OBJECT EXTRACTION FOR DIGITAL PHOTOGRAMMETRIC WORKSTATIONS

Helmut Mayer
Institute for Photogrammetry and Cartography, Bundeswehr University Munich, D-85577 Neubiberg, Germany
Helmut.Mayer@UniBw-Muenchen.de

KEY WORDS: Object, Extraction, Digital, Photogrammetry, Semi-automation, Modeling, Statistics, Test

ABSTRACT

This paper deals with the state and with promising directions of automated object extraction for digital photogrammetric workstations (DPW). A review of the state of the art shows that there are only few success stories. Therefore, important areas for a practical success are identified. A solid and most important powerful theoretical background is the basis. Here, we advocate particularly statistical modeling. Testing makes clear which of the approaches are best suited and how useful they are for praxis. A key for commercial success is user interaction, an area where much work still has to be done. As the means for data acquisition are changing, new promising application areas such as extremely detailed three-dimensional (3D) urban models for virtual television or mission rehearsal evolve.

1 INTRODUCTION

Digital photogrammetric workstations (DPW) (Heipke, 1995) have been introduced in the market on a larger scale at the middle / end of the nineties and have become the standard for photogrammetric processing. While tasks with a high redundancy such as orientation have reached a high degree of automation and robustness, this is only partially the case where the redundancy is not so high such as for the generation of digital surface models (DSM) or digital elevation models (DEM). For the latter two, laser-scanning has become an attractive alternative.

While it was a matter of a few decades to highly automate the above tasks, the situation is much more difficult for automated object extraction. There are only few (semi-) automated systems which are used with success in the market. (Baltsavias, 2004) cites most prominently the systems for building extraction InJect of INPHO GmbH (Gülch et al., 1999) and CC-Modeler of CyberCity AG (Grün and Wang, 2001). Additionally, the systems for road update and verification ATOMIR (Zhang, 2004) and WIPKA-QS (Gerke et al., 2004) are on the verge of becoming operational.

This paper addresses reasons for this deficit, but also points on issues we think are important to improve the situation and introduce object extraction on a larger scale in practical applications.

Legend has it, that in the 1950ies scientists from the field of artificial intelligence thought, that the solution of the vision problem was a matter of a graduate student project. This estimation then shifted from five years to twenty years and then to much much longer. Today, there is a large body of knowledge in different fields as diverse as psychology (Kosslyn, 1994) and the use of geometry in computer vision (Hartley and Zisserman, 2000), but still we might be only at the beginning of understanding the basic problems.

There is still progress not only in the high level understanding, i.e., interpretation, area, but also in the basic understanding of the image function. E.g., (Köthe, 2003) has shown that the well known operator (Förstner and Gülch, 1987) does not take into account the frequency doubling implicit in the squaring of the Hessian matrix (some people also call it the structure tensor). The SIFT operator of (Lowe, 2004) offers scale and rotation invariant features which can be robustly matched under affine distortion, noise, and illumination changes. (Pollefeys and Van Gool, 1999, Pollefeys et al., 2002) have shown that it is possible to fully automatically reconstruct the pose and calibration of images of a camera of which the only thing known is, that it is perspective. They also demonstrated the importance of redundancy in matching, an issue recently propagated by (Gruber et al., 2003) for robust DSM / DTM generation by means of digital aerial cameras. In (Nistér, 2003) a direct solution for the five-point relative orientation problem is given. Finally, the test of (Scharstein and Szeliski, 2002) on stereo matching has sparked a large number of new approaches for matching, using, e.g., the powerful graph cut technique (Kolmogorov and Zabih, 2001), or cooperative disparity estimation (Mayer, 2003).

This paper rests on a recent survey (Baltsavias, 2004) which summarizes important points for the practical use of object extraction. We will not repeat the contents of this survey, but rather deepen some points, yet giving enough overview of the area to make this paper self-contained. We are mainly concerned with aerial imagery and laser-scanner data. Although focusing on the former two sources, we also deal with satellite imagery and other data such as hyper spectral data or terrestrial video sequences and laser-scanner data. To limit the scope, we do not consider radar data.

The prerequisite for productive object extraction is modeling (cf. Section 2), which in our case comprises also the strategy, data sources including data from geographic information systems (GIS), statistics with and without geometry, and learning. While a lot of basic scientific work ends at this point, there is a recent tendency to evaluate the performance of the approaches, possibly also in comparison with each other, in different tests. We think that testing as presented in Section 3 is a must for bringing object extraction into praxis as it not only makes clear which models and strategies are superior to others, but it also shows what is possible with object extraction. After a successful test, the next step is to design the user interaction in semi-automated systems. We discuss the state of the art and several issues in Section 4. As all technical developments are nothing without markets, we give in Section 5 an idea about future markets and what other areas, particularly visualization from computer science, envisage. The paper ends up with conclusions.

2 MODELING

Modeling is the key issue for the performance of any approach for automated or also semi-automated object extraction. Basically, modeling consists of knowledge about the objects to be extracted. Additionally, in most cases it is necessary to analyze their mutual spatial and topologic relations as well as their relations to additional objects, which a customer might not be interested in to extract, but which give important clues for the recognition of an object. E.g., even though one is just interested into roads in city centers, one will only find them, when one knows, where the cars are (Hinz, 2003).

The modeling of the objects is the key issue. But instead of analyzing the assets and drawbacks of individual approaches, as, e.g., in (Mayer et al., 1998, Mayer, 1999), we will in the remainder of this paper concentrate on a number of issues we consider as important to improve object extraction for DPW. Overall it is our firm believe, that only by a detailed modeling of many objects and their relations of the scene, it will be ultimately possible to mostly reliably extract objects from imagery, laser-scanner data, etc.

2.1 Strategy and Multiple Scales

Even though the objects and their relations are the necessary core of modeling, experience shows, that the sequence of operations employing the knowledge about the objects and their relations is a, often even the key factor for an efficient, but also powerful extraction. E.g., it is well known that markings are an important clue to find roads. Unfortunately, in images with a ground pixel size of about 0.25 m the markings very often correspond to very faint bright lines. When trying to extract them in open rural space one will in most cases extract millions in the fields and meadows leading to an infeasible grouping problem. On the other hand, one can first produce hypotheses for roads in the form of lines in images of a reduced resolution, i.e., images in a higher level of an image pyramid. Then one verifies the roads in the form of directed homogeneity such as in (Baumgartner et al., 1999). Inside the generated hypotheses for roads the markings can be extracted and grouped reliably, giving the hypotheses a high evidence for being actually roads.

We term the basic concepts behind a sequence of operations controling the extraction the “strategy”. Ideally, there exist objects

- which are easy to extract,
- can be extracted reliably, and
- which have a large positive influence on the interpretation of the whole scene.

The idea is to find cues for objects which allow to focus the attention to specific areas, such as hypotheses for roads to extract markings (cf. above). Unfortunately, this kind of objects does not always exist and if so, they are not always easy to identify.

In the above example on roads, scale plays an important role. Coarse to fine approaches have long been used in orientation determination and in image matching (Heipke, 1995). For linear objects it was shown in (Mayer and Steger, 1998), that by means of changing scale from fine to coarse by means of linear scale-space (Lindeberg, 1994), one can often eliminate interfering objects such as cars and trees together with their shadows from roads. Other means are irregular pyramids, as, e.g., implemented in eCognition of Definiens GmbH (Benz and Schreier, 2001). A comparison of different means is given in (Blaschke and Hay, 2001).

Our experience is, that a multi-scale approach is in many cases useful. Depending on the type of object, smoothing with the linear scale-space, eliminating interfering details by means of gray-scale morphology (Köthe, 1996), or a combination of both such as in (Kimia et al., 1995) is most suitable.

2.2 Data Sources and GIS Data

DPW have included in recent years means to deal with high resolution satellite imagery such as IKONOS or Quickbird together with aerial imagery, possibly digital, e.g., from Leica‘s ADS 40 (Fricker, 2001), Vexcel‘s Ultracam (Leberl et al., 2003), or Z/I imaging‘s DMC (Hinz et al., 2001).

To use data which comprise explicit information suitable for the problem can be a very efficient means to make extraction more robust and reliable. These are most importantly color, or more generally spectral data, as well as three dimensional (3D) data. (McKeown et al., 1999, Mikhail, 2000) show the advantages of using aerial hyperspectral data allowing for reasoning about the materials of the objects. Both make also use of DSM.

For 3D, highly reliable data from laser-scanners are the data source of choice. Early experiments with the extraction of buildings from laser-scanner data where done by (Weidner and Förstner, 1995). Recently, laser-scanner data are more and more fused with aerial imagery. For it, the establishment of a common reference frame plays an important role to arrive at rich features (Schenk and Csathó, 2002). Work such as (Rottensteiner, 2003) uses additionally to the integration with aerial imagery sophisticated segmentation methods and a consistent model estimation. In (Straub, 2003) DSM data from laser-scanners partially together with reflection properties in the infrared are used for the extraction of individual trees.

A very important source often neglected in more theoretical work are GIS data. (Brenner, 2000) use two dimensional (2D) polygons from which they generate straight skeletons in conjunction with laser-scanner data to efficiently and reliably extract buildings. In (Gerke et al., 2004) a two stage process is employed to verify given road data. After extracting reliable roads in the first stage using strict parameters, topologic information is used to restrict the further analysis so that relaxed parameters can be used leading to a more complete verification and therefore to a higher efficiency. (Zhang, 2004) employ color and stereo data together with extensive modeling, comprising, e.g., context, occlusions, and shadows, making heavy, yet intelligent use of the GIS data, leading to an impressive performance.

Even though the above papers show the potential of using GIS information, it is essential to keep in mind, that one cannot rely absolutely on it, as it might be outdated and unprecise and therefore lead to wrong conclusions. There is always a trade-off to be made between accepting wrong, because changed objects, and rejecting many correct, because unchanged objects. Therefore, even when using additional information from GIS, in most cases approaches are needed which model reality so deeply, that they can extract the objects at hand even under complex circumstances as, e.g., (Zhang, 2004) demonstrates it.

2.3 Statistical Modeling

The deficits of a mainly deterministic modeling, for instance based on semantical networks, e.g., (Niemann et al., 1990), have been known for a long time. There have been heuristic attempts by adding for instance believe values, but more sound ways of including statistical modeling have been used only recently, for instance Bayesian networks in (Growe et al., 2000) or (Kim and Nevatia, 2003). The work on dynamic Bayesian networks (Kulschewski, 1999) has been interesting in terms of modeling objects and their relations. Though, manually generated ideal data were used, and thus the feasibility of the approach to cope with real world, noisy, and unreliable data is hard to judge.

Until recently, semantical modeling was also lacking the capability to visualize the actual contents of the knowledge modeled. The quality of the modeling, e.g., by a semantical network, could only be judged by looking at interpretation results and it was not clear how much a component contributed to the results.

By the advent of reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) there is a means for statistical modeling which can also be used for simulation. The jumps in RJMCMC make it possible not only to use distributions for the parameters of objects and relations, but also to introduce new objects or relations and to delete them. The latter is the reason why the jumps are called reversible: For every jump generating a new object there needs to exist a backward jump, allowing to eliminate the object. Because of this, RJMCMC has the following outstanding features:

- The modeling is extended in a sound way to deal with the uncertainty of objects as well as their relations even when it is not known beforehand, which and how many objects exist.
- It is possible to sample into the distribution allowing to simulate objects and their relations according to the model. Thus, one can check from the outcome, if the given model really describes what it is supposed to describe. I.e., in stark contrast to most modeling schemes, one can check the model without analyzing given data.

That the ideas of RJMCMC are practically feasible and meaningful was shown by work on facade interpretation (Dick et al., 2002), road extraction (Stoica et al., 2004), and vegetation extraction (Andersen et al., 2002). The former two demonstrate that one can produce realistically looking facades or roads, respectively, by starting from a few basic primitives, such as a window and a door, or a road piece, and then sampling into the distribution. For the roads and the vegetation, sampling is done for the extraction in conjuction with simulated annealing, avoiding local minima, but also resulting in a very high complexity of the approach.

Another issue of statistical modeling is self-diagnosis. (Förstner, 1996) introduced the “traffic light” paradigm. Results which are correct (green) are distinguished from certainly incorrect results (red) and results, which might be correct, but should be checked (yellow). The idea is that a calling routine will get back information if it can rely on a result (green), if the result might be correct (yellow), or if there was no meaningful result (red). Self-diagnosis is based on statistical modeling. The more one knows about the deterministic and stochastic structure of the problem, the more reliable self-diagnosis will be. (Gerke et al., 2004) have built their approach for road verification on top of the traffic light paradigm.

2.4 Geometry and Statistics

An area of statistics linked to problems often geometrical in nature is concerned with the large number of blunders in the data, the vision community always has to deal with, especially when using matching algorithms. This has sparked the development of techniques which approach the problem differently from how most photogrammetrists do this. Especially popular is the random sample consensus, or short RANSAC approach of (Fischler and Bolles, 1981) and its variants such as the geometric information criterion (GRIC) (Torr, 1997). The basic idea is to take a larger number of random samples with the minimum number of observations necessary to solve the problem. All these samples lead to solutions which are then checked against the rest of the observations. Finally, the solution is taken, which is in correspondence with the largest portion of observations. This technique is extremely useful for applications such as the estimation of the epipolar geometry (Hartley and Zisserman, 2000), for aero-triangulation (Schmidt and Brand, 2003), or to find planes in a large number of 3D points (Bauer et al., 2003).

Computer vision has understood many of the geometric problems of the imaging process over the last decade very well. Early results are summarized in (Faugeras, 1993), while the state of the art is given by (Hartley and Zisserman, 2000, Faugeras and Luong, 2001). In the last years a focus has been on geometric algebras. An important ingredient is Grassman-Cayley Algebra as proposed by (Faugeras and Papadopoulo, 1997). Recently, (Rosenhahn and Sommer, 2002) have extended the scope of geometric modeling significantly, allowing, e.g., to deal with articulated objects linearly. (Heuel, 2001) has presented work where traditional statistics is linked with geometric algebras making it possible to propagate stochastic information.

2.5 Learning

From a practical, but also from a theoretical point of view automatic learning, i.e., the automatic generation of models from given data or even experience, is of big importance as it avoids the tedious manual process of model generation. The latter is one of the most important reasons, why an automated extraction of objects with a wider variety of appearances does not seem to be feasible yet.

For learning one has to distinguish between very different degrees ranging from the mere adaptation of parameters to the fully automatic generation of models for objects such as buildings including their parts, their structure, and their geometry, as, e.g., in (Englert, 1998).

Unfortunately, learning is, after standard textbooks have been introduced a long while ago (Michalski et al., 1984, Michalski et al., 1986), still not advanced enough to deal well with real world problems as complex as object extraction. Yet, this is not a surprise as object extraction is a large part of the overall vision problem which is even after a lot of research by extremely skilled humans not really understood.

Also for learning statistics might come to help. Hidden Markov Models (HMM) have made possible a breakthrough in the interpretation of written and spoken text. Instead of describing words and their relations structurally (grammar) and semantically, it was found for many applications enough just to analyze the statistical dependencies of very few neighboring words based on HMM (Ney, 1999). Similar ideas have been introduced also into image processing, but the much higher complexity makes progress much more difficult.

Finally, concerning another popular means also used for learning, namely artificial neural networks, we refer to the discussion in a recent survey on statistical pattern recognition (Jain et al., 2000). There it is stated, that "many concepts in neural networks, which were inspired by biological neural networks, can be directly treated in a principled way in statistical pattern recognition." On the other hand, it is noted that "neural networks, do offer several advantages such as, unified approaches for feature extraction and classification and flexible procedures for finding good, moderately nonlinear solutions."

3 TESTING

A key factor for the practical use of a technique in many areas is thorough testing. Yet, this is only useful after having obtained a profound theoretical understanding of the problem. There are different issues, where testing can help significantly:

- It becomes evident what the best approaches can achieve and therefore, what the state of the art is.
- The strengths but also the weaknesses of competing approaches become clearly visible and the whole area can flourish by focusing on promising directions, abandoning less promising ones, and by identifying unexplored territory.
- Testing usually gives a large push to all people involved. By trying to outperform other approaches one learns much about the possibilities but also the limits of one's owns approach.

Unfortunately, it is not always easy to define what to actually test. This is most critical for practical issues, such as the effectiveness of semi-automated approaches compared to the manual approaches. It depends on many factors some of them needing lots of efforts for optimization if the real potential of an approach is to be obtained. But also for automated approaches there is a large number of factors which influence the test and by this also which approaches perform well and which not. For roads, e.g., the preferred characteristics of the terrain plays an important role while for buildings, the situation is even worse. There, approaches exist, assuming at least 4-fold image overlap, while others rely on laser-scanner data only, both possibly modeling different types of buildings, e.g., flat roofs versus polyhedral objects.

Our experience shows that for many applications two basic measures are suitable for testing, namely "correctness" and "completeness" (Heipke et al., 1997). Other people use different names for these concepts, but what we mean is the percentage of extracted object information which can be matched to given ground truth data (correctness), as well as the percentage of ground truth data that can be matched to the extracted information (completeness). As one can see, the matching of the object information to ground truth data is an important issue. Road axes can be seen to match as long as they are inside the actual area of the road or inside a buffer generated from specifications for the precision of the acquisition. For buildings it is more complicated as one can match ground truth data and extracted information in 2D and in 3D. Usually, the computation is done in image space (pixels) or 3D voxel space (Shufelt, 1999). To separate orientation errors from object extraction errors, individual objects can be optimally transformed before this computation by means of matching. Depending on the application, another possibility is to consider a valid match to be achieved as long as ground truth data and extracted object information have any overlap.

For approaches relevant for DPW, testing must always be done against real world data, and not against simulated data. The question if ground truth data should be gathered from the 3D reality, or if ground truth data should be manually digitized from the image used also for the automated interpretation is from our point easy to answer: If the goal is to evaluate the whole production chain, the former is appropriate. In many cases one just wants to know how much worse than a human the automated system is. Then, benchmarking against given manually digitized ground truth data is the way to go. To avoid a bias from an operator, one can match against the results of more than one operator such as in (Martin et al., 2004).

Together with Emmanuel Baltsavias of ETH Zurich we have recently set up a test on "Automated extraction, refinement, and update of road databases from imagery and other data" (http://www.bauv.unibw muenchen.de/institute/inst10/eurosdr) under the umbrella of EuroSDR (European spatial data research; formerly known as OEEPE). On one hand, we want to learn the data specification needs of important data producers, mainly national mapping agencies (NMA) and their customers. On the other hand, existing semi- and fully-automated systems for road extraction will be evaluated based on high quality image data against given, manually digitized ground truth data.

4 USER INTERACTION

To limit the scope, we do not deal with multi-spectral classification, which is well understood and for which powerful commercial products such as ERDAS IMAGINE from Leica Geosystems or ENVI from Research Systems Inc. are available. Closer to our intentions is eCognition of Definiens GmbH as it deals with objects, not pixels. Because it aims more at similar applications as the former two products assuming larger ground pixel sizes than DPW, we will not treat it here either.

For general purpose DPW as well as GIS, automated functionality for object extraction is very limited. According to (Baltsavias, 2004), the only more widely known systems actually useful for practice because offering the most automation are the systems InJect of INPHO GmbH (Gülch et al., 1999) and CC-Modeler of CyberCity AG (Grün and Wang, 2001). Though, both are limited with this respect, that they are dedicated to building extraction only.

(Baltsavias, 2004) points out, that it is clear why full automation is not feasible today, but asks "why are important-for-the-practice semi-automated approaches so rare?" We will give some ideas why this is the case, but we will also point on ways how to change it.

Basically, as pointed out above, automated object extraction is extremely difficult and therefore error-prone. Only a limited number of the approaches developed over the last two decades has been developed so far that they work for a larger number of data sets and are ready for testing (cf. Section 3). But even if there was a larger number of approaches with reasonable performance in real world tests, there is another issue which makes the preparation of an approach for practice even more problematic than the usual 1 : 10 : 100 relation between proof of concept : stable prototype : product level: This is the dependence of the user interaction on the performance level and the strategy of object extraction in the system.

This means, that to build a highly effective interactive system, the interaction needs to be tailored to a fixed level of automation. If the level of automation improves, it is not too unlikely, that the interaction of the system will have to be considerably different, implying larger changes for the software, but also possibly for the production chains of the customers. Seen the other way around more positively, (Baltsavias, 2004) recommends to design the control including human interaction to build systems that are useful for practice.

A reaction to the difficulties of fully-automated object extraction is a restriction to problems where the computer directly assists the user in real-time. This is the case for InJect but only partly for CC-Modeler. For roads, this idea has been promoted early, e.g., by (Grün and Li, 1994, Heipke et al., 1994), but nowadays it seems that roads are, e.g., in open rural areas, so easy to extract, that it is a good idea to do it fully-automated. On the other hand, in urban areas, but also in shadows or at complex crossings, they are so difficult to extract, that only fully-automated non-real-time-processing can deal with them today.

For practically relevant systems, we believe, that the human has to be in the loop. We also think that in many cases it is beneficial to use one or two automated off-line processes, probably preceded or interrupted, but in any case followed by manual interaction. The generation of workflows defining the offline-phases, but also very importantly the information to be given to the automated procedure by a user interaction preceding it, is essential for the overall performance.

There are several steps needed to develop a system useful for practice, to be embedded into a DPW. The basis are thorough theoretical understanding and testing. For an efficient user interaction, the key is an appropriate trade-off between completeness and correctness / reliability. It is usually more costly in terms of user interaction time to eliminate complex failures. Therefore, it is a good idea, to use as a basis for human interaction a version where the completeness is still high, but where very few complicated errors, especially in terms of topology occur.

A related issue is self-diagnosis. In this context it is not the same as in statistical modeling as it makes use of additional knowledge about the strengths and weaknesses of human interaction. For a semi-automated system it is extremely important that the correct objects (green) are actually correct with a very high probability, so that they do not have to be checked any more. For the "yellow" results, the situation is more complicated. It should be avoided to offer the operator a lot of objects, which are plainly wrong. Also results with complex topologic errors, which might take more time to heal than to acquire the whole feature manually, should not be presented to the operator. Helpful might be, though, to offer a number of choices, one of which is with relatively high probability correct.

An efficient semi-automated system should comprise real-time tools, which help to improve the results obtained fully automatically. The best way, yet needing also the most effort to implement and again depending on the current state of automated systems, is to make use of the results of automated extraction.

Finally, testing, this time on a very practical level, comes into play again. Only by customizing the system for specific customers working on real data will make clear the strengths but also weaknesses of the whole complex chain of (semi-) automated object extraction approaches. The latter includes models and strategies as well as their integration with suitable work-flows and real-time tools for user interaction. The overall goals are maximal efficiency and, often even more important, minimal cost.

Because of the large costs, the high risks, the dependence on in-depth knowledge, as well as on specific production environments to be tuned for, practical semi-automated object extraction is and will be in many cases first developed in cooperation of academia and data producers, especially NMA. Only after reasonable success and especially versatility will have been demonstrated, the main DPW developers will probably join in. Yet, even the above cooperation of academia and NMA on a larger scale would be a large achievement, because as (Baltsavias, 2004) notes, at academia there is often a "lack of practical spirit."

5 APPLICATION AREAS

The traditional market for DPW consists of the acquisition of 3D topographic information, such as buildings and roads. DPW have included means to efficiently handle high resolution satellite imagery together with aerial imagery, but also multi- and hyperspectral as well as laser-scanner data. Yet, to our knowledge there is not a strong tendency to integrate also tools to handle terrestrial imagery or close-range laser-scanner data. We think that this is a deficit and we will explain why as well as which additional application areas we see especially concerning vegetation in the remainder of this section.

That close range data is not considered in DPW is in contrast to a recent issue of the IEEE Journal of Computer Graphics and Applications focusing on 3D reconstruction and visualization (Ribarsky and Rushmeier, 2003). The paper starts with the statement "We have entered an era where the acquisition of 3D data is ubiquitous, continuous, and massive." Highly detailed 3D city models from high resolution terrestrial images, dense video sequences, and terrestrial laser-scanner data are seen to be useful for virtual television, tourism, but also mission rehearsal for fire fighting or security and rescue scenarios.

Even though there is one photogrammetric paper (Rottensteiner, 2003) on building extraction from laser-scanner data also in conjunction with aerial imagery in the above IEEE journal issue, the survey on large-scale urban modeling (Hu et al., 2003) shows, that the awareness of the work done in photogrammetry is not too big. As usual, this can be only changed by submitting papers in this area, but also by going to the particular conferences.

Recently, there is a large interest into producing highly detailed 3D city models. One of the first and largest projects in this area is the city-scanning project at MIT (Teller, 1999). Two of the most advanced approaches using images only are (Dick et al., 2002) and (Werner and Zisserman, 2002). (Dick et al., 2002) use advanced statistical modeling in the form of RJMCMC (cf. Section 2.3) allowing for the reconstruction of complete models from partial samples of the object. (Werner and Zisserman, 2002) show what can be achieved assuming that an object is made up of planes (facades or roofs), which are partially vertically oriented, have some parallel structures in front of them (columns) or behind them (windows, doors), and which can be symmetrical (roofs of dormer window).

Terrestrial laser-scanners can be used for very complex historic sites (Allen et al., 2003). Other approaches combine terrestrial and aerial imagery as well as laser-scanner data to produce 3D models with a good fidelity seen from the top but also from the ground (Früh and Zakhor, 2003).

An area where not too much research has been done is the extraction of vegetation outside forests, especially in cities. While it is useful information for city administrations, it is extremely important for the generation of realistic visualizations. (Bacher and Mayer, 2000) use the shadow projection of the tree in an aerial image together with the fact, that the vertical trunk of the tree points to the nadir point. In (Andersen et al., 2002) RJMCMC is used to find trees in aerial laser-scanner DSM employing knowledge about the sensing process as well as the spatial interaction of individual trees modeled by a Markov process. (Straub, 2003) model the shape of trees to extract them from aerial laser-scanner DSM possibly together with reflection properties in the infrared.

Even less work on vegetation extraction has been done in the close range. In (Shlyakhter et al., 2001) the hull of the tree is determined from its silhouette in several images. The 3D medial axis is computed for the hull and from it a representation based on an L-system (Měch and Prusinkiewicz, 1996) is derived. By this means one is able to derive a tree model with which one cannot only deal with occlusions, but which can also be animated, e.g., to simulate wind, and which can be adapted to the seasons. And, the model corresponds closely to the actual tree at the given position.

6 CONCLUSIONS

We have presented a number of issues we think are important to make automated object extraction a part of DPW. These are naturally the models and strategies of the automated processes. To improve them, a thorough testing is needed, promoting also competition between approaches, making clear what way should be taken. Most importantly, though, one should start, or at least start to think about how, to integrate the semi-automated systems into DPW to build efficient systems for practice. Finally, we have shown that automated object extraction offers new possibilities such as highly-detailed 3D models in cities including new objects such as vegetation, which can be animated, e.g., by wind.

ACKNOWLEDGMENTS

We want to thank Christian Heipke for giving us the possibility, but also for pushing us to talk about the issues in this paper. We are grateful to Christian Heipke and Uwe Bacher for proofreading.

REFERENCES

Allen, P., Troccoli, A., Smith, B., Murray, S., Stamos, I. and Leordeanu, M., 2003. New Methods for Digital Modeling of Historic Sites. IEEE Computer Graphics and Applications 23(6), pp. 32–41.

Andersen, H.-E., Reutebuch, S. and Schreuder, G., 2002. Bayesian Object Recognition for the Analysis of Complex Forest Scenes in Airborne Laser Scanner Data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3A, pp. 35–41.

Bacher, P. and Mayer, H., 2000. Automatic Extraction of Trees in Urban Areas from Aerial Imagery. In: International Archives of Photogrammetry and Remote Sensing, Vol. (33) B3/1, pp. 51–57.

Baltsavias, E., 2004. Object Extraction and Revision by Image Analysis Using Existing Geodata and Knowledge: Current Status and Steps Towards Operational Systems. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 129–151.

Bauer, J., Karner, K., Schindler, K., Klaus, A. and Zach, C., 2003. Segmentation of Building Models from Dense 3D Point-Clouds. In: 27th Workshop of the Austrian Association for Pattern Recognition.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W. and Ebner, H., 1999. Automatic Road Extraction Based on Multi-Scale, Grouping, and Context. Photogrammetric Engineering & Remote Sensing 65(7), pp. 777–785.

Benz, U. and Schreier, G., 2001. Definiens Imaging GmbH: Object Oriented Classification and Feature Detection. IEEE Geoscience and Remote Sensing Society Newsletter pp. 16–20.

Blaschke, T. and Hay, G., 2001. Object-Oriented Analysis and Scale-Space: Theory and Methods for Modeling and Evaluating Multiscale Landscape Structure. In: International Archives of Photogrammetry and Remote Sensing, Vol. (34) 4/W5, pp. 22–29.

Brenner, C., 2000. Towards Fully Automatatic Generation of City Models. In: International Archives of Photogrammetry and Remote Sensing, Vol. (33) B3/1, pp. 85–92.

Dick, A., Torr, P. and Cipolla, R., 2002. A Bayesian Estimation of Building Shape Using MCMC. In: Seventh European Conference on Computer Vision, Vol. II, pp. 852–866.

Englert, R., 1998. Acquisition of Complex Model Knowledge by Domain Theory-Controlled Generalization. In: International Archives of Photogrammetry and Remote Sensing, Vol. (32) 3/1, pp. 317–324.

Faugeras, O., 1993. Three-Dimensional Computer Vision. MIT Press, Cambridge, USA.

Faugeras, O. and Luong, Q.-T., 2001. The Geometry of Multiple Images. MIT Press, Cambridge, USA.

Faugeras, O. and Papadopoulo, T., 1997. Grassman-Cayley Algebra for Modeling Systems of Cameras and the Algebraic Equations of the Manifold of Trifocal Tensors. Rapport de Recherche 3225, INRIA, Sophia Antipolis, France.

Fischler, M. and Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24(6), pp. 381–395.

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

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

Fricker, P., 2001. ADS40 – Progress in Digital Aerial Data Collection. In: Photogrammetric Week, Wichmann Verlag, Heidelberg, Germany, pp. 1005–116.

Früh, C. and Zakhor, A., 2003. Constructing 3D City Models by Merging Aerial and Ground Views. IEEE Computer Graphics and Applications 23(6), pp. 52–61.

Gerke, M., Butenuth, M., Heipke, C. and Willrich, F., 2004. Graph-Supported Verification of Road Databases. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 152–165.

Green, P., 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika 82, pp. 711–732.

Growe, S., Schröder, T. and Liedtke, C.-E., 2000. Use of Bayesian Networks as Judgement Calculus in a Knowledge Based Image Interpretation System. In: International Archives of Photogrammetry and Remote Sensing, Vol. (33) B3/1, pp. 342–350.

Gruber, M., Leberl, F. and Perko, R., 2003. Paradigmenwechsel in der Photogrammetrie durch digitale Luftbildaufnahme. Photogrammetrie – Fernerkundung – Geoinformation 4/03, pp. 285–297.

Grün, A. and Li, H., 1994. Semi-Automatic Road Extraction by Dynamic Programming. In: International Archives of Photogrammetry and Remote Sensing, Vol. (30) 3/1, pp. 324–332.

Grün, A. and Wang, X., 2001. News from CyberCity Modeler. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), Balkema Publishers, Lisse, The Netherlands, pp. 93–101.

Gülch, E., Müller, H. and Läbe, T., 1999. Integration of Automatic Processes into Semi-Automatic Building Extraction. In: International Archives of Photogrammetry and Remote Sensing, Vol. (32) 3-2W5, pp. 177–186.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK.

Heipke, C., 1995. State-of-the-Art of Digital Photogrammetric Workstations for Topographic Applications. Photogrammetric Engineering & Remote Sensing 61(1), pp. 49–56.

Heipke, C., Englisch, A., Speer, T., Stier, S. and Kutka, R., 1994. Semi-Automatic Extraction of Roads from Aerial Images. In: International Archives of Photogrammetry and Remote Sensing, Vol. (30) 3/1, pp. 353–360.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O., 1997. Evaluation of Automatic Road Extraction. In: International Archives of Photogrammetry and Remote Sensing, Vol. (32) 3-2W3, pp. 47–56.

Heuel, S., 2001. Points, Lines and Planes and Their Optimal Estimation. In: Pattern Recognition – DAGM 2001, Springer-Verlag, Berlin, Germany, pp. 92–99.

Hinz, A., Dörstel, C. and Heier, H., 2001. DMC – Digital Modular Camera: Systemkonzept und Ablauf der Datenverarbeitung. Photogrammetrie – Fernerkundung – Geoinformation 3/01, pp. 189–198.

Hinz, S., 2003. Integrating Local and Global Features for Vehicle Detection in High Resolution Aerial Imagery. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3/W8, pp. 119–124.

Hu, J., You, S. and Neumann, U., 2003. Approaches to Large-Scale Urban Modeling. IEEE Computer Graphics and Applications 23(6), pp. 62–69.

Jain, A., Duin, P. and Mao, J., 2000. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), pp. 4–37.

Kim, Z. and Nevatia, R., 2003. Expandable Bayesian Networks for 3D Object Description from Multiple Views and Multiple Mode Inputs. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(6), pp. 769–774.

Kimia, B., Tannenbaum, A. and Zucker, S., 1995. Shapes, Shocks, and Deformations I: The Components of Two-Dimensional Shape and the Reaction-Diffussion Space. International Journal of Computer Vision 15(3), pp. 189–224.

Kolmogorov, V. and Zabih, R., 2001. Computing Visual Correspondence with Occlusions Using Graph Cuts. In: Eighth International Conference on Computer Vision, pp. 508–515.

Kosslyn, S., 1994. Image and Brain. MIT Press, Cambridge, USA.

Köthe, U., 1996. Local Appropriate Scale in Morphological Scale-Space. In: Fourth European Conference on Computer Vision, Vol. I, pp. 219–228.

Köthe, U., 2003. Edge and Junction Detection with an Improved Structure Tensor. In: Pattern Recognition – DAGM 2003, Springer-Verlag, Berlin, Germany, pp. 25–32.

Kulschewski, K., 1999. Recognition of Buildings Using a Dynamic Bayesian Network. In: Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI ’99, München, pp. 121–132.

Leberl, F., Gruber, M., Ponticelli, M., Bernoegger, S. and Perko, R., 2003. The Ultracam Large Format Aerial Digital Camera System. In: ASPRS Annual Convention.

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

Lowe, D., 2004. Distintive Image Features from Scale-Invariant Keypoints. Accepted for publication in International Journal of Computer Vision.

Martin, D., Fowlkes, C. and Malik, J., 2004. Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, pp. in press.

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., 2003. Analysis of Means to Improve Cooperative Disparity Estimation. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3/W8, pp. 25–31.

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

Mayer, H., Baumgartner, A. and Steger, C., 1998. Road Extraction from Aerial Imagery. In: CVonline: On-line Compendium of Computer Vision – Applications – Remote Sensing – Road Detection, Extraction and Matching, http://homepages.inf.ed.ac.uk/rbf/CVonline/.

McKeown, D., Cochran, S., Ford, S., McGlone, C., Shufelt, J. and Yocum, D., 1999. Fusion of HYDICE Hyperspectral Data with Panchromatic Imagery for Cartographic Feature Extraction. IEEE Transactions on Geoscience and Remote Sensing 37(3), pp. 1261–1277.

Měch and Prusinkiewicz, P., 1996. Visual Models of Plants Interacting with Their Environment. In: SIGGRAPH '96, pp. 397–410.

Michalski, R., Carbonell, J. and Mitchell, T., 1984. Machine Learning – An Artificial Intelligence Approach. Springer-Verlag, Berlin, Germany.

Michalski, R., Carbonell, J. and Mitchell, T., 1986. Machine Learning II – An Artificial Intelligence Approach. Springer-Verlag, Berlin, Germany.

Mikhail, E., 2000. Multi-Source Feature Extraction and Visualization in Urban Environments. In: International Archives of Photogrammetry and Remote Sensing, Vol. (33) B3/2, pp. 592–607.

Ney, H., 1999. Stochastic Modelling: From Pattern Classification to Speech Recognition and Language Translation. In: Mustererkennung 1999, Springer-Verlag, Berlin, Germany, pp. 325–337.

Niemann, H., Sagerer, G., Schröder, S. and Kummert, F., 1990. ERNEST: A Semantic Network System for Pattern Understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(9), pp. 883–905.

Nistér, D., 2003. An Efficient Solution to the Five-Point Relative Pose Problem. In: Computer Vision and Pattern Recognition, Vol. II, pp. 195–202.

Pollefeys, M. and Van Gool, L., 1999. Stratified Self-Calibration with the Modulus Constraint. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8), pp. 707–724.

Pollefeys, M., Verbiest, F. and Van Gool, L., 2002. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In: Seventh European Conference on Computer Vision, Vol. II, pp. 837–851.

Ribarsky, A. and Rushmeier, H., 2003. 3D Reconstruction and Visualization. IEEE Computer Graphics and Applications 23(6), pp. 20–21.

Rosenhahn, B. and Sommer, G., 2002. Adaptive Pose Estimation for Different Corresponding Entities. In: Pattern Recognition – DAGM 2002, Springer-Verlag, Berlin, Germany, pp. 265–273.

Rottensteiner, F., 2003. Automatic Generation of High Quality Building Models from Lidar Data. IEEE Computer Graphics and Applications 23(6), pp. 42–50.

Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision 47(1), pp. 7–42.

Schenk, T. and Csathó, B., 2002. Fusion of LIDAR Data and Aerial Imagery for a More Complete Surface Description. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3A, pp. 310–317.

Schmidt, R. and Brand, R., 2003. Automatic Determination of Tie Points for HRSC on Mars Express,. In: ISPRS / EARSeL Joint Workshop on High Resolution Mapping from Space.

Shlyakhter, I., Rozenoer, M., Dorsey, J. and Teller, S., 2001. Reconstructing 3D Tree Models from Instrumented Photographs. IEEE Computer Graphics and Applications 21(3), pp. 53–61.

Shufelt, J., 1999. Performance Evaluation and Analysis of Monocular Building Extraction from Aerial Imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(4), pp. 311–326.

Stoica, R., Descombes, X. and Zerubia, J., 2004. A Gibbs Point Process for Road Extraction from Remotely Sensed Images. International Journal of Computer Vision 57(2), pp. 121–136.

Straub, B.-M., 2003. Automatic Extraction of Trees from Aerial Images and Surface Models. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3/W8, pp. 157–164.

Teller, S., 1999. Automated Urban Model Acquisition: Project Rationale and Status. Bulletin Société Française de Photogramétrie et Télédétection (SFPT) 153, pp. 56–63.

Torr, P., 1997. An Assessment of Information Criteria for Motion Model Selection. In: Computer Vision and Pattern Recognition, pp. 47–53.

Weidner, U. and Förstner, W., 1995. Towards Automatic Building Extraction from High Resolution Digital Elevation Models. ISPRS Journal of Photogrammetry and Remote Sensing 50, pp. 38–49.

Werner, T. and Zisserman, A., 2002. New Techniques for Automated Architectural Reconstruction from Photographs. In: Seventh European Conference on Computer Vision, Vol. II, pp. 541–555.

Zhang, C., 2004. Towards an Operational System for Automated Updating of Road Databases by Integration of Imagery and Geodata. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 166–186.