ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 193–203

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A graph cut optimization guided by 3D-features for surface height recovery

N. Chehata [a,b,*], F. Jung [c], G. Stamon [d]

[a] Institut EGID, Université of Bordeaux, GHYMAC Lab: 1 allée F.Daguin, 33607 Pessac, France
[b] Institut Géographique National, MATIS lab: 2 av pasteur, 94165 St Mandé, France
[c] DDE de Seine maritime, Cité administrative Saint Sever 76032 Rouen, France
[d] Université Paris Descartes, CRIP5-SIP lab: 45 rue des saints-pères 75006 Paris, France

ARTICLE INFO

Article history:
Received 21 July 2006
Received in revised form
8 October 2008
Accepted 14 October 2008
Available online 2 December 2008

Keywords:
Stereoscopic
Urban
High resolution satellite images
Surface modeling
Graph cut optimization

ABSTRACT

This paper aims to present a new approach for automatic urban scene modeling from high-resolution satellite images with focus on building areas. The input data consist of a panchromatic stereo pair of satellite images, with a submetric resolution of 50–70 cm and a low Base to Height ratio B/H [0.05–0.2]. Since a detailed extraction and description of building roofs is complex in a satellite context, we propose to describe the scene by means of a 3D-surface that provides either raster or vector information using different description levels. The main contribution of our approach is the use of 3D-features such as 3D-segments and 3D-facets to guide the optimization process. 3D surface modeling can be formulated as a matching problem that can be solved by graph cut minimization. The novelty consists in the original construction of the graph to combine input 2D data and 3D feature constraints to control the final surface. Complementary features are used. 3D-segments modelize discontinuities and 3D-facets help to regularize the surface by planar patches. The proposed automatic system provides a surface height map with subpixellar precision. Moreover, the system is generic and extensible to other data such as aerial and terrestrial images or to a multiple view context. External databases can also be easily added to the process to constrain the optimization.

© 2008 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

1.1. Context

Photogrammetric building reconstruction from aerial and satellite stereopairs of images occupies a prominent place in cartography and remote sensing. These last years, the launching of submetric resolution commercial satellites [Ikonos, Eros, QuickBird] has provided new scale satellite images whose resolution ranges from 60 cm to 1 m. Consequently, building reconstruction from satellite images represents a new challenge and has been of more and more interest to the scientific community.

For this study, the input data consists of a panchromatic stereopair of satellite images, at a very high resolution of 50–70 cm and a low Base to Height ratio B/H ranging between [0.05–0.2]. It is assumed that both geometry and stereo system calibration are already determined. Hence, the correspondence between epipolar lines in the two images is supposed to be known.

This paper is focused on urban scene 3D-surface modeling from a very high-resolution satellite stereopair. Our system handles only building roofs since facade extraction is too complex in satellite context and especially with a low B/H ratio. Moreover, some simplifications of shapes have to be done and are acceptable when taking into account relatively low image resolution. Consequently, dormer windows, chimneys or building recesses will not be treated. In satellite context, a low B/H ratio has the advantage of reducing geometric distortions between images and then makes image matching easier. However, it leads to poor 3D altimetric precision. Furthermore, an automatic system in stereoscopic context has to deal with discontinuities, occlusions and presence of vegetation. These difficulties make a 3D-polyhedral building model reconstruction too complex. The goal is to provide an automatic and a generic method for 3D-surface modeling concerning roofs and ground surface. The 3D-surface has to be dense for applications such as virtual tourism or orthophoto production.

1.2. General scheme

In the last decade, automatic building reconstruction, especially on aerial images, has received much attention from researchers. Developed 3D reconstruction methods should be adapted to satellite context. 3D-building automatic reconstruction approaches can

* Corresponding author at: Institut EGID, Université of Bordeaux, GHYMAC Lab: 1 allée F.Daguin, 33607 Pessac, France. Tel.: +33 662007836.
E-mail address: nesrine.chehata@ign.fr (N. Chehata).

0924-2716/$ – see front matter © 2008 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2008.10.003

be classified as data-based (bottom-up) approach and model-based (top-down) approach. In the former one, authors do not restrict the set of available roof structures. 3D-feature extraction leads to a 3D model. In general, only one kind of features is used; 3D-segments in Baillard and Zisserman (1999), corners in Heuel et al. (2000) and planar patches in Ameri and Fritsch (1999). To yield a 3D building model, automatic methods were proposed such as perceptual organization (Price and Nevatia, 2002) and polyhedral approaches (Baillard and Zisserman, 2000). These approaches cannot handle properly under-detection and over-detection problems of feature extractors. Conversely, the model-based approaches (Fischer et al., 1998; Fuchs and Le-Men, 2000; Lafarge et al., 2006) are more robust to over- and under-detection but they restrict roof structures to some prefixed building models.

In this study, all the image features (points, segments, facets) can not be extracted due to a relatively low resolution of satellite images, so a mixed strategy is adopted. First, 3D reliable and complementary features are extracted from images. The use of different features (segments, planar facets) is a key step to generating building hypothesis. The extraction of 3D-features is detailed in previous works (Chehata et al., 2002, 2003). A top-down approach, based on a global optimization process yields a 3D dense surface. Surface modeling is formulated as a matching problem that can be solved by graph cut minimization. The novelty consists in the use and the construction of the graph in order to combine input 2D data and 3D feature constraints. The final surface can be either represented in 2.5-D as a DEM or as a hybrid 3D-surface that provides different description levels (raster/vector).

In the first section, the state of art of stereo matching is presented and graph cut minimization is detailed. The global optimization process with graph construction is detailed in Section 3. Section 4 presents the 3D-feature hard constraints. Finally, results and discussions are presented in Section 5.

2. State of art

2.1. Stereo matching

Stereo matching algorithms that produce dense disparity maps can be classified as local or global optimization methods. In local methods, the disparity value of each pixel is chosen independently of other pixels. Square windows of fixed size Okutomi and Kanade (1991) or adaptative windows Kanade and Okutomi (1994) can be used. Geiger et al. (1995) use a multiple window method to choose the best correlation score. Since these local aggregation methods assume a constant disparity in these windows, their behavior is not optimal in regions of depth discontinuities. Global optimization methods attempt to overcome this problem by minimizing an energy function that is a combination of a "data" term and a "smoothness" term. The goal is to find the disparity/depth function that minimizes this energy:

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{(p,q) \in \mathcal{N}} V_{pq}(f_p, f_q) \quad (1)$$

where $\mathcal{N} \subset P \times P$ is a neighboring system on pixels, $D_p(f_p)$ is the data term and consists in a penalty for the pixel p to have the value f_p. $V_{pq}(f_p, f_q)$ measures the cost of assigning the labels f_p, f_q to adjacent pixels p, q and is used to impose smoothness. However, the smoothness term has also to deal with large discontinuities, that are present in urban scenes.

Several methods for energy minimization were used including simulated annealing (Geman and Geman, 1984), Bayesian networks, belief propagation (Tappen and Freeman, 2003), relaxation labeling (Szeliski, 1990) and non-linear diffusion of support (Scharstein and Szeliski, 1996). Some methods tend to find a minimum separately for each scan line (via dynamic programming) and minimization is processed in one dimension, along epipolar lines (Belhumeur, 1996; Geiger et al., 1995; Bobick and Intille, 1999). These approaches are not globally optimal. Other methods, based on graph formulation, minimize directly 2D-functions representing the disparity surface (Roy and Cox, 1998; Ishikawa and Geiger, 1998; Boykov et al., 1999; Veksler, 1999; Kolmogorov and Zabih, 2002; Paris et al., 2006). They are based on graph cut formulation. The matching problem is considered as a graph labeling problem where labels correspond to disparities. The objective is to find the consistent labeling of the given graph under scene model assumptions. This global approach to stereo analysis provides a more accurate and coherent depth map than the traditional line-by-line stereo. In this study, computing a 3D surface from a stereopair images is formulated as a graph cut minimization. A detailed state of art is presented in the following section.

2.2. Graph cut minimization

Graph cut minimization proved to be a useful multidimensional optimization tool that can enforce smoothness and deal with discontinuities (Scharstein et al., 2002). It appeared with Greig et al. (1989) in the context of Binary Markov Random Fields where each image pixel is given a binary label. In Roy and Cox (1998), authors introduced the first approach using graph cuts for more than two labels and convex continuities. A local coherence constraint ensures that neighboring pixels have similar disparities.

This method has been extended with ordering constraint (Ishikawa and Geiger, 1998) in the context of MRFs, but this solution suffered both from linear discontinuity penalty and from streaking lines due to their use of ordering constraint. Furthermore, the energy used in both approaches is not discontinuity preserving. Most improvements mentioned in the literature concern discontinuity preserving and occlusion explicit modeling. In Boykov et al. (1999), authors present an approximate algorithm applied to binary sub-problems that guarantee discontinuity preserving. They give rise to multi-way cut problems. This has been generalized by Kolmogorov and Zabih (2001) to enforce uniqueness and to deal explicitly with occlusions. In Scharstein et al. (2002); Boykov et al. (1999), smoothness energy depends on intensity differences to favor discontinuities. In Zitnick and Kanade (2000), occlusions are labeled by thresholding correlation scores. In addition, newer cooperative algorithms based on graph cuts have emerged. In Agrawal and Davis (2004), fixed local windows and global energy minimization are confronted by allowing two disparities to be active within the same window. Another algorithm (Lin and Tomasi, 2004) alternates iteratively between segmenting the image in planar regions and fitting a smooth disparity surface to each region.

While, previous algorithms propose optimization on pixellar information, the novelty consists in combining image information and 3D-features as constraints to preserve discontinuities and ensure 3D surface smoothness. The algorithm is processed to recover height maps on urban scenes.

2.2.1. Graph cut formulation

Let us consider a weighted graph $\mathcal{G} = \langle V, E \rangle$, $E \subset V \times V$ oriented edge set. $V = \{s, t\} \cup P$ contains two terminal nodes, the source s and the sink t and a set of non-terminal nodes P. Each graph edge is assigned a non-negative weight or cost $w(p, q) \in \mathbb{N}$, $(p, q) \in E$.

A cut $C_{\mathcal{G}}$ is a set of edges such that terminals are separated into two disjoint sets $s \in S$ and $t \in T$ (cf. Fig. 1).

The cost of a $C_{\mathcal{G}}$ cut, denoted $|C_{\mathcal{G}}|$, is equal to the sum of its edge weights (Eq. (2)). An edge $(p, q) \in C_{\mathcal{G}}$ if $(p, q) \in S \times T$ or $(p, q) \in T \times S$.

$$|C_{\mathcal{G}}| = \sum_{\substack{(p,q) \in S \times T \\ (p,q) \in T \times S}} w(p, q). \quad (2)$$

Download English Version:

https://daneshyari.com/en/article/557415

Download Persian Version:

https://daneshyari.com/article/557415

Daneshyari.com