EXTRACTION OF BUILDING HEIGHTS FROM VHR SAR IMAGERY USING AN ITERATIVE SIMULATION AND MATCH PROCEDURE

Dominik Brunner*† and Guido Lemoine*

Department of Information Engineering and Computer Science
University of Trento
38050 Trento, Italy

ABSTRACT

The new spaceborne very high resolution (VHR) SAR sensors onboard the TerraSAR-X and COSMO-SkyMed satellites have a spatial resolution of up to 1 meter. In VHR SAR data, features from individual urban structures (like buildings) can be identified in their characteristic settings in urban settlement patterns. In this paper, we present a novel methodology for the height estimation for generic man made structures from single power SAR data. The proposed approach is based on the definition of a hypothesis on the height of the building and on the simulation of a SAR image for that hypothesis. Then a matching procedure is applied between the estimated and the actual SAR images in order to validate the height assumption. The process is iterated for different initial height assumptions until the matching function is satisfied and thus the building height is estimated. The efficiency and the properties of the proposed method are demonstrated for the height estimation of flat- and gable roof buildings from a VHR airborne SAR scene and for a pyramid in Giza (Egypt) from VHR TerraSAR-X SPOT beam data.

Index Terms— 3D reconstruction, height extraction, building detection, VHR SAR.

1. INTRODUCTION

Until recently spaceborne SAR systems were only capable of imaging the earth surface with a spatial resolution higher than 9 meter. This changed when the new VHR SAR sensors onboard the TerraSAR-X and Cosmo-SkyMed satellites having spatial resolutions up to 1 meter were launched in 2007. In VHR imagery, features from individual urban structures, like buildings, can be identified in their characteristic settings in urban settlement patterns (e.g. residential areas, city centres, industrial parks).

The 3D reconstruction of a building, or in more general terms of a man made structure, is a key issue for information retrieval from VHR SAR in urban areas. Existing methodologies are either based on shadow analysis [1] or on interferometric data [2]. However, the calculation of the interferogram fails if all of the roof backscattering is sensed before the double bounce area and therefore superimposes with the ground scattering in the layover region, which is usually the case for high buildings [2]. To tackle the problem of signal mixture from different altitudes, [3] proposes two models, one for interferometric and one for polarimetric data. Instead of using interferometric data, [4] emphasizes the usage of stereoscopic SAR. Recently, also methods based on multi-aspect data, where the same area is measured from different flight paths, were proposed [5]. [6] demonstrates height extraction by radiometric analysis of the double bounce area of a building using an electromagnetic model based on the Kirchhoff approximations for a simplified rectangular flat roofed building structure.

In this paper we propose a novel 3D reconstruction method capable to extract the height of man made structures from single detected SAR (power) imagery under the assumption that i) a map with the localization of the building is available; and ii) the core structural characteristics of the buildings in the investigated area are known. The method is based on a "hypothesis generation - rendering - matching" procedure. A series of hypotheses are generated and rendered by a SAR imaging simulator taking into account the acquisition parameters of the actual VHR SAR data. The simulations are compared to the actual VHR SAR data; the estimated height corresponds to the hypothesis, whose simulation matches best with the actual scene. The main novelty of the method consists in the use of detected single VHR SAR images instead of multi-dimensional data (e.g. interferometric, polarimetric, multi-aspect). On a scientific level, the method complements existing approaches based on the use of multi-dimensional SAR data. It is worth noting that the use of a detected single VHR SAR image for height estimation can support a wide range of current applications including the use of existing spaceborne SAR sensors.

The proposed method is flexible with respect to the shape of the object for which the height shall be estimated and its aspect angle, i.e. the angle between the front wall of the object and the flight track of the sensor. One of the key characteristics of the proposed procedure is the simultaneous consideration of the major scattering characteristics of the man made structure in SAR (i.e. layover- and shadow areas, multi bounce contributions) for estimating the height. We demonstrate the performance and the properties of our approach for a flat roof and a gable roof building in submeter VHR airborne SAR data from Dorsten, Germany, and for the Menkaure Pyramid in Giza, Egypt, in meter resolution TerraSAR-X SPOT beam data.

The remainder of the paper is structured as follows. In Sec. 2 we describe the methodology and its main components in detail, while we introduce the test data in Sec. 3. Sec. 4 shows the results of the method before we finish with some conclusions and proposals for future work in Sec. 5.

2. METHODOLOGY

Let \( \hat{h} \) and \( h_{\text{true}} \) be an estimation of the height and the true height, respectively, of the analyzed building. \( \textbf{X} \) and \( \hat{\textbf{X}} \) denote the true and a
simulated SAR image, respectively, of the analyzed object. We formulate the problem of height extraction from single detected VHR SAR images as:

\[
\min \left \{ (\hat{h} - h_{\text{true}})^2 \right \},
\]

In order to estimate \( \hat{h} \) according to the mean square error criterion we define a “hypothesis generation - rendering - matching” approach, where a building is simulated at different heights and compared to the actual scene. The final estimated height \( h_{\text{est}} \) of the object corresponds to the hypothesis which matches best with the actual scene and is given by:

\[
h_{\text{est}} = \arg \max_{\hat{h}} \left \{ M[\hat{X}(\hat{h}), X(h_{\text{true}})] \right \},
\]

with \( M \) as matching function. The highest value of \( M \) corresponds to the best match between the hypothesis and the actual scene. To calculate the match between the simulation and the actual scene, both images need to be co-registered. In practice, co-registration and matching are similar tasks which can be executed at the same time. The value of the measure for which the best co-registration between simulation and actual scene is achieved, is also the final match value for this pair. Since the viewing configuration by which the object under investigation was sensed in the actual scene is modeled by the SAR simulator, only translations are considered as transformation. Hence, the one dimensional optimization problem of Eq. 2 becomes a three dimensional problem:

\[
h_{\text{est}} = \arg \max_{\hat{h}, s} \left \{ M[\hat{X}(\hat{h}), X(h_{\text{true}})] \right \},
\]

where \( \hat{X} \) denotes the translation of the image \( X \) by two dimensional vector \( s = (dx, dy)^T \).

To solve the optimization problem of Eq. 3 we use the multi-dimensional Nelder-Mead (or downhill simplex) method. The use of Simulated-Annealing [7] would be a better choice to avoid local maxima, though it would imply an increased computational costs.

In the following subsections we describe each of the three steps (Hypotheses Generation, Rendering and Matching) in detail.

2.1. Hypotheses Generation

A hypothesis consists of a type of object (e.g. rectangular building with gable- or flat roof) with a certain width, length and height. For the height estimation process, only the height is variable, while the type of the object and its width and length are constant throughout the estimation procedure. Planar dimensions (width, length) are derived from a GIS database (e.g. cadastral maps, digitized maps from independent ancillary data). A number of hypotheses are generated for the same building by iterating the height parameter.

2.2. Rendering

For evaluating which hypothesis matches best with the actual scene, a SAR imaging simulator is employed. It renders the hypothesis by taking into the account the viewing configuration (i.e. local incidence angle, aspect angle) by which the object in the actual scene was sensed by the VHR SAR sensor.

Our radar imaging simulator uses ray tracing to determine which surfaces of a generic object are visible. It can handle any complex object which is composed by spheres, planes, and triangles or any arbitrary combination of one or several instances of these objects. An adjustable mixture of Lambertian and specular scattering is used as model to calculate the backscattering from the surface and model. The simulator optionally includes multiple bounce scattering, and can therefore distinguish between single and dual bounce reflections.

2.3. Matching

For image matching, two types of methods exist: Area-based and feature based methods [8]. Area-based methods calculate directly the correlation between all or a subset of samples in the two corresponding images. Feature based methods instead first extract structural information such as lines and edges from the images to be compared, and then in the second step match them in the feature space.

Our matching task is characterized by two challenges: i) we compare the actual SAR data with speckle to synthetic images without speckle, i.e. the local statistics of the images in the comparison are different; ii) the absolute radiometry of the simulated image do not match exactly with those of the actual scene.

To tackle the problem we proposed in [9] a feature based method, which is based on the extraction of shadow areas and edges. As match criterion we used the normalized cross-correlation coefficient. The drawback of feature based methods is the dependence of the effectiveness and stability of the feature extraction procedures on parameter settings, which is especially true for SAR images. Therefore, we propose in this paper an area based method using Mutual Information (MI) for \( M \) in Eq. 3, a measure derived from information theory, due to its suitability for multi-modality image matching/registration tasks [10].

The mutual information \( MI(X, \tilde{X}) \) between \( X \) and \( \tilde{X} \) is given by:

\[
MI(X, \tilde{X}) = H(X) + H(\tilde{X}) - H(X, \tilde{X}),
\]

where \( H(X) \) and \( H(\tilde{X}) \) are the entropies of \( X \) and \( \tilde{X} \) and \( H(X, \tilde{X}) \) is their joint entropy. The entropies can be calculated from the marginal and joint probability mass functions, which can be derived from the joint histogram. The reason for the independence of MI to the absolute intensity values of the two images is the fact that it is solely based on the joint histogram.

3. TEST DATA SET

As test data set we choose an airborne VHR SAR scene from Dorsten, Germany, which was acquired by the AeS-1 sensor from Intermap Technologies. The imagery was acquired in X-band on 13 March 2003 with a 16 cm azimuth and 38 cm slant range resolution in HH polarisation. The incidence angle ranges from 28° to 52° (near range) to 52° (far range). We preprocessed the data by averaging the image by 4 samples in azimuth and 2 samples in range direction, to achieve approximately the same pixel spacing in both directions. The resulting equivalent number of samples is 2.59. We speckle-filtered the image with a Gamma MAP filter. As ground truth we used manually collected data in combination with a Lidar DSM with 10 cm vertical resolution. Furthermore, we acquired an Orthophoto

<table>
<thead>
<tr>
<th>Building</th>
<th>Flat Roof</th>
<th>Gable Roof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width</td>
<td>12.4 m</td>
<td>32.0 m</td>
</tr>
<tr>
<td>Length</td>
<td>35.9 m</td>
<td>10.3 m</td>
</tr>
<tr>
<td>Aspect Angle</td>
<td>16.3°</td>
<td>52.0°</td>
</tr>
<tr>
<td>Local Incidence Angle</td>
<td>45.5°</td>
<td>40.2°</td>
</tr>
<tr>
<td>Real Height</td>
<td>12.5 m</td>
<td>9.5 m</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of test buildings.
from 09 June 2006 with 30 cm resolution which was used to retrieve the ancillary data needed to initialize the hypotheses.

To demonstrate the potential performance and the generic capabilities of our method we choose two individual, medium sized apartment buildings, one with a flat roof and one with a gable roof, which had different aspect angles in the SAR acquisition. Since our current model does not take into account any effects related to objects which are surrounding the buildings, we chose test candidates which were not affected by a high density of trees. Fig. 1 shows the two building candidates [(a),(b) - flat roof; (c),(d) - gable roof] in an Orthophoto and in the VHR SAR data. The main characteristics of the buildings (i.e. the dimensions and the SAR viewing configuration) are listed in Tab. 1.

4. RESULTS

The result of the height estimation method for the flat roof building is shown in Fig. 2. Plot (a) shows the normalized MI values for hypotheses with an height interval from [3,20] meters in 0.1 m steps. The global maximum is at 12.7 m which is 0.2 m (or 1.6%) higher than the true height. Fig. 2 (b) shows the simulation of the building with the estimated height. For the comparison between the simulation with the estimated height and the actual SAR scene we show in Fig. 2 (c) the subset of the center-left image of Fig. 1, which overlaps with the co-registered simulation.

In Fig. 3 we show the corresponding result for the gable roof building. The estimated height is 9.6 m, which is 0.1 m or respectively 1.1% higher than the true height. Fig. 3 (c) is the subset of Fig. 1 (d), which overlaps with the co-registered simulation of the gable roof building with 9.6 m height shown in Fig. 3 (b).

To get a more realistic and statistically more meaningful performance statement about the proposed method, we run initial tests on a set of 29 objects, comprising flat and gable roof buildings with various local incidence angles and aspect angles with different degree of objects in the immediate surrounding. The overall mean absolute difference for this test set is 2.0 m or 20.8% of estimation error. The error is largely due to underestimates, i.e. the method tends to yield a lower height than the true height of the building.

In a demonstration of the generic nature of our method, we show the results of the height estimation from TerraSAR-X SPOT beam data for the Menkaure Pyramid in Giza, Egypt in Fig. 4. The actual SAR image was acquired on 2 July 2007 in descending orbit with right looking direction in HH polarization with 1.4 meter range resolution and 53 degree incidence angle. The pyramid has a square base of 103.4 meter and an inclination angle of about 51 degree and is made out of limestone and granite blocks, which gives it a staircase like structure. The pyramid has a 5 degree aspect angle with respect to the azimuth direction of the TerraSAR-X image. The true height of the Pyramid is at present 62 m (original height was 65.5 meter), while the estimate of the proposed method is 61.4 m. Given the difference in shape and dimension, in comparison to the residential buildings in the Dorsten set, this example demonstrates the flexibility and the potential of this novel approach.

5. DISCUSSION AND CONCLUSION

In the previous section we showed the first results of the proposed novel method for building height estimation from single VHR SAR detected images. The two building examples have different aspect
and incidence angles and were selected in such a way that they fit well with our model, which assumes that the objects are not immediately surrounded by other objects (e.g. trees). To demonstrate the generic nature of the method we showed the height extraction for the Menkaure Pyramid from VHR TerraSAR-X data. The method was able to extract the height of the presented objects accurately, thus providing an indication of the overall feasibility of the presented approach. Initial tests on an extended dataset of 29 buildings which do not necessarily fit our model (i.e. different degree of objects in the immediate surrounding) showed an average estimation error of about 20%.

For the generation of the hypotheses we require information (i.e. footprint and type of building) from ancillary data. These information can either be available as cadastral maps or can be directly extracted by processing the SAR image or other VHR optical data. With the growing global availability of VHR data from urban areas, these requirements, while demanding, appear to be realistic.

We are currently analyzing the results of the extended data set in detail. Furthermore, we will test the set against a second airborne VHR SAR scene from the same measurement campaign, which was flown approximately perpendicular to the scene used in this paper. From this analysis, we will be able to test the robustness of the approach, especially with respect to invariance to the viewing configuration (i.e. different local incidence and aspect angles). We also plan to test the method with ascending and descending spaceborne TerraSAR-X SPOT beam data already acquired over the Dorsten scene.

**Acknowledgment**

The authors would like to thank René Günzkofer, Marcus Schwäbisch and Hugh MacKay from Intermap Technologies GmbH, Germany, for providing the VHR airborne SAR data and Richard Chiles from NPA Group for providing the LIDAR DSM. Furthermore, we would like to thank Harm Greidanus from the EC-JRC for fruitful discussions on the methodology.

**6. REFERENCES**


