Robust Optical-to-SAR Image Matching Based on Shape Properties

Yuanxin Ye, Member, IEEE, Li Shen, Ming Hao, Jicheng Wang, and Zhu Xu

Abstract—Although image matching techniques have been developed in the last decades, automatic optical-to-synthetic aperture radar (SAR) image matching is still a challenging task due to significant nonlinear intensity differences between such images. This letter addresses this problem by proposing a novel similarity metric for image matching using shape properties. A shape descriptor named dense local self-similarity (DLSS) is first developed based on self-similarities within images. Then a similarity metric (named DLSC) is defined using the normalized cross correlation (NCC) of the DLSS descriptors, followed by a template matching strategy to detect correspondences between images. DLSC is robust against significant nonlinear intensity differences because it captures the shape similarity between images, which is independent of intensity patterns. DLSC has been evaluated with four pairs of optical and SAR images. Experimental results demonstrate its advantage over the state-of-the-art similarity metrics (such as NCC and mutual information), and show the superior matching performance.

Index Terms—Dense local self-similarity (DLSS), optical-to-synthetic aperture radar (SAR) image matching, shape properties, similarity metric.

I. INTRODUCTION

IMAGE matching aims to detect control points (CPs) or correspondences between images, and it is a crucial preliminary step for many subsequent remote sensing image processing, such as image registration, image fusion, and change detection [1]. Acquired by different imaging modalities and from various spectra, optical and synthetic aperture radar (SAR) images usually have significant intensity differences, namely, the two types of images present very different intensity patterns despite they cover the same scene (Fig. 1). These differences result in the CP detection becomes much more difficult than a case when multisensor optical images should be matched. Therefore, automatic matching of optical and SAR images could be challenging.

Methods of optical-to-SAR image matching can be generally classified into two categories [2]: feature-based and area-based methods. Feature-based methods use similarities of features extracted from images to achieve CPs between images. Such features mainly include point features [3], line or edge features [4], and contour or region features [5]. Recently, local invariant features have been developed rapidly in the computer vision. Some representative local invariant features such as the scale invariant feature transform and shape context have been used and improved for optical-to-SAR image matching because of their invariance to image rotation and scale changes [6], [7]. However, these methods depend on detecting highly repeatable common features between images, which is a tough task for optical and SAR images due to significant intensity differences, thus degrading their matching performance [8].

As another type of matching methods, area-based methods usually utilize similarity metrics to detect CPs between images through a template matching strategy. Compared with feature-based methods, area-based methods have the following advantages: 1) they avoid to extract common features between images and 2) they can detect CPs within a small search region because most remote sensing images can be directly georeferenced by applying navigation instruments, which make images only have position offsets of several or dozens of pixels [9].

The selection of similarity metrics is crucial for area-based methods. Common similarity metrics include the normalized cross correlation (NCC), the mutual information (MI) [8], and the matching by tone mapping (MTM) [10]. NCC is not adapted to matching of optical and SAR images because it is vulnerable to nonlinear intensity differences. By comparison, MI is more robust to intensity changes and has been applied for optical-to-SAR image matching [8]. However, the MI-based methods often produce many local extrema in the
matching, which may result in some mismatches. MTM is a new similarity metric in the computer vision, and handles intensity differences between images using a tone mapping approximated by a piecewise constant function. Nonetheless, the piecewise constant function cannot exactly fit the intensity relationship between optical and SAR images because of its complexity. In general, these similarity metrics cannot effectively handle nonlinear intensity differences between optical and SAR images because they only use intensity information to detect CPs. Compared with intensity information, shape features or geometric structures of images are more robust to nonlinear intensity changes, as shown in Fig. 1. The shape features between the optical and SAR images are quite similar to some degree though their intensity information is significantly different. Recently, Ye and Shen [11] and Ye et al. [12] proposed a similarity metric based on structural properties, which has been effectively applied in the matching of multisensor remote sensing images. However, this method requires to compute phase congruency features with multiscale and multidirection, resulting in a computationally time-consuming procedure. Differently from this way, this letter will build a shape similarity metric on the basis of local self-similarity (LSS) that captures local shape properties of images [13], [14].

We should note that it can be difficult to employ LSS for optical-to-SAR image matching directly because local shape properties between such images are different (Fig. 1). By comparison, shape properties of a large image region are quite similar (Fig. 1). Accordingly, we develop a similarity metric which can capture the shape similarity between large regions of images. First, a feature descriptor named the dense LSS (DLSS) is built by integrating LSS descriptors of multiple small regions. DLSS aims to capture shape features of a large image region. Then, a similarity metric named DLSC is developed using the NCC of the DLSS descriptors. Finally, DLSC is used to detect CPs between images by a template matching strategy. This letter extends our earlier work [15] by presenting a more principled derivation of this similarity metric. We also give a more thorough evaluation using more experimental data. The code is available in this Web site.\(^1\)

II. METHODOLOGY

This letter aims to build a similarity metric that is robust to significant nonlinear intensity differences between optical and SAR images. The proposed approach is based on the assumption that optical and SAR images share similar shape properties despite they have quite different intensity patterns. In this section, we develop a dense shape descriptor named DLSS and define a similarity metric based on this descriptor.

A. Dense Shape Descriptor

DLSS is inspired from the LSS descriptor that captures internal geometric layouts of local self-similarities within images. LSS describes statistical co-occurrence of surrounding small image patches in an image region. An example of this descriptor generation is illustrated in Fig. 2. First, a correlation surface \(S_q(x, y)\) is calculated by the sum of square differences between all surrounding image patches \(p_i\) and the patch centered at \(q\) in a local region. \(S_q(x, y)\) is normalized by the maximum value of the image patch intensity variance and noise, which is a constant that corresponds to acceptable photometric variations (in illumination or due to noise). It is defined as

\[
S_q(x, y) = \exp\left(-\frac{\text{SSD}_q(x, y)}{\max(\text{var}_{\text{noise}}, \text{var}_{\text{auto}}(q))}\right). \tag{1}
\]

Then, the correlation surface \(S_q(x, y)\) is transformed into a log-polar representation and is partitioned into bins (e.g., 10 angles and 3 radial intervals). The maximal correlation value of each bin is used to form the LSS descriptor associated with the pixel \(q\). Finally, the LSS descriptor is linearly stretched to the range of \([0...1]\) to achieve invariance to intensity variations of different patches in the local region.

LSS is able to capture shape properties of local image regions, but it cannot provide enough information for optical-to-SAR image matching because local shape properties between such images are different (Fig. 1). In contrast, their shape properties of larger image regions are quite similar (Fig. 1). Accordingly, we develop a dense shape descriptor (named DLSS) to capture these properties by integrating the LSS descriptors of multiple local image regions in a dense sampling grid. Fig. 3 illustrates the main processing chain for extracting the DLSS descriptor, and each step of which is as follows.

1) The first step is to select a template window in an image. All pixels within the template window are utilized to build the DLSS descriptor.

2) The second step is to divide the template window into some spatial regions called “cells.” Each cell contains \(n \times n\) pixels, and has an overlapping region of half a cell width with the neighboring cell. This process forms the fundamental framework of the DLSS descriptor.

3) The third step is to extract the LSS descriptor of each cell, and each descriptor is normalized by the L2 norm to achieve a better invariance to intensity changes.

4) Finally, DLSS is built by collecting the LSS descriptor from all cells within a dense overlapping grid covering the template window into a combined feature vector.

B. Similarity Metric Based on Dense Shape Descriptor

As mentioned above, DLSS is a feature descriptor that captures shape features of large image regions. Accordingly, this

\(^1\)https://www.dropbox.com/s/o6f7j0igdiadlvs/DLSC%20publish%20code.rar?dl=0
Fig. 3. Main processing chain for extracting the DLSS descriptor.

Fig. 4. DLSS descriptors of a pair of synthetic images in the corner and edge regions.

descriptor can be used to match two images with significant intensity differences as long as they have similar geometric shape layouts. Fig. 4 shows that the DLSS descriptors computed from a pair of synthetic images with nonlinear intensity differences and noise. One can clearly observe that the two DLSS descriptors are quite similar despite the different intensity patterns between the two images.

Considering the shape similarity of large regions between optical and SAR images, the NCC of the DLSS descriptors is used as the similarity metric (named DLSC) for image matching.

To illustrate its advantage in matching multimodal images, DLSC is compared to NCC, MTM, and MI by the similarity curve. A pair of visible and SAR images with a high resolution is selected in the test. The significant nonlinear intensity differences can be clearly observed between the two images. A template window with a size of $68 \times 68$ pixels is first selected from the visible image. Then, NCC, MTM, MI, and DLSC are computed for the horizontal-direction translations ($-10$ to $10$ pixels) within a search region of the SAR image, respectively.

Fig. 5 shows the similarity curves of NCC, MTM, MI, and DLSC. NCC fails to detect the CP, and MTM and MI also have some location errors caused by the significant intensity differences. By comparison, DLSC not only detects the CP correctly but also presents a more distinguishable curve peak. This example preliminarily illustrates that DLSC is more robust than the other similarity metrics to complex intensity changes. A detailed analysis on the performance of DLSC will be given in Section III.

III. EXPERIMENTS

To evaluate the matching performance of DLSC, it is compared with the three state-of-the-art similarity metrics including NCC, MTM, and MI. The test data sets, implementation details, and experimental analysis are as follows.
A. Date Sets

Four pairs of optical and SAR images are selected as the test data, which are divided into two categories: high-resolution images and medium-resolution images. Fig. 6 shows the image pairs of test sets. All of these image pairs have been systematically rectified by the directly georeferencing technique, and also resampled in the same ground sample distance (GSD). Thus there are almost no obvious rotation and scale differences, only the offsets of a few pixels existing between the reference and sensed images. However, different intensity characteristics can be clearly observed between these images due to the differences of their imaging mechanism (Fig. 6). Table I presents the description of the test sets, and other characteristics of each set are described below.

High-resolution data sets: Test 1 and test 2 are two pairs of optical and SAR images with a high resolution. Since they are located in urban areas, these images have rich shape and contour features. In addition, since there is a temporal difference of 14 months between the images of test 2, some ground objects have changed during this period. These differences in this test make it quite difficult to match the two images.

Medium-resolution data sets: Test 3 and test 4 are two pairs of optical and SAR images with medium resolutions, which are located in the suburban areas. One can observe from Fig. 6 that the images of test 3 have the more significant shape and structure features compared with these of test 4.

B. Implementation Details

In the experiments, the block-based Harris operator [14] is used to detect the 200 evenly distributed interest points in the reference image, and then utilize NCC, MTM, MI, and DLSC to detect CPs within a search region with a fixed size (20 × 20 pixels) of the sensed image using a template matching strategy. In the matching process, template windows with different sizes (from 36 × 36 to 124 × 124 pixels) are used to detect CPs to analyze the sensitivities of these similarity metrics with respect to changes in the template size.

The correct matching ratio (CMR) is chosen as the evaluation criterion, where $CMR = CM/C$, $CM$ is the number of correctly matched point pairs in the matching results, and $C$ is the total number of match point pairs. The $CM$ is determined by the following strategy. For each image pair, 40–60 evenly distributed points were manually selected as the check points, which are used to calculate the residual error for each matched point pair by a projective transformation model. The point pair with residual error less than 1.5 pixels is regarded as the correct match. Moreover, the root-mean-square error (RMSE) [15] of the correct matched point pairs is used to evaluate the matching accuracy.

There are three key parameters in the DLSC, i.e., cell size $n \times n$ pixels, angle number $\beta$, and radial interval number $\alpha$. Their influences have been tested by ten pairs of optical and SAR images. The test results showed the DLSC with the parameters (cell size with $7 \times 7$ pixels, 9 angles, and 2 radial intervals) achieved the optimal CMR value. Therefore, these parameter settings are used in the following experiment.

C. Experimental Analysis

Fig. 7 shows the CMR values of all the experiments. DLSC performs the best, followed by MI and MTM. NCC achieves the lowest CMR values in most tests because it is quite sensitive to nonlinear intensity differences. MTM applies a piecewise linear function to fit the intensity relationship between images to handle intensity differences. However, the intensity relationship between optical and SAR images is too complicated to be exactly fit by a piecewise linear function. Accordingly, MTM cannot also effectively detect CPs between optical and SAR images. In addition, it can be seen that DLSC’s performance is less influenced by template sizes compared with MI, the performance of which is sensitive to template size changes.

The CMR values of MI are usually low when the template size is small (less than $60 \times 60$ pixels). This is because MI needs to calculate the joint entropy between images, which is very sensitive to sample sizes (namely, template sizes). By comparison, DLSC presents the more stable performance when the template size changes, and can achieve a substantial CMR value even in a small template size.

On the other hand, these four similarity metrics present the different matching performance for different test sets due to the differences of image characteristics. For test 1 and test 2 consisting of high-resolution images covering urban areas, DLSC achieves the much higher CMRs than the other similarity metrics [Fig. 7(a) and (b)]. This is because there are the similar shape and contour features such as buildings and roads between the images. Therefore DLSC, capturing the
shape similarity between images, has an obvious advantage compared with the others. For test 3 and test 4 including the medium-resolution images covering suburban areas, DLS's superiority over other similarity metrics declines slightly in comparison to the first two tests. This may be attributed to that the images of the two tests do not contain so rich shape and structure information as well as these of test 1 and test 2. Moreover, for test 4 including the images with quite poor shape and contour features, the CMR value of DLS present an obvious drop compared with the first three test sets. This indicates that the performance of DLS depends on shape properties of images, and could decline when images contain few shape or contour information. However, DLS still performs better than other similarity metrics.

Fig. 8 shows the RMSEs of the four similarity metrics of correct matched point pairs which are detected in the template window with a size of 100 × 100 pixels. One can clearly observe that DLS achieves the highest matching accuracy. The above experiments demonstrate that DLS is a robust similarity metric for optical-to-SAR image matching.

IV. CONCLUSION

In this letter, a novel similarity metric based on shape properties of images is proposed for optical-to-SAR image matching, to address complex nonlinear intensity differences between such images. We first build the DLSS descriptor on the basis of LSS, and then utilize the NCC of such descriptors to define a similarity metric (named DLS) that represents the shape similarity between images. A template matching strategy is employed to detect CPs between images. DLS has been evaluated using four pairs of optical and SAR images, and compared to the traditional alternatives—NCC, MTM, and MI. The experimental results demonstrate that DLS is robust to nonlinear intensity differences between optical and SAR images, and outperforms the other similarity metrics. However, it should be noted that the performance of DLS may decline if images contain few shape or contour information because it depends on shape features of images. This issue will be further addressed in the future work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and good suggestions, and Chatfield et al. [16] for the public local self-similarity code.

REFERENCES