Matching and Co-Registration of Satellite Images Using Local Features

Mohamed Tahoun[1,3], Abd El Rahman Shabayek[1], Aboul Ella Hassanien[2,3]
[1]Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt
[2]Faculty of Computers and Information Sciences, Cairo University , Cairo, Egypt
[3]Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net
tahoun@informatik.hu-berlin.de, a.shabayek@ci.suez.edu.eg, aboitcairo@cu.edu.eg

***Abstract*—Satellite image matching and co-registration are two key stages in image registration, fusion and super-resolution imaging processes where images are taken from different sensors, viewpoints or at different times. This paper presents: (1) An evaluation for the co-registration process using local features, (2) A registration scheme for registering optical images taken from different viewpoints in addition to radar images taken at different times. The selected feature detectors have been tested during the key point extraction, descriptor construction and matching processes. The framework suggests a sub-sampling process which controls the number of extracted key points for a real time processing and for minimizing the hardware requirements. After getting the pairwise matches between the two images, a registered image is composited by applying bundle adjustment and image warping enhancements. The results showed a good performance level for SURF over both SIFT and ORB detectors in terms of higher number of inliers and repeatability ratios. The Experiments were done on different optical and radar images from Rapid-Eye, TerraSAR-X, and ASTER satellite data for some areas in Germany and Egypt.**

***Index Terms*—SURF, SIFT, ORB, BRISK, feature extraction and matching, satellite images.**

I. Introduction.

Detection and matching of features from satellite images taken from different sensors, viewpoints, or at different times are important tasks when manipulating and processing remote sensing data for many applications[6]. Depending on the application of the remote sensing data, some achievements have been remarked but there is still a need for improving these two processes for an accurate alignment and co-registration of satellite images with different modalities [5], [7]. Huge number of techniques have been proposed for this purpose with the aim to correctly detect and extract the important features and objects from images. A set of correspondent features or matching points between two images is used later to know how they both are related to each other. The features extracted from images could be local or global. They are usually represented by points, edges, corners and contours or by other features[2]. Once the features are extracted from images, the matching process starts by comparing the feature descriptors of the extracted keypoints. A Final set of inliers or tie points should be determined in order to co-registrating the input images[3]. Section (II) will briefly touch on different tested feature detectors. The detailed layout of the experiments is introduced in section (III). A discussion about the results is given in section (IV) and finally some conclusions and findings will be presented in section (V).

II. Feature Detection

Feature detection is the first step in image matching and registration. Local invariant features allow to find local image structures and to represent them invariantly to a range of image transformations. The purpose is to find a sparse set of local measurements that captures the essence of the input images. Two important criterias should be fulfilled in feature extractors. The first is to be precise and repeatable in order to make sure that the same features are extracted from different images. The second criteria is to be distinctive so that different image structures can be recognized from each other. It is also required to obtain a sufficient number of feature regions to cover important objects in the image. The general procedure for the extraction and matching processes includes the following steps:

- Finding distinctive keypoints.
- Taking a region around each keypoint in an invariant manner (e.g scale or affine invariant).
- Extracting and normalize the region content.
- Computing a descriptor for the normalized region.
- Matching the obtained local descriptors.

The keypoints are firstly detected from the input images then depending on the number of detected keypoints, a sub-sampling process may start if the number of the detected keypoints exceeds a user predefined number of keypoints. The aim of this step is to overcome the problem of huge keypoints from high resolution images which requires high configuration hardware and take much processing time. In the sub-sampling process, the keypoints are sorted according to their response. Keypoints with the best response are chosen within a predefined number of keypoints. Based on the new keypoints list, the descriptors are built and ready for the matching step. Four detectors have been chosen and evaluated in our experiments. ORB and BRISK detectors use binary visual descriptors while others use vector-based feature descriptors as in SIFT and SURF. The rest of this section gives a brief description for each detector by briefly underlining the most important features and the basic ideas.

SIFT (Scale Invariant Feature Transform) has been presented by Lowe in the year 2004 [9]. It has four major steps

including: scale-space extrema detection, keypoint localization, orientation assignment and finally building the keypoint descriptor. In the first step, points of interest are identified by scanning both the location and the scale of the image. The difference of Gaussian (DoG) is used to perform this step and then the candidates of the points are localized to sub-pixel accuracy. Then the orientation is assigned to each keypoint in local image gradient directions to obtain invariance to rotation. In the last step a 128- keypoint descriptor or feature vector is built and ready for the matching process. SIFT gives good performance but still have some limitations against strong illumination changes and big rotation angles.

SURF (Speeded-Up Robust Features) is a local invariant Interest point or blob detector [1]. It is partly inspired by the SIFT descriptor and is used too in static scene matching and retrieval. It is invariant to most of the image transformations like scale and illumination changes in addition to small changes in viewpoint. It uses Integral Images or an intermediate representation for the image and contains the sum of gray scale pixel values of image. Then a Hessian-based interest point localization is obtained using Laplacian of Gaussian of the image.

ORB (Oriented BRIEF-Binary Robust Independent Elementary Features) is a local feature detector based on binary strings [4]. It depends on a relatively small number of intensity difference tests to represent a patch of the image as a binary string.The construction and matching of this local feature is fast and performs well as long as invariant to large in-place rotations is not required. ORB is basically a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to enhance the performance. First it use FAST to find keypoints, then apply Harris corner measure to find top N points among them. For descriptor matching, multi-probe LSH which improves on the traditional LSH, is used [4].

BRISK (Binary Robust Invariant Scalable Keypoints) depends on easily configurable circular sampling pattern from which it computes brightness comparisons to form a binary descriptor string [8]. The authors claim it to be faster than SIFT and SURF.

III. Proposed Framework

As mentioned in the previous section, the main steps of the experiments include: Feature extraction, descriptor construction and matching, and finally returning the number of inliers or tie points between two or more images. Each detector has been run with different parameters and variables. Five types of descriptors are tested in the experiments including: SIFT, SURF, ORB, BRIEF and BRISK. The descriptor vectors are constructed from the sub-sampled extracted keypoints of the two input images then they are matched using similarity measurements. Different similarity measurements have been also tested including Euclidean and FLANN ((Fast Library for Approximate Nearest Neighbors).

The layout of the experiments enables the user to choose both the descriptor and the matcher types (figure 1). To remove outliers, RANSAC has been used then we get the final number of matches including the number of inliers and outliers. For each keypoint different parameters like 2-D location, scale and rotation are specified. An evaluation process starts where repeatability, correspondence and precision and recalls values are computed with each detector. Different performance evaluation measurements have been used to test the invariance of the tested feature detectors. Good features should be invariant to all possible changes that can exist between images.

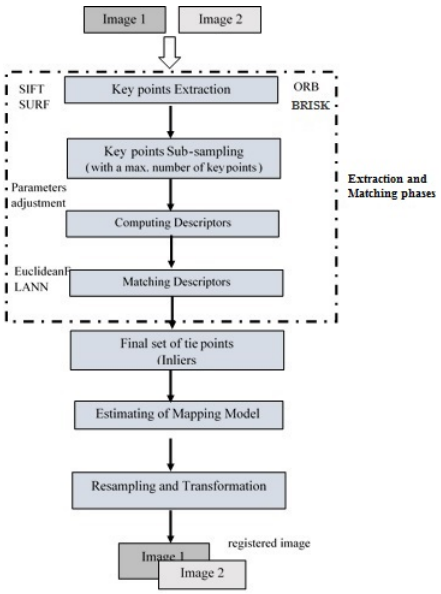

Figure 1. The general framework of the experiments

The general steps to match two input images in our experiments are as follows:

(1) Extract the keypoints from the input images and compute their descriptors (KP1 and KP2 if there are two input images). If the number of the detected keypoints exceeds a predefined number of the key points, then the detected keypoints are sub-sampled by sorting them according to their response. A keypoint response strength is measured according to its cornerness.

(2) Once the keypoint descriptors are built, they are ready for the matching process. In this step, the nearest neighbor is considered as the keypoint with minimum Euclidean distance for the invariant descriptor. Different image matchers or similarity measurements have been tested in the experiments. The Minkowski form distance is defined based on the Lp norm as:

$$D_p(S,R) = (\sum_{i=0}^{N-1}(S_i - R_i)^p)^{1/p} \quad (1)$$

Where $D_p(S,R)$ is the distance between the two feature vectors$S = S_1, S_2, ...S_{N-1}$ and $R = R_1, R_2, ...R_{N-1}$ representing the descriptors of the extracted keypoints from the input images. Euclidean distance ($p = 2$) recorded the most stable and acceptable results compared to Manhattan distance ($p = 1$) distance and FLANN matchers. Filter the matches using RANSAC to exclude the inconsistent matches and help getting the list of tie points which are the actual matches between the two input images (inliers).

In order to evaluate the previous matching process, we

applied the following steps: (1) Compute the homography between the filtered keypoints (H12 from first group of keypoints to the second group), (2) Compute the number of overlaps between KP1 and the transformed keypoints from the second image (using inverse of H12) KP2'. This number of overlaps is called Corresponding Count $C(I_1, I_2)$ [10]. (3) Divide $C(I_1, I_2)$ by the mean of the number of detected keypoints (d_1 and d_2). This value is called Repeatability (Re)[6]:

$$Re_{1,2} = \frac{C(I_1, I_2)}{mean\,(d_1, d_2)} \tag{2}$$

Synthesize Test: In this test, an input image (S) is transformed to a new image (N) by applying a perspective transformation as follows :

$$N(x,y) = S(\frac{M_{11}x + M_{12}y + M_{13}}{M_{31}x + M_{32}y + M_{33}}, \frac{M_{21}x + M_{22}y + M_{23}}{M_{31}x + M_{32} + M_{33}}) \tag{3}$$

where M is the transformation matrix (3x3). The homography matrix values are randomly generated within a uniform distribution. The same procedure for the extraction and matching processes is applied on both the input and new image. The same steps could be repeated between the input image and any transformed version from it. Figure 3 illustrates the inliers between an original ASTER image for Suez Canal area and its synthesized version after applying homography changes.

Precision and recall graphs**:** They are commonly used in image processing and information retrieval fields. Precision is considered as the ratio between the number of relevant points and the total number of the matches while recall is the the ratio between the number of relevant divided by the total number of the matches in the data. This graph displays the relation between the average precision and recall at any selected point of all feature detectors.

Processing time**:** after reading the input images, the running time of each stage of the experiments is computed. Firstly, the time of the keypoint detection is computed. secondly, the keypoints sub-sampling time is also computed in case the detected keypoints exceed the predefined limit. Then, the time of constructing the descriptor of each image is computed followed by the finally the matching time of both input image descriptors. The total time of the whole process is simply equal to the cumulative sum of all the computed times in all the stages.

IV. Experimental Results

The experiments have been run on rapid-Eye and ASTER images in addition to TerraSAR-X images with different resolution (5m, 0.75m, 0.5m) and dimensions (for example 5000x5000, and 10000x10000) (figure 2). The hardware and software configuration is: Intel-core™2 quad CPU 2.66GHz and 8GB RAM, Windows 7 (64bit) and Visual Studio 10.0 and OpenCV 2.46. This section includes some results of applying our scheme on optical and radar images for some areas in Egypt and Germany.

Different comparisons have been done amongst SURF, SIFT, ORB and BRISK. They are tested using different descriptor and matcher types. The number of keypoints / matches

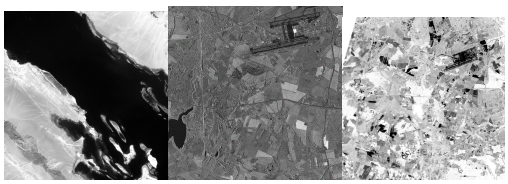

Figure 2. Different sample images including Aster image for an area in Suez Canal (Egypt) (left) and TerraSAR-X radar (middle) and Rapid-Eye image (right) of Berlin Brandenburg airport area

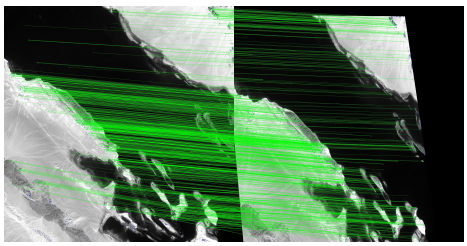

Figure 3. The inliers between an original ASTER image for Suez Canal area (left) and its synthesized version after applying homography changes

changes when different homography evaluation (transformation changes) are randomly applied to an input image and compared with the original image. The experiments have showed that SURF and SIFT give good results over ORB and BRISK with different samples but ORB still gives good results with optical images. SIFT has scored a good performance with TerraSAR-X radar images (include much noise) compared to the other detectors. Furthermore, ORB and BRISK have scored very low performance with TerraSAR-X data in terms of the inliers ratios and repeatability (figures 4 and 5).

The similarity measurements: Euclidean distance, Manhattan and FLANN have been also tested in our experiments. They affect some how on the similarity check between descriptors. Euclidean distance matcher has recorded the most stable and accurate results with different descriptors. SIFT has given the most stable performance with higher numbers of inliers compared to the other descriptors. SIFT and BRISK still take longer time to be built while SURF and ORB take less time. SIFT detector usually requires more memory with high resolution images while the sub-sampling process works good to solve the huge number of detected points with other detectors. Once the pairwise matching is done, a bundle adjustment and image warping enhancements are done

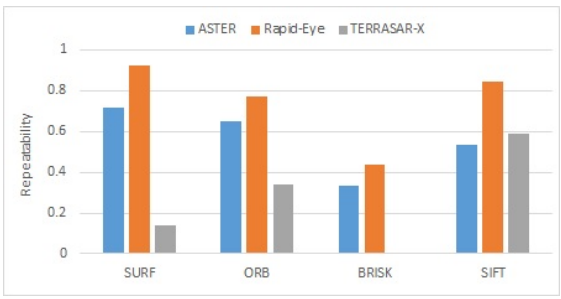

Figure 4. The general repeatability averages of the tested feature detectors

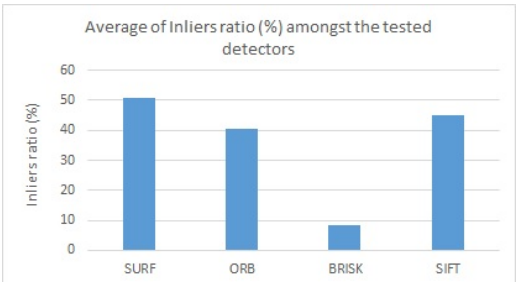

Figure 5. The general inliers ratios (%) of the tested feature detectors showing better performance for SURF over the other detectors.

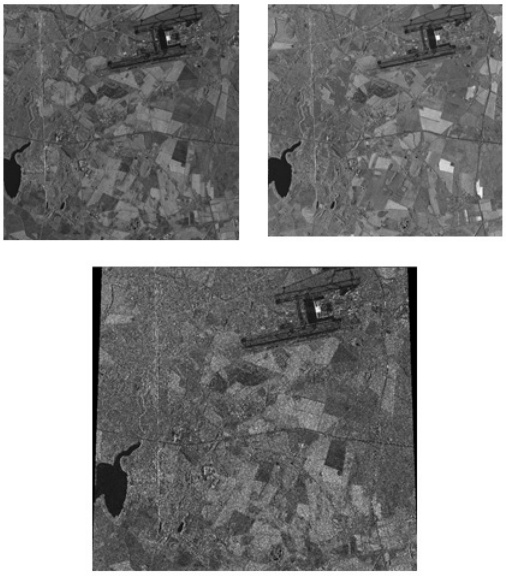

Figure 6. Two TerraSAR-X images for Berlin Brandenburg airport taken at different times (25.07/25.08.2010) (up) and their registered version (down) using SURF

Figure 7. Two Rapid-Eye images for the area of Berlin Brandenburg airport (left) and their registered version (right) using SURF

before compositing the registered or stitched image. In the compositing process the input images are re-sized and blended before getting the final registered version. Figures 6 and 7 give an example of the registration process where TerraSAR-X radar images taken at different times and the the Rapid-Eye optical ones are taken from different view points. They are some parameters like the confidence of the matching steps and the bundle adjustment cost function need to be adjusted for each tested satellite data for the sake of reaching the best performance. The system also enables the users to modify and adjust different flags during the compositing stage like warping surface type, seam estimation method and the resolution of the compositing step (in Megapixels).

V. Conclusions

In this paper, some local feature detectors have been evaluated on satellite images. The suggested framework uses a keypoints sub-sampling process and variety of parameters to enhance both matching and co-registration processes. The detected keypoints are sub-sampled (in case of high resolution images) in order to reduce the running time during the extraction and the matching stages. SURF has recorded the best stable performance and running time compared to the other tested detectors including SIFT, ORB and BRISK. SIFT detector recorded the best inliers ratios on TerraSAR-X data while it has a weak performance with other tested images for example the Rapid-Eye images. Although SIFT descriptor still takes longer time but it gives the most stable inliers ratios among the other tested descriptors like SURF, BRIEF and ORB. The findings of this work are important for a full and efficient registration scheme for registering both optical and radar images taken from different sensors.

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van GoolHY. Surf: Speeded up robust features. In *Proceedings of the ninth European Conference on Computer Vision*, May 2006.

[2] Rochdi Bouchiha and Kamel Besbes. Automatic remote-sensing image registration using surf. *International Journal of Computer Theory and Engineering*, 5(1):88–92, 2013.

[3] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[4] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. *11th European Conference on Computer Vision (ECCV)*, 2010.

[5] T. D. Hong and R. A. Schowengerdt. A robust technique for precise registration of radar and optical satellite images. *Photogrammetric Engineering and Remote Sensing*, 71(5):585–594, May 2005.

[6] Luo Juan and Oubong Gwon. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4):143–152, 2009.

[7] Nabeel Younus Khan, Brendan McCane, and Geoff Wyvill. SIFT and SURF performance evaluation against various image deformations on benchmark dataset. In Andrew P. Bradley and Paul T. Jackway, editors, *DICTA*, pages 501–506. IEEE, 2011.

[8] Stefan Leutenegger, Margarita Chli, and Roland Siegwart. Brisk: Binary robust invariant scalable keypoints. *ICCV*, pages 2548–2555, 2011.

[9] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 2:91–110, 2004.

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.