

# *Building extraction in urban scenes from high-resolution InSAR data and optical imagery*

Jan Dirk Wegner, Uwe Soergel

IPI Institute of Photogrammetry and GeoInformation  
Leibniz Universität Hannover,  
30167 Hannover, Germany  
{wegner, soergel}@ipi.uni-hannover.de

Antje Thiele

FGAN-FOM Research Institute for Optronics and Pattern  
Recognition  
76275 Ettlingen, Germany  
thiele@fom.fgan.de

**Abstract**— Modern space borne SAR sensors provide geometric resolution of one meter, airborne systems even higher. In data of this kind many features of urban objects become visible, which were beyond the scope of radar remote sensing only a few years ago. However, layover and occlusion issues inevitably arise in undulated terrain and urban areas because of the side-looking SAR sensor principle. In order to support interpretation, SAR data are often analyzed using additional complementary information provided by maps or other remote sensing imagery. The focus of this paper is on building extraction in urban scenes by means of combined InSAR data and optical aerial imagery.

## I. INTRODUCTION

In the last decades, synthetic aperture radar (SAR) has become a key remote sensing technique due to its all-weather capability and independence of daylight. However, the coarse geometric resolution of tens of meters of satellite data limited its use to large scale applications until recently. In contrast, today's state-of-the-art space borne sensors like TerraSAR-X and Cosmo-SkyMed provide geometric resolution of one meter, while airborne SAR sensors acquire imagery of even finer grid. In such very high resolution imagery the geometric extent of individual ground objects like buildings, bridges and roads becomes visible and can hence be exploited for various applications. One main application arises in crisis situations when the immediate acquisition of an urban scene is required for rapid hazard response. However, the SAR typical effects layover and shadowing complicate the interpretation in urban scenes. Three-dimensional objects are situated closely together and therefore small buildings are occluded by higher ones while facades overlap with trees and cars on the streets. Additionally, the appearance of a particular building in the image depends on the aspect angle of the sensor. Buildings that are not oriented in azimuth direction with respect to the sensor are often hard to detect. This drawback can be partly overcome by using Interferometric SAR (InSAR) acquisitions from two orthogonal flight directions [1]. InSAR has the advantage of providing height information which facilitates detection and three-dimensional modeling of buildings. Multi-aspect InSAR data from four aspects is used for model-based building detection and reconstruction in [2]. Height information of buildings can also be deduced from stereoscopic measurements

[3]. Large rectangular flat-roofed buildings are detected and modeled based on characteristic combinations of extracted line features. Another approach for the detection and three-dimensional reconstruction of buildings with rather simple shapes, based on parallel line features extracted from SAR intensity images of four different aspects, is presented in [4]. In [5] model-based building detection from a single SAR amplitude image is carried out using stochastic geometry.

None-the-less, automatic urban scene analysis based on SAR data alone is hard to accomplish. As a consequence, it is helpful to analyze SAR data in combination with additional information from GIS databases or high-resolution optical imagery. Optical imagery has the advantage of being widely available. In [6] high-resolution InSAR data is combined with an optical aerial image in order to reconstruct bridges over water. Height information is deduced from the InSAR data while the exact position of the bridges and their horizontal extent is determined using the corresponding optical image. Automatic extraction of building footprints based on line features by means of a SAR amplitude image and an optical aerial image is presented in [7]. Furthermore, an aerial photo is used in [8] in order to regularize outcomes from radargrammetric processing of a SAR image pair by means of Markov Random Fields applied to a region adjacency graph.

This paper presents an approach for building detection in dense urban areas combining line features from mono-aspect InSAR data with classification results from an optical aerial image. First concepts are discussed and further ideas for three-dimensional modeling of urban scenes are outlined.

## II. INSAR SIGNATURE OF BUILDINGS

The signature of buildings in InSAR imagery depends on sensor parameters, on properties of the imaged object itself, and on its direct environment. This chapter focuses on the characteristics of different building types in magnitude images and in phase images.

### A. Signature in magnitude data

The building signature in optical and SAR images is characterized by the different mapping geometries, the central projection and the slant range projection, respectively. Examples of flat-roofed and gable-roofed buildings in optical

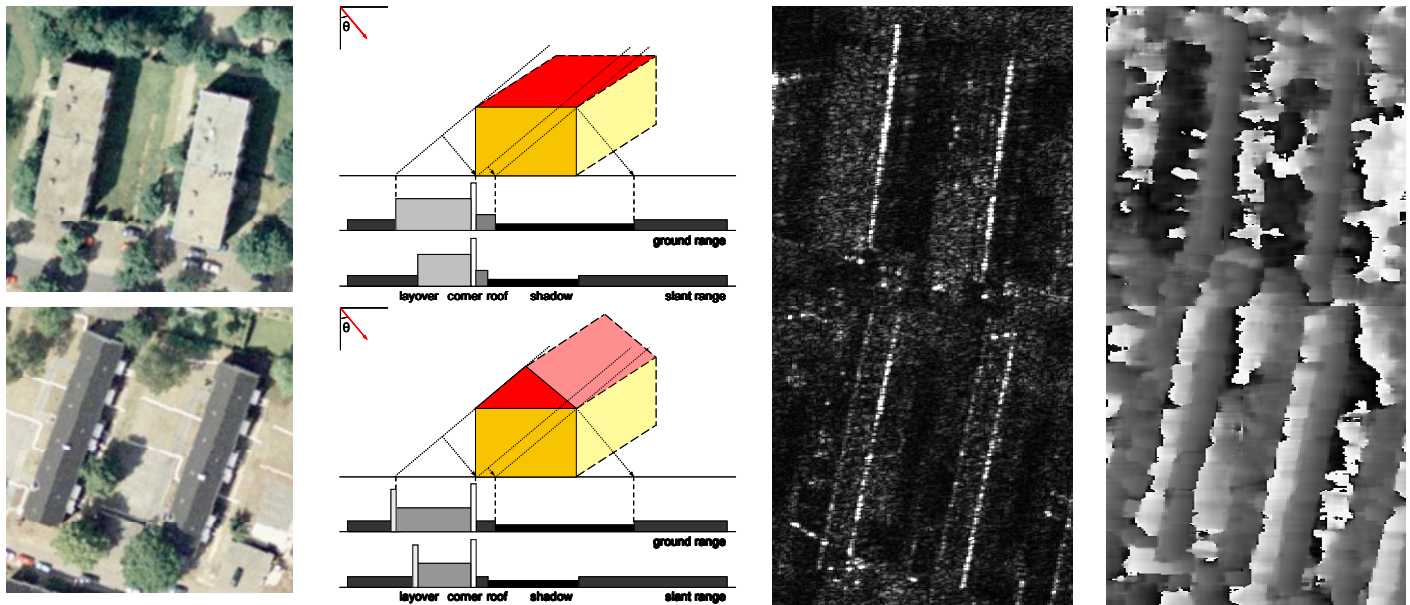


Figure 1 Signature of flat- (first row) and gable-roofed (second row) buildings in optical data (first column), in SAR magnitude data (third column) and InSAR phase data (forth column); schematic view of magnitude signature in SAR data (second column)

and InSAR data are given in Fig. 1. In the second column from the left, the expected appearance of these two building types is displayed schematically. Effects that occur if the scene is illuminated from two orthogonal flight directions are investigated in [9]. The appearance of a building in a SAR image is characterized by the side looking viewing geometry of the sensor, the range measurement, and poor resolution in elevation direction. Its signature consists of layover, corner reflector between ground and building wall, roof signal, and radar shadow.

The layover area is the building signal situated the closest to the sensor because it has the smallest distance to the sensor. It usually appears bright due to superposition of backscatter from ground, façade, and roof. By comparing the layover of the flat-roofed and the gable-roofed example, a subdivision of the layover area is possible according to building dimensions and illumination geometry [1]. This is also observable in the real magnitude data (Fig. 1 third column). The bright line of the flat-roofed signature and the second line (relative to the sensor position) of the gable-roofed signature is the corner reflector line. It is caused by a dihedral corner reflector spanned by ground and building wall. This line is part of the building footprint and can be distinguished from other lines of bright scattering using the InSAR phases (see next paragraph). The subsequent single backscatter signal of the building roof is usually scattered away from the sensor. Thus, it is rarely observable, depending on the roof structure and illumination geometry. Behind the building, the ground is partly occluded by the building shadow, which appears as a dark region. Magnitude changes of the building signature due to illumination direction and building geometry are described in more detail for flat-roofed buildings in [1] and for gable-roofed buildings in [9], respectively.

### B. Signature in phase data

Different building types lead to specific patterns in the interferometric phase data (Fig. 1 fourth column). The phase value of a single range cell results from a mixture of the backscatter of different contributors, such as ground, façade, and roof in the layover area. The phase signature of flat-roofed buildings is characterized by a layover region (in Fig. 1, last column colored in dark gray), and a homogeneous roof region (in Fig. 1 not observable because of the low building width). Gable-roofed buildings also feature a phase signature showing the layover area, which is colored in lighter gray because of the smaller building height. The phase of the terrain enclosing the building signature is displayed slightly darker. A similar phase value is calculated at the building corner location, which is used for the detection of building footprints. Behind the layover area a shadow area, characterized by a noisy phase distribution, is observable. Dependencies of the phase signature on illumination direction and building geometry are described in more detail in [10].

## III. BUILDING DETECTION IN INSAR DATA

The recognition of buildings in InSAR data is based on the detection of parts of the building footprint in this approach. First, bright lines are segmented from the magnitude data. Subsequently, the lines caused by a dihedral corner reflector spanned by ground and building wall are detected based on a local InSAR height analysis. Finally, these building hints are projected into the same ground range geometry as the optical data.

### A. Extraction of building features

The segmentation of primitives exploits the previously discussed bright lines in the magnitude signature of flat-roofed and gable-roofed buildings. The so-called corner line, part of the building footprint, is distinctive for both signatures. The

## 2009 Urban Remote Sensing Joint Event

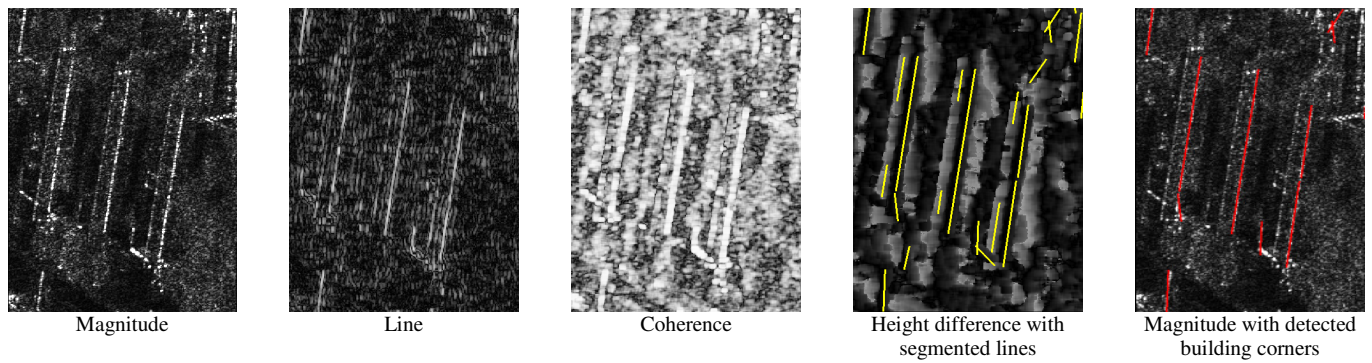


Figure 2 First row shows step by step process of building corner segmentation in slant range geometry; second row shows steps of filtering the InSAR heights to filter and project building corners from slant into ground range geometry

full process of corner line detection is shown in Fig. 2, first row.

The line detection is carried out in slant range geometry. An adapted ratio line detector according to [11] is applied to the original magnitude image (Fig. 2 “Magnitude”). This template detector determines the probability of a pixel of belonging to a line. In this case, eight different template orientations are considered. The probability image for a vertical template orientation is shown in Fig. 2 “Line”. Thereafter, the line segments are filtered based on the eight probability images and their respective window orientation. The resulting segments are fitted to straight lines and edges, respectively, by linear approximation and subsequent prolongation (yellow lines in Fig. 2).

### B. Geocoding of building features

After line extraction, the interferometric heights are calculated [1]. In order to discriminate between lines caused by direct reflection, lines due to double-reflection between ground and wall or lines as a result of double-reflection between roof and substructures, we investigate the local InSAR heights. Only areas of high coherence value and an InSAR height value close to the global mean terrain height are considered for further processing. Based on this filter, a local InSAR height value is calculated over an area of 50 m x 50 m in ground range geometry. This step results in a DTM (Digital Terrain Model). Thereafter, the height differences between InSAR DSM (Digital Surface Model) and DTM are calculated. A line is considered a building corner line if there is a region with significant height difference located next to it. The outcome are real corner lines (Fig. 2 red lines), caused by double-bounce reflection between ground and building wall. The projection of the corner lines from slant range geometry to ground geometry is carried out using the InSAR heights. The resulting geographic position of the corner lines superimposed onto the optical image is displayed in Fig. 3e and Fig. 4c.

## IV. BUILDING DETECTION IN OPTICAL IMAGERY

Building roofs in the optical image are extracted using a model-based approach introduced in [12] which will be outlined in this chapter. It consists of a low-level and a high-level image processing step. The low-level step includes transformation of the image to HSI (Hue Saturation Intensity)

representation, a segmentation of the intensity image and the application of morphological operators in order to close holes. Features are extracted for all regions and the final classification takes place at the end of the following high-level step.

First of all, the input image in RGB (Red Green Blue) color space is transformed to HSI space. The intensity of the image and its hue are of interest for further processing while saturation is not used in this case. A region growing algorithm is applied to the intensity channel in order to find homogeneous roof regions. It is initialized with seed points that are distributed over the image with a raster size adapted to the expected roof size. The assumption is made that building roofs usually appear red, brownish, or grey in optical imagery from airborne sensors. Hence, roofs are expected to show high values in the red channel of the RGB image. Consequentially, a histogram of the red channel of each seed point region is calculated and a pixel with the maximum reddishness is chosen as the seed point. In order to prevent seed points from falling into shadow regions, a threshold is applied to the histogram of the red channel. In case no pixel with a grey value greater than the threshold is present in the region, the region is not considered as a potential roof and no seed point is set.

Secondly, opening and closing is carried out to prepare the regions for feature extraction and classification. Thereafter, regions belonging to the same roof are merged. This step is particularly important for gable-roofed buildings because very often the roof is split at the roof ridge due to different illumination of the two roof sides. The outcomes of this step are roof hypothesis, regions that potentially are building roofs.

Third, features are extracted for each roof hypothesis in order to prepare for classification. Four different feature types are used, based on geometry, shape, radiometry, and structure. Geometric features are the region size and its perimeter. The shape of a building region is described by its compactness, length, and characteristic angles which are of interest because they often distinguish rectangular buildings from natural objects like trees. Since most building roofs are considered to be rather grey or red, radiometry is useful to reject regions with different colors. Hence, the hue angle is determined in order to cancel out regions with a high percentage of green pixels. Structural features that are useful to evaluate a building roof hypothesis are neighboring building regions and shadows cast by the potential building. In urban areas, building rows are

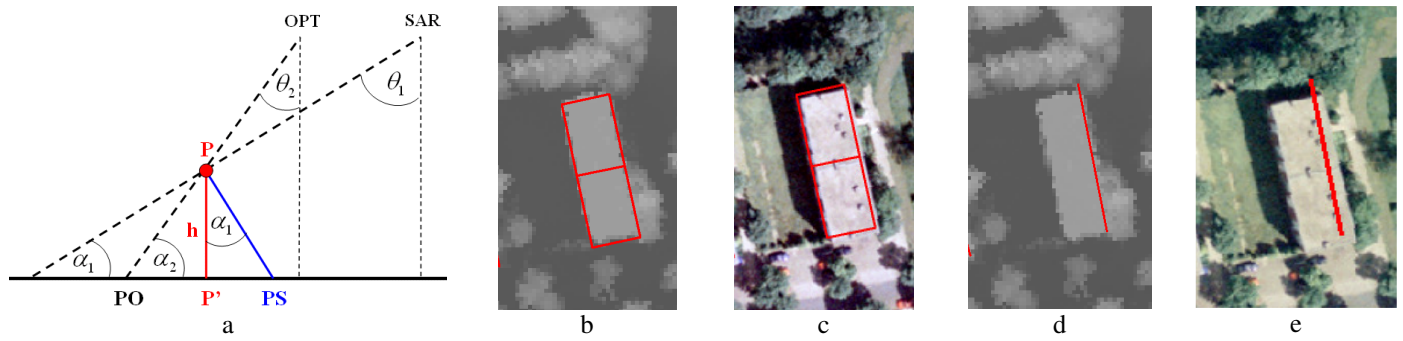


Figure 3 Comparison of SAR and optical viewing geometry (a); LIDAR data (b) and optical data (c) overlaid with cadastral building footprint; LIDAR data (d) and optical data (e) overlaid with detected building corner

often aligned along one side of a street. Rarely, buildings stand alone without any close neighbors. Shadows are good hints for elevated objects. In order to not take into account shadows cast by trees, only shadows with relatively straight borders are considered as belonging to buildings.

Finally, a classification based on the previously determined feature vector takes place (see chapter V.B. for details). All necessary evaluation intervals and thresholds were learned from manually classified training regions.

### V. FUSION OF DETECTION OUTCOMES

In order to accurately combine features from the InSAR data and the optical image, different sensor geometries and projections have to be considered carefully. It is required that both feature sets are projected to the same geometry, i.e., all data have to be transformed to a common coordinate system [13]. In addition, a fusion and classification framework for combining the detection outcomes from the optical image and from the InSAR data has to be set up.

#### A. Sensor geometries

Optical sensors are passive sensors that may acquire images with small off-nadir angles. The imaging principle consists of measuring angles while no direct distances to the objects on the ground are measured. Building roofs that are not located directly in nadir view are displaced away from the sensor. The higher the building is and the further away it is located from the nadir axis of the optical sensor, the more the building roof is displaced.

By contrast, SAR sensors are active sensors acquiring images with high off-nadir angles. Their imaging technique relies on measuring slant ranges to ground objects while their angular resolution in elevation direction is very poor. This leads to foreshortening, layover and shadowing effects (see chapter II). Hence, elevated objects are displaced towards the sensor. The degree of displacement depends on the object height and the off-nadir angle.

Fig. 3a shows the basic difference between SAR and optical viewing geometry under the assumption of locally flat terrain. In the optical image, the elevated object P gets mapped to point PO while it is mapped to PS in the SAR image. The higher P, the further away P from the optical nadir axis and the greater the off-nadir angle  $\theta_1$  becomes, the longer the distance between PO and PS will get.

The optical image was ortho-rectified by means of a DTM before the building detection algorithm explained in chapter IV was applied. Hence, distortions introduced by undulated terrain are decreased. However, distortions due to elevated objects that are not present in the terrain model, as the buildings themselves, remain in the image and building facades are visible. As a consequence, extracted roof regions tend to be slightly shifted away from the sensor if only the building roof has been detected. The displacement effect can be seen in Fig. 3c to 3e. In Fig. 3b the cadastral boundaries of a building are superimposed onto the corresponding LIDAR heights. Such boundaries fit very well to the LIDAR heights but not to the building roof in the optical image (Fig. 3c). In Fig. 3d the LIDAR heights are overlaid with the corner line extracted from InSAR data. The corner line is situated on the edge of the building footprint as expected. In Fig. 3e, the same corner line is superimposed onto the optical image. It falls onto the building roof due to the building's displacement away from the sensor. Such effect is of high interest and can be exploited for three-dimensional modeling of the scene [14, 15] since the distance between the corner line and the building edge comprises height information.

#### B. Joint classification framework

After having projected both the optical features and the InSAR features to the same geometry, a joint classification of both feature sets is conducted. A variety of data fusion approaches exist that are usually applied on pixel level. In [16] and [17] data fusion based on Bayesian inference is proposed. Another possibility is to use the Dempster-Shafer evidential theory [17]. Both approaches are usually based on the assumption that an object is represented identically in the different sensor outputs, i.e., exactly the same region is found in both datasets but with slightly different classification results. This assumption is not readily valid in our case because a combination of regions and lines is carried out. Therefore, a very simple method is used for feature fusion and joint classification.

The extracted lines from the InSAR data are integrated into the linear regression classifier that was used for detection of roof regions in the optical image. Each region detected by the algorithm outlined in chapter IV is evaluated based on several features from the optical image and the InSAR double bounce lines. The optical features have already been mentioned in the previous chapter. For integration of the InSAR double bounce



Figure 4 Results of building corner line segmentation based on InSAR data (a, c), of building detection based on optical data (b), and of building detection based on InSAR and optical data (d)

lines, all building regions are first enlarged by two dilations, i.e., the entire region is enlarged by a two pixel wide buffer enclosing the original region. This corresponds to approximately 0.6 meters in object space. A region is considered to be a roof region if a certain length of a double bounce line lies inside the region.

A quality measure is assigned to each region. It has to be highlighted that this quality measure cannot be interpreted as a probability. Initially, this value is set to one. Subsequently, an evaluation of each region based on its feature vector is done. In case a certain feature does not completely support the building hypothesis, the quality measure is multiplied by a reduction value between one and zero. The new value of the quality measure is again multiplied by another reduction value if the next feature does not completely support the building hypothesis and so on. The exact reduction value for each feature is learned from a training data set of manually classified buildings. The evaluation procedure is done separately for optical features and InSAR features. The overall quality measure is obtained by summing up the optical quality measure and the InSAR quality measure after both measures have been multiplied separately by weighting factors. Since much more

information about the building regions is derived from the optical image, the optical data is believed to contribute two thirds to the final result while the InSAR data contributes one third. Hence, the weight for the optical quality measure is set to 0.666 and the InSAR quality measure weight is set to 0.333. In case neither an optical feature nor an InSAR feature decreases the quality measure, both quality measures sum up to one. Finally, all regions that have an evaluation value greater than a threshold are classified as buildings. Such threshold is determined empirically. It was set to 0.6 in this case. Hence, a region may be classified as a building even if there is no hint from the InSAR data, but strong evidence from the photo.

## VI. RESULTS

The investigated single look complex InSAR data set was recorded by the AeS-1 sensor system of Intermap Technologies with a spatial resolution of about 38 cm in range and 16 cm in azimuth direction. The two X-Band sensors operated with effective baseline  $B \approx 2.4$  m. The mapped residential area in the city of Dorsten in Germany is characterized by a mixture of flat-roofed and gable-roofed buildings and low terrain undulation.

Results of the presented approach are displayed in Fig. 4. In Fig. 4a a LIDAR DSM of the investigated area is overlaid with the corner lines from the InSAR data. It can be seen that some of the lines are not situated on building edges. This effect is due to erroneous InSAR heights used for the projection from slant range geometry to ground geometry. Fig. 4c shows the same corner lines superimposed onto the optical image. Some of the buildings do not show any corner line close to them. This is the case for the row of gable-roofed buildings with reddish roofs at the bottom of the image. Missing corner lines are first and foremost due to the aspect the InSAR data was acquired with. A second reason is that a building may be occluded by other elevated objects like trees or another building.

Fig. 4b shows the detected buildings using the classifier from [12]. Results are rather poor which is, above all, due to the assumption that roofs do not split up into more than two regions during the region growing step. As a consequence, several gable-roofed buildings with reddish roofs in lower right of the image are not detected. Some flat-roofed buildings in the upper part of the image are not detected based on optical data alone since their color and shape are similar to those of street segments. Thus, their evaluation value does not exceed the threshold. Final building detection results are enhanced considerably by the integration of the InSAR corner lines (Fig. 4d). The detection rate of buildings could be increased significantly while the number of false alarms was reduced notably. Some buildings that are hardly recognizable in the optical image are detected due to a corner line, e.g. the garage between two vertically oriented gable-roofed buildings in the lower left of Fig. 4d. Again, the vertically oriented gable-roofed buildings with reddish roofs in the lower right of Fig. 4d are not detected although strong hints from corner lines are present. This is due to the previously outlined over-segmentation during the region growing step.

## VII. CONCLUSION AND OUTLOOK

In this paper, an approach for the combination of features from optical imagery and InSAR data with the goal of building detection in urban areas was presented. It has been shown that features from InSAR data contribute significant information to building detection. The rather poor results from the optical classification could very much be improved using InSAR corner lines.

One main disadvantage of the presented classification approach is that its quality measures are not interpretable as probabilities in a Bayesian sense. Although many parameters have been learned from training data, parts of the approach appear rather ad-hoc. A next step will thus be the integration of the presented approach into a Bayesian framework.

Another issue to deal with is that the final building detection is completely dependent on the very first segmentation of the optical image. More advanced processing steps have to be developed that correct for ill-shaped regions. One idea would be to use the corner lines for shape improvements.

## REFERENCES

- [1] A. Thiele, E. Cadario, K. Schulz, U. Thoenessen and U. Soergel, "Building Recognition From Multi-Aspect High\_resolution InSAR Data in Urban Areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45(11), pp. 3583-3593, 2007
- [2] R. Bolter and F. Leberl, "Detection and Reconstruction of Human Scale Features from High Resolution Interferometric SAR Data," *Proceedings of the International Conference on Pattern Recognition 2000 (ICPR2000)*, pp. 291-294, 2000
- [3] E. Simonetto, H. Oriot and R. Garello, "Rectangular Building Extraction From Stereoscopic Airborne Radar Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43(10), pp. 2386-2395, 2005
- [4] F. Xu and Y.-Q. Jin, "Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45(7), pp. 2336-2353, 2007
- [5] M. Quartulli and M. Datcu, "Stochastic Geometrical Modeling for Built-Up Area Understanding From a Single SAR Intensity Image With Meter Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42(9), pp. 1996-2003, 2004
- [6] U. Soergel, A. Thiele, E. Cadario, U. Thoenessen, "Fusion of High-Resolution InSAR Data and optical Imagery in Scenes with Bridges over water for 3D Visualization and Interpretation," *Proceedings of Urban Remote Sensing Joint Event 2007 (URBAN2007)*, 6 pages, 2007
- [7] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, pp. 71 – 82, 2003
- [8] F. Tupin and M. Roux, "Markov Random Field on Region Adjacency Graph for the Fusion of SAR and Optical Data in Radargrammetric Applications", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42(8), pp. 1920-1928, 2005
- [9] A. Thiele, E. Cadario, K. Schulz, U. Thoenessen, and U. Soergel, "Reconstruction of residential buildings from multi-aspect InSAR data", *Proceedings of ESA-EUSC Workshop, March 2008, Frascati, Italy*, available [http://earth.esa.int/rtd/Events/ESA-EUSC\\_2008/](http://earth.esa.int/rtd/Events/ESA-EUSC_2008/), 6p.
- [10] A. Thiele, E. Cadario, K. Schulz, U. Thoenessen, and U. Soergel, "InSAR Phase Profiles at Building Locations," *Proc. of ISPRS Photogrammetric Image Analysis*, vol. XXXVI, part 3/W49A, pp. 203-208, 2007.
- [11] F. Tupin, H. Maitre, J-F. Mangin, J-M. Nicolas, and E. Pechersky, "Detection of Linear Features in SAR Images: Application to Road Network Extraction", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 2, 1998, pp. 434-453.
- [12] S. Mueller and D. W. Zaum, "Robust Building Detection in Aerial Images," *IntArchPhRS*, vol. XXXVI, Part B2/W24, pp. 143 – 148, 2005
- [13] A. Thiele, K. Schulz, U. Thoenessen and E. Cadario, "Orthorectification as Preliminary Step for the Fusion of Data from Active and Passive Sensor Systems," *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems 2006 (MFI2006)*, 6 pages, 2006
- [14] J. Inglada and A. Giros, "On the possibility of Automatic Multisensor Image Registration", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42(10), pp. 2104-2120, 2004
- [15] J.D. Wegner and U. Soergel, "Bridge height estimation from combined high-resolution optical and SAR imagery," *IntArchPhRS*, vol. XXXVII, Part B7-3, pp. 1071-1076, Beijing 2008
- [16] C. Bishop: *Pattern Recognition and Machine Learning*. 1st edition, New York, NY: Springer, 2006, p.46.
- [17] L. A. Klein, *Sensor and Data Fusion-A Tool for Information Assessment and Decision Making*, 3rd ed., Bellingham, WA: SPIE Press, 2004, pp.127-181.