ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 490–500

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions

Uwe Soergel [a], Eckart Michaelsen [b,*], Antje Thiele [b], Erich Cadario [b], Ulrich Thoennessen [b]

[a] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, 30167 Hannover, Germany

[b] FGAN-FOM Research Institute for Optronics and Pattern Recognition, 76275 Ettlingen, Germany

ARTICLE INFO

Article history:
Received 31 October 2007
Received in revised form
15 October 2008
Accepted 16 October 2008
Available online 20 February 2009

Keywords:
SAR images
Stereo
Building extraction
Production system
Perceptual grouping

ABSTRACT

SAR stereo image analysis for 3D information extraction is mostly carried out based on imagery taken under same-side or opposite-side viewing conditions. For urban scenes in practice stereo is up to now usually restricted to the first configuration, because increasing image dissimilarity connected with rising illumination direction differences leads to a lack of suitable features for matching, especially in the case of low or medium resolution data. However, due to two developments SAR stereo from arbitrary viewing conditions becomes an interesting option for urban information extraction. The first one is the availability of airborne sensor systems, which are capable of more flexible data acquisition in comparison to satellite sensors. This flexibility enables multi-aspect analysis of objects in built-up areas for various kinds of purpose, such as building recognition, road network extraction, or traffic monitoring. The second development is the significant improvement of the geometric resolution providing a high level of detail especially of roof features, which can be observed from a wide span of viewpoints. In this paper, high-resolution SAR images of an urban scene are analyzed in order to infer buildings and their height from the different layover effects in views taken from orthogonal aspect angles. High level object matching is proposed that relies on symbolic data, representing suitable features of urban objects. Here, a knowledge-based approach is applied, which is realized by a production system that codes a set of suitable principles of perceptual grouping in its production rules. The images are analyzed separately for the presence of certain object groups and their characteristics frequently appearing on buildings, such as salient rows of point targets, rectangular structures or symmetries. The stereo analysis is then accomplished by means of productions that combine and match these 2D image objects and infer their height by 3D clustering. The approach is tested using real SAR data of an urban scene.

1. Introduction

Automatic building recognition from remote sensing data is an important field of research today (Baltsavias et al., 2001). The required 3D information is usually derived either indirectly by stereo analysis of optical image pairs, which are collected passively by aerial and satellite frame cameras and scanners, or directly by runtime measurement of active airborne laser-scanning devices. Limited by diffraction the aperture angle α of any image forming sensor depends on the ratio of its wavelength λ and aperture D. Both mentioned sensor types operate in the visible or near-infrared frequency domains and a high spatial resolution ∂_α in the decimeter scale can be achieved for typical parameters settings of λ, D, and distance to the scene r:

$$\partial_\alpha \propto \alpha \cdot r = \frac{\lambda}{D} \cdot r. \tag{1}$$

A crucial restriction of optical and laser sensors is their dependency on clear sky conditions, because of atmospheric signal loss due to scattering by aerosols or raindrops. In order to ensure all-weather mapping capability the microwave frequency domain is an alternative of still growing importance, since the power of a typical radar wavelength signal in the order of centimeters is hardly attenuated by propagation through haze or clouds. On the other hand, according to Eq. (1) for given distance r and aperture D the angular resolution ∂_α linearly worsens with rising wavelength. As a consequence, imaging in nadir view using radar sensors is in practice restricted to low altitude platforms (Klare et al., 2006).

Imaging radar from airborne sensor carriers at higher altitudes or from satellites requires oblique illumination in order to separate the backscatter from objects of different distances to the sensor

* Corresponding author. Tel.: +49 7243 992 335; fax: +49 7243 992 299.
E-mail address: michaelsen@fom.fgan.de (E. Michaelsen).

0924-2716/$ – see front matter © 2009 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2008.10.007

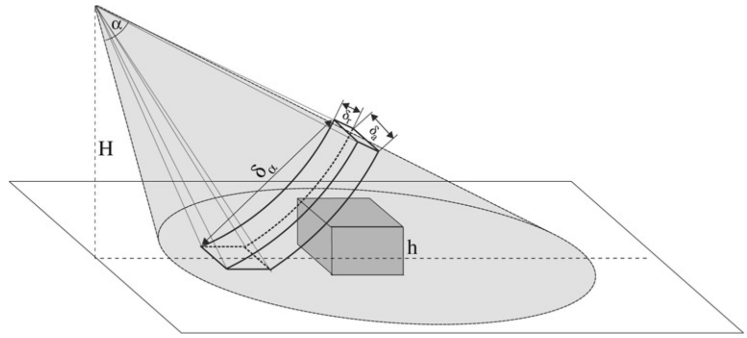

Fig. 1. SAR principle of scene mapping giving rise to layover at building location.

located in the antenna footprint exploiting the runtime of the incoming signal. The resolution ∂_r in range direction (coinciding with the beam antenna's 3 dB main lobe axis of angle α) is not a function of the distance and depends only on the pulse length τ, which is inverse proportional to the signal bandwidth B. In the case of a single pulse measurement only the resolution in the direction perpendicular to range axis would still be ∂_α. The synthetic aperture radar (SAR) principle overcomes this limitation (Schreier, 1993). The scene is illuminated obliquely and side-looking repeatedly with high spatial overlap of subsequent coherent pulses, usually orthogonal to the carrier path. High azimuth resolution ∂_a is achieved by signal processing of the entire set of pulses along the flight path which cover a scene point. The resolutions of a SAR image in range and azimuth (stripmap mode) directions are given as:

$$\partial_r = \frac{c \cdot \tau}{2}, \qquad \partial_a \to \frac{D_a}{2}, \tag{2}$$

with velocity of light c and antenna size in azimuth direction D_a.

Consider the representation of the 3D world in a cylindrical coordinate system. If the coordinates are chosen such that the z-axis coincides with the sensor path, each pulse emitted by the beam antenna in the range direction would intersect a cone of this cylinder and by subsequent measurements the set union of these pulses represents a wedge-shaped subset of the world. A SAR image can be thought of as a projection of this 3D representation (azimuth $=$ z, range, and elevation angle coordinates) on a 2D plane (range, azimuth axes) in elevation angle direction. In other words, the poor angular resolution ∂_α of a real aperture radar system is still valid for the elevation coordinate. This is the reason for the so-called layover phenomenon: all signal contributions of objects inside the antenna beam which share the same range and azimuth coordinates are integrated into the same 2D resolution cell of the SAR image even though if they differ in elevation angle. Owing to vertical facades layover is ubiquitous in urban scenes (Dong et al., 1997); the sketch in Fig. 1 visualizes the described mapping process of a 3D volume into a SAR image resolution cell using the example of signal mixture of a building and the ground in front of it. One possibility to overcome the layover problem is multi-baseline processing of sets of SAR images of suitable tracks; the key idea is to establish a second synthetic aperture orthogonal to flight path and to achieve in this manner a real 3D imaging of the scene. This technique, which is referred to as SAR tomography as well, was already demonstrated for airborne (Reigber and Moreira, 2000) and space borne scenarios (Fornaro et al., 2005). Besides layover, the side-looking illumination leads to significant amount of occlusion in form of radar shadow behind buildings in built-up areas (Soergel et al., 2006). In addition, due to the preferred rectangular alignment of objects mostly consisting of piecewise planar surface facets, multi-bounce signal propagation is frequently observed.

There exist several methods for Digital Elevation Model (DEM) extraction from SAR data. A comprehensive overview of extraction methods from satellite SAR data is given in Toutin and Gray (2000). The most important and common is SAR Interferometry (InSAR), where the phase difference between two SAR images of the same scene is exploited to derive height information. A prerequisite of InSAR is significant overlap of the object spectra in the SAR images. As a consequence, InSAR is in general restricted to similar illumination directions. The illumination direction is described in terms of viewing and aspect angles; the former is defined here by the off-nadir angle to the antenna beam axis, the latter by the horizontal angle between a given reference direction and the projection of the antenna beam axis on the ground connecting a sensor track nadir point and the object of interest. For example, the critical baseline of ERS satellites operating in orbits of about 780 km altitude is about 1.1 km. InSAR data were for instance used in Gamba et al. (2000) for detection and extraction of large buildings applying methods developed originally for machine vision applications. In another recently published work (Tison et al., 2007) urban height map and classification were jointly derived from airborne high-resolution InSAR data by merging several input information sources into a Markovian framework.

A different method for 3D information extraction from images is stereo analysis (in case of Radar images known as radargrammetric stereo), which relies on the amplitude data only and can therefore by applied for larger viewing or aspect angles differences. The standard process for optical images is based on correlation of overlapping images according to epipolar geometry. This principle is transferable to SAR data, this will be discussed in the next section in more detail. In general, the larger the change in viewing geometry is, the more dissimilar the SAR images appear; in particular in urban areas this effect is observed (Leberl, 1990). Therefore, usually image pairs recorded with parallel trajectory and same aspect but different viewing angle are used for radargrammetric stereo analysis (Simonetto et al., 2003). On the other hand, the severe layover and radar shadow problems, which may cover large parts of urban scenes, call for multi-aspect SAR data acquisition and analysis schemes. Soergel et al. (2005) determined sets of optimal illumination directions for a typical central European city by simulations based on a given high-resolution DEM and systematic variation of aspect and viewing angles. Two applications were considered: maximization of the roof area and respectively road area from up to four SAR images such that at least in one image no layover or occlusion caused by adjacent elevated objects would occur. It turned out that in both cases the optimal complementing image for the best single image would be taken from an orthogonal aspect.

Pixel based techniques are not suitable for stereo analysis in case of significant aspect angle difference, because of strong radiometric differences. Instead, object oriented approaches should be applied to extract corresponding features in the images.