

Digital processing of map data in raster format

Doctoral Thesis

Author(s):

Ilg, Markus

Publication date:

1986

Permanent link:

<https://doi.org/10.3929/ethz-a-000360935>

Rights / license:

In Copyright - Non-Commercial Use Permitted

DIGITAL PROCESSING OF MAP DATA IN RASTER FORMAT

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Technical Sciences

presented by

Markus Ilg

dipl.El.Ing.ETH

M.Sc. in Computer and Systems Engineering
Rensselaer Polytechnic Institute, Troy, New York

born May 13 1953

citizen of Salenstein (TG)

accepted on the recommendation of

Prof. Dr. D. Steiner, examiner

Prof. Dr. O. Kübler, co-examiner

ABSTRACT

To efficiently and automatically convert analog map information to computer-accessible digital form is still a problem. This work presents the design and the implementation of an overall new procedural sequence to treat the frequently occurring case of polygonal outline maps that show the boundaries of administrative and/or statistical (spatial) units. It is the geometric information contained in such maps with regard to its incorporation into a geographic information system we are concerned with here.

The methods described in the literature so far try to solve the problem by thinning the boundary lines in the scanned raw data to unit pixel width and by subsequent line following by means of a template. We use a different approach which, in a way, is the dual to the common solution: Instead of emphasizing the lines we consider primarily the areas enclosed by those lines (area-centered view). After having labeled all area pixels by a unique area code (color), the areas are being expanded uniformly and "simultaneously" on all sides. This is continued until neighboring areas touch each other. The boundaries between areas are then a sequence of edges between pixels of different color and are thus implicitly given. These sequences are extracted and archived in a compact topological data structure. The advantages of the method used are twofold: First, no tedious tests are needed to guarantee the correctness of the digitized line network and, second, the required size of templates for line extraction is 2 by 2 pixels only instead of the traditional 3 by 3. As a consequence, the space needed in the main memory becomes smaller while the algorithms are faster and also more stable. In addition, a representation of the spatial units of concern in the raster mode is obtained as a by-product. From the resulting compact data structure derived structures suitable for particular applications can be generated.

Methods well-known in picture processing and computer science are borrowed to solve the problem at hand as elegantly as possible. All algorithms adhere to a predetermined processing sequence, the so-called scan line sequence, which follows the way a TV frame is built up. This makes it possible to combine several logically separate algorithms into one more comprehensive algorithm.

Throughout the whole work careful consideration is given to the possibility of applying results from modern software engineering. In particular, a clear distinction is made between methodical and implementational concerns; the latter are treated in a separate chapter. The fact that we are concerned with implementational details demonstrates our goal to develop a system which can be readily used in a production environment.

KURZFASSUNG

Eine gesamthaft gesehene neue Verfahrenssequenz wird entworfen und implementiert, um damit einen häufigen Fall der digitalen Erfassung von existierenden Karten besonders rationell und weitgehend automatisiert abdecken zu können. Es handelt sich dabei um sogenannte Polygonkarten von administrativen und/oder statistischen (räumlichen) Einheiten, die als Linienzeichnungen vorhanden sind. Dabei interessiert vor allem die darin enthaltene geometrische Information, im Hinblick auf ihre mögliche Einfügung in ein geographisches Informationssystem.

Im Gegensatz zu den in der Literatur beschriebenen Methoden, bei denen in den abgetasteten Rohdaten die Linien auf die Dicke eines Bildpunktes verdünnt werden, um diese anschliessend mit geeigneten Schablonen zu verfolgen, wird hier ein gewissermassen dualer Ansatz verfolgt. Anstatt die begrenzenden Linien in den Vordergrund zu stellen, werden primär die von den Linien eingeschlossenen Gebiete betrachtet. Nachdem diese mit einem eindeutigen Code versehen worden sind, werden sie gleichmässig und "gleichzeitig" nach allen Seiten expandiert, bis sie mit benachbarten Gebieten in Berührung kommen. Die jetzt nur noch implizit vorhandenen Grenzen, d.h. die Ränder von Rasterzellen unterschiedlicher Farbe, werden extrahiert und in einer kompakten topologischen Datenstruktur im Sinne der Archivierung abgelegt. Der Vorteil dieser Methode ist einmal, dass keine aufwendigen Tests mehr notwendig sind, um den Linienzusammenhang zu gewährleisten, dann aber auch dass die benötigte Grösse der Schablonen zur Linienextraktion nur ein 2x2-Fenster im Gegensatz zu einem 3x3-Fenster für die traditionelle Methode ist. Dies hat zur Folge, dass der benötigte Platz im Hauptspeicher minimiert werden kann und somit schnellere und überdies noch stabilere Algorithmen ermög-

licht werden. Zusätzlich entsteht quasi als Nebenprodukt auch eine Darstellung der räumlichen Einheiten im Rasterformat. Ausgehend von dieser kompakten Datenstruktur lassen sich die für beliebige Anwendungen am besten geeigneten Datenstrukturen ableiten.

Bekannte Methoden aus Bildverarbeitung und Informatik werden angewendet und erweitert, um die vorliegende Fragestellung möglichst sinnvoll und elegant zu lösen. Alle Algorithmen arbeiten in einer zum voraus bestimmten Ablaufsequenz, der sogenannten Scanline-Sequenz (vgl. Aufbau eines Fernsehbilds). Diese erlaubt es, mehrere logisch separate Algorithmen zu einem umfassenderen Algorithmus zu vereinigen.

Während der ganzen Arbeit wird grosser Wert darauf gelegt, dass Erkenntnisse aus dem modernen Software-Engineering eingesetzt werden. Insbesondere werden implementationstechnische sauber von methodischen Belangen getrennt und separat behandelt. Mit der Berücksichtigung der Frage der Implementation wird angedeutet, dass ein System entwickelt werden soll, das sich für den produktionsmässigen Einsatz eignet.