanet (1996): Strategies for automated generalization; Proceeding of the 7th International Symposium on Spatial Data Handling, Delft (NL); 6.1-6.17

Ruas, A. and W. Mackaness (1997): Strategies for urban map generalisation; Proceedings 18th ICA/ACI, 3, Stockholm; 1387-1394

Shapiro, B., J. Pisa and J. Sklansky (1981): Skeleton generalization from x,y boundary sequences; Computer Graphics and Image Processing, 15 (2); 136-151

Stoer, M. and F. Wagner (1997): A simple Min-Cut algorithm; Journal of the ACM, 44 (4); 585-591

Thapa, K. (1989): Data compression and critical points detection using normalized symmetric scattered matrix; Proceedings of AUTO-CARTO 9; 78-89

Tsai, V, 1993, Fast topological construction of Delaunay triangulation and Voronoï diagrams. Computer & Geosciences Vol 19 n10 pp. 1463-1474

Wang, Z. and J.C. Muller (1993): Complex coastline generalization; Cartography and GIS, 20 (2); 96-106

Wang, Z. and J.C. Muller (1998): Line generalization based on high-level characteristics; Cartography and GIS, 25 (1)

Weibel, R. (1996): A Typology of Constraints to Line Simplification; Proceeding of the 7th International Symposium on Spatial Data Handling, Delft (NL)

Wentz, E.A. (1997): Shape Analysis in GIS; Proceedings of AUTO-CARTO 13; 204-213

Werschlein, Th. (1996): Frequenzbasierte Linienpräsentationen für die kartographische Generalisierung; Diplomarbeit, Geographisches Institut der Universität Zürich; 122 S.

Zahn, Ch.T. (1971): Graph-theoretical methods for detecting and describing Gestalt clusters; IEEE Transactions on Computers, C-20 (1); 68-86

10 Appendix

10.1 Measures

10.1.1 Vector Displacement and Areal Displacement

Name	Vector Displacement and Areal Displacement
Concept	Indication for the 'geometric shift' of a line due to simplification
Extrinsic or intrinsic	Extrinsic
References	McMaster (1986)
Location in the Process	Evaluation of the process that led to the geometric shift
Short description	Vector Displacement: At each vertex of the original line compute the distance to the closest point on the weeded line. Compute the total length of vector differences per inch of line.
	Areal Displacement: Compute the total areal distance between the base line and its simplification.
Input data types	Two matching polylines
Output data types	Real number
Tools required	none
Pre-processing required	
Parameters	
Present state	Used by Jasinski and by Cromley, but code non-available
Drawbacks	
Possible Improvements	
Similar measures	
Remarks	Measure is intended to express the goodness of a weeding. Esp. 'Areal Displacement' might be used to evaluate all kind of shifts (e.g. displacement).

10.1.2 A maxmin-center

Name	A maxmin-center
Concept	Calculate the central point of the skeleton within a polygon to describe its center-point
Extrinsic or intrinsic	Intrinsic
References	Cromley (1992)
Location in the Process	
Short description	Find the longest path of the skeleton between two extremities of the graph. The maxmin center is the midpoint of this path
Input data types	Polygon
Output data types	Point
Tools required	Skeleton
Pre-processing required	-
Parameters	No parameter
Present state	Code for skeletonisation is available (Lull/LSL) but the final measure is not impletemented.
Drawbacks	
Possible Improvements	
Similar measures	See Cromley (1992) for further measures to compute a central point of a polygon.
Remarks	

10.1.3 Axis of inertia

Name	Axis of inertia
Concept	Orientation of a polygon (angle of the axis of inertia)
Extrinsic or intrinsic	Intrinsic
References	Boesch (1993); Reeves A.P. and A. Rostampour (1981): Shape analysis of segmented objects using moments; IEEE, Proc. of the Conf. Pattern Recognition Image Processing; 171-174.
Location in the Process	Characterization of a polygon
Short description	Compute the moments of 1st. and 2nd. order. Use formula to compute orientation.
Input data types	Polygon
Output data types	Real number, expressing the angle of inertia
Tools required	
Pre-processing required	None
Parameters	No parameter
Present state	code non available
Drawbacks	If polygon is not elongated and has not a main orientation, the result might change drastically due to minor distortions.
Possible Improvements	use in conjunction with measure of elongation
Similar measures	Ballard (1982)
Remarks	

10.1.4 Absolute Orientation of a Building

Name	Absolute Orientation of a Building
Concept	Orientation of a building
Extrinsic or intrinsic	Intrinsic
Algorithm	Makes each side of the building's contour start from a same origin. Computes the best-fitting orthogonal cross.
Value	Vector
References	JF Hangouët PhD thesis (1998) - JF Hangouët OEEPE Measure Grid (1996) - JF Hangouët proceedings of InterCarto2, Irkutsk 1996
Location in the process	Contributes to the enriched description of buildings.
Short description	The orientation of the building is defined as that of its (most-of-the-time orthogonal) walls
Input data types	1 building
Output data types	Pair of coordinates
Tools required	None
Pre-processing required	None
Parameters	None
Present state	Coded (LeLisp – Stratège)
Drawbacks	Works only on buildings or shapes with roughly right corners.
Possible improvements	None
Similar measures	Longest side, longest diameter, etc. Nothing as phenomenological as this one. From the Gestalt point of view, see Nicolas Regnauld PhD thesis (1998)
Remarks	What is written in the « drawback » section is not a drawback of the method, since it is its very principle.

10.1.5 Orientation of a Bend

Name	Orientation of a bend
Concept	Computes the main orientation of a bend extracted from a line
Extrinsic or intrinsic	Intrinsic
Algorithm	Emmanuel Fritsch, 1997 PhD thesis
Value	Vector
References	
Location in the process	Contributes to the enriched description of a polyline
Short description	
Input data types	1 polyline
Output data types	Real between 0 an □ defining the angle Real between 0 and 1 : Confidence of the previous value (how far an orientation can be defined)
Tools required	None
Pre-processing required	extraction of the bend
Parameters	None
Present state	Coded in Lull
Drawbacks	
Possible improvements	
Similar measures	
Remarks	Algorithm is more described in the DC2 AGENT report.

10.1.6 Coalescence

Name	Coalescence
Concept	Graphical errors : evaluate how much one line is coalesced
Extrinsic or intrinsic	Intrinsic
Algorithm	-Evaluates the length of the boundary between two close parts of a polyline
Value	Length of the line which is coalesced
References	Mustière, S, 1998, An Algorithm for Legibility Evaluation of Symbolised Lines, Intercarto 4 proceedings, Barnaul July 98, pp 43-47
Location in the process	Characterisation of legibility conflicts
Short description	Given a line L, and a symbolisation width W, compute the location of the symbol edges of L defined as the locus of the points at a distance exactly equal to W/2 to L. Then compute the length of the line where the distance from a point P of L to at least one of the symbol edge is bigger than W/2 * Tolerance
Input data types	a polyline
Output data types	Length
Tools required	Symbol edges computation (= buffer computation)
Pre-processing required	None
Parameters	The parameter Tolerance has been definitely set to 1.7 by comparison of the algorithm evaluation and a human evaluation.
Present state	Coded in ADA
Drawbacks	Still some small bugs Does not detect noise (soft sinuosity)
Possible improvements	Cf. detection of sinuosity (forthcoming)
Similar measures	(Perkal 66)
Remarks	It is very close to a human evaluation of coalescence

10.1.7 Coalescence Conflicts Detection

Name	Coalescence Conflicts Detection
Concept	Split a polyline in homogenous parts regarding the coalescence (a split part is totally or not at all coalesced).
Algorithm	-
References	Mustière, S, 1998, An Algorithm for Legibility Evaluation of Symbolised Lines, Intercarto 4 proceedings, Barnaul July 98, pp 43-47
Location in the process	Can be used to focus on parts of the lines which are homogeneous and where we can apply a specific operator. It is a focalisation (or pre-treatment)
Short description	Given a line L, and a symbolisation width W, compute the location of the symbol edges of L defined as the locus of the points at a distance exactly equal to W/2 to L. Then compute the length of the line where the distance from a point P of L to one of the symbol edge is bigger than W/2+Tolerance
Input data types	Polyline
Output data types	Set of polylines (a partition of the input polyline)
Tools required	Buffer computation
Pre-processing required	None
Parameters	The parameter Tolerance has been definitely set to 1.7 by comparison of the algorithm evaluation and a human evaluation.
Present state	Coded in ADA
Drawbacks	Still some small bugs Does not detect noise (soft sinuosity)
Possible improvements	Cf. detection of sinuosity (forthcoming)
Similar measures	-
Remarks	It is the basis of the GALBE process (described in D1)

10.1.8 Turning Distance

Name	Turning Distance
Concept	Evaluate the shape similarity by computing a mathematical distance between the shape of 2 polygons.
Algorithm	Based on turning function
Value	positive real number, between 0 and $\sqrt{3\pi}$
References	Arkin et al. 1991 : Arkin, E.M., Chew, L. P., Huttenlocher, D. P., Kedem, K., & Mitchell, J. S. B. (1991). An efficient computable metric for comparing polygonal shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (3), 209-216. Bel Hadj Ali 1997 : Bel Hadj Ali, A. (1997). Appariement géométrique des objets géographiques et étude des indicateurs de qualité (Mémoire de fin de stage, DEA SIG / GIS Masterthesis, ed. Vauglin) laboratoire COGIT, IGN, Paris, France.
Location in the process	Validation : Comparison between 2 polygons = extrinsic
Short description	The Turning Function θ_a and θ_b is computed for both polygons A and B. The Turning Distance between polygon A and polygon B $d_\theta(A,B)$ is $\sqrt{\min_{\omega \in [0;2\pi[, t \in [0;1]} \int_0^1 \lvert \theta_a(s+t) - \theta_b(s) + \omega \rvert^2 ds}$. This distance is an actual mathematical distance verifying all the criteria of a distance.
Input data types	2 polygons
Output data types	A real number
Tools required	Turning Function
Pre-processing required	Computation of the Turning Function for both polygons
Parameters	
Present state	Coded in C/C++
Drawbacks	It is highly sensitive to small details if they are not homogeneously distributed on the curvilinear absciss, but is very efficient to compare global shape if the small details (i.e. noise) is homogeneous.
Possible improvements	Further study to define threshold helping to match areal features
Similar measures	Angular distance, radial distance
Remarks	This definition restricts this measure on polygons and can not then be applied on polylines.

10.1.9 Radial Distance

Name	Radial Distance
Concept	Distance between the radial shape of 2 polygons
Extrinsic or intrinsic	Extrinsic
Algorithm	Based on radial function
Value	Positive real number
References	Bel Hadj Ali 1997.
Location in the process	Validation : Comparison between 2 polygons
Short description	The radial function f_a and f_b is computed for both polygons A and B. The Radial Distance between polygon A and polygon B $d_r(A,B)$ is $\int_0^1 \lvert f_a(t) - f_b(t) \rvert dt$. This distance is an actual mathematical distance verifying all the criteria of a distance.
Input data types	2 polygons
Output data types	A real number
Tools required	Radial Function
Pre-processing required	Computation of the Radial Function for both polygons
Parameters	-
Present state	Coded in C/C++
Drawbacks	It is not sensitive to small details, is not very efficient to compare global shape, properties need further study
Possible improvements	Further study to define threshold helping to match areal features. It seems that the mean squared error is a theoretically valid threshold for that purpose
Similar measures	Angular distance, Turning distance
Remarks	-

10.1.10 Measures of angularity

Name	Measures of angularity
Concept	Evaluate the reduction of 'microwiggliness' after simplification
Extrinsic or intrinsic	Extrinsic
References	McMaster (1986)
Location in the Process	Evaluation of simplification algorithm
Short description	The percentage change in angularity may be expressed as the sum of the angles between consecutive vectors on the simplified line divided by this sum on the base line.
Input data types	Two lines
Output data types	Dimensionless real number
Tools required	
Pre-processing required	
Parameters	
Present state	Code non available
Drawbacks	
Possible Improvements	adjustments are needed if input is a polygon
Similar measures	See McMaster (1986) for more angularity measures
Remarks	% angularity, angularity per vertex and angularity per unit length are all used; each measures a somewhat different concept

10.1.11 Curvilinearity measure

Name	Curvilinearity measure
Concept	Compare segments of same curvilineaity
Extrinsic or intrinsic	Extrinsic
References	McMaster (1986)
Location in the Process	Evaluation of simplification algorithms
Short description	Curvilinear segments are portions of a line in which all angles are in the same positive or negative direction. Count and compare the number of curvilinear segments before and after simplification.
Input data types	Two polylines
Output data types	Dimensionless integer
Tools required	
Pre-processing required	
Parameters	No parameter
Present state	Code non available
Drawbacks	
Possible Improvements	
Similar measures	McMaster (1986) describes 4 different measures of curvilinearity
Remarks	can be expressed per unit length or per vertex; related to “zero-crossing number”

10.1.12 Slope Density Function

Name	Slope Density Function
Concept	Characterisation of the angularity
Extrinsic or intrinsic	Intrinsic
References	Ballard (1982); Nahin, P.J. (1974): The theory and measurement of a silhouette descriptor for image processing and recognition; Pattern Recognition, 6 (2)
Location in the Process	Characterisation of a polygon
Short description	The SDF is the histogram or frequency distribution of psi collected over the boundary.
Input data types	Polygon
Output data types	Histogram
Tools required	Psi-s curve
Pre-processing required	
Parameters	No Parameter
Present state	Code non available
Drawbacks	There are additional techniques necessary to interpret the histogram
Possible Improvements	
Similar measures	
Remarks	

10.1.13 Richardson-Plot

Name	Richardson-Plot
Concept	Self similarity
Extrinsic or intrinsic	Intrinsic
References	Buttenfield (1989)
Location in the Process	Characterisation of a line
Short description	Step along a length of a line in equal increments and compute its length. Set this measure as a point in the Richardson-plot (x-axis: log Length of step; y-axis: log Total length). Decrease step size and iterate. Analyse the final plot.
Input data types	Polyline
Output data types	Richardson-Plot
Tools required	
Pre-processing required	
Parameters	min step, max step, step size or ratio
Present state	Code non available
Drawbacks	Result is a plot that needs further analysis (see A.13)
Possible Improvements	?
Similar measures	See measures on Fractal dimension
Remarks	linearity of plot indicates self-similarity, and a plateau on the left may indicate the largest scale / smallest resolution at which the digitisation is useful

10.1.14 Fractal Dimension

Name	Fractal Dimension
Concept	complexity of a line
Extrinsic or intrinsic	Intrinsic
Algorithm	-
Value	Fractal dimension
References	Jasinsky, M.J. 1990 The comparison of Complexity Measures for Cartographic lines, Rapport NCGIA 90. Müller, J.C. 1987 Fractal and Automated Line Generalisation, The cartographic Journal Vol 24 pp 27-34. Gouyet, J.F. 1992 « Géométrie Fractales » dans Physique et structures fractales, Ed. Masson pp 1-38. Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN See also Dutton (1981, 1998), Lam (1993), Longley (1989)
Location in the process	Characterisation of a line
Short description	The line is travelled by the method of Walking divider : we obtain the number N of segments of length ε necessary to walk along the line (the intersection between the line and the circle centered on the current point is the following point). Then, the fractal dimension is lim (1-log N)/logε when ε → 0. If s is the slope of the function log ε → log N, the fractal dimension is also 1-s.
Input data types	a polyline
Output data types	positive number : continuous, ordered (the greater the value, the more complex the line)
Tools required	line sharing method (Van horn, Walking divider, ...), regression analysis, sig. tests
Pre-processing required	None
Parameters	No parameter
Present state	Analysed
Drawbacks	ε has to be chosen not too small (> resolution) and not too big (< half length of the line)
Possible improvements	Automation of the choice of ε
Similar measures	-
Remarks	Not very useful or hard to use with human lines like roads. More useful but hard to use with some rivers and with coast line as it is usually determined via regression (Richardson plot), there is a significance value to be considered. Also, the robustness of estimates depends on degree of self-similarity; at some scales this may change, so that D is possibly scale-dependent.

10.1.15 Entropy

Name	Entropy
Concept	order, regularity
Extrinsic or intrinsic	Intrinsic
Algorithm	-
Value	Without dimension
References	Bjørke, J.T. & T. Midtbø 1993 Generalisation of digital surface models. 9th International Cartographic Conference. Cologne. pp 363-371
Location in the process	Characterisation of a line => intrinsic
Short description	If αi is the angle between the ith and the i+1th segments of the line and $p(\alpha)$the probability for the angle to belong to an interval, we have : $H = \Sigma[0,2\pi]\ p(\alpha) \log(p(\alpha))$.
Input data types	a polyline
Output data types	positive number : continuous, ordered (the greater the value, the less ordered the line)
Tools required	None
Pre-processing required	None
Parameters	angular size of interval over which to sum
Present state	Analysed
Drawbacks	As entropy is an indicator of local sinuosity, it is very sensitive to digitising defaults.
Possible improvements	A small smoothing is necessary to avoid this drawback
Similar measures	-
Remarks	measure described only characterises angles; a similar measure can be computed for segment lengths.

10.1.16 Dutton's Sinuosity

Name	Dutton's Sinuosity
Concept	Sinuosity of a line
Extrinsic or intrinsic	Intrinsic
References	Dutton (1999)
Location in the Process	Characterisation of a line
Short description	Sinuosity is computed for each vertex along a polyline by constructing a ratio of distance +/- k vertices along the line to the length of an anchor line connecting the first and last vertices
Input data types	Polyline
Output data types	Ratio of Real numbers
Tools required	-
Pre-processing required	-
Parameters	k, the topological distance around each vertex
Present state	
Drawbacks	Needs a classification of the dimensionless number for further use
Possible Improvements	
Similar measures	See fractality measures
Remarks	Can be adapted to measure how similar the local simuosity at a vertex is to that of a specified neighborhood around it

10.1.17 Regnauld-Elongation

Name	Regnauld-Elongation
Concept	Gives information on the shape of a building, not taking the concavity into account
Extrinsic or intrinsic	Intrinsic
References	Phd Nicolas Regnauld 1998 : Generalisation du bati : structure spatiale de type graphe et representation cartographique
Location in the process	It takes place during the building simplification process. In case of rectangle (or assimilated to), to know the length and the width of the final shape. In case of complex shape, its used with the squareness to know what makes the shape different from a square (elongation or concavity)
Short description	This is the ratio between the width and the length of the minimum bounding rectangle of a building. It varies between 0 for a line to 1 for a square
Input data types	A polygon
Output data types	A real
Tools required	none
Pre-processing required	None
Parameters	no parameter
Present state	Implemented in Lull
Drawbacks	
Possible improvements	
Similar measures	
Remarks	Useful for buildings with 4 points. Then knowing the elongation and the area is enough to determine the final shape.

10.1.18 Convex deficiency

Name	Convex deficiency
Concept	Complexity and Compactness of an object
Extrinsic or intrinsic	Intrinsic
References	Boesch (1993); Phillips, T.H. (1985): A shrinking technique for complex object decomposition; Pattern recognition letters 3, 271-277.
Location in the Process	Characterisation of polygons
Short description	Formula to compute a number of convex deficiency form the polygons area and its Convex Hull.
Input data types	Polygon
Output data types	dimensionless real number
Tools required	Convex Hull
Pre-processing required	-
Parameters	-
Present state	
Drawbacks	
Possible Improvements	
Similar measures	For measures on concavity and a representation called 'concavity-tree', see also Sklansky J. (1972): Measuring concavity on a rectangular mosaic, IEEE Trans. Computers 21.
Remarks	This measure is of use to see if a polygon is too complex and should therefore be split into simpler objects.

10.1.19 Squareness (compactness for buildings)

Name	Squareness (compactness for buildings)
Concept	Gives information on the shape of a polygon. Return a value between 0 for a segment and 1 for a square, and can increase until 4/pi (1.27) for a circle. The value decreases with the elongation and the concavity of the shape.
Extrinsic or intrinsic	Intrinsic
References	Phd Nicolas Regnauld 1998 : Generalisation du bati : structure spatiale de type graphe et representation cartographique
Location in the process	It is useful for characterising buildings. When the value is close to 1, it means that the building can be simplified as a square, with no need for further measures.
Short description	(16 * area) / (perimeter * perimeter)
Input data types	A polygon
Output data types	A real between 0 and 1.27
Tools required	Area and perimeter
Pre-processing required	None
Parameters	no parameter
Present state	Implemented in Lull
Drawbacks	Low values need other measure to know if the low value is due to elongation or concavity (or both)
Possible improvements	
Similar measures	Comes from the indice iso-perimetrique from Coster and Chermant 1989, adapted to have the value 1 for squares instead of circles
Remarks	

10.1.20 Wall squareness

Name	Wall squareness
Concept	Measures the angles at each corner of a boundary, and statistical information on it. Gives information on the shape of a building, with regard to the angles of its walls.
References	Templates of measures (Edinburgh report on LaserScan exercice)
Extrinsic or intrinsic	Intrinsic
Location in the process	It takes place during the building simplification process. In case of rectangle (or assimilated to), to know the length and the width of the final shape. In case of complex shape, its used with the squareness to know what makes the shape different from a square (elongation or concavity)
Short description	Computes mean, standard deviation and minimum/maximum value with the angles of a boundary
Input data types	A polygon
Output data types	3 real numbers
Tools required	None
Pre-processing required	None
Parameters	No parameter
Present state	Implemented in Lull
Drawbacks	
Possible improvements	
Similar measures	
Remarks	Useful for buildings with 4 points. Then knowing the elongation and the area is enough to determine the final shape. If the number of points in the geometry are known the mean is fairly meaningless since it is the same as the 2*pi/number of points. The other statistical information is the most useful.

10.1.21 Neck Searching

Name	Neck Searching
Concept	Identify necks in a coastline
Extrinsic or intrinsic	Intrinsic
References	Wang and Mueller (1993)
Location in the Process	Construction of a river-network
Short description	Two points that have a distance falling below a small threshold value but are non-contiguous in the coordinate string may form the neck of a river/bay. Some additional checks are needed
Input data types	Polyline (coastline)
Output data types	Pairs of coordinates building necks
Tools required	-
Pre-processing required	?
Parameters	?
Present state	?
Drawbacks	Usually the rivers are already explicitly stored in the database and need therefore no identification. This identifies local shapes which presumably are not feature-coded
Possible Improvements	?
Similar measures	-

Remarks	- can be achieved via rolling-ball/buffer; it is a potentially expensive test

10.1.22 Minimal width parts

Name	Minimal width parts
Concept	Detects the closest couple of non adjacent parts of a polygon.
Extrinsic or intrinsic	Intrinsic
References	AGENT D C2
Location in the Process	Detection of conflict sources due to a building shape
Short description	Shape is here supposed to be not self intersecting. Distance defining the closest two parts is involving either an edge and a vertex or two vertices. then, distance from any vertex to any edge (including its two vertices) is computed. the smallest is stored. The two involved parts are supposed to be non adjacent. The closest vertex implies its two edges (and then the two neighbouring vertices) and an edge its two vertices. Two parts are non adjacent if they have not any vertex in common.
Input data types	Non overlapping polygon
Output data types	real : minimal distance integer : number of the vertex in the polygon real : characterising the point of the edge from where the vertex is the closest. It can be on of its vertices.
Tools required	
Pre-processing required	
Parameters	
Present state	Coded in Lull
Drawbacks	Especially designed for buildings, it can not be applied on overlapping shapes.
Possible Improvements	
Similar measures	
Remarks	

10.1.23 Bend Shape

Name	Bend Shape
Concept	Shape
Extrinsic or intrinsic	Intrinsic
Algorithm	arrange inflection points in a 2-levelled hierarchy
References	Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN
Location in the process	Micro-characterisation of lines shape : characterisation for roads
Short description	With 2 values of σ (gauss parameter), a small and a great one, it is possible to get 2 levels of importance for the inflection points, level 1 and 2. Then, the vertices are computed (there also are of 2 levels) and a bend (part of the line comprised between 2 level-1 inflection points) is described as the succession of its characteristic points (i for inflection point and s for vertex). Each of them associated with its level, 1 or 2.
Input data types	a polyline
Output data types	set of strings for each bend of the polyline
Tools required	Gaussian smoothing Detection of characteristic points (inflection points and vertices)
Pre-processing required	None
Parameters	No parameter
Present state	Coded in ADA
Drawbacks	-
Possible improvements	-
Similar measures	Simple measures as 1) height/base (base = distance between inflection points) 2) difference of slope between the inflection points 3) Shape of the curvature function between the inflection points 4) Ratio between area and convex hull
Remarks	As roads are artificial (contrary to mountains, rivers, ...), the number of bend shapes is limited.

10.1.24 Bend Description (and Repetition of Bends)

Name	Bend Description (and Repetition of Bends)
Concept	Description of a bend
Extrinsic or intrinsic	Intrinsic
References	Wang and Mueller (1998)
Location in the Process	Characterisation of a bend
Short description	Used after the bend-identification method presented by the same authors. The description is based on the size of a bend (=area of the polygon enclosed by the bend and its baseline) and its shape (=compactness index of bend-polygon)
Input data types	Bend from bend-detection (Wang and Mueller, 1998)
Output data types	2 real numbers
Tools required	Bend-detection (Wang and Mueller, 1998)
Pre-processing required	-
Parameters	No Parameter
Present state	?
Drawbacks	?
Possible Improvements	?
Similar measures	-
Remarks	The resulting descriptors might be used for further analysis of a series of bends (e.g. detect similar bends in series; detect the most important bend)

10.1.25 Bend Height

Name	Bend Height
Concept	Description of a polyline
Extrinsic or intrinsic	Intrinsic
References	DC2 Agent report
Location in the Process	Characterisation of a bend
Short description	Bend is here defined as a continuous part of a smoothed polyline between two inflexion points. A bend height is then computed as the Hausdorff distance between the bend and the two inflexion points segment.
Input data types	1 polyline
Output data types	1 real describing the distance 1 real between 0 and 1, the relevance of bend orientation value
Tools required	Inflexion points computation, Hausdorff Distance,
Pre-processing required	Inflexion point detection
Parameters	Flatness of the analysed bend in order to also handling exceptional shapes
Present state	Coded in Lull
Drawbacks	
Possible Improvements	
Similar measures	bend Description
Remarks	Actually, two approaches are used in this measure, depending on the flatness of the bend.

10.1.26 Max Bend Height

Name	Maximal Bend Height
Concept	Description of a polyline
Extrinsic or intrinsic	Intrinsic
References	DC2 Agent report
Location in the Process	Characterisation of a bend
Short description	Bend is here defined as a continuous part of a smoothed polyline between two inflexion points. Each bend height is computed and the highest is noticed.
Input data types	1 polyline
Output data types	1 real describing the distance 1 real between 0 and 1, the relevance of bend orientation value
Tools required	Inflexion points computation, Hausdorff Distance, Bend height
Pre-processing required	Inflexion point detection
Parameters	Flatness of the analysed bend in order to also handling exceptional shapes
Present state	Coded in Lull
Drawbacks	
Possible Improvements	
Similar measures	bend Description
Remarks	cf. Bend Height

10.1.27 Number of Bends

Name	Number of Bends
Concept	Description of a polyline
Extrinsic or intrinsic	Intrinsic
References	Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN
Location in the Process	Characterisation of a bend
Short description	Bend is here defined as a continuous part of a smoothed polyline between two inflexion points. All bends of a polyline are then counted.
Input data types	1 polyine
Output data types	1 positive or null Integer
Tools required	
Pre-processing required	Inflexion point detection
Parameters	Sigma for smoothing the initial polyline real for defining digitalisation accuracy of the polyline
Present state	Coded in Lull
Drawbacks	Result is strongly correlated to the sigma value. The more line is smoothed, the less number of bends is decreased
Possible Improvements	
Similar measures	bend Description
Remarks	

10.1.28 Characterisation using a distance-direction matrix

Name	Characterisation using a distance-direction matrix
Concept	Characterisation of a line on the basis of angularity and sinuosity
Extrinsic or intrinsic	Intrinsic
References	McMaster (1995)
Location in the Process	Characterisation of the polyline
Short description	A matrix is centred over the start point of each segment. The cell of the matrix, where the endpoint of the segment is placed, gets incremented. This results in a distance-direction matrix. This matrix can be analysed
Input data types	Polyline
Output data types	The type of a line, depending on the predefined groups.
Tools required	-
Pre-processing required	-
Parameters	Size of the matrix
Present state	
Drawbacks	
Possible Improvements	
Similar measures	
Remarks	The result is of use to group polylines with similar matrices using clustering techniques.

10.1.29 Number of points

Name	Number of points
Concept	Gives information on the shape of a building. A building with 4 points is (nearly always) a rectangle.
Extrinsic or intrinsic	Intrinsic
References	Phd Nicolas Regnauld 1998 : Generalisation du bati : structure spatiale de type graphe et representation cartographique
Location in the process	First measure before simplifying a building. Depending on the result, more complex measures can be needed or not.
Short description	Compute the number of non-colinear points of the boundary of a polygon
Input data types	A polygon
Output data types	An integer
Tools required	none
Pre-processing required	None
Parameters	no parameter
Present state	Implemented in Lull
Drawbacks	
Possible improvements	
Similar measures	
Remarks	Useful for buildings with 4 points. Then knowing the elongation and the area is enough to determine the final shape.

10.1.30 Hausdorff Distance

Name	HAUSDORFF DISTANCE
Concept	Relative position : distance
Extrinsic or intrinsic	Both
Algorithm	2 algorithms : without Voronoï and from Voronoï
Value	Meters or other linear units
References	« Computation of the Hausdorff distance between plane vector polylines » - Hangouët - AutoCarto12, 1995, pp.1-10 / Hausdorff's distance between polylines – Hangouët, OEEPE report, 1996 / Approximate Matching of Polygonal Shapes - Alt, Behrends, Blömer, Proceedings of the 7th ACM Symposium on Computational Geometry, pp.186-193, 1991.
Location in the process	Comparison between 2 states of maps => Extrinsic Comparison between 2 adjacent features => Intrinsic Used by other measures (in intrinsic contexts, e.g. computation of objects remoteness)
Short description	The Hausdorff distance between two objects is a true mathematical distance expressing their locational remoteness.
Input data types	LINEAR => 2 polylines AREAL => 2 polygons MIXED => 1 polyline and a polygon COMPLEX => curvature function of 2 polylines, ... ? ? ? ?
Output data types	Positive number : continuous, ordered
Tools required	alg. in AutoCarto12 : no tool required / alg. by Alt et al. : Voronoï diagram
Pre-processing required	None
Parameters	No parameter
Present state	Coded in LeLisp (AutoCarto) / C++ (from Voronoï). Some bugs due to computational mathematics.
Drawbacks	None (it performs what it's meant to)
Possible improvements	None
Similar measures	Fréchet Distance (for locational ⊕ morphological remoteness)
Remarks	Approximate (i.e. exact in most cases) computations are straightforward. With the current programs : Haphazard success with Voronoï, bugs with AutoCarto.

10.1.31 Frechet Distance

Name	Fréchet Distance
Concept	Relative position : distance
Extrinsic or intrinsic	Both
Algorithm	Fréchet
Value	Meters or other linear units
References	Computing the Fréchet distance between two polygonal curves - Alt, Godau, International Journal of Computational Geometry & Applications, vol.5 1995 pp.75-91 ; Compte rendu de réunion Généralisation, Hangouët IGN-DT/970290
Location in the process	Comparison between 2 states of maps => Extrinsic Comparison between 2 adjacent features => Intrinsic Used by other measures (in intrinsic contexts, e.g. computation of objects proximity)
Short description	A true mathematical distance that expresses morphological ⊕ locational remoteness
Input data types	LINEAR => 2 polylines (or 2 contours) COMPLEX => curvature function of 2 polylines, ... ? ? ? ?
Output data types	Positive number : continuous, ordered
Tools required	None
Pre-processing required	None
Parameters	No parameter
Present state	Discussions
Drawbacks	None
Possible improvements	None
Similar measures	Hausdorff Distance (that expresses strict locational remoteness)
Remarks	Involves not only proximity but also shape

10.1.32 Algebraic Area

Name	ALGEBRIC AREA
Concept	Relative position : area
Extrinsic or intrinsic	Intrinsic
Algorithm	Very basic algorithm
Value	Square meters or other areal units
References	-
Location in the process	Comparison between 2 states of maps => Extrinsic Used by other measures (in intrinsic contexts, e.g. computation of objects proximity)
Short description	The 2 polylines constitute a polygon. The algebric area is the sum of the trapezes delimited par each segment of the polygon, i.e. Σ(xi+1-xi)*(yi+1+yi)/2
Input data types	LINEAR => 2 polylines COMPLEX => curvature function of 2 polylines, ...
Output data types	Number : continuous, ordered but be careful (see drawbacks)
Tools required	None
Pre-processing required	Join end-points
Parameters	No parameter
Present state	Coded in ADA
Drawbacks	Necessity of joint end-points May be zero though lines are very different
Possible improvements	See ABSOLUTE AREA
Similar measures	Absolute Area,
Remarks	-

10.1.33 Absolute Area

Name	Absolute Area
Concept	Relative position : area
Extrinsic or intrinsic	Extrinsic
Algorithm	Mc Master
Value	Square meters or other areal units
References	Mc Master, R.B., 1986, A Statistical Analysis of Mathematical Measures for Linear Simplification, The American Cartographer 13(2) pp103-117
Location in the process	Comparison between 2 map states Used by other measures(in intrinsic contexts, e.g. computation of objects proximity)
Short description	Computation of the intersections between the lines Computation of the area separating the lines between 2 consecutive intersection points Addition of the absolute values of these areas
Input data types	LINEAR => 2 polylines COMPLEX => curvature function of 2 polylines, ...
Output data types	Positive number : continuous, ordered
Tools required	None
Pre-processing required	fuse the end-points
Parameters	No parameter
Present state	Coded in ADA
Drawbacks	When complex intersections between the lines exist, it may be not pertinent
Possible improvements	-
Similar measures	Relative Area
Remarks	Divided by the length of one of the lines, it can give an approximation of the mean distance between them

10.1.34 Conflict Indicator

Name	Conflict Indicator
Concept	Two contextual measures to 1/ spot proximity conflict between features and 2/ evaluate the importance of the conflict. A proximity conflict occurs when the minimal interdistance is smaller than the allowed threshold 2.δ. The conflict is important when it is repeated along the features' sides that face each other.
Extrinsic or intrinsic	Intrinsic
Algorithm	1/ select Voronoï edges between features (interface edges) whose attribute « dist-min-to-the-figure » is smaller than δ. => each pair of features involved is conflicting. 2/ for a pair of conflicting features, compute λ, length of the conflicting Voronoï interface between them (ie. where « d-to-the-sites » is smaller than δ), and Σ, the area backboned by the conflicting interface and bounded by the features' contours and the segments that project the conflicting interface's endpoints onto each feature. Ratio : GC = Σ / λ_, when small, indicates an important conflict (features are close to each other - λ - and face each other over a long distance - Σ / λ.
Value	1/ boolean 2/ real (no unit)
References	Voronoï Diagrams on Segments - Properties and Tractability for Generalization Purposes – Technical Report for Agent - JF Hangouët - Cogit - March 3, 1998
Location in the process	Contextual measure For generalization decision purposes
Short description	cf. 'concept' and 'algorithm' above
Input data types	1/ all features 2/ a pair of conflicting features
Output data types	1/ boolean 2/ real, positive, roughly ordered (if at threshold δ1 pair AA is conflicting with importance GC1aa, and pair BB with importance GC1bb, with GC1aa < GC1bb, then at the larger threshold δ2, if AA and BB are isolated from other features, we still have GC2aa < GC2bb).
Tools required	Voronoï program
Pre-processing required	Voronoï diagram
Parameters	δ is the only parameter : resolution allowed for separation between features (e.g. Related through scale to visual separation power).
Present state	Dependent on the Voronoï program's success. Coded in LeLisp.
Drawbacks	None (computes what it's meant to).
Possible improvements	None
Similar measures	None
Remarks	this is a contextual measure

10.1.35 Parallelism of lines

Name	Parallelism of lines
Concept	Detect parallelism between two lines
Extrinsic or intrinsic	Extrinsic
References	Ip and Wong (1997)
Location in the Process	Characterisation of situation
Short description	Step 1: Decomposition of curves into simple segments; Step 2: Prescanning for possible coupling between simple segments; Step 3: Force-driven coupling; Step 4: Parallelism verification;
Input data types	Two polylines
Output data types	Boolean
Tools required	
Pre-processing required	
Parameters	
Present state	
Drawbacks	
Possible Improvements	
Similar measures	
Remarks	Algorithm not yet studied in detail

10.1.36 Centering

Name	Centering
Concept	Relative position : evaluate how much one line is centered on another one
Extrinsic or intrinsic	Extrinsic
Algorithm	-
Value	Between 0 and 1 : 1 => a line is centered on the other one 0 => a line is shifted from the other one (no intersections)
References	Mustière, S, 1995, Mesures de la qualité de la généralisation du linéaire, Rapport de stage DESS, Univ. Paris I - ENSG
Location in the process	Comparison between 2 states of maps => Extrinsic (validation)
Short description	Based on area computation of McMaster : value = 2*max (area on the left, area on the right) / absolute area
Input data types	Two versions of a polyline
Output data types	number in [0,1] : continuous, ordered
Tools required	Area computation
Pre-processing required	None
Parameters	No parameter
Present state	Coded in ADA
Drawbacks	-
Possible improvements	-
Similar measures	-
Remarks	Not sufficient by itself : has to be used with other measures evaluating relative position

10.1.37 Relative Orientation

Name	Relative Orientation
Concept	The absolute orientation of the building is expressed not in the natural system of coordinates, but in that of the building's access street section.
Extrinsic or intrinsic	Extrinsic
Algorithm	Computation of the building's absolute orientation Projection of the building onto the street section it belongs and computation of the street's local tangent. Rotation.
Value	Vector.
References	JF Hangouët PhD thesis (1998) - JF Hangouët OEEPE Measure Grid (1996) - JF Hangouët proceedings of InterCarto2, Irkutsk 1996
Location in the process	Contributes to the description of relationships between buildings and streets. Intrinsic.
Short description	See algorithm above.
Input data types	2 objects
Output data types	angle
Tools required	None
Pre-processing required	Absolute orientation + association of the building to its access street section.
Parameters	no parameter
Present state	coded (LeLisp - Stratège)
Drawbacks	None
Possible improvements	Useless
Similar measures	
Remarks	

10.1.38 Nearest Neighbour technique

Name	Nearest Neighbour technique
Concept	Compare Distribution with 'Standard'-Distribution
Extrinsic or intrinsic	Intrinsic
References	Campbell (1993), Clarke (1995)
Location in the Process	Characterisation of point distribution
Short description	The mean of the distance observed between each point and its nearest neighbour is compared with the expected mean distance that would occur if the distance would be random
Input data types	Set of points
Output data types	Real number
Tools required	-
Pre-processing required	-
Parameters	No Parameter
Present state	Was developed in 1960's and extensively used in pre-GIS times
Drawbacks	
Possible Improvements	
Similar measures	Data analysis
Remarks	

10.1.39 Quadrat analysis

Name	Quadrat analysis
Concept	Compute distribution of points
Extrinsic or intrinsic	Intrinsic
References	Campbell (1993)
Location in the Process	Characterisation of point distribution
Short description	A uniform grid is drawn over the distribution of interest. Number of Points occurring within each quadrat is recorded. The variance (=number of points in each grid cell with the average number of points over all cells) is compared to the char. of random scatter
Input data types	Set of points
Output data types	Real number
Tools required	-
Pre-processing required	-
Parameters	The size and orientation of the grid does highly influence the result
Present state	Was developed in 1960's and extensively used in pre-GIS times
Drawbacks	The size and orientation of the grid does highly influence the result
Possible Improvements	
Similar measures	Data analysis
Remarks	

10.1.40 Mean Values on a Group of Buildings

Name	Mean Values on a Group of Buildings
Concept	Buildings in a city-block may usually be generalized in isolation from what happens in other areas on the map. Displacement or selection within the city-block however cannot be performed blindly, lest the geographical distribution be distorted. The selection and displacement of a building will be decided on its contribution to the (ir)regularity of the group. The independent qualities of some importance for the regularity of a group of buildings have been identified as follows : semantics size shape orientation relative orientation on the road distance to the road distance between buildings These qualities' mean values are computed on the group, and comparing a given buildings qualities against the mean value will make it possible to detect whether it is remarkable or not (next entry). In addition, the neighbouring relationships between buildings has to be retrieved, either from a simple chaining (when the group is aligned) or from the more elaborate usage of a Minimal Spanning Tree (MST).
Extrinsic or intrinsic	Intrinsic / contextual
Algorithm	(depends on the quality measured, most of the time : mathematical mean)
Value	vector of mean values (e.g. code for semantics, areal unit for size, number of corners for shape, 2D-vectors for orientation and relative orientation, length unit for distance to road and for distance between buildings)
References	JF Hangouët PhD Thesis - N. Regnauld PhD Thesis
Location in the process	description enrichment
Short description	semantics is measured from the feature's 'nature' attribute. size can be measured by the area of the building shape can be measured by the number of corners absolute and relative orientations can be measured as described somewhere else in the document distance to the road is both minimum distance from contour to road and distance from centroid to road distance between buildings is both minimum distance between contours and distance between centroids.
Input data types	set of buildings + their delimiting streets
Output data types	vector of mean values, vector of standard deviations
Tools required	tools for the computation of the qualities above
Pre-processing required	Identification of city-blocks and groups of buildings.
Parameters	no parameter
Present state	exists in several programmes.
Drawbacks	-
Possible improvements	Each quality can be computed with many different measures and algorithms.
Similar measures	-
Remarks	There's an interesting discussion on qualities / measures / algorithms for describing

	measures in Hangouët's PhD thesis pp.198 sqq.

10.1.41 Qualification of homogeneity in a group

Name	Qualification of homogeneity in a group
Concept	Given a collection of values (in a generalisation context, the collection is a set of geographical objects and we study one of their character), gives information on homogeneity. When a group is qualified as homogeneous, then typical value and exceptions are computed
Extrinsic or intrinsic	Intrinsic / contextual
References	Phd Nicolas Regnauld 1998 : Generalisation du bati : structure spatiale de type graphe et representation cartographique
Location in the process	Used at a meso-level to qualify a structure, in order to determine a particular operation in case where homogeneity has to be taken into account.
Short description	Compute the mean and the standard deviation. Then remove one by one all most atypical values until the remaining values are similar. Then compare the number of remaining values with the proportion parameter to determine if the group is homogeneous or not.
Input data types	A set of values (real)
Output data types	A Boolean to specify if the group is homogeneous or not If homogeneous, specifies the typical value, the exceptional values if any, and the standard deviation.
Tools required	Mean and standard deviation
Pre-processing required	The group to study and the values of the characteristic it is wished to study
Parameters	2 parameters : the proportion of exception allowed. Default value = 0.2 (20 % of exceptions allowed) the maximum deviation allowed (a coefficient : difference between the value and the median divided by the standard deviation). Default value = 1
Present state	Implemented in LISP on Stratege
Drawbacks	
Possible improvements	Could be improved by providing degree of homogeneity or uniqueness rather than a boolean
Similar measures	entropy of distributions
Remarks	

10.1.42 Individual Density

Name	Individual Density
Concept	Two measures : 1/ Measure of the individual density of a feature (or tightness within its surroundings) 2/ Qualification of the importance of the tightness in function of the separation threshold allowed
Extrinsic or intrinsic	Intrinsic / contextual
Algorithm	1/ Divide the feature's area by that of its Voronoï region 2/ (see Reference for the description of this tricky operation)
Value	1/ real, > 0 2/ real ,> 0
References	Voronoï Diagrams on Segments - Properties and Tractability for Generalization Purposes - Technical Report for Agent - JF Hangouët - Cogit - March 3, 1998
Location in the process	Both contextual and subjective measures For generalization decision purposes. Intrinsic.
Short description	See 'concept' and 'algorithm' above
Input data types	a set of objects
Output data types	a set of densities
Tools required	Voronoï program & diagram
Pre-processing required	no
Parameters	The second measure involves 2 parameters: separation threshold (between features) and perception threshold (for a surface feature).
Present state	simulated
Drawbacks	None
Possible improvements	None
Similar measures	None
Remarks	This is a contextual measure

10.1.43 Remarkable Aspect

Name	Remarkable Aspect
Concept	A building in a group can be remarkable : 1/ because of its position in the row (at one of its end-positions) or in the group (at a fork in the MST) 2/ because one or several of its qualities are outstanding (either much smaller or much greater than the average on the group, see entry above)
Extrinsic or intrinsic	Intrinsic / contextual
Algorithm	1/ is computed from the ordinal number of the building in the row, or from the indication of a division in the MST. 2/ is computed on the belonging of the measured building's quality to the mean and deviation interval of the quality on the group
Value	code for its distinctive characteristic
References	Jef Hangouët / Nicolas Regnauld
Location in the process	Before any generalisation operations is decided. Intrinsic
Short description	(see 'algorithm' above)
Input data types	a feature and the description of the average values of the group where it belongs
Output data types	mark
Tools required	nothing extraordinary
Pre-processing required	Computation of the mean values on the group
Parameters	-
Present state	Coded very differently by very different people
Drawbacks	-
Possible improvements	-
Similar measures	-
Remarks	The mean quality values on the group can be computed only from the actual values of each individual feature. They're better off stored away, fit for future computations such as this computation of remarkability.

10.1.44 Bifurcation Ratio

Name	Bifurcation Ratio
Concept	Compare characteristics of different basins
Extrinsic or intrinsic	Intrinsic or Extrinsic
References	Campbell (1993)
Location in the Process	Characterisation of drainage networks
Short description	The bifurcation ratio is the ratio between the number of links of one order and the number of links of the next higher order. The number can be compared between different stream networks.
Input data types	Number of links at each level of stream-order
Output data types	Real number
Tools required	Ordering of network, e.g. Strahler, semi-log regression
Pre-processing required	-
Parameters	No Parameter
Present state	?
Drawbacks	The measure is very 'global', as it compares different drainage basins. Such holistic measures are not often of use.
Possible Improvements	?
Similar measures	drainage density
Remarks	useful in retaining character of network that is being pruned

10.1.45 Importance of Roads

Name	Importance of Roads
Concept	Build order of roads by computing the importance of roads
Extrinsic or intrinsic	Intrinsic
References	Morisset and Ruas (1997)
Location in the Process	Characterisation of roads
Short description	The use of roads is studied by simulating movement of traffic
Input data types	Road network
Output data types	Number for each road within the network
Tools required	Handling of graphs; Quickest path; Monte Carlo simulation
Pre-processing required	
Parameters	
Present state	
Drawbacks	
Possible Improvements	
Similar measures	
Remarks	A very specific algorithm

10.1.46 Shortest edge

Name	Shortest edge
Concept	Detect the shortest edge of a polyline.
Extrinsic or intrinsic	Intrinsic
Algorithm	Computes distance of each edge and store the smallest
References	
Location in the process	Detection of conflict sources in the building generalisation process.
Short description	
Input data types	a polyline
Output data types	integer : index of the first vertex of the shortest edge.
Tools required	
Pre-processing required	None
Parameters	No parameter
Present state	Coded in Lull
Drawbacks	-
Possible improvements	-
Similar measures	
Remarks	

Structuring Functions

10.1.47 Segmentation

Name	Segmentation
Concept	Split a polyline in homogeneous sinuous parts
References	Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN
Location in the process	Can be used to analyse a line or to focus on homogeneous parts for an adaptive process
Short description	It is based on inflection points detection, their resistance to a smoothing of the line and on some basic measures of a single bend analysis : each bend of the line is characterised and homogeneous clusters are built
Input data types	a polyline
Output data types	Set of polylines (partition of the original polyline)
Tools required	Inflection points and vertices detection, bend shape
Pre-processing required	None
Parameters	There is 1 parameter, linked to a ground distance (it is linked to the size of the bends that we consider as relevant bends). When the parameter increase, the number of split parts decrease (or the shape analysis goes less in details),
Present state	Coded in ADA
Drawbacks	Difficulty to set the parameter
Possible improvements	A better inflection points detection would give better results
Similar functions	Sinuosity estimation (Dutton 1998, 1999)
Remarks	-

10.1.48 Buildings Gathering

Name	Buildings Gathering
Concept	Gather buildings using criteria relative to the theory of Gestalt (visual grouping perception)
References	Phd Nicolas Regnauld 1998 : Generalisation du bati : structure spatiale de type graphe et representation cartographique
Location in the process	Used at a meso-level on a road partition when the density of building is too high to preserve all buildings, and not high enough to amalgamate all the buildings in a built-up area. It's a first stage for an algorithm of global building typification.
Short description	Compute a minimum spanning tree on the buildings inside a partition cell. Then segment this graph to isolate groups having some perceptual characteristics (regular spacing, similar size, similar orientation).
Input data types	A set of buildings. When used for typification, buildings are those included in a same partition cell.
Output data types	Group of buildings characterised with regard to their pattern, size similarity and orientation similarity.
Tools required	Distance (minimum distance between two buildings), area, orientation, elongation computation, plus statistical (mean and standard deviation) computation.
Pre-processing required	Space partitioning using the road network
Parameters	No parameters. Could add one to vary the granularity of the grouping
Present state	Implemented in LISP on Stratege
Drawbacks	More efficient when buildings pattern is linear.
Possible improvements	Auto detect when it is not suitable to the current situation
Similar measures	
Remarks	

10.2 Representations

10.2.1 Turning Function

Name	Turning Function
Concept	Evolution of the changes of slope of the line defining a polygon
References	Arkin et al. 1991 / Bel Hadj Ali 1997
Location in the process	Characterisation of areas (or lines)
Short description	Choose an initial point anywhere on the contour ; this point will be the origin of curvilinear abscissas. The function of the curvilinear abscissa that gives the angle between a horizontal vector and the contour is the Turning Function.
Input data types	AREAL => 1 polygon LINEAR => 1 polyline
Output data types	Function of the curvilinear abscissa, positive, continuous, has values on a 2π long interval.
Tools required	None
Pre-processing required	None
Parameters	No parameter
Present state	Coded in C/C++
Drawbacks	Gives a description of the shape only (does not depend on position)
Possible improvements	-
Similar functions	Angular function, Radial function
Remarks	Has not been tested on lines, only on areas. It is not sure that it can be useful for lines.

10.2.2 Radial Function

Name	Radial Function
Concept	Relative position : distance
References	Bel Hadj Ali 1997
Location in the process	Characterisation of areas
Short description	Choose a central point anywhere within the contour ; we always choose the center of gravity when it is inside the polygon. Compute the distance between that central point and any point of the contour : this distance given in function of the curvilinear abscissa is the Radial Function when the origin of abscissas is chosen on the first intersection between the contour and the horizontal half-line starting on the central point and going in the same direction as increasing x.
Input data types	AREAL => 1 polygons
Output data types	Function of the curvilinear abscissa, positive, continuous.
Tools required	None
Pre-processing required	None
Parameters	Choice of the central point
Present state	Coded in C/C++
Drawbacks	Gives a description of the shape only (does not depend on position)
Possible improvements	A better understanding of the influence of the choice of the central point on the result
Similar functions	Angular function, Turning function
Remarks	-

10.2.3 Voronoi Diagram on Segment

Name	Voronoi Diagram on Segment
Concept	The Voronoï diagram assigns each element of a set of 'sites' (points and segments) with a cell in the plane so that any point inside the cell is closer to its originator-site than to any other. Cells have linear or parabolic edges.
References	Voronoï Diagrams on Segments - Properties and Tractability for Generalization Purposes - Technical Report for Agent - JF Hangouët - Cogit - March 3, 1998
Location in the process	Data structure (similar to topology)
Short description	the Voronoï diagram on segments is the exact dual of the features' geometry. A great variety of basic and difficult measures between features (contextual measures) or on a given feature can be computed easily, systematically and exactly from it.
Input data types	Features and their geometry
Output data types	Voronoï edges and regions
Tools required	Voronoï programs
Pre-processing required	None
Parameters	No parameters
Present state	Haphazard success with the incremental Voronoï program used (bugs and errors entailed by computational mathematics)
Drawbacks	None
Possible improvements	None
Similar functions	None
Remarks	Voronoï or Delaunay on points are mere intuitive and uncontrolled approximations of Voronoï on segments. The Voronoï on segments contributes to the understanding of what is usually intended with Voronoï and Delaunay on points

10.2.4 Delaunay Triangulation

Name	Delaunay Triangulation
Concept	Triangulation on a set of points : you obtain a set of triangles linking the points so that no point of the set is included in the circumcircle of each triangle.
References	Tsai, V. 1993 Fast topological construction of Delaunay triangulation and Voronoï diagrams. Computer & Geosciences Vol 19 n10 pp. 1463-1474 Ruas, A. 1995 Multiple paradigms for automating map generalization. Autocarto 12 Vol 4 pp. 69-78
Location in the process	Analysis of neighbouring and proximities => contextual (intrinsic)
Short description	1) Computation of the convex hull 2) Triangulation of the convex hull 3) The points are added successively
Input data types	An array of points with integer coordinates
Output data types	A list of the triangles that were built on these points OR A list of the points and edges (included in the triangles) with a graph structure
Tools required	None
Pre-processing required	None
Parameters	No parameter
Present state	Coded in ADA on PLAGE and in LISP on STRATEGE
Drawbacks	Only the list of triangles is returned, but some functions are provided that enable to retrieve the neighbours of a given point
Possible improvements	Return not only a list of triangles but also a list of edges
Similar functions	-
Remarks	Possibility of false constraints in order to have a given segment in the structure returned by densifying the initial set of points

10.3 Support Functions

10.3.1 Modelling by Arc of a Circle and Cubic

Name	Modelling by Arc of a Circle and Cubic
Concept	Modelling in a context of lines generalisation by geometric comparison with lines already generalised
References	Affholder, JG, 1997, Géneralisation du Linéaire : une approche nouvelle - Rapport interne IGN Affholder, JG, 1998, Points d'inflexion et Sommets d'une polygonale - Rapport interne IGN
Location in the process	Before statistical research of generalisation rules (this research is based on comparison between 2 homologous polylines corresponding to different scales)
Short description	Each initial polyline, after a small smoothing in order to get rid of digitising errors, are segmented into portions limited by 2 consecutive inflection points. Sometimes (the most simple case), between 2 inflection points, there is only a vertex (maximal curvature point). These portions are modelled by 2 straight segments at the extremities, an arc of a circle at the vertex and 2 cubics in order to join the segments to the arc of a circle.
Input data types	Polylines library
Output data types	Modelled portions library
Tools required	Gaussian smoothing
Pre-processing required	None
Parameters	None
Present state	Coded in FORTRAN
Drawbacks	The statistical research of this modelling is supposed to be very long Does not take into account contextual generalisation
Possible improvements	-
Similar functions	-
Remarks	-

10.3.2 Cubic Spline Interpolation

Name	Cubic Spline Interpolation
Concept	Modelling
References	IGN Cogit : Xavier Barillot
Location in the process	Possibly before measures or algorithms
Short description	1) Acute vertices along the polyline are widened into obtuse angles by adding a point on both sides of each acute vertex. 2) slope at a vertex is defined as the line joining its 2 neighbours. 3) Then, in a given coordinate system, there is only one cubic joining 2 consecutive points and constrained by the computed slopes. So, we obtain a sequence of cubic portions and the final polyline is a n-unit sampling of these portions. Let's notice the coordinate system is chosen in order to obtain a curve close to the original segment.
Input data types	- a polyline - a real (n = the step of the sampling)
Output data types	- the interpolated polyline - the parameters of each cubic - the angle between the chosen coordinate system and the initial one - the inflection points computed from the cubic
Tools required	None
Pre-processing required	Disangularization of the polyline (Cf. 1) in short description above)
Parameters	n (sampling step) has to be smaller than the minimal length of the segments of the polyline
Present state	Coded in ADA - well tested
Drawbacks	The second derivative is not continuous ; nevertheless, the inflection points are well located
Possible improvements	1) The second derivative becomes continuous with a small smoothing (advantage : suppression of not significant inflections) 2) With 4-degree polynomial function, it is possible to get a continuous second derivative, but the result is not constrained 3) See polynomial interpolation
Similar functions	See other modelling
Remarks	None

10.3.3 Polynomial Interpolation

Name	Polynomial Interpolation
Concept	Modelling
References	IGN Cogit : Xavier Barillot
Location in the process	Possibly before any treatment
Short description	First, the coordinates of the curvature centre at each point of the polyline is computed (possibly by several ways : centre of the circle / intersection of the mediatrix / ...). Then, in a given coordinate system, there is only one 5-degree polynomial function joining 2 consecutive points and constrained by the computed curvature centres. So, we obtain a sequence of portions of polynomial function and the final polyline is a pas-unit sampling of these portions. In order to
Input data types	a polyline
Output data types	- the interpolated polyline - the parameters of each polynomial function
Tools required	Computation of curvature centre
Pre-processing required	None
Parameters	pas (sampling unit) has to be smaller than the minimal distance of a segment of the polyline
Present state	Coded in ADA - not very well debugged
Drawbacks	Instability : there sometimes are some unnecessary oscillations
Possible improvements	If a 6-degree polynomial function is used, it is possible to choose a coordinate system where all roots are negative and then to avoid oscillations
Similar functions	See other modelling
Remarks	None

10.3.4 Weighted Parabolic Interpolation

Name	Weighted Parabolic Interpolation
Concept	Modelling
References	IGN Cogit : Xavier Barillot
Location in the process	Possibly before any treatment
Short description	In a given coordinate system, there is only one parabola joining 3 consecutive points. Then, a segment (2 points) of a polyline belongs to 2 parabolas. The interpolated line is a weighted sum of these 2 parabolas (with respect to the distance to the central point of the parabola). So, we obtain a sequence of portions of sum of parabolas and the final polyline is a pas-unit sampling of these portions.
Input data types	- a polyline - the unit of the sampling (pas)
Output data types	- the interpolated polyline
Tools required	None
Pre-processing required	
Parameters	pas (sampling unit) has to be smaller than the minimal distance of a segment of the polyline
Present state	Coded in ADA - not tested
Drawbacks	The second derivative is not continuous
Possible improvements	See other interpolations
Similar functions	See other modellings
Remarks	None

10.3.5 Curvature

Name	Curvature
Algorithm	Curvature computation by convolution of the slope with a gaussian smoothing
References	Barrault, M. 1995 Etude de la Courbure d'un objet linéaire. Rapport interne IGN Barrault, M. 1998 PhD, Le placement cartographique des écritures…, Université de Marne la Vallée,.
Location in the process	Segmentation of polylines Characterisation of line shape => intrinsic
Short description	The slope at each point is the one of the segment of its neighbours.
Input data types	a polyline σ (Gaussian parameter)
Output data types	a function
Tools required	None
Pre-processing required	Eventually smoothing with decomposition with a given step
Parameters	The greater σ, the less inflections it will remain
Present state	Coded in ADA
Drawbacks	Because of smoothing, small inflections does not remain (and the problem is that we do not know what small mean)
Possible improvements	-
Similar functions	See other curvatures
Remarks	-

10.3.6 Curvature

Name	Curvature
Algorithm	Computation by convolution
References	Emmanuel Fritsch, 1997 PhD thesis
Location in the process	Segmentation of polylines Characterisation of line shape => intrinsic
Short description	The initial difficulty lies in the definition of curvature on polyline, since for a reason of angularity, polyline curvature is not defined. We have considered curvature as a distribution (sum of Dirac at each point of the polyline), whose smoothing by gaussians -and other regular functions- is a well-defined function.
Input data types	a polyline σ (gauss parameter) a value of curvilinear abscissa
Output data types	the value of the curvature for the curvilinear abscissa
Tools required	None
Pre-processing required	None
Parameters	The greater σ, the less inflections it will remain
Present state	Coded in ADA
Drawbacks	Edge effect Because of smoothing, small inflections does not remain (and the problem is that we do not know what small mean)
Possible improvements	-
Similar functions	See other curvatures
Remarks	-

10.3.7 Curvature

Name	Curvature
Algorithm	Curvature computation from a mathematical interpolation of the polyline
References	IGN Cogit : Xavier Barillot
Location in the process	Segmentation of polylines Characterisation of line shape => intrinsic
Short description	The curvature is computed from the cubic or polynomial modelling of the polyline, either from the parameters (*) of the polynomial function or from the final discrete line (**) (for instance with circles comprising 3 consecutive points).
Input data types	a polyline sigma of gauss-smoothing
Output data types	a function
Tools required	Modelling by cubic spline or polynomial function
Pre-processing required	Modelling by cubic spline or polynomial function
Parameters	sigma : to remove small inflections
Present state	Coded in ADA
Drawbacks	(*) the shape of the curvature is a bit broken and the higher the degree of the polynomial function, the flatter the shape at the extremities (**) the importance of inflections is uniform (no hierarchy between inflections)
Possible improvements	(*) smoothing (**) put a weight from the difference of slopes between 2 consecutive inflection points
Similar functions	See other curvature computations
Remarks	None

10.3.8 Inflection Points

Name	Inflection Points
Concept	Computation of inflection points
Algorithm	-
Value	Set of points
References	Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN
Location in the process	Characterisation of shape of lines => intrinsic
Short description	After a σ-gaussian smoothing, the sign variations of the angle between 2 consecutive segments of the line are detected. The inflection point is chosen at the middle of the segment.
Input data types	a polyline σ (gauss parameter)
Output data types	Inflection points
Tools required	Gaussian smoothing
Pre-processing required	None
Parameters	The greater σ, the more important the inflection points; but, there is no stable correlation between the degree of importance and the value of σ : it depends on the nature of the line. Then, this parameter is not easy to choose.
Present state	Coded in ADA
Drawbacks	Cf. Parameters above. Moreover, a gaussian smoothing moves the line and the inflection points so; It is then sometimes hard to retrieve the correspondent inflection points on the original line. Inversely, this parameter has to be high enough, otherwise the micro-inflections would be kept.
Possible improvements	1) Smooth the line constituted by the angles between 2 consecutive segments and only keep those greater than a given threshold. 2) Cancel inflections very close of another one (but depending on the direction of walking along the line)
Similar measures	See cubic-spline interpolation
Remarks	-

10.3.9 Vertices

Name	Vertices
Concept	Computation of vertices
Algorithm	-
Value	Set of points
References	Plazanet C. 1996 Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes). Rapport de Thèse, Univ. Marne-la-Vallée / IGN
Location in the process	Characterisation of shape of lines => intrinsic
Short description	Between 2 consecutive inflection points (see template above), there is only one vertex : it is the point of maximal curvature (see Curvature of Barrault).
Input data types	a polyline
Output data types	Vertices
Tools required	Detection of inflection points [Plazanet, 96] Curvature computation [Barrault, 95]
Pre-processing required	None
Parameters	-
Present state	Coded in ADA
Drawbacks	An important gaussian smoothing (Cf. template above) moves the line and of course also the vertices.
Possible improvements	1) Smooth the line with a great value of σ for the inflection points computation but with a small value for the vertices computation.
Similar measures	See cubic-spline interpolation
Remarks	-

10.3.10 Additional Formulas

Name	Reference	Concept	Formula	Explanation
Brown-Eccentricity	Ballard and Brown (1982)	Eccentricity	$e = \frac{(M_{20} - M_{02})^2 + 4M_{11}}{area}$	M define the Moments: $M_{ij} = \sum_{x \ in \ R} (x_0 - x)^i (y_0 - y)^j$
Ratio of Maximum Chords	Ballard and Brown (1982)	Eccentricity	$e = \frac{A}{B}$	Ratio of the length of maximum chord A to maximum chord B perpendicular to A
Elongation	Boesch (1993)	Elongation	$el = \frac{M_{\max} - M_{\min}}{M_{\max} + M_{\min}}$	M define the Moments (see Brown-Eccentricity)
Spreadness	Boesch (1993)	Spreadness	$spd = \frac{M_{\max} - M_{\min}}{(Area)^2}$	M define the Moments (see Brown-Eccentricity)
Circularity	Davis (1986)	Elongation	$c_1 = \sqrt{\frac{lw}{l^2}}$ $c_3 = \frac{4A}{lp}$ $c_5 = \sqrt{\frac{D_i}{D_c}}$ $c_2 = \frac{4A}{p^2}$ $c_4 = \sqrt{\frac{A}{A_c}}$	l: Length of long axis w: Width of object perpendicular to long axis A: Area of object p: Perimeter of object Ac: Area of smallest enclosing circle Di: Diameter of largest inscribed circle Dc: Diameter of smallest enclosing circle
Ellipticity	Davis (1986)	Elongation	$E = \frac{w}{l}$	l: Length of long axis w: Width of object perpendicular to long axis
Bending Energy	Young, I.T. et al. (1974): An analysis technique for biological shape I; Information and Control 25.	Compactness	$E = \sum_{k=-\infty}^{\infty} (kw_0)^2 \left(\lvert X_k \rvert^2 + \lvert Y_k \rvert^2 \right)$	$X_k = (X_k, Y_k)$ are the Fourier descriptor coefficients (see Ballard and Brown, 1982) for further details).
Miller's Measure	Campbell (1993)	Compactness	$C = \frac{Area \ of \ polygon}{Area \ of \ circle \ with \ same \ perimeter}$	
Boyce-Clark radial shape indexy	Campbell (1993)	Compactness	$C = \sum_{i=1}^{n} \left\lvert \left(r_i \cdot 100 \Big/ \sum_{k=1}^{n} r_k \right) - \left(100 \big/ n \right) \right\rvert$	The indices is based on the length of radials extending outward from a node at the center of the shape. A set of equally spaced radials (n radials) is then drawn outward from the center to the perimeter of the shape.
Compactness measure	Davis (1986)	Compactness	$K_1 = \frac{2\sqrt{\pi A}}{p}$ $K_2 = \frac{p^2}{4\pi A}$	p: Perimeter of object A: Area of object