Neural network algorithm for optical-SAR image registration based on a uniform grid of points

V.V. Volkov [1], E.A. Shvets [1]

[1] Institute for Information Transmission Problems (IITP RAS),
127051, Moscow, Russia, Bolshoy Karetny per. 19, build. 1

Abstract

The paper considers the problem of satellite multimodal image registration, in particular, optical and SAR (Synthetic Aperture Radar). Such algorithms are used in object detection, change detection, navigation. The paper considers algorithms for optical-to-SAR image registration in conditions of rough image pre-alignment. It is known that optical and SAR images have an inaccuracy in registration with georeference (up to 100 pixels with a spatial resolution of 10 m/pixel).

This paper presents a neural network algorithm for optical-to-SAR image registration based on descriptors calculated for a uniform grid of points. First, algorithm find uniform grid of points for both images. Next, the neural network calculates descriptors for each point and finds descriptor distances between all possible pairs of points between optical and SAR images. Using obtained descriptor distances, a matching is made between the points on the optical and SAR images. The found matches between points are used to calculate the geometric transformation between images using the RANSAC algorithm with a limited (to combinations of translation, rotation and uniform scaling) affine transformation model.

The accuracy of the proposed algorithm for optical-to-SAR image registration was investigated with different distortions in rotation and scale.

Keywords: image registration, optical-to-SAR, resnet18, neural network descriptor.

Citation: Volkov VV, Shvets EA. Neural network algorithm for optical-SAR image registration based on a uniform grid of points. Computer Optics 2024; 48(4): 610-618. DOI: 10.18287/2412-6179-CO-1426.

Introduction

Image registration is the alignment of images of the same scene obtained at different times, from different angles and/or using different sensors. The task of comparing images obtained from different sensors is called the multimodal image registration. Optical-to-SAR (Synthetic Aperture Radar) image registration is a particular case of multimodal image registration. Optical-to-SAR image registration is widely used in remote sensing tasks, such as object detection [1], change detection [2], navigation [3].

Optical images are well interpreted by humans and do not contain speckle noise, while SAR images are not affected by the atmosphere and day/night illumination [4 – 5]. It is also easier to distinguish built-up areas on SAR images, due to the high intensity of their pixels caused by multiple re-reflection mechanisms [1, 6].

The difficulty of comparing optical and SAR images is that (i) the images are exposed to different types and strength of noise. For example, speckle noise presents on both optical and SAR images, but more expressed on SAR images. Moreover (ii) the intensity values of optical and SAR images in some areas may not be correlated even without the noise. In addition, (iii) the geometric position of sensors (for example, the angle to the Earth’s surface) in space may differ and as a result some lines or shapes of three-dimensional objects may look different (this is especially noticeable in images with mountain terrain).

In literature, algorithms of optical-to-SAR image registration are applied to images of different size from 256×256 pixels [7] to about 10000×10000 pixels [8]. In this paper, the image registration is considered with images 256×256 pixels (512×512 pixels before downscale), which differ from each other in scale and rotation.

Optical-to-SAR image registration has problems with public availability of datasets and code. Most papers considering optical-to-SAR image registration use datasets which non available for public. Moreover, most datasets are small (about 10 pairs of images or less). Also papers not include code realization for published algorithms. This fact makes it difficult to find algorithm for fair comparison.

Optical and SAR images, due to the parameters describing the position of the sensor, can be roughly aligned based on the georeference. However, due to the inaccuracy of these parameters, the alignment accuracy is tens or even hundreds of pixels [9 – 10]. In the work [8], authors made an estimation of inaccuracy of image registration only with georeference for images obtained from Sentinel-1 (SAR) and Sentinel-2 (optical) satellites. Images from these satellites are used in this work. The reason for the significant geometry distortions between these images is that terrain correction process is not conducted for SAR images from the Sentinel-1 satellite, as well as the side-looking geometry for SAR sensor, i.e. the angle of Sentinel-1 sensor is not a perpendicular to the Earth surface contrary to Sentinel-2 sensor. This result in significant geometric distortions relative to optical Sentinel-2 images, which can reach about 100 pixels (at 10 m/pixel). Additionally, it is worth to note the paper in which the authors indicate that georeference leaves large translation distortions differences between images when small distortions in rotation and scale [9].

Optical-to-SAR image registration methods are divided into two categories: area-based methods (or intensity-based methods) [11 – 13] and feature-based methods [14 – 21]. The area-based methods are based on the calculation of two corresponding (i.e. «similar» by some metric) subimages. The examples of such methods are normalized cross-correlation [11] and mutual information function [12]. Such models may have high (subpixel) accuracy, however, they have great computational complexity and often require preliminary approximate alignment of images [13]. Also, these models may be ineffective for pairs of images with large geometric differences [12].

Feature-based methods are based on (i) finding some “special” points or image elements (called features) that are easily distinguishable both in the optical and SAR image; (ii) subsequent calculating of a geometric transformation that compares the found features. For example, such methods can use following features: points [14 – 16], corners [17 – 19], lines [20] and objects (e.g. roads) [21]. Feature-based methods are computationally simpler and often more efficient [13] than area-based methods in cases where there are significant geometric distortions between images. However, feature-based methods may be worse than area-based methods in maximum image registration accuracy [13].

Consider feature-based methods based on point features. Such algorithms mainly consists of following steps: (i) detection keypoints; (ii) calculation descriptors for each keypoint; (iii) matching keypoints by comparing descriptors; (iv) image registration using matched keypoints (for example with RANSAC algorithm). For example, in [22] authors used SIFT-like detector and modificated SIFT descriptor with using phase consistency. In [23] used SURF detector and descriptor.

Optical-to-SAR image registration algorithms also could use neural networks. As example, algorithm based on sequence of two neural networks: the deep translation network which transform images into a single type (optical or SAR) and U-Net++ neural network based on the U-Net which used the result of first network for change detection [24]. This algorithm processes full images and ignoring mentioned steps. Another example, Siamese neural network with dot product similarity metric [25]. Algorithm considers the task of patches registration with fixed size about 200×200 pixels. In [26] proposed RotNET neural network which predicts rotation between optical-SAR images based on a histogram of gradients. Image size 256×256 pixels. It should be noted that unlike [22 – 23] algorithms in works [25 – 26] algorithms compare full patches. In other words it solve steps (ii) – (iii). For registration images with bigger size than supported patch size it still needed to detect patches which could be compared. And accuracy of full image registration algorithm will be depended on the accuracy of keypoint or patch detection – keypont or patch must be on both images to be matched.

In this paper, we propose a new neural network algorithm for optical-to-SAR image registration in conditions of rough pre-alignment based on the calculation of neural network descriptors, but without the use of keypoint detectors. The images submitted to the algorithm are assumed to be roughly aligned by georeference with an accuracy of ±100 pixels. In this study, image registration is performed using a limited affine model (limited to uniform scale, translation, rotation). We suggest an approach which similar to keypoint-matching approach [22 – 23] with neural network descriptors. The difference is that we avoid using keypoint detectors and assume that keypoints correspond to uniform grid of points with step of 8 pixels. In general case it is impossible to find ideal match between two points (images are not aligned), but instead of it almost each point could be matched with small pixel inaccuracy in point matches. Proposed algorithm based on siamese neural network with cossine similarity metric for descriptor matching. Geometric transformation between images calculated by RANSAC algorithm with a limited affine model using matched points.

The accuracy of the image registration algorithm is investigated for different distortions in scale and rotation. The algorithm was tested on a dataset of 58 optical-SAR images converted from a dataset containing images of the Sentinel-1A and Sentinel-2A satellites (one of two satellites of the Sentinel-1 and Sentinel-2 family, respectively) and published in [27]. The characteristics of the original and converted datasets are described in section 1. The accuracy was compared with another keypoint-matching algorithm based on SIFT detector and SIFT descriptor.

1. Dataset description

The dataset used in this work is transformed from the source dataset consisting of 100 aligned pairs of optical-SAR images of 1024×1024 pixels and published in [27]. The images used in the source dataset are collected from open sources, and licenses to them allow them to be distributed and modified [28]. The Copernicus Open Access Hub site was used as a source of optical and SAR data [29].

1.1. Parameters of source dataset

This subsection describes the parameters of the source dataset before modification to the form used by the proposed algorithm. Images from the Sentinel-1A satellite were used as SAR images with the following parameters: product type Level-1 Ground Range Detected (GRD), Interferometric Wave (IW) swath mode sensor mode, VH polarization. Three-channel RGB satellite images from the Sentinel-2A satellite (product type: S2MSI1C) were used as optical images. Each pair of optical-SAR images was aligned relative to each other using georeference. To further increase the alignment accuracy, a manual image registration using a projective transformation was performed. The resulting images are aligned mainly with sub-pixel accuracy, but sometimes there are small areas where the accuracy is lower (error up to two pixels). The dataset mainly contains images of cities and fields. Images of both types have spatial resolution of 10 meters/pixel. For example, roads as one of the noticeable landmark could reach width of 3-9 pixels for big roads and about one pixel for roads between building blocks. The ready-to-use dataset and its accompanying metadata are available for download and published in [27]. The data can also be downloaded manually from the resources listed above. The source dataset was randomly divided into three parts: the training (58 pairs), validation (13 pairs) and test (29 pairs) parts. Uniform distibution in subclasses (city, town, field, forest, mountain, wasteland, water region, coast, illumination) was taken into account.

1.2. Used dataset

This paper considers the problem of optical-to-SAR image registration under conditions of rough pre-alignment with a limited affine model transformation (limited to translation, rotation and uniform scale). The accuracy of image registration was investigated for different distortions in scale and rotation. To do this, datasets with different scale and rotation limitations were generated using images from the source dataset (subsection 1.1).

Consider process of generating dataset with following limitations: maximum relative scale of the SAR image to the optical one was determined as scaling multiplier s_{max} when maximum relative rotation was determined as r_{max}. In this work we investigated 12 combinations of limitations: 3 values for scale: $s_{max}=0, 0.1, 0.2$ and 4 values for rotation: $r_{max}=0°, 10°, 20°, 30°$.

Firstly, all images were downscaled to the size of 512×512 pixels. Next, the SAR images were scaled again with a random for each image scaling multiplier value s ($s\in[1-s_{max}, 1-s_{max}]$) with step 0.05. Further, the unite rotation r_u ($r_u\in[-90°, 90°]$) and relative rotation r ($r_u\in[-r_{max}, r_{max}]$) are randomly determined for each pair of images with an accuracy of up to a degree. The optical images were rotated by the angle r_u, while the SAR images were rotated by the angle r_u+r (used single rotation operation to reduce the loss). At the end, a 256×256 pixel fragment was cut out from the center of the image. The reason to prevent appearing «empty» areas in the image formed after transformations (for example, after rotating, there will be «empty» areas in the corners, since the image of the corresponding area was outside the image).

As an augmentation for the training and validation parts of the dataset, 4 sets of parameters s, r_u and r determined randomly by a uniform distribution were applied to each pair of images. For the test part, 2 sets were used.

Thus, the size of the dataset used in this paper reaches 228 pairs for the training part (4 pairs were an uniform sea surface and therefore were discarded), 52 pairs for validation and 58 pairs for test part. An example of images from dataset with different rotations shown in Fig. 1 and with different scale multiplier in Fig. 2.

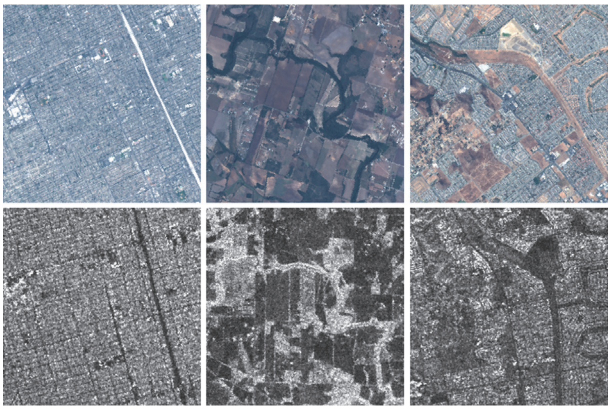

Fig. 1. Examples of different rotations between optical (top) and SAR (bottom) images from left to right: $-10°$, $18°$, $-29°$

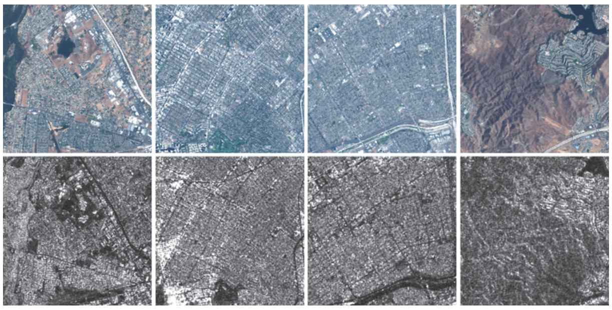

Fig. 2. Examples of different scale multiplier between optical (top) and SAR (bottom) images from left to right: 0.8, 0.9, 1.1, 1.2

2. Proposed image registration algorithm

This paper proposes the neural network algorithm for optical-to-SAR image registration based on neural network descriptors and not using keypoint detectors. The algorithm is designed to match roughly pre-aligned images. When developing the algorithm, pre-aligned images with an accuracy of ±100 pixels, differing in scale and rotation, were considered.

2.1. Proposed neural network architecture

The first stage of the matching algorithm was a neural network that calculated a uniform grid of descriptors for optical and SAR images. Conventionally, this could be represented as a calculation of descriptors for a uniform grid of points. In this work, the step between adjacent points in the grid was 8 pixels. Thus, for an image with size of 256×256 pixels network calculated 32×32 descriptors ($256/8=32$) (total 1024 descriptors). Next, the neural network calculated the descriptor distances between all possible pairs of descriptors. For correctly matched descriptors (i.e. the matched points refer to the same geographical location) the distance tends to be 0, otherwise it tends to be 1.

The algorithm for obtaining input optical and SAR images is described in subsection 1.2. Let's consider the algorithm for obtaining ground truth labels of point matches for neural network training. For an optical image, we calculated the coordinates of a uniform grid of points taken with a certain step between them (in this work, the step was taken 8 pixels). For the SAR image, we calculated the coordinates of the same grid of points, but taking into account the transformation between a pair of images used in generating the dataset. In other words we represented the coordinates of points on SAR image in the coordinate system of the optical image. Next, we calculate the L2 distances between pixel coordinates of all possible pairs of points (between points on optical image and transformed points on SAR image). For each point on the optical image, the matched pair will be with the transformed point on the SAR image that will be the closest, provided that the euclidean distance between them does not exceed the step between the points in the grid. The ground truth labels of the matched pairs are represented as a 1024×1024 matrix, where the row corresponds to the index of the point on the optical image, and the column corresponds to the index of the point on the SAR image. Matched pairs are marked as 0, not matched as 1.

At the input, the neural network receives two images with the size of 256×256 pixels: optical RGB and SAR. At the output, the neural network produces descriptor distances between all possible pairs of descriptors in the form of a 1024×1024 matrix (the number of descriptors for one image is 1024), where the row index corresponds to the index of the descriptor (and the corresponding grid point) on the optical image, and the column index corresponds to the index of the descriptor on the SAR image.

The neural network architecture was a siamese neural network with different weights for each branch which was based on the first two blocks (out of four) of the resnet18 neural network (Fig. 3). The first branch processed an optical three–channel image, the second – a grayscale SAR image. The result of both branches was 32×32 descriptors with length of 128 numbers. Next, we calculated the descriptor distances between all possible pairs of descriptors. The descriptor distances were calculated using the following formula (presented for two descriptors).

$$similarity = 1 - similarity_{cos} = 1 - \frac{x_1 x_2}{max(\| x_1 \| \| x_2 \|, \varepsilon)}, \quad (1)$$

where $similarity_{cos}$ – cosine similarity, x_1, x_2 – two vector-descriptors, ε – small value to avoid division by zero ($\varepsilon = 10^{-8}$).

Thus, the neural network calculated descriptors so that for the corresponding pairs of descriptors the *similarity* metric was close to 0, and for non corresponding ones it was close to 1. Before discussing the loss function let's consider the search window for candidate points for matching.

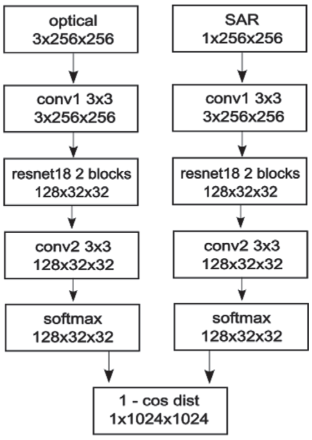

Fig 3. Proposed neural network architecture for optical-to-SAR image registration

The paper examines the image registration with pre-alignment. Accordingly, we used a restriction on which points could be considered as matched. This allowed to discard some of the obviously incorrect point matches, which simplified the formation of correct matches. Fig. 4 shows a pair of not aligned images. Let's say we are looking for a matched point on the SAR image for a yellow dot on the optical image. The yellow dot on the optical and SAR image have the same pixel coordinates, but not corresponded to each other (because the images are not perfectly aligned). The geographically corresponding point on the SAR image is the green dot. Since it is known that the images are roughly aligned, there is no need to look for a paired point throughout the full image – it is enough to consider some neighborhood around the intresting point. This neighborhood is called the search window (green square on Fig. 4) and is formed around a point in the SAR image with the same pixel coordinates as the interesting point (yellow dot). Candidates for matching are searched only inside the search window. In Fig. 4 the images have a size of 256×256 pixels with a window radius of 50 pixels (half of side of green square).

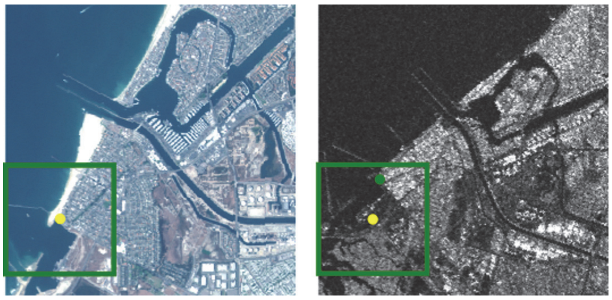

Fig. 4. Example of a search window for searching candidate points for matching. The yellow dot on the optical image and the green dot on the SAR correspond to the same geographical location. The green square corresponds to the search window centered on the yellow dot

The search window was used in the loss function as a mask M of valid matches. The mask size corresponded to the size of the matches matrix calculated by the neural network (1024×1024 in our case). Matching points with indices i and j were allowed if $M(i,j)=1$ and were not allowed if $M(i,j)=0$. Valid matches were those pairs of points whose coordinates on any image axis didn't differ by more than the radius of the search window.

The loss function was calculated using the following formula.

$$l_{all}(x,y) = \\ = \begin{cases} w * l_{good}(x(i,j), \\ y(i,j)) + l_{bad}(x(i,j), y(i,j)), M(i,j)=1 \\ 0, M(i,j)=0 \end{cases} \tag{2}$$

$$l_{good}(x,y) = (1-y) * x^2,$$

$$l_{bad}(x,y) = y * (y - min(x+t,1))^2,$$

where w – weight (parameter of the algorithm), x – network response, y – ground truth labels {0, 1}, t – margin (parameter of the algorithm), M – mask of valid matches.

When descriptors are not matching, we want the cosine similarity ($similarity_{cos}$) to be 0, which is achieved only when the descriptor vectors are perpendicular or at least one of them is a null descriptor. This is difficult to achieve simultaneously for all unmatched descriptors. Therefore, the margin parameter t was used in the loss function for the case of unmatched descriptors. Fig. 5 shows the probability-normalized histograms of matching pairs of points with different margin parameters t. The histogram is based on a test sample matches of descriptor-points. The histogram shows the descriptor distances calculated by the neural network: the blue histogram corresponds to matched pairs of points, the red one corresponds to unmatched. Candidates for comparison will be those pairs of points whose descriptor distance is less than a certain threshold (parameter). Therefore, from the two examples presented in Fig. 5, the case with $t=0.35$ is better than the case with $t=0.2$, since the descriptor distance for correct matches has become much closer to zero, while for unmatched pairs the change is not so significant with descriptor distances less than 0.4. This allows, for example, in the case of $t=0.35$ with a threshold of a descriptor distance of 0.4, to obtain a sufficient number of correct matches with a small proportion of incorrect matches, while in the case of $t=0.2$, to obtain the same number of correct matches, it is necessary to increase the threshold of the descriptor distance, which increases the proportion of incorrect matches.

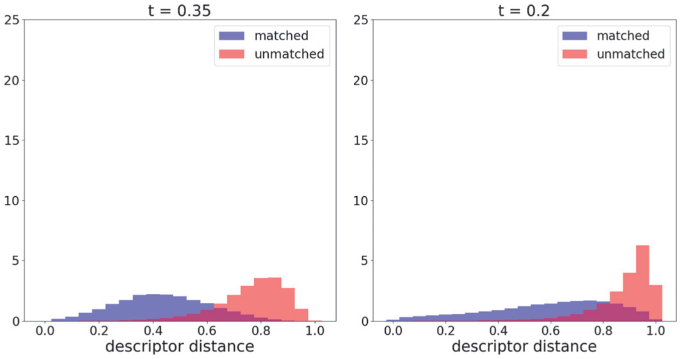

Fig. 5. Examples of histograms of matching pairs of points with different margin parameter t. On the left t = 0.35, on the right t = 0.2

2.2. Image registration algorithm

This subsection describes the full proposed image registration algorithm. The algorithm parameters are as follows: parameters of the neural network loss function (weight w for cases of correct matches, margin t for cases of incorrect matches, the size of the search window), parameters for forming matches between points (the size of the search window that can be different from the search window in the neural network; the threshold of the descriptor distance), parameter of the RANSAC algorithm (threshold for determining inliers). The complete algorithm for optical-to-SAR image registration looks like this:

1. Formation of a uniform grid of points for optical and SAR images (the step between the points is 8 pixels).
2. Obtaining descriptor distances between various pairs of points using the neural network described in subsection 2.1.
3. Forming matches between points by descriptor distances.
4. Calculation of the geometric transformation between images using the RANSAC algorithm based on the found matches between points.

The first two steps were described earlier. Consider the third step of the algorithm. It is based on the nearest neighbor method with some additions. First, the search window was used – we will filter out those pairs of points, which get outside of the search window. The radius of the candidate search window was considered in the range from 50 to 100 pixels. Then, among the remaining points, for each point on the optical image, a point on the SAR image with the minimum descriptor distance was searched. If the converse was also true (for a point on the SAR image, the same point was the closest among all points on the optical image), then a pair of points was considered as candidate for matching. And finally, from the remaining pairs of points were chosen pairs with descriptor distance less than the threshold of the descriptor distance.

After finding the matches between the points, the RANSAC algorithm with a limited affine transformation model (limited to translation, rotation and uniform scale) was used to find a geometric transformation between the images. The RANSAC algorithm randomly iterates through the minimal sets (3 pairs which enough for affine transformation) of matched pairs of points from which it calculates the geometric transformation. The final transformation will be the one that corresponds to the largest number of inliers –- pairs of points that will be determined on each iteration of RANSAC algorithm by applying geometric transformation to all points of one image. If after point transformation its location will be closer to its matched point than inlier threshold then pair considered as inlier. Closer means that the Euclidean distance between points is less than the threshold of inliers (in this work it is taken to be 10 pixels).

To estimate the accuracy of the matching algorithm, the number of correctly matched pairs of images was calculated. The images were considered matched if for each corner of the SAR image $p_i^{SAR}(x_i, y_i)$ following condition is met:

$$\| p_i^{gt} - Hp_i^{SAR} \| \le T,$$

where p_i^{gt} – coordinates of ground truth position of corner of SAR image after transformation, p_i^{SAR} – coordinates of corner of SAR image before transformation, i – corner index {1,2,3,4}, H – matrix of calculated geometric transformation, T – distance threshold, in this work equal to 10 pixels.

The example of successfully matched images is shown in Fig. 6. The lines connect the points which were matched by the algorithm and used for geometric transformation calculation.

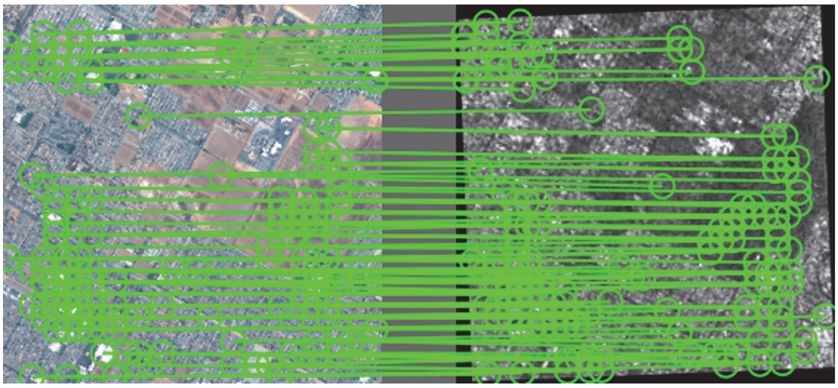

Fig. 6. The example of successful optical-to-SAR image registration

3. Baseline algorithm

As was mentioned before most of optical-to-SAR algorithms published in literature used no public available datasets and most of them consisted of few images (about 10 pairs or less). Also we couldn't find code realization for published algorithm so we chose «classic» algorithm based on SIFT detector and descriptor as baseline. The algorithm is following:

1. Find keypoints using SIFT detector for each image (used OpenCV function).
2. Calculate SIFT descriptors for each keypoint (used OpenCV function).
3. Calculate descriptor distances between all possible pairs of descriptors using L2 metric.
4. Forming matches between points by descriptor distances using nearest neighbor algorithm.
5. Calculation of the geometric transformation between images using the RANSAC algorithm based on the found matches between points.

The difference of this SIFT-based algorithm from our proposed algorithm was in using keypoints which led to different matches. For detecting SIFT keypoints we used OpenCV function (cv2.xfeatures2d.SIFT_create) with binning and NMS (non-max-suppression) filtration. In [30] showed that binning and NMS filtration improve accuracy of optical-to-SAR image registration. Binning based on dividing an image on disjoint cells (we used cells with size 128×128 pixels). The detector was applied to each cell independently. Next, detected keypoints from each cell were united. This allowed to get more evenly distributed keypoints on the image which improve the quality of image registration by RANSAC algorithm [31]. Maximum number of keypoints detected for each cell was no more than 200 keypoints. NMS algorithm filters keypoints based on «score» value calculated by detector. So if the euclidean distance between two keypoints of the same image is less than threshold (we used 5 pixels) then NMS algorithm deletes keypoint with lower «score» value. It prevents from getting clusters of keypoints.

On step 2 we also used OpenCV function for calculating descriptors. Next, we calculated descriptor distances between all possible pairs of descriptors using L2 metric (step 3). Steps 4 – 5 were the same as in proposed algorithm described in section 2.

4. Results

This section describes the results of the proposed algorithm for optical-to-SAR image registration. In this work we investigated the accuracy of matching images for different distortions in rotation and scale (12 cases). Rotation and scale were random for each pair of images, but limited for a single distortion case (subsection 1.2). All kinds of cases were considered with a rotation up to $\{0°, 10°, 20°, 30°\}$ in any direction and three cases for the scale: without changing the scale, with a scale multiplier in range [0.9, 1.1] and with a scale multiplier in range [0.8, 1.2]. The results shown in Tab. 1, 2. The first column of the Tab. 1, 2 shows the distortion boundaries (*s* is the scale boundary, *r* is the rotation boundary). The second column indicates the maximum possible shift of the image point achieved with the selected rotation and scale limit.

Tab. 1. The accuracy of optical-to-SAR image registration algorithms for different distortions with fixed algorithm parameters optimized on the s1.10_r10 distortion case

Distorsions scale_ rotation	max shift, pixels	SIFT	network1 trained on s1.10_r10	network2 trained on s1.20_r30
s1.00_r0	0	18/58	**55/58**	53/58
s1.00_r10	24	11/58	**52/58**	49/58
s1.00_r20	52	5/58	**43/58**	40/58
s1.00_r30	81	3/58	**45/58**	37/58
s1.10_r0	13	10/58	**50/58**	44/58
s1.10_r10	35	4/58	**51/58**	43/58
s1.10_r20	58	0/58	**41/58**	40/58
s1.10_r30	85	1/58	**35/58**	**35/58**
s1.20_r0	22	2/58	**43/58**	32/58
s1.20_r10	56	1/58	**44/58**	36/58
s1.20_r20	75	0/58	**40/58**	34/58
s1.20_r30	89	1/58	25/58	**35/58**

Tab. 2. The accuracy of optical-to-SAR image registration algorithms for different distortions. Individual optimal parameters were determined for each case of distortion

Distorsions scale_ rotation	max shift, pixels	SIFT	network1 trained on s1.10_r10	network2 trained on s1.20_r30
s1.00_r0	0	18/58	55/58	**56/58**
s1.00_r10	24	11/58	**54/58**	51/58
s1.00_r20	52	5/58	**51/58**	42/58
s1.00_r30	81	3/58	**45/58**	41/58
s1.10_r0	13	10/58	**51/58**	45/58
s1.10_r10	35	4/58	**51/58**	43/58
s1.10_r20	58	0/58	**46/58**	43/58
s1.10_r30	85	1/58	39/58	**40/58**
s1.20_r0	22	2/58	**44/58**	37/58
s1.20_r10	56	1/58	**46/58**	38/58
s1.20_r20	75	0/58	**40/58**	35/58
s1.20_r30	89	1/58	33/58	**35/58**

For the proposed algorithm, two neural networks trained on datasets with varying distortion's levels were considered. The results were compared with the «classical» image registration algorithm based on the SIFT detector and descriptor, described in section 3. The first neural network *network1* was trained on small distortions on the dataset s1.10_r10, the second neural network *network2* was trained on large distortions on the dataset s1.20_r30. Two cases were considered for each neural network: in the first case, the same algorithm parameters optimized for s1.10_r10 were used for all distortion cases (Table 1), in the second case, a separate optimization of parameters was carried out for each distortion case (Tab. 2). The largest values for each distortion case were highlighted by bold font.

The parameters used for *network1* were as follows: $t = 0.35, w = 30$, the size of the search window in the loss function 80 pixels. Parameters of the image registration algorithm when evaluating on s1.10_r10: the threshold of the descriptor distance was 0.4, the threshold of the RANSAC inliers was 10 pixels, the radius of the search window was 50 pixels. For *network2*, the optimal parameters when evaluating on s1.10_r10 were the same, but the radius of the search window in the loss function was 100 pixels. This was necessary because *network2* was trained for large distortions for which previous search window radius was not enough. The optimal parameters of the image registration algorithm were also the same, except for the threshold of the descriptor distance which was 0.5. For baseline algorithm based on SIFT detector and descriptor we used descriptor distance threshold 160, threshold of the RANSAC inliers 4 pixels, the radius of the search window 50 pixels.

Table 1 shows the results in a case with fixed parameters optimized on the s1.10_r10 dataset. These results reflected the accuracy of image registration for the same fixed algorithm for different distortion cases. For most distortion cases *network1* showed greater matching accuracy than *network2*. On the other hand, *network2* has more stable results for different distortions, which could be explained that *network1* was not trained on large distortions.

In the Tab. 2 were shown the results with individual optimal parameters determined for each distortion case. For most distortion cases, *network1* also showed greater image registration accuracy compared to *network2*. At the same time *network2* won with large distortions (s1.10_r30 and s1.20_r30). Also in this case, *network2* showed stable image registration accuracy for a fixed scale change (40 – 45 matched pairs and 35 – 37 matched pairs), except the case of no scale distortion (41-56 matched pairs). In both cases and for any distortion case, the proposed algorithm (with two networks) showed greater image registration accuracy compared to the algorithm based on the SIFT detector and descriptor (Tab. 1, 2). The results showed that *network1* was better than *network2* even on large distortion cases except the largest. But the difference was small. Hypothesis is that *network2* had the lack of training data because *network2* gave unstable results for small and large distortions in spite of all distortion cases were in limitations of trained dataset.

Conclusion

The paper proposes the neural network algorithm for optical-to-SAR image registration designed to compare roughly pre-aligned images. The accuracy of the pre-alignment was considered up to 100 pixels. This pre-alignment accuracy was chosen because the images from the Sentinel-1 and Sentinel-2 satellites have inconsistencies of up to 100 pixels when aligned by georeference. The proposed algorithm was based on the neural network that calculates descriptors corresponding to a uniform grid of points, as well as calculated descriptor distances between all possible pairs of descriptors. Using uniform grid of points allow to avoid using keypoint detectors which have errors with negative influence on overall accuracy and guarantee that almost each point could be matched with accuracy based on grid step. By the obtained descriptor distances the nearest neighbor method was used to determine the matching between the points. The resulting comparisons were submitted to the RANSAC algorithm with a limited affine geometric model (limited to translation, rotation and uniform transform) to calculate the geometric transformation between images.

The neural network was based on siamese neural network with branches based on the first two blocks of the resnet18 neural network. The loss function was based on the cosine similarity. Since it is difficult to achieve a cosine similarity equal to 0 for all unmatched pairs of points simultaneously we suggested to use a margin value t in the loss function which improved the training of the neural network by better dividing descriptor distances for matched and unmatched descriptors. The algorithm was tested on 58 pairs of images with varying levels of distortions in rotation and scale. The proposed algorithm surpassed the algorithm based on the SIFT detector and descriptor. With distortions within ±50 pixels, the matching accuracy is at least 43/58 pairs with successful matching criterion that the image corner after geometric transformation was matched with an accuracy of at least 10 pixels.

Acknowledgements

This work was supported by the Russian Science Foundation (project №20-61-47089).

References

[1] Errico A, Angelino CV, Cicala L, Persechino G, Ferrara C, Lega M, Vallario A, Parente C, Masi G, Gaetano R, Scarpa G. Detection of environmental hazards through the feature-based fusion of optical and SAR data: A case study in southern Italy. Int J Remote Sens 2015; 36(13): 3345-3367. DOI: 10.1080/01431161.2015.1054960.

[2] Plank S, Twele A, Martinis S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. Remote Sens 2016; 8(4): 307. DOI: 10.3390/rs8040307.

[3] Yu Q, Ni D, Jiang Y, Yan Y, An J, Sun T. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor. ISPRS J Photogramm Remote Sens 2021; 171: 1-17.

[4] Ye SP, Chen CX, Nedzved A, Jiang J. Building detection by local region features in SAR images. Computer Optics 2020; 44(6): 944-950. DOI: 10.18287/2412-6179-CO-703.

[5] Hamdi I, Tounsi Y, Benjelloun M, Nassim A. Evaluation of the change in synthetic aperture radar imaging using transfer learning and residual network. Computer Optics 2021; 45(4): 600-607. DOI: 10.18287/2412-6179-CO-814.

[6] Sidorchuk DS, Volkov VV. Fusion of radar, visible and thermal imagery with account for differences in brightness and chromaticity perception [In Russian]. Sensory Systems 2018; 32(1): 14-18. DOI: 10.7868/S0235009218010031.

[7] Schmitt M, Hughes LH, Zhu XX. The SEN1-2 dataset for deep learning in SAR-optical data fusion. arXiv Preprint. 2018. Source: <https://arxiv.org/abs/1807.01569>.

[8] Ye Y, Yang C, Zhu B, Zhou L, He Y, Jia H. Improving co-registration for Sentinel-1 SAR and Sentinel-2 optical images. Remote Sens 2021; 13(5): 928.

[9] Wang Z, Yu A, Zhang B, Dong Z, Chen X. A fast registration method for optical and SAR images based on SRAWG feature description. Remote Sens 2022; 14(19): 5060.

[10] Hansson N. Investigation of registration methods for high resolution SAR-EO imagery. Master of Science Thesis in Electrical Engineering. Linköping, Sweden: Linköping University; 2022.

[11] Shi W, Su F, Wang R, Fan J. A visual circle based image registration algorithm for optical and SAR imagery. 2012 IEEE Int Geoscience and Remote Sensing Symposium 2012; 2109-2112. DOI: 10.1109/IGARSS.2012.6351089.

[12] Suri S, Reinartz P. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. IEEE Trans Geosci Remote Sens 2009; 48(2): 939-949. DOI: 10.1109/TGRS.2009.2034842.

[13] Gong M, Zhao S, Jiao L, Tian D, Wang S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. IEEE Trans Geosci Remote Sens 2013; 52(7): 4328-4338. DOI: 10.1109/TGRS.2013.2281391.

[14] Fan B, Huo C, Pan C, Kong Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. IEEE Geosci Remote Sens Lett 2012; 10(4): 657-661. DOI: 10.1109/LGRS.2012.2216500.

[15] Ma W, Wen Z, Wu Y, Jiao L, Gong M, Zheng Y, Liu L. Remote sensing image registration with modified SIFT and enhanced feature matching. IEEE Geosci Remote Sens Lett 2016; 14(1): 3-7. DOI: 10.1109/LGRS.2016.2600858.

[16] Paul S, Pati UC. Optical-to-SAR image registration using modified distinctive order based self-similarity operator. 2018 IEEE Int Students' Conf on Electrical, Electronics and Computer Science (SCEECS) 2018: 1-5. DOI: 10.1109/SCEECS.2018.8546950.

[17] Xiang Y, Wang F, You H. OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas. IEEE Trans Geosci Remote Sens 2018; 56(6): 3078-3090. DOI: 10.1109/TGRS.2018.2790483.

[18] Paul S, Pati UC. Automatic optical-to-SAR image registration using a structural descriptor. IET Image Process 2019; 14(1): 62-73. DOI: 10.1049/iet-ipr.2019.0389.

[19] Xiong X, Xu Q, Jin G, Zhang H, Gao X. Rank-based local self-similarity descriptor for optical-to-SAR image matching. IEEE Geosci Remote Sens Lett 2019; 17(10): 1742-1746. DOI: 10.1109/LGRS.2019.2955153.

[20] Wang H, Wang C, Li P, Chen Z, Cheng M, Luo L, Liu Y. Optical-to-SAR image registration based on Gaussian mixture model. Int Arch Photogramm Remote Sens Spat Inf Sci 2012; 39: 179-183.

[21] Kunina I, Panfilova E, Gladkov A. Matching of SAR and optical images by independent referencing to vector map. Proc SPIE 2019; 11041: 1104102. DOI: 10.1117/12.2523132.

[22] Zhang W, Zhao Y. SAR and optical image registration based on uniform feature points extraction and consistency gradient calculation. Appl Sci 2023; 13(3): 1238.

[23] Kouyama T, Kanemura A, Kato S, Imamoglu N, Fukuhara T, Nakamura R. Satellite attitude determination and map projection based on robust image matching. Remote Sens 2017; 9(1): 90.

[24] Li X, Du Z, Huang Y, Tan Z. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. ISPRS J Photogramm Remote Sens 2021; 179: 14-34.

[25] Merkle N, Luo W, Auer S, Müller R, Urtasun R. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. Remote Sens 2017; 9(6): 586.

[26] Li Z, Zhang H, Huang Y. A rotation-invariant optical and SAR image registration algorithm based on deep and Gaussian features. Remote Sens 2021; 13(13): 2628.

[27] Volkov VV, Shvets EA. Dataset and method for evaluating optical-to-SAR image registration algorithms based on keypoints [In Russian]. Information Technologies and Computing Systems 2021; 2: 44-57. DOI: 10.14357/20718632210205.

[28] Terms of the Copernicus Data Hub portals and Data supply conditions. 2024. Source: <https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions>.

[29] Copernicus Open Access Hub website. 2024. Source: <https://scihub.copernicus.eu/>.

[30] Volkov V. Modification of the method of detecting and describing keypoints SIFT for optical-to-SAR image registration [In Russian]. Sensory Systems 2022; 36(4): 349-365. DOI: 10.31857/S0235009222040060.

[31] Tropin DV, Shemiakina JA, Konovalenko IA, Faradjev IA. Localization of planar objects on the images with complex structure of projective distortion [In Russian]. Information Processes 2019; 19(2): 208-229.

Authors' information

Vladislav Vladimirovich Volkov (b. 1994) graduated from Moscow Institute of Physics and Technology in 2022. Currently he works as junior researcher at Institute for Information Transmission Problems (Kharkevich Institute) RAS. Research interests are image registration, image processing, multispectral image visualization.
E-mail: *volkov-vl-v@yandex.ru*

Evgeny Aleksandrovich Shvets (b. 1990) graduated from the Moscow Institute of Physics and Technology, Moscow, Russia, then received the Ph.D. degree in Technology from the Institute for Information Transmission Problems, Moscow in 2017. His Ph.D. thesis focused on distributed control of a robotic swarm for distributed area surveillance. His research interests include image processing, image registration and deep learning, including leverage of synthetic data for zero-shot training. Evgeny currently works as Chief AI Officer in NVI Research. E-mail: *vortexd77@gmail.com*

Code of State Categories Scientific and Technical Information (in Russian – GRNTI)): 28.23.15
Received September 16, 2023. The final version February 01, 2024.