On the appropriate feature for general SAR image registration postprint

Authors: Li, Dong, Zhang, Yunhua

Date: 2017-01-04T00:00:00+00:00

Abstract

An investigation to the appropriate feature for SAR image registration is conducted. The commonly-used features such as tie points, Harris corner, the scale invariant feature transform (SIFT), and the speeded up robust feature (SURF) are comprehensively evaluated in terms of several criteria such as the geometrical invariance of feature, the extraction speed, the localization accuracy, the geometrical invariance of descriptor, the matching speed, the robustness to decorrelation, and the flexibility to image speckling. It is shown that SURF outperforms others. It is particularly indicated that SURF has good flexibility to image speckling because the Fast-Hessian detector of SURF has a potential relation with the refined Lee filter. It is recommended to perform SURF on the oversampled image with unaltered sampling step so as to improve the subpixel registration accuracy and speckle immunity. Thus SURF is more appropriate and competent for general SAR image registration. 2012 SPIE.

Full Text

Preamble

On the Appropriate Feature for General SAR Image Registration

Dong Li*, Yunhua Zhang
Key Lab of Microwave Remote Sensing, Center for Space Science and Applied Research, Chinese Academy of Sciences, No.1 Nanertiao, Zhongguancun, Haidian District, Beijing 100190, China

Abstract

This paper investigates the most appropriate feature for synthetic aperture radar (SAR) image registration. Commonly-used features—including tie points, Harris corners, scale-invariant feature transform (SIFT), and speeded-up robust features (SURF)—are comprehensively evaluated according to several criteria:

geometrical invariance of features, extraction speed, localization accuracy, geometrical invariance of descriptors, matching speed, robustness to decorrelation, and flexibility to image speckling. Results demonstrate that SURF outperforms the alternatives. Notably, SURF exhibits excellent flexibility to image speckling because its Fast-Hessian detector has a potential relationship with the refined Lee filter. We recommend applying SURF to oversampled images with an unaltered sampling step to improve subpixel registration accuracy and speckle immunity. Therefore, SURF is more appropriate and competent for general SAR image registration.

Keywords: Feature detector, feature descriptor, image registration, speeded-up robust feature (SURF), subpixel accuracy, synthetic aperture radar (SAR).

1. Introduction

Benefiting from numerous SAR missions, the availability of imagery for any given region has increased dramatically, making joint processing of multiple images for accurate scene sensing and understanding possible. This has become a promising direction for SAR image processing. Since SAR images may be acquired from different imaging geometries and/or by different sensors, geometrical warping always exists between images and must be aligned before further processing. Image registration aims to estimate the warp function between images so that the same pixel position in each image maps to the same target position in a global coordinate system.

Numerous SAR image registration algorithms have been proposed to date. This paper focuses solely on feature-based registration algorithms. Contours [1], [2], regions [3], [4], lines [5], [6], and points [7], [8] are commonly-used features for SAR image registration. The first three features and their combinations (i.e., multi-features [9], [10]) are typically used for registering images from different modalities, such as SAR and optical cameras. For SAR image registration, which involves inherent speckle and distortion, point features are much clearer and easier to extract. Physical points, tie points, corners, and keypoints are commonly used for SAR image registration. Physical points refer to distinctive existing targets in the real world, such as road junctions and crossroads [11], [12], building corners [13], ground control points [14]-[16], isolated point scatterers [17], and even temporarily or partially coherent targets [18], [19]. However, available physical features are usually scarce in general SAR images, hindering the application of more sophisticated warp functions needed to model and correct severe geometrical distortion. Therefore, image-based features that can be automatically and abundantly extracted from images—albeit with less physical significance—have been proposed, such as tie points [20]-[23]. Generally, any point pair that can geometrically tie two images together qualifies as tie points. In SAR image registration, tie points typically refer to features extracted from tie patches or regions. The tie patches are first matched using region-based algorithms such as cross-correlation (CC) [4], after which tie points are constructed by extracting specific point positions from the matched patches, such as geometrical centers or centroids.

Corners represent another type of image-based feature extracted by analyzing local texture. A corner is a point with two dominant and different edge directions in its local neighborhood. Many corner detectors have been devised, with the Harris corner [24] being commonly used for SAR image registration [7], [8], [21]. The Harris measure is the second moment matrix, which describes the local neighboring gradient distribution of a point. The corner response function is the weighted sum of the determinant and squared trace of this matrix. A pixel is selected as a corner if its response exceeds a given threshold.

Keypoints, also called blob features, refer to points in an image that differ in properties like brightness or color compared to their surroundings [25]. Keypoints provide complementary description of image structure in terms of regions that cannot be obtained from corners [25]. Scale-invariant feature transform (SIFT) [26] and speeded-up robust features (SURF) [27] are commonly-used keypoint features for SAR image registration. SIFT was proposed by Lowe to extract scale-invariant features based on Lindeberg' s automatic scale selection theory [28]. Lindeberg found that under various reasonable assumptions, the only possible scale-space kernel is the Gaussian function. He experimented with both the trace of the Hessian matrix (i.e., the Laplacian of Gaussian, LoG) and the determinant of Hessian (DoH) matrix to detect blob-like structures. To efficiently extract keypoints, Lowe proposed simplifying LoG with the difference of Gaussian (DoG), demonstrating that this approximation successfully improves speed with only slight accuracy loss.

SIFT functions as both a feature detector and descriptor. The SIFT descriptor is a 128-dimensional vector of gradient and orientation information. In [29], the authors presented a comparative study of ten different local descriptors and identified SIFT as the most robust algorithm for treating common image deformations, achieving the best performance. SIFT has been widely used for SAR image registration [30]-[35]. Chen et al. [32] systematically assessed SIFT' s application to SAR and demonstrated its usefulness for image registration. Based on multisensor, multitemporal, and different viewpoint SAR images, Schwind et al. [34] further evaluated SIFT for SAR image registration, showing it to be a potentially robust alternative for point feature-based registration, achieving subpixel consistency for most tested datasets. SIFT' s primary bottleneck is speed [27], [32], [34], which hinders its application for general SAR image registration because SAR data volumes are typically large. To accelerate SIFT, Schwind et al. [34] proposed skipping features detected at the first octave of the scale space pyramid (SSP), finding that very few matches were found at the first octave and that these matches had the highest matching false alarm rate (MFAR) among all octaves for their selected SAR dataset. They showed this could significantly reduce the number of detected keypoints and processing time without decreasing the number of correct matches. The first scale octave in SIFT' s SSP refers to the original or doubled image with the highest resolution. Features extracted from this octave are more accurate for image registration because localization precision at higher scale octaves is lower than at lower octaves due to increased smoothing and coarser resolution [35]. Therefore, discarding matches from the first octave may impact final registration accuracy.

SURF, proposed by Bay et al. [27], follows the same scheme as SIFT but employs novel detection, description, and matching approaches that simplify the original algorithms to their essentials. SURF extracts features based on DoH rather than its trace because DoH has slightly better scale selection properties under non-Euclidean affine transformation than LoG [28], thus responding less to elongated and ill-localized structures [27]. Bay et al. proposed a Fast-Hessian detector to approximate DoH using box filters. The SURF descriptor is a 64-dimensional vector composed of Haar wavelet responses from the square region around the keypoint and is invariant to affine changes in illumination. For SAR images, this means the descriptor is invariant to linear alterations in signal intensity, making it robust to intensity undulations from imaging geometry, depolarization, and decorrelation to some extent. This indicates SURF’ s potential for registering images from multisensor, multisource, and multimodal sources. Although both SURF and SIFT descriptors focus on spatial gradient distribution, SURF is less sensitive to noise because it integrates gradient information within subpatches while SIFT depends on individual gradient orientations [27]. SURF has been demonstrated to approximately match or even outperform SIFT in terms of speed, repeatability, distinctiveness, and robustness [27]. Recently, SURF has been applied in remote sensing for multispectral satellite image registration [36], seabed recognition based on sonar image texture analysis [37], and SAR image registration [38], [39].

Although many feature-based registration algorithms for SAR images have been proposed, many appear to be adapted from optical image registration, raising open problems that remain unsolved. This paper investigates the most appropriate feature for SAR image registration, providing a detailed evaluation of commonly-used features—tie points, Harris corners, SIFT, and SURF—according to criteria including geometrical invariance of features, extraction speed, localization accuracy, geometrical invariance of descriptors, matching speed, robustness to decorrelation, and flexibility to image speckling. SURF emerges as the best performer. Notably, we find SURF is flexible to image speckling due to a potential relationship between its Fast-Hessian detector and the refined Lee speckle filter. We further observe that applying SURF to oversampled SAR images while keeping the sampling step unchanged greatly improves subpixel registration accuracy, correct correspondence number, and MFAR. Thus, SURF is more appropriate for SAR image registration.

The remainder of this paper is organized as follows: Section 2 evaluates commonly-used features to identify the optimal feature for general SAR image registration. Section 3 experimentally assesses the identified optimal feature for high-accuracy image registration. Section 4 concludes the paper.

2. Evaluation of Commonly-Used Features for SAR Registration

Instead of developing a novel feature, we identify the most appropriate feature from existing widely-used image-based features—tie points, Harris corners, SIFT, and SURF—through comprehensive evaluation based on several key factors affecting SAR image registration: geometrical invariance of features, extraction speed, localization accuracy, geometrical invariance of descriptors, matching speed, robustness to decorrelation, and flexibility to image speckling.

Geometrical Invariance of Features. Feature invariance indicates the degree of warping under which the same feature can still be stably and distinctively extracted by the detector from transformed images. Cross-correlation is sensitive to image rotation and scaling, making CC-based tie points invariant only to translation. The second moment matrix in the Harris measure is sensitive to image scaling, making extracted Harris corners invariant only to translation and rotation. Since SIFT and SURF were designed to achieve scale invariance, they ensure stable feature extraction under higher-order distortion compared to tie points and Harris corners. Although SIFT and SURF features are not fully affine-invariant like those extracted by Harris-Affine and Hessian-Affine [40], it has been claimed that affine frames in Harris-Affine and Hessian-Affine are more sensitive to noise than those of scale-invariant detectors. Consequently, affine features have lower repeatability than scale-invariant features in practice unless affine distortion exceeds approximately a 40-degree tilt of a planar surface [26], [40]. Large viewpoint changes may cause significant decorrelation and even different image content because scattering is sensitive to imaging geometry, potentially limiting applications. For general SAR applications, scale-invariant features such as SIFT and SURF are sufficient.

Feature Extraction Speed. Extraction speed is heavily affected by each detector’s computational load. Tie points are extracted by exhaustively traversing all potential offsets in two image directions to calculate cross-correlation. For subpixel extraction, images must be further oversampled, leading to quadratic increases in dataset size and resulting in very heavy computational loads. Harris points are extracted by analyzing the determinant and trace of the second moment matrix at each pixel position. Calculating this matrix only requires first-order image derivatives, which can be easily obtained through simple addition operations, enabling very fast performance. SIFT and SURF extract scale-invariant features by first constructing the scale space pyramid (SSP), which consists of several octaves, each containing a constant number of scale levels. In SIFT, a scale level is obtained by smoothing the image with a Gaussian filter. The next scale level is obtained by applying the same filter to the previously filtered layer, generating a series of scale levels. Nearby layers are then subtracted to compute the DoG, which approximates LoG. For the next octave, images are subsampled and the same procedure is repeated. Final keypoints are extracted by selecting points with extreme DoG values using non-maximum suppression in a $3 \times 3 \times 3$ neighborhood in scale space. SIFT detection is slower than

Harris because it extracts features in 3D space rather than 2D space. However, when extracting an equal number of subpixel features, SIFT is still faster than CC-based tie points because the latter requires exhaustive search.

SURF extracts features based on DoH. For a point $\mathbf{x} = (x, y)$ in an image I, the Hessian matrix $\mathcal{H}(\mathbf{x}, \sigma)$ at $\mathbf{x}$ and scale σ is defined as:

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

where $L_{xx}(\mathbf{x}, \sigma)$, $L_{yy}(\mathbf{x}, \sigma)$, and $L_{xy}(\mathbf{x}, \sigma)$ are respectively the convolutions of Gaussian second-order derivatives in the x-, y-, and xy-directions with I at position $\mathbf{x}$. The scale function DoH can then be obtained by:

$$\mathrm{DoH} = L_{xx} L_{yy} - L_{xy}^2$$

When applied in practice, Gaussians must be discretized and cropped. The corresponding discretized and cropped L_{xx}, L_{xy}, and L_{yy} at the lowest scale of 1.2 are shown in the top half of Figure 1 [Figure 1: see original paper]. Inspired by SIFT' s successful simplification of LoG using DoG, Bay et al. proposed a Fast-Hessian detector to approximate DoH with box filters. The box filters D_{xx}, D_{xy}, and D_{yy} corresponding to L_{xx}, L_{xy}, and L_{yy} at scale 1.2 are shown in the bottom half of Figure 1. This approximation has been shown to perform comparably or even better than the original discretized and cropped Gaussians [27]. As Figure 1 shows, the approximation gives equal weight to pixels within certain windows, enabling convolution calculations at very low computational cost through integral images. Another advantage of box filters is that the SSP can be constructed by applying box filters of any size directly to the original image at exactly the same speed, because integral images make computation independent of filter size. Therefore, instead of iteratively reducing image size and using cascade filtering, SURF' s SSP is built by up-scaling filter size, as shown in Figure 2 [Figure 2: see original paper]. Consequently, SURF detection is much faster than SIFT. The scale function DoH can then be approximated as [27]:

$$\mathrm{DoH}_{\mathrm{approx}} = D_{xx} D_{yy} - (0.9 D_{xy})^2$$

where the coefficient 0.9 balances the expression for the Hessian' s determinant. This function is then applied to locate keypoints in scale space using the same non-maximum suppression technique.

Feature Localization Accuracy. Localization accuracy refers to how precisely features can be extracted, which significantly impacts final registration accuracy. Tie points extracted by cross-correlation can achieve subpixel accuracy by oversampling image patches [16], [41] or oversampling the CC map obtained from coarse registration [4]. However, the localization accuracy of tie points depends on the oversampling rate [23]. Higher sampling rates yield higher accuracy but result in larger datasets and heavier computational loads. In practice, the oversampling rate should not be too high because increasing aliasing may destroy offset estimation. Harris corner accuracy is limited to the pixel level, making it suitable only for coarse registration. SIFT and SURF achieve equivalent subpixel localization accuracy because they use the same localization approach. Both extract extrema through non-maximum suppression and further locate them to subpixel and sub-scale accuracy by fitting a 3D quadratic to the scale function in scale space. The scale functions are DoG for SIFT and approximated DoH for SURF. The scale function f at a point $\mathbf{X} = (x, y, \sigma)$ in scale space can be expanded in a Taylor series up to quadratic terms with $\Delta\mathbf{X}$ shifting from a detected extremum $\mathbf{X}_0 = (x_0, y_0, \sigma_0)$ [42]:

$$f(\mathbf{X}) = f(\mathbf{X}_0) + \frac{\partial f}{\partial \mathbf{X}}^T \Delta\mathbf{X} + \frac{1}{2}\Delta\mathbf{X}^T \frac{\partial^2 f}{\partial \mathbf{X}^2}\Delta\mathbf{X}$$

where $\Delta\mathbf{X} = (\mathbf{X} - \mathbf{X}_0)$ is the offset from the extremum. The subpixel localization $\hat{\Delta}\mathbf{X}$ is found by taking the derivative of the equation with respect to $\Delta\mathbf{X}$ and setting it to zero [42], [43]:

$$\hat{\Delta}\mathbf{X} = -\frac{\partial^2 f}{\partial \mathbf{X}^2}^{-1} \frac{\partial f}{\partial \mathbf{X}}$$

The obtained $\hat{\Delta}\mathbf{X}$ is accepted if it is less than 0.5 in all dimensions, and the final extremum location is obtained by adding $\hat{\Delta}\mathbf{X}$ to its sampling position. Theoretically, the localization expression can be infinitely accurate without limitation, meaning SIFT and SURF can achieve the highest possible accuracy. However, although subpixel feature localization is a prerequisite for accurate image registration, it does not guarantee subpixel image registration. Therefore, for high-accuracy SAR image registration, features require further evaluation, which is presented experimentally in Section 3.

Geometrical Invariance of Descriptor. Feature descriptors are vectors that describe neighboring information around features. In image registration, descriptors are used to construct correspondences from extracted features. Descriptor invariance indicates the degree of image warping under which two features can still be correctly matched. Tie points and Harris corners have no assigned descriptors, but from a matching perspective, both use template matching, where the image patch centered at the feature serves as the descriptor. The commonly-used square window template is invariant only to image translation, meaning tie points and Harris corners can only be correctly matched under slight image distortion. SIFT and SURF descriptors offer a good compromise between feature complexity and robustness to common deformations, with affine transformations such as shearing and anisotropic scaling also covered to some degree by the overall descriptor robustness [26], [27]. Thus, robust matching across substantial ranges of affine distortion and 3D viewpoint changes can be achieved.

Feature Matching Speed. Matching speed reflects the efficiency of correspondence construction. Feature matching is typically performed using a merit function that compares a feature point in the master image with one in the slave image by calculating the merit between their descriptors. For feature-based registration, the merit function is usually either cross-correlation to maximize similarity [4] or Euclidean distance to minimize differences [26], [27]. Generally, a correspondence is detected if it optimizes the merit function. For SIFT and SURF, an optimal correspondence is selected only if its merit also exceeds a certain multiple of the second-best merit. Given extracted features, the dominant factor affecting matching speed is merit calculation. For tie points and Harris corners, the merit function typically refers to cross-correlation, which can be computed using complex data (coherent CC) or magnitude data only (incoherent CC) [41]. Coherent CC has been shown to achieve much higher final registration accuracy than incoherent CC [44]. If $\mathbf{D}_1$ and $\mathbf{D}_2$ are image patches centered at an initial match, coherent CC is calculated as:

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(D_1(i)-\mu_1)(D_2(i)-\mu_2)^*}{\sqrt{\sum_{i=1}^{N}|D_1(i)-\mu_1|^2\sum_{i=1}^{N}|D_2(i)-\mu_2|^2}}$$

where N is the patch size and μ_1 and μ_2 denote the means of $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. The superscript $*$ denotes complex conjugation. Zero-mean operation is necessary because it ensures the obtained CC is invariant to affine radiometric warping. Computing this equation requires approximately $10N^2$ operations, including $7N^2$ additions and $3N^2$ multiplications. For SIFT and SURF, the merit function is often Euclidean distance. If $\mathbf{D}_3$ and $\mathbf{D}_4$ are descriptors of an initial match, the merit is calculated by:

$$\mathrm{Dist}(\mathbf{D}_3,\mathbf{D}_4) = \sqrt{\sum_{i=1}^{L}(D_3(i)-D_4(i))^2}$$

where L is the descriptor length. This computation requires about $3L$ operations, including $2L$ additions and L multiplications. For SURF, Bay et al. [27] found that the Laplacian sign can further distinguish features from background and enable fast indexing during matching. Merit is computed only if the considered initial match shares the same Laplacian sign. The two features composing an initial match can be considered statistically independent because they are extracted from different images. Therefore, assuming equal probability distribution of positive and negative Laplacian signs, SURF’ s merit computation requires only $1.5L$ operations. With descriptor lengths of 128 for SIFT and 64 for SURF, merit computation requires 384 operations for SIFT and 96 for

SURF, making SURF four times faster than SIFT in matching. To achieve equivalent computation for tie points and Harris corners, the required patch size N would be about 6 or 3, respectively. Such small patch sizes may result in biased CC estimation and poor feature localization and matching due to insufficient sampling.

Robustness to Decorrelation. SAR image decorrelation sources can be classified into two categories: geometrical warping and radiometric warping. Geometrical warping induces decorrelation because CC is invariant only to translation, making correlation estimates inaccurate and causing decorrelation under higher-order geometrical warping. This relates to geometrical invariance of features, which has been discussed above. Here we consider only radiometric warping-induced decorrelation. Radiometric warping introduces decorrelation because CC is invariant only to affine changes in scattering. In the microwave band, target scattering is sensitive to frequency, bandwidth, and polarization. Conversely, scattering also differs when distributed targets change, such as natural motions of water surfaces, vegetated lands, and snow-covered ground, not to mention obvious changes like developing foliage, moving vehicles, and constructing buildings. All these introduce complex nonlinear radiometric warping that decorrelates and degrades SAR information, aggravating image registration difficulty. Decorrelation severely impacts tie point localization accuracy because tie points are extracted based on correlation estimation. The achievable localization accuracy for tie points is given by the error standard deviation σ_L [17], [44], [45]:

$$\sigma_L = \frac{\sqrt{3}}{\pi} \frac{\sqrt{1-\gamma^2}}{\gamma} \frac{N}{\mathrm{osr}}$$

where γ is the correlation between two tie patches, N is the patch size, and osr is the data oversampling rate. This equation indicates that localization accuracy is directly related to correlation: higher coherence means higher localization accuracy, while higher decorrelation means worse localization accuracy and thus worse registration accuracy.

Nonlinear functions can be approximated with a series of linear functions, and zero-mean CC is invariant to affine radiometric warping. Therefore, a good method to improve robustness to decorrelation is using smaller image patches, but this reduces localization accuracy as shown in the equation above. Consequently, tie points are not robust to decorrelation. Similarly, decorrelation' s influence on CC-based matching of Harris corners is unavoidable. However, Harris, SIFT, and SURF locate features based on geometrical texture—such as first- and second-order image derivatives—rather than correlation. Additionally, feature extraction in Harris and the lower scale octaves of SIFT and SURF involves only small neighborhoods around features, reducing decorrelation influence. For higher scale octaves of SIFT and SURF, images are greatly smoothed, reducing decorrelation to some extent. SIFT and SURF feature matching uses local descriptors related to local gradients and invariant to affine scattering changes. Therefore, SIFT and SURF features are more robust to decorrelation.

Flexibility to Image Speckling. SAR obtains images by actively measuring and coherently processing electromagnetic scattering signals from targets. Coherent interference from scatterings reflected by different diffuse scatterers within each resolution cell causes pixel-to-pixel intensity variation, resulting in speckle. For tie points based on image correlation, the assumption that scattering is locally stationary and ergodic may not hold under speckle, biasing correlation estimation and causing inaccurate feature localization and incorrect matching. For geometrical texture-based detectors such as Harris, SIFT, and SURF, speckle may create false textures and high MFAR. To extract stable features from speckle-contaminated SAR images, a conceivable approach is to suppress speckle with filtering before feature extraction. Schwind et al. [34] proposed using the ISEF filter to reduce speckle influence, but noted that image smoothing by ISEF or any other speckle filter may slightly affect feature localization and final registration quality, making evaluation difficult without ground truth. Therefore, an ideal strategy is to conduct speckle filtering concurrently with feature extraction, requiring the detector to be flexible to image speckling.

The Harris detector extracts features directly based on first-order image derivatives, making it difficult to be immune to image speckling. Consequently, Harris corners may extract many features but yield only a small number of correct matches because most extracted features are speckles. This influence was also observed for SIFT by Schwind et al. [34] when evaluating SIFT' s applicability to SAR. They found very few matches at the first SSP octave despite an extensive number of extractable features, and matches from this octave had the highest MFAR among all octaves. This influence weakens at higher octaves because greater image smoothing reduces speckle to some extent. The first scale octave refers to original or double-sized images with the highest resolution and largest number of extractable feature points. The few matches and highest MFAR at this octave clearly indicate SIFT' s poor flexibility to speckle. However, SURF handles speckle very well due to the potential relationship between its Fast-Hessian detector and the refined Lee speckle filter, as discussed below.

The primary goal of speckle filtering is to reduce speckle without sacrificing image content. The most commonly-applied technique is the boxcar filter, which replaces a pixel with the average of its windowed neighborhood. This filter is easily implemented and works well for homogeneous areas but has a major drawback: spatial resolution degradation from indiscriminately averaging pixels from inhomogeneous media [54], which blurs edges and smears bright point targets. An ideal speckle filter should adaptively smooth speckle, retain edge sharpness and feature boundaries, and preserve subtle but distinguishable details such as thin linear features and point targets [46]. Many filtering techniques have been proposed, among which the refined Lee filter is one of the most commonly-used. The Lee filter [47] uses local statistics such as mean and variance for despeckling. To reduce speckle without degrading the image, neighboring pixels with texture characteristics similar to the center pixel are selected. Lee proposed matching edge direction using eight nonsquare edge-aligned windows, as shown in Figure 3 [Figure 3: see original paper]. During filtering, one nonsquare window is selected based on edge direction to calculate local statistics, after which the minimum mean square algorithm is applied. This filter has been shown to effectively reduce speckle without degrading edges [46], [47].

As mentioned, SURF extracts features based on DoH, which can be simplified using box filters. Using box filters not only accelerates feature extraction but also provides an excellent approach for extracting edge features while reducing speckle. The D_{xx} in Figure 1 shows that speckle can be reduced using a 5×3 window to average pixels, then extracting vertical edge features via second-order partial derivatives in the x-direction with convolution template $[1\ \ -2\ 1]$. This is equivalent to filtering speckle with Lee' s nonsquare windows (a) and (e). Similarly, D_{yy} indicates speckle reduction using a 5×3 window and horizontal edge feature extraction via second-order partial derivatives in the y-direction with convolution template $[1\ \ -2\ 1]^T$, equivalent to filtering with Lee' s nonsquare windows (c) and (g). The D_{xy} indicates speckle reduction using a 3×3 window and extraction of 135° edge features via second-order partial derivatives in the negative xy-direction with convolution template $\begin{bmatrix} 0 & -1 \\ -1 & 4 \end{bmatrix}$, equivalent to filtering with Lee' s nonsquare windows (b) and (f). The $-D_{xy}$ indicates speckle reduction using a 3×3 window and extraction of 45° edge features via second-order partial derivatives in the xy-direction with convolution template $\begin{bmatrix} -1 & 0 \\ 4 & -1 \end{bmatrix}$, equivalent to filtering with Lee' s nonsquare windows (d) and (h). Instead of selecting the optimal edge to estimate local statistics, SURF combines all four speckle-reduced edge features into a new characteristic value through $D_{xx} \times D_{yy} + 0.9D_{xy} \times 0.9(-D_{xy})$, corresponding to the approximated DoH and used for further keypoint localization. The SSP shown in Figure 2 indicates that a series of windows of different sizes can filter speckle and extract features at different scales. Therefore, SURF is very flexible in dealing with speckle.

Based on the above analysis, Table 1 provides a comprehensive evaluation of commonly-used point features for SAR image registration according to several criteria. From these results, we can summarize for general SAR image registration that: (1) SURF performs best across the selected evaluation criteria; (2) SIFT may be applicable when speed requirements are not strict; (3) Harris is suitable only for coarse registration; and (4) Tie points are appropriate for slightly distorted and decorrelated images but require heavy computation.

3. Evaluation of SURF for SAR Image Subpixel Registration

Based on the above evaluation, we conclude that SURF is more appropriate and competent for general SAR image registration. Nevertheless, SURF requires further evaluation with real data, particularly regarding registration accuracy.

SAR applications typically have strict requirements for registration accuracy, especially for interferometric SAR-based elevation or deformation estimation. To ensure acceptable estimation, registration must achieve subpixel accuracy [48]. To evaluate SURF's capability for subpixel image registration, we designed a comparative experiment using contrived SAR image pairs. Figure 4 [Figure 4: see original paper] shows a SAR image of South Phoenix, AZ, acquired by RadarSat-2 on May 4, 2008. We treat this image as the master and generate the slave image through affine warping:

$$\begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix}$$

where $(x, y, 1)^T$ denotes homogeneous image coordinates, subscripts s and m indicate slave and master, respectively, and $\mathbf{A}$ is an affine matrix with parameters a, b, c, d, and translations t_x and t_y. This yields an image pair with controllable registration parameters.

Bay et al. proposed two versions of the Fast-Hessian detector for SURF: FH-9(-1) with an initial filter size of 9×9 on the original image, and FH-15(-2) with an initial filter size of 15×15 on the doubled image with doubled sampling step. The number in parentheses indicates image size: "1" denotes extraction on the original image and "2" denotes extraction on the doubled image. FH-15(-2) has been shown to outperform FH-9(-1) in repeatability [27]. We applied both detectors to extract point correspondences, then used robust extended fast least trimmed squares [49] to retrieve the warp matrix. To compare the two detectors for SAR registration, we considered four criteria: average transfer error (ATE), number of correct matches, MFAR, and warp matrix estimation error (WMEE).

ATE evaluates the consistency of extracted features with retrieved parameters and is defined as:

$$\text{ATE} = \frac{1}{N} \sum_{i=1}^{N} \left\| \begin{bmatrix} \hat{x}_{mi} \\ \hat{y}_{mi} \end{bmatrix} - \begin{bmatrix} x_{mi} \\ y_{mi} \end{bmatrix} \right\|$$

where $\hat{\mathbf{A}}$ denotes the estimated warp matrix from all constructed correspondences, (x_{si}, y_{si}) and (x_{mi}, y_{mi}) denote the i-th correct correspondence in slave and master images, respectively, and N is the number of correct matches selected by:

$$\text{threshold} > \left\| \begin{bmatrix} x_{mi} \\ y_{mi} \end{bmatrix} - \mathbf{A} \begin{bmatrix} x_{si} \\ y_{si} \end{bmatrix} \right\| \Rightarrow \text{correct match}$$

$$\text{threshold} < \left\| \begin{bmatrix} x_{mi} \\ y_{mi} \end{bmatrix} - \mathbf{A} \begin{bmatrix} x_{si} \\ y_{si} \end{bmatrix} \right\| \Rightarrow \text{mismatch}$$

where $\mathbf{A}$ is the true warp matrix. In our experiments, the threshold is set to 5 pixels, meaning a correspondence is treated as a mismatch if its transfer error in any image direction exceeds 5 pixels. MFAR, also called 1-precision [29], is defined as:

$$\text{MFAR} = \frac{\#\text{mismatches}}{\#\text{matches}} = 1 - \frac{\#\text{correct matches}}{\#\text{matches}}$$

where # denotes "the number of." MFAR represents the percentage of mismatches among constructed correspondences and is primarily influenced by geometrical warping, radiometric warping, and speckle. For a given contrived SAR image pair with controlled geometrical and radiometric warping, the combination of correct match number and MFAR can evaluate detector flexibility to image speckling. WMEE, defined as:

$$\text{WMEE} = \frac{\|\hat{\mathbf{A}} - \mathbf{A}\|_F}{\|\mathbf{A}\|_F}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, further evaluates the consistency and accuracy of the estimated warp matrix relative to its true value.

We exemplified four image pairs with different transformations. Table 2 lists the estimated registration parameters, ATE, correct match number, MFAR, and WMEE for both detectors. Results show that FH-15(-2) extracts more correct matches with lower MFAR than FH-9(-1), demonstrating SURF's greater flexibility to image speckling because FH-15(-2) performs feature extraction on the doubled image with more severe speckle. FH-15(-2) also shows smaller ATE, demonstrating greater consistency with retrieved parameters. However, FH-15(-2) does not improve registration accuracy for all four cases as expected; clear inconsistency remains between the estimated warp matrix and its true value. We believe the primary reason is that when FH-15(-2) doubles the image size, the sampling step is also doubled. Although using different integral images, FH-15(-2) feature detection is geometrically equivalent to detection on the original image with an initial filter size of 7.5 × 7.5 (i.e., FH-7.5(-1)). Smaller filter size indicates higher resolution and more extractable features, as shown in Table 2. However, the doubled sampling step causes sampling to still occur at the same equivalent pixel positions rather than at subpixel image positions. For example, if (x_0, y_0) is a sampled pixel position in the original image, its corresponding position in the doubled image is $(2x_0, 2y_0)$. The doubled sampling step means this pixel position is still sampled but not at $(2x_0 \pm 1, 2y_0 \pm 1)$, which correspond to subpixel positions $(x_0 \pm 0.5, y_0 \pm 0.5)$ in the original image. These subpixel features have doubled position accuracy and contribute more to subpixel image registration.

Therefore, we propose performing detection with an initial filter size of 9 × 9 and an unchanged sampling step, but on oversampled images. We denote this detector as FH-9(-Fs), where Fs indicates the sampling rate. We recommend using linear interpolators such as bilinear interpolation for sampling to avoid nonlinear image aliasing. Table 2 further lists registration results based on FH-9(-2) through FH-9(-5) detectors. The ATE, correct match number, MFAR, and WMEE of FH-9(-2) are perfectly improved compared to FH-9(-1) and FH-15(-2), while registration accuracy also improves as oversampling rate increases because more feature correspondences can be extracted with higher localization accuracy. This makes higher-accuracy SAR image registration possible. Image oversampling increases dataset size and computational load, but this remains acceptable. For high-accuracy registration, we recommend oversampling the image three or four times to achieve a compromise among accuracy, robustness, and computational complexity.

4. Conclusion

SAR imaging geometry and mechanisms inherently introduce unavoidable geometrical distortion and speckle into the resulting images, making SAR image registration significantly more complex than that of optical images. Many existing feature-based algorithms appear to be adapted from optical image registration, raising open problems that remain unsolved. This paper investigates the most appropriate feature for SAR image registration, providing a comprehensive evaluation of commonly-used features—tie points, Harris corners, SIFT, and SURF—according to criteria including geometrical invariance of features, extraction speed, localization accuracy, geometrical invariance of descriptors, matching speed, robustness to decorrelation, and flexibility to image speckling. Results demonstrate that SURF outperforms the alternatives. Among these criteria, we particularly address feature flexibility regarding unavoidable SAR speckle, which degrades image information and must be suppressed. Since speckle filtering may alter feature positions and consequently impact subpixel feature localization, a robust feature detector should therefore be resilient to speckle. We find that SURF’ s Fast-Hessian detector has a potential relationship with the refined Lee speckle filter, indicating that SURF can extract image features at different scales even under speckle influence. This observation is validated when oversampling the image with increasing sampling rates: unlike SIFT, the number of correct correspondences extractable by SURF and the matching false alarm rate are nevertheless improved, despite encountering more severe speckle. For applications with strict registration accuracy requirements, we recommend using the original Fast-Hessian detector on oversampled images with an unaltered sampling step for feature extraction. Registration experiments on SAR image pairs demonstrate that the proposed SURF approach significantly improves subpixel registration accuracy and speckle immunity. Therefore, SURF is more appropriate and competent for general SAR image registration.

References

[1] Li, H., Manjunath, B. S. and Mitra, S. K., “A contour-based approach to multisensor image registration,” IEEE Trans. Image Process. 4(3), 320-334 (1995).

[2] Meddeber, L., Berrached, N. E. and Taleb-Ahmed, A., “An automatic registration and mosaicking system based on contour features and wavelet transformation for remote sensing images,” Proc. Int. Conf. Signals, Circuits Syst., 1-7 (2009).

[3] Gabriel, A. K. and Goldstein, R. M., “Crossed orbit interferometry: Theory and experimental results from SIR-B,” Int. J. Remote Sens. 9(5), 857-872 (1988).

[4] Li, F. K. and Goldstein, R. M., “Studies of multibaseline spaceborne interferometric synthetic aperture radars,” IEEE Trans. Geosci. Remote Sens. 28(1), 88-97 (1990).

[5] Sakurai-Amano, T. et al., “Scene feature matching analysis of JERS-1 SAR images,” Proc. IGARSS 2, 1037-1039 (1998).

[6] Zalmanson, G. H., Doytsher, Y. and Schenk, T., “Towards a new paradigm for registration of SAR images using linear features,” Proc. ISPRS Congr. 3, 41-45 (2004).

[7] Marinkovic, P. and Hanssen, R. F., “Advanced InSAR coregistration using point clusters,” Proc. IGARSS 1, 489-492 (2004).

[8] Bentoutou, Y., Taleb, N., Kpalma, K. and Ronsin, J., “An automatic image registration for applications in remote sensing,” IEEE Trans. Geosci. Remote Sens. 43(9), 2127-2137 (2005).

[9] Dare, P. M. and Dowman, I. J., “Automatic registration of SAR and SPOT imagery based on multiple feature extraction and matching,” Proc. IGARSS 7, 2896-2898 (2000).

[10] Dare, P. M. and Dowman, I. J., “An improved model for automatic feature-based registration of SAR and SPOT images,” ISPRS J. Photogramm. Remote Sens. 56(1), 13-28 (2001).

[11] Growe, S. and Tonjes, R., “A knowledge based approach to automatic image registration,” Proc. IEEE Int. Conf. Image Process. 3, 228–231 (1997).

[12] Dell’ Acqua, F., Gamba, P. and Lisini, G., “Coregistration of multiangle fine spatial resolution SAR images,” IEEE Geosci. Remote Sens. Lett. 1(4), 237-241 (2004).

[13] Zhang, X., Xi, R., Li, H., Zhang, X. and Yang, T., “Image registration based on rectangle pattern,” Proc. ICIG, 319-324 (2010).

[14] Gelautz, M., Mitteregger, E. and Leberl, F., “Automated acquisition of ground control using SAR layover and shadows,” Proc. IGARSS 1, 468-470 (1997).

[15] Toutin, T., Hoja, D., Hoeppner, E., Remond, A. and King, C., “GCPs selection from multi-source data over mountainous topography,” Proc. IGARSS 5, 2339-2341 (1998).

[16] Rufino, G., Moccia, A. and Esposito, S., “DEM generation by means of ERS tandem data,” IEEE Trans. Geosci. Remote Sens. 36(6), 1905–1912 (1998).

[17] Serafino, F., “SAR image coregistration based on isolated point scatterers,” IEEE Geosci. Remote Sens. Lett. 3(3), 354-358 (2006).
[18] Zhang, L., Ding, X. and Lu, Z., “Ground settlement monitoring based on temporarily coherent points between two SAR acquisitions,” ISPRS J. Photogramm. Remote Sens. 66(1), 146-152 (2011).
[19] Perissin, D. and Wang, T., “Repeat-pass SAR interferometry with partially coherent targets,” IEEE Trans. Geosci. Remote Sens. 50(1) 271–280 (2012).
[20] Homer, J. and Longstaff, I. D., “Minimising the tie patch window size for SAR image coregistration,” Electron. Lett. 39(1), 122-124 (2003).
[21] Zou, W., Li, Y., Li, Z. and Ding, X., “Improvement of the accuracy of InSAR image coregistration based on tie points: A review,” Sensors 9(2), 1259-1281 (2009).
[22] Giles, A. B., Massom, R. A. and Warner, R. C., “A method for subpixel scale feature-tracking using Radarsat images applied to the Mertz Glacier Tongue, East Antarctica,” Remote Sens. Environ. 113(8), 1691-1699 (2009).
[23] Li, D. and Zhang, Y., “A fast offset estimation approach for InSAR image subpixel registration,” IEEE Geosci. Remote Sens. Lett. 9(2), 267-271 (2012).
[24] Harris, C. and Stephens, M., “A combined corner and edge detector,” Proc. Alvey Vis. Conf., 141-151 (1988).
[25] “Blob Detection,” http://en.wikipedia.org/wiki/Blob_detection.
[26] Lowe, D. G., “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis. 60(2), 91-110 (2004).
[27] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., “Speeded-up robust features (SURF),” Comput. Vis. Image Understand. 110(3), 346–359 (2008).
[28] Lindeberg, T., “Feature detection with automatic scale selection,” Int. J. Comput. Vis. 30(2), 77-116 (1998).
[29] Mikolajczyk, K. and Schmid, C., “A performance evaluation of local descriptors,” IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615-1630 (2005).
[30] Wessel, B., Huber, M. and Roth, A., “Automatic image-to-image registration of near real-time SAR images,” Proc. ENVISAT Symp., SP-636 (2007).
[31] Li, F., Zhang, G. and Yan, J., “Coregistration based on SIFT algorithm for synthetic aperture radar interferometry,” Proc. ISPRS Congr. 1, 123-128 (2008).
[32] Chen, E. -X., Li, Z. -Y., Tian, X. and Li, S. -M., “Application of scale invariant feature transformation to SAR imagery registration,” Acta Automatica Sinica 34(8), 861-868 (2008).
[33] Li, D. and Zhang, Y., “Geometric feature-based image coregistration approach for InSAR,” Proc. APSAR, 1026-1030 (2009).
[34] Schwind, P., Suri, S., Reinartz, P. and Siebert, A., “Applicability of the SIFT operator to geometric SAR image registration,”Int. J. Remote Sens. 31(8), 1959-1980 (2010).
[35] Wang, S., You, H. and Fu, K., “BFSIFT: A novel method to find feature matches for SAR image registration,” IEEE Geosci. Remote Sens. Lett. 9(4), 649-653 (2012).
[36] Teke, M. and Temizel, A., “Multi-spectral satellite image registration using scale-restricted SURF,” Proc. Int. Conf. Pattern Recognit., 2310-2313 (2010).

[37] Nguyen, H. -G., Fablet, R., Ehrhold, A. and Boucher, J. -M., "Keypoint-based analysis of sonar images: Applications to seabed recognition," IEEE Trans. Geosci. Remote Sens. 50(4), 1171-1184 (2012).
[38] Liu, R. and Wang, Y., "SAR image matching based on speeded up robust feature," Proc. WRI Global Congr. Intell. Syst. 4, 518-522 (2009).
[39] Lee, W. -K. and Kim, A. -L., "An efficient automatic geo-registration technique for high resolution spaceborne SAR image fusion," Proc. IGARSS, 3566-3569 (2011).
[40] Mikolajczyk, K. and Schmid, C., "An affine invariant interest point detector," Proc. ECCV 1, 128-142 (2002).
[41] Li, Z. and Bethel, J., "Image coregistration in SAR interferometry," Proc. ISPRS Congr. 1, 433-438 (2008).
[42] Brown, M. and Lowe, D. G., "Invariant features from interest point groups," Proc. BMVC, 253-262 (2002).
[43] Zhang, N., "Computing parallel speeded-up robust features (P-SURF) via POSIX threads," Proc. Int. Conf. Intell. Comput., 287-296 (2009).
[44] Bamler, R. and Eineder, M., "Accuracy of differential shift estimation by correlation and split-bandwidth interferometry for wideband and delta-k SAR systems," IEEE Geosci. Remote Sens. Lett. 2(2), 151-155 (2005).
[45] Bamler, R., "Interferometric stereo radargrammetry: Absolute height determination from ERS-ENVISAT interferograms," Proc. IGARSS 2, 742-745 (2000).
[46] Lee, J. -S. and Pottier, E., [Polarimetric Radar Imaging: From Basics to Applications], CRC Press, Boca Raton, FL, 143-177 (2009).
[47] Lee, J. -S., "Refined filtering of image noise using local statistics," Comput. Graphics Image Process. 15(4), 380-389 (1981).
[48] Hanssen, R. F., [Radar Interferometry: Data Interpretation and Error Analysis], Kluwer Academic Publishers, Dordrecht, The Netherlands, 45-47 (2001).
[49] Li, D. and Zhang. Y., "The appropriate parameter retrieval algorithm for feature-based SAR image registration," SPIE Conf. SAR Image Anal. Model. Tech., 2012.

Note: Figure translations are in progress. See original paper for figures.