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.

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