Image registration of SAR and Optical based on salient image sub-patches

Yuchen Ge*, Zhaolong Xiong and Zuomei Lai
Southwest China Institute of Electronic Technology, Chengdu, 610036, China
*Corresponding author’s e-mail: gyc913@sina.com

Abstract. As a fundamental and critical task in multi-source image fusion, the registration of optical image and synthetic aperture radar (SAR) image can identify corresponding identical or similar structures from two heterogeneous images. Although Pseudo Siamese network has achieved notable success in matching heterogeneous images, the network prone to mismatch when there are blurred, duplicate or similar scenes appeared in the search image. To improve the performance, in this paper, the OPT-to-SAR image pair is cut into sub-patch pairs through a pre-defined sliding window. Then a two-stage image filtering mechanism is proposed to maintain candidate sub-patches with ideal texture information. After sending all the qualified sub-patch pairs into the Pseudo Siamese network, the final matching result will be passed through a RANSAC module. In this way, the interference from invalid image areas can be reduced and the model robustness can be ensured by using the statistic information of the whole image pair. A series of experiments conducted under various data scenarios proved the effectiveness of the method.

1. Introduction
Among the multi-source image data, SAR and optical images are the most typical: optical images conform to the visual characteristics of human eyes, while synthetic aperture radar (SAR) images have the characteristics of all-weather and high-resolution. Therefore, the registration of SAR and optical images helps to describe the same scene more comprehensively and objectively. Since that optical images usually contain significant geo-location errors, geocoding cannot be directly relied on to provide accurate matching with SAR satellite images, so it is necessary to rely on heterogeneous image registration technology for revision. Heterogeneous image registration is one of the key technologies of multi-source image processing, which will directly affect the accuracy and effectiveness of image fusion, change detection and target recognition. In the process of generating optical images, it is easy to produce blurry noise due to the influence of clouds and light, while SAR images belong to coherent imaging of oblique-distance projection, and inevitably produce coherent speckle noise, stagnation, perspective, and radar shadow. Therefore, in the process of matching optical and SAR satellite images, it is faced with the problem of how to deal with the significant geometric and radiation differences between the two data sources. In addition, because the image texture is scarce in areas such as the sea surface and lake waves, even if the interferences of imaging noises are excluded, the image in such area does not have significant texture characteristics. Therefore, the direct application of the traditional matching method cannot achieve satisfied result.

With the success of deep learning technology, neural network convolution operators can obtain more complex and robust features than artificial operators and have the ability to extract strong illumination changes and strong radiation differences between SAR and optical images [1-3].
However, the feature extraction network only obtains the feature vector of the entire image pair, and the spatial statistical information of the image block is lost. Besides, when blurry, repetitive or similar scenes appear in the search image, the network is prone to mismatch.

In this paper, the OPT-to-SAR image pair is cut into sub-patches through a pre-defined sliding window. Then a two-stage image filtering mechanism is proposed to maintain candidate sub-patch pairs with ideal texture information. After sending all the qualified sub-patch pairs into the Pseudo Siamese network, the final matching result will be passed through the Random Sampling and Consensus (RANSAC) module. In this way, the interference from invalid image areas is reduced and the model robustness is guaranteed by utilizing statistic information of the whole image pairs.

2. Model

2.1. Overall Architecture

As shown in Figure 1, the overall architecture of the proposed model mainly consists of 4 parts: sub-patch pairs for the original OPT-to-SAR image pair by sliding window, qualified sub-patch pairs selection, Pseudo Siamese network for sub-patch pairs and RANSAC operation to get the final registration result. Details of these parts will be described as follow.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The overall architecture of the proposed model.}
\end{figure}

2.2. Sliding window

The sliding window method is used to sample the SAR and optical images at equal interval to collect sub-patch pairs. Specifically, since that the prior geographic information of the SAR satellite image is more accurate, the imaging range is larger, and the acquired image is more uniform than the spliced optical image, so the SAR images is more suitable for the search patches. In this solution, the SAR and optical images pre-processed by geocoding will be intercepted according to the established window size \( w \times w \) and step size \( s \) to obtain a sequence of sub-patch pairs of length \( N \):

\[
[(P_{\text{sar}1, \text{opt}1}), (P_{\text{sar}2, \text{opt}2}), \ldots, (P_{\text{sar}N, \text{opt}N})].
\]

2.3. 2-stage image patch filter

In order to improve the robustness of the matching, it is necessary to search through all the sub-patch pairs, and only retain candidates with significant texture in both the SAR domain and the optical domain. In order to provide a fast search solution, a two-stage image patch filter algorithm is proposed.

The first step is to quickly block out blurred areas covered by clouds, fog, and rain as well as large areas without obvious features, such as the sea or lake in optical domain. As a robust criterion for describing different IR background [4], especially the sea-sky scenes, the variance weighted information entropy (WIE) is adopted [5,6].

Let \( S \) donate the set of grey-values in an optical image patch with 256 grey levels, \( P_s \) be the probability of grey-values \( s \) occurring in the set \( S \), \( \bar{s} \) is the mean of grey-values in the optical image patch, the variance WIE of the patch can be formulated as:

\[
H(S) = -\sum_{s=0}^{255} (s - \bar{s})^2 \cdot p_s \log p_s \quad \text{When } p_s = 0, \text{ let } p_s \log p_s = 0
\]
In order to increase the calculation rate, the input optical blocks with size $s_w \times s_h$ are resized to a fixed size $k \times k$, where $k < s_w$ and $k < s_h$. After getting all the variance WIE scores in optical image patches, threshold $Th_{WIE}$ is set to filter out block pairs $(P_{sar}, P_{opt})$ with $WIE(P_{opt}) < Th_{WIE}$.

The second step to select sub-patch pairs with significant texture feature in both the SAR domain and the optical domain through Saliency detection. To give a fast selection on gray images, the LC algorithm is adopted [7]. The saliency value of a pixel $I_k$ in an image $I$ is defined as:

$$Sal(I_k) = \sum_{I \in I_k} || I_k - I ||$$ (2)

Where the saliency value of a pixel $I_k$ is in the range of [0, 255], and $|| \cdot ||$ represents the color distance metric. To get an exact value that describes the extent of sub-patch saliency, $Score_{sal}$ is defined in this paper as follow:

$$Score_{sal} = \sum_{I \in I_k} I_{255} - \frac{Sal(I_k) - \min(Sal(I))}{\max(Sal(I)) - \min(Sal(I))} \times 255$$ (3)

$(P_{sar}, P_{opt})$ are both sent into the second step to get $(Score_{sal}(P_{sar}), Score_{sal}(P_{opt}))$. Threshold $Th_{sal,sar}$ and $Th_{sal,opt}$ are set to filter out patch pairs $(P_{sar}, P_{opt})$ with $Score_{sal}(P_{sar}) < Th_{sal,sar}$ or $Score_{sal}(P_{opt}) < Th_{sal,opt}$.

2.4. Pseudo-Siamese networks

The next step is to find a matching key point between $(P_{sar}, P_{opt})$. Inspired from [1-3], A Pseudo-Siamese network is adopted to take each $(P_{sar}, P_{opt})$ qualified by the 2-stage image patch filter into the two branches, extract features from two conjoined neural networks that do not share weights and undergo a correlation calculation module to obtain a matching thermal map of SAR and optical sub-patch pairs. To promote model convergence, this paper performs Gaussian operation on the single-point heat map label in the training phase. It should be noticed that the Gaussian operator does not affect finding the final matching result with the largest correlation score.

![Figure 2. The Pseudo-Siamese network](image)

2.5. Outlier remover

Considering that the matching results of sub-patch pairs cut out from the same original OPT-to-SAR image pair should be statistically consistent. This paper uses Random Sampling and Consensus (RANSAC) [8] module to eliminate abnormal matching pairs and obtain the final registration result.
3. Experiment Settings

3.1. Data description

Our experimental data consists of two parts. The optical image data comes from Google Maps and the SAR image data comes from Sentinel 1[9]. The original dataset is perfectly matched. In order to make the test sample, the original opt-SAR image pair are cut into $4000 \times 4000$ and artificially created with a translation of 128 pixels on each of the x and y axes.

3.2. Implementation Details

After geocoding preprocessing, the original SAR and Optical images are cut into training sample through a sliding window. Specifically, the window size is set to $256 \times 256$ with step = 64. While the SAR sub-patches can be got directly, the optical sub-patches are randomly cut into $128 \times 128$ within the region and the Gaussian labels are set accordingly. To get a convincing threshold for the 3 thresholds in 2-stage image patch filter. We performed sampling and numerical statistics on the image patch pairs.

![Image](a)

(a)

![Image](b)

(b)

![Image](c)

(c)

Figure 3. The histogram of entropy and Saliency score. (a) The histogram of variance WIE on sampled optical sub-patches. (b) The histogram of Saliency value on sampled SAR sub-patches. (c) The histogram of Saliency value on sampled optical sub-patches

According to the distribution of the histogram as well as the sampling data itself, the threshold $Th_{WIE}$ is set to 4.5, $Th_{sal_{opt}}$ is set to $1.8e^6$ and $Th_{sal_{sar}}$ is set to $3.9e^6$. After selection, 8000 OPT-to-SAR image patch pairs are sent into the model as training samples. The training batch is set to 16 and the optimizer is Adam with $1e^{-4}$ learning rate. After 500 epochs, the model is finally converged.

3.3. Experimental Results

In order to show the effectiveness of the saliency texture filtering algorithm, Figure 4 shows the ablation study of the registration result for all sub-patch pairs after performing RANSAC on 9 test
images. Test number 0 and 8 are the images with large sea scenes. Test number 1 and 6 contain rich textures throughout the image. Test numbers 2, 3, 4, 5, and 7 are images with partial lake views, or images with similar repeated textures on rice fields and buildings. The unfiltered registration results are shown in dashed lines, and the filtered results are shown in solid lines. It can be seen that the unfiltered matching error is usually higher than the filtered matching error.

![Figure 4. The ablation study of the registration result for all sub-patch pairs after RANSAC.](image)

In order to observe the matching effect of filtering preprocessing in different scenarios, three matching result with lakes and sea scenes are shown in Figure 5. It can be seen that sub-patches with insignificant textual information do not contribute to the registration process, and the matching outliers are excluded by the RANSAC method. The average error value of coordinates for image Figure 5 (b), Figure 5 (d), Figure 5 (f) after matching are 14.35, 13.9 and 11.68.
Figure 5. The matching result. (a), (c), (e) give the registration result of all qualified opt-SAR image sub-patches on three different scenes; (b), (d), (f) gives the registration result after RANSAC respectively.

4. Conclusion
In this paper, we seek to find an OPT-to-SAR registration algorithm based on Pseudo Siamese network by cutting the image pair into sub-patches and selecting qualified salient sub-patches with rich texture information. The RANSAC module is used to eliminate abnormal matching pairs. In future work, we plan to conduct a more in-depth study on the statistical analysis of the fusion of multi-dimensional matching results and the improvement of model structure.

References
[1] Lhh A, Dm B, Slb C (2020). A deep learning framework for matching of SAR and optical imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 169:166-179.
[2] Merkle, N., Luo, W., Auer, S., Müller, R., Urtasun, R., (2017) Exploiting deep matching and SAR data for the Geo-localization accuracy improvement of optical satellite images. Remote Sensing 9 (6), 586
[3] Hoffmann, S., Brust, C.-A., Shadaydeh, M., Denzler, J., (2019) Registration of high-resolution SAR and optical satellite imagery using fully convolutional networks. In: Proc. IEEE International Geoscience and Remote Sensing Symposium. IEEE, 5152–5155.
[4] Yang L, Yang J, Yang K (2004). Adaptive detection for infrared small target under sea-sky complex background. Electronics Letters, 40(17):1083-1085.
[5] Yang L, Zhou Y, Yang J (2006) Variance WIE based infrared images processing[J]. Electronics Letters, 42(15): 857-859.
[6] Cao Z, Ge Y, Feng J, (2014) Fast target detection method for high-resolution SAR images based on variance weighted information entropy[J]. Eurasip Journal on Advances in Signal Processing, (1):45.
[7] Zhai, Yun & Shah, Mubarak. (2006). Visual attention detection in video sequences using spatiotemporal cues. Proceedings of the 14th Annual ACM International Conference on Multimedia, 815-824.
[8] Martin A. Fischler, Robert C. Bolles (1987), Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Readings in Computer Vision, 726-740.
[9] Attema E, Davidson M, Floury N (2007). Sentinel-1 ESA's new European SAR mission.Proceedings of SPIE - The International Society for Optical Engineering.