Available online at www.sciencedirect.com

ScienceDirect

ISPRS Journal of Photogrammetry & Remote Sensing 63 (2008) 213–222

PHOTOGRAMMETRY
& REMOTE SENSING

www.elsevier.com/locate/isprsjprs

Object extraction in photogrammetric computer vision☆

Helmut Mayer *

Institute of Photogrammetry and Cartography, Bundeswehr University Munich, D-85577 Neubiberg, Germany

Received 29 August 2006; received in revised form 25 July 2007; accepted 27 August 2007
Available online 29 October 2007

Abstract

This paper discusses state and promising directions of automated object extraction in photogrammetric computer vision considering also practical aspects arising for digital photogrammetric workstations (DPW). A review of the state of the art shows that there are only few practically successful systems on the market. Therefore, important issues for a practical success of automated object extraction are identified. A sound and most important powerful theoretical background is the basis. Here, we particularly point to statistical modeling. Testing makes clear which of the approaches are suited best and how useful they are for praxis. A key for commercial success of a practical system is efficient user interaction. 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.

© 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Automatic object extraction; Digital photogrammetry; Semi-automation; Modeling; Statistics; Testing

1. Introduction

It took a few decades to highly automate (i.e., minimize human work) orientation determination and the generation of digital surface models (DSM) or digital elevation models (DEM). This has led to digital photogrammetric workstations (DPW) (Heipke, 1995), which 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. Compared to this, the situation is much more difficult for object extraction. There are only few successful (semi-) automated systems 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, particularly, WIPKA-QS (Gerke et al., 2004) are on the verge of becoming operational.

This paper addresses reasons for this deficit of viable practical systems, but also points on issues we consider important to improve the situation and introduce object extraction on a larger scale also in practical applications. To begin with, we show how the difficulties of object extraction have been underestimated in (photogrammetric) computer vision from the very beginning but also point to recent developments in this context. While some of the latter are mostly important only for close range applications, which we see as an evolving market for DPW (cf. Section 4), advances in the exploitation of

---

☆ Revised and extended version of a Commission II invited paper on "Object Extraction for Digital Photogrammetric Workstations" presented at the ISPRS congress, Istanbul, Turkey, 12–23 July 2004.

* Tel.: +49 89 6004 3429; fax: +49 89 6004 4090.
E-mail address: Helmut.Mayer@unibw.de.
URL: http://www.unibw.de/ipk.

0924-2716/$ - see front matter © 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2007.08.008

redundancy and in stereo matching are of importance for DSM/DEM generation in topographic applications.

Legend has it, that in the 1950s 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 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 with the milestone "cookbook" (Hartley and Zisserman, 2003), but still we might be only at the beginning of understanding the basic problems.

There is progress not only in the high level understanding, i.e., interpretation, area, but also in the modeling of the image function. Köthe (2003) has for instance shown that the well known operator of Förstner and Gülch (1987) does not take into account the frequency doubling implicit in the squaring of the Hessian matrix. For detailed structures this can lead to missing exactly those points one is interested in or to a bad localization of the points. The SIFT operator of Lowe (2004) offers scale and rotation-invariant features which can be robustly matched under affine distortion, noise, and illumination changes, largely extending the scope of matching procedures. One particularly relevant example using it is the commercial "Autostitch" program for the construction of panoramas insensitive to the ordering, orientation, scale and illumination of the employed images (Brown and Lowe, 2003).

Pollefeys et al. (2004) have shown that it is possible to fully automatically reconstruct the pose and calibrate images of cameras of which the only thing known is, that they are perspective. This opens up new application areas particularly in close range and gives additional flexibility. Pollefeys et al. (2004) also demonstrated the importance of redundancy in matching, an issue recently propagated by Leberl and Thurgood (2004) for robust DSM/DTM generation from images of digital aerial cameras, claiming that one can obtain results with a quality similar or even superior to laser scanning. Nistér (2004) presents a direct solution for the five-point relative orientation problem allowing for real-time orientation without approximate values by making use of given calibration information. Particularly the possibility to generate approximate values is very helpful for close range as it allows for full automation also for a freely moving camera without any markers. 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 of Kolmogorov and Zabih (2001), or cooperative disparity estimation as in (Mayer, 2006), opening ways for obtaining meaningful DSM also in complex urban areas.

This paper rests on a recent survey of Baltsavias (2004) which summarizes important points for the practical use of object extraction. Our goal is to deepen some points, yet give enough overview of the area to make the paper self-contained. Although focusing on aerial imagery and aerial laser-scanner data, we also deal with satellite imagery, hyper spectral data, and terrestrial video sequences or laser-scanner data. To limit the scope, we do not consider radar data.

The prerequisite for highly productive object extraction is appropriate modeling (cf. Section 2), which in our case comprises 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 with the visual presentation of specific examples, there is a recent tendency to evaluate the performance of the approaches by means of different tests giving way to the design of the user interaction for semi-automated systems described in Section 3. As technical developments are useless without markets, Section 4 gives 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 for the performance of any approach for automated or semi-automated object extraction. Basically, modeling consists of knowledge about the objects to be extracted. Additionally, in most cases it is necessary to analyze also 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 might only find them, when one knows, where the cars are (Hinz, 2003).

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

2.1. Strategy and scale

Experience shows, that the sequence of operations employing the knowledge about the objects and their relations is extremely important for an efficient, but also successful extraction. It is for instance well known, that markings are an important clue to find roads. Unfortunately, in images with a ground pixel size of about 0.25 m markings very often correspond to very faint bright lines. Extracting the latter in open rural space results in 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 of for instance 2 m, i.e., images on 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 the markings can be extracted and grouped reliably (Hinz and Baumgartner, 2003), giving the hypotheses a high evidence for actually being roads.

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

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

The basic 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, it is not always easy to identify them.

In the above example on roads, scale plays an important role. Coarse to fine approaches have long been used for orientation determination and in image matching (Heipke, 1995). For linear objects it was shown in (Mayer and Steger, 1998), that by changing scale from fine to coarse by means of the 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 for instance 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, elimination of 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. In (Mayer and Steger, 1998) we give an example for the application of the latter on road extraction, while in (Mayer, 1998) we show that this is advantageous because it preserves the elongatedness of roads while at the same time suppressing most other objects.

2.2. Data sources and GIS data

In recent years, DPW have included means to deal with high resolution satellite imagery such as IKONOS or Quickbird together with aerial imagery, possibly digital, for instance from Leica's ADS40 (Fricker, 2001), Vexcel's Ultracam (Leberl et al., 2003), or Z/I imaging's DMC (Madani et al., 2004). The advantages of digital imagery including economic aspects are highlighted in (Leberl and Thurgood, 2004).

To use data which comprise information more explicit for the problem is a very efficient means to make the 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) and 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 information.

Concerning 3D, often highly reliable data from laser scanners are the data source of choice. Early experiments on the extraction of buildings from laser-scanner data where conducted 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 scheme. In Straub (2004) DSM data from laser scanners partially together with reflection properties in the infrared is used for the extraction of individual trees.

A very important source often neglected in more theoretical work are GIS data. Brenner (2000) uses two dimensional (2D) polygons, from which straight skeletons are generated, 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 in a way that relaxed parameters can be used leading to a more complete verification and therefore a higher efficiency. Zhang (2004) employs color and stereo data together with extensive modeling, comprising, e.g., context, occlusions, and shadows. He is making heavy, yet intelligent use of given GIS data, leading to an impressive performance for the update of road data.

One reason that the potential of using GIS information is not exploited is, that one cannot absolutely rely on it, as it might be outdated and unprecise and therefore might lead to wrong conclusions. There is always a trade-off between accepting wrong, because changed objects, and rejecting correct, because unchanged objects. Therefore, as for instance (Zhang, 2004) demonstrates, even when using additional information from GIS, reality needs to be modeled very deeply, so that the objects can be extracted reliably also under complex circumstances. Additional reasons hindering the widespread use of GIS data are the plethora of data models and formats and the different policies and legal requirements of cities, regions, and countries.

2.3. Statistical modeling

The deficits of a mainly deterministic modeling, for instance based on semantical networks (Niemann et al., 1990), have been known for a long time. There have been heuristic attempts by adding for instance believe values. Yet, more sound ways of including statistical modeling have been proposed for object extraction only recently, for instance Bayesian networks in (Growe et al., 2000) or (Kim and Nevatia, 2003). The work on dynamic Bayesian networks of Kulschewski (1999) has been interesting in terms of statistically 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. Taillandier and Deriche (2004) proposed a Bayesian approach penalizing complex models via minimum description length (MDL). A very generic model with only a few constraints is employed with MDL being used to balance good fit to the data with the need to generalize according to architectural models much more abstract than the images. The potential of this approach is shown with impressive results for 45 buildings in a block of six images.

Until recently, semantical modeling was also often 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 difficult to judge how much a component contributed to the results.

By the advent of Reversible Jump (RJ) Markov Chain Monte Carlo (MCMC) (Green, 1995) there is a means for statistical modeling which can also be used in a generative way 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 by being able to change the number of parameters, which is not feasible for standard MCMC Neal (1993). The jumps are called reversible, because 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 building facade interpretation of Dick et al. (2004) as well as Mayer and Reznik (2007), road extraction by Stoica et al. (2004), and vegetation extraction by Andersen et al. (2002). Dick et al. (2004) and Stoica et al. (2004) 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.

Another important 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, automation in vision often 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 would do this. Especially popular is the random sample consensus, or short RANSAC approach of Fischler and Bolles (1981) and its variants for instance based on the geometric information criterion (GRIC) of Torr (1997).

The basic idea is to take a larger number of random samples consisting of the (assumed to be small) minimum number of observations necessary to solve the problem. All these samples lead to solutions which are then checked with a given acceptance criterion 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, 2003), for aero-triangulation (Schmidt and Brand, 2003), or to find planes (and also spheres and cones) in a large number of 3D points (Bauer et al., 2003; Schnabel et al., 2006), as it can deal with data of which only a few percent are correct, far below the breakdown point of 50% for (traditional) robust statistical approaches such as M-estimators. This is feasible, because of the restrictive assumptions of a small minimum number of parameters needed to solve the problem as well as of a given acceptance criterion to decide if an observation is correct.

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 (Faugeras and Luong, 2001) and (Hartley and Zisserman, 2003). Heuel (2004) has presented work where statistics is linked with geometric algebras making it possible to statistically test uncertain geometric relations such as incidence, parallelism or orthogonality, which is used for the reconstruction of buildings in the form of polyhedra.

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 from 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 (topological) structure, and their geometry (Englert, 1998).

Unfortunately, learning is, after standard textbooks have been introduced a long while ago (Michalski et al., 1984, 1986), still not advanced enough to reliably 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 well understood.

Also for learning statistics might come to help. Hidden Markov Models (HMM) have made possible a break-through 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 to be sufficient to analyze the statistical dependencies of very few neighboring words based on HMM (Ney, 1999).

For image understanding the higher complexity makes learning much more difficult. Yet, recently, inspired by the success of the statistical learning of local relations for text, approaches based on learning the appearance of small image patches and their mutual spatial relations have given means to extract objects such as cars (Agarwal et al., 2004). In Leibe and Schiele (2004) the segmentation of the detected objects as well as scale-invariance was introduced. The system of Fei-Fei et al. (2006) based on Bayesian decisions can learn more than one hundred object categories incrementally from a small number of training examples. The Bayesian decisions allow to improve recognition by integrating a given prior from earlier experiments, possibly done on other, unrelated objects.

Concerning another popular means for learning, namely artificial neural networks, additionally to neuro-fuzzy extensions used for 3D object recognition and reconstruction as in (Samadzadegan et al., 2005), we particularly refer to the discussion in a recent survey on statistical pattern recognition by 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 and user interaction

Key factors determining the practical usefulness of a system are thorough testing as well as an optimized user interaction. Yet, testing 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 strongly motivates all people involved. By trying to outperform other approaches one learns much about the possibilities but also the limits of one's own 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 compared to manual approaches. It depends on many factors some of them needing lots of effort for optimization if the real potential of an approach is to be revealed. But also for automated approaches there is a large number of factors which influence the test and, therefore, also which approaches perform well and which not. For roads for instance the preferred characteristics of the terrain plays an important role.

Our experience shows that for many applications two basic measures are suitable for evaluation of the test results, namely "correctness" and "completeness" (Heipke et al., 1997).

For approaches relevant for DPW, testing must be done against real world data. If the goal is to evaluate the whole production chain, ground truth data should be gathered from the 3D reality. In many cases one just wants to know how much worse than a human the automated system is. Then, benchmarking against manually digitized ground truth data is the way to go. To avoid the bias of individual operators, 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 set up a test on "Automated extraction, refinement, and update of road databases from imagery and other data" under the umbrella of EuroSDR (European spatial data research — http://www.eurosdr.net; formerly known as OEEPE). On one hand, we wanted to learn the data specification needs of important data producers, mainly national mapping and cadastral agencies (NMCA) and their customers. On the other hand, existing semi- and fully-automated systems for road extraction have been evaluated based on high quality image data against given, manually digitized ground truth data (Mayer et al., 2006a,b).

After thoroughly testing an approach, the next step towards an operational system is often to design an optimized user interaction. To limit the scope, we do not deal here 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, we will not treat it here either.

For general purpose DPW as well as GIS, functionality for automated 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 restricted to building extraction.

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 additional ideas why this is the case, but we will also point on ways how to change this situation.

Basically, as pointed out above, automating 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. above). 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 in industry 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 used by the system.

With this we mean, that to build a highly effective interactive system, the interaction needs to be tailored for a fixed level of extraction. If the level of extraction improves, it is not unlikely, that the interaction of the system will have to be considerably different, implying larger changes to 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 without user interaction is a restriction to problems where the computer assists the user on-line. This is the case for InJect, but only partly for CC-Modeler. For roads, this idea has been promoted early (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, roads are so difficult to extract, that only fully-automated offline processing can be imagined to 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 fully-automated offline processes, probably preceded or interrupted, but in any case followed by manual interaction. The generation of work-flows defining the offline-phases, but also very importantly the information to be given to the automated procedure by user interaction preceding it, is essential for the overall performance.

It is often more costly in terms of user interaction time to correct complex failures, than to manually acquire a situation from scratch. 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 slightly different from the ideas presented on statistical modeling above (cf. Section 2.3) as it makes use of additional knowledge about the strengths and weaknesses of human interaction. For a semi-automated system the correct objects (green) have to be actually correct with a probability sufficient for the application, 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. Also results with a high likelihood of complex topologic errors should not be presented to the operator. Helpful might be, though, to offer a small number of choices, one of which is with relatively high probability correct.

An efficient semi-automated system could comprise real-time tools, which help to improve the results obtained fully automatically. A good way, yet needing much effort to implement and again depending on the current state of an automated system, is to make use of the results of automated extraction.

Eventually, testing, this time on a very practical level, comes into play again. Only by customizing the system for specific customers will make clear the strengths but also weaknesses of the whole complex chain of semi-automated object extraction. 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 NMCA. The main DPW developers will probably join in only after reasonable success and especially versatility will have been demonstrated. The above cooperation of academia and NMCA on a larger scale would be a large achievement, because as Baltsavias (2004) notes, at academia there is often a "lack of practical spirit."

4. Application areas

Recently, DPW have included means to efficiently handle high resolution satellite imagery, multi- and hyperspectral, as well as laser-scanner data together with aerial imagery. There is also an interest to integrate tools to handle terrestrial imagery. We give examples for current work in this area which shows the potential particularly for highly-detailed 3D city models possibly including vegetation.

In the ISPRS Istanbul congress there has been considerable interest into detailed 3D visualization. Haala (2004) deals with the orientation of a panoramic sensor with a large field of view and high resolution to efficiently obtain texture for given 3D building data including its automation based on approximate values for orientation from GPS and an inertial sensor. Böhm (2004) presents the generation of occlusion free texture for facades by removing moving and also static objects given a small number of images based on background estimation. In (von Hansen et al., 2004) it is shown how given planar facades can be enriched with depth information by guided matching in several terrestrial images to improve the realism of visualization. Jülge and Brenner (2004) present a semi-automatic process for the extraction of window hypotheses from terrestrial laser-scanner data, e.g., suitable for the registration of different scans.

In the vision and graphics area a recent issue of the IEEE Journal of Computer Graphics and Applications edited by Ribarsky and Rushmeier (2003) focuses on 3D reconstruction and visualization. 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 by 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 by Hu et al., (2003) shows, that the awareness of the work done in photogrammetry is not big. As usual, this can only be changed by submitting papers in this area, but also by going to the particular conferences.

One of the first and largest projects in the area of the production of highly-detailed 3D city models from terrestrial images and laser-scanner data is the city-scanning project at MIT initiated by Teller (1999). Some of the most advanced approaches using images only are (Dick et al., 2004), (Werner and Zisserman, 2002), and (Mayer and Reznik, 2006, 2007). Dick et al. (2004) use advanced statistical modeling in the form of RJMCMC (cf. Section 2.3) allowing for the faithful reconstruction of complete models from samples of parts 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 (e.g., the two roofs of a dormer window). Mayer and Reznik (2006, 2007) derive facade planes from 3D points, detect and delineate windows via an appearance based approach (Leibe and Schiele, 2004), and estimate their depth based on plane sweeping (Werner and Zisserman, 2002). Other approaches such as (Früh and Zakhor, 2003) 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.

A complementary area is the extraction of vegetation in cities. While it is useful information for city administrations, it is extremely important for the generation of realistic visualizations. In Andersen et al. (2002) RJMCMC is used to find trees in aerial laser-scanner DSM employing knowledge about the spatial interaction of individual trees. Straub (2004) models the shape of trees to extract them from aerial laser-scanner DSM possibly together with reflection properties in the infrared. Shlyakhter et al. (2001) obtain tree models from terrestrial image sequences based on an L-System (Měch and Prusinkiewicz, 1996) which can be animated for instance by wind, and adapted to the seasons. In (Huang and Mayer, 2007) MCMC is used in conjunction with L-systems to reconstruct the 3D branching structure from terrestrial image sequences of unfoliaged deciduous trees.

5. Conclusions

We have presented a number of issues we consider important to make automated object extraction become a part of DPW. These are naturally the models and strategies of the automated processes. To improve them, thorough testing is needed, promoting 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. We have also shown that automated object extraction offers new possibilities such as highly-detailed 3D models in cities including additional objects such as vegetation.

Finally, for practical applications we recommend to consider the following three techniques, which overcome limits in their respective areas and have been widely found to work reliably for a broad range of situations: The SIFT operator (Lowe, 2004) allows robust scale- and rotation-invariant point extraction and matching, extending the range of problems where automatic point correspondences can be used. The 5-point algorithm (Nistér, 2004) for calibrated direct orientation gives approximate values for relative orientation under many circumstances. RANSAC (Fischler et al., 1981) is a versatile tool for robust estimation even for well below 50% correct data. It is suitable for many problems with few parameters and for which a (local) acceptance criterion for correct data can be defined, such as image orientation or the determination of planes, spheres, and cones from 3D point clouds.

Acknowledgments

We want to thank Christian Heipke for giving us the opportunity to write down our ideas on this subject. We are grateful to the anonymous reviewers for their helpful comments.

References

Agarwal, S., Awan, A., Roth, D., 2004. Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11), 1475–1490.

Andersen, H.-E., Reutebuch, S., Schreuder, G., 2002. Bayesian object recognition for the analysis of complex forest scenes in airborne laser scanner data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3A), 35–41.

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 (3), 129–151.

Bauer, J., Karner, K., Schindler, K., Klaus, A., Zach, C., 2003. Segmentation of building models from dense 3D point-clouds. 27th Workshop of the Austrian Association for Pattern Recognition, pp. 253–258.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W., Ebner, H., 1999. Automatic road extraction based on multi-scale, grouping, and context. Photogrammetric Engineering & Remote Sensing 65 (7), 777–785.

Benz, U., Schreier, G., 2001. Definiens imaging GmbH: object oriented classification and feature detection. IEEE Geoscience and Remote Sensing Society Newsletter 9, 16–20.

Blaschke, T., Hay, G., 2001. Object-oriented analysis and scale-space: theory and methods for modeling and evaluating multiscale landscape structure. International Archives of Photogrammetry and Remote Sensing 34 (Part 4/W5), 22–29.

Böhm, J., 2004. Multi image fusion for occlusion-free façade texturing. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B5), 867–872.

Brenner, C., 2000. Towards fully automatic generation of city models. International Archives of Photogrammetry and Remote Sensing 33 (Part B3/1), 85–92.

Brown, M., Lowe, D., 2003. Recognising panoramas. Ninth International Conference on Computer Vision II, pp. 1218–1225.

Dick, A., Torr, P., Cipolla, R., 2004. Modelling and interpretation of architecture from several images. International Journal of Computer Vision 60 (2), 111–134.

Englert, R., 1998. Acquisition of complex model knowledge by domain theory-controlled generalization. International Archives of Photogrammetry and Remote Sensing 32 (Part 3/1), 317–324.

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

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

Fei-Fei, L., Fergus, R., Perona, P., 2006. One-shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (4), 594–611.

Fischler, M., 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), 381–395.

Fischler, M., Tenenbaum, J., Wolf, H., 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. Computer Graphics and Image Processing 15 (3), 201–223.

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.

Förstner, W., Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data. Interlaken, Switzerland, pp. 281–305.

Fricker, P., 2001. ADS40 — progress in digital aerial data collection. Photogrammetric Week. Wichmann Verlag, Heidelberg, pp. 1005–1116.

Früh, C., Zakhor, A., 2003. Constructing 3D city models by merging aerial and ground views. IEEE Computer Graphics and Applications 23 (6), 52–61.

Gerke, M., Butenuth, M., Heipke, C., Willrich, F., 2004. Graph-supported verification of road databases. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3), 152–165.

Green, P., 1995. Reversible jump Markov Chain Monte Carlo Computation and Bayesian Model determination. Biometrika 82 (4), 711–732.

Growe, S., Schröder, T., Liedtke, C.-E., 2000. Use of Bayesian Networks as judgement calculus in a knowledge based image interpretation system. International Archives of Photogrammetry and Remote Sensing 33 (Part B3/1), 342–350.

Grün, A., Li, H., 1994. Semi-automatic road extraction by dynamic programming. International Archives of Photogrammetry and Remote Sensing 30 (Part 3/1), 324–332.

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

Gülch, E., Müller, H., Läbe, T., 1999. Integration of automatic processes into semi-automatic building extraction. International Archives of Photogrammetry and Remote Sensing 32 (Part 3–2W5), 177–186.

Haala, N., 2004. On the refinement of urban models by terrestrial data collection. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 564–569.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision — Second Edition. 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), 49–56.

Heipke, C., Englisch, A., Speer, T., Stier, S., Kutka, R., 1994. Semi-automatic extraction of roads from aerial images. International Archives of Photogrammetry and Remote Sensing 30 (Part 3/1), 353–360.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. International Archives of Photogrammetry and Remote Sensing 32 (Part 3-2W3), 47–56.

Heuel, S., 2004. Uncertain projective geometry: statistical reasoning for polyhedral object recognition. Vol. 3008 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany.

Hinz, S., 2003. Integrating local and global features for vehicle detection in high resolution aerial imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3/W8), 119–124.

Hinz, S., Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. ISPRS Journal of Photogrammetry and Remote Sensing 58 (1), 83–98.

Hu, J., You, S., Neumann, U., 2003. Approaches to large-scale urban modeling. IEEE Computer Graphics and Applications 23 (6), 62–69.

Huang, H., Mayer, H., 2007. Extraction of the 3D branching structure of unfoliaged deciduous trees from image sequences. Photogrammetrie – Fernerkundung – Geoinformation 2007 (6), 429–436.

Jain, A., Duin, P., Mao, J., 2000. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1), 4–37.

Jülge, K., Brenner, C., 2004. Object extraction from terrestrial laser scan data for a detailed building description. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 1079–1084.

Kim, Z., 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), 769–774.

Kimia, B., Tannenbaum, A., Zucker, S., 1995. Shapes, shocks, and deformations I: the components of two-dimensional shape and the reaction–diffusion space. International Journal of Computer Vision 15 (3), 189–224.

Kolmogorov, V., Zabih, R., 2001. Computing visual correspondence with occlusions using graph cuts. 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. Fourth European Conference on Computer Vision I, pp. 219–228.

Köthe, U., 2003. Edge and junction detection with an improved structure tensor. Pattern Recognition — DAGM 2003. Springer-Verlag, Berlin, Germany, pp. 25–32.

Kulschewski, K., 1999. Recognition of buildings using a dynamic bayesian network. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI '99. Department for Photogrammetry, University of Bonn, Germany, pp. 121–132.

Leberl, F., Gruber, M., Ponticelli, M., Bernoegger, S., Perko, R., 2003. The ultracam large format aerial digital camera system. Proceedings ASPRS Annual Convention (on CD-ROM).

Leberl, F., Thurgood, J., 2004. The Promise of softcopy photogrammetry revisited. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 759–763.

Leibe, B., Schiele, B., 2004. Scale-invariant object categorization using a scale-adaptive mean-shift search. Pattern Recognition — DAGM 2004. Springer-Verlag, Berlin, Germany, pp. 145–153.

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

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2), 91–110.

Madani, M., Dörstel, C., Heipke, C., Jacobsen, K., 2004. DMC practical experience and accuracy assessment. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B2), 396–401.

Martin, D., Fowlkes, C., 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 (5), 530–549.

Mayer, H., 1998. Automatische Objektextraktion aus digitalen Luftbildern. Deutsche Geodätische Kommission, series C, vol. 494. C.H. Beck'sche Verlagsbuchhandlung, München, Germany.

Mayer, H., 1999. Automatic object extraction from aerial imagery — a survey focusing on buildings. Computer Vision and Image Understanding 74 (2), 138–149.

Mayer, H., 2006. Cooperative disparity estimation and its improvement. Photogrammetrie – Fernerkundung – Geoinformation 2/06, pp. 139–151.

Mayer, H., Bacher, U., Baltsavias, E., 2006a. Automated extraction, refinement, and update of road databases from imagery and other data. European Spatial Data Research (EuroSDR) Official Publication vol. 50, 217–280.

Mayer, H., Hinz, S., Bacher, U., Baltsavias, E., 2006b. A test of automatic road extraction approaches. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3), 209–214.

Mayer, H., Baumgartner, A., Steger, C., 1998. Road extraction from aerial imagery. CVonline: on-line compendium of computer vision–applications–remote sensing–road detection, extraction and matching. http://homepages.inf.ed.ac.uk/rbf/CVonline1998/ (Last accessed September 4, 2007).

Mayer, H., Reznik, S., 2006. MCMC linked with implicit shape models and plane sweeping for 3D building facade interpretation in image sequences. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3), 130–135.

Mayer, H., Reznik, S., 2007. Building facade interpretation from uncalibrated wide-baseline image sequences. ISPRS Journal of Photogrammetry and Remote Sensing 61 (6), 371–380.

Mayer, H., Steger, C., 1998. Scale-space events and their link to abstraction for road extraction. ISPRS Journal of Photogrammetry and Remote Sensing 53 (2), 62–75.

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

Měch, Prusinkiewicz, P., 1996. Visual models of plants interacting with their environment. SIGGRAPH '96, pp. 397–410.

Michalski, R., Carbonell, J., Mitchell, T., 1984. Machine Learning — an Artificial Intelligence Approach. Springer-Verlag, Berlin, Germany.

Michalski, R., Carbonell, J., 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. International Archives of Photogrammetry and Remote Sensing 33 (Part B3/2), 592–607.

Neal, R., 1993. Probabilistic inference using Markov Chain Monte Carlo methods. Technical Report CRG-TR-93-1. Department of Computer Science, University of Toronto.

Ney, H., 1999. Stochastic modelling: from pattern classification to speech recognition and language translation. Mustererkennung 1999. Springer-Verlag, Berlin, Germany, pp. 325–337.

Niemann, H., Sagerer, G., Schröder, S., Kummert, F., 1990. ERNEST: a semantic network system for pattern understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (9), 883–905.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (6), 756–770.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., 2004. Visual modeling with a hand-held camera. International Journal of Computer Vision 59 (3), 207–232.

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

Rottensteiner, F., 2003. Automatic generation of high quality building models from LIDAR data. IEEE Computer Graphics and Applications 23 (6), 42–50.

Samadzadegan, F., Azizi, A., Hahn, M., Lucas, C., 2005. Automatic 3D object recognition and reconstruction based on neuro-fuzzy modelling. ISPRS Journal of Photogrammetry and Remote Sensing 59 (5), 255–277.

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47 (1), 7–42.

Schenk, T., Csathó, B., 2002. Fusion of LIDAR data and aerial imagery for a more complete surface description. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3A), 310–317.

Schmidt, R., Brand, R., 2003. Automatic determination of tie points for HRSC on Mars express. ISPRS/EARSeL Joint Workshop on High Resolution Mapping from Space (on CD-ROM).

Schnabel, R., Wahl, R., Klein, R., 2006. Shape detection in point clouds. Tech. Rep. CG-2006-2. Universität Bonn, Bonn, Germany, January.

Shlyakhter, I., Rozenoer, M., Dorsey, J., Teller, S., 2001. Reconstructing 3D tree models from instrumented photographs. IEEE Computer Graphics and Applications 21 (3), 53–61.

Stoica, R., Descombes, X., Zerubia, J., 2004. A Gibbs point process for road extraction from remotely sensed images. International Journal of Computer Vision 57 (2), 121–136.

Straub, B.-M., 2004. A multiscale approach for the automatic extraction of tree tops from remote sensing data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 418–423.

Taillandier, F., Deriche, R., 2004. Automatic building reconstruction from aerial images: a generic bayesian framework. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 343–348.

Teller, S., 1999. Automated Urban model acquisition: project rationale and status. Bulletin Société Française de Photogramétrie et Télédétection (SFPT) 153 (1), 56–63.

Torr, P., 1997. An assessment of information criteria for motion model selection. Computer Vision and Pattern Recognition, pp. 47–53.

von Hansen, W., Thönnessen, U., Stilla, U., 2004. Detailed relief modeling of building facades from video sequences. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 967–972.

Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high resolution digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 50 (4), 38–49.

Werner, T., Zisserman, A., 2002. New techniques for automated architectural reconstruction from photographs. Seventh European Conference on Computer Vision 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 (3), 166–186.