'Photogrammetric Week '99' D. Fritsch & R. Spiller, Eds., Wichmann Verlag, Heidelberg, 1999.

Trends and Perspectives of Automated GIS Data Collection

HELMUT MAYER, München

ABSTRACT

In this paper we discuss trends and perspectives of automated data collection for Geographic Information Systems (GIS). We focus on aerial imagery and employ road extraction for the examples. The paper is split into three parts: In the first we present major findings for the (fully) automatic extraction of buildings, roads, and other topographic objects in the form of a model and a strategy. In the second part, we give an impression of the strengths but also of the problems by presenting results of the evaluation of the automated extraction of roads from aerial imagery. This leads to the third and main part of the paper where we highlight trends for automatic data collection and set them into relation to different ways to integrate human interaction in the process. Finally, we derive perspectives for automated GIS data collection.

1. INTRODUCTION

Data collection is still the major bottleneck for the wide-spread use of Geographic Information Systems (GIS). Therefore, any kind of automation is welcome. Besides mobile mapping systems based on cars, aerial imagery is a standard data source for which automation seems to be possible. Unfortunately, the research of more than the last quarter of the century has resulted in only few systems for specialized tasks which are close to be useful in practice.

In this paper we aim at trends and perspectives of automated data collection for GIS. To do so, in Section 2 we first present an overview in the form of a model and a strategy for automatic, i.e., without human interaction, GIS data collection. Then, by giving results for an evaluation of automated GIS data collection in Section 3 we give an impression of the strengths but also the weaknesses of existing approaches in this case for automated road extraction from aerial imagery. This gives way to trends for automatic data collection and a discussion of how to integrate human interaction. This is the main body of this paper and is presented in Section 4. Finally, we derive perspectives for automated GIS data collection in Section 5.

2. MODEL AND STRATEGY FOR AUTOMATIC GIS DATA COLLECTION

A detailed description of a model and a strategy for the automatic extraction of buildings and roads can be found in (Mayer 1999) and (Mayer et al. 1998), respectively. Here, we only summarize the major findings for the model and the strategy separately.

2.1. Model

Globally, the model can be split into more or less general parts, which are valid for various topographic objects, and specific parts for buildings and roads, respectively. Important general parts of the model comprise:

- Characteristic properties often arise from the *function of objects* (e.g., usefulness for humans).
- If *material properties* are modeled, the interpretation is less affected by sensor properties.
- *Two dimensional (2D) regularities* such as parallelism are characteristic for many object types.
- With a *detailed image model* the information content of the image is much better exploited.

- By the *abstraction capabilities of scale spaces* (Lindeberg 1994), a globally consistent interpretation by an elimination of disturbances in coarse scales can be combined with making full use of the detailed information (e.g., road markings) in fine scale.
- The context defined by the *geometric/topologic neighborhood* of objects (e.g., road – driveway – building) reduces ambiguities arising from misinterpretations of single objects significantly. It can be split into geometrically local arrangements (*local context*) which are constrained by a global partitioning (*global context*), e.g., into suburb_urban, forest and open_rural.
- *Structures of parts*, such as cars on a road, or doors and windows in a wall are local evidence for an object.
- Functional and deterministic modeling can be extended by *statistic modeling*.

Additionally, there are specific parts for buildings and roads. For buildings these are:

- *Shadows* and *Walls* (vertical edges) are very good evidence in mono images.
- The *three dimensional (3D)-geometry in two or more images* gives an important indication for the existence of 3D-structure characteristic for buildings.
- With a *generic 3D-model* consisting of surfaces and constructive solid geometry (CSG) modeling, complex building structures can be described properly.
- With *aspects* derived from a generic description based on building terminals and connectors, a parallel modeling in 2D and in 3D is feasible by allowing for an explicit transition.

For roads there are the following specific parts of the model:

- *Lines* are employed to model roads in coarse and medium scales/resolutions.
- The *road pavement* which is mainly described by parallel edges is suitable for fine scales.
- By the *road network* consisting of connections and intersections the roads are tied together.
- The *global network* extends the local connections by a criterion of optimality for the whole network. This is extremely useful to bridge gaps.

2.2. Strategy

Here we mean by "strategy" a collection of principles of how to proceed when extracting objects. Like the model, it is split into general and specific parts. The general parts consist of:

- *Appearance-based methods* avoid the explicit transition from image to object domain, e.g., by template-matching. This avoids the complex explicit modeling for objects such as trees.
- *Grouping*, i.e., the search for geometric/topologic regularities, allows to focus on parts of objects.
- By *focusing on different scales*, the extraction is sped up and improved at the same time by using reliable structures in coarse scale to focus in fine scale.
- *Hypotheses generation and search/re-segmentation based on spatial context* is done by predicting objects based on known spatial relations to given objects. Many objects receive their semantics only this way.
- *By focusing on contexts*, the distinction in global and local context (see above) is used for a further improvement of hypotheses generation.
- *Structures of parts generate* evidence and thereby improve the probability of hypotheses.
- By *balancing image information versus the geometric modeling*, i.e., by using techniques such as snakes, the geometry of objects with an already clear semantics can be improved but also additional evidence can be gained for specific object types.

'Photogrammetric Week '99' D. Fritsch & R. Spiller, Eds., Wichmann Verlag, Heidelberg, 1999.

- The *fusion of data and algorithms* combines on one hand color and multi-spectral imagery with images from other sensors. On the other hand, different algorithms give additional valuable information for many tasks.
- By using *GIS*, a speed-up is obtained by focusing onto relevant areas. Complex hypotheses can be avoided in many cases.

Specific parts of the strategy for buildings are:

- *When primitives are matched in several images*, valuable information about the 3D-geometry of parts of buildings can be obtained. As an approximation, a digital surface model (DSM) is very useful.
- By *employing aspects*, i.e., the sequence 3D-points – buildings-part – building – matching of building-part to image primitive, a direct transition from object to image and vice versa is feasible.
- Results for an *extraction of hypotheses for buildings from a DSM* are not too precise and also unreliable, but useful for many applications and especially as a robust approximation for further refinement.

For roads, the strategy comprises:

- *Road tracking*, e.g., based on profiles, makes full use of the location and the direction of a road by employing techniques such as the Kalman filter. The drawback is the computational complexity.
- By *grouping for the construction of the road network*, gaps are locally closed and intersections are constructed.
- *Grouping based on the network* extends the previous item by taking into account the global structure and especially the hierarchy of the network.

3. EVALUATION OF AUTOMATED GIS DATA COLLECTION

The model and the strategy we have introduced in the previous section are only partly realized in the various approaches available especially for building and road extraction. For building extraction, the most advanced systems are (Fischer et al. 1998, Henricsson 1998). We concentrate on road extraction and give an impression of the state-of-the-art by presenting results for an evaluation of different approaches.

In (Heipke et al. 1998) an evaluation of the approaches developed at Technische Universität München (TUM), Germany (Baumgartner et al. 1999) and at Institut Géographique National (IGN), Paris, France (Ruskoné and Airault 1997) is presented. Table 1 gives some of the most important results for the evaluation of the results of the TUM approach based on multi-scale, grouping, and context, optimized for rural areas, and the IGN approach making use of context for the image "Marchetsreut". This image can be characterized as "flat, agricultural, easy". The pixel size on the ground is 0.225 m. To give an idea about the actual outcome of the approach, the result for the TUM approach is shown in Figure 1.

	TUM	IGN
Completeness	0.91	0.81
Correctness	0.99	0.91
Quality	0.90	0.78
RMS [m]	0.28	0.56

Table 1: Evaluation results for image “Marchetsreut”.

In Table 1, completeness is the percentage of the reference data which is explained by the extracted data. Its optimum value is 1. Correctness represents the percentage of correctly extracted road data. The optimum value is again 1. Quality combines completeness and correctness in one value with an optimum value of 1. At last, the RMS difference expresses the average distance between the matched extracted and the matched reference network.

Figure 1: Result of automatic road extraction for image “Marchetsreut” using the TUM approach (white lines: correct result; black lines: incorrect result, i.e., not matched to ground truth).

'Photogrammetric Week '99' D. Fritsch & R. Spiller, Eds., Wichmann Verlag, Heidelberg, 1999.

The results of the evaluation in Table 1 look very good. Especially the TUM approach gives nearly no wrong answers and also the completeness is so high that a combination of the automated extraction and a manual correction and completion looks promising. Unfortunately, these good results are only obtained for easy, i.e., mostly rural areas. In Table 2 results are shown for the image “Erquy”. The results of the evaluation of the fully automatic extraction are still relatively good, but in this case it is not really obvious that a simple combination of automatic extraction and manual completion is much more efficient than standard manual plotting. Still, this is only a problem of completeness and correctness: The RMS value for the correctly located roads is in nearly all cases better than needed for most applications.

	TUM	IGN
Completeness	0.79	0.45
Correctness	0.94	0.62
Quality	0.75	0.35
RMS [m]	0.53	0.95

Table 2: Evaluation results for image “Erquy”.

Things become worse in urban areas. In (Harvey 1999) results for an evaluation of the approach for road extraction of the Carnegie Mellon University, Pittsburgh, USA are presented. With a fully automatic process an overall completeness of 0.56, a correctness of 0.27, and a quality factor of 0.25 was reached. When using manually chosen starting points, the completeness reaches up to 0.76, the correctness 0.91, and the quality 0.73.

Summing up, automatic extraction has improved in recent years and this has been demonstrated by the good results when evaluating the approaches. Though, still missing is a more widespread and general evaluation with a wider range of images. This makes, however, only sense when the approaches will make further use of the trends presented in the next section.

4. TRENDS AND THE INTEGRATION OF HUMAN INTERACTION

There are five trends for automatic GIS data collection whose outstanding importance has become clear only recently.

First, *scale* is not only important because objects can only be seen at some minimum resolution. More importantly, by abstraction in coarser scales, e.g., generated by scale-spaces (Lindeberg 1994), features like lines can directly be linked to objects such as roads (Mayer and Steger 1998).

Second, *context* with its spatial organization is a highly effective means to structure the knowledge. This makes it feasible to construct large consistent models and strategies taking into account the high variability and complexity of objects and their relations (Baumgartner et al. 1999).

Third, the *3D-structure* of vegetation and especially buildings is the key to their recognition. There are two ways which should be combined:

- DSM are extremely useful to detect vegetation and buildings. Their reliability is improved very much when DSM from active laser scanning, e.g., (Lohr 1999) is used.
- By matching features in two and more images using the information from a given DSM as an approximation and a detailed image model, highly reliable 3D-structures are obtained which can be combined with the knowledge from the model (Fischer et al. 1998, Henricsson 1998).

Fourth, *fusion* of data from different sensors can help to reduce ambiguities very much. For instance, trees can be discriminated from roofs and grass on the ground based on color information from aerial imagery and height in the form of laser-scanner data (Haala 1999).

Fifth, *GIS data* can be used to focus the extraction (Plietker 1997, Haala 1999, Walter 1999), increasing the speed of the extraction as well as the reliability.

Besides a further improvement of the automation by deepening of the model and the strategy, one of the major challenges is the integration of the human. The rationale behind this is that nobody expects fully automatic systems with an acceptable error rate for the next ten or twenty years. Thus, the human is needed to improve the results by interacting with the system. Basically, there are two ways to proceed:

- The automatic procedure works off-line and the data and results are improved/extended before or after it.
- An automatic component runs on-line while the operator interacts with the system. This is what many people understand by a “semi-automatic system”.

As the results of the automatic systems did look much worse some years ago, the development of semi-automatic systems seemed to be a good idea. For buildings, the system described in (Gülch et al. 1998) is maybe the most advanced one. It is based on CSG-trees and uses techniques like RANSAC (random sample consensus) and clustering to compute parameters such as gutter height, width, and length. During its development, approximately a factor 4 of speed-up has been achieved in the last five years.

Semi-automatic systems for roads rely either on starting points and a given direction (Airault et al. 1996), or start- and end-point as well as some intermediate points are given (Grün and Li 1997). In the first case, start point and direction are used to track the road, e.g., based on the homogeneity of the road. In the second case, the intermediate parts are reconstructed, for instance, using optimization techniques such as “snakes”. This can be done in 3D when two or more images are available. Then it is even possible to take into account occlusions in some of the images.

Meanwhile, the attitude about human computer interaction has changed. Many people believe that there is no sharp distinction between fully automatic systems with pre- and post-editing and semi-automatic systems, but that they complement each other. While it is a good idea to pre-edit the data and then use a fully automatic process, in many cases focussed by GIS data, to arrive at a 90+% solution, it is wise to base the post-editing on semi-automatic processing supported by (partial) results of the automatic processing. One example for this attitude is presented in (Haala 1999). After automatically extracting CSG-trees for buildings from laser-scanner data based on GIS-data (ground-plans), they are refined taking into account the laser-scanner data based on 2D data given by the operator. Another example, in this case for road extraction, is presented in (Airault et al. 1996). There, an automatic system complemented with given GIS data and restricted via hard thresholds to areas where it is reliable, is combined with the semi-automatic approach presented above.

5. PERSPECTIVES OF AUTOMATED GIS DATA COLLECTION

As shown in the preceding section, there are two main ways to proceed: The improvement of automatic components and the smooth integration of human interaction. For the automatic components, the trends presented in the last section should be followed, but additionally the following issues should be further investigated:

- The behavior of objects in different scale-spaces should be investigated. This includes characteristic textures.

'Photogrammetric Week '99' D. Fritsch & R. Spiller, Eds., Wichmann Verlag, Heidelberg, 1999.

- Local contexts should be modeled in more depth.
- For structures of parts it should be investigated how important details like cars, markings, or windows are.
- The function of objects could be modeled more explicitly, e.g., based on autonomous agents.
- The usefulness of different sensors and algorithms for fusion is an important question. For sensor fusion, material properties should be modeled. Imaging laser-scanners might help because they are more or less independent of lighting conditions.
- Machine learning can help to improve the versatility of object extraction. Reference data might come from a GIS.
- To treat the variability of objects sufficiently, more than one appearance of objects should be modeled which leads to a “multi-model”. A strategy might be to tackle the most prominent objects in an area first.
- Very important is a more detailed modeling. Not only knowledge about dormer windows and gutters, but also, e.g., about architecture should be incorporated. This is to be linked with statistical modeling. In speech recognition, only after the introduction of a statistical interaction of objects via hidden Markov models, systems useful for practice became feasible.
- Statistical modeling is not only needed to decide about the semantics of objects, but also to control object extraction.
- In principle, GIS data can only be verified when enough evidence is found in an image via an elaborated model. In practice, it might be enough for certain types of objects like big buildings or highways when part of them are found, because one knows that the parts which could not be verified have to be there. Additionally, it might be enough for certain applications when the system is right in 99 % or even less of the cases.

An effective interaction of the human with the automatic system is highly influenced, in the first place, by the error rate of the system. In principle, the less errors the system makes, the faster the operator can eliminate them and therefore complete data collection. In practice, things are not that simple: There are errors, especially in topology, which take much more time to get rid of by a human operator than others. Most important, when the system cannot tell with a very high level of confidence that a result is actually correct, one might have to scan the whole result for errors and cannot only focus on some parts where the system is not sure about. In (Förstner 1996) this is called the traffic light paradigm, where green means that the system is sure about its decision, yellow means that inspection might be needed, and red tells you that the system is not sure at all. But still, this information is only useful when one can trust in it.

Putting things together, it is critical not to start too early with the integration of an automatic system with human interaction. This is mostly due to the fact that an effective interaction depends very much on the type and the quality of the results. Thus, when the automatic part is improved, it is very likely, that the interactive part, mostly consisting of interactions before and after the automatic parts, has to be totally redesigned. This was not too critical if the systems were used in practice and the effort would pay off. But as we are still in a research state and the interaction component is, due to its inherent complexity, not a key research issue, one should better stick to (international) evaluation contests of the automatic systems and for the practical part focus on semi-automatic systems where the extraction is directly controlled. Though, integrating fully automatic parts is pretty interesting because only with them, an extreme speed-up and cost reduction, needed for instance to make the wide use of detailed 3D-city models a reality, seems to be practical.

6. REFERENCES

Airault, S., Jamet, O. and F. Leymarie (1996): From Manual to Automatic Stereoplotting: Evaluation of Different Road Network Capture Processes. International Archives of Photogrammetry and Remote Sensing (31) B3/III, pp. 14-17.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W. and H. Ebner (1999): Automatic Road Extraction Based on Multi-Scale, Grouping, and Context. Accepted for Publication in "Photogrammetric Engineering & Remote Sensing, July 1999".

Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Förstner, W., Plümer, L. and V. Steinhage (1998): Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D. Computer Vision and Image Understanding (72) 2, pp. 185-203.

Förstner, W. (1996): 10 Pros and Cons Against Performance Characterization of Vision Algorithms. European Conference on Computer Vision, Workshop on Performance Characteristics of Vision Algorithms, pp. 13-29.

Grün, A. and H. Li (1997): Semi-Automatic Linear Feature Extraction by Dynamic Programing and LSB-Snakes. Photogrammetric Engineering & Remote Sensing (63) 8, pp. 985-995.

Gülch, E., Müller, H., Läbe, T. and L. Ragia (1998): On the Performance of Semi-Automatic Building Extraction. International Archives of Photogrammetry and Remote Sensing (32) 3/1, pp. 331-338.

Haala, N. (1999): Combining Multiple Data Sources for Urban Data Acquisition. These Proceedings.

Harvey, W. A. (1999): Performance Evaluation for Road Extraction. Proceedings, ISPRS Workshop on "3D Geospatial Data Production: Meeting Application Requirements", Paris, France, April 7-9, 1999, pp. 175-184.

Heipke, C., Mayer, H., Wiedemann, C. and O. Jamet (1998): External Evaluation of Automatically Extracted Road Axes. Photogrammetrie – Fernerkundung – Geoinformation 2/1998, pp. 81-94.

Henricsson, O. (1998): The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction. Computer Vision and Image Understanding (72) 2, pp. 163-184.

Lindeberg, T. (1994): Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston, USA.

Lohr, U. (1999): High Resolution Laser Scanning for 3D city models. These Proceedings.

Mayer, H. (1999): Automatic Object Extraction from Aerial Imagery – A Survey Focusing on Buildings. Computer Vision and Image Understanding (74) 2, pp. 138-149.

Mayer, H., Baumgartner, A. and C. Steger (1998): Road Extraction from Aerial Imagery: http://www.photo.verm.tu-muenchen.de/CVonline_roads/. In CVonline: On-line Compendium of Computer Vision: http://www.dai.ed.ac.uk/CVonline/.

Mayer, H. and C. Steger (1998): Scale-Space Events and Their Link to Abstraction for Road Extraction. ISPRS Journal of Photogrammetry and Remote Sensing (53), pp. 62-75.

Plietker, B. (1997): Automatisierte Methoden zur ATKIS-Fortführung auf der Basis von digitalen Orthophotos. Photogrammetric Week 1997, Wichmann, Karlsruhe, pp. 135-146.

Ruskoné, R. and S. Airault (1997): Toward an Automatic Extraction of the Road Network by Local Interpretation of the Scene. Photogrammetric Week 1997, Wichmann, Karlsruhe, pp. 147-157.

Walter, V. (1999): Automated GIS Data Collection and Update. These Proceedings.