MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

KNOWLEDGE-BASED INTERPRETATION OF ROAD MAPS BASED ON SYMMETRICAL SKELETONS

Markus Ilg and Robert Ogniewicz
IKT/Image Science Laboratory

ETH-Zentrum, CH-8092 Zürich

Abstract

The objective of our research reported here is to show the various processing steps on all levels of a computer vision system required to interpret and describe road maps. First, the foundation for the recognition process is laid by computing two sets of distance skeletons. They are termed endo- and exo-skeleton and are symmetrical with respect to the boundary between foreground and background of a binary raster image. The skeletons are derived from a novel implementation of Blum's concept of the MAT in the semi-continuum which preserves important features such as Euclidean metric and correct topology. To eliminate noise and quantization effects a new regularization method has been devised based on which the MAT is pruned to its stable inner branches. After the removal of artefacts and further simplification of the skeleton the recognition of meaningful complex structures is accomplished in several steps with the aid of a growing amount of domain-specific knowledge. Scene knowledge helps to verify (local) hypotheses in a larger context and to arrive at a consistent interpretation throughout the scene. The selection of the object-oriented programming paradigm proved to be very effective.

1 Introduction

Document image analysis has become a very attractive and active research area. Its goal is the automated interpretation and description of the visual information contained on a 2D medium (mostly paper) comprising any mixture of texts, graphics, and pictures. The merits are obvious considering the ease with which the information can be manipulated once the information is in digital form. There exists a considerable amount of publications in the area of document image analysis. For a representative selection of papers on theory and applications consult [1] and also [3]. In this paper we concentrate on the subarea of 'graphics recognition'. A good overview is given in [2].

Although the understanding of documents is an important topic per se and has immediate practical applications there is also another motivation behind this work. It is considered to be a rather excellent training step toward the even more challenging task found in general machine vision and scene analysis [6]. All known problems in computer vision are addressed in a nontrivial manner while allowing to concentrate on the higher (AI) levels.

We analyze first typical situations from a single thematic layer of a road map and try later to expand on the insight gained so far to tackle other line images as well.

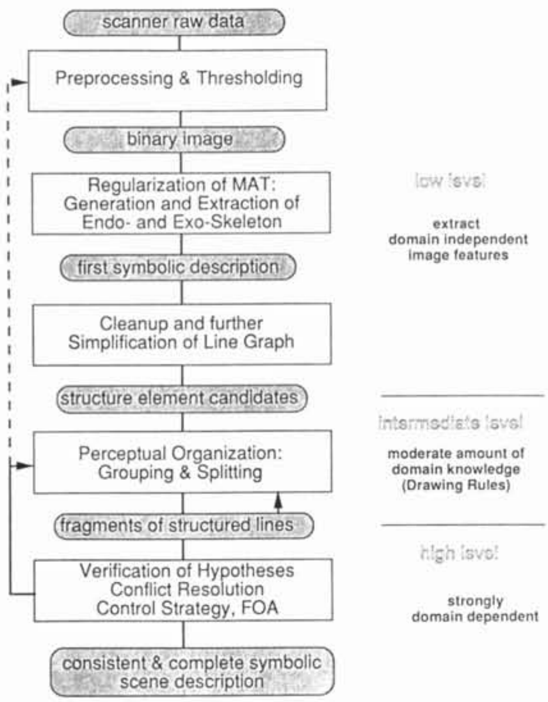

Figure 1: Architecture of the MAPS System

The selected layer contains essentially road information. It is a line drawing (i.e., 1D lines embedded in a 2D space) consisting of linearly and/or laterally structured curved lines forming a network. Each line style corresponds to a different road type (see Fig. 4a).

2 System Overview

The overview in Fig. 1 shows the essential phases of the understanding process. From the scanned (iconic) data a complete and compact symbolic description of the geometry of the image entities as well as their type and relationships is computed.

The corresponding levels of abstraction are depicted in Fig. 2 and are described in detail in [6]. With these levels hypotheses can be associated about candidate interpretations for entities on this level. The partial ordering implied by the 'part-of' links indicates the allowed transitions up and down the hierarchy: Bottom-up (data-driven) by grouping entities from one level and post a hypothesis on the next higher level, or top-down (expectation-driven) by placing a hypothesis on a lower level about missing parts of a given entity.

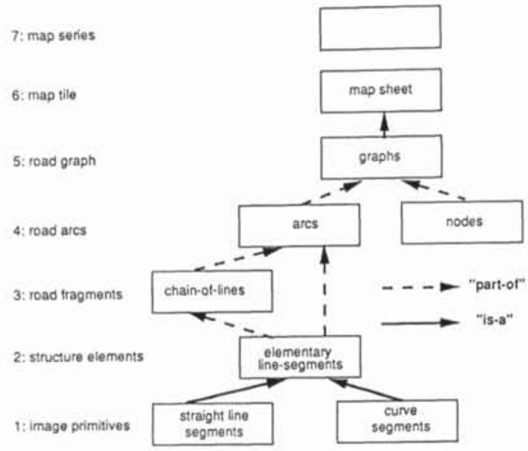

Figure 2: Hierarchical Levels of Abstraction

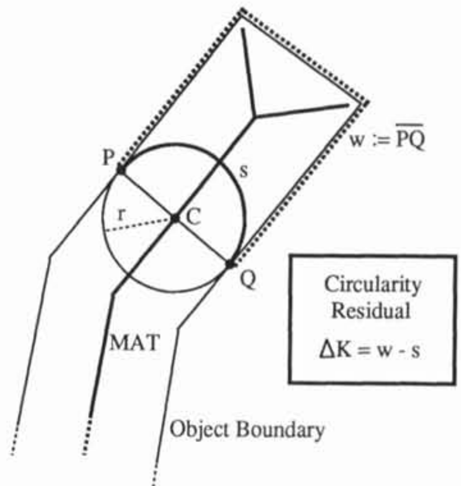

Figure 3: Regularization

3 Regularizing the MAT: Generation of Endo- and Exo-Skeleton

The majority of the roads are defined by parallel lines or dually by their CHaracteristic Elongated Enclosed Regions ('cheers') in background[1]. The middle axes (skeletons) of those cheers lay the foundation for a description of the course of laterally structured lines. For solid lines or linearly structured lines (e.g., dashed lines) the skeletons of the lines themselves form the middle axes. As a consequence, it is necessary to extract the skeletons of both foreground (lines) and background components as will be shown next.

Traditionally, the representation of a (2D) object is strictly related to the underlying tesselation of the discrete plane, usually a square grid with integer coordinate values. Needless to say, most of the published skeletonization methods work on such a grid [7,10]. This approach introduces, however, a lot of impediments: The computation of the correct Euclidean distances requires relatively complicated or highly parallel algorithms which in turn significantly slow down the execution speed; Thinning algorithms often produce counter-intuitive results and tend to overemphasize the influence of artefacts at object boundaries. The latter observation leads to a central problem of the skeletonization: Given a (not necessarily Euclidean) skeleton, how can we distinguish between stable parts that are insensitive to noise and affine transformations and parts that were generated rather incidentally and should be removed prior to further processing?

In [4] a new regularization method has been presented to overcome just this problem. A simple relationship between the needed degree of pruning and the estimated amount of disturbances along the boundary points is given as well. A brief review of our regularization method follows.

Instead of computing the distance transformation on the regular grid, we calculate the *Voronoi Diagram* (VD) [9] of selected boundary elements (Fig. 4 b). The boundaries of the convex Voronoi polygons describe the zone of influence for each boundary element and represent the incarnation of Blum's concept of the MAT [8] in the semi-continuum. The VD is represented as an abstract data structure known as DCEL (doubly connected edge list) attributed with the exact proximity information. To obtain a stability measure for an arbitrary point of the VD, we assess its influence on the reconstruction of the original shape from the attributed VD (see Fig. 3).

According to the definition of the VD, each maximally filling disk (radius r) with respect to the object is centered at one point of the VD and shares at least two points P and Q with the object boundary. The so-called circularity residual ΔK is the difference between the length of the boundary from P to Q and the length of the circular arc between P and Q. A point of the VD is considered stable if ΔK surpasses a certain threshold value, usually 2.0. The resulting reduced VD is henceforth called *Voronoi Skeleton* (Fig. 4 c–e). It has been pointed out in [4], that ΔK is a function which grows monotonically as the center C of the disk moves along an edge toward the interior of the object. The topology of the Voronoi skeleton will thus be preserved for any threshold and multiple pruning passes.

Another important aspect of the Voronoi skeleton shows up if the boundaries are interpreted as truly 1D edges (*cracks*) rather than pixel strings. This allows a symmetrical handling of foreground and background. Thus endo-skeletons and exo-skeletons with Euclidean metric and correct topology can be computed in a single scan over the image.

4 Grouping Activity

As a result of the skeleton extraction explicit information is available about local image features (connectivity and geometry) and about their local relationships which are likely to be usable in the subsequent grouping process. The primitive symbolic descriptions are domain-independent and define each a line graph attributed with the distance and references to the neighboring elements. After the removal of artefacts the structure is further simplified in two steps by eliminating branches that obviously do not correspond to a 'cheer': (1) Any line arc whose variance of the radii along its course exceeds a preset threshold is removed, since this indicates that the bounding lines of the corresponding 'cheer' are not sufficiently parallel; (2) Any line arc having a point whose distance value is greater than half of the expected maximal width of any 'cheer' is discarded.

Generic grouping knowledge (e.g., similarity, proximity, and good continuity) is supplemented by further knowledge to account for the specific contents of the image under consideration. The latter is obtained by inferring the draw-

[1] similar to the concept of a 'ribbon'

ing rules from the map. The cleaned primitives are then grouped accordingly in several steps [5] to constructs on always higher (more abstract) levels (levels 3 and 4 in Fig. 2).

5 Attaching Semantics to Structures

Processing at this level requires a model of the expected scene. It must, however, not be about a particular map scene since it would not be feasible to store such a model for each map sheet. It should rather contain generic knowledge about the inherent structure of map scenes, about map entities along with their relationships and about layout rules. Other kinds of knowledge sources used are: Control knowledge to allow focusing on a particular area of the scene to direct the processing to the locations that are most promising at some point in time; and strategy knowledge which directs the execution of the most appropriate control strategy.

The understanding process can be seen as a mapping from the hypothesized data (data panel) onto the road model (model panel) such that a consistent interpretation throughout the map section can be obtained. The difficulty with the model of the expected scene is that it is rather vague.

Some of the recognition steps are exemplified in the image sequence in Fig.4 f–i. Further details can be taken from [5].

6 Extensions to Other Line Images

Lines are important entities with respect to the analysis of a wide range of scenes. While in images of maps, plans, etc., the information is inherently contained in lines there are others, namely gray-valued images from which lines bearing relevant information can be derived by applying suitable edge detectors. In the latter case the output most likely consists of contour fragments attributed with a measure of their evidence. Since the regularization method does not demand closed contours our future research will also focus on the definition of skeletons in gray-valued images. Furthermore, the tesselation defined by the underlying VD allows to direct the grouping into hierarchies of the line fragments, where successive steps analyze increasingly larger areas of interest.

Extensions have been made to include the recognition of not just other plans but also medical images (see [6]).

7 Implementation

Low-level procedures have been coded in C while code pertaining to higher levels has been written in Common Lisp and New Flavors. Scene knowledge which is not so clear-cut and is subject to frequent changes/additions during the development phase is put into a rule-based expert system shell. This allows for more flexibility and less side effects if changes become necessary. Since the data representations had to be general across different applications but also extendable to new applications the choice to using object-oriented programming proved to be very effective.

Extensive use is made of the inheritance capability to build a set of generic operators which support the computation of the properties of point and line features and their respective relationships.

To better manage the large amount of data spatial and feature indexing has been introduced.

Some figures in Table 1 illustrate the practicality of our approach, especially for the low-level part. The higher levels take on the order of ten minutes.

Machine: DecStation 3100		
Operation	Elapsed Time in secs	Size in MBytes
Crack Extraction	1.8	0.6
VD Computation	11.2	3.7
Regularization	6.7	4.3

Table 1: Time and Space Requirements of a 512x512 Image

8 Summary and Conclusion

We have shown that symmetrical skeletons lend themselves readily for subsequent recognition processes for the following reasons. The endo-skeleton can be used to extract the lineal features and classify them into classes of different width. The exo-skeleton on the other hand retains the spatial relationships between the foreground entities and lays the foundation for the grouping activity. The introduction of a new regularization method allows to successfully cope with noise and other artefacts and to generate robust Voronoi skeletons. Their computation is very efficient and fast. The multi-step grouping procedures exploit the neighborhood information contained in the exo-skeleton as well as knowledge inferred from drawing rules. Finally, global consistency is achieved by reassessing (local) interpretation results in a larger context.

To exemplify the generic character of the procedures developed the application area has been extended to edges extracted from gray-valued images.

Using object-oriented programming techniques proved to be very advantageous for the required genericness and extendability of the code.

References

[1] *Pre-Proc. of the IAPR Workshop on Syntactic and Structural Pattern Recognition SSPR 90*, New Jersey, June 1990.

[2] Kasturi R., R. Raman, Ch. Chennubhotla, and L. O'Gorman, *Document Image Analysis: An Overview of Techniques for Graphics Recognition*, in: pre-proceedings SSPR 90, pp. 192–230, 1990.

[3] *IEEE Computer*, Vol. 22, No. 12, 1989.

[4] Ogniewicz, R., and M. Ilg, Skeletons with Euclidean Metric and Correct Topology and Their Application in Object Recognition and Document Analysis, *Proceedings 4th Int'l Symposium on Spatial Data Handling*, Zurich, 1990.

[5] Ilg, M., Knowledge-Based Interpretation of Road Maps, *Proceedings 4th Int'l Symposium on Spatial Data Handling*, Zurich, 1990.

[6] Ilg, M., Knowledge-Based Understanding of Road Maps and other Line Images *Proceedings 10th ICPR 90*, Atlantic City, pp. 282–284, 1990.

[7] Arcelli C., Sanniti di Baja B.. A width-independent fast thinning algorithm, *IEEE PAMI-7/4*, pp. 463–474, 1985.

[8] Blum H., A Transformation for Extracting New Descriptors of Shape, from: *Models for the Perception of Speech and Visual Form*, W. Wathen-Dunn, Ed., MIT Press Cambridge MA, 1967.

[9] Preparata F.P., Shamos M.I., Computational Geometry, *Texts and Monographs in Computer Science*, Springer Verlag, 1986.

[10] Xia Y., Skeletonization Via the Realization of the Fire Front's Propagation and Extinction in Digital Binary Shapes, *IEEE PAMI-11/10*, pp. 1076–1086, 1989.

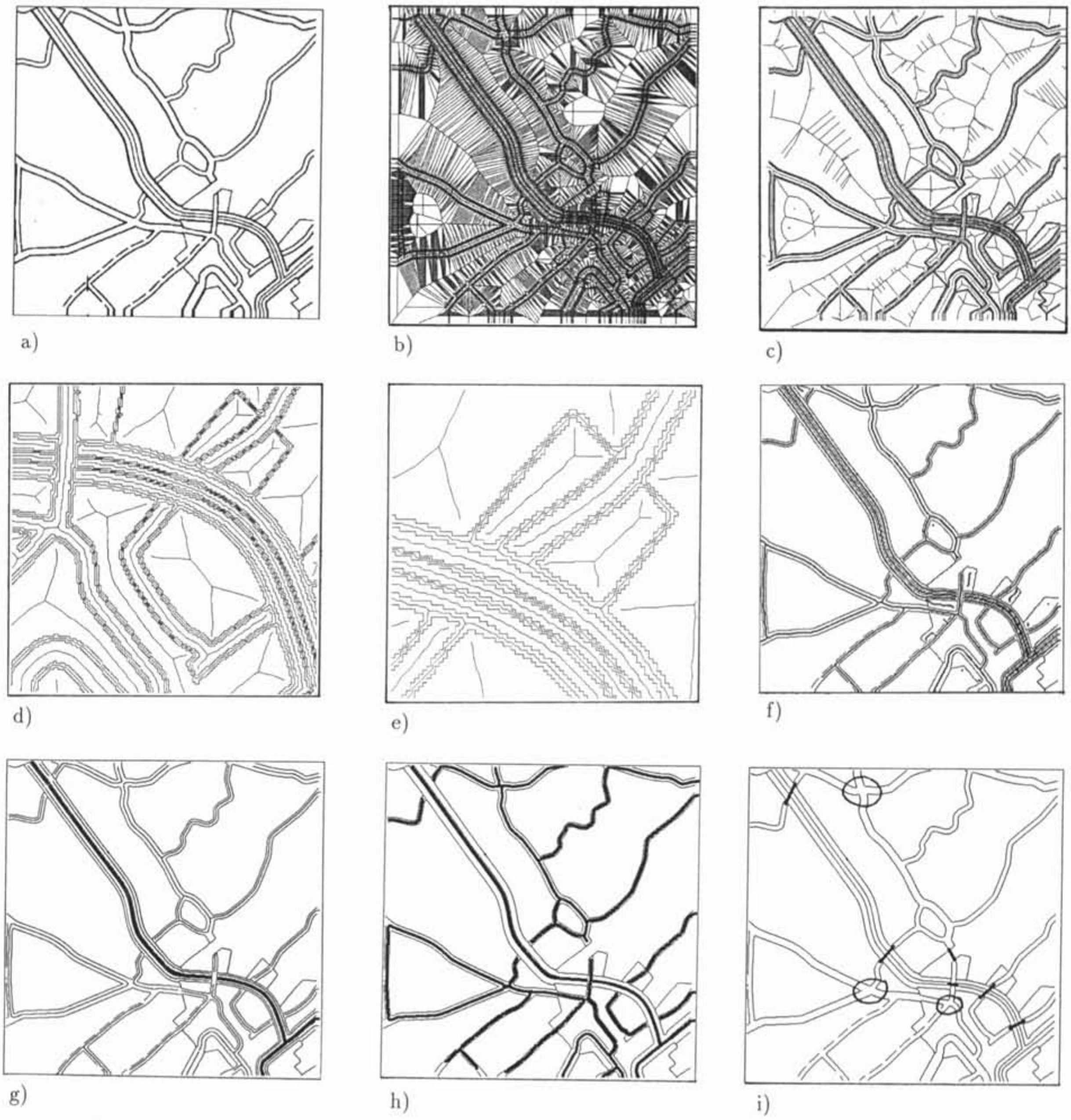

Figure 4: Different Processing Stages on a Section of a Road Map

a) Binary raster image of original, b) Voronoi Diagram VD on raster cracks of original, c) Voronoi skeleton (i.e., VD regularized with threshold T = 2), d)–e) Enlarged portions of c) show endo- and exo-skeleton of the dark (line) components in the original, f) Exo-skeleton further simplified to the branches that represent most likely a 'cheer' g) Road axes of fragments resulting from the understanding process. The axis of the highway (quadruple parallel line) is represented by a bold line. Note that the interpretation is incomplete, h) Extension of road axes into road junctions, i) Bridging gaps stemming from overpasses by means of 'good continuity' and detection of virtual 4-road-junctions (encircled).