A SIFT-BASED MODE-SEEKING PROCEDURE FOR EFFICIENT, ACCURATE REGISTRATION OF REMOTELY SENSED IMAGES

Benny Kupfer[1], Nathan S. Netanyahu[2,3], Ilan Shimshoni[4]

[1]Department of Mathematics, Bar-Ilan University, Ramat-Gan 52900, Israel
[2]Department of Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel
[3]Center for Automation Research, University of Maryland at College Park, MD 20742
[4]Department of Information Systems, Haifa University, Haifa 31905, Israel

ABSTRACT

Several image registration methods, based on the *scaled-invariant feature transform* (SIFT) technique, have appeared recently in the remote sensing literature. All of these methods attempt to overcome problems encountered by SIFT in multi-modal remotely sensed imagery, in terms of the quality of its feature correspondences. The method presented in this paper performs mode seeking (in transformation space) to eliminate outlying corresponding key-points (i.e., features) and improve the overall match obtained. Preliminary experimental results seem to indicate that our method achieves high accuracy and is rather fast in a variety of test cases.

Index Terms— Remotely sensed images, image registration, feature correspondence, mode-seeking SIFT.

1. INTRODUCTION

Image registration (IR) of multi-temporal/spectral/sensor images is most essential in a variety of remote sensing applications, e.g., environmental monitoring, image mosaicking and classification, change detection, etc. Although many approaches have been proposed over the years for IR of remotely sensed data, the problem is still rather challenging, prompting various ongoing innovations in an attempt to obtain enhanced performance, vis-à-vis accuracy, running time, etc. In this paper we propose an efficient variant based on the *scale invariant feature transform* (SIFT) [1]. It performs reliable filtering of outlying feature correspondences by *mode seeking* of scale ratios, rotation differences (and eventually horizontal and vertical shifts) between all corresponding SIFT key-points. Our mode-seeking SIFT (MS-SIFT) is very simple and fast, and appears to achieve sub-pixel accuracy.

2. PRIOR WORK

In previous work [2, 3], we used edge-like wavelet features to perform feature matching by hierarchical searching in transformation space. An initial bounding box was used to reduce the elaborate search required at the higher resolution levels and the *partial Hausdorff distance* (PHD) was chosen as a similarity measure. Our new method is designed to alleviate, to a significant extent, the need for the above described search process. Other researchers also draw strongly on the notion of SIFT for image registration of remotely sensed data. However, since SIFT often results in inaccurate (if not incorrect) matching when applied to multi-modal remotely sensed imagery, the main thrust is to obtain first a reliable set of corresponding key-points. Li *et al.* [4] proposed to refine the SIFT key-point orientations and assign multiple orientations to each key-point. Outlier filtering is based on the ratio between the Euclidean distance to the closest neighbor and the second closest neighbor (as proposed in [1]), followed by similar pruning with respect to the so-called *joint distance* (JD). An iterative search between all orientation differences is used to find the best match (in a JD sense) for the resulting key-points. Teke *et al.* [5] proposed *orientation restricted SIFT* (OR-SIFT). Orientations with opposite directions are binned together to compensate for inversions in the gradient orientations, and matches are based on nearest-neighbor (NN) distances between SIFT features, where false matches are excluded if their SIFT key-points scale distance is larger than a predefined threshold. Sedagat *et al.* [6] proposed the *uniform robust SIFT* (UR-SIFT) algorithm, where extracted SIFT key-points are distributed evenly in both the scale and image spaces. Key-points with low principal curvature are rejected; further rejection is obtained by checking each correspondence in a global transformation model between the reference and sensed images. Li *et al.* [7] performed matching using the rotation-invariant distance between SIFT key-points in a polar grid. Matching is done via RANSAC [8], followed by a final transformation computation. Hasan *et al.* [9] proposed a 2-step procedure which also rejects

outliers according to the distance ratio between the first and second NNs. RANSAC is used to exclude remaining outliers, and a global transformation is computed according to so-called primary and secondary matched feature points. Finally, Hasan *et al.* [10] proposed numerous modifications in the SIFT procedure, e.g., preserving every SIFT key-point, limiting gradient values to reduce the effect of strong edges, using a larger window for the SIFT descriptor, etc.

3. PROPOSED METHOD

The above described methods require typically thousands of SIFT key-points in both the reference and sensed images, even for a standard image size. Also, some of them require exhaustive search and matching. Registration run-times are rarely reported, and those reported vary from tens to hundreds of seconds (for complete registration). In contrast, our method uses a relatively high threshold to detect initially only (up to) hundreds of SIFT key-points for each image. Key points of the reference and sensed images are then matched according to NNs of corresponding SIFT descriptors. We abandon the traditional approach of filtering outliers according to the distance ratio between the first and second NNs [1]. Much in the spirit of a Hough-like voting scheme, we exploit, instead, the inherent information of each SIFT key-point (i.e., scale and position orientation) to compute a prospective transformation for each match (i.e., corresponding key points). In principle, we perform mode seeking in 4-D space (assuming a similarity transformation), which is done in practice for each of the four components separately. This is followed by effective pruning of outlying correspondences and a refined computation of the transformation. We start by computing a histogram of the scale ratios for all of the SIFT key-point matches and find its mode scale, s_{mode}. Similarly, we compute a histogram of the orientation differences for all the matches and find its mode rotation difference, $\Delta\theta_{\text{mode}}$. Our experimental results show that for a variety of multi-temporal and multi-spectral images the histogram modes are unique and evident (at least 40% higher than the next peak). We now use the scale ratio and rotation difference found to perform mode seeking of the horizontal and vertical translations as follows. Let (x, y) and (x', y') denote, respectively, the coordinates of a SIFT key-point in the reference image and its corresponding key point in the sensed image. Each pair of corresponding key points defines the following horizontal and vertical shifts:

$$\Delta x = x - s_{\text{mode}}\left(x'\cos(\Delta\theta_{\text{mode}}) - y'\sin(\Delta\theta_{\text{mode}})\right) \tag{1.1}$$

$$\Delta y = y - s_{\text{mode}}\left(x'\sin(\Delta\theta_{\text{mode}}) + y'\cos(\Delta\theta_{\text{mode}})\right). \tag{1.2}$$

We now compute two additional histograms of Δx and Δy for all corresponding key-points, for which we find the mode values, Δx_{mode} and Δy_{mode}, respectively. The quadruple obtained, $< s_{\text{mode}}, \Delta\theta_{\text{mode}}, \Delta x_{\text{mode}}, \Delta y_{\text{mode}} >$, is used (as a transformation approximation) to eliminate outlying key-point pairs according to the following logical filters:

$$F_1 : \left|\Delta x - \Delta x_{\text{mode}}\right| \geq \Delta x_{\text{thresh}} \tag{2.1}$$

$$F_2 : \left|\Delta y - \Delta y_{\text{mode}}\right| \geq \Delta y_{\text{thresh}} \tag{2.2}$$

where Δx and Δy are given in Eqs. (1.1-1.2), and Δx_{thresh} Δy_{thresh} denote, respectively, thresholds of horizontal and vertical differences, in terms of corresponding histogram bin widths (measured in pixels). All corresponding pairs $(x, y) \Leftrightarrow (x', y')$ for which F_1 or F_2 hold will be considered outliers and thus rejected. Although this filters out typically 80% — 90% of the initial correspondences (for the data sets we have experimented with), the resulting set of correspondences is very reliable, so that it suffices to employ at this stage a 1-step *ordinary least squares* (OLS) procedure. In a nutshell, this is done by first computing the transformation that aligns the centroids of the (remaining) point sets, then computing the scale factor that aligns their spatial variances, and finally computing the rotation that minimizes the sum of squared distances [3].

4. EXPERIMENTAL RESULTS

We have implemented our MS-SIFT procedure in C using a single thread style, and tested it on several remotely sensed image pairs. Performance (in terms of accuracy) was evaluated in each case via manual ground truth (GT), using the *root mean square error* (RMSE) criterion. Picking manually N corresponding points $(x_i, y_i) \Leftrightarrow (x_i', y_i')$ from the reference and sensed images, the RMSE is computed according to:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2} \tag{3}$$

where $(\tilde{x}_i, \tilde{y}_i)$ denotes the transformed coordinates of (x_i', y_i').

We present the following results. Figs. 1(a) and 1(b) show two 512×512 Landsat images acquired at different times. The underlying transformation clearly consists of substantial rotation and translation and a mild scale. The SIFT threshold gave rise to 763 and 806 key-points for the reference and sensed images, respectively. Figs. 1(c) and 1(d) depict scale ratio and rotation difference histograms (with bin widths of 0.05 and $9°$), respectively. The modes are easily located at $s_{\text{mode}} = 0.991$ and $\Delta\theta_{\text{mode}} = 17.92°$. Figs. 2(a) and 2(b)

show the histograms of the horizontal and vertical shifts, computed by Eqs. (1.1-1.2), with equal bin widths of 7.5 pixels. The modes obtained at $\Delta x_{\text{mode}} = 93.42$ and $\Delta y_{\text{mode}} = 313.25$ are clearly evident. Employing the outlier filter(s) (Eqs. (2.1)-(2.2)) with horizontal and vertical shift thresholds of a bin size results in 112 points (out of 763 correspondences), i.e., a rejection of ~85%. Finally, employing the 1-step OLS yields a bottom-line transformation $< s, \theta, t_x, t_y > = < 0.99, 15.02°, -82.05, 307.49 >$. The RMSE in this case was 0.314 pixels (for 96 pairs).

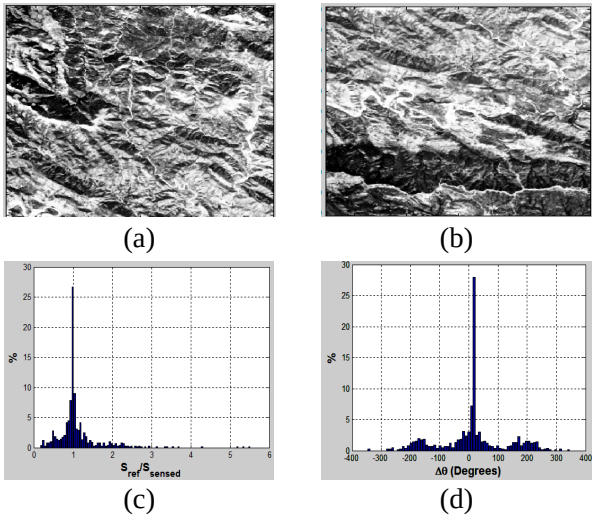

Fig. 1: (a)-(b) Reference and sensed Landsat images (source: UCSB website of image processing and vision research labs), and (c)-(d) scale ratio and orientation difference histograms.

Fig. 2(c) illustrates the registration output. The above registration took 2.31[sec] on an old PC (Intel Q8200 with 3[GB] RAM and Vista OS) and 1.56[sec] on a powerful laptop (HP Elitebook with Core i-7 and 12[GB] RAM with Windows-7 OS).

Figs. 3(a) and 3(b) depict an image pair of size 312×312 over the Cascades site acquired by Landsat ETM+ and IKONOS in the near infra-red band (NIR), respectively. (Wavelet decomposition was utilized here to bring the data to a similar spatial resolution, i.e., the IKONOS image was transformed to a spatial resolution of 32 meters, applying three levels of decomposition. Thus the scaling expected for the given images is roughly 1.07.) In this case we had 223 initial SIFT correspondences. Figs. 3(c) and 3(d) depict the scale ratio and rotation difference histograms, respectively. Figs. 4(a) and 4(b) show the corresponding histograms of the horizontal and vertical shifts respectively. Fig. 4(c) depicts the registration result. The transformation obtained was $< s, \theta, t_x, t_y > = < 1.064, -0.09°, 8.73, 10.27 >$; this was computed from 142 inliers (63% of the initial number of correspondences). The RMSE in this case was 0.88 pixels and the run-time on our standard PC was 0.92[sec].

For all of the data sets tested thus far, we achieved sub-pixel accuracy for consistent histogram binning. (See www.cs.biu.ac.il/~nathan/IGARSS_13/supplementary.pdf where registration results for 14 additional image pairs are presented.)

5. CONCLUSIONS

In this paper we presented a simple SIFT-based variant for image registration of remotely sensed images. Our proposed method performs, essentially, mode-seeking in 4-D space (assuming a similarity transformation), followed by effective pruning of outlying SIFT key-point correspondences. Preliminary results show good promise; in particular, the method seems to perform accurately and fast. As part of future research, we intend to further validate the performance of our MS-SIFT module by experimenting with a large, diverse data set of multi-temporal/spectral/sensor images. Specifically, it would be of interest to detect automatically when the method fails and to study such cases more extensively.

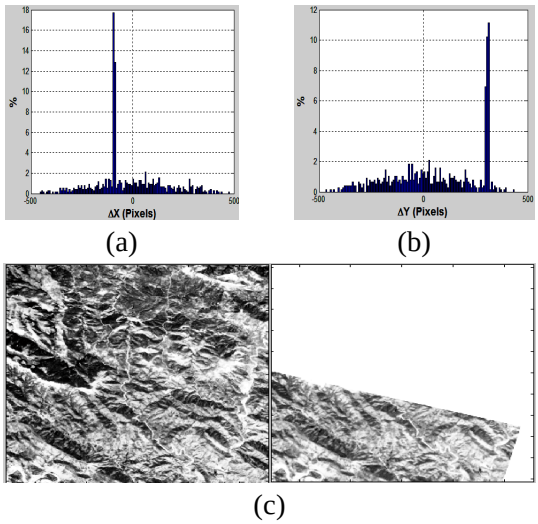

Fig. 2: Further processing of first image pair: (a)-(b) Histograms of horizontal and vertical shifts, and (c) superimposed images after registration.

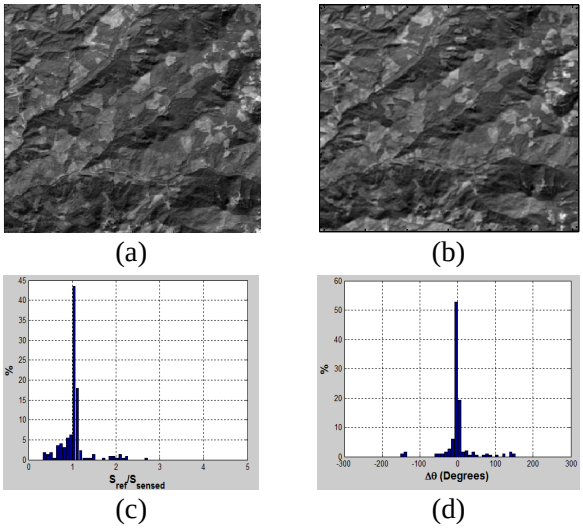

Fig. 3: (a)-(b) Reference and sensed images over Cascades (source: MODIS Validation Core Sites), and (c)-(d) scale ratio and orientation difference histograms.

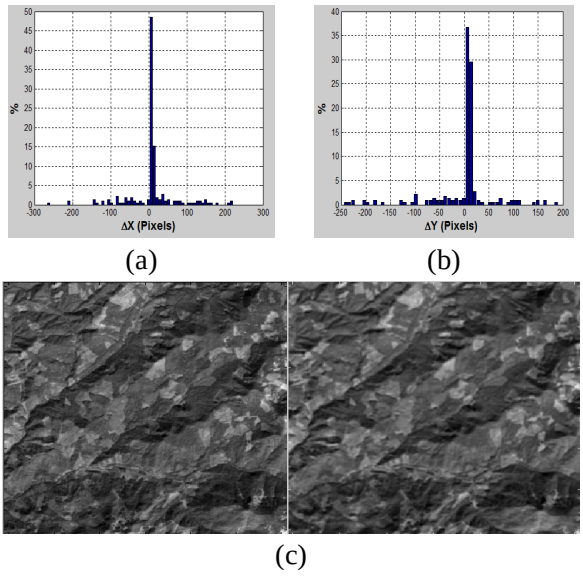

Fig. 4: Further processing of second image pair: (a)-(b) Histograms of horizontal and vertical shifts, and (c) superimposed images after registration.

6. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments and suggestions. Also, we are grateful to Ardeshir Goshtasby, Jacqueline LeMoigne, and Dave Mount for many useful and stimulating discussions regarding SIFT-based image registration for remote sensing. Finally, we are indebted to Andrea Vedaldi for his SIFT code (in MATLAB and C).

7. REFERENCES

[1] D.G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," *International. Journal of Computer Vision.*, vol. 60, no. 2, pp. 91—110, November 2004.

[2] N.S. Netanyahu, J. LeMoigne, and J.G. Masek, "Georegistration of Landsat Data via Robust Matching of Multiresolution Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1586—1600, July 2004.

[3] D.M. Mount, N.S. Netanyahu, and S. Ratanasanya, "New Approaches to Robust, Point-Based Image Registration," in *Image Registration for Remote Sensing*, J. LeMoigne, N. S. Netanyahu, and R. D. Eastman, Eds., pp. 179—199, Cambridge University Press, March 2011.

[4] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287—291, April 2009.

[5] 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," *Journal of Applied Remote Sensing*, vol. 5, no. 1, pp. 053553-1—053553-9, 2011.

[6] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform Robust Scale–Invariant Feature Matching for Optical Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4516—4527, November 2011.

[7] Q. Li, H. Zhang, and T. Wang, "Multispectral Image Matching Using Rotation-Invariant Distance," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 406—410, May 2011.

[8] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of The ACM*, vol. 24, no. 6, pp. 381—395, June 1981.

[9] M. Hasan, X. Jia, A. Robles-Kelly, J. Zhou, and M.R. Pickering, "Multi-Spectral Remote Sensing Image Registration via Spatial Relationship Analysis on SIFT Keypoints," *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing*, pp. 1011—1014, July 2010.

[10] M. Hasan, X. Jia, and M. R. Pickering, "Modified SIFT For Multi-Modal Remote Sensing Image Registration," *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing*, pp. 2348—2351, July 2012.