AUTOMATIC COREGISTRATION FOR MULTIVIEW SAR IMAGES IN URBAN AREAS

Yuming Xiang\textsuperscript{a,b}, Wenchao Kang\textsuperscript{a}, Feng Wang\textsuperscript{a}, Hongjian You\textsuperscript{a,b}

\textsuperscript{a} Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing, China
\textsuperscript{b} University of Chinese Academy of Sciences, Beijing, China

\textit{Commission I, WG I/3}

KEY WORDS: SAR; Image registration; Urban areas; Building detection

ABSTRACT:

Due to the high resolution property and the side-looking mechanism of SAR sensors, complex buildings structures make the registration of SAR images in urban areas becomes very hard. In order to solve the problem, an automatic and robust coregistration approach for multiview high resolution SAR images is proposed in the paper, which consists of three main modules. First, both the reference image and the sensed image are segmented into two parts, urban areas and nonurban areas. Urban areas caused by double or multiple scattering in a SAR image have a tendency to show higher local mean and local variance values compared with general homogeneous regions due to the complex structural information. Based on this criterion, building areas are extracted. After obtaining the target regions, L-shape structures are detected using the SAR phase congruency model and Hough transform. The double bounce scatterings formed by wall and ground are shown as strong L- or T-shapes, which are usually taken as the most reliable indicator for building detection. According to the assumption that buildings are rectangular and flat models, planimetric buildings are delineated using the L-shapes, then the reconstructed target areas are obtained. For the original areas and the reconstructed target areas, the SAR-SIFT matching algorithm is implemented. Finally, correct corresponding points are extracted by the fast sample consensus (FSC) and the transformation model is also derived. The experimental results on a pair of multiview TerraSAR images with 1-m resolution show that the proposed approach gives a robust and precise registration performance, compared with the original SAR-SIFT method.

1. INTRODUCTION

With the rapid development of the spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X/TanDEM-X and COSMO-SkyMed, numerous high resolution (HR) SAR images have been widely used in commercial and civil areas (Suri and Reinstart, 2010). As the fundamental task of many image applications, SAR image registration is still a challenging task due to the speckle noise and the side-looking imaging mechanism. In this paper, multiview SAR images refer to two SAR images of the same area obtained from the satellite’s ascending and descending orbits. For SAR images in urban or suburban areas, building structures in these areas become visible owing to the high resolution property, offering new possibilities for analyzing and monitoring these areas (Wegner et al., 2010). Consequently, jointly using SAR images from different views of the same area is important and helpful for better building detection, characterization and classification results (Dell’Acqua et al., 2004).

However, visual interpretation of building areas is still challenging due to geometric distortions such as foreshortening, layover, and electromagnetic wave interactions between adjacent targets, together with speckle noise (Zhang et al., 2011). In addition, different building facades may be superimposed since they are on the equidistant line in slant range with respect to the SAR sensor. A lot of efforts have been made to solve these issues in building detection and reconstruction (Xu and Jin, 2007, Ferro et al., 2013, Shahzad and Zhu, 2016), most of them focus on the the line of bright scattering caused by dihedral corner reflector between ground and building wall, which is considered as the only feature that is nearly invariant to the configuration changes (Thiele et al., 2007). However, few research effort considering the complex building problems was put in the image registration topic. Here, in this paper, we propose an automatic coregistration technique for multiview SAR image in building areas. We only assume the buildings as rectangular and flat models; coregistration of more complex building structures will be explored in future work.

As shown in Figure 1, two multiview SAR images are presented. The bright lines are mainly L-shaped and T-shaped structures. It can be observed that the L-shapes from SAR images of different views provides either the front or the rear facade of the same building. Consequently, the L-shapes correspond to the same building structure appear differently. Since the building areas contain plentiful structure information, numerous keypoints can be detected, which may lead to mismatching problems.

In order to solve this problem, an automatic coregistration framework is proposed based on the L-shape detection and reconstruction. Aiming at finding a intermediate description to reduce the difference between the aforementioned L-shapes, we first detect the L-shapes in the reference and sensed images based on a robust ridge detector, then each L-shape is reconstructed to a rectangle structure based on the aforementioned assumption and its geometrical property. Finally, the two reconstructed images are matched by a feature-based registration method.

2. RELATED WORK

Numerous research efforts have been devoted to SAR image registration. The existing algorithms can be roughly divided into...
two categories: area-based and feature-based. Area-based tech-
niques first define a template in the sensed image, then search for
the optimal correspondence in the reference image using different
kinds of similarity measures. The most common used simi-
larity metric is the mutual information as for example used in
(Suri and Reinartz, 2010), where SAR (TerraSAR-X) and optical
(Ikonos) imagery of urban areas is matched. Though the mutual
information has been successfully applied in the coregistration of
multispectral or multisensor remote sensing images because of
its invariance to non-linear intensity changes (Mellor and Brady,
2005), this similarity measure may lead to local extrema and high
computation load (Keller and Averbuch, 2006).

Feature-based techniques first detect distinctive features, such as
points, lines and regions, both in the reference and sensed im-
ages, then construct descriptors based on the local image neigh-
borhood and attempt to find corresponding features(Hänisch et al.,
2016). Among the feature-based methods, the SIFT-like algo-
rithms are the most widely used techniques in SAR image reg-
istration due to the efficient performance and invariance to scale,
rotation and illumination changes (Dellinger et al., 2015). For ex-
ample, Schwind et al. (Schwind et al., 2010) proposed SIFT-OCT
to extract feature points starting from the second octave. Fan et
al. (Fan et al., 2013) improved the matching performance by skip-
ping the dominant orientation assignment when the matching im-
ages do not have rotation transformation. Wang et al. (Wang
et al., 2012) replaced the Gaussian filter with the Bilateral Filter
SIFT (BFSIFT). Dellinger et al. (Dellinger et al., 2015) proposed
the SAR-SIFT algorithm specifically dedicated to SAR images
by utilizing the ROEWA instead of a differential to calculate a
gradient.

However, as no exact keypoint correspondences can be found
among the building areas, the aforementioned methods are no
longer applicable. Only a few researches have considered the
coregistration problem in urban areas. They mainly focus on
the invariant features in the multiview SAR images, or ignore
the variant building structures areas. For example, Acqua et al.
(Dell’Acqua et al., 2004) extracted the cross-roads and road junc-
tions from high-resolution urban SAR images, then uses corre-
sponding pairs of these features to drive the subsequent coregis-
tration. Though the road junctions are invariant in multiview SAR
images, the road detection heavily depends on the performance
and parameters selection of detectors, which are still challenging
works. Han et al. (Han and Byun, 2015) only found correspond-
ing points in homogeneous regions by extracting heterogeneous
regions using a quadtree-based segmentation method. Because
keypoints are mainly detected in heterogeneous regions, the cor-
responding points in homogeneous regions may be not enough
to derive a robust transformation model. Wang et al. (Wang
and Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds in urban area. This method is the most similar to our pro-
posed framework, while it only focuses on multiview TomoSAR
point clouds.

3. METHODOLOGY

The flowchart of the proposed framework is shown in Figure 2. We
do not need to apply a preprocessing method for SAR images,
such as speckle reduction, since the techniques we used for L-
shape detection and image matching are both robust to speckle
noise. As a consequence, no information of the original SAR
image is lost, and the resolution is not degraded (Wegner et al.,
2010).

![Figure 1. Two multiview SAR images of the same area and
comparisons. (a) Ascending orbit. (b) Descending orbit.](image)

![Figure 2. The flowchart of the proposed method.](image)

In the detection process, linear features are first extracted by com-
bining the SAR phase congruency detector (Xiang et al., 2017)
and the hysteresis thresholding method. However, the thresholds
are strongly affected by the template size and image scene, which
needs to be manually set, while it is difficult to define a proper
threshold to distinguish good linear features from bad ones. In
order to solve this issue, an image segmentation method is first
applied before detection to extract target area. Building areas
cauased by double or multiple scattering in SAR images have a
tendency to show higher local mean and local variance values
compared with general homogeneous regions (Esch et al., 2011).
Hence, we use the local mean and local variance of the areas to
drive a robust transformation model. Wang et al. (Wang and
Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds.

In order to solve this issue, an image segmentation method is first
applied before detection to extract target area. Building areas
cauased by double or multiple scattering in SAR images have a
tendency to show higher local mean and local variance values
compared with general homogeneous regions (Esch et al., 2011).
Hence, we use the local mean and local variance of the areas to
drive a robust transformation model. Wang et al. (Wang and
Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds.

In the detection process, linear features are first extracted by com-
bining the SAR phase congruency detector (Xiang et al., 2017)
and the hysteresis thresholding method. However, the thresholds
are strongly affected by the template size and image scene, which
needs to be manually set, while it is difficult to define a proper
threshold to distinguish good linear features from bad ones. In
order to solve this issue, an image segmentation method is first
applied before detection to extract target area. Building areas
cauased by double or multiple scattering in SAR images have a
tendency to show higher local mean and local variance values
compared with general homogeneous regions (Esch et al., 2011).
Hence, we use the local mean and local variance of the areas to
drive a robust transformation model. Wang et al. (Wang and
Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds.

In the detection process, linear features are first extracted by com-
bining the SAR phase congruency detector (Xiang et al., 2017)
and the hysteresis thresholding method. However, the thresholds
are strongly affected by the template size and image scene, which
needs to be manually set, while it is difficult to define a proper
threshold to distinguish good linear features from bad ones. In
order to solve this issue, an image segmentation method is first
applied before detection to extract target area. Building areas
cauased by double or multiple scattering in SAR images have a
tendency to show higher local mean and local variance values
compared with general homogeneous regions (Esch et al., 2011).
Hence, we use the local mean and local variance of the areas to
drive a robust transformation model. Wang et al. (Wang and
Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds.

In the detection process, linear features are first extracted by com-
bining the SAR phase congruency detector (Xiang et al., 2017)
and the hysteresis thresholding method. However, the thresholds
are strongly affected by the template size and image scene, which
needs to be manually set, while it is difficult to define a proper
threshold to distinguish good linear features from bad ones. In
order to solve this issue, an image segmentation method is first
applied before detection to extract target area. Building areas
cauased by double or multiple scattering in SAR images have a
tendency to show higher local mean and local variance values
compared with general homogeneous regions (Esch et al., 2011).
Hence, we use the local mean and local variance of the areas to
drive a robust transformation model. Wang et al. (Wang and
Zhu, 2015) used the end points of the building edges where the
two point clouds close to match two multiview TomoSAR point
clouds.
where $T_i$ and $I$ are the pixel sets of the $i$th region and the entire image, respectively, $\mu$ and $\sigma$ are the mean and variance. We utilize the quadrature-based splitting method in (Han and Byun, 2015). After obtaining target areas, we employ a robust detector based on the phase congruency to extract line features in these areas, followed by the global Hough Transform. Considering the geometrical property of L-shapes, restrictions on orientation and size are applied to extract the location of L-shape. According to the assumption made in Section 1, L-shapes are recovered to rectangle models. Then the recovered target areas and the homogenous areas are combined to form the recovered image.

We use the SAR-SIFT algorithm (Dellinger et al., 2015) as the matching technique. The SIFT-like algorithms have shown good performance for multi-aspect SAR image registration (Schwind et al., 2010). Among them, the SAR-SIFT is the state-of-the-art method, which contains three major steps: keypoints detection, orientation assignment and descriptor extraction, keypoints matching. First, instead of constructing the Gaussian image pyramid, the algorithm starts with constructing a Harris scale space. Replacing the original gradient with the logarithmic ratio of the exponentially weighted averages (ROEW A) operator (Fjortoft et al., 1998), the gradient by ratio is given as follows:

$$G_{b,\alpha} = \log \left( \frac{\sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} I(x+i,y+j)e^{-\frac{|i+i|}{\alpha}}}{\sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} I(x+i,y+j)} \right)$$

$$G_{c,\alpha} = \log \left( \frac{\sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} I(x+i,y+j)e^{-\frac{|i+i|}{\alpha}}}{\sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} I(x+i,y+j)} \right)$$

where $M$ and $N$ denote the size of the neighborhood, $(x, y)$ stands for the location of the feature point, $\alpha$ is the exponential weight parameter, $I$ represents the pixel intensity. Then the SAR-Harris function is computed at different scales, resulting in the Harris scale space. Local extreme in the scale space are selected as keypoints candidates. Second, dominant orientation is assigned to each keypoint to maintain the rotation invariance, which corresponds to the highest peak in the scale-dependent gradient orientation histogram. A circular neighborhood (size of 6x6) and log-polar sectors are employed to construct the descriptor. Third, the keypoint matching stage is similar to the SIFT algorithm, which refers to the Nearest Neighbor Distance Ratio (NNDR) method. Finally, the correct corresponding keypoints are extracted by the fast sample consensus method (Wu et al., 2015), and the transformation model is also derived.

4. EXPERIMENTS

In this section, in order to assess the performance of the proposed registration method, a pair of TerraSAR images is used. The images are obtained under ascending and descending orbits, corresponding to two multi-view SAR images. The parameters of the quadrature-based splitting method and the SAR phase congruency method follow their authors’ instructions, same as the SAR-SIFT method. Moreover, we use the SAR-SIFT method on the original reference and sensed images as a comparison.

Figure 3 show the L-shape detection and reconstruction results of the reference and sensed images. It can be observed that almost all L-shapes have successfully been detected and reconstructed.

5. CONCLUSION

In the paper, an automatic and robust coregistration approach for multiview high resolution SAR images is proposed. Aiming at reducing the difference of building structures from SAR images of different views, we focus on the L-shapes caused by the double-bounce scattering. First, both the reference image and the sensed image are segmented into two parts, urban areas and nonurban areas. Urban areas caused by double or multiple scattering in a SAR image have higher local mean and local variance values compared with general homogeneous regions due to the complex structural information. Based on this criterion, building areas are extracted and L-shape structures are detected in these areas. For both the original areas and the reconstructed areas, the SAR-SIFT
matching algorithm is employed. Finally, correct corresponding points are extracted by the fast sample consensus (FSC) and the transformation model is also derived. Experimental results on a pair of multiview TerraSAR images show that the proposed approach gives a robust and precise registration performance, compared with the original SAR-SIFT method.

ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Project under Grant 41-Y20A13-9001-15/16.

REFERENCES


