Large-scale Building Height Estimation from Single VHR SAR image Using Fully Convolutional Network and GIS building footprints

Yao Sun[(1)], Yuansheng Hua[(1)], Lichao Mou[(2)], Xiao Xiang Zhu[(1,2)]

(1)*Remote Sensing Technology Institute (IMF)*
German Aerospace Center (DLR)
82234 Weling,Germany

(2)*Signal Processing in Earth Observation (SiPEO)*
Technische Universität München
80333 Munich, Germany

Email: yao.sun@dlr.de, yuansheng.hua@dlr.de, lichao.mou@dlr.de, xiaoxiang.zhu@dlr.de

***Abstract*—Height reconstruction of large-scale buildings from single very high resolution (VHR) SAR image is of great interest especially in applications with temporal restrictions. The problem is highly challenging due to the inherent complexity of SAR images, e.g., side-looking geometry and different microwave scattering contributions. In this work, we present a framework to estimate large-scale building heights from single VHR SAR image. The individual buildings are defined by GIS data, and deep neural network is used to segment wall area in SAR image. The wall layover length is then converted to height and assigned to each building footprint. Experiment in center Berlin area shows results of overall instance height accuracy around 3.51 meters.**

***Index Terms*—building heights, large-scale, SAR, GIS, deep neural network**

I. Introduction

As the spatial resolution of SAR image increases to meter, even sub-meter level, urban structures became visible in SAR image, thus brings urban applications to research interests of many users. Among the applications, height reconstruction is a hot topic since 3-D information is contained in SAR image however is difficult to interpret due to the side-looking geometry and different microwave scattering contributions.

Using SAR images, conventional methods for height reconstruction in large-scale urban area often require multiple acquisitions, e.g., TomoSAR reconstruction [1]. However, multiple images coverage is not available for most areas, especially in the case of emergency response which is of particular interest, only one image can be acquired during the short time. With less SAR images, urban area interpretation is highly challenging. The signals from different buildings and from different building parts, i.e., roofs, walls, are often mixed, as shown in Fig. 1, which brings large uncertainty in building information generation.

In literature, some researchers first detect and extract building features, such as double bounce lines and different back scattering areas, then derive the building hypothesis [2], [3]. Some other researchers assume certain prior information or hypothesis of buildings [4]–[6]. The geometric or radiometric priors are used for simulation. The desired building parameters are progressively achieved by minimizing the difference between simulated data and real data. These aforementioned researches mainly focus on isolated buildings, or building with simple shapes, such as rectangle shape or L-shape. In recent years, deep neural network has been increasingly popular and shown success in many applications including classification and segmentation in remote sensing images. The authors in [7] detect building areas using a fully convolutional network (FCN) with labeled data generated from TomoSAR data. However, the positioning accuracy of the used TomoSAR point cloud is not enough for tasks such as building reconstruction and building height estimation. Thus, it remains a problem to estimate building heights in large-scale urban area using small number of SAR images.

In this work, we propose a framework to estimate large-scale building heights using single SAR image and GIS building footprints. GIS data provides the number, shape, and location of individual buildings. Registered in SAR coordinate with ground height value, GIS data also locates the far-range boundaries of building walls. An operational workflow is proposed for generating labeled data for SAR image segmentation and for building height estimation from wall segmentation result of FCN. Next, we present the methodology of the proposed approach in Section II. The results and evaluation are shown in Section III, which are discussed and lead to a short conclusion in Section IV.

II. Methodology

The workflow contains three parts, as shown in Fig. 2: 1) data preparation, 2) wall segmentation using FCN, 3) layover length estimation and height conversion.

Building heights consist of wall height and roof height. In this work, building height is approximated by wall height for the following reasons: 1), roof in general shows less signatures in SAR image than wall, as the high intensity value mostly comes from corner reflectors, which are often regularly distributed on walls, e.g., windows and balconies. 2), the layovers of wall and roof are often partially or completely

978-1-7281-0009-8/19/$31.00 ©2019 IEEE

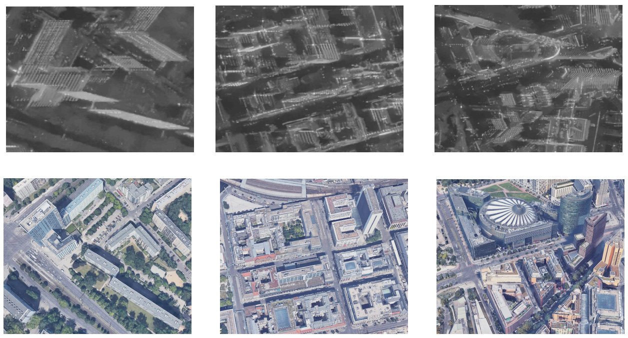

Fig. 1: Examples of building area in: (up) VHR SAR images and (down) optical image (©Google Earth).

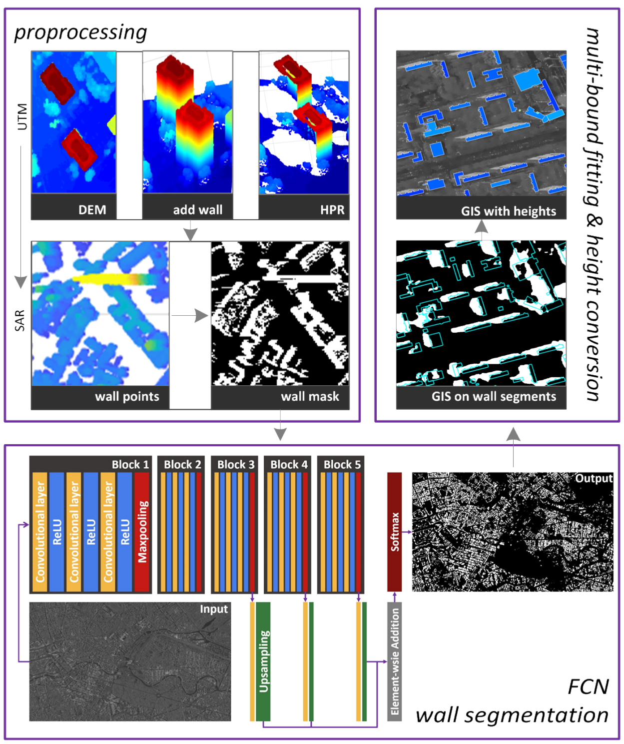

Fig. 2: General Workflow

mixed and are hard to separate. While the ground location of wall is given by GIS building footprints, the boundary of roof is difficult to decide, as shown in case (a) and (b) in Fig. 3. 3), when the layovers of roof and wall are not mixed, the roof height can not be estimated from its layover length without knowing roof inclination angle, as shown in case (c) in Fig. 3: when $\alpha > \theta$, and lr and lw are not mixed, the calculation of length hr needs α besides lr, even if lr is measurable.

Thus, we estimate building height using wall layover length: $hw = lw/cos(\theta)$, knowing that roof height usually is small comparing to building height. Thus the approximation error is within an acceptable range.

A. Data preparation

First, we generate a wall mask in SAR image coordinate as training data for segmentation. For this purpose, an accurate digital elevation model (DEM) is utilized.

Giving each DEM pixel $x - y$ coordinate with its geolocation, and z coordinate with its height value, the DEM data can be seen as a nadir-looking point cloud that vertical structures, such as building walls, are implied in the vertical data gaps from side-view. To obtain the representation of building walls, we locate walls using GIS building footprint polygons and add wall points at these vertical gaps. These added wall points are partially invisible for SAR sensor due to its side-looking geometry and should not be included in wall mask generation. Thus they are removed afterwards using the hidden points removal (HPR) algorithm [8].

Then, the added wall points in DEM point cloud are to be projected to SAR image coordinate to generate the wall mask. However, direct projection between geo-coordinate and SAR coordinate always results in misalignment, due to the fact that the absolute geolocation of SAR images is subjected to various effects such as atmospheric delay and earth tides, and precise absolute geolocation can only be achieved by sophisticated signal processing. For accurate registration, we use the procedure proposed in [9] with an auxiliary data: TomoSAR point cloud in geo-coordinate. By matching DEM point cloud and TomoSAR point cloud, the transformation parameters are solved and are utilized for correct projection.

B. Building Wall segmentation using FCN

Efficient building wall segmentation plays a key role in building height estimation, especially in large scale. We employ a deep neural network (i.e., FCN [10]) to perform wall segmentation for its capabilities of extracting high-level features and exploiting spatial texture in a VHR SAR image for the following considerations: 1) the spatial pattern of backscatters, especially backscatters in building walls, is complicated and difficult to be discovered with conventional methods, and 2) large-scale segmentation requires a robust and high-level segmenter, which is capable of digging out semantic information.

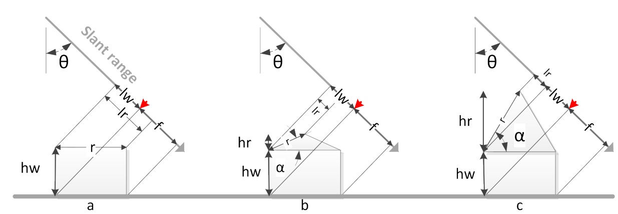

Fig. 3: SAR imaging geometry of flat and gabled roof building. r, hw, hr are sensor visible roof, wall height, and roof height; lw, lr, f are wall layover, roof layover and footprints on SAR image; θ and α are SAR incidence angle and roof inclination angle. The read arrows mark the wall correspondence point of wall signature in SAR image and sensor visible side of building footprints.

Notably, unlike FCN proposed in [10], we attach convolutional layers, where the number of filters is 2, and the size is 1×1, to pooling layers in block3, block4, and block5. Outputs of these convolutional layers are then upsampled to the original resolution before element-wise adding them for generating segmentation maps.

In order to train an effective and robust segmentation network, we first crop the SAR image as well as its corresponding ground truth into patches of 256×256 pixels with a stride of 128 pixels. Afterwards, data is augmented by flipping cropped patches vertically and horizontally. Eventually, 26442 patches are obtained, where 19816 are utilized to train the network, and 9626 are selected to build the testing set. Considering the difference between SAR images and nature optical images, we train the network from scratch instead of employing a pre-trained model. All convolutional filters are initialized with a Glorot uniform initializer, and weights are optimized with an Nestrov Adam optimizer. Parameters of the optimizer are set as recommended: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\theta = 1e08$. The learning rate is set as $2e04$, and decayed by 0.1 when the validation accuracy is saturated. The loss of the network is simply defined as categorical crossentropy. We implement the network on TensorFlow and train it on one NVIDIA Tesla P100 16GB GPU for 100 epochs. Considering that each pixel is a training sample, the size of training batch is set as 5.

C. Multi-bound fitting and height conversion

After wall segmentation, the segment areas are assigned to individual buildings. For each wall segment S, A is its area, B is its bounding box area with a buffer. First, GIS data are aligned to TomoSAR point cloud and then projected in SAR image coordinate. For all footprints polygons in B, and their sensor-visible edges are extracted [11] $E = e_i$. As wall layovers in SAR image always shift from the actual position to near-range side, forming a parallelogram with one pair edges parallel to range direction, to estimate layover length is to find l_i for each e_i, so that S is maximally non-overlapped covered by parallelograms formed by e_i and l_i. It is a constrained multivariable optimization for l_i:

$$E(l) = \min_{l} |abs(A - \sum \vec{e_i} \times \vec{l_i})|,\ lb \leq l_i \leq ub. \quad (1)$$

where lb and ub are lower and upper boundaries, corresponding to minimum and maximum expected building heights.

Height is converted from l_i: $h_i = l_i lw / cos(\theta)$.

III. Experimental results and Evaluation

A. Test results

The test site is at center of Berlin. One TerraSAR-X image in High Resolution SpotLight mode is used, with incidence angle 36 degree. The GIS data is downloaded from Berlin 3D-Download Portal [12]. Fig. 4 (left) shows the test area of GIS data overlaid on SAR image. The DEM has been created from aerial UltraCam-D images in 2006 using stereo processing by semi-global matching and mutual information [13] and provided by the DLR Institute of Robotics and Mechatronics. Since the data are not acquired at same time, the buildings that are not contained in DEM are detected and removed.

Fig. 4 (left) shows the GIS building footprints in study area overlaid on SAR image, while the wall segmentation result of FCN are shown in Fig. 4 (right) . In Fig. 5, the height estimation result is plotted: heights for 4573 out of 6201 building are estimated. Four sub-areas are selected and zoomed in for visualization. The result shows that height is estimated for different buildings regardless of the building shape and the area density. The heights are not estimated for the rest 1628 buildings (footprint polygons plotted in magenta in Fig. 5) mainly for three reasons: 1) the footprint size is too small or narrow that it does not have enough structures on walls to be recognized in SAR image, 2) the buildings are occluded by other buildings in front of them, or 3) the buildings have large footprints but small height, thus the walls are not the significant signatures in SAR image.

B. Evaluation

FCN gives wall segmentation accuracy of 81.50%. For height evaluation, the accurate DEM is used as ground truth. The mean height accuracy is 3.51m for the estimated 4573 buildings: 48.5% with accuracy within 1 m, 73.6% within 3 m, and 9.4% with accuracy bigger than 10 m. The histogram of estimation error is shown in Fig 6.

IV. Conclusion

In this work, we propose a framework to estimate building heights using single VHR SAR image and GIS building footprints. The workflow is tested in Berlin area, and the result shows mean building instance height accuracy of 3.51m for detected buildings. To our knowledge, this is the first time that building height is obtained from single SAR image in this large scale.

There are several concerns we would like to point out. First, the method uses auxiliary data in data matching and training data generation, which may not be available in most cases, thus the extension of the method to general cases is limited. In this respect, data matching techniques should be studied in future for registration of SAR coordinate and geo-coordinate. Second, the building height estimation problem is simplified by neglecting roof height and assigning one height to one building. The neglected height difference should be analyzed. Also, the achievable level of details (LoDs) of reconstructed building parameters from certain used data should be studied. Third, over 25% buildings are not detected: whose footprints are too small or narrow, or show barely any signature in SAR image. These situation should be distinguished, and GIS data should be aggregated or simplified, to quantify the buildings can not be detected in single SAR image.

Other future work includes: improve segmentation results; perform detailed evaluation to understand factors that influence the achievable height accuracy and the LoDs. Last but not least, the wall and roof layovers should be analyzed. Unmixing layovers from different objects remains an open question.

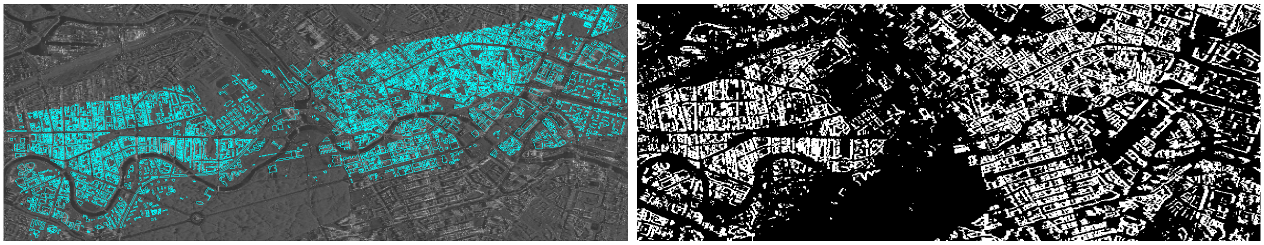

Fig. 4: (left) test area: GIS building footprints on SAR image, and (right) wall segmentation result from FCN.

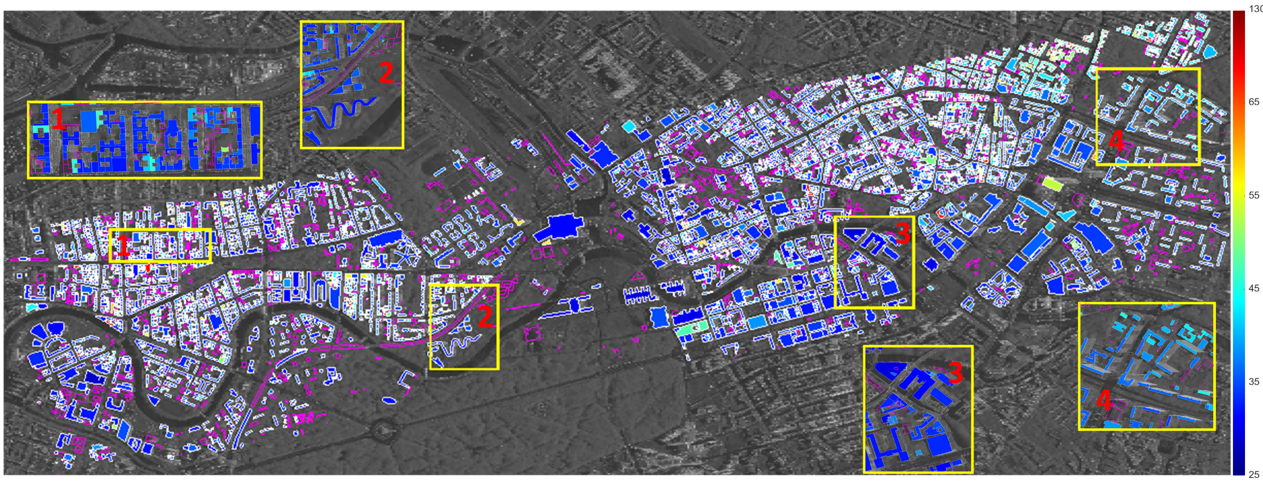

Fig. 5: Height estimation result in test area. Color is height-coded and assigned to each footprints. Buildings footprints that height could not be estimated in the workflow in magenta.

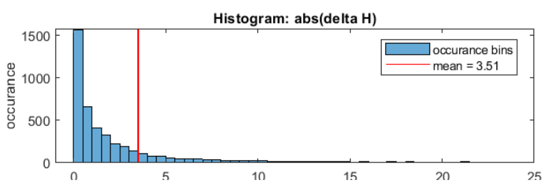

Fig. 6: Histogram of height accuracy of building instances.

ACKNOWLEDGMENT

The authors acknowledge the Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeo.bgu.tum.de), and the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No [ERC-2016-StG-714087], Acronym: So2Sat).

REFERENCES

[1] X. X. Zhu and R. Bamler, “Very high resolution spaceborne sar tomography in urban environment,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, 2010.

[2] 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, no. 7, pp. 2336–2353, 2007.

[3] A. Ferro, D. Brunner, and L. Bruzzone, “Automatic detection and reconstruction of building radar footprints from single vhr sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 935–952, 2013.

[4] E. Simonetto, H. Oriot, and R. Garello, “Rectangular building extraction from stereoscopic airborne radar images,” *IEEE Transactions on Geoscience and remote Sensing*, vol. 43, no. 10, pp. 2386–2395, 2005.

[5] R. Guida, A. Iodice, and D. Riccio, “Height retrieval of isolated buildings from single high-resolution sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2967–2979, 2010.

[6] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, “Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1487–1504, 2010.

[7] M. Shahzad, M. Maurer, F. Fraundorfer, Y. Wang, and X. X. Zhu, “Buildings detection in vhr sar images using fully convolution neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[8] S. Katz, A. Tal, and R. Basri, “Direct visibility of point sets,” in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 24.

[9] Y. Wang, X. X. Zhu, B. Zeisl, and M. Pollefeys, “Fusing meter-resolution 4-d insar point clouds and optical images for semantic urban infrastructure monitoring,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 14–26, 2017.

[10] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[11] Y. Sun, M. Shahzad, and X. X. Zhu, “Building height estimation in single sar image using osm building footprints,” in *Urban Remote Sensing Event (JURSE), 2017 Joint*. IEEE, 2017, pp. 1–4.

[12] Berlin 3D - Downloadportal. [Online]. Available: http://www.businesslocationcenter.de/berlin3d-downloadportal/

[13] H. Hirschmuller, “Stereo processing by semiglobal matching and mutual information,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2008.