ROBUST MULTIMODAL IMAGE MATCHING BASED ON MAIN STRUCTURE FEATURE REPRESENTATION

Yin Fu[1], Yuanxin Ye[1,*], Guoxiang Liu[1,2,*], Bo Zhang[1], Rui Zhang[1,2]

[1] The Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China;
[2] State-Province Joint Engineering Laboratory of Spatial Information Technology of High-Speed Rail Safety, Southwest Jiaotong University, Chengdu, China;

TC III, WG III/6

KEY WORDS: Image matching, Multimodal images, Nonlinear intensity differences, Main structure

ABSTRACT:

Image matching is a crucial procedure for multimodal remote sensing image processing. However, the performance of conventional methods is often degraded in matching multimodal images due to significant nonlinear intensity differences. To address this problem, this letter proposes a novel image feature representation named Main Structure with Histogram of Orientated Phase Congruency (M-HOPC). M-HOPC is able to precisely capture similar structure properties between multimodal images by reinforcing the main structure information for the construction of the phase congruency feature description. Specifically, each pixel of an image is assigned an independent weight for feature descriptor according to the main structure such as large contours and edges. Then M-HOPC is integrated as the similarity measure for correspondence detection by a template matching scheme. Three pairs of multimodal images including optical, LiDAR, and SAR data have been used to evaluate the proposed method. The results show that M-HOPC is robust to nonlinear intensity differences and achieves the superior matching performance compared with other state-of-the-art methods.

1. INTRODUCTION

Multimodal remote sensing images (e.g., optical, LiDAR, SAR, et al.) can reflect different properties of ground objects, and provide complementary information for surface observation and analysis (Zitova and Flusser 2003). The quantity and quality of information extracted from multimodal data can be increased significantly compared with that obtained from single-modal data. To integrate multimodal images for Earth resources and environment monitoring, one fundamental preliminary task is image registration, which can ensure the spatial consistency of these images. For image registration, the critical step is image matching which detects control points (CPs) between images. Thus, this letter aims to develop a robust matching method for multimodal images.

In general, geometric distortions and nonlinear intensity differences are the main difficulties for multimodal image matching. Current satellite sensors can coarsely correct remote sensing images by on-board global positioning system (GPS) receivers and rigorous physical models, which can eliminate the most global geometric distortions such as rotation and scale differences and only a few pixels offset between images (Bunting, Labrosse, and Lucas 2010; Gonçalves et al. 2012). Accordingly, this paper mainly addresses the matching difficult caused by nonlinear intensity differences between multimodal images. Traditional image matching methods are usually divided into two groups: feature-based methods and area-based methods (Zitova and Flusser 2003). Feature-based methods are the process of extracting a large number of common features (such as points, lines, and regions) from the reference and sensed images, and then match them for image registration. However, it is difficult to extract highly repeatable common features because of significant nonlinear intensity differences between multimodal images, which degrades the matching performance. By contrast, area-based methods perform feature detection and matching simultaneously, avoiding the demand for highly repetitive detection of common features. In general, they define a template window of a certain size and then detect CPs in the corresponding window by similarity measures. In this case, it is important to determine a suitable similarity measure for CP detection. The normalized cross correlation (NCC) and the mutual information (MI) are commonly used as the similarity measures (Gong et al. 2014). However, NCC cannot effectively handle nonlinear intensity differences (Hel-Or, Hel-Or, and David 2014). Although MI can address nonlinear intensity differences to some degree (Viola and Wells 1997; Suri and Reinartz, 2009), it is computationally expensive and prone to mismatches (Ye et al. 2017), which limits its widespread application for multimodal image matching.

Despite great differences in intensity and texture information between multimodal images, it has been found that structure properties of images remain stable in different modalities and can be used as similarity measures for multimodal images matching (Fan et al. 2018; Li et al. 2016;Ye et al. 2019). Recently, Ye et al. (2017) proposed a feature descriptor based on image geometric properties, named Histogram of Orientated Phase Congruency (HOPC), which significantly improve the matching performance. HOPC utilizes all the geometric information to construct the feature descriptor. However, as shown in Figure 1, there hardly exists a complete one-to-one correspondence of structure information between multimodal images. Local structural details, such as small plaque structures, are quite different. Consequently, this feature descriptor has a large amount of redundancy, which degrades the matching performance to some extent. In comparison, the main structures of large contours maintain a stable similarity between images. Therefore, this letter constructs a novel feature descriptor based on the main structure, which is named Main Structure with Histogram of Orientated Phase Congruency (M-HOPC).

* Corresponding author,Yuanxin Ye, yeyuanxin@home.swjtu.edu.cn, Guoxiang Liu, rsgxliu@swjtu.edu.cn

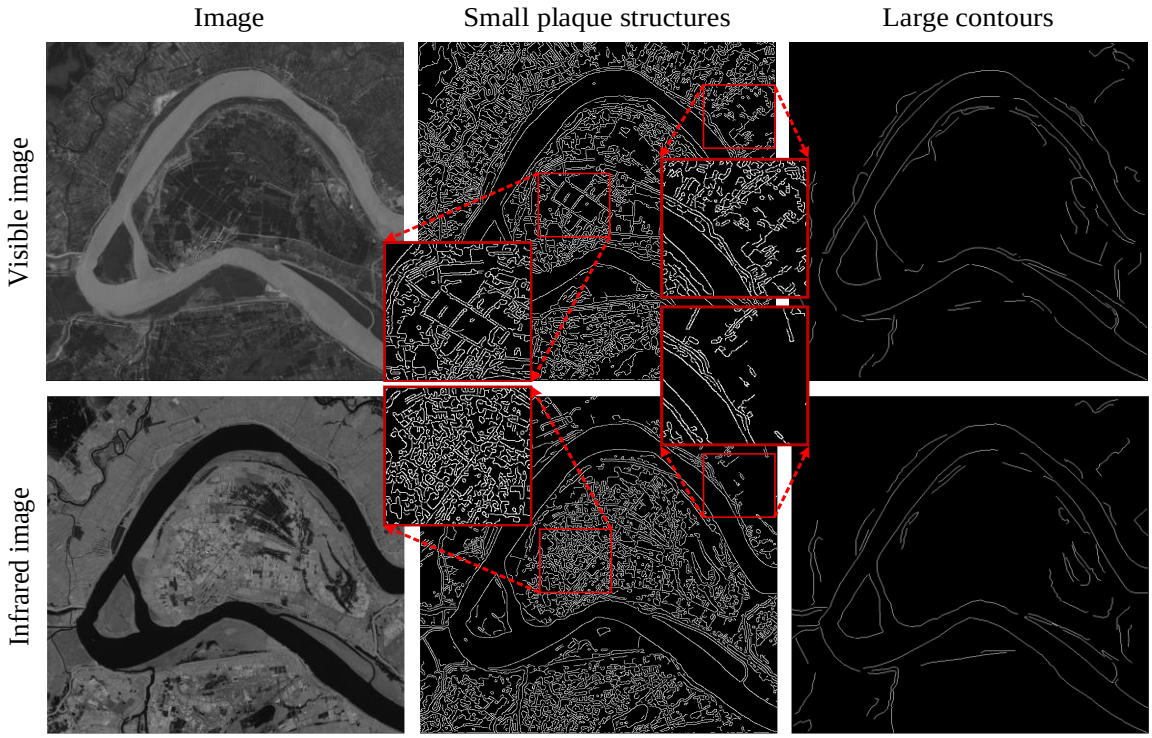

Figure 1. Comparison of small plaque structures and main structure of the large contours. The local details of small plaque structures are quite different (in red frame), whereas the main structures look similar between multimodal images.

Specifically, M-HOPC reinforces the structure similarity between images by weighting the HOPC descriptor at the pixel level using the main structure. The NCC of M-HOPC is subsequently used as the similarity measure for image matching.

2. METHODOLOGY

This section presents a new feature representation method (named M-HOPC), which reinforces the main structure information in the construction of feature descriptor and aims to precisely capture common structure properties between multimodal images. Then, a similarity measure is defined based on M-HOPC, followed by a template matching scheme to detect CPs between images.

2.1 Main structure detection

The main structure is inspired by Guo et al. (2007). They proposed an algorithm for tracking a sketch map, which is actually an extension of the Canny edge detector to model the line segments in the image structure domain (Canny 1986). However, when there are a couple of complicated object features in remote sensing images, the algorithm would extract some broken edge segments, which may interfere with the representation of the structure properties in the feature description. In order to precisely remove the small plaque structure, the main structure acquired procedure is divided into two parts: the sketch pursuit processing and the isolated branch removal. The sketch pursuit processing mainly consists of three phases. In phase 1, for each pixel, multi-scale and multi-orientation Gabor filters are used to initialize the sketch map. In phase 2, one of the maximum intensity edge-lines on the entire image is taken as the initial point, and the remaining edge-line points are searched within a certain area to connect them into a line segment. Repeat this process until no more edge-line points can be connected. In phase 3, a set of predetermined graphical operators is used to correct defects in the corners and connect the current sketch map, which can improve its spatial organization. More details about the sketch pursuit algorithm can be found in Guo, Zhu, and Wu (2007).

Prior to isolated branch removal processing, it is necessary to dislodge some line segments, which are too short in length and scattered over the sketch map obtained in the above process. This step could prevent these line segments from complicating the extraction of the main structural features. The isolated branch removal relies on the processed sketch map. During implementation processing, the linking function is used to generate lists of edge-line points with two nodes (starting and ending points) for each line segments in the sketch map. After that, an adjacency matrix is built for each node to determine which edge-line points lists are linked to the node. This will be used to clean up the isolated segments and branches that are shorter than a set length. Finally, these remaining lists are transferred back into the binary edge image to acquire the main structure consisting of edge contours.

Figure 2 shows an example. (a) is the original image. (b) is the edge detected by the canny detector. Since the subsequent processing is difficult to separate the large contours from the small plaque structures and selectively delete the small patch edges may cause the loss of the main structure information, it is not suitable for extracting the image main structure. (c) is the sketch map, where the large contours can be easily distinguished from other line segments. However, due to the richness of the remote sensing image features, the obtained sketch map is too broken to be directly applied to construct the feature descriptor. Therefore, it is necessary to remove the isolated branch for further processing, and the resulting main

structure, shown in (d), is applied to the subsequent feature description.

2.2 Description of M-HOPC

M-HOPC is constructed by referring to the framework of HOPC, which is based on the finding that structure properties of the same scene in different modalities generally maintain a stable similarity. HOPC has achieved better results than traditional methods. However, there are still some flaws in its feature representation. For example, it exploits all structure information of images to construct the feature descriptor.

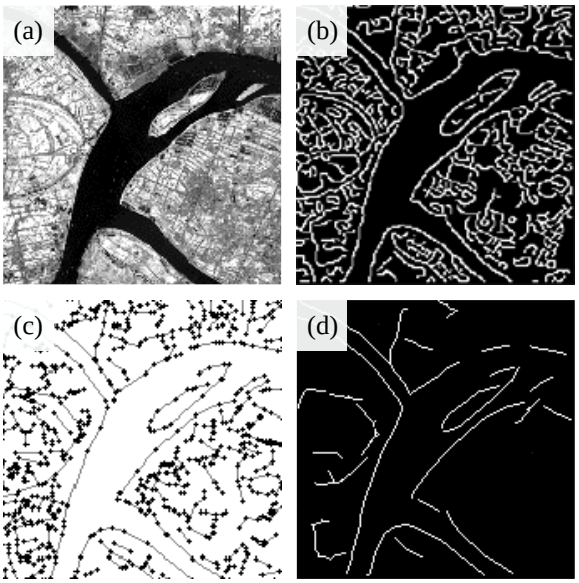

Figure 2. Comparison of different edge detection algorithms, (a) original image, (b) canny edge detection, (c) primal sketch, (d) main structure.

This deteriorates the matching performance because geometric structures are not a complete one-to-one correspondence between multimodal images. In order to strengthen the similarity of structure properties in the feature representation, M-HOPC uses the main structural information as a mask to assign an independent weight to each pixel in the construction of feature descriptor, thereby increasing the proportion of large contours and reducing the adverse impact of local details such as small plaque structures (Padfield 2012). Figure 3 shows the processing chain of the M-HOPC descriptor.

(1) The process starts from the extraction of the original image (Figure 3(a)), which yields a binary edge image of the main structure consisting of 0 and 1 (Figure 3(b)). In Figure 3(c), the distance from the main structure to each pixel is obtained by distance transformation. The image visualization can be represented as: the brightness is proportional to the distance, i.e., the black represents the near distance while the white represents the far distance. Particularly, the distance value of the pixel on the main structure is 0, and it has the darkest brightness in the image. Next, Gaussian distribution is used to calculate the weight of each pixel, which can be expressed as:

$$w(t) = A \cdot exp^{-\frac{d(t)^2}{2\sigma^2}} + c \quad (1)$$

Where A is a coefficient of the normal distribution, and c is the constant that avoids a weight of zero. $d(t)$ denotes the distance from the main structure to each pixel. The weight of each pixel decreases as it is away from the main structure. Therefore, each pixel t has an independent weight $w(t)$ in the template window. A weight template is shown in Figure 3(d), with brightness ranging from black to white and weight values ranging from 0 to 1. In particular, the weight of the pixel on the main structure is 1.

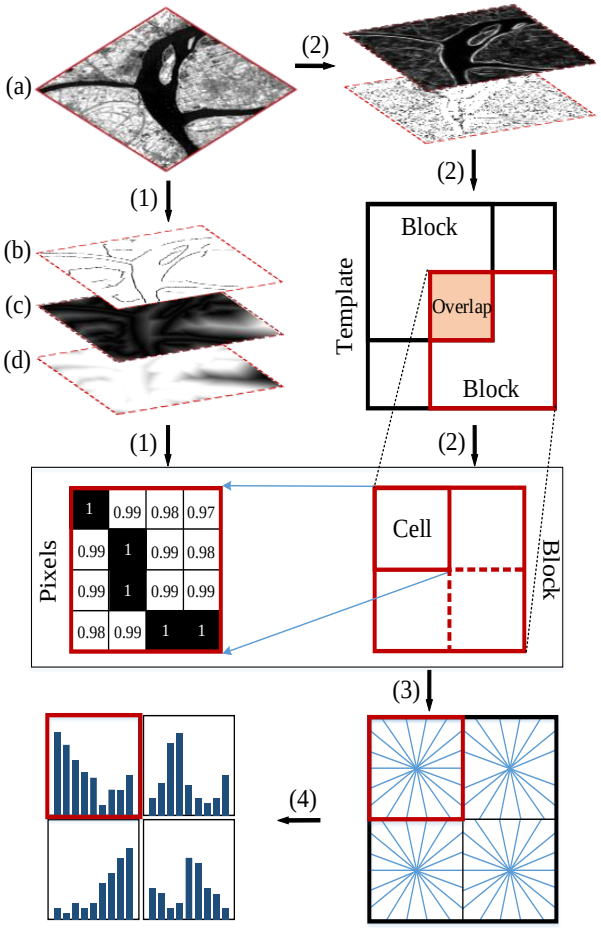

Fig.3 Main processing chain for M-HOPC descriptor.

(2) Following this stage, the phase congruency amplitude and orientation of each pixel are calculated in a template window of a certain size to provide feature information for M-HOPC. The template window is subsequently divided into some overlapping blocks, each block is composed of some fixed-size units, namely “cells”.

(3) In the third stage, an orientation histogram is formed in each cell of the overlapping blocks, where each pixel contributes to the histogram by the weight obtained in (1). Then, the histograms are collected and normalized within blocks. This process forms the M-HOPC descriptor for each block.

(4) At the final stage, the M-HOPC descriptors of all the overlapping blocks in the template window are collected into a feature vector, which can be used to construct the similarity measure for image matching.

2.3 Image Matching Scheme

First, some evenly distributed salient points on the reference image are detected as the interest points through the block-based Harris operator (Ye and Shan 2014). Then, the M-HOPC descriptors are extracted for the template windows centered at

these interest points. Finally, each of these template windows slides pixel by pixel in the search area, and the NCC of M-HOPC is used as the similarity measure to detect CPs on the sensed images.

3. EXPERIMENTS

In this section, the matching performance of M-HOPC is evaluated by comparing with three state-of-the-art similarity measures (i.e., NCC, MI, and HOPC). The image sets, implementation details, evaluation criterion, and experimental analysis are as follows.

3.1 Image Sets

Three pairs of multimodal images are chosen for our experiments. These images have been systematically corrected by rigorous physical models and resampled to the same ground sampling distance (GSD), which removes obvious geometric distortions such as rotation and scale differences between these images. However, due to different imaging mechanisms, there are significant nonlinear intensity differences between images. Table 1 presents the descriptions of the image sets.

Table 1. Descriptions of the test images

Test	Image pair	Size (pixels) and GSD	Date	Characteristic
Test 1	Daedalus visible	512×512, 0.5m	4/2000	Urban area
	Daedalus infrared	512×512, 0.5m	4/2000	
Test 2	LiDAR intensity	550×550, 2m	10/2010	Urban area with high building and noise
	WorldView2 visible	550×550, 2m	10/2011	
Test 3	Google Earth visible	500×500, 3m	3/2009	Urban area with significant noise
	TerraSAR-X SAR	500×500, 3m	1/2008	

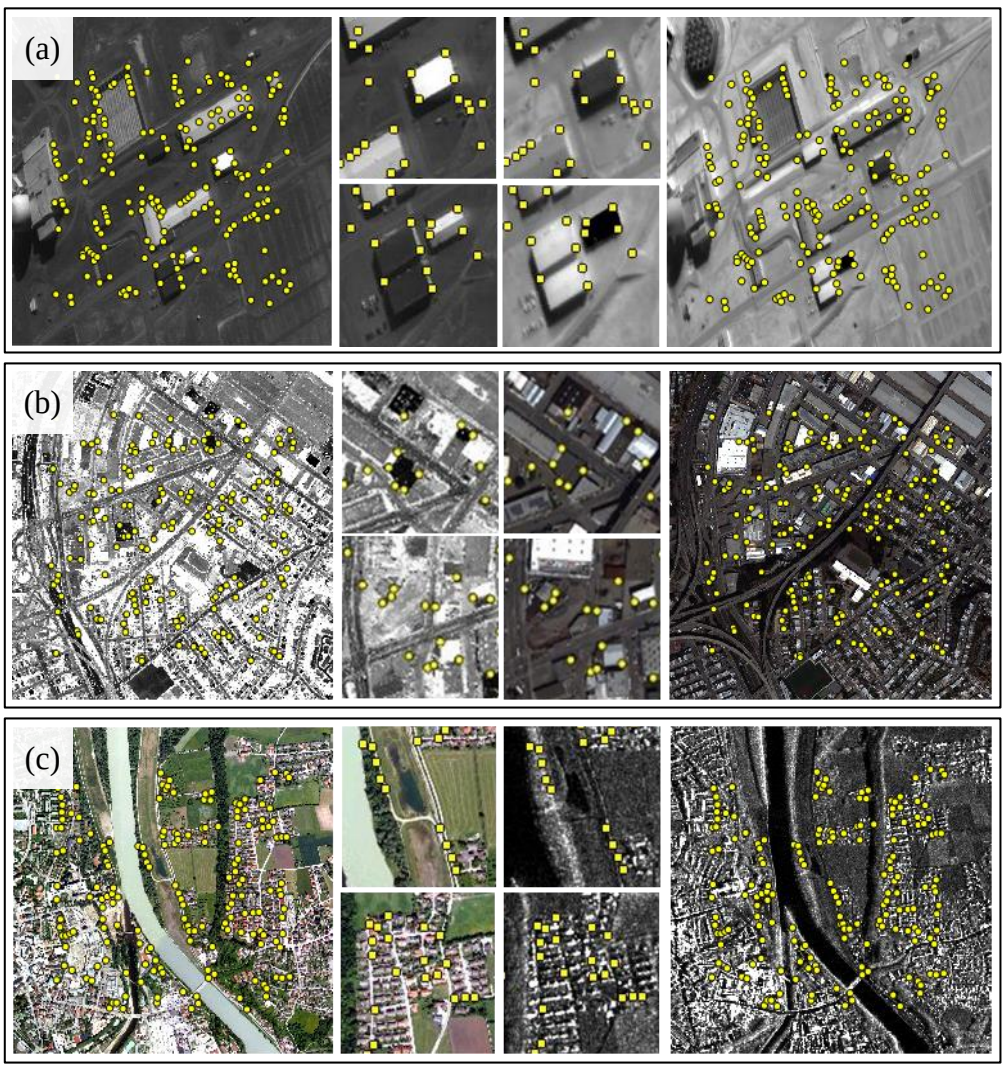

Figure 4. The CPs identified by M-HOPC (template size 100×100 pixels) for image pairs.
(a) Test 1: Visible-to-Infrared. (b) Test 2: LiDAR-to-Visible. (c) Test 3: Visible-to-SAR.

Figure 4(a) is a pair of visible and infrared images, which locates in urban areas with rich structural features. Because of different imaging spectral ranges between images, some ground objects there show the intensity inversion. Figure 4(b) is a pair of LiDAR intensity and visible images, in which the imaging mechanism of LiDAR causes significant noise. Figure 4(c) is a pair of visible and SAR images. Some ground features have changed due to a 14-month temporal difference. All of these differences can cause serious difficulties for image matching.

3.2 Implementation Details

In the experiment, 200 evenly distributed interest points are extracted on the reference image by a block-based Harris operator (Ye and Shan 2014). Subsequently, NCC, MI, HOPC, and M-HOPC are respectively applied to CP detection in the search region (20×20 pixels) of the sensed image, which employs a template matching strategy with different window sizes (Ye et al. 2017) (from 10×10 to 100×100 pixels at 10-pixel intervals). Then the subpixel positions are determined through fitting a similarity surface using a quadratic polynomial model (Ma, Chan, and Canters 2010). Based on the results, we analyse the influence of template size changes on the matching performance of these similarity measures. During the matching process, HOPC and M-HOPC are set to the same parameters for a fair comparison.

3.3 Evaluation Criterion

The similarity surface and profile, the correct match rate (CMR), and the root mean square error (RMSE) are used as evaluation criteria during this procedure.

(1) Similarity surface and profile: normally, the similarity surfaces and profiles reach its peak values when the CP pair is precisely aligned.

(2) CMR: correct match rate is selected as the evaluation criterion, which is defined as CMR=CM/N×100%. CM is the number of correct CP pairs, and N is the number of total CP pairs. If the positioning error is less than 1.5 pixels, the CP pair is considered as a CM.

(3) RMSE: the RMSE of correct CP pairs is used for accuracy evaluation. The matching results with small RMSE values are more accurate than those with large RMSE values.

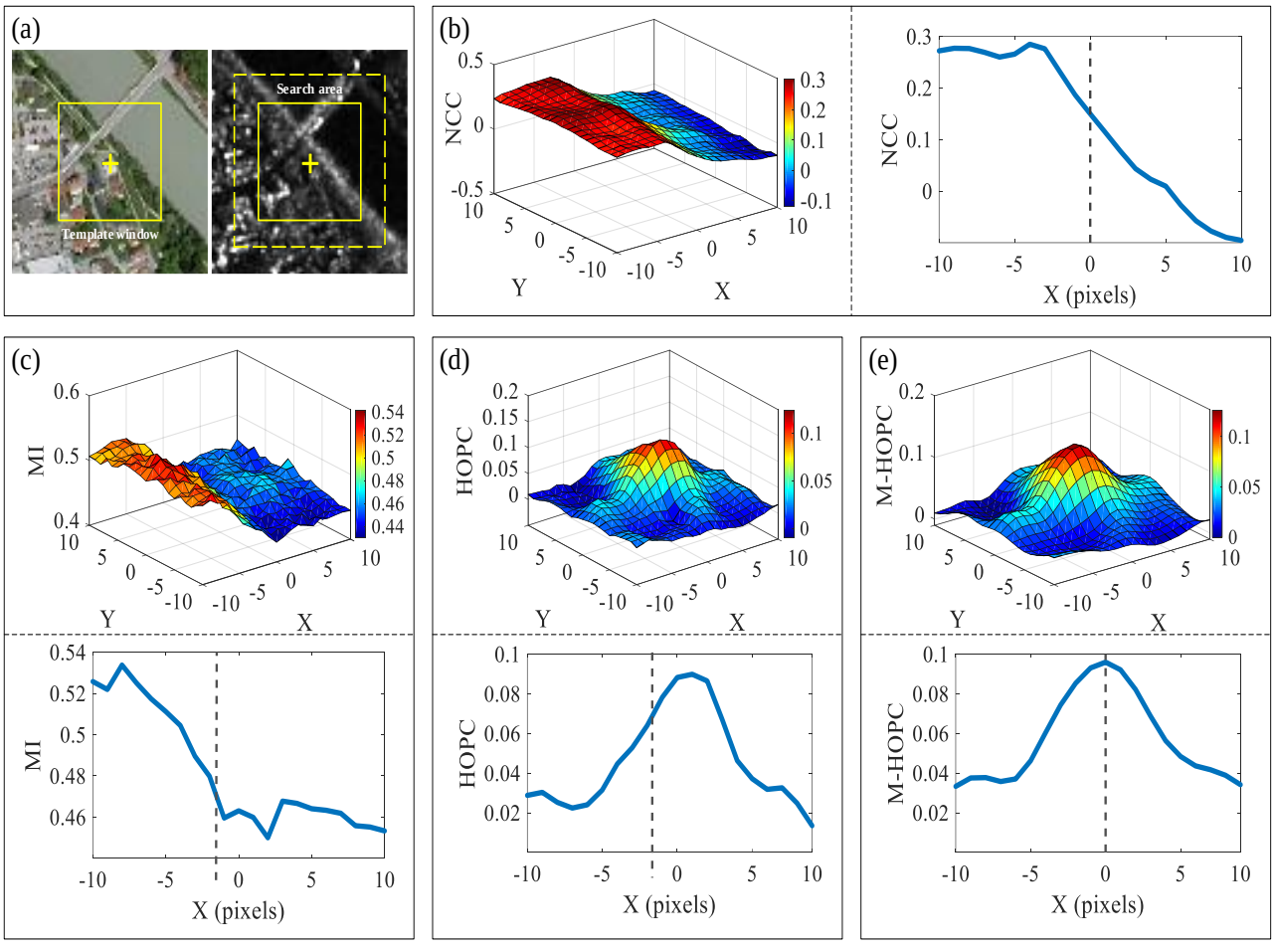

Figure 5. Similarity surfaces and profiles of different similarity measures.
(a) Visible and SAR images. (b) Similarity surface and profile of NCC. (c) Similarity surface and profile of MI.
(d) Similarity surface and profile of HOPC. (e) Similarity surface and profile of M-HOPC.

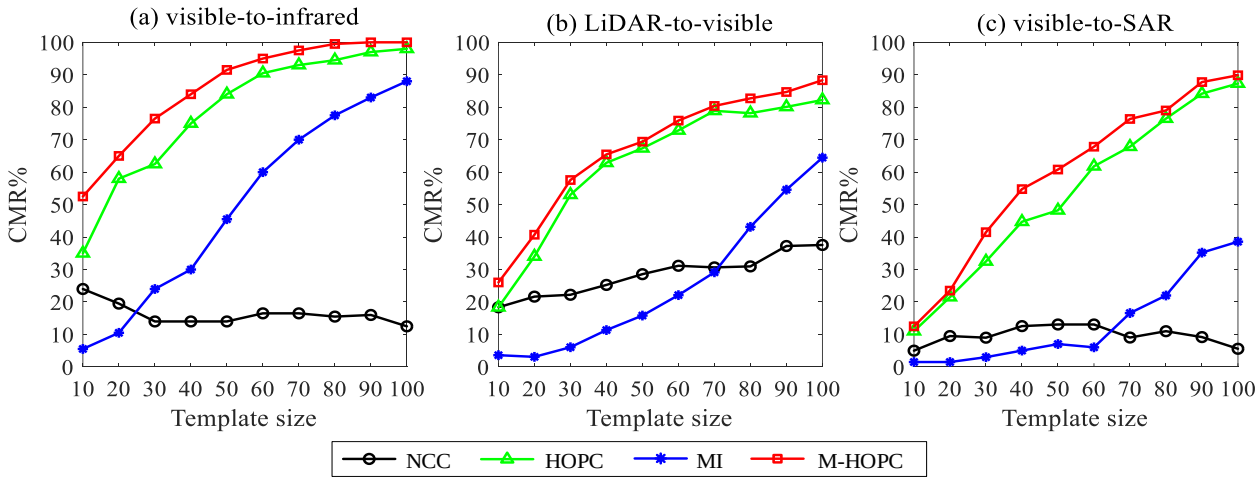

Figure 6. CMR values versus the template sizes of different similarity measures.
(a) Test 1: Visible-to-Infrared. (b) Test 2: LiDAR-to-Visible. (c) Test 3: Visible-to-SAR.

3.4 Experiment Analysis

Figure 5 shows the similarity surfaces and profiles of these similarity measures for a pair of visible and SAR images. NCC and MI present the peak values in the wrong positions, and the peak of HOPC slightly deviates, which indicates that their matching results are inaccurate. Instead, M-HOPC reaches the maximum value at the accurate matching position, and its similarity surface and profile are quite smooth. These results preliminary illustrate that M-HOPC is more robust to nonlinear intensity differences.

Figure 6 shows the CMR values of the four similarity measures for three image sets. M-HOPC performs the best, followed by HOPC and MI. NCC performs poorly and has the lowest CMR value because it is vulnerable to nonlinear intensity differences (Hel-Or, Hel-Or, and David 2014). In addition, with the change of template sizes, the matching performance of MI fluctuates greatly. Taking Test 1 (Visible-to-Infrared) as an example, MI has the lowest CMR value of only 5.5%(with the template size of 10×10 pixels). The reason is that the joint entropy of images calculated by MI is sensitive to the template size (Hel-Or, Hel-Or, and David 2014). In contrast, M-HOPC and HOPC, which based on structure properties, exhibit stable performance. Especially, M-HOPC can achieve a maximum CMR value at 100%, which is more than six times the level of the NCC, likewise higher than MI (83%) and HOPC (97%). This is mainly because M-HOPC reinforces the main structure information to precisely capture similar structure features.

In general, the matching performance of M-HOPC varies for different image sets. Since the high-resolution image set (Visible-to-Infrared) has clear structure and shape information (such as the contour of buildings and edge lines), which is beneficial for the extraction of the main structure features, M-HOPC performs better for this image set than the lower resolution image sets. For Test 2 (LiDAR-to-Visible) and Test 3 (Visible-to-SAR), the performance of M-HOPC is decreased by the existence of noise and insufficient structure information. However, it also achieves higher CMR values than other similarity measures. The CPs detected by M-HOPC are shown in Figure 4.

Figure7 shows the RMSEs of the correct CP pairs detected by the four similarity measures in the template size of 100 ×100 pixels. It can be observed that the M-HOPC has the smallest RMSE value, thus achieves the highest matching accuracy. Overall, these results demonstrate the superiority of M-HOPC in multimodal image matching.

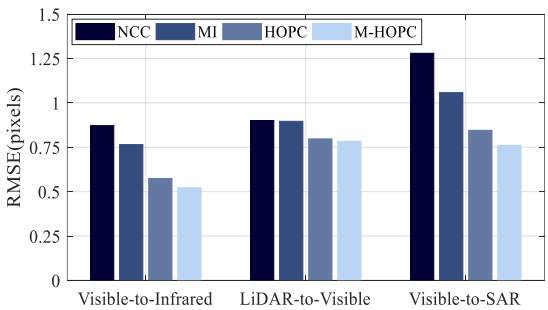

Figure 7. RMSEs detected in the template size of 100 ×100 pixels for the correct CP matches of the four similarity measures.

4. CONCLUSIONS

In this letter, we propose a novel feature representation based on the main structure information of images, named M-HOPC, which aims to address the matching difficulties caused by nonlinear intensity differences. In the construction of M-HOPC, we reinforce the structure feature descriptor by the main structure extracted from images, which enables M-HOPC to precisely capture the structure similarity between images. Then the NCC of M-HOPC is used as a similarity measure to detect CPs. The results of the experiment show that M-HOPC outperforms the other state-of-the-art methods, and improves the matching performance. However, this algorithm suffers from some limitations. If images have strong geometric distortions (such as large rotation and scale differences), it is still not possible to obtain the satisfactory matching results. These restrictions will be further addressed in the near future.

REFERENCES

Bunting, P., Labrosse, F., and Lucas, R. 2010. "A multi-resolution area-based technique for automatic multi-modal image registration." Image and Vision Computing 28 (8): 1203-1219. doi:10.1016/j.imavis.2009.12.005.

Canny, J. 1987. "A computational approach to edge detection." Readings in computer vision 184-203. doi: 10.1016/B978-0-08-051581-6.50024-6.

Dai, X. and Khorram, S. 1998. "The effects of image misregistration on the accuracy of remotely sensed change detection." IEEE Transactions on Geoscience and Remote sensing 36 (5): 1566-1577. doi:10.1109/36.718860

Fan, J., Wu, Y., Li, M., Liang, W. and Cao, Y. 2018. "SAR and Optical Image Registration Using Nonlinear Diffusion and Phase Congruency Structural Descriptor." IEEE Transactions on Geoscience and Remote Sensing 56 (9): 5368-5379. doi:10.1109/TGRS.2018.2815523.

Gonçalves, H., Gonçalves, J.A., Corte-Real, L. and Teodoro, A.C. 2012. "CHAIR: Automatic image registration based on correlation and Hough transform." International journal of remote sensing 33 (24): 7936-7968. doi:10.1080/01431161.2012.701345.

Gong, M., Zhao, S., Jiao, L., Tian, D. and Wang, S. 2013. "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information." IEEE Transactions on Geoscience and Remote Sensing 52 (7): 4328-4338. doi:10.1109/TGRS.2013.2281391

Guo, C.E., Zhu, S.C. and Wu, Y.N. 2007. "Primal Sketch: Integrating Structure and Texture." Computer Vision and Image Understanding 106 (1): 5-19. doi:10.1016/j.cviu.2005.09.004.

Hel-Or, Y., Hel-Or, H. and David, E. 2013. "Matching by tone mapping: Photometric invariant template matching." IEEE transactions on pattern analysis and machine intelligence 36 (2): 317-330. doi:10.1109/TPAMI.2013.138.

Li, Z., Mahapatra, D., Tielbeek, J.A., Stoker, J., van Vliet, L.J. and Vos, F.M. 2015. "Image registration based on autocorrelation of local structure." IEEE transactions on medical imaging 35 (1): 63-75. doi:10.1109/TMI.2015.2455416.

Ma, J., Chan, J.C.W. and Canters, F. 2010. "Fully automatic subpixel image registration of multiangle CHRIS/Proba data." IEEE transactions on geoscience and remote sensing 48 (7): 2829-2839. doi:10.1109/TGRS.2010.2042813.

Padfield, D. 2011. "Masked Object Registration in the Fourier Domain." IEEE Transactions on image processing 21 (5): 2706-2718. doi:10.1109/TIP.2011.2181402.

Suri, S. and Reinartz, P. 2009. "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas" IEEE Transactions on Geoscience and Remote Sensing 48 (2): 939-949. doi:10.1109/TGRS.2009.2034842.

Viola, P. and Wells III, W.M., 1997. "Alignment by maximization of mutual information." International journal of computer vision 24 (2): 137-154. doi:10.1023/A:1007958904918.

Ye, Y., Bruzzone, L., Shan, J., Bovolo, F. and Zhu, Q., 2019. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9059-9070. doi: 10.1109/TGRS.2019.2924684

Ye, Y. and Shan, J. 2014. "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences." ISPRS Journal of Photogrammetry and Remote Sensing 90: 83-95. doi:10.1016/j.isprsjprs.2014.01.009.

Ye, Y., Shan, J., Bruzzone, L. and Shen, L. 2017. "Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity." IEEE Transactions on Geoscience and Remote Sensing 55 (5): 2941-2958. doi:10.1109/TGRS.2017.2656380.

Zitova, B., and Flusser, J. 2003. "Image registration methods: A survey." *Image and vision computing* 21 (11): 977-1000. doi:10.1016/s0262-8856(03)00137-9.