Scan-Vectorisation, Classification and Pattern Recognition as a Means for Transforming Analogue Maps into GIS

Prof. Hans F. Kern
Fb Geoinformationswesen
Fachhochschule Karlsruhe
Moltkestr. 4
D 76133 Karlsruhe

Prof. Dr. Rolf Klauer
Fb Vermessungswesen und Kartographie
Fachhochschule München
Karlstr. 6
D 80333 München

Abstract

There is a need to transform existing analogue maps into digital form. Cadastral offices, energy and water suppliers, planning authorities have usually large amounts of mostly color separated quasi layer oriented "sheets" that must be incorporated into Geo Information Systems. By a combination of automatic and interactive steps this task can be most effectively solved.

The first step is to scan the sheets. Automatic filtering procedures then improve the scanned raster image without changing the topology of nodes, lines, areas, symbols and text. The automatic classification and pattern recognition has to be prepared by the programming of "production lines" that take into account the special symbology and design structure of the sheets under consideration. This classification and pattern recognition delivers a 50% to 90% correct result in dependence of the map type and the efforts invested into the formulation of the "production lines". Finally interactive manual work solves the problems that have not been handled automatically.

The results can be imported into GIS by a number of exchange formats.

There will be shown the general procedure as well as examples from cadastral maps, gas supply documentations, planning maps, water conservation area maps etc.

1 Introduction

Geographic Information Systems (GIS) are widely available and the Query Languages (SQL) for retrieving information by highly sophisticated questions are steadily improving. The acceptance of such systems is growing and of no problem any longer. The real bottleneck is to fill the GIS with information, geometric as well as alphanumeric. In traditional cartographic offices and enterprises there is a vast amount of colour separated, quasi layer oriented, analogue "sheets".
It is a challenging problem to convert these analogue information into a digital form. Whereas the raster oriented approach can be solved quite easily, the vectorisation of analogue maps is not yet an accepted and broadly used procedure.

2 Overview over the process

The transforming of analogue maps into digital GIS-information requires black and white drawings, i.e. transparent films. Printed multi colour maps are not yet usable as the software at present handles only binary images.

The first step is the scanning of the different colour separated films. Because in cartography the colour separated films are also a rough separation into object classes the classification process starts already with preclassified information. The scanning resolution is usually of minor importance, so a resolution of 500 dpi will suffice. When there are numerous small adjacent symbols it is advisable to have 800 dpi in order to reduce their glueing. The result of scanning are raster data in the form of comprised tiff-files.

The next task is filtering. The aim is to improve the raster data without changing the topology espacially of line drawings. Filtering suppresses tiny black areas in white surrounding or white areas in black surroundings (dust) or smoothes the contours of lines and areas.

The third step comprises the change of raster data to vector data. It is the so called rough vectorisation; rough vectorisation as there is nearly a one to one correspondence between raster and vector data. For its own these vector data are yet of little value, as they are to short and to crumpled and have no attributes.

So the next exercise - and that is the heart piece - is the classification and pattern recognition. The tool is OGS (Objektbasierte graphische Sprache - objectbased graphical language).

Of course we do not rwach a 100 percent solution; there are misclassifications and unrecognized parts. So it is neccessary to control the result by visual inspection and to correct the errors of the automated process.

The last step is the export of the data into a final GIS.

All together we have the following parts:

1. scanning
2. filtering
3. rough vectorisation
4 classification and pattern recognition
5 interactice inspection and correction
6. export into final GIS.

With the exception of point 5 all steps - even scanning - can be highly automated so that manual operation can be avoided. As with manual digitization the bottle-neck remains at the human interface, but with a highly increased effectivity.

3. Example for intermediate steps in the recognition process

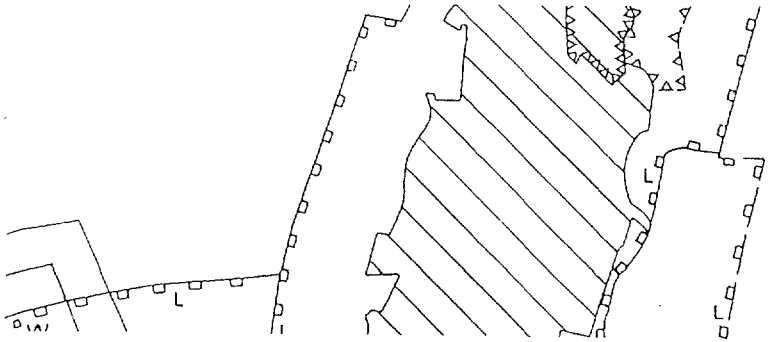

Figure 1: Original from a planning cadastre

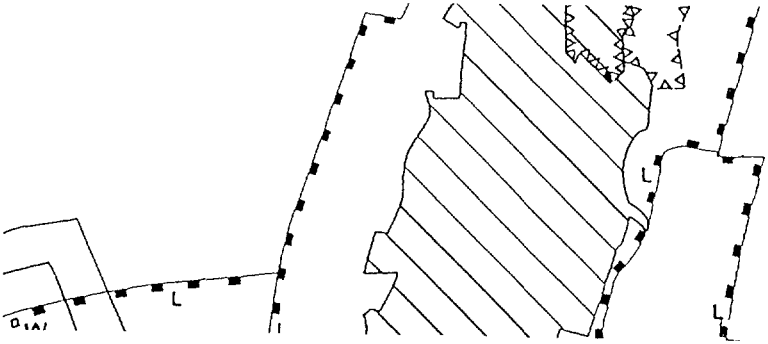

Figure 2: Classification of specific small areas

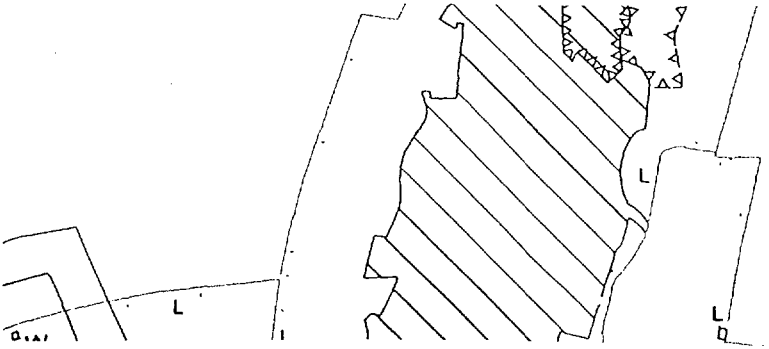

Figure 3: Lines with line following rectangles are recognized as solid and dashed

4. Example for line, character and symbol recognition

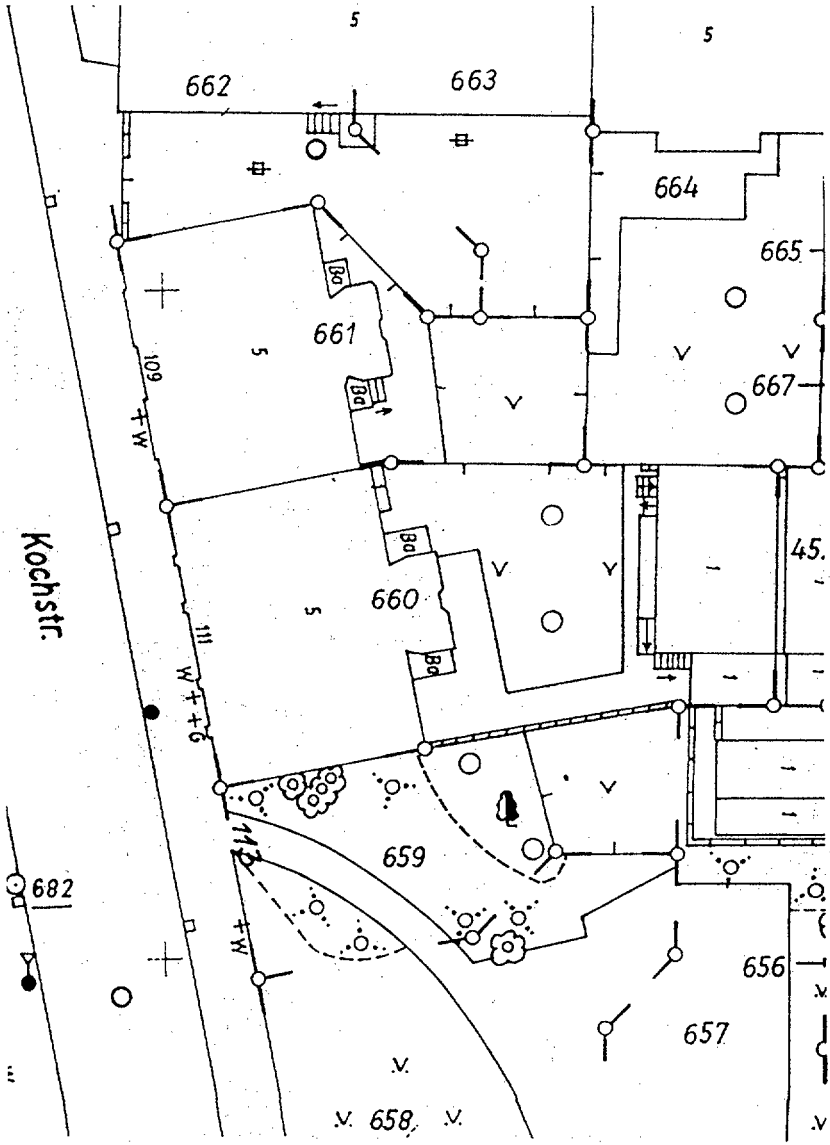

Figure 4: Original from a municipal cadastre

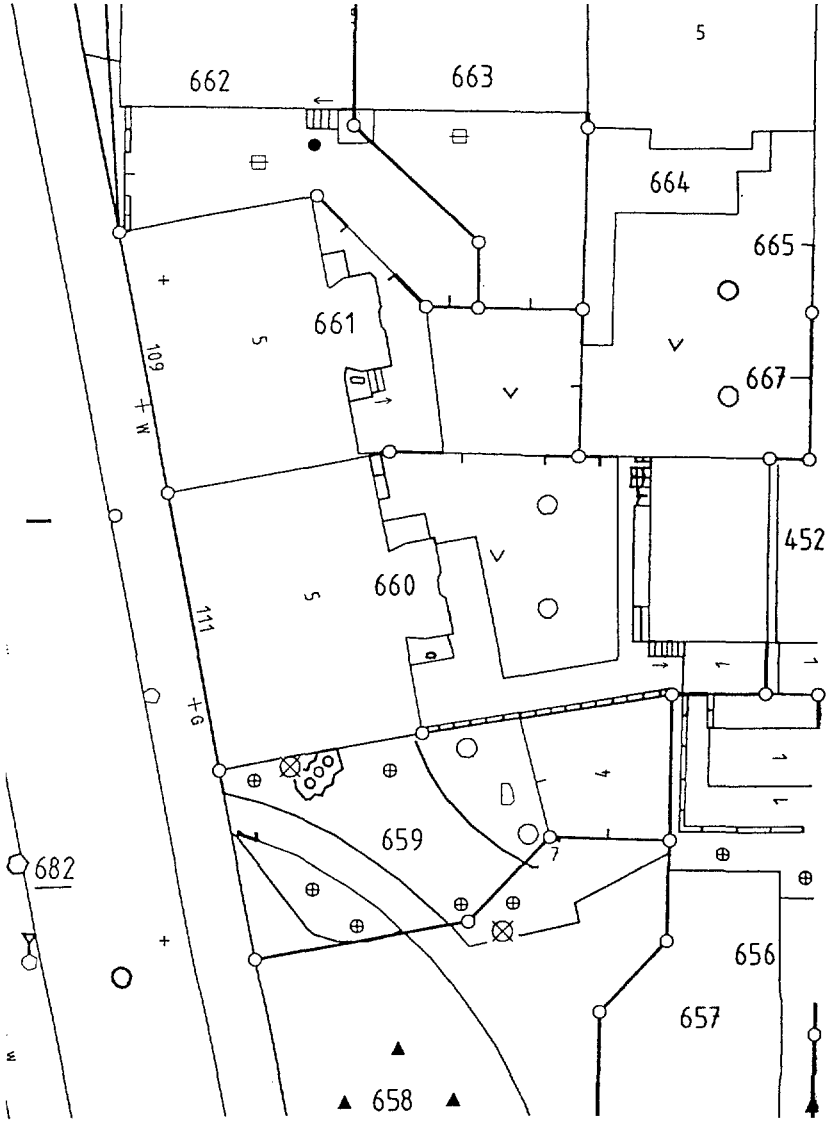

Figure 5: Boundary stones, fat lines between boundary stones, characters (numbers), arrows and other symbols recognized