Dense Matching Method for UAV SAR Images Without Epipolar Rectification

Yunhao Chang, Xin Xiong, Qing Xu, Guowang Jin, Guoping Zhang, and Ruibing Cui

Abstract— Dense matching is an important step in radargrammetry. As approximating epipolar lines in synthetic aperture radar (SAR) images is difficult, conventional dense image matching (DIM) algorithms are unsuitable for these images. This study proposes a DIM algorithm for unmanned aerial vehicle (UAV) SAR images that does not require epipolar rectification. The proposed algorithm uses tie-point matching results to construct a search window for corresponding points and utilizes an improved DAISY descriptor incorporating the ratio of exponentially weighted averages (ROEWAs) operator for cost calculation, which suppresses errors caused by speckle noise. Cost aggregation was performed using computationally efficient superpixel segmentation and the guided filter (GF) algorithm, and the winner-takes-all (WTA) strategy was applied for dense matching. Finally, experiments were performed on six pairs of UAV SAR images containing different terrains and ground objects, and an average root mean square error (RMSE) of 4.3 pixels was obtained, demonstrating that the proposed method is superior to conventional DIM algorithms and has excellent precision and accuracy.

Index Terms— DAISY descriptor, dense matching, guided filter (GF), superpixel segmentation, synthetic aperture radar (SAR), unmanned aerial vehicle (UAV).

I. Introduction

WITH the advent of civilian unmanned aerial vehicles (UAVs) and the miniaturization of synthetic aperture radar (SAR) systems, radargrammetry using UAV-borne SAR systems has become possible. These systems combine the flexibility, specificity, and cost efficiency of UAVs with the 24-h all-weather capabilities of SAR [1], making them tremendously useful in applications, such as geographic and emergency surveys, as well as environmental monitoring [2]. Dense image matching (DIM) is an important step in radargrammetry for constructing digital surface models (DSMs) or digital elevation models (DEMs) [3]. As such, the precision of DIM directly determines that of the DSM or DEM. UAV SAR images tend to have a large relative image distortion owing to their unique imaging characteristics, which makes DIM a challenging task. Therefore, a DIM method suitable for UAV SAR images is urgently needed.

DIM methods for optical images have matured, with algorithms falling into three categories: local matching, global matching, and semi-global matching (SGM). SGM algorithms combine the strengths of local and global matching algorithms and are thus accurate and computationally efficient [4]; therefore, most commercial DIM software packages use SGM algorithms. The procedures of an SGM algorithm include epipolar rectification, cost calculation, cost aggregation, disparity computation, and disparity refinement. The characteristics of SAR images based on optical DIM algorithms have been extensively researched. Some scholars' research is based on SGM to improve, such as using normalized cross correlation (NCC), absolute difference [5], phase-only correlation [6], sum of adaptive NCC [7] and other algorithms for cost calculation, or adding hierarchical pipeline [8] for DIM. There are also algorithms, such as improved scale-invariant feature transformation (SIFT) and multiscale coherent point drift [9] to estimate deformation field parameters, or geometric constraints and region matching [10], or improved Horn–Schunck flow algorithms [11] to complete pixel-by-pixel matching between SAR images.

Most of these methods require the generation of epipolar images prior to dense matching. However, unlike optical images, SAR images are captured in the slant range and therefore have object-dependent epipolar lines [12]. The epipolar correction of spaceborne SAR images is based on the assumption that the epipolar lines are straight lines, but this assumption does not apply to UAV SAR images [13].

Therefore, this letter works on UAV SAR images to explore how to achieve dense matching with better results without epipolar correction, and a method is designed. In radargrammetry, tie-point matching is performed before dense matching, and the results can be used to define the cost-search window, thus avoiding the process of epipolar correction. The method incorporates the ratio of exponentially weighted averages (ROEWAs) operator into the DAISY [14] descriptor's cost calculations, enhancing speckle noise suppression [15]. Superpixel segmentation and a guided filter (GF) are employed for cost aggregation, producing smoother and more reliable results while maintaining computational efficiency.

Yunhao Chang and Ruibing Cui are with the Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou, Henan 450001, China, and also with Henan College of Surveying and Mapping, Zhengzhou, Henan 450000, China (e-mail: cyh19910216@outlook.com; cuiruibingzz@163.com).

Xin Xiong is with the Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou, Henan 450001, China, and also with the Key Laboratory of Smart Earth, Beijing 100094, China (e-mail: xiongxinhbhh@163.com).

Qing Xu, Guowang Jin, and Guoping Zhang are with the Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou, Henan 450001, China (e-mail: 13937169139@139.com; guowang_jin@163.com; zaczhang96@whu.edu.cn).
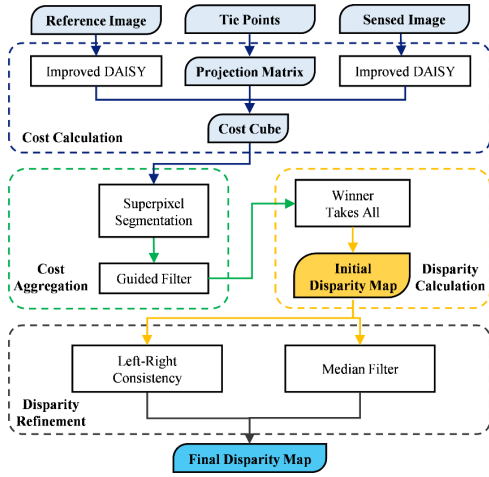
Fig. 1. Process flow of the DIM algorithm, including cost calculation, cost aggregation, disparity calculation, and disparity refinement.

This letter is organized as follows. Section II describes the proposed DIM algorithm, while Section III provides and analyzes the experimental results obtained using this algorithm. The conclusions are given in Section IV.

II. METHODS

A DIM method typically includes cost calculation, cost aggregation, disparity calculations, and disparity refinement. First, the DAISY descriptor (including the ROEWA operator) is used to describe the reference and sensed images. After one descriptor is obtained for each pixel, cost-search calculations are performed on the projection matrix H derived from tie-point matching, thus producing the cost cube. Cost aggregation is then performed on each layer of the cost cube using super-pixel segmentation and GF, using the winner-takes-all (WTA) strategy to calculate the cost and disparity of each pixel. Finally, left-right consistency checking (LRC) and median filtering are applied on the disparity matrix to produce the final disparity map. The process flow is shown in Fig. 1.

A. Cost Calculation With Improved DAISY Descriptor

In the DAISY descriptor, gradients in specific directions are convoluted with several Gaussian filters to compute the weighted sum of gradient norms [14]. As SAR images often contain high speckle noise, our algorithm uses the ROEWA operator for gradient norm calculations, which suppresses speckle noise in SAR image matching algorithms, such as SAR-SIFT.

To obtain DAISY descriptors, the eight orientation maps $G_o(1 \leq o \leq 8)$ for the given input image $I(x, y)$ must first be computed, where $G_o(x, y)$ is the gradient norm for direction o. G_o is written as $G_o(x, y) = (\partial I/\partial o)^+$, where $(\cdot)^+ = \max\{\cdot, 0\}$. These orientation maps are then convolved with σ-parameterized Gaussian kernels to obtain convolved orientation maps for different sized regions, $O_i^\sigma(x, y) = G_i(x, y) * \mathcal{G}_\sigma(x, y)$. The local descriptor is defined as coefficients generated by convolved orientation maps located on concentric circles

$$h_\sigma(x, y) = [\mathcal{O}_1^\sigma(x, y), \mathcal{O}_2^\sigma(x, y), \ldots, \mathcal{O}_8^\sigma(x, y)]. \quad (1)$$
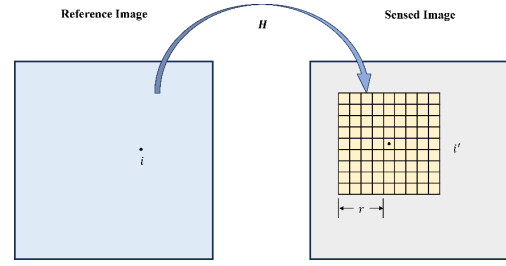

Fig. 2. Search and calculation of matching cost. The grid in the sensed image represents the search range, and H is the projection matrix.

Since the DAISY descriptor has M circular cells, its full expression can be obtained by concatenating the normalized vectors

$$
\begin{aligned}
D(x, y) = \big[& h_{\sigma_1}^T(x, y), \\
& h_{\sigma_1}^T(d_1(x, y, R_1)), \ldots, h_{\sigma_1}^T(d_n(x, y, R_1)), \\
& h_{\sigma_2}^T(d_1(x, y, R_2)), \ldots, h_{\sigma_2}^T(d_n(x, y, R_2)), \\
& \vdots \\
& h_{\sigma_M}^T(d_1(x, y, R_M)), \ldots, h_{\sigma_M}^T(d_n(x, y, R_M)) \big] \quad (2)
\end{aligned}
$$

where $d_j(x, y, R_i)$ is the location with distance R_i from (x, y) in direction j, $j = 1, 2, \ldots, n$. By using the improved DAISY descriptor to describe the image pair, descriptions for every pixel in these images are obtained.

After tie-point matching between the reference and sensed images was performed (as in the SAR-SIFT algorithm [15]), the projection matrix H can be computed. Any point i in the reference image may be mapped to a point i' in the sensed image with coordinates (x, y) and (x', y'), respectively. A square window with an edge length of $2r + 1$ centered on i' can be constructed, which defines the cost-search window (see Fig. 2).

The matching cost of all pixels in the search window within the sensed image with respect to i was then calculated. Matching cost is given by

$$C(x, y, (dx, dy)) = E(D_R(x + dx, y + dy), D_L(x, y)) \quad (3)$$

where $C(x, y, (dx, dy))$ is the matching cost for the two points; $D_L(x, y)$ and $D_R(x+dx, y+dy)$ are the descriptors for the points on the reference and sensed images, respectively; E is the Euclidean distance between the two descriptors; dx and dy are the difference in coordinates between the two points, i.e., the disparity between the two directions. Calculating the cost of each pixel in the reference image yields the cost cube, with the horizontal dimensions of the reference image and height equivalent to the number of pixels in the cost-search window. The cost cube stores the matching cost of any pixel in the reference image with respect to all pixels in the cost-search window.

B. Cost Aggregation Based on Superpixel Segmentation and the GF Algorithm

As cost calculation is solely concerned with local correlation, the results are very sensitive to noise. The cost correlation results cannot be directly used to calculate the

optimal disparity; therefore, cost aggregation is necessary. This is performed by applying filters to the cost cube. In this work, cost aggregation was performed using superpixel segmentation and the GF algorithm.

Superpixel segmentation is a process where adjacent and perceptually similar pixels are grouped to form local structures that are approximately uniform in size, which are then used to replace the image's grid structure. This algorithm makes use of the image's textural information to facilitate the representation of its characteristics by a relatively small number of super-pixels instead of many pixels, which simplifies subsequent procedures. In this work, the simple linear iterative clustering superpixel algorithm was used to segment the reference image into superpixel blocks [17].

GF is a nonlinear filter used for image smoothing [18], which can be used for cost aggregation on superpixels. The filtered result is represented by the following equation:

$$q_i = I_i \cdot \frac{1}{|w|} \sum_{i \in N} \frac{\sigma_i^2}{\sigma_i^2 + \epsilon} + \frac{1}{|w|} \sum_{i \in N} \left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \epsilon} \right) \mu_i \quad (4)$$

where q_i is the filtered result, I is the guidance image, N is the window in a layer of the cost cube, $|w|$ is the number of pixels in the window, σ_i^2 is the variance in N, μ_i is the mean inside N, and ϵ is the smoothing coefficient. A new cost cube is obtained by applying GF to the superpixel blocks in each layer of the cost cube.

C. WTA-Based Calculation of Disparity

Every pixel in the cost cube has a cost for every disparity (dx, dy) after cost aggregation. Since then, the WTA strategy is used to eliminate excess disparity by searching the minimum cost value.

The disparity calculation produces a pair of disparity maps with the same dimensions as the reference image, wherein the pixel values correspond to disparities in two distinct directions.

D. Disparity Refinement With LRC and Median Filtering

The disparity maps require further refinement, as they may contain errors and noise. To this end, LRC and median filtering are used for refinement.

LRC is based on the uniqueness constraint for disparity, i.e., each pixel can only have one correct disparity. LRC is performed by swapping the image pair after its initial disparity map was computed and then recomputing the disparities. If the sum of calculated disparities is lower than a given threshold, they are assumed to be correct; otherwise, the calculated disparity is invalidated and the disparity value at this pixel will be obtained by interpolation.

Besides erroneous disparities, the disparity map may contain a number of outlier values, i.e., noise. After multiple trials, we determined that smoothing the disparity map via median filtering with a window size of 11×11 was effective for reducing noise. In addition, bilinear interpolation fills the gaps left by the invalidation of erroneous disparities. These procedures yield the final disparity map.

TABLE I
DETAILS OF UAV-SAR IMAGES

Image pair No.	Central angle of incidence	Terrain	Ground object	Date of images
1	61.8° 55.6°	Flat	Residential areas	March 2022
2	62.7° 57.3°	Flat	Farms + vegetation + roads	March 2022
3	56.2° 47.8°	Hilly	Vegetation + roads	March 2022
4	67.1° 58.8°	Flat	Residential areas + roads	June 2021
5	66.7° 58.3°	Flat	Vegetation + roads	June 2021
6	51.7° 36.2°	Hilly	Farms + vegetation	November 2021

TABLE II
ACCURACIES OF THE DIM METHODS (IN PIXELS)

Method	Our method	DAISY	Improved DAISY-SGM	SGM
Image pair 1	**4.2**	5.1	17.8	33.4
Image pair 2	**5.3**	6.6	7.4	16.4
Image pair 3	**3.2**	3.3	7.6	9.4
Image pair 4	**4.4**	5.2	10.6	36.0
Image pair 5	**5.1**	6.1	9.1	30.3
Image pair 6	**3.5**	3.5	7.1	27.5

III. EXPERIMENTS

Dense matching experiments were performed on six pairs of UAV SAR images taken in Dengfeng, Zhengzhou. These images were obtained using a drone-borne Ku-band MiniSAR system. All experimental images had a resolution of 1000 × 1000 and contained various types of terrain and ground objects. The sampling interval was 0.12 m in range and 0.06 m in azimuth. The basic information of the experimental images is provided in Table I.

To test the performance of our method, it was compared with the original DAISY descriptor, an improved DAISY-SGM descriptor, and the original SGM algorithm. The DAISY descriptor was configured with default parameters, where the cost-search radius was ten pixels. The time complexity of the first two methods is $O(N^2 D)$ and the latter two is $O(ND)$, where N is the number of pixels in the reference image and D is the number of pixels searched in the sensed image. DIM was performed using the four aforementioned methods to generate disparity maps for the six pairs of images, as shown in Fig. 3.

All four methods include the entirety of the DIM process, but our method produced smoother-looking disparity maps. Due to the characteristics of SAR images, SAR disparity maps cannot show the approximate contours of the terrain and ground objects, unlike those of optical images. To assess the accuracy of the DIM methods, over 40 pairs of control points were manually selected for each pair of images, and the root mean square error (RMSE) of each method was calculated. The results are shown in Table II.

Table II shows that our method had the best accuracy with errors generally less than six pixels. Compared with the comparative method, the average accuracy of this method is at
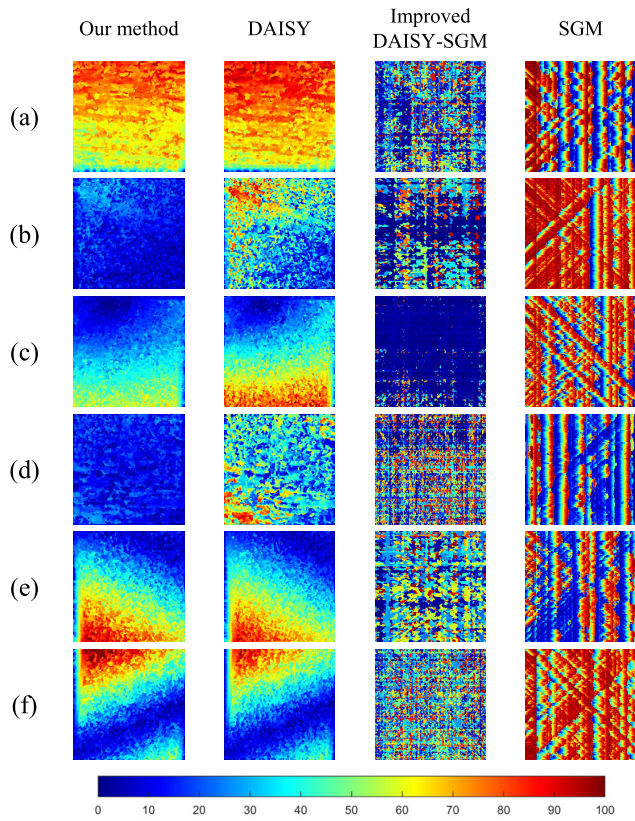
Fig. 3. Disparity maps generated from different DIM methods, where (a)–(f) represent image pairs 1–6. The colorbar at the bottom represents the corresponding relationship between the color and disparity value (in pixels).

least 13% improved. Therefore, the disparity maps produced by our method are good reproductions of the image pairs' disparities. Comparing the results of our method to those of the original DAISY operator confirms that the inclusion of the ROEWA operator improved our method's suitability for UAV SAR images. Furthermore, the results of SGM and the improved DAISY-SGM algorithm show that our descriptor can also be used with the SGM algorithm, and the overall structure is better than the traditional DIM algorithm that requires epipolar lines estimation and correction.

There is no clear ranking between the six image pairs in terms of DIM accuracy. Although image pairs 3 and 6 had somewhat higher DIM accuracies, the DIM accuracies of pairs 1, 2, 4, and 5 were comparable to each other. Although image pairs 3 and 6 were the images of hilly regions, they contained no occluded areas or small central angles of incidence. The ground objects in these images were farms, roads, and trees, which have relatively mild variations in elevation and clearly defined textures. Image pairs 1 and 4 had large central angles of incidence and ground objects with large variations in elevation, such as buildings and roads. Furthermore, some houses had repetitive textures and occluded each other. Image pairs 2 and 5 also had large central angles of incidence, and their ground objects included vegetation and trees. However, there were large differences in elevation between the trees and roads, and image pair 2 also contained a high-voltage tower; these ground-objects caused sudden changes in disparity, which increased matching
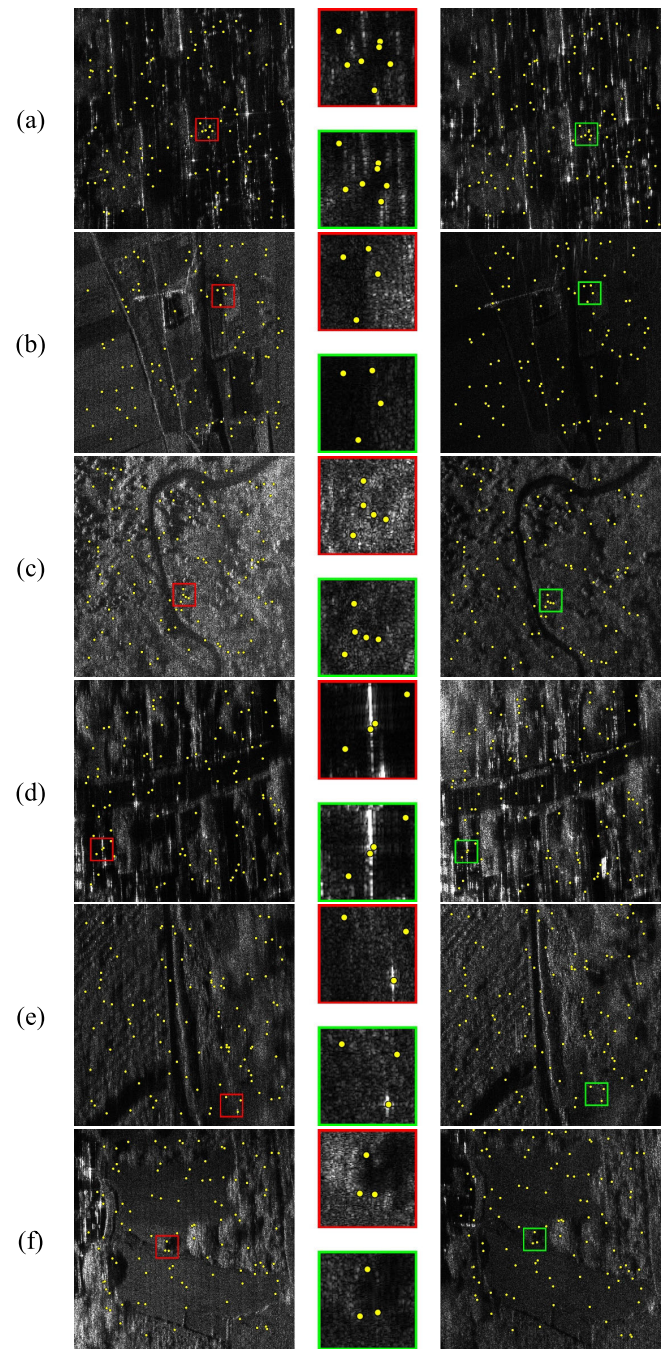

Fig. 4. Random selection of densely matched corresponding points, where (a)–(f) represent image pairs 1–6.

errors. Overall, DIM accuracy depends on the image's central angle of incidence, types of ground objects, and variations in elevation and texture. Accuracy generally decreased with an increased central angle of incidence, variability in elevation, and monotonity in texture.

To intuitively illustrate the matching results of our method, 100 points were randomly selected on each reference image, and their corresponding coordinates in the sensed images were then calculated using disparity maps, as shown in Fig. 4.

The left and right sides in Fig. 4 correspond to the reference and sensed images. The magnified parts of these images are shown in red and green rectangles, respectively. Matching was
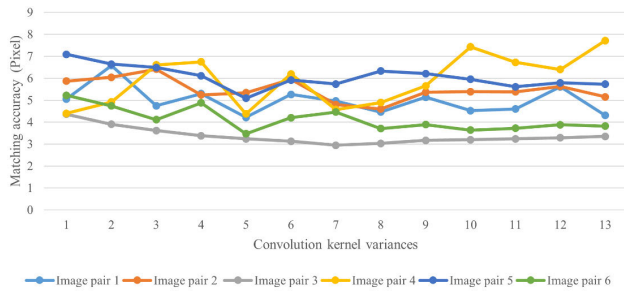
Fig. 5. Matching accuracy with different convolution kernel variances.

better in areas where the features were pronounced. However, each image was taken from different strips, with different angles of incidence, brightness, ground objects, and levels of occlusion, which all contributed to matching errors. Our method succeeded in the DIM of corresponding points in UAV SAR images in most situations. However, its accuracy decreased in areas with monotonous textures and poorly defined features.

The improved DAISY descriptor used in this study contains many configurable parameters, which may alter the performance of the algorithm [14], particularly the variance of the convolution kernel. To assess the effects of convolution kernel variance, different variances were compared for the six image pairs. The results are shown in Fig. 5.

As shown in Fig. 5, matching accuracy varied somewhat with variance, albeit without a clear pattern. Nonetheless, overall accuracy was found to be highest with a variance of five, implying that the descriptor is most stable and generalizable with this value of variance. Therefore, a variance of five was used in all experiments.

IV. CONCLUSION

This study proposes a DIM method that can produce the disparity map of two UAV SAR images without epipolar rectification. In this method, the results of tie-point matching were used to estimate the initial positions of the corresponding points and define the cost-search window, which avoided errors and information loss caused by the estimation and resampling of epipolar lines. The DAISY descriptor was used with the ROEWA operator in the algorithm, which suppressed the effects of speckle noise. The average RMSE of our method can reach 4.3 pixels through experiments on six pairs of UAV SAR images, which contained different ground objects and were taken at different times and locations. This letter provides substantial theoretical support for DIM of UAV SAR images and makes certain contributions to UAV SAR image processing. Future work will focus on further improving the accuracy and efficiency of the algorithm to adapt to applications in large amounts of data.

REFERENCES

[1] F. Xu et al., "Heuristic path planning method for multistatic UAV-borne SAR imaging system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8522–8536, 2021.

[2] L. Pallotta, G. Giunta, and C. Clemente, "SAR image registration in the presence of rotation and translation: A constrained least squares approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1595–1599, Sep. 2021.

[3] S. Méric, F. Fayard, and É. Pottier, "Radargrammetric SAR image processing," in *Geoscience and Remote Sensing*, H. Pei-Gee, Ed. 2009, pp. 421–454. [Online]. Available: https://www.intechopen.com/books/3345

[4] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[5] S. Arunagiri et al., "Stereo matching: Performance study of two global algorithms," in *Radar Sensor Technology XV*, vol. 8021, Jun. 2011, pp. 530–546. [Online]. Available: https://spie.org/Publications/Proceedings/Volume/8021#_=_ and https://www.spiedigitallibrary.org/conference-proceedings-of-spie/8021/80211Z/Stereo-matching-performance-study-of-two-global-algorithms/10.1117/12.887317.short

[6] K. Insfran, K. Ito, and T. Aoki, "Accurate 3D measurement from two SAR images without prior knowledge of scene," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Brussels, Belgium, Jul. 2021, pp. 4814–4817.

[7] H. Ding, Z. Jixian, H. Guoman, and Z. Jianjun, "Multi-image matching of airborne SAR imagery by SANCC," *Acta Geodaetica et Cartographica Sinica*, vol. 44, no. 3, pp. 274–281, 2015.

[8] J. Wang et al., "Radargrammetric DSM generation by semi-global matching and evaluation of penalty functions," *Remote Sens.*, vol. 14, no. 8, p. 1778, Apr. 2022.

[9] Q. Yu et al., "SAR pixelwise registration via multiscale coherent point drift with iterative residual map minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5200919.

[10] P. Capaldo, M. Crespi, F. Fratarcangeli, A. Nascetti, and F. Pieralice, "High-resolution SAR radargrammetry: A first application with COSMO-skymed spotlight imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1100–1104, Nov. 2011.

[11] J. Wang, X. Lv, Z. Huang, and X. Fu, "An epipolar HS-NCC flow algorithm for DSM generation using GaoFen-3 stereo SAR images," *Remote Sens.*, vol. 15, no. 1, p. 129, Dec. 2022.

[12] D. Li and Y. Zhang, "Epipolar geometry comparison of SAR and optical camera," in *Proc. 2nd ISPRS Int. Conf. Comput. Vis. Remote Sens. (CVRS)*, Xiamen, China, vol. 9901, 2015, pp. 197–205.

[13] K. Gutjahr, R. Perko, H. Raggam, and M. Schardt, "The epipolarity constraint in stereo-radargrammetric DEM generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5014–5022, Aug. 2014.

[14] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

[15] L. P. Garcia, L. Pallotta, C. Clemente, G. Giunta, and J. J. Soraghan, "A cross-cross-correlation based method for joint coregistration of rotated multitemporal synthetic aperture radar images," *IET Radar, Sonar Navigat.*, vol. 18, no. 1, pp. 198–209, Jan. 2024, doi: 10.1049/rsn2.12493.

[16] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2014, doi: 10.1109/TGRS.2014.2323552.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[18] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jan. 2013.