High-resolution optical-to-SAR image registration using mutual information and SPSA optimisation

Sourabh Paul1 | Umesh C. Pati2

1 Department of Electronics and Communication Engineering, Madanapalle Institute of Science and Technology (MITS), Madanapalle, Angulu, India
2 Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela, India

Correspondence
Sourabh Paul, Department of Electronics and Communication Engineering, Madanapalle Institute of Science and Technology (MITS), Madanapalle, Angulu, India.
Email: sourabhpaul26@gmail.com

Abstract
The high-resolution optical and synthetic aperture radar (SAR) images are widely used in many remote sensing application areas such as image fusion and change detection where image registration is a fundamental step. The latest high-resolution optical and SAR satellites and airborne systems provide geometrically corrected images which do not contain global deformations. Though the images do not have global differences, still, registration differences exist between these optical and SAR images. These registration differences should be minimised through an automatic registration method before using the images for the aforementioned applications. However, an automatic optical-to-SAR image registration is a challenging task due to the presence of significant nonlinear intensity differences as well as local geometric distortions between the images. In order to solve these problems, an automatic optical-to-SAR image registration method is proposed which can effectively handle the registration differences between the globally corrected high-resolution images. In the proposed method, initially, a coarse registration is performed by using a discrete simultaneous perturbation stochastic approximation (SPSA) optimisation. Then, a smooth continuous SPSA optimisation is utilised for the fine registration of the images. Experiments are performed on six sets of high-resolution optical-SAR image pairs and the results show the effectiveness of the proposed method.

1 | INTRODUCTION

The modern optical and synthetic aperture radar (SAR) satellites and the airborne system provide very high-resolution images which are used in variety of applications such as early warning system, vegetation monitoring, and change detection [1]. The passive optical sensors such as IKONOS and Orbview-3 capture the remote sensing images in the day time whereas the passive SAR sensors such as TerraSAR-X have the capability to acquire images at the night and even through the cloud also. In many remote sensing applications, it is necessary to integrate the complementary information provided by the optical and SAR images. In these applications such as image fusion, change detection, image mosaicing [2, 3], image registration is a vital step. It performs the alignment of two or more images which are captured by same/different sensors, at different times or from different view points [4]. In optical-to-SAR image registration, the SAR image is accurately aligned with the optical image.

The latest satellites are providing the georeferenced remote sensing images which do not have any orientation and scaling differences [1]. The recent technologies are capable of producing the geometrically corrected images which have an offset of dozen or so pixels [5, 6]. These alignment differences between the images should be minimised through image registration before using these images in many remote sensing applications such as image fusion and change detection. Though these differences are in the range of dozen or so pixels, still, optical-to-SAR image registration is a challenging task as the images have significant nonlinear intensity differences as well as high local distortions. Moreover, the influence of multiplicative speckle noise in SAR image affects the performance of registration.

Image registration is generally classified into two categories: feature-based methods and intensity-based methods. The feature-based methods contain the following steps: feature extraction, feature matching, transformation model estimation, and image transformation. The main advantage of feature-based
methods is that these are comparatively faster than the intensity-based methods. Scale-invariant feature transform (SIFT) [7], Daisy [8] and Speeded up robust feature (SURF) [9] are the well known approaches used to register the remote sensing images. However, the performance of the standard SIFT algorithm is not satisfactory for optical-to-SAR image registration as the number of final matches is very less [10]. In order to improve the performance of SIFT algorithm, Fan et al. proposed an improved SIFT (I-SIFT). In [10], the dominant orientation step of the SIFT features is skipped and multiple supported regions are considered to construct the descriptor. Li et al. [11] extracted the contours from the optical and SAR images and then, contour matching is performed to register the images. In [12], phase congruency (PC) model is used along with Gaussian gamma shaped bi-directional window (GGSBW) [13]-based descriptor to register the optical and SAR images. The PC is used to increase the repeatability rate of the extracted features in the input images and GGSBW is utilised to improve the distinctiveness of the descriptor. Ye et al. [14] proposed an optical-to-SAR image registration method using block-based Harris features [15] and an improved local self similarity (LSS) descriptor, known as dense LSS (D-LSS). The LSS [16] utilises the local shape property of the image to construct the descriptor which can handle the nonlinear intensity differences between images. In [17], a distinctive order-based self-similarity operator (DOBSS) is proposed to handle the nonlinear intensity differences between remote sensing images. Paul et al. [18] proposed a modified DOBSS (M-DOBSS) descriptor to further improve the performance of the DOBSS operator. Sui et al. [19] presented a optical-to-SAR image registration algorithm which is very effective for the high-resolution images. In this method, the line segments are extracted from the SAR and optical images and these are matched through Voronoi diagram and multi-resolution analysis. However, its performance is limited to the cases when linear features are available in the input images. Fan et al. [20] presented an optical-to-SAR registration algorithm by using corner features and PC-based structural descriptor. In [20], homogeneously distributed corner features are extracted from the input images and structural descriptors are formed for the extracted features. Then, feature matching is performed to identify the correct matches. Ye et al. [5] proposed a registration algorithm based on Harris features extraction and structural similarity. In [21], a modified SIFT algorithm is proposed to register the remote sensing data sets which include optical and SAR images. In [22], an edge-based control point selection method is proposed to register the remote sensing images. In this method, template matching is used to identify the corresponding control points between the input images. Feng et al. [23], developed a remote sensing registration algorithm by combining feature and area-based methods. In the feature-based stage, the weighted feature points are used along with a block-weighted projective transformation model to obtain the local geometric relation between images. In the area-based stage, an outlier-insensitive model is proposed to further improve the registration result obtained from the feature-based stage. In [24], a registration algorithm is proposed for the remote sensing mountain images based on improved optical flow estimation. Feature selection plays a significant role in feature-based image registration. In [25], Pathak et al. proposed a levy flight-based grey wolf optimisation method to select the prominent feature from image by reducing the redundant and irrelevant features. Though the feature-based methods perform well for the low or middle resolution optical-SAR image pairs, most of the methods are less effective to register the high-resolution optical and SAR images. The main reason is that it is very difficult to extract conjugate features from the high-resolution optical and SAR images by using a feature-based method. Figure 1 shows a pair of high-resolution optical SAR images covering the same scene. It can be observed the images have large intensity differences and the speckle noise effect is significant in the SAR image. Moreover, the images have significant local distortions as the image capturing mechanism of SAR sensors is completely different from optical sensors. It is obvious that the high-resolution optical-to-SAR image registration is a challenging task. So, in this paper, an intensity-based method is proposed to register the high-resolution optical-SAR image pairs having significant intensity differences as well as high local distortions.

Generally, in intensity-based method, a similarity matrix is computed between the input images to obtain the registered image. Cross-correlation (CC) [26] and mutual information (MI)
are the well known similarity matrices of intensity-based methods. The MI-based methods are more effective to handle significant intensity difference between images compared to CC [4]. A proper selection of joint histogramming technique and optimiser play significant role in MI-based image registration [28]. In [29–31], different types interpolation methods are presented to estimate the joint histogram in MI-based registration. Optimisation techniques are generally used to speed up the registration process. The continuous simultaneous perturbation stochastic approximation (C-SPSA) [32] is popularly used optimiser in MI based registration [33–36]. However, in [28], it is reported that C-SPSA fails many times in optical-to-SAR image registration if the optimum point is far away from the starting point. Suri et al. [1] presented an automatic coarse-to-fine registration algorithm to register the globally corrected high-resolution optical and SAR images. In this cited method, coarse registration is performed by an exhaustive search technique and the fine registration is carried out by MI-based method along with C-SPSA optimisation. However, the presented exhaustive search technique needs significant computational time. Moreover, the accuracy of the fine registration technique can be further improved.

Though a number of registration algorithms are proposed in recent years to register the globally corrected optical-SAR image pairs, the high-resolution optical-to-SAR image registration is still a challenging task in remote sensing. In this paper, we have proposed a MI-based coarse-to-fine registration method to register the high-resolution globally corrected optical-SAR images. Initially, a coarse registration is performed by utilising a discrete SPSA (D-SPSA) optimiser [37]. Then, a smooth version of C-SPSA is introduced for MI-based fine registration. The multi-resolution framework is adapted to speed up the proposed MI-based registration. The translation values obtained at the finer resolution level are projected to the finest level to estimate the global translations at the original resolution level of the images. Then, a fine registration method is presented by using a smooth C-SPSA optimiser to estimate the local translational difference between the images. In this case also, multi-resolution decomposition is performed. Finally, the SAR image is transformed by using the estimated transformation to obtain the registered image. The detailed description of the proposed method is presented below.

(i) A novel high-resolution optical-to-SAR image registration algorithm is proposed which is comparatively faster than the popularly used intensity-based registration methods.

(ii) A coarse registration is performed by utilising a D-SPSA optimiser to estimate the global translational differences between the input images. This MI-based coarse registration is performed in the lower resolution of input images which saves a significant computational time for intensity-based registration.

(iii) A smooth version of the C-SPSA optimisation is introduced for MI-based fine registration. The fine registration is performed to handle the local distortions between the optical and SAR images.

The main advantage of the proposed method is that it is robust to significant non-linear intensity differences between the optical and SAR images. Moreover, it is an automatic method which is very effective to register the high-resolution optical-SAR image pairs even if the SAR images contain significant speckle noise.

2 | METHODOLOGY

In this section, the proposed MI-based coarse-to-fine registration method is discussed. Figure 2 shows frame work of the proposed method. At first, a MI-based coarse registration is performed by using a D-SPSA optimisation to estimate the global translational differences between the optical and SAR images. Multi-resolution decomposition is performed to reduce the computational time in registration. The global translation differences are estimated at the finer resolution level rather than the finest level to further improve the computational speed. The translation values obtained at the finer resolution level are projected to the finest level to estimate the global translations at the original resolution level of the images. Then, a fine registration method is presented by using a smooth C-SPSA optimiser to estimate the local translational difference between the images. In this case also, multi-resolution decomposition is performed. Finally, the SAR image is transformed by using the estimated transformation to obtain the registered image. The detailed description of the proposed method is presented below.

2.1 | Coarse registration: Estimation of the global translational differences

It is already mentioned that the latest satellites and the current technologies can provide geometrically corrected images which do not have any orientation and scaling differences. But, still the images have registration differences of dozen or so pixels. So, a MI-based coarse registration is performed by using a D
d
2.1.1 | Mutual information (MI)

Let, $I_1$ and $I_2$ be the two input images. Then, the MI between the images is defined as

$$S(I_1, I_2) = \frac{1}{M} \sum_{i_1} \sum_{i_2} b_{i_1,i_2}(i_1,i_2) \log \left( \frac{M b_{i_1,i_2}(i_1,i_2)}{b_{i_1}(i_1)b_{i_2}(i_2)} \right),$$

(1)

where $b_{i_1}(i_1)$ and $b_{i_2}(i_2)$ are the entropies of the image $I_1$ and $I_2$, respectively. $b_{i_1,i_2}(i_1,i_2)$ denotes the joint entropy. $\frac{b_{i_1,i_2}(i_1,i_2)}{b_{i_1}(i_1)b_{i_2}(i_2)}$ is computed by the joint histogram of the two images. $M$ is the sum of all the entries in the histogram. The MI-based global translational differences estimation can be considered as an optimisation problem which is defined as

$$T^* = \arg\text{opt } S(T),$$

(2)

where $S$ is the MI and $T^*$ denotes optimal transformation parameters which give the maximum MI value. Multi-resolution decomposition is performed to speed up the translational differences estimation process.

Generally, in MI-based methods, the multi-resolution decomposition is performed to speed up the registration process. The input images are decomposed into a certain number of levels and the maximum similarity point is searched from the coarse resolution level to the high-resolution level. At each level, the MI values between the images are computed and maximised. The search is performed on an interval around the optimum transformation found at the previous level and is refined at the next level. So, the accuracy gets increased from the coarse resolution level to the higher resolution level. Maximisation of the MI value at each level is performed by an optimisation technique which can efficiently find the optimal transformation parameters $T^*$. However, if the lower resolution levels do not have sufficient information, the optimiser can diverge from the solution.

2.1.2 | D-SPSA optimisation

The D-SPSA optimiser proposed by Wang et al. [38] calculates the objective function only at the integer positions. The optimiser needs only two measurements of the loss function at each iteration. The D-SPSA consists of the following steps.

(i) An initial guess of the parameters value is given as $\hat{T}_0$

(ii) A vector $\Delta_k = [\Delta_{k1}, \Delta_{k2}, ..., \Delta_{kp}]^T$ is generated where the components of $\Delta_k$ are Bernoulli random variables taking the values +1 or -1 with probability 1/2.

(iii) A function $\pi(\hat{T}_k)$ is calculated as

$$\pi(\hat{T}_k) = \left( \hat{T}_k + \frac{\Delta_k}{2} \right),$$

(3)

where $\Delta_k$ is a $p$-dimensional vector with all the components being unity, and $\left( \hat{T}_k \right) = \left[ [\hat{T}_{k1}], ..., [\hat{T}_{kp}] \right]^T$. Here, $[\_]$ denotes the floor function. Actually, the function $\pi(\hat{T}_k)$ computes the middle value between the floor and ceiling values of $\hat{T}_k$.

(iv) The objective function $L$ is computed at $\pi(\hat{T}_k) + \frac{\Delta_k}{2}$ and $\pi(\hat{T}_k) - \frac{\Delta_k}{2}$ and the gradient vector $\hat{g}_k(\hat{T}_k)$ is estimated as

$$\hat{g}_k(\hat{T}_k) = \left[ L\left( \pi(\hat{T}_k) + \frac{\Delta_k}{2} \right) - L\left( \pi(\hat{T}_k) - \frac{\Delta_k}{2} \right) \right] \Delta_k^{-1},$$

(4)

where $\Delta_k^{-1} = [\Delta_k^{-1}, ..., \Delta_k^{-1}]$.

(v) The vector $\hat{T}_k$ is updated as

$$\hat{T}_{k+1} = \hat{T}_k - a_k \hat{g}_k(\hat{T}_k),$$

(5)

where the gain parameter $a_k$ is given as

$$a_k = \frac{a_d}{(A_d + k + 1)^\alpha},$$

(6)

where $a_d$ and $A_d$ are the positive scalers and $0.5<\alpha<1$.

(vi) Repeat the steps (ii) to (v) for $K$ times. So, the approximated optimum solution is $\hat{T}_K$.

2.1.3 | MI-based coarse registration using D-SPSA optimisation

The C-SPSA is a well known optimiser for MI-based remote sensing image registration. However, it fails frequently to identify the optimum point for high-resolution optical-to-SAR image registration when the optimum point is far away from the initial point. As C-SPSA needs the MI values at the integer as well as the non-integer positions, it is very difficult to get a very smooth MI function starting from the initial point to the optimum point. On the other hand, D-SPSA needs the objective function values only at the integer points. So, it does not require a very smooth objective function to find the optimum point within a certain number of iterations. D-SPSA optimisation method searches the discrete optimum point. As D-SPSA is a discrete optimiser, the accuracy obtained by D-SPSA is in pixel level which is well enough for coarse registration. Therefore, the D-SPSA optimiser is utilised in our proposed method to estimate the global translational differences between the optical and SAR images.

In the proposed D-SPSA-based global translational differences estimation method, the initial guess of the translational parameters is set to 0. As the values of two translational parameters need to be estimated, a two dimensional vector $\Delta_k$ is generated using Bernoulli distribution, that is, $\Delta_k = [\Delta_{k1}, \Delta_{k2}]^T$. The vector $\Delta_k$ in equation (3) contains the two translational parameters. MI value computed between the optical and the SAR image is the objective function $L$ for the D-SPSA-based global translational difference estimation. The translation parameters are updated by using Equation (5).

Three resolution levels are used to estimate the global translational differences. Three resolution levels can easily handle the global translational difference of dozen or so pixels.
between images because the joint histogram of the MI function is estimated by the cubic b-spline. However, the optimum point is searched at the coarsest and the finer resolution levels. The finest resolution level represents the original size of the input images. Searching in the original size of the image requires a significant computational time whereas the accuracy does not improve significantly for discrete optimiser. So, in our proposed coarse registration method, no searching is performed for the finest resolution. The transformational values obtained in the finer resolution level are directly projected to the finest resolution level to estimate the global translational differences. Generally, in multi-resolution frame work, the finest resolution images are down-sampled by a factor of 2 to generate the finest resolution level. The finest resolution level represents the original size of the input images. In order to handle these distortions, a MI-based fine registration is performed by using a smooth C-SPSA optimiser.

### 2.2.1 Smooth C-SPSA

The C-SPSA requires only two function measurements at each iteration. The update rule of the C-SPSA optimisation is given as

$$
\phi_i = \phi_{i-1} + a_ng_{i-1},
$$

where the gradient vector $g_i = [g_1, g_2, \ldots, g_p]^T$ for the $p$-dimensional parameter space is estimated as

$$
(g_i)_j = \frac{I_L(\phi_i + \epsilon_j \Delta_n) - I_L(\phi_i - \epsilon_j \Delta_n)}{2\epsilon_j \Delta_n},
$$

for $i=1,2,\ldots,p$.

Here, $L$ is the objective function. According to Bernoulli’s distribution, every element $(\Delta_n)_j$ of the vector $(\Delta_n)$ takes a value of +1 or -1. $a_n$ and $\epsilon_j$ are the positive sequences of the form

$$
a_n = \frac{a}{(4+n)^{\alpha}},
$$

and

$$
\epsilon_j = \frac{\epsilon}{n^{\alpha}},
$$

where $0 < \alpha < 1$. The smooth version of the C-SPSA introduced by Spall [39] gives better performance in optimisation problems compared to C-SPSA. It also needs only two function measurements at each iteration. The update rule for smooth C-SPSA is given as

$$
\phi_i' = \phi_{i-1}' + a_n G_n,
$$

where $G_n$ is the smooth approximation of $g_n$.

The smooth approximation $G_n$ is defined as

$$
G_n = \rho_n G_{n-1} + (1-\rho_n)g_n(\phi_{i-1}'),
$$

where $0 \leq \rho_n \leq 1$.

$\rho_n$ is defined as

$$
\rho_n = \frac{\epsilon_j}{n^{\alpha}},
$$

More details of smooth C-SPSA can be found in [39]. The smooth version of C-SPSA shares the almost sure convergence property like C-SPSA following some specific conditions [39, 40]. It can be used in optimisation problem to find the global maximum by an iterative process. Motivated by its performance, we have utilised the smooth C-SPSA optimiser for MI-based fine registration.

### 2.2 Fine registration: Estimation of the local translational differences

The geometrically corrected high-resolution optical-SAR images have significant local distortions [1]. Although the proposed coarse registration method globally aligned the optical and SAR images, still local distortions are there between the images.

**Algorithm 1** D-SPSA-based coarse registration (DCR)

**Input:**

$[t_x, t_y] = [0, 0]$ : Initial value of the translation parameters for the first resolution level.

$I_s$: Optical image.

$I_c$: SAR image.

**Output:** $I_{cs}$: Coarse registered SAR image.

1. for $i$ = 1: number of resolution levels
2. if $i$ is not the finest resolution level, do
3. find the initial value $T_i$ of translation parameters for resolution level $i$.
4. for $k$ = 1: number of iterations
5. Generate $\Delta_k = [\Delta_{x1}, \Delta_{x2}]^T$ using Bernoulli’s distribution.
6. Compute function $\pi([\Delta_{x1}, \Delta_{x2}]^T)$ using (3).
7. Compute objective function $L$, that is, mutual information between $I_s$ and $I_c$ using (4).
8. Update the translational parameters $[t_{x1}, t_{x2}]$ using (5).
9. end for
10. $T_{i1} = [2 \ast t_{x1}, 2 \ast t_{x2}]$.
11. else $T_i = T_{i+1}$.
12. end if
13. end for
14. Transformed $I_c$ by using the transformation value $T_i$ and obtain the coarse registered SAR image $I_{cs}$.
2.2.2 MI-based fine registration using smooth C-SPSA optimisation

When an optimisation technique is used for MI-based registration, the input image is transformed many times over the reference image grid [1]. The transformed input image grid may not coincide with the grid of the reference image every time. Therefore, it is very important to compute an exact joint histogram which can give a smooth MI function. There are two types of joint histogramming techniques which can be found in the literature: one and two step joint histogram [1]. The two-step joint histogram techniques produce interpolation artefacts in most of the cases which reduce the registration accuracy. The one-step joint histogram estimated by the higher order interpolation function is more effective to produce a smooth MI function. The smooth C-SPSA optimisation is able to find the optimum transformation value if a higher order one-step joint histogram technique is utilised to compute the value of MI. So, in our proposed method, the cubic b-spline interpolation function is used to estimate the one-step joint histogram.

In order to estimate the local distortions between the optical image and coarse registered SAR images, these are divided into regular windows of size \( N \times N \) [1]. The local translational differences are estimated for the corresponding windows of the two images by using MI-based fine registration along with the smooth C-SPSA optimisation. In this case also, multi-resolution decomposition is performed to speed up the process. In our proposed smooth C-SPSA-based fine registration, the objective function \( I \) is the mutual information computed between the optical an SAR image. As the two translation parameters are estimated to handle the local distortions, \( \phi_t \) represents a two dimensional vector contains the two translation parameters \( [t_x, t_y]^T \).

Figure 3 shows an example of the regular divisions of the input images (left: optical image, right:SAR image). In this figure, images are divided into four sub-windows and the translation differences are computed for the four corresponding windows. The centre of every window of the optical image is considered as tie point and the corresponding tie point in SAR image is obtained by adding the estimated translation values to the centre of the corresponding window in SAR image. Let the co-ordinates of the centre of a window in optical image be \((X_t, Y_t)\) and the estimated translation values between the corresponding windows be \((\delta t_x, \delta t_y)\). Therefore, the co-ordinates of the corresponding tie point in SAR image are \((X_t + \delta t_x, Y_t + \delta t_y)\). The corresponding tie points are obtained for every corresponding window of the optical and SAR images. RANSAC [41] algorithm is used to eliminate the incorrect tie points estimated in fine registration. Local affine transformation model is utilised along with the RANSAC to identify the correct tie points in local regions. These tie points are used to estimate the transformation between the optical and SAR images by using a piece wise linear model [42]. The piece wise linear transformation is very effective to handle the local distortions between the optical and SAR images. Algorithm 2 presents the process of the proposed fine registration method.

3 EXPERIMENTS AND RESULTS

3.1 Selected data sets

Six high-resolution globally corrected Optical-SAR image pairs [43, 44] are selected for the analysis of the experimental results. The detailed information of the data sets is provided below.

3.1.1 Data set 1

The optical image of the data set 1 is captured by the Google Earth (resolution of 2 m) on 31 December 2010 over the area of Barcelona, Spain. The image has the size of 800 × 800 pixels.
Algorithm 2 Smooth-C-SPSA-based fine registration (SCFR)

Input:
\[ [t_x, t_y] = [0, 0]: \text{Initial value of the translation parameters for the first resolution level.} \]
\[ I_o: \text{Optical image.} \]
\[ I_{cs}: \text{Coarse registered SAR image.} \]

Output: \( I_{fs}: \) Fine registered SAR image.

1. Divide \( I_o \) and \( I_{cs} \) into regular windows of size \( N \times N \)
2. for \( m = 1: \) number of windows
3. for \( i = 1: \) number of resolution levels
4. Compute the initial value \( T_i \) of translation parameters for resolution level \( i \) as \( T_i = [2 \ast t_{(i-1)}, 2 \ast t_{(i-1)}] \).
5. for \( k = 1: \) number of iterations
6. Compute objective function \( G_i \), that is, mutual information between \( I_o \) and \( I_{cs} \) and approximate gradient \( g_i \) by using (8).
7. Compute the smooth approximation of \( G_i \) by using (12).
8. Update the translational parameters \([t_{xk}, t_{yk}]\) using (11).
9. end for
10. end for
11. Find the co-ordinates \((X_m, Y_m)\) of the centre of the current window in \( I_o \). Compute the corresponding tie point (TP) in \( I_{cs} \) as \((X_m + t_{xk}, Y_m + t_{yk})\). Store \( \text{opticalTP}(m) = (X_m, Y_m) \) and \( \text{SARTP}(m) = (X_m + t_{xk}, Y_m + t_{yk}) \).
12. end for
13. Use RANSAC to eliminate the outliers from \( \text{opticalTP} \) and \( \text{SARTP} \).
14. Estimate the transformation \( T_f \) between \( I_o \) and \( I_{cs} \) by piece wise linear model using the remaining \( \text{opticalTP} \) and \( \text{SARTP} \).
15. Transform the coarse registered SAR image \( I_{cs} \) by using \( T_f \) and obtain the fine registered SAR image \( I_{fs} \).

FIGURE 4 (a) Optical image and (b) SAR image of data set 1

The SAR image is acquired by the TerraSAR-X sensor on May 15, 2011 with the spatial resolution of 2 m. The image also has the dimension of 800 \( \times \) 800 pixels and it is captured in HH mode. Figure 4 shows the two images.

3.1.2 Data set 2

The optical image of the data set 2 is taken by the Google Earth (resolution of 2 m) on 15 September 2008 over an area of Toronto, Canada. The image has the size of 800 \( \times \) 800 pixels. The SAR image of the data set is captured by the TerraSAR-X sensor on 15 December 2007 (resolution of 2 m). The image has the size of 800 \( \times \) 800 pixels and it is taken in HH mode. Figure 5 shows the two images.

3.1.3 Data set 3

The optical image of the data set 3 is captured by the Orbview-3 sensor (resolution of 2 m) on 23 June 2004 over an area of Toronto, Canada. The image has the size of 800 \( \times \) 800 pixels. The SAR image is acquired by the TerraSAR-X sensor on December 15, 2007 (resolution of 2 m). The image has the size of 800 \( \times \) 800 pixels and it is taken in HH mode. Figure 6 shows the two images.

3.1.4 Data set 4

The optical image of the data set 4 is captured by the Orbview-3 sensor (resolution of 2 m) on 23 June 2004 over an area of Toronto, Canada. The image has the size of 800 \( \times \) 800 pixels. The SAR image is acquired by the TerraSAR-X sensor on 15 December 2007 (resolution of 2 m). The image has the size of 800 \( \times \) 800 pixels and it is taken in HH mode. Figure 7 shows the two images.
3.1.5 | Data set 5

The optical image of the data set 5 is captured by the Google Earth (resolution of 1.5 m) on 9 September 2009 over an area of Pearson International Airport, Canada. The image has the size of 600 × 600 pixels. The SAR image is acquired by the TerraSAR-X sensor on September 18, 2010 (resolution of 1.5 m). The image has the size of 600 × 600 pixels and it is taken in HH mode. Figure 8 shows the two images. The images have radiometric variations.

3.1.6 | Data set 6

The optical image of the data set 6 is captured by the Google Earth (resolution of 2.5 m) on 27 November 2011 over an area of Shanghai, China. The image has the size of 1000 × 1000 pixels. The SAR image is acquired by the TerraSAR-X sensor on 9 January 2013 (resolution of 2 m). The image has the size of 1000 × 1000 pixels and it is taken in HH mode. Figure 9 shows the two images. The images have radiometric variations.

3.2 | Qualitative assessment parameters

The following parameters are used to evaluate the performance of the proposed method.

(a) Mutual information (MI): It is the measurement of the statistical similarity between two images. The definition of MI is already given in (1). Its value increases with the increase in the position accuracy in registration. So, its value should be as high as possible.

(b) Root mean square error (RMSE): It is used to measure the position accuracy between registered images. In order to measure the RMSE value, the input images are divided into 5×5 windows and one corresponding points pair is selected manually from every corresponding window pair. Therefore, 5×5×1 = 25 uniformly distributed corresponding points are obtained between the input images. The transformation value obtained by a registration method is applied on these points and the residual values are calculated. The root mean square of these residual values is the RMSE value.

3.3 | Parameter setting

In the proposed global translational differences estimation method, when the optimum point is far away from the initial starting point, the changes of MI value is very low. As a result, more iterations are required to approximate the solution when the standard D-SPSA optimisation is used. In order to solve the problem, the change of the transformation parameters is considered as 0.5 for the first certain number of iterations (k_f). The value of k_f is set to k/2, where, k is the number of iterations in every resolution level. Using the above approach, the optimum point can be approximated with fewer iterations.

The number of iterations (k) in D-SPSA is decided by the experimental results analysis of the six data sets. A global error ($E_{global}$) value is computed to show the performance of the D-SPSA optimisation. The original global translational differences between the images are computed in the finer resolution level by an exhaustive search method presented in [1]. According to [1], the sensed image is translated over the reference image in vertical and horizontal directions and the MI values are computed in pixel as well as sub-pixel positions. The translation differences corresponding to the highest MI value, are considered as the original global translational differences between the images in that resolution level. The translational differences
estimated by the proposed method are compared with the original global translational differences to verify the performance of the D-SPSA. The global error value of the finer level is computed as

$$E_{glob} = \sqrt{(X_o - X_c)^2 + (Y_o - Y_c)^2},$$  
(14)

where $X_o$ and $Y_o$ are the original translational differences between the images in vertical and horizontal directions, respectively. $X_c$ and $Y_c$ are the computed translational differences by the proposed D-SPSA based translational differences estimation method.

Figure 10 shows the global error values with respect to the number of iterations for the six data sets by using the proposed approach. From Figure 10, it can be observed that the proposed D-SPSA based global translational difference estimation method achieves around 1 pixel global error within 20 iterations. If the number of iterations increases, the global error value oscillates between 0 and 1 due to the stochastic nature of D-SPSA.

3.4 Comparative results analysis

A comparative analysis is performed between the proposed D-SPSA-based coarse registration method and an exhaustive search method presented in [1]. A discrete exhaustive search method can also be used to determine the discrete translational differences between the input images by translating the SAR image over the optical image from $[-5$ to $5]$ pixels both in horizontal and vertical directions with an interval of 1 pixel (using multi-resolution approach). However, the discrete exhaustive search method needs significant computational time in MI-based registration.

### TABLE 1
Comparison of the computational times needed for discrete exhaustive search and proposed D-SPSA-based method

| Method                    | Computational Time(s) |
|---------------------------|------------------------|
| Discrete exhaustive search| 112                    |
| Proposed D-SPSA           | 65                     |

### TABLE 2
Quantitative results of different methods

| Data Set | Method | MI     | RMSE | Time(s) |
|----------|--------|--------|------|---------|
| 1        | LSS    | -      | -    | 42      |
|          | I-SIFT | -      | -    | 48      |
|          | DOBSS  | -      | -    | 59      |
|          | M-DOBSS| -      | -    | 71      |
|          | D-SPSA | 0.1921 | 2.322| 65      |
| 2        | LSS    | -      | -    | 41      |
|          | I-SIFT | -      | -    | 46      |
|          | DOBSS  | -      | -    | 60      |
|          | M-DOBSS| -      | -    | 74      |
|          | D-SPSA | 0.4083 | 2.448| 65      |
| 3        | LSS    | 0.3779 | 1.864| 42      |
|          | I-SIFT | 0.3788 | 1.780| 49      |
|          | DOBSS  | 0.3802 | 1.698| 60      |
|          | M-DOBSS| 0.3810 | 1.651| 76      |
|          | D-SPSA | 0.3628 | 2.387| 65      |
| 4        | LSS    | 0.4911 | 2.124| 14      |
|          | I-SIFT | 0.4923 | 2.003| 16      |
|          | DOBSS  | 0.4945 | 1.932| 29      |
|          | M-DOBSS| 0.4962 | 1.817| 43      |
|          | D-SPSA | 0.4820 | 2.624| 38      |
| 5        | LSS    | 0.3368 | 1.984| 58      |
|          | I-SIFT | 0.3376 | 1.895| 67      |
|          | DOBSS  | 0.3393 | 1.774| 86      |
|          | M-DOBSS| 0.3411 | 1.674| 117     |
|          | D-SPSA | 0.3261 | 2.368| 102     |

3.4 Comparative results analysis

A comparative analysis is performed between the proposed D-SPSA-based coarse registration method and an exhaustive search method presented in [1]. A discrete exhaustive search method can also be used to determine the discrete translational differences between the input images by translating the SAR image over the optical image from $[-5$ to $5]$ pixels both in horizontal and vertical directions with an interval of 1 pixel (using multi-resolution approach). However, the discrete exhaustive search method needs significant computational time in MI-based registration.
Table 1 presents the average computational time required for the MI-based registration using D-SPSA and discrete exhaustive search technique for the first four selected data sets. The experiments are performed on a computer with an Intel core i7 3.40-GHz CPU and 4 GB of physical memory, using MATLAB 2014.

From Table 1, it can be observed that D-SPSA takes nearly (1/2) of the computational time required for the discrete exhaustive search to estimate the global translational differences. The saving of the computational time can be more significant in D-SPSA compared to discrete exhaustive search with the increase of the size of the image.

In order to verify the performance of the proposed D-SPSA-based coarse registration method, it is compared with four feature-based methods LSS [16], I-SIFT [10], DOBSS [17], and M-DOBSS [18]. The block-based Harris features [15] are extracted for these feature-based methods. Table 2 presents the quantitative results of different methods for the six selected data sets. The LSS, I-SIFT, DOBSS, and M-DOBSS algorithm fail to register the first, second, and the fourth data sets as no correct matches are found for these data sets. These algorithms only register the third, fifth, and the sixth data set. On the other hand, the proposed D-SPSA-based method successfully registers all the image pairs. As the LSS, I-SIFT, and DOBSS descriptors are feature-based methods, these methods need comparatively less computational times than proposed D-SPSA-based method. Though M-DOBSS is a feature-based method, still it requires more time than the proposed method because it uses multiple patches to construct its descriptor which requires significant computational time. In our proposed D-SPSA-based method, the RMSE values are comparatively more than the other methods because it only estimates the global translational differences between the input images. However, it is very effective for the coarse registration of images as it successfully registers all the optical-SAR images. The D-SPSA-based registration needs nearly 65 s for every image pair to estimate the global translational differences. Therefore, this method is very useful to perform the coarse registration of high-resolution optical-SAR images.

The feature-based method I-SIFT fails to register the selected images in many cases as it is less effective to handle the non-linear intensity differences of optical-SAR image pairs. Though
LSS and DOBSS, M-DOBSS descriptors can handle the non-linear intensity differences, these methods also fail to find the correct matches in most of the used data sets. The main reasons are as follows: (i) these descriptors are less effective to handle the significant noise effect as these are constructed directly on the original image and (ii) the repeatability of the extracted features is very less for these selected images. An example is shown in Figure 11 to illustrate that the repeatability of the features is very less for these images. In this figure, the extracted features are shown for two corresponding sections of the data set 1. From this figure, it can be observed that most of the extracted features are not located at the corresponding positions. Therefore, the repeatability of the features is very less. As a result, these feature-based methods fail to find correct matches for the selected data sets.

Table 3 presents the performance of the proposed method where coarse and fine registrations are performed by D-SPSA and S-C-SPSA-based methods, respectively. The proposed method is compared with another method where the coarse registration is performed by D-SPSA-based method and fine registration is by C-SPSA-based method [1, 33]. From Table 3, it can be observed that the RMSE values are comparatively lower and MI values are higher in case of our proposed method. It indicates that better position accuracy is obtained in our method. Moreover, it takes same computational time like the D-SPSA+C-SPSA method to give better quantitative results. The proposed method provides better performance compared to the D-SPSA+C-SPSA method because the smooth version of C-SPSA optimiser often gives better accuracy than the standard C-SPSA optimiser [39, 40]. The proposed S-C-SPSA handles the local distortions between the images which can not be handled by the D-SPSA optimisation.

The computational complexity of the proposed DCR algorithm is $O(N_r N_p)$, where $N_r$ is the number of resolution levels considered in coarse registration and $N_p$ is the number of iterations used in D-SPSA optimisation. The proposed SCFR algorithm has the computational complexity of $O(N_w N_r N_p)$, where $N_w$ is the number of windows, $N_r$ is the number of resolution levels considered in fine registration and $N_p$ is the number of iterations used in S-C-SPSA optimisation. Figure 12 shows the visual results obtained by our proposed method for

**FIGURE 12** Checker board mosaiced images for the six data sets. (a) Data set 1, (b) data set 2, (c) data set 3, (d) data set 4, (e) data set 5, and (f) data set 6

| Data Set | Method       | MI   | RMSE  | Time(s) |
|----------|--------------|------|-------|---------|
| 1        | D-SPSA+C-SPSA| 0.2379| 1.291 | 190     |
|          | D-SPSA+S-C-SPSA | 0.2414 | 1.168 | 190     |
| 2        | D-SPSA+C-SPSA | 0.4615| 1.408 | 190     |
|          | D-SPSA+S-C-SPSA | 0.4742 | 1.302 | 190     |
| 3        | D-SPSA+C-SPSA | 0.4203| 1.225 | 190     |
|          | D-SPSA+S-C-SPSA | 0.4322 | 1.112 | 190     |
| 4        | D-SPSA+C-SPSA | 0.4244| 1.315 | 190     |
|          | D-SPSA+S-C-SPSA | 0.4393 | 1.207 | 190     |
| 5        | D-SPSA+C-SPSA | 0.5011| 1.582 | 106     |
|          | D-SPSA+S-C-SPSA | 0.5032 | 1.478 | 106     |
| 6        | D-SPSA+C-SPSA | 0.3503| 1.287 | 295     |
|          | D-SPSA+S-C-SPSA | 0.3514 | 1.174 | 295     |
the six data sets. From these checker board mosaiced images, it can be observed that the image areas and the objects are perfectly aligned.

4 | CONCLUSION

In this paper, we have proposed a novel coarse-to-fine registration method to register the geometrically corrected high-resolution optical and SAR images. Initially, the global translational differences between the images are computed by a MI-based coarse registration using discrete SPSA optimisation. The proposed coarse registration method can achieve pixel level accuracy within a less computational time. Then, a MI-based fine registration is performed by utilising a smooth version of the continuous SPSA optimisation. The fine registration can handle the local distortions between the images. The proposed optical-to-SAR registration method is robust to register the high-resolution geometrically corrected images. From the analysis of the experimental result, it is obvious that the proposed method gives comparatively better performance than the other optical-to-SAR registration methods. However, the proposed method is designed for the globally corrected images. So, it is not effective for the images having large global geometric differences. The high-resolution optical-to-SAR image registration having significant global geometric differences between these images can be considered as future work.

ORCID
Sourabh Paul https://orcid.org/0000-0003-2421-0232

REFERENCES
1. Suri, S., Reinartz, P.: Mutual-information-based registration of terrasar-x and ikonos imagery in urban areas. IEEE Trans. Geosci. Remote Sens. 48(2), 939–949 (2010)
2. Diamat, R., Kasaee, S.: Change detection in optical remote sensing images using difference-based methods and spatial information. IEEE Geosci. Remote Sens. Lett. 7(1), 215–219 (2009)
3. Li, X., et al.: Remote sensing image mosaicking: Achievements and challenges. IEEE Geosci. Remote Sens. Mag. 7(4), 8–22 (2019)
4. Zitova, B., Flusser, J.: Image registration methods: A survey. Image Vision Comput. 21(11), 977–1000 (2003)
5. Ye, Y., et al.: Robust registration of multimodal remote sensing images based on structural similarity. IEEE Trans. Geosci. Remote Sens. 55(5), 2941–2958 (2017)
6. Gonçalves, H., et al.: Chair: Automatic image registration based on correlation and hough transform. Int. J. Remote Sens. 33(24), 7936–7968 (2012)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
8. Tola, E., et al.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. 32(5), 815–830 (2010)
9. Bay, H., et al.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110(3), 346–359 (2008)
10. Fan, B., et al.: Registration of optical and SAR satellite images by exploring the spatial relationship of the improved sift. IEEE Geosci. Remote Sens. Lett. 10(4), 657–661 (2013)
11. Li, H., et al.: A contour-based approach to multisensor image registration. IEEE Trans. Image Process. 4(3), 320–334 (1995)
12. Chen, M., et al.: Gsor: A Gaussian-gamma-shaped bi-windows based descriptor for optical and SAR images matching. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3683–3686. IEEE, Piscataway (2015)
13. Shui, P.L., Cheng, D.: Edge detector of SAR images using Gaussian-gamma-shaped bi-windows. IEEE Geosci. Remote Sens. Lett. 9(5), 846–850 (2012)
14. Ye, Y., et al.: Robust optical-to-SAR image matching based on shape properties. IEEE Geosci. Remote Sens. Lett. 14(9), 646–650 (2017)
15. Ye, Y., Shan, J.: A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. ISPRS J. Photogramm. Remote Sens. 90, 83–95 (2014)
16. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR’07., pp. 1–8. IEEE Conference on. IEEE, Piscataway (2007)
17. Sedaghat, A., Ebadi, H.: Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching. ISPRS J. Photogramm. Remote Sens. 108, 62–71 (2015)
18. Paul, S., Pati, U.C.: Optical-to-sarSAR image registration using modified distinctive order based self-similarity operator. In: 2018 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–5. IEEE, Piscataway (2018)
19. Sui, H., et al.: Automatic optical-to-sarSAR image registration by iterative line extraction and vVoronoi integrated spectral point matching. IEEE Trans. Geosci. Remote Sens. 53(11), 6058–6072 (2015)
20. Fan, J., et al.: SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor. IEEE Trans. Geosci. Remote Sens. 56(9), 5368–5379 (2018)
21. Ma, W., et al.: Remote sensing image registration with modified sift and enhanced feature matching. IEEE Geosci. Remote Sens. Lett. 14(1), 3–7 (2017)
22. Bentoutou, Y., et al.: An automatic image registration for applications in remote sensing. IEEE Trans. Geosci. Remote Sens. 43(9), 2127–2137 (2005)
23. Feng, R., et al.: Robust registration for remote sensing images by combining and localizing feature-and area-based methods. ISPRS J. Photogramm. Remote Sens. 151, 15–26 (2019)
24. Feng, R., et al.: Mountainous remote sensing images registration based on improved optical flow estimation. ISPRS Ann Photogramm. Remote Sens. Spat. Inf. Sci. 4, 479–484 (2019)
25. Pathak, V., et al.: Feature selection for image steganalysis using levy flight-based grey wolf optimization. Multim. Tools Appl. 78(2), 1473–1494 (2019)
26. Shi, W., et al.: A visual circle based image registration algorithm for optical and SAR imagery. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2109–2112. IEEE, Piscataway (2012)
27. Gong, M., et al.: A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information. IEEE Trans. Geosci. Remote Sens. 52(7), 4328–4338 (2014)
28. Suri, S.: Automatic image to image registration for multimodal remote sensing images. Thesis, Technische Universität München (2010)
29. Maes, F., et al.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag. 16(2), 187–198 (1997)
30. Chen, H.M., Varshney, P.K.: Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. IEEE Trans. Med. Imaging 22(9), 1111–1119 (2003)
31. Chen, H.M., et al.: Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. Geosci. Remote Sens. 41(11), 2445–2454 (2003)
32. Spall, J.C.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans. Autom. Control 37(3), 332–341 (1992)
33. Cole-Rhodes, A.A., et al.: Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Trans. Image Process. 12(12), 1495–1511 (2003)
34. Cole-Rhodes, A., et al.: Image registration using a second order stochastic optimization of mutual information. In: Proceedings of 2003 IEEE International Geoscience and Remote Sensing Symposium, IGARSS, pp. 4038–4040. IEEE, Piscataway (2003)
35. Li, Q., et al.: Simultaneous perturbation stochastic approximation algorithm for automated image registration optimization. In: IEEE
36. Li, Q., et al.: Affine registration of multimodality images by optimization of mutual information using a stochastic gradient approximation technique. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007, pp. 397–400. IEEE, Piscataway (2007)

37. Wang, Q., Spall, J.C.: Discrete simultaneous perturbation stochastic approximation on loss function with noisy measurements. In: American Control Conference (ACC), 2011. pp. 4520–4525. IEEE, Piscataway (2011)

38. Wang, Q., Spall, J.C.: Rate of convergence analysis of discrete simultaneous perturbation stochastic approximation algorithm. In: American Control Conference (ACC), 2013. pp. 4771–4776. IEEE, Piscataway (2013)

39. Spall, J.C., Cristion, J.A.: Nonlinear adaptive control using neural networks: Estimation with a smoothed form of simultaneous perturbation gradient approximation, vol. 4, pp. 1–27. Institute of Statistical Science, Academia Sinica 4(1), 1–27, Taipei (1994)

40. Spall, J.C., Cristion, J.A.: Nonlinear adaptive control using neural networks: Estimation with a smoothed form of simultaneous perturbation gradient approximation. In: American Control Conference, vol. 3, pp. 2560–2564. IEEE, Piscataway (1994)

41. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In: Readings in Computer Vision, pp. 726–740. Morgan Kaufmann Publishers, Inc., Los Altos (1987)

42. Goshtasby, A.: Piecewise linear mapping functions for image registration. Pattern Recognit. 19(6), 459–466 (1986)

43. EarthExplorer, http://earthexplorer.usgs.gov

44. Sample imagery, https://www.intelligence-airbusds.com/en/8262-sample-imagery

How to cite this article: Paul S, Pati UC.
High-resolution optical-to-SAR image registration using mutual information and SPSA optimisation. IET Image Process. 2021;15:1319–1331.
https://doi.org/10.1049/ipr2.12107