Building Extraction Based on Stereo Analysis of Orthogonal View High-resolution SAR Data

E. Michaelsen[1], U. Soergel[2], A. Thiele[1]

[1] FGAN-FOM, Gutleuthausstrasse 1, 76275 Ettlingen, Germany, mich@fom.fgan.de
[2] Institute of Photogrammetry and GeoInformation, University of Hanover, Nienburger Strasse 1, 30167 Hanover, Germany

In this paper high-resolution SAR images of urban sceneries are analyzed in order to infer buildings and their height from the different lay-over effects in perpendicular taken views. Due to the strong dependence of the appearance of objects on the lighting and viewing direction it is unlikely that a simple image-matching method would succeed. Instead, higher level object matching is proposed. Here a knowledge-based approach is applied, a production system. The images are analyzed separately for the presence of image objects frequently appearing on buildings, such as salient rows, rectangular structures or symmetries. The stereo step is than coded by means of productions that combine and match these image objects and infer the height. Experiments with real data are reported.

Introduction

Automatic building recognition from high-resolution multi-view SAR data is an important challenge today. Two significant advantages of SAR are the independence of the daytime due to an active scene illumination and the insensitivity to weather conditions resulting from the large signal wavelength. The SAR principle requires an oblique and side-looking viewing direction [9]. This leads to some disadvantages like occlusions, layover and multi-bounce signal propagation, occurring frequently in urban areas [2]. These effects make it difficult to use SAR stereoscopy for urban scene analysis [5]. Small changes in the illumination aspect can lead to strong image dissimilarity which hinders the search of suitable tie points. Therefore, image pairs recorded with parallel trajectory and viewing direction but different depression angle were recently used for radargrammetric stereo approach [8]. Modern experimental airborne SAR systems are able to provide data with a spatial resolution in the decimeter scale [1]. Many features of urban objects can be identified in such data, which were beyond the scope of radar remote sensing before. This suggests that using radargrammetric stereo approach matching these objects also for data recorded with non parallel trajectory may be feasible [7].

In this contribution an image pair with almost orthogonal viewing directions is investigated and we assume the image data to be given in ground range and geo-referenced. Such image pair is shown in Figures 1 and 2 respectively.

Fig. 1. Campus of the University of Karlsruhe – viewed by high-resolution SAR, lighted from North.

Geo-reference means here that a scatterer on ground would appear at the same position in both images. A scatterer located above ground must appear higher in the image lighted from North and further left in the image lighted from West. One may render the images on a display in a alternating way and would perceive elevated objects flipping in a diagonal way up-right to down-left and back.

Fig. 2. Campus of the University of Karlsruhe – viewed by high-resolution SAR, lighted from West.

This resembles the situation in stereo vision from ordinary pictures. However the "epipolar lines" are oblique and the "disparities" have to be interpreted a little differently. The following section will discuss these issues in more detail, before we treat our experiments and conclude.

Problems in SAR Stereo

In [5,8] the term radargrammetry is used for elevation-from-displacement approaches such as the one outlined above. There is general agreement that in SAR-data – and in high resolution SAR-data in particular – the appearance of objects very much depends on the lighting direction. Most buildings in our images appear in fact very differently when lighted from North as compared to lighted from West. Therefore in [8] data are used where the lighting direction is the same and only the depression angle varies – giving layover-displacement in the same direction but of different magnitude.

The classical occlusion problem of stereo-vision also occurs here: Some objects appearing in one view may not be visible in the other. Additionally there is a mixture problem: In layover areas different objects may appear in the same location. The signals are just summed. All these effects make it rather unlikely that the same grey-levels appear at corresponding positions along the epipolar lines in our image pair. However, human observers can perceive corresponding objects. It results that – while things like correlation based pixel-wise correspondence is useless – image-object correspondence may be feasible. The image-objects just have to be abstract enough.

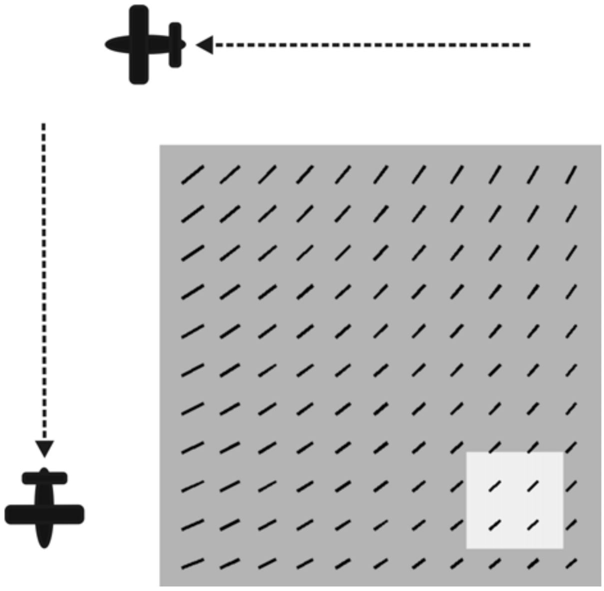

Fig. 3. SAR-geometry: Flight paths and epipolar lines depending on the position; the white square indicates the image region

Desirable is a level of abstraction that describes the object and locates at its reference position independently of the lighting direction. The position displacement d is given in Equation (1)

$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = h_{object} \begin{pmatrix} \tan(\arctan((near_x + x)/h_{west})) \\ \tan(\arctan((near_y + y)/h_{north})) \end{pmatrix} \qquad (1)$$

where h_{object} denotes the height of the object, $near_x$ and $near_y$ are the distances of the left and upper image margin to the flight-paths and h the corresponding flight heights respectively. The epipolar search lines result from rotating d by π/2. For each pixel position $(x,y)^T$ there is a slightly different displacement for the same h_{object}. And also the same disparity means a greater object height in the upper left corner than in the lower right. However, images like the presented examples display a fairly small section. In our case they are located more than 2000m away from the flight paths and show only about 500m by 500m of the scene. So the non-linearity effects from equation (1) are not very strong – as can be seen in Fig. 3, where for a hundred systematically chosen positions and a maximal height of 24m the epipolar lines are shown.

Stereo Productions on Different Abstraction Levels

A suitable computational environment for swift testing declarative knowledge – such as the relations described above (in particular in (1) and Fig. 3) – on given data such as the SAR-image pair shown in Figs. 1 and 2 – are *production systems* [6, 3, 10].

Two classes of *primitive objects* are extracted from the images by suitable filter operations: **Pixel** objects are considerably brighter than their surrounding; they indicate the presence of salient scatteres. **Line** objects are constructed where the image gradient is large and stable; they indicate the presence of an edge or line-object.

According to their format (without considering their specific content) the productions used for this work can be assigned to the following production types:

Clustering productions group a set of objects of the same type which are mutually affirming into a single higher order object: **BrightSpot** objects are a result of grouping adjacent **Pixel** objects; **LongLine** objects are formed of a set of **Line** objects that happen to lie co-linear; **SyCluster** objects are constructed from a set of **Symmetry** objects that vote locally for a similar symmetry axis.

Pair wise construction productions produce a composed object of two parts which fulfill a given geometric constraint: **Angle** objects are constructed from two **LongLine** objects provided they are arranged in an L-shape; **Symmetry** objects are constructed from symmetrically arranged **Angle** objects.

Recursive grouping needs one production for initialization and one for the recursion step: A trivial **Row** object with just one member is formed from a **BrightSpot** object provided there is a contextual **LongLine** object in the vicinity giving a hint for a direction; longer **Row** objects with one more member are formed of **Row** objects by appending another **BrightSpot** object that suits the regularity of the row.

Stereo productions infer the height of an object in the scene by establishing correspondence between objects in the different views and using (1): **3D-Angle** objects are inferred from **Angle** objects; **3D-Row** objects are inferred from **Row** objects; **3D-Cluster** objects are inferred from **SyCluster** objects.

The object class definitions and the corresponding productions are purely declarative knowledge. The control must be seen completely separated from this. For production systems like the one presented above trivial exhaustive search is not a feasible control strategy. Instead, assessment driven control strategies are recommended as they have been proposed e.g. in [6, 3, 10]. Provided such suitable control tool is given it is quite easy to have a system like that adapted to a particular kind of data. In the next section we show an example for the recognition capabilities that result from such knowledge coding.

Results and Conclusion

The pictures presented in Figures 1 and 2 have been processed by our system and we show the resulting numbers of objects in Tab. 1.

Table 1. Growth of Object Numbers

Object class	100kc	200kc	500kc
Line	39617	39617	39617
Pixel	20697	20697	20697
LongLine	7990	10658	16207
BrightSpot	9118	9827	9827
Angle	710	10803	39001

Row	8630	9602	12048
Symmetry	952	23067	91824
SyCluster	152	3869	16868
3D-Angle	5	697	3208
3D-Row	564	629	655
3D-Cluster	42	5165	18394

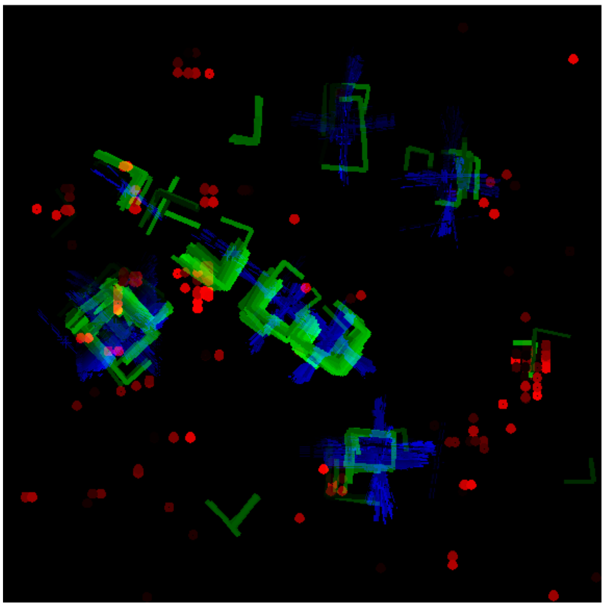

Fig. 4. Objects Angle3D and Row3D obtained after 500.000 interpretation cycles.

The resulting objects **3D-Row** are displayed in Fig. 3 in red color where the brightness indicates the disparity i.e. the height of the objects. All these objects are trivial rows that only contain one salient scatterer. Although there were many non-trivial **Row** objects found in the pictures with many scatterers (up to seven), it has not been possible to achieve correspondence between them after 500.000 interpretation cycles. Presumably, this is due to the wide difference in view directions. **3D-Angle** objects are displayed in the green channel – with again the height coded as brightness. Thus yellow color indicates the presence of row objects and angle objects at that location. The blue channel was used to display the **3D-Cluster** objects correspondingly. These objects are indicated with a small cross at their location and a dotted line giving the symmetry axis. Obviously, some properties of the scene and in particular of the buildings in the scene are recognized. We see that – while the more primitive objects such as the trivial 3D-rows tend to fail and give rather arbitrary response and a lot of presumably erroneous correspondences - the higher level objects in particular the 3D symmetry clusters give a better building detector. Moreover, they exhibit elements of building understanding in terms of inferring features like location, main orientation and symmetry and also height.

Acknowledgment

We want to thank Prof. Dr. Ender and Dr. Brenner (FGAN-FHR Research Institute for High Frequency Physics and Radar Techniques) for providing the SAR data acquired with the PAMIR sensor [1].

References

1. A.R. Brenner, J.H.G. Ender. First experimental results achieved with the new very wideband SAR system PAMIR // Proc. of EUSAR. - 2002 - P. 81-86.
2. Y. Dong, B. Forster, C. Ticehurst. Radar Backscatter Analysis for Urban Environments. // Int. Journal of Remote Sensing. - 1997. - Vol. 18, No. 6, P. 1351-1364.
3. B. Draper, R. Collins, J. Brolio, A. Hanson, E. Riseman. The Schema System // IJCV. – 1989. - Vol. 2. – P. 209-250.
4. J.H.G. Ender, A.R. Brenner. PAMIR - a wideband phased array SAR/MTI system // IEE Proceedings - Radar, Sonar, Navigation. - 2003. -Vol. 150, No.3. - P. 165-172.
5. F. Leberl. Radargrammetric Image Processing. // Artech House, Boston. - 1990.
6. T. Matsuyama, V. S.-S. Hwang. Sigma a Knowledge-based Image Understanding System // Plenum Press, New York. - 1990.
7. E. Michaelsen, U. Soergel, U. Thoennessen. Potential of building extraction from multi-aspect high-resolution amplitude SAR data // In: Stilla U, Rottensteiner F, Hinz S (Eds) CMRT05. - IAPRS 2005. - Vol. XXXVI, Part 3/W24. - P. 149-154.
8. E. Simonetto, H. Oriot, R. Garello. Radargrammetric Processing for 3-D Building Extraction from High-Resolution Airborne SAR Data. // IEEE Int. Geoscience and Remote Sensing Symosium. - IGARSS 2003. - Toulouse, Vol. 3, P. 2002-2004.
9. G. Schreier. Geometrical properties of SAR images. // G. Schreier (ed.), SAR geocoding: Data and Systems: - 1993. - pp. 103-134. Karlsruhe: Wichmann.
10. U. Stilla, E. Michaelsen. Semantic modelling of man-made objects by production nets // In: Gruen A., Baltsavias EP., Henricsson O. (eds). Automatic extraction of man-made objects fromaerial and space images (II) // Birkhäuser, Basel. – 1997. - P. 43-52.