Journal of
Applied Remote Sensing

High-resolution multispectral satellite image matching using scale invariant feature transform and speeded up robust features

Mustafa Teke
M. Firat Vural
Alptekin Temizel
Yasemin Yardımcı

M. Teke, M. F. Vural, A. Temizel, and Y. Yardimci, "High-resolution multispectral satellite image matching using scale invariant feature transform and speeded up robust features", J. Appl. Remote Sens. 5, 053553 (2011)

http://dx.doi.org/10.1117/1.3643693

SPIE

High-resolution multispectral satellite image matching using scale invariant feature transform and speeded up robust features

Mustafa Teke,[a] M. Firat Vural,[b] Alptekin Temizel,[a] and Yasemin Yardımcı[a]

[a] Middle East Technical University, Graduate School of Informatics, Ankara, 06531 Turkey
mustafa.teke@gmail.com, atemizel@ii.metu.edu.tr, yardimy@ii.metu.edu.tr

[b] Eindhoven University of Technology, Software Technology, 5612 Eindhoven, Netherlands
m.f.vural@tue.nl

Abstract. Satellite images captured in different spectral bands might exhibit nonlinear intensity changes at the corresponding spatial locations due to the different reflectance responses for these bands. This affects the image registration performance negatively as the corresponding features might have different properties in different bands. We propose a modification to the widely used scale invariant feature transform (SIFT) method to increase the correct feature matching ratio and to decrease the computation time of this algorithm for the multispectral satellite images. We also apply scale restriction to SIFT and speeded up robust features (SURF) algorithms to increase the correct match ratio. We present test results for variations of SIFT and SURF algorithms. The results show the effectiveness of the proposed improvements compared to the other SIFT- and SURF-based methods. © *2011 Society of Photo-Optical Instrumentation Engineers (SPIE)*. [DOI: 10.1117/1.3643693]

Keywords: image registration; image matching; satellite image processing; SIFT; SURF.

Paper 10166R received Oct. 27, 2010; revised manuscript received May 27, 2011; accepted for publication Sep. 7, 2011; published online Sep. 23, 2011.

1 Introduction

Each band of remotely sensed images contains unique information due to the different electromagnetic reflection properties of the surfaces at different wavelengths. Satellites offer such data via images obtained using cameras with the ability of observing different bands. The images might be acquired from cameras having different parameters or might be taken at different times and from different view-points. Images might have been gathered from different sources or satellites and with different resolutions. In such cases, the images from different bands need to be registered to facilitate combining of distinct information contained in each band.

Multispectral satellite images have nonlinear intensity changes.For these reasons, moment-based scale invariant methods are more favorable for the registration of multispectral images.

Scale invariant feature transform (SIFT) is an effective and popular feature-based technique[1,2] and it has largely replaced the corner detectors that have commonly been used for feature-based registration. Another popular feature-based technique is the more recently introduced speeded up robust features (SURF).[3] Both SIFT and SURF features are invariant to rotation, translation, scaling, and linear intensity changes. These methods essentially consist of two stages. First, potential features that can be used for matching are determined in both images. Then, the best matches between feature points in these images are obtained via comparison of the feature vectors accompanying these feature points. These characteristic feature vectors mainly include the histogram of the intensity gradient spatial directions around the feature points.

1931-3195/2011/$25.00 © 2011 SPIE

SIFT assumes similar intensity changes around the corresponding features between the images to be matched. However, in multimodal images these points and their vicinities could have different characteristics, and hence the accompanying feature vectors for these points would need to be modified for correct matching. The gradient orientation modification SIFT (GOM-SIFT) technique[4] has been developed to overcome this problem by feature orientation reversal, treating the intensity changes in both directions (dark-to-light or light-to-dark) in the same manner. Nevertheless, the performance of this method decreases significantly with the increasing rotation between the images. Another method aiming to increase correct match ratio by descriptor refinement and applying scale-orientation restriction criteria similar to GOM-SIFT is given in Ref. 5. In Ref. 6, a method that combines maximally stable extremal regions (MSER) detector and the SIFT descriptor is introduced.

SURF is known to be faster than the SIFT method and it is claimed to be more repeatable and less sensitive to noise.[3] The SURF detector is based on the Hessian matrix. The Hessian matrix is approximated by using box filters of second order Gaussian derivatives. As opposed to SIFT, SURF increases filter size instead of downsampling images to extract features at different scales.

Upright-speeded up robust features (U-SURF) is a variety of SURF that can be used when there is no or little rotation (up to 15°). U-SURF is faster to compute as it disregards the orientation information.[3]

Methods such as SIFT and SURF find features in different scales. However, a constant scale difference between the images can be assumed when there is an affine relation between the target image and the reference image. Scale restriction (SR)[4] utilizes this information to reject features matched at different scales. Restricting scale difference histograms increases performances of registration by removing incorrect matches found at distant scale ratios.

Mutual information (MI) is commonly used for multimodal image registration because it is robust to intensity changes. An algorithm that employs mutual information is described in Ref. 7. As MI does not provide scale and rotation invariance, they used the orientation and scale estimates obtained via SIFT to achieve rotation and scale invariance. Intensity invariance of the algorithm is demonstrated on a synthetically inverted image pair. Another MI-based algorithm that uses SIFT and MI for multisensor SAR images is described in Ref. 8. MI is used to obtain course estimates of the registration parameters which are utilized at the SIFT feature matching stage to improve consistency of the matches.

In this paper, we propose a modification to the SIFT vector representation to increase the successful matching ratio of SIFT features between images of different modalities for the same feature vector bin size. The proposed method could also be used to reduce the computation time at a similar correct matching rate by using a reduced number of bins. We also propose a parameter independent scale restriction method. We applied this method to SIFT and SURF. We evaluated and compared matching performances of the variations of the proposed method with other SIFT- and SURF-based methods aiming multispectral image matching in the literature as well as features from accelerated segment test (FAST)+ histogram of oriented gradients (HOG).

2 Proposed Method

2.1 Orientation-Restricted (OR)-SIFT

SIFT assumes that the feature point areas have gradients in the same direction, while GOM-SIFT works by inverting the gradients when the gradient direction is more than 180°, and hence accounting for inversions in gradient directions between corresponding images. However, there could be more variation in the intensity changes between multimodal images than simple gradient negation as in GOM-SIFT. The same surface on different modality images may be observed with inversed intensities as demonstrated in Fig. 1. Such nonlinear variations between the images adversely affect SIFT performance as inverted feature orientations also change the order of the descriptor vector.

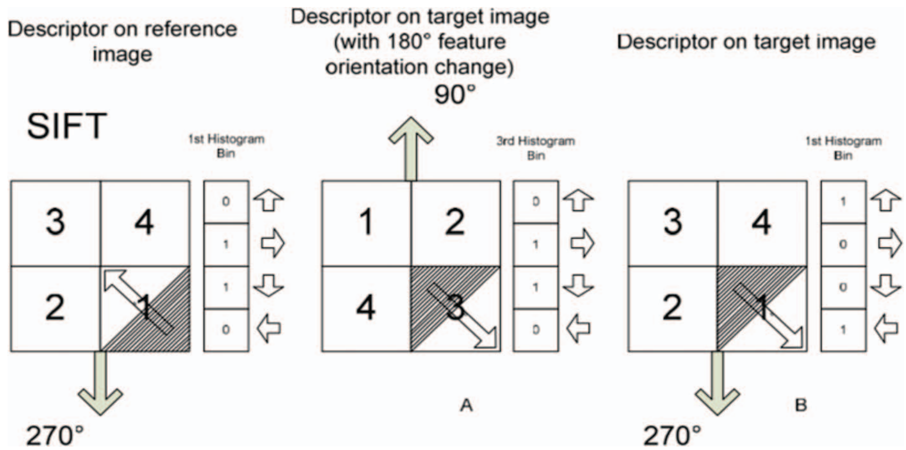

Fig. 1 Reaction of SIFT descriptors to gradient orientation change.

Orientation-restricted (OR)-SIFT aims to minimize this error. In order to allow for more variation, we construct the feature vectors by merging the histogram bins corresponding to the opposite directions. For example, the histogram bin containing the 0° to 45° gradient direction and the bin containing the 180° to 225° information are merged in a single bin as the gradients in both these directions indicate the same edge albeit with different gradient directions, resulting in a single bin containing the summation of opposite bins. The feature vector of the feature given in Fig. 2(a) with orientation as in Fig. 2(b) will be represented in four bins instead of eight as shown in Fig. 2(c).

Figure 3 illustrates the behavior of the OR-SIFT descriptor vectors for the same feature in Fig. 1. As can be seen from Fig. 3, the feature vectors are the same for inverted feature directions and hence matching of these features is facilitated.

The proposed scheme increases the robustness of SIFT for image pairs of different modalities, which can be observed by comparing Fig. 1 to Fig. 3. Moreover, the OR-SIFT method reduces the feature vector size by one-half and, as a result, the computation time is reduced since shorter feature vectors need to be matched. We call this method "orientation-restricted SIFT" or OR-SIFT.

2.2 Scale Restriction

Often there is a constant scale factor between the images and the matching performance can be increased by restricting the scale differences between the corresponding features.[4] A match is

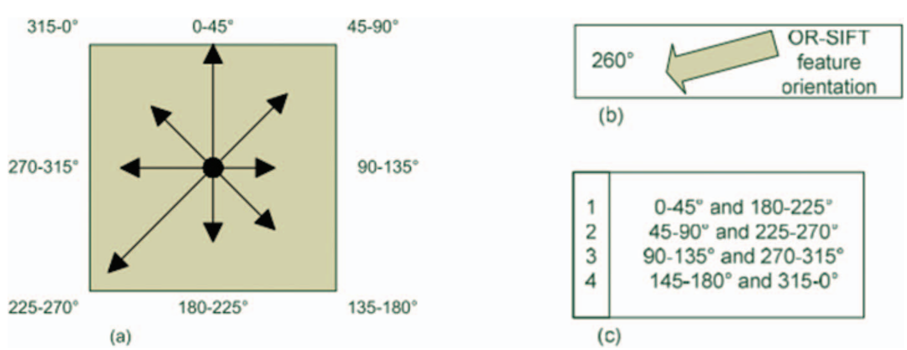

Fig. 2 (a) Presentation of 1 of 16 sub-regions' normalized histogram. (b) Modified orientation of feature according to OR-SIFT. (c) Information contained in different bins in OR-SIFT.

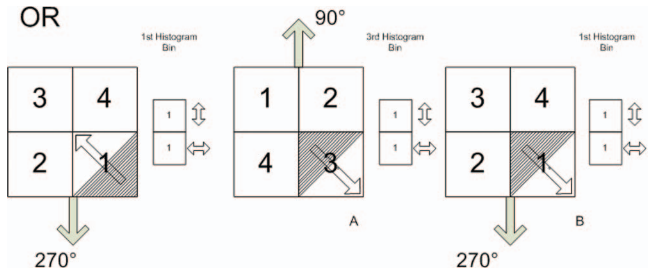

Fig. 3 Reaction of OR-SIFT descriptors to gradient orientation change.

rejected if the scale difference of the matched features does not satisfy the SR criteria. Scale difference (SD) for a key point pair $P_1(x_1, y_1, \sigma_1, \theta_1)$ and $P_2(x_2, y_2, \sigma_2, \theta_2)$ is defined as:

$$\mathrm{SD}(P_1, P_2) = |\sigma_1 - \sigma_2| \tag{1}$$

where x and y are the spatial locations, σ is the scale, and θ is the rotation of the key point. SD is calculated by taking the absolute value of the difference to account for scale differences. Scale restriction rejects the ones outside the range:

$$\overline{\mathrm{SD}} - W < \mathrm{SD} < \overline{\mathrm{SD}} + W. \tag{2}$$

For SIFT, $\overline{\mathrm{SD}}$ is selected as the peak of the histogram of all SDs.[9] Experiments show that the maximum correct match ratio is achieved with a threshold W between 0.10 and 0.13. However, the selection of $\overline{\mathrm{SD}}$ and W are image dependent.

Another scale restriction method is defined in Ref. 10 based on scale ratios of Harris interest points. This method finds the mean of scale difference histogram, $\mathrm{SD}_{\mathrm{MEAN}}$. Then the matches with scale difference SD are selected such that

$$0.5\mathrm{SD}_{\mathrm{MEAN}} \leq \mathrm{SD}_{\mathrm{MEAN}} \leq 1.2\mathrm{SD}_{\mathrm{MEAN}}. \tag{3}$$

However, this algorithm is not as successful as SR in Ref. 4 when the peak value of scale difference histogram is close to zero. SD histograms are not always distributed in the same way to allow constant thresholds; as a result, performance of this approach is dependent on the shape/distribution of the scale difference histogram.

We have applied a different method to determine $\overline{\mathrm{SD}}$ and W. In our experiments, we have observed that the scale difference of correct matches is clustered around a constant value; hence we propose to assign $\overline{\mathrm{SD}}$ to the mean value of the scale differences of all matches. Then we assign the standard deviation of scale differences to W.[11] This allows automatic calculation of the scale restriction parameter independent of scale and rotation. Figure 4 shows histograms of scale differences versus number of matches for images having scaling about 1.6. $\overline{\mathrm{SD}}$ is calculated as 1.25 and W is calculated as 0.95. Since most of the correct matches have scale differences between $\overline{\mathrm{SD}} - W$ and $\overline{\mathrm{SD}} + W$, the correct match ratio is increased when we apply SR. For this image, the correct match ratio increases to 84% from 77% when SR is applied.

2.3 Algorithm Steps

Figure 5 shows the main steps of the SIFT and SURF algorithms. The first step is preprocessing. Images taken from different bands of satellite sensors might have different contrast ranges. The preprocessing step equalizes contrast ranges of the input images. In this step, either contrast stretching or histogram equalization could be used. In our experiments we have used histogram equalization as it provided better results, especially for blue and near-infrared (NIR) band registration. In the second step, an image matching task is performed: the descriptors are extracted and matched. Finally, scale restriction is applied to improve correct match rates.

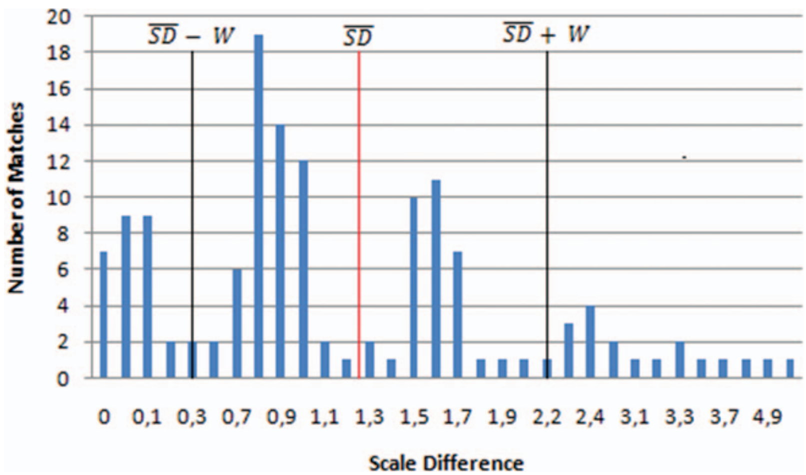

Fig. 4 Scale difference histogram of matches for NIR and Red image pair having scaling ratio of 1.6 with $\overline{SD} = 1.25$ and $W = 0.95$.

3 Experimental Results

We used the red, green, NIR band, and panchromatic images from 35 different QuickBird high resolution data. The images are from urban areas with different characteristics, each image having a size of 500×500 pixels. The results have been obtained by averaging the correct match percentages for all the images. The corresponding images have the same orientation and they are at the same scale to eliminate the effects of these transformations. SIFT and SURF comparisons are performed by having to produce the same number of total matches. The results are shown in Table 1 where the ratio of the correct match percentage is denoted by *TCM*. Verification of correct matches is performed by checking if the matches are in 3 pixels neighborhood of each other. If the points in a matched pair are in 3 pixels neighborhood of each other, it is marked as a correct match.

In our tests, we used existing SIFT (Ref. 12) and SURF (Ref. 13) implementations and made the proposed modifications to these. We also included a HOG (Ref. 14) feature descriptor. The HOG descriptor uses gradients around a point and combines a number of directional gradient information as a descriptor. In our tests, we selected directional gradients 20° apart. HOG descriptors have been calculated around interest points detected by the FAST (Ref. 15) corner detection method. FAST is reported to be a robust and accurate corner detection method. We have used the code developed for the work in Ref. 16 for testing.

As can be seen in Table 1, SR increases the performance regardless of the base algorithm used, so SR usage is suggested to reduce the false matches when there is a constant scale difference between the images. The proposed SR method produces better results, especially when used in conjunction with OR-SIFT. SR increases the correct match ratio for SURF significantly.

Table 1 shows that OR-SIFT generates comparable results when only half the number of bins are used with the other methods that it is compared with, namely the original SIFT method[2]

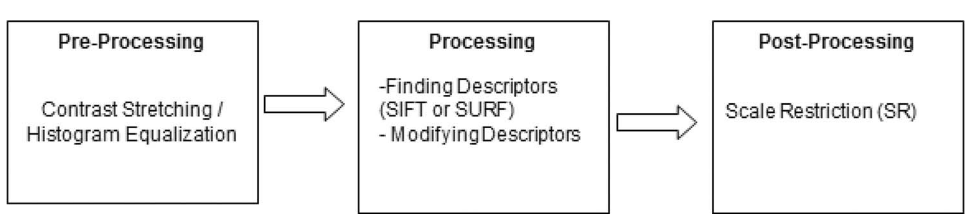

Fig. 5 Main steps of feature matching using SIFT or SURF.

Table 1 Comparison in terms of the ratio of the correctly matched features to all the features.

	No. of bins	NIR and green band *TCM*%	NIR and red band *TCM*%	NIR and panchromatic *TCM*%
SIFT (Ref. 2)	8	87.17	88.4	91.25
	16	91.02	91.76	93.03
SIFT and SR (Ref. 2)	8	94.23	95.03	95.42
	16	95.82	96.7	96.33
SIFT and SR	8	96.36	96.99	96.96
	16	95.84	96.6	96.28
GOM-SIFT (Ref. 3)	8	87.15	88.9	91.23
	16	90.76	91.8	93.33
GOM-SIFT and SR (Ref. 9)	8	94.46	95.35	95.56
	16	95.87	96.56	96.58
GOM-SIFT and SR	8	96.22	96.8	97.01
	16	95.44	96.52	96.36
OR-SIFT (Ref. 9)	8	90.6	92.09	92.83
	16	92.62	93.68	94.03
OR-SIFT and SR (Ref. 9)	8	95.53	96.67	96.45
	16	96.58	97.45	97.1
OR-SIFT and SR	8	98.38	98.67	98.95
	16	98.15	98.54	98.72
SURF	–	83.06	87.49	87.98
SURF and SR	–	86.65	90.12	90.75
U-SURF	–	87.03	88.32	88.80
U-SURF and SR	–	90.34	91.24	91.16
FAST + HOG		62.29	71.64	54.44

and GOM-SIFT.[3] OR-SIFT generates better results than the other methods when the same number of bins is used. SURF method variants have lower performance than SIFT. Results for FAST + HOG-based matching are lower compared to SIFT and SURF. This is due to the corners detected using HOG descriptors in different modalities having different characteristics. In the tests, the NIR band is compared with other bands. Comparison of RGB bands with each other produced *TCM* around 99% to 100%. For this reason we have not included visible band test results. Figure 6 shows the results in Table 1 as a graph for 8-bin SIFT methods, SURF and FAST + HOG, while results for 16-bin SIFT methods are given in Fig. 7.

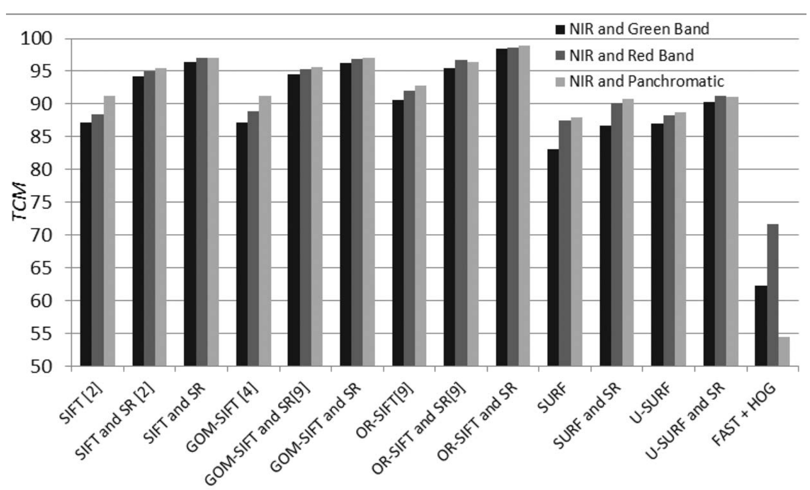

Fig. 6 Comparison in terms of the ratio of the correctly matched features to all the features for 8-bin SIFT methods, SURF, and FAST + HOG.

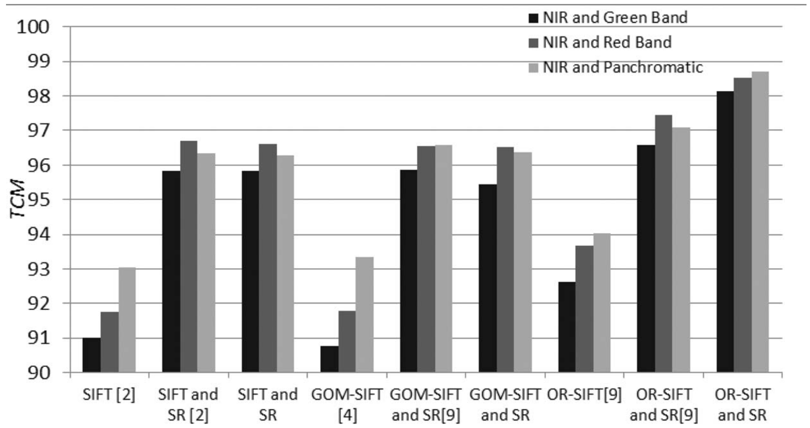

Fig. 7 Comparison in terms of the ratio of the correctly matched features to all the features for 16-bin SIFT methods.

Table 2 shows the processing times for SIFT,[12] SURF,[13] and upright-speeded up robust features (U-SURF).[13] Several tests to compute performances of descriptors are performed using different SIFT and SURF parameters. All 35 images are used in the calculations. The test results are obtained on a computer with Intel Core 2 Duo P6750 CPU and Windows 7. The average computation time for each descriptor is 0.58, 0.25, and 0.13 ms, respectively. So, when a similar number of descriptors is calculated, SURF is observed to be 2.32 times faster than SIFT, while U-SURF is 4.46 times faster. It has to be noted that using a halved number of bins for SIFT decreases the total processing time between 15% and 25% depending on the number of extracted feature points on the image. Also, doubling the SURF descriptor size to 128 decreases SURF performance by 10% due to computation of additional descriptor components.

In Fig. 8, QuickBird NIR and visible red band images of the same region are shown. The nonlinear intensity differences between the NIR band image and red band image can easily be observed in these images. All feature points that are correctly matched by the proposed OR-SIFT method are indicated by isolated green dots in the figure. The incorrect matches are shown by lines connecting the matched feature pairs. For this patch, of the 137 matches made, GOM-SIFT (Ref. 4) results in 77 correct matches while OR-SIFT results in 80 matches. The correct match ratio increases from 56.20% to 58.39% with the proposed method.

Figure 9 shows the change in the *TCM* with the increasing rotation between the images for these three methods. *TCM* values, when there is no rotation, are taken as reference and the change as ratio of this value is given in the *y*-axis. As can be seen from Fig. 9, GOM-SIFT performance degrades significantly with the increasing rotation. Almost 60% of the correct matches are lost at 60° rotation. Even though OR-SIFT performance degrades slightly more compared to SIFT, it is robust and holds itself well with the increasing rotation. Repeatability of SURF decreases as rotation increases, while a constant *TCM* is maintained. U-SURF is not robust against rotation for angles exceeding 15 deg. SURF is less robust to rotation when compared to SIFT.

All three SIFT methods respond similarly to scaling and are observed to be robust against scaling between the images. It has been observed that repeatability of SURF and U-SURF

Table 2 Comparison of the methods in terms of processing time.

	SIFT	SURF	U-SURF
Total time (ms)	46260	24140	12812
Total number of descriptors	79165	97630	97137
Time/descriptor (ms)	0.58	0.25	0.13

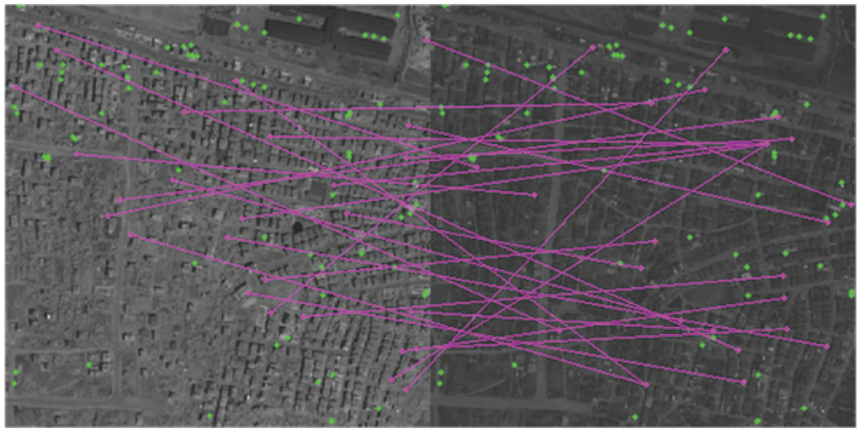

Fig. 8 NIR band image (left) and red band image (right). The correctly matched features are shown by green dots and the incorrect matches are connected with lines.

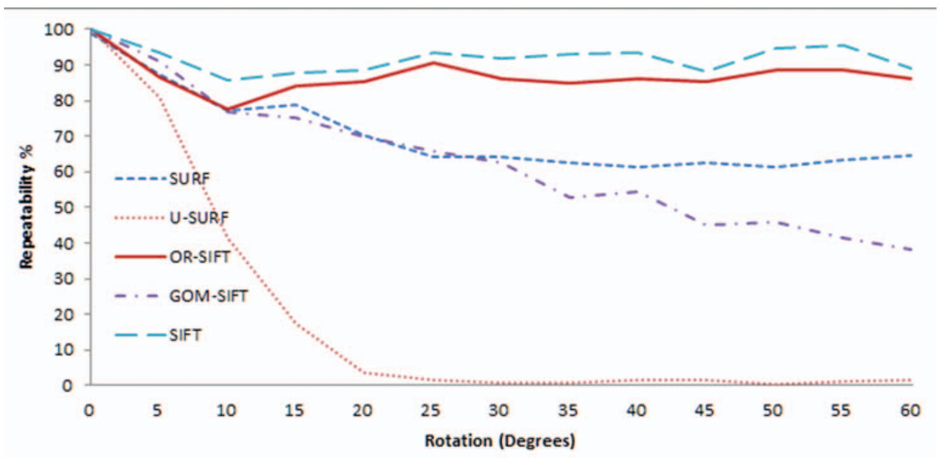

Fig. 9 Responses of SIFT and SURF algorithms against rotation.

methods decreases as the scaling ratio between images increases while a constant *TCM* value is maintained.

4 Conclusion

We proposed methods based on SIFT and SURF to facilitate matching between images with different modalities. The algorithm has been shown to have a higher correct matching ratio than SIFT and GOM-SIFT when the same number of bins is used. Alternatively, a similar performance with a lower-complexity is obtained by halving the number of bins. It has also been shown to be more robust against rotation compared to GOM-SIFT.

The proposed scale restriction method is a general method applicable to both SIFT and SURF and produces better matching performance compared to the previous methods. Also, it is adaptive to scale and rotation changes since its parameters are calculated for each case independently.

SURF and U-SURF have been observed to underperform when a similar number of total matches as SIFT is used. Scale restriction increases the performance of both SURF and U-SURF. U-SURF performs better than SURF when there is no or small rotation. On the other hand, while SURF and U-SURF produce lower *TCM* values, they are faster to compute.

In a future work, orientation compensation can also be achieved by employing a normalized correlation coefficient measure for SIFT descriptor matching instead of the Euclidian distance between the orientation restricted descriptors.

References

1. D. G. Lowe, "Object recognition from local scale invariant features," in *International Conference on Computer Vision*, Corfu (1999).
2. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
3. H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
4. Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *IET Elect. Lett.* **44**(2), 107–108 (2008).
5. Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett* **6**(2), 287–291 (2009).
6. L. Cheng, J. Gong, X. Yang, C. Fan, and P. Han, "Robust affine invariant feature extraction for image matching," *IEEE Geosci. Remote Sens. Lett.* **5**(2), 246–250 (2008).
7. H. H. Jeon, A. Basso, and P. F. Driessen, "A global correspondence for scale invariant matching using mutual information and the graph search," *IEEE International Conference on Multimedia and Expo* (2006).
8. S. Suri, P. Schwind, P. Reinartz, and J. Uhl, "Combining mutual information and scale invariant feature transform for fast and robust multisensor SAR image registration," in *Proceedings of the 75 ASPRS Annual Conference* (2009).
9. M. F. Vural, Y. Yardimci, and A. Temizel, "Registration Of multispectral satellite images with gradient-corrected SIFT," in *Proceedings of IGARSS'09* (2009).
10. X. J. Liu, J. Yang, and H. Shen, "Automatic image registration by local descriptors in remote sensing," *Opt. Eng.* **47**, 087206 (2008).
11. M. Teke and A. Temizel, "Multi-spectral satellite image registration using scale-restricted SURF," in *Proceedings of the International Conference on Pattern Recognition* (2010).
12. A. Vedaldi, "SIFT++," Online. Available: http://www.vlfeat.org/∼vedaldi/code/siftpp.html. (Accessed: April 2010).
13. H. Bay, A. Ess, and G. Willems, "SURF Implementation," [Online]. Available: http://www.vision.ee.ethz.ch/∼surf/index.html. (Accessed: April 2010).
14. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2005).
15. E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1), 105–119 (2010)
16. O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: an application to pedestrian detection," in *12th International IEEE Conference On Intelligent Transportation Systems*, Vol. 1, pp. 432–437, St. Louis (2009).

Biographies and photographs of the authors not available.