

AUTOMATIC DIGITIZATION OF CONVENTIONAL MAPS
AND CARTOGRAPHIC PATTERN RECOGNITION

Werner Lichtner
Institute of Cartography (IfK)
University of Hannover
FRG
Commission IV

Abstract:

One important task of the establishment of a Geo-Information System (GIS) is or can be the digitization of existing maps. Digitization of conventional maps is a costly and labour-intensive task. Generally it must be completed before other processes can commence. That's the reason for the market-wide interest in automatic digitization techniques of conventional maps.

Under the leadership of the author scientists at the Institute of Cartography at Hannover University are engaged in automatic digitization of maps, especially urban basic maps and cadastral maps.

After scanning an original map sheet data processing starts with the binarization and raster-to-vector conversion which generated unstructured vector data. This process is realized by a successfully tested program RAVEL. Some characteristics (e.g. the line width, line length, perimeter of a node pattern, heading of line segments) which are helpful for pattern recognition procedures are extracted and stored. This program is also used for data reduction in a lot of other applications (e.g. printing industry, storage of construction drawings). The main task of the research group is the development of a software system for structuring the data by means of pattern recognition. Classification is performed hierarchically. It starts with basic map features like symbols and letters. Features has to be separated by a so-called segmentation process before they can be supplied to a cartographic recognition system.

The described software system is based on the strategy of so-called knowledge based pattern recognition. The author explains the fundamental principles and the stage of development. Finally he evaluates the automatic digitization from an economic point of view in comparision to alternative digitization methods.

1. Introduction

For the representation of the geometry of features in a Geo-Information System (GIS) vector data are preferred generally. Vector data can be calculated by original measurement elements. Many users of GIS decided to use this procedure because they don't have maps of necessary resolution or existing map editions are not of the expected or wanted quality.

In many european contries and urban districts we have a very dense series of official topographic and thematic maps since a lot of decades. Depending from the necessary resolution of the models in a GIS and the actuality of the map content these

maps can be used as a source for data acquisition. The task is now to digitize the map content (see Fig. 1).

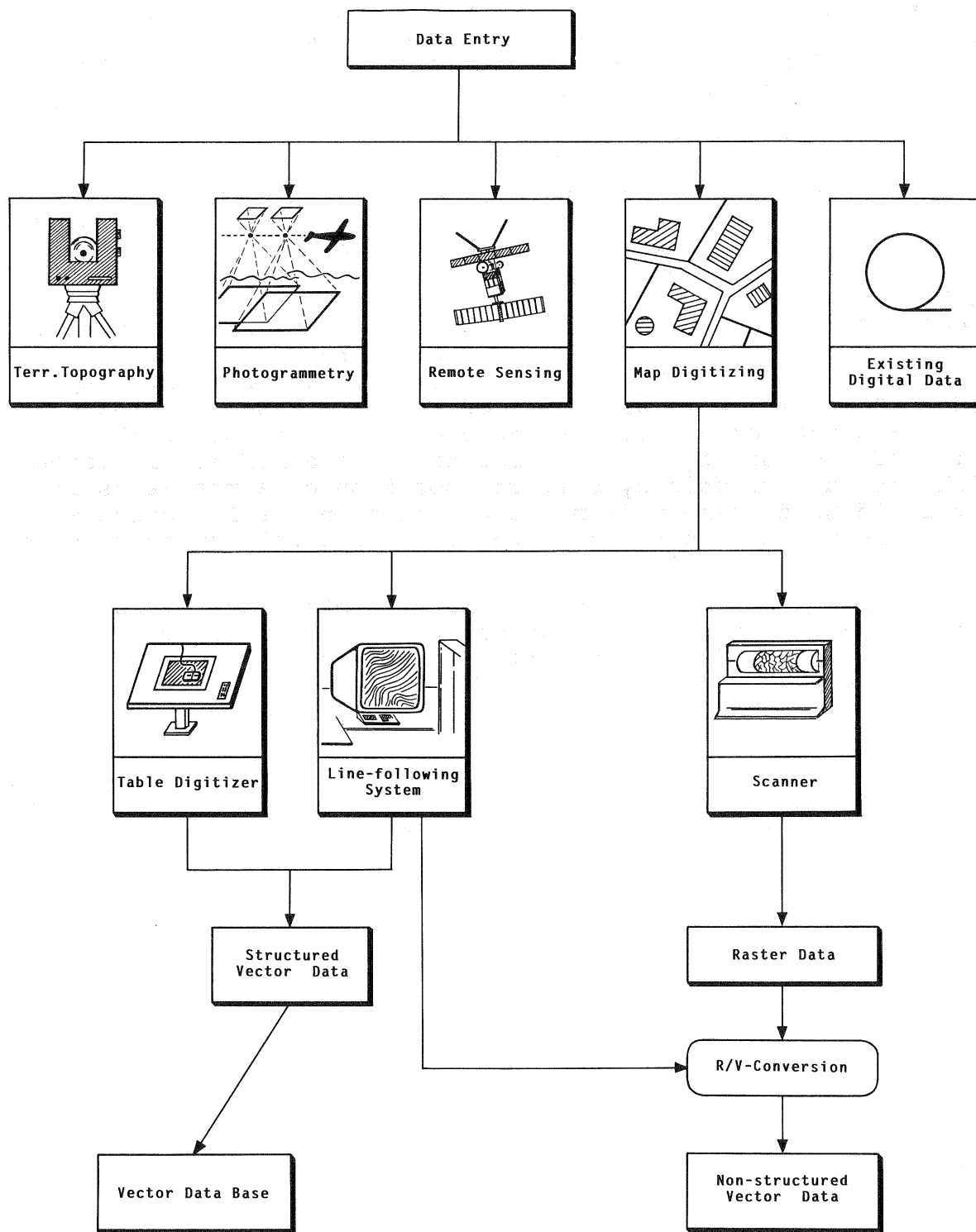


Fig. 1: Alternative sources for digital data and alternative methods for map digitization.

Today there do exist three alternative methods for map digitization (Fig.1). Table digitizers are mostly in use. Other possibilities offer line-following systems (semi-automatic digitizing) and scanners. In contrast to table digitizers and line-followers, scanners produce raster as opposed to vector data (Fig. 1). Because of the feature-related data structure of many GIS the raster data have to be converted on structured vector data. This necessary transformation and structuring is a very complex and difficult process. Till today this task is the objective of research and development.

2. Geometry Models in a GIS.

Geometric information can be represented in a GIS in two different models (see Fig.2). There are the so-called Digital Landscape Models (DLM) in which are the visible and non visible objects of the surface of the earth in their area relation in quality and quantity. They consist of the Digital Planimetric Model (DPM), the Digital Elevation Model (DEM) and other possible Digital Thematic Models. For cartographic representation the information of a DLM is prepared in a Digital Cartographic Model (DCM), which is also part of the data base. A certain "road"-object for example is represented in the DLM with its axis and its quality characteristics. In the DCM it is stored with its cartographic symbol. The position of a road-axis in DCM could be different from its position in the DLM because of the influence of a certain cartographic generalization process (i.g. displacement).

In the case of automatic structuring of vector data by cartographic pattern recognition procedures this connectivity has to be investigated very detailed in advance.

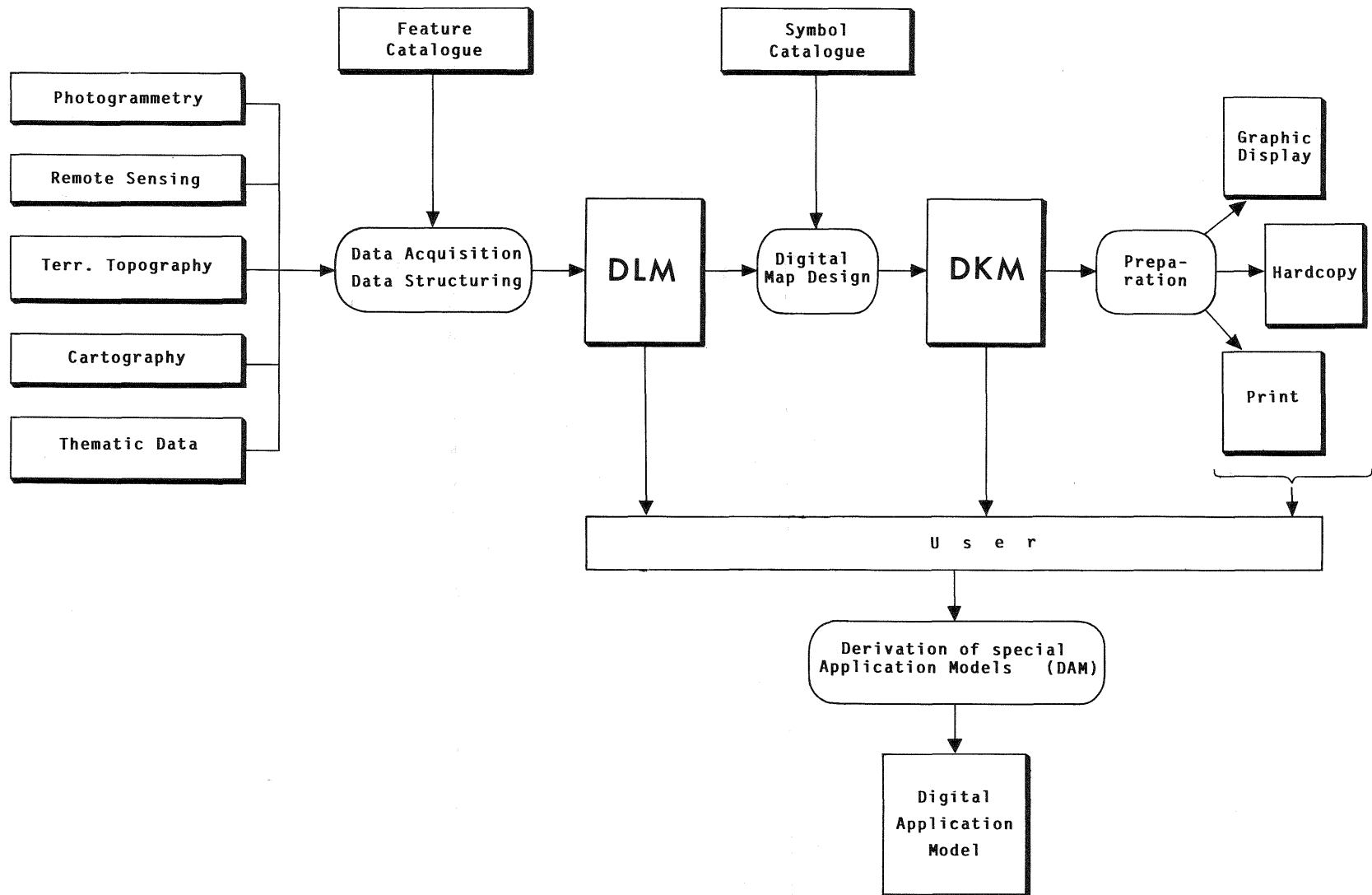


Fig. 2: The establishment of a topographic and cartographic data base in a GIS.

3. Knowledge-based Cartographic Pattern Recognition Activities of IfK.

In the Institute of Cartography at Hannover University a group of scientists is working under the leadership of the author on the field of cartographic pattern recognition. These specialists are engaged with the development of a so-called knowledge-based cartographic pattern recognition system for line maps like cadastral maps or city maps of different thematic topics.

The principle of such a system is described in Fig. 3 graphically.

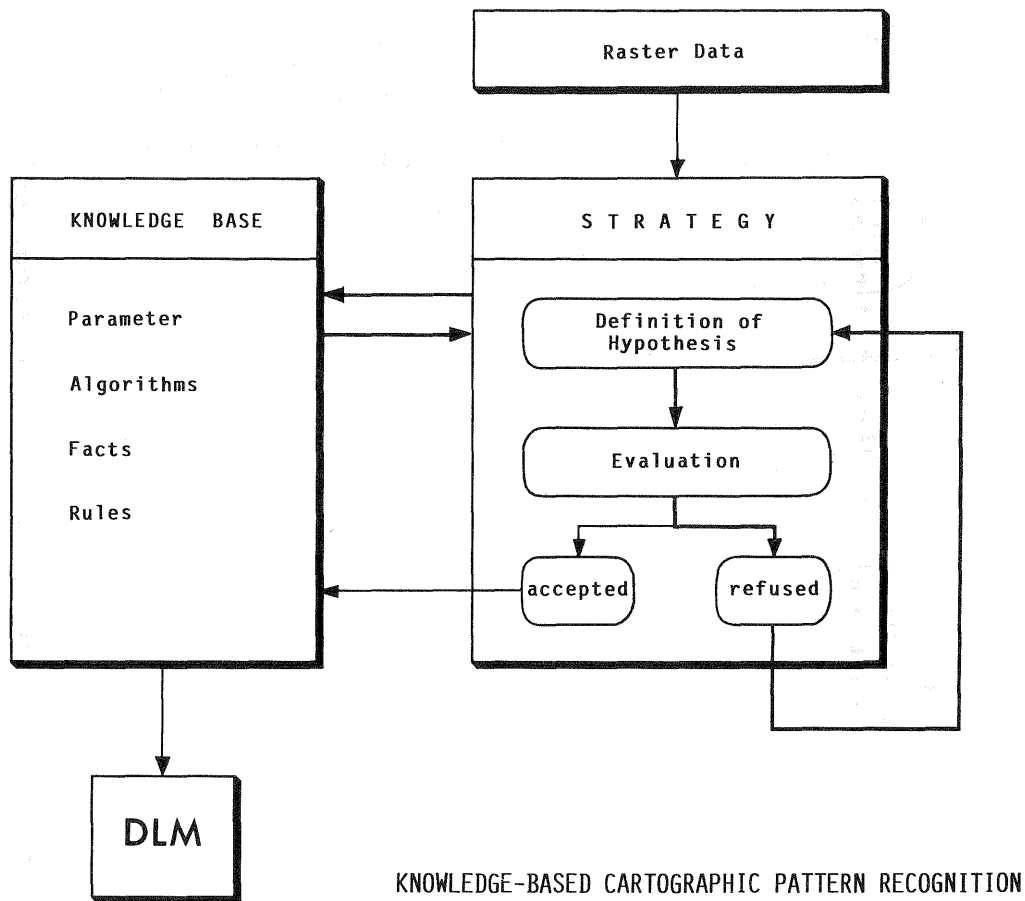


Fig. 3: The principal procedure of a Knowledge-based Cartographic Pattern Recognition System.

The knowlegde is represented in the knowledge base in form of

Parameters,
Algorithms,
Rules and
Facts.

The set of tools represented in form of algorithms is very complex. There do exist a very effectful Raster/Vector Conversion and most of the numerical and syntactical classification algorithms well-known from the literature.

The system starts with raster data as a result of a scanning process. After a certain preprocessing the data are reduced to binary raster data. These data are the input for the raster-vector conversion realized in the modul RAVEL /LICHTNER, 1987/. Fig. 4 represents the principal functions of RAVEL.

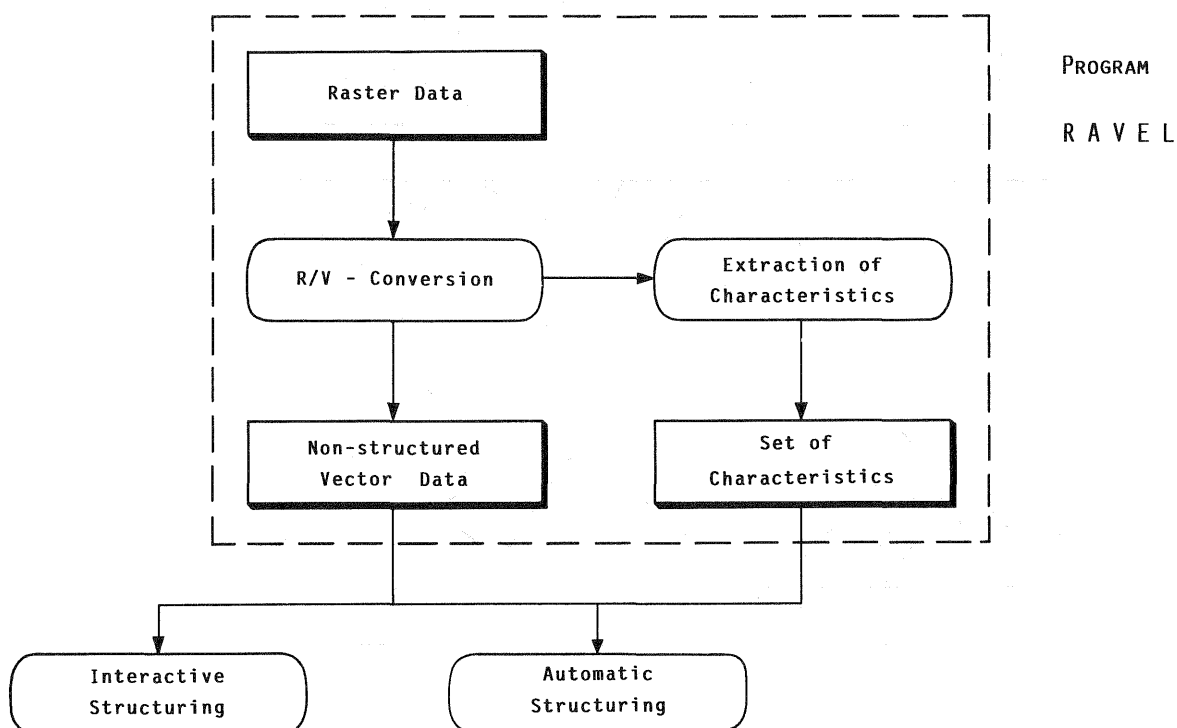


Fig. 4: Principal functions and results of the Raster-Vector-Conversion modul RAVEL.

The result of RAVEL are non-structured vector data and a set of characteristics which can be extracted during the conversion process. Typical characteristics are the line width, the length of line elements, nodes, perimeter of a node pattern, heading chain of a line element. The modul RAVEL is described in detail in /LICHTNER, 1987/.

Different strategies are now in test to get experiences with certain algorithms and rules. Each strategy starts with the definition of a hypothesis. Then the hypothesis is evaluated by certain cartographic criterions, rules and parameters. After that a decision is made to accept or to refuse the hypothesis. It's the objective of the research and development to recognize about 95-98% of cartographic features and context correct. More would be -from our point of view- an illusion.

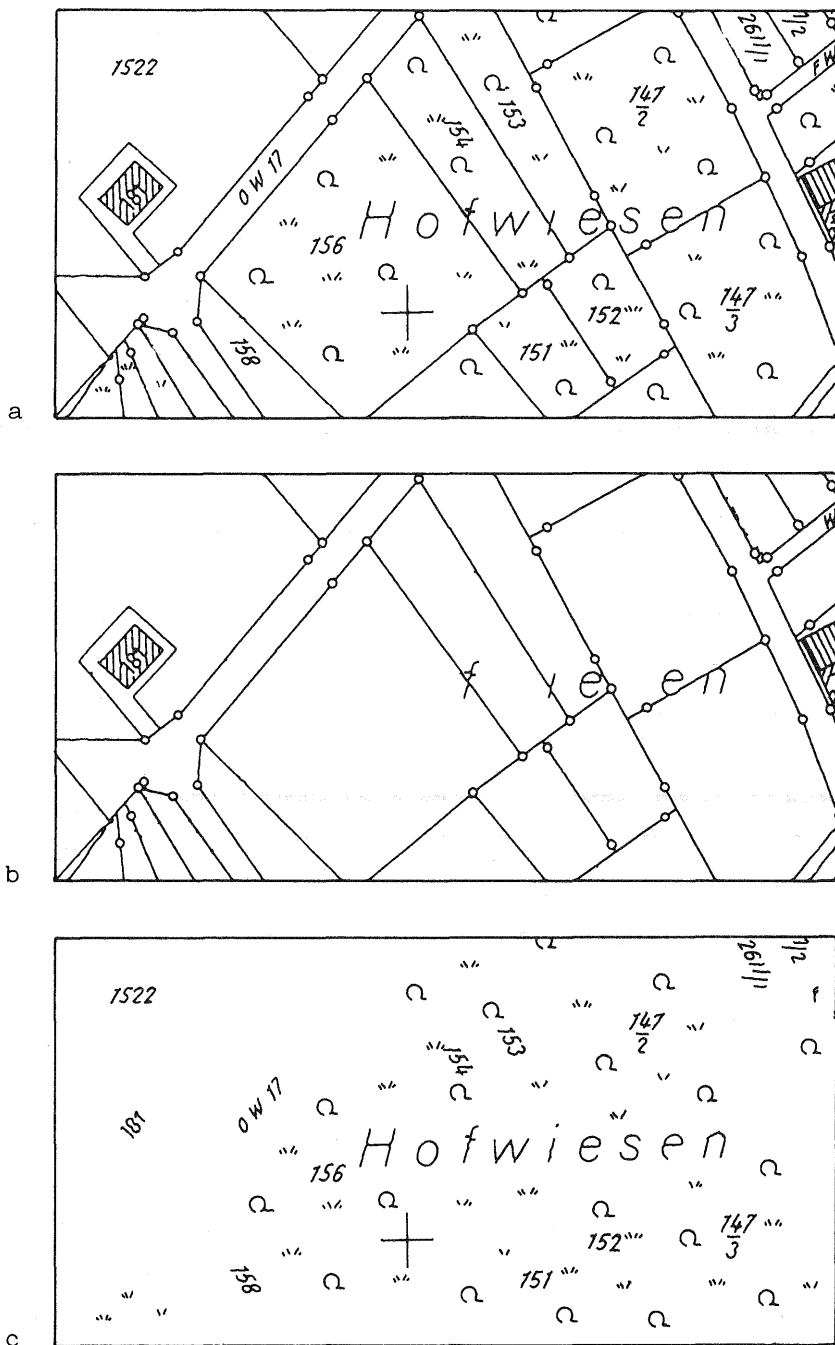


Fig. 5: a) Set of non-structured vectors,
 b) Networks with intersected texts,
 c) Final set of separated symbols and texts.

It is not easy or quite impossible to demonstrate a certain stage of the system development in a paper, especially the recognition of a cartographic symbol, a number or a letter. Therefore the author avoid an attempt to try this and restrict himself to a printed figure (Fig.5) representing the separation of intersecting lines by means of line width, straightness and topology produced by two members of his pattern recognition research group /ILLERT-YANG, 1987/.

A further very important advantage of the investigations and research work on the field of cartographic pattern recognition is the fact that the scientists detect step by step cartographic rules of a new quality. These rules are not describing in a general form more intuitive processes like many of the cartographic standard literature do, but in a very concrete form small steps of cartographic processes and relations. They are practical usable for the development of computer-assisted cartographic software.

4. Brief evaluation of alternative digitizing methods.

If vector data are preferred generally three alternative digitizing methods can be used (Fig.1). There is no doubt that manual digitizing is the most universal and cheapest method for the mentioned purpose, but also the most time-consuming one. Line-follower-systems are not successful on the market. They are too expensive and can demonstrate real advantages only in the case of isoline-digitizing. Large format scanners are very expensive too and produce raster data. The original digitizing process is very fast, but the necessary post-processing extreme intensive.

Successful pattern recognition software could be or is very helpful. But scanning and automatic structuring of vector data cannot substitute the manual digitizing completely. Therefore a digitizing company interested in cartographic pattern recognition software has to invest money in both. The efficiency of pattern recognition software is in cartography profitable if

- a) the weight factor of time is very high or
- b) the costs of cartographic pattern recognition software are not more then the price of about one digitizing workstation.

These conditions are a realistic problem for the introduction of qualified cartographic pattern recognition software systems on the market today.

5. Conclusion.

The development of knowledge-based cartographic pattern recognition software is going on successfully. First finished developments for large-scaled line maps can be expected in 2-3 years.

6. References.

- ILLERT, A./
YANG, J. Extraction of Texts and Symbol from Unstruc-
 tured Vector Data.
 Nachrichten aus dem Karten u. Vermessungswesen,
 Reihe II, Heft Nr. 46, Frankfurt 1987
- LICHTNER, W.: RAVEL-Complex Software for Raster-Vector
 Conversion.
 Proceedings EURO-CARTO VI, Brno (CSSR), 1987.