EnviroInfo 2008 (Lüneburg)
Environmental Informatics and Industrial Ecology
 ISBN: 978-3-8322-7313-2

Full Automatic Calculation of Settlement Structure on Base of Topographic Raster Maps

Gotthard Meinel[1]

Abstract

The paper describes a method for full automatic calculation of settlement structure on base of simple topographic raster maps. In a first step extracted all buildings, they are mixed with streets, scripts and signatures in the maps. After vectorization we calculated building variables such as area, length, width, shape complexity and distance to next building. On base of this description we classified all buildings and realized a statistical analysis on building block level. We estimated 16 indicators (e. g. building and inhabitant density), visualized indicators in a GIS in optimized pre-defined legends and calculated a statistical report for the study area in different spatial resolution. The full automatic procedures called SettlementAnalyzer (SEMENTA®) have been implemented in ArcGIS under additional using of the image processing software HALCON.

Keywords

Spatial data infrastructures (SDIs), data mining, spatial data mining

1. Introduction

County, regional, spatial and expert planners require finely detailed information to the building stock, type of use and occupancy rates of settlement areas. These information become even more important when planning the provision of public services under conditions of urban contraction, especially in the context of settlement structures which are less resource intensive. On the other hand planning requires also demographic and socio-economic statistical data regarding economic conditions and the environment.

On the lowest level of administrative divisions in Germany are communes. However, this spatial unit is too inexact to answer many questions concerning settlement structures and their development. Therefore a discussion is presented in Germany about introducing a regional statistic e. g. on the basis of geographic grid units. Such products are available for Denmark, Finland, Switzerland and Austria [Wonka, 2006]. But a regional statistics will not be available in Germany before 2013, following the planned population census in year 2011 [Eschwege et al. 2006, Szibalski 2006].

The idea of the present paper is the using of geodata in form of simplest (binary) topographic raster maps, making the geodata intelligent by building extraction, surveying and classification and using these data for spatial information in combination with statistical data to generate relevant local information. The topics data extraction from scanned topographic or thematic maps and dasymetric mapping and attributing census data to geoobjects have been a research area for a long time [Mennis 2003, Langford et al. 1994, Dharet al. 2006].

[1]Leibniz Institute of Ecological and Regional Development (IOER) , Weberplatz 1, 01217 Dresden, Germany, Email: g.meinel@ioer.de

2. Input Datasets and Workflow

The basic unit of information for settlement structure is the building-unit with its attributes of physical extension and special forms of use. Related datasets must include all significant buildings. However, minor generalisations are acceptable for the planned application, such as simplification of building form or merging of several buildings. In view of the guarantee of regular updating to preserve the future validity of datasets, the choice of data sources is restricted to geodata sets from national survey bureaus. In Germany, the digital geodata is offered by the Federal Agency for Cartography and Geodesy (BKG). To geodata sets are interesting for the processing: digital topographic map 1 : 25,000 and the ATKIS Basis DLM (Digital Basic Landscape Model of the Official Topographic-Cartographic Information System in scale 1:25,000).

The digital topographic map 1 : 25,000 (DTK25-V) is currently being used for the extraction of individual buildings. The building stock is fully portrayed by the DTK25-V with only few exceptions. Partially generalised buildings (minimum size, simplification of building profiles, merging of adjoining side buildings into the main building) and minor shifting of location (through expanded depiction of road networks) do not compromise the data's suitability. Apart from the representation of buildings, the black footprint layer (also called the settlement layer) of the DTK25-V also includes cartographic annotation and other point-like, linear or polygonal signatures such as borders, streets, vegetation signatures or power lines. These elements are merged in a binary raster layer and cannot be separated. This provides a challenge for the unambiguous selection of buildings.

There are three kinds of topographic maps in Germany: analogue version (TK25, all maps before 1990), digital version in old style (DTK25-V, scanned version of TK25, maps between 1990-2008) and digital version in new style (DTK25, maps since 2008). In TK25 and DTK25-V buildings are mixed with other ground information, in DTK25 the building layer is separate.

In addition to the cartographic representation of buildings, separate building blocks must be spatially distinguished. This allows the derived parameters from individual building-units to be aggregated to the next higher spatial layer. Normally, a building block is designated as a residential area clearly delimited by roadways. This conforms to the digital classification in ATKIS for the object group "built-up area". ATKIS specifies the main function of a block and differentiates between "residential area", "industrial and commercial area", "mixed-use area", as well as "area with particularly functional character".

3. Processing Workflow and Selection of Buildings from Raster Maps

Figure 1 shows the whole processing workflow. The heart of the procedure and as well first step taken is the extraction of all buildings (residential and non-residential houses) from topographical raster maps in scale 1 : 25,000 [Meinel et al. 2006]. In the settlement layer of the map, buildings are stored together with transport and boundary lines, vegetation and other signatures in a binary layer.

The selection of buildings is a fully-automated multi-stage process using digital image processing. Subsequently, transport and boundaries lines are removed by morphological operations (opening).

The removal of textual annotation cannot be realised by text recognition software (OCR) because the type and size of script in the individual maps is too diverse, while the short town names and abbreviations do not form unitary areas of text. Thus a special object recognition feature using image pyramids (stepped image resolution) was introduced for text and signatures removal.

The removal of script is based on an automated parameter analysis of all objects in the segmented binary image regarding their morphological characteristics, such as compactness, convexity, anisometry and orientation. Special structure element parameters were introduced for the analysis of signature objects, since signatures are only partially retained in the resulting layer and cannot be clearly distinguished. Subsequently, still existing signatures such as churches, towers, mines, transformer stations are extracted and removed from the original image. Variations in the quality and characteristics of the DTK25-V maps from the various German states are a problem for the determination of image processing parameters. It is not possible to reconstruct the few buildings completely or partially masked in the DTK maps by textual annotation. As the advanced image processing operations are not implemented in GIS or RS software, it was necessary to employ the image analysis software HALCON [Halcon 2007].

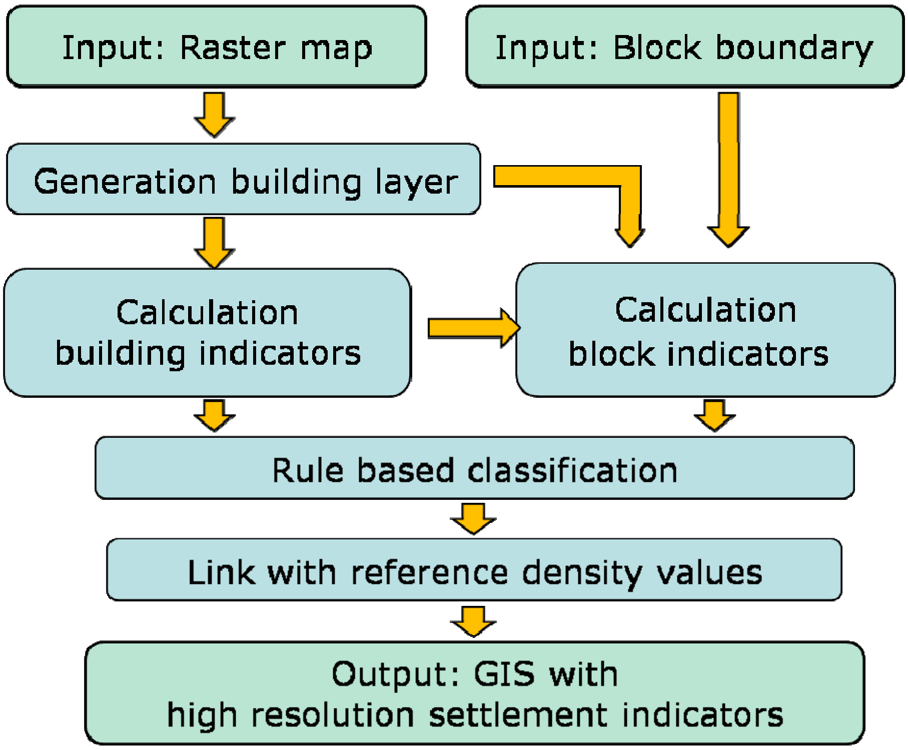

Fig. 1: Processing Workflow in SEMENTA®

4. Classification of Buildings and Blocks

After extraction and vectorization, buildings are represented in the form of unclassified polygons. The second processing step is therefore to classify the entire building stock into predefined types. Table 1 shows the chosen two-level hierarchical typology of buildings. Level 1 (main level) roughly distinguishes three classes of residential buildings and one class of non-residential buildings according to structural characteristics: block structures and open structures in multi-family housing, detached and semi-detached housing organised in smaller groupings, as well as irregular structures in non-residential buildings. Further characteristics are distinguished at Level 2 according to the size of the individual building complexes.

In order to realise an automated rule-based classification, it is necessary to survey the entire stock of buildings only in terms of shape, location and orientation (e.g. area of building, its perimeter, compactness, distance to neighbouring buildings, distance to block perimeter etc.) and link these attributes to the building geometry.

A rule-based decision-making network was created for the automated classification process. This has been proved extremely efficient in the processing of data, while allowing great flexibility in the modification of the underlying rules in comparison to other examined classification strategies (multi-nomial logistical regression, compromise optimization using fuzzy methods). A further advantage of rule-based classification is that an acceptable result can be obtained in a short time using only a few rules. The elements of a rule-based system are the rules present in the initial conditions (premises) and the action phase (conclusions). The characteristic parameters of buildings can be used for classification using if-then rules with the aid of statistically determined threshold values.

Level 1	Level 2
MFH in closed block structure	Multi-family house (MFH), traditional in closed structure
	Multi-family house, free-standing (traditional or new)
MFH in open block structure	Multi-family house - traditional in rows
	Multi-family house - industrial in rows
	Tower block >50m
Detached, semi-detached and row house	Detached and semi-detached house
	Row house
	Traditional village-style house
Non-residential house	Industrial or commercial house
	House with functional profile such as administration, health, social, education, research, culture etc.

Tab. 1: Selected typology of buildings

The classification process was realised in two phases. In the first phase, individual buildings are classified according to their shape (e.g. small, linear, large breadth or complex). Parameters related to neighbouring buildings or blocks are not used at this level – the focus is solely on parameters which describe the basic building shape. This means, however, that the buildings cannot be clearly classified into one of the predefined types (Table 1). Building type can only be finally determined in the second classification phase, when further information is incorporated on neighbouring buildings, block-related parameters, the object type in ATKIS and the calculated areas of individual classes from phase 1. The decision-making network makes use of 63 variables, of which 17 are related to individual buildings and 46 to blocks. An optimization of the statistically determined threshold values is achieved by comparison with orthophotos and reference datasets. Figure 2 shows results of building selection and classification for a part of study area Dresden/Germany.

5. Determination of Settlement and Density Parameters

Classification of the building stock allows the derivation of basic data on settlement structure. The following block-based planning parameters are calculated automatically:

- Block type (7 residential and 2 non-residential types)
- Number of buildings and building density
- Developed area and density of developed area
- Average number of floors (weighted average of the number of floors of all buildings in a block)
- Floor space and floor space density (sum of all building footprints multiplied by the average number of floors)
- Building volume and density of building volumes
- Number of flats and flat density
- Number of residents and density of population.

Fig. 2: top left: Footprint Layer of Map, top right: Building Extraction, mid left: Building Classification, mid right: Block Classification, down: Ortho-photo (Example of Dresden)

The determined ratios of building types in each block, in terms of area and number, give detailed information on the exact composition of buildings in any block. This high-resolution structural data can greatly aid planning processes. For the cities of Dresden and Bonn, it was possible to classify 76% of blocks. While a very high level of classification could be achieved for closed structures (94%), detached structures

(91%) and tower blocks (80%), only a relatively low level of classification could be achieved for physiognomically similar structures such as row and linear classes. Further work is necessary to improve these results.

Visual interpretation of study areas Dresden and Bonn supplied a classification of buildings (totalling 13,120 and 25,176 respectively), whilst block-related residential and housing data were available from intra-municipal statistical sources. To determine typical reference density values for inhabitants and flats in buildings, blocks were then chosen which were largely homogenous in terms of building types. A relationship was established between the number of flats and the sum of building footprints for each building type in a block, so that clear values for density based on building type could be realised from several measurements. The resulting building-specific reference value flat density in building area (flats/m²) is linked in the calculation process to the classified buildings. Introducing building type and floor space allows an estimation to be made of the number of flats in all buildings. Summing all flats in a block gives the number of flats or flat density for each block (1/ha). These estimated values are subsequently summed to the next largest spatial unit offering statistical data (generally municipalities, or parts thereof if available), compared with the statistical data on flats and, following necessary corrections, once again applied to the buildings or blocks. The deviance of the initial estimation (without correction) for the study area Dresden was +1.1% of the reference value for the flats.

In a similar fashion a determination of the population density in building area (residents/m²) allows an estimation of the number of residents per building and thus per block. The density estimation can be corrected by comparison with the value of the lowest know statistical unit.

Fig. 3: Population density for city Bonn: reference by intra municipal statistic (left), prediction by SEMENTA (right)

Figure 3 shows the results for the local population densities in Bonn. The divergence of the initial estimate (with-out correction) in comparison to the reference population figure was +10.4 %.

6. Implementation

The described pre-processing, building extraction, object surveying and classification processes, as well as the derivation of parameters on settlement structures, have been completed automatically with the aid of a specially developed GIS software add-on. An intuitively designed user interface was developed in the form of a toolbar for ArcMap (ESRI), in order to make the operation and control of the different modules as simple as possible. The add-on was written in the C# programming language as a dynamic-link library (DLL). Applications have been made to patent the program design as well as the program itself, and the add-on has been given the name SEMENTA® (SettlementAnalyzer).

The menu-driven program allows selection of basic data (DTK25, ATKIS-Basis-DLM and statistical data for comparison and correction) and offers diverse program settings. Also, the program parameters building density of population EGD ($1/m^2$), building density of flats WGD (flats/m^2) as well as the average number of floors and floor heights (m) of defined building types can be modified.

After processing, SEMENTA® supplies shape data and EXCEL compatible statistical data. An ArcMap project is also opened automatically, offering visualization of the most important layer of results with predefined legends (figure 4).

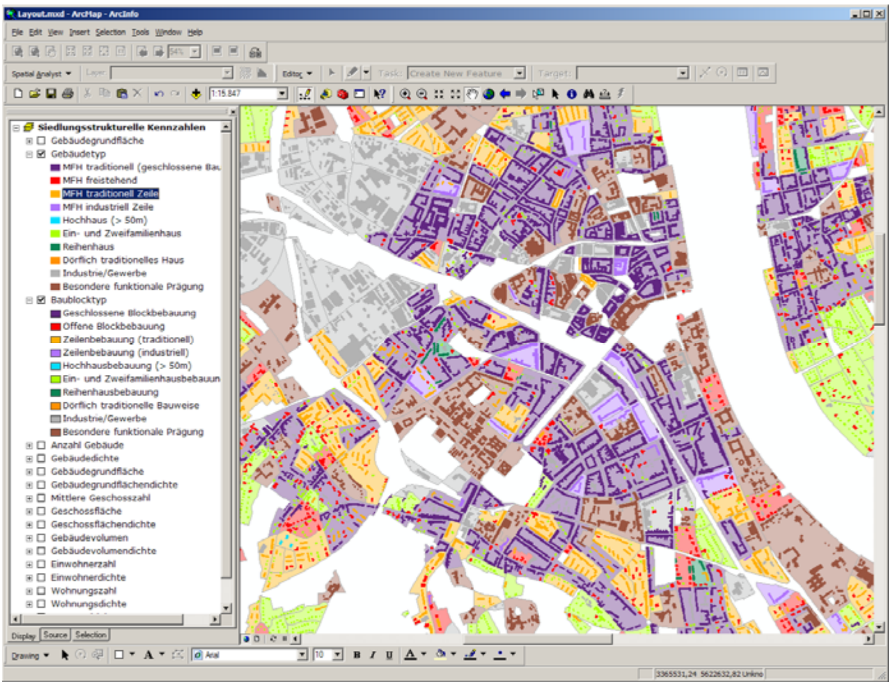

Fig. 4: ArcMap project as automatic results of SEMENTA® and toolbar

7. Application and Prospects

The presented procedure has a large area of application in the planning sector. For the first time, a fully-automated method of surveying and classifying building stock allows a very exact description of the settlement structure using a system of quantitative parameters, down to the level of individual blocks. This fundamentally improves the information base used in planning and decision-making processes. The presented process replaces the tradition of assuming and generalising data with low spatial or informative detail, as well as “direct” data acquisition (which of course is only possible over limited areas). The developed procedure can also be an important aid for inter-urban transport planning (Federal Transport Plan). Thus a large number of applications are feasible for this new procedure:

- State, regional and urban planning
- Roadways planning

- Infrastructure planning
- Hazard maps
- Geomarketing.

The developed procedure has great potential in view of the general availability and regular up-dating of the data it uses. The data can also be utilized for planning processes by introducing geobasis data produced at great cost by regional survey offices. In the future, individual planning specialists should strongly influence the further development of geobase data, as their work makes them predisposed to welcome an expansion of the relevant datasets.

The increasing availability of maps in the new graphical form (such as DTK25 maps directly derived from ATKIS) greatly improves the precision of the results: buildings are shown more exactly (digitized in greater detail) and presented in separate layers, thus obviating the time-consuming and occasionally error-prone process of building extraction.

The procedure could also offer a quantitative description of changes in the building stock (down to the level of individual building-units) by processing maps of different time slots. This would provide technical support when realising the aims of national political strategies in housing sustainability. Although up to now, development has focussed on the construction of housing, with SEMENTA® the stock of industrial facilities and commercial premises can also be separately examined. Such buildings are clearly characterized by their specific geometry, in particular the frequent large breadth (unusual in housing due to the requirement of natural lighting) and can be differentiated using building classification codes.

In the near future the program should also be upgraded to support the calculation of the environmental efficiency of settlement structures. Adaption of the building extraction function of SEMENTA® to accommodate older or historical topographic maps would enable an evaluation of previous settlement structures, and thereby a retrospective analysis of the development of such structures. Finally the program is a basic tool in future for an indicator calculation in a nation-wide monitor for settlement and open space development in Germany. The program could also be adapted for international use by making it compatible with the diverse range of national topographic raster maps in scale 1 : 25,000.

References

Wonka, E. (2006): Regionalstatistik in Österreich, Salzburger Geographische Arbeiten, Band 39, Hrsg. Strobl, J., Salzburg/Wien (2006)

Eschwege A.V., Heidrich-Riske, H (2006): Nutzung des Raumbezuges in der amtlichen Statistik, Statistisches Bundesamt, Wirtschaft und Statistik, 2/2006, pp. 118-135

Szibalski, M. (2006): Karten in der amtlichen Statistik, Statistisches Bundesamt - Wirtschaft und Statistik 3/2006, pp. 205-21

Mennis, J. (2003): Generating Surface Models of Population Using Dasymetric Mapping, The Professional Geographer 55(1), pp. 31-42

Langford, M., Unwin, D. (1994): Generating and Mapping Population Density Surfaces within a geographical Information System, The Cartographic Journal 31(1), pp. 21-26

Dhar D.B., Chanda, B. (2006): Extraction and Recognition of Geographical Features from Paper Maps, International Journal on Document Analysis and Recognition, 8(4), pp. 232-245

Meinel, G., Herold, H., Hecht, R. (2006): Automatische Ableitung siedlungsstruktureller Grundlagendaten auf Basis digitaler Bildverarbeitung, GIS und räumlicher Statistik. In: Strobl, J.; Blaschke, Th.; Griesebner, G. (Hrsg.) Beiträge zum 18. AGIT-Symposium Salzburg/Heidelberg, Wichmann, pp.423-429

Halcon (2007) – Machine Vision Program, MVTec Software GmbH - Machine Vision Technologies http://www.mvtec.com