Multi-Modal Registration Using a Block-Wise Coarse-to-Fine Approach

MUHAMMAD TARIQ MAHMOOD, (Senior Member, IEEE),
AND IK HYUN LEE, (Member, IEEE)
1School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan 31253, South Korea
2Department of Mechatronics Engineering, Korea Polytechnic University, Sihieung 15073, South Korea
Corresponding author: Ik Hyun Lee (ihlee@kpu.ac.kr)
This work was supported in part by the Basic Science Research Program under Grant 2017R1D1A1B03033526 and Grant 2016R1D1A1B03033860, and in part by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1A6A1A03015562.

ABSTRACT Multi-modal registration is a fundamental problem in medical imaging, remote sensing and computer vision applications. Moreover, it is significantly challenging task as compared to mono-modal registration due to the complicated relationships between various modalities. Conventional approaches for registering multi-modal images suffer from limited correct matched points, high computational cost and poor registration accuracy. In this paper, we propose a multi-modal registration method using block-wise processing and coarse-to-fine approach. The proposed method efficiently extracts larger number of distinctive key-points and consequently provides better registration accuracy. In addition, it is able to handle weak illumination conditions effectively. In order to evaluate the effectiveness of the proposed method, a number of experiments are carried out using well-known datasets. The performance of the proposed scheme is then compared with the state-of-the-art methods using precision qualitative measure. The comparative analysis demonstrates that the proposed method has achieved higher precision values as compared to the obtained from conventional methods. In addition, the proposed scheme has provided more accurate registration results through lower computational complexity.

INDEX TERMS Multi-modal image, image registration, feature extraction.

I. INTRODUCTION
Image registration is an important step in medical imaging, remote sensing and image fusion systems. It is a procedure for spatial alignment of the same area acquired from same or different views and sensors. Multi-modal registration [1]–[4] i.e. alignment of images from more than one sensor is significantly difficult task as compared to mono-modal registration [5]–[8]. The main challenges of using two or more sensors at the same time include limited corresponding features extractions, complicated relationship between pixel intensities, variations in contrast of images. For instance, in medical imaging, the alignment of color fundus photography (CFP) and scanning laser ophthalmoscopy (SLO) images is a challenging task. Similarly, in remote sensing, registration of images from infrared sensor and electro-optics (EO) sensor is a demanding task.

Color fundus photography (CFP) and scanning laser ophthalmoscopy (SLO) have been well-known imaging procedures for investigating fundoscopic abnormalities. The both imaging procedures have different characteristics and mechanisms. In CFP, fundus camera records color images of the condition of the interior surface of the eye including the retina, retinal vasculature, optic disc, macula, and posterior pole. On the other side, the SLO imaging technique is capable of capturing fine structures at hefty magnification and high frame rate. It also allows accurate diagnosis of retinal structures that are poorly seen by ordinary fundus cameras. It uses low levels of light exposure and improved contrast [1], [2]. For better diagnosis, invaluable characteristics of both CFP/SLO images are incorporated. Therefore, multi-modal registration with higher accuracy is important.

Similarly, electro-optics (EO) and Infrared (IR) sensors are widely used to acquire visual information for studies in including geography, land surveying and most earth science
disciplines. Electro-optics (EO) image is acquired by using reflection and radiation of visible rays, whereas infrared (IR) image absorbs the radiation energy and temperature information. Generally, infrared sensors provide richer information than EO sensors, particularly in poor lighting conditions, smoke, fog, and cloudy weather. Generally, single-modal registration is applied for many applications using visible range sensors during day time and in indoor environments. However, many critical applications including military & defense and night vision demand utilization of both EO/IR sensors for richer information. Hence, the multi-model registration is applied in alignment of EO/IR images for accurate information.

In general, the registration methods can be divided into two main categories: (1) area-based methods [1], [2] and (2) feature-based methods [3]–[8]. Similarity measurement is key factor for accuracy in these two methods. Area-based methods use the pixel intensities of corresponding regions to extract correct corresponding features between two or more images. On the other hand, feature-based methods use features including points, curves, lines, branches and regions. Previously, many attempts have been done in multi-modal images by using area-based methods and feature-based methods. The original mutual information (MI) based method provides poor results due to differences between distributions of the multi-modal images. Rosin et al. [1] suggested the window based MI that performed better than non-window based MI. Another approach based on MI is proposed by Legg et al. [2] that has incorporated the features from the neighborhood. In the proposal of Berger et al. [3], first, edges were extracted by using Canny edge detection, then partial Hausdorff distance was applied to register images. Chen et al. [5] proposed another feature-based method using bifurcation structure of vessels. Another method is proposed by Chen et al. [4] using partial intensity invariant feature descriptor. These approaches provide reasonable registration accuracy when edges or bifurcation points are well preserved. In addition, area-based methods have high computational complexity for measuring the similarity between multi-modal images. Therefore, these methods are not suitable for registering CFP/SLO and EO/IR images. In addition, the feature-based methods suffer from extracting limited number of distinctive features, while area-based methods also suffer from weak illumination conditions [9], [10].

In this paper, we propose a registration method for multi-modal images by using block-wise processing and coarse-to-fine approach. The proposed method overcomes the problems with feature based and area based methods. It extracts more distinctive features efficiently and is also able to handle weak illumination conditions effectively. Experiments are performed using well-known datasets and the performance of the proposed method is evaluated using quantitative measure. The comparative analysis demonstrates the effectiveness of the proposed registration method for multi-modal images. In the remainder of the paper, Section 2 presents the proposed method. The experimental results are covered in Section 3. Finally, the section 4 concludes this study.

II. PROPOSED METHOD

The block diagram for the proposed framework is represented in Fig.1. Features such as points, curves, lines, branches, and regions are important factors for registration accuracy. In medical imaging, extracted features for scanning laser ophthalmoscopy (SLO) image are comparatively smaller than color fundus photography (CFP) image due to the image size (narrow field of view (FOV)). In remote sensing, extracted features using infrared (IR) image are comparatively smaller than Electro-optics (EO) image due to the low contrast. In non-block processing [11], the computational cost is higher for image re-sampling, local transformation using B-spline [12] and image re-sampling using bicubic interpolation [13]. The proposed coarse-to-fine approach handles these issues. In coarse approach, first, images are down sampled using Wavelet Packet transform. Then key-points are extracted using Harris corner using block-wise processing. Finally, descriptors are refined using random sample consensus (RANSAC) before transformation estimation and the registration process is performed. In fine approach, all steps for registration are repeated with up-sampled images. The subsequent subsections provide detail about the coarse and fine approaches.

A. COARSE STEP

1) IMAGE DOWN-SAMPLING USING WAVELET PACKET

The first step of the proposed approach is to get the signal at the lower scale by down sampling. The image features can be obtained more efficiently at lower scale, as the image size becomes smaller and consequently computational cost is reduced. In the proposed method, down sampling is carried out by using Wavelet packet transform (WPT) and is useful tool for denoising, compression and analysis of the signals. It is a generalization of wavelet transform (WT) [14], [15] that can be generated from the translation of a dilation function; which are defined as \( \psi_{s,\tau} (t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right) \) with discrete values for the parameters \( s = 2^j \) and \( \tau = k t_0 s^j \). The discrete wavelet of a given signal \( x (t) \) can be obtained as,

\[
W (j, k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x (t) \psi \left( \frac{t - k_0 2^j}{2^j} \right) dt \tag{1}
\]
Comparing with the WT, WPT uses number of filters to decompose the detail part (the high frequency components) at each level to get high resolution in the high frequency regions [14]. Wavelet packet decomposition can be described as

$$W_{2n}(x) = \sqrt{2} \sum_k h_k W_n(2^p x - k)$$

$$W_{2n+1}(x) = \sqrt{2} \sum_k g_k W_n(2^p x - k).$$

where $p$ is a scale index and $k$ is the translation index, $h$ is low pass filter, $g$ is the high pass filter, $W_0(x) = \phi(x)$ is the scaling function and $W_1(x) = \psi(x)$ is the basic wavelet function. Wavelet packets are well localized in both time and frequency domains and thus provide an attractive alternative to frequency analysis.

Wavelet packet decomposition differs from the normal Wavelet transform. In wavelet decomposition, at each level, approximation part (low frequency components) is further decomposed into approximation and detail parts (high frequency components) that provides richer analysis of high frequency components. The Wavelet packets procedure results in a large number of decompositions and requires a criterion to find the optimal decomposition. Among others, the selection of tree nodes on the base of best entropy values can be a reasonable criterion. In this work, we have used Shannon entropy measure to calculate the entropy that can be calculated as:

$$E(S_i) = - \sum s_i^2 \log(s_i^2).$$

Using the above measure, an optimal and compact tree can be constructed. In our procedure, a node will split into two or more children nodes if and only if the sum of the entropy of children nodes is less than the entropy of the parent node. This is a local criterion based only on the information available at the parent node.

From the Fig. 2, the difference between the best and the original tree can be observed. Figure 2 (a) represents the Wavelet packet tree. The best tree is shown in Fig. 2 (c) that is obtained by using the Shannon entropy measure. Figure 2 (b) and (d) shows the decomposed input image. For feature extraction, the best tree after calculating the wavelet packets of the repository images up to the second level is used.

2) FEATURE EXTRACTION USING HARRIS CORNER

The Harris detector is widely used in many computer vision applications for corner detector. It uses the second moment matrix and autocorrelation matrix. It also determines intensity change along a window patch. This detector is described as follow

$$M = \begin{bmatrix} < I_x^2 > & < I_x I_y > \\ < I_x I_y > & < I_y^2 > \end{bmatrix}.$$ 

where $I_x^2$, $I_y^2$ and $I_x I_y$ are the second-order derivatives of the image intensities in the $x$, $y$, and $xy$ directions. Harris corner detector determines the corner by analyzing eigenvalues of matrix $M$ as follow

$$R(M) = \det(M) - \kappa \text{trace}^2(M) = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2.$$ 

where $\kappa$ value is in the range 0.04 to 0.15 empirically in the literature [16]. If the eigenvalues $\lambda_1$, $\lambda_2$ has large positive values, the candidate points are considered as corner.
Difference between a scale-invariant feature transform (SIFT) [17] and Harris detector is detected feature type. SIFT extracts blob and region, whereas Harris detects corner. Feature detector should extract corresponding features in both CFP/SLO images, but SIFT is not appropriate detector due to the lack of blob similarity between CFP/SLO images. Corner detector can extract features, even low relation of pixel intensity between CFP/SLO images.

Figure 3 shows the extracted features using SIFT, Harris corner. From the figure, we can observe that Harris corner detection provides more distinctive features than SIFT.

3) FEATURE DESCRIPTOR AND MATCHING

To extract matching points, SIFT based descriptor used in the study, Mikolajczyk and Schmid [18] evaluated the performance of local descriptors and showed that SIFT provides the best performance with respect to other descriptors. The descriptors are histogram of the image gradient and orientations.

- Histogram: For orientation invariance, The sampling grid for the histograms is rotated to the main orientation of each keypoint. The grid is a \( 4 \times 4 \) array of \( 4 \times 4 \) sample cells of 8-bin orientation histograms, which produces 128-dimensional feature vectors.
- Normalization: To achieve the invariance of illumination changes, the descriptor is normalized with respect to the unit length.
- Gaussian weighting: This function is applied to give less importance to gradients farther from descriptor center and to avoid sudden changes.

Putative matching points are obtained by using nearest-neighborhood distance ratio (NNDR) [18], which uses the threshold in the ratio between the first and the second nearest neighbor descriptors. NNDR can be defined as:

\[
NNDR = \frac{\|\text{descr}_A - \text{descr}_B\|}{\|\text{descr}_A - \text{descr}_C\|},
\]  

where \( \text{descr}_A - \text{descr}_B \) and \( \text{descr}_A - \text{descr}_C \) are distances to the nearest and second nearest neighbors, \( \text{descr}_A \) is the base descriptor, and \( \text{descr}_B \) and \( \text{descr}_C \) are its closest two neighbors.

4) OUTLIER REMOVAL USING RANSAC

After determining putative matching points, a set of data that consists of inliers and outliers are given. To eliminate the outliers, random sample consensus (RANSAC) [19] is used to cope with a large proportion of outliers. The objective, in the absence of prior noise distribution is to provide an initial matching point set. The RANSAC [20] is applied to the initial matching points set to estimate the homography. The sample size is four because four match pairs determine a homography. The number of samples is set adaptively as the proportion of outliers is determined from each consensus set.

5) TRANSFORMATION ESTIMATION AND IMAGE RE-SAMPLING

Image transformation is the process to register sensed image to reference image. One of the basic and widely used global transformation models is the affine transformation. This transformation includes image translation, rotation, and scaling. In affine model, a point \( x = (x, y) \) can be scaled or rotated respectively. The affine model is defined as follows.

\[
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix} =
\begin{bmatrix}
s \cos \theta & s \sin \theta & 0 \\
-s \sin \theta & s \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
u \\
v \\
1
\end{bmatrix}
\]  

(9)

The transformed sensed image is interpolated with 2 times magnification via image re-sampling. Among various methods, pixel created in bicubic interpolation [13] is using \( 4 \times 4 \) neighborhood pixels. Thus, this interpolation technique can decrease variations in pixels as it provides different weight to both side. Therefore, there is possibility for further improvement in image quality, however other interpolation techniques such as nearest neighbor interpolation, bilinear interpolation require higher computational time. In conclusion, in spite of drawback mentioned above, the proposed interpolation technique provides higher accuracy and smoother surface in reasonable computational time. Thus, this interpolation technique can decrease variations in pixels as it provides different weight to both side. Therefore, there is possibility for further improvement in image quality, however other interpolation techniques such as nearest neighbor interpolation, bilinear interpolation require higher computational time. In conclusion, in spite of drawback mentioned above, the proposed interpolation technique provides higher accuracy and smoother surface in reasonable computational time.

B. FINE STEP

High resolution image contains more detailed information than low resolution image. It also has repetitive structures such as building, house, etc. Repetitive structures cause false matching points. Aerial images contain local distortions due to different sensors having different paths, angles. Therefore, the number of accurate matching points affects the accuracy and performance. Block processing [11] helps to reduce false matching points with geometrical constraint. Even block processing is successfully applied to register multi-modal images; this approach is accurate only having same paths.
TABLE 1. Performance comparison of CFP/SLO pairs using original SIFT, improved SIFT (PSO-SIFT), and proposed block-based coarse-to-fine approach (BCF).

| Corresponding Points | Sample 1 | Sample 2 |
|----------------------|----------|----------|
|                      | SIPT     | PSO-SIFT | Harris | BCF Coarse | BCF Fine | SIFT     | PSO-SIFT | Harris | BCF Coarse | BCF Fine |
| Correct matching     | 102      | 1,347    | 118    | 53        | 141      | 116      | 1,451    | 140    | 44        | 197      |
| precision            | 0.0098   | 0.0000   | 0.1271 | 0.1132    | 0.2766   | 0.0345   | 0.0000   | 0.1429 | 0.3636    | 0.3706   |

| Corresponding Points | Sample 3 | Sample 4 |
|----------------------|----------|----------|
|                      | SIFT     | PSO-SIFT | Harris | BCF Coarse | BCF Fine | SIFT     | PSO-SIFT | Harris | BCF Coarse | BCF Fine |
| Correct matching     | 98       | 1,413    | 69     | 17        | 97       | 21       | 1,500    | 87     | 29        | 94       |
| precision            | 0.0000   | 0.0000   | 0.1739 | 0.2353    | 0.2474   | 0.0476   | 0.0000   | 0.1264 | 0.1724    | 0.1915   |

Figure 4. Image blocks for 20% overlapped block processing: (a) CFP image, (b) SLO image.

and/or angle of sensors. This causes a sizable displacement between reference image and sensed image with no mating points. To overcome this limitation of block processing, overlapped block processing is applied to improve registration accuracy. Overlapping rate between reference and sensed image is determined heuristically. Figure 4 shows the overlapped blocks of entire image. More detailed block based analysis can be found in [11], [21].

Feature points are extracted by using overlapped block-based Harris corner detector. Then, features are matched by using nearest-neighborhood distance ratio (NNDR) [18] based SIFT descriptor for block matching. Next, outliers are removed by random sample consensus (RANSAC) [19]. Finally, transformation is estimated using B-spline and image is re-sampled using the bicubic interpolation. The B-Spline method proposed by Rueckert et al. [12] serves as an alternative to Demons [22]. This method requires a rough pre-alignment to bring the images together such that only local deformations are present. Then these local deformations can be modeled using B-Spline. While comparing with Demons [20], this method is more robust to noise and is not that much dependent on texture. In contrast to thin-plate splines [23], B-splines are locally controlled, which makes them computationally efficient even for a large number of correct matching points.

Figure 5 demonstrates the effectiveness of the coarse-to-fine approach. Figure 5 (a) shows the correct matching points (13 points) from SLO images using coarse step, while Fig. 5 (b) shows the correct matching points (73 points) from IR image after applying fine step. It means, that the fine step has increased the extracted and matched points. In addition, Fig. 5 (c) shows the correct matching points (20 points) using Harris detector without using block-wise coarse-to-fine approach. We can observe that the coarse step has provided smaller number of points than the Harris detector, however the fine step has significantly increased the number of correct matching points and are almost 3.5 times more than Harris detector. It can be concluded that the coarse-to-fine approach extracts more distinctive features and improves the registration accuracy.

III. RESULTS AND DISCUSSION

In this work, datasets for two different multi-modal images are used for conducting experiments and evaluating accuracy. Retinal Images (RI) consist of four CFP/SLO pairs [2], while remote sensing dataset consists of three EO/IR images. The performance of the proposed method is evaluated using precision metric along with number of matched points and number of correctly matched points. Precision metric computes the
FIGURE 6. Matching points comparison: (a) - (d) SIFT, (e) - (h) improved SIFT (PSO-SIFT), (i) - (l) Harris without block-wise coarse-to-fine approach, (m) - (p) proposed method; CFP (left side) and SLO (right side).
TABLE 2. Performance comparison of EO/IR pairs using original SIFT, improved SIFT (PSO-SIFT), and proposed block-wise coarse-to-fine approach (BCF).

| Items                  | Sample 1         | Sample 2         | Sample 3         | Sample 4         |
|------------------------|------------------|------------------|------------------|------------------|
|                        | SIFT | PSO-SIFT | Harris | BCF | SIFT | PSO-SIFT | Harris | BCF | SIFT | PSO-SIFT | Harris | BCF |
| Corresponding points   | 189  | 724     | 653    | 318 | 834  | 198   | 3.728  | 461  | 175  | 617     |
| Correct matching points| 2    | 0       | 251    | 114 | 374  | 7     | 0      | 223  | 65   | 305     |
| precision              | 0.0106 | 0.0000  | 0.3844 | 0.3585 | 0.4484 | 0.0354 | 0.0000  | 0.4837 | 0.3714 | 0.4943 |

Figure 7. Matching points (a) - (b) SIFT, (c) - (d) proposed method; (a), (c) EO, (b), (d) IR.

In addition, it can also be observed that PSO-SIFT has provided larger number of corresponding points, however, this method does not produce enough correct matching points and consequently providing low registration accuracy. We have repeated the same experiment by using a different dataset and similar conclusions are obtained. Table 2 compares the registration accuracy in terms of precision along with the number of extracted corresponding points and number of correct matching points for four pairs of EO/IR images. Again, it can be observed that, the number of correct matching points extracted in coarse step are less than the points extracted from non-block-wise coarse-to-fine approach. However, correct matched points are increased after applying the fine step. Again, PSO-SIFT is providing larger number of corresponding points, however, yielding lower accuracy due to lesser number of correct matching points. In conclusion, from the Table 1 and the Table 2, it is clear that the proposed method has provided higher precision values i.e. higher registration accuracy than the conventional approaches. In addition, it can also be observed that the improved accuracy is due to the fact that the proposed coarse-to-fine approach provides higher number of correctly matched points.

Figure 6 shows the examples for the number of correct matching points for the pairs of CFP/SLO images obtained through the original SIFT (Fig. 6 (a) to (d)), improved SIFT (PSO-SIFT) (Fig. 6 (e) to (h)), Harris detector without block-wise coarse-to-fine approach (Fig. 6 (i) to (l)), and the proposed coarse-to-fine approach (Fig. 6 (m) to (p)). From the figure, it can be observed that the proposed method has provided higher number of correct matching points as compared to the conventional methods which leads to the registration with higher accuracy. The same experiment is repeated with pairs of EO/IR images and results are shown in Fig. 7. Again, it can be observed from the Fig. 7 that the number of correct matching points obtained from the proposed coarse-to-fine approach are higher than that obtained through the conventional approaches SIFT and Harris detector.

number of true matching points relative to the total number of matched points. Its higher values represent higher accuracy.” More detailed description about the precision metric can be found in [18].

Table 1 compares the registration accuracy in terms of precision for pairs of CFP/SLO images. The number of extracted corresponding points, number of correct matching points and the precision values are computed by using SIFT, PSO-SIFT [24], Harris detector and the proposed block-wise coarse-to-fine (BCF) approach for all four samples from CFP/SLO pairs. It can be observed that, in coarse step, the number of correct matching points are less than the points extracted from non-block-wise coarse-to-fine approach. However, after applying the fine step, the extracted number of points are more than Harris approaches.
Figure 8 shows the pseudo color images (Red: registered SLO image, Green: original CFP image, Blue: original CFP image). The resultant images from the original SIFT, improved SIFT (PSO-SIFT), Harris detector and the proposed scheme are shown in Fig. 8 ((a) to (d)), Fig. 8 ((e) to (h)), Fig. 8 ((i) to (l)), and Fig. 8 ((m) to (p)), respectively. The original SIFT method is failed to register images in cases shown in Fig. 8 ((a), (c), (d)) due to fewer number of points as affine transform needs at least three pairs of points. The original SIFT method appears successful in case of shown in Fig. 8 (b), however its magnified image shown in Fig. 8 (q) depicts misaligned results. Similarly, misalignments are also observable in results obtained through Harris detector. The improved SIFT (PSO-SIFT) is also failed in all cases due to the limited number of correctly matched points. Whereas, it can be observed that the proposed coarse-to-fine scheme provided enough correctly matched points and registered the images with higher accuracy. Hence, visual results show the superiority of the proposed scheme over the conventional approaches. Similarly, Fig. 9 shows the
superimposed image (background: EO image, Yellow: registered IR image). Here, Canny edge of registered IR image is used for making superimposed image. The output image by using the proposed method is shown in Fig. 9 second column. From the figure, it can be observed that the proposed method has registered the images accurately.

IV. CONCLUSION

In this paper, a block-wise coarse-to-fine approach is proposed for multi-modal image registration. Due to the different characteristics of two sensors, original SIFT fails to register images accurately, whereas the proposed method is able to extract more distinctive features and provides accurate results. The comparative analysis with respect to precision measure and the visual outputs demonstrate the effectiveness of the proposed method. We believe that the proposed approach can be used in other multi-modal registration scenarios including CFP/optical coherence tomography (OCT), CFP/fundus fluorescein angiogram (FFA) images.

REFERENCES

[1] P. L. Rosin, D. Marshall, and J. E. Morgan, “Multimodal retinal imaging: New strategies for the detection of glaucoma,” in Proc. Int. Conf. Image Process., vol. 3, Sep. 2002, p. 3.
[2] P. A. Legg, P. L. Rosin, D. Marshall, and J. E. Morgan, “Feature neighbourhood mutual information for multi-modal image registration: An application to eye fundus imaging,” Pattern Recognit., vol. 48, no. 6, pp. 1937–1946, Jun. 2015.
[3] J. W. Berger, M. E. Leventon, N. Hata, W. Wells, and R. Kikinis, “Design considerations for a computer-vision-enabled ophthalmic augmented reality environment,” in Proc. CVRMed-MRCAS. Berlin, Germany: Springer, 1997, pp. 399–408.
[4] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, “A partial intensity invariant feature descriptor for multimodal retinal image registration,” IEEE Trans. Biomed. Eng., vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
[5] L. Chen, Y. Xiang, Y. Chen, and X. Zhang, “Retinal image registration using bifurcation structures,” in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 2169–2172.
[6] C. Huo, C. Fan, L. Huo, and Z. Zhou, “Multilevel SIFT matching for large-size VHR image registration,” IEEE Geosci. Remote Sens. Lett., vol. 9, no. 2, pp. 171–175, Mar. 2012.
[7] X. Chang, S. Du, Y. Li, and S. Fang, “A coarse-to-fine geometric scale-invariant feature transform for large size high resolution satellite image registration,” Sensors, vol. 18, no. 5, p. 1360, 2018.
[8] M. I. Patel, V. K. Thakar, and S. K. Shah, “Image registration of satellite images with varying illumination level using HOG descriptor based SURF,” Procedia Comput. Sci., vol. 93, pp. 382–388, 2016.
[9] B. Zitová and J. Flusser, “Image registration methods: A survey,” Image Vis. Comput., vol. 21, no. 11, pp. 977–1000, Oct. 2003.
[10] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable, “Retinal image analysis: Concepts, applications and potential,” Prog. Retinal Eye Res., vol. 25, no. 1, pp. 99–127, 2006.
[11] I. Lee, D.-C. Seo, and T.-S. Choi, “Entropy-based block processing for satellite image registration.” Entropy, vol. 14, no. 12, pp. 2397–2407, 2012.
[12] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: Application to breast MR images,” IEEE Trans. Med. Imag., vol. 18, no. 8, pp. 712–721, Aug. 1999.
[13] R. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.
[14] A. Laine and J. Fan, “Texture classification by wavelet packet signatures,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 11, pp. 1186–1191, Nov. 1993.
[15] R. X. Gao and Y. Yan, Wavelets: Theory and Applications for Manufacturing. New York, NY, USA: Springer, 2010.
[16] C. Harris and M. Stephens, “A combined corner and edge detector,” in Proc. Proceedings Alvey Vis. Conf., Manchester, U.K., vol. 15, 1988, p. 50.
[17] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.
[18] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[19] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol. 24, no. 6, pp. 385–395, 1981.
MUHAMMAD TARIQ MAHMOOD (Senior Member, IEEE) received the M.C.S. degree in computer science from The University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan, in 2004, the M.S. degree in intelligent software systems from the Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree from the School of Information and Mechatronics, Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2011. He is currently an Associate Professor with the School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan, South Korea. His research interests include image processing, 3-D shape recovery from image focus, computer vision, pattern recognition, and machine learning.

IK HYUN LEE (Member, IEEE) received the B.S. degree in control and instrument engineering from Korea University, South Korea, in 2004, and the M.S. and Ph.D. degrees from the School of Information and Mechatronics, Gwangju Institute of Science and Technology, South Korea, in 2008 and 2013, respectively. He was a Postdoctoral Researcher with Media Lab, Massachusetts Institute of Technology, and a Senior Researcher with the Korea Aerospace Institute of Research. He is currently an Assistant Professor with the Department of Mechatronics Engineering, Korea Polytechnic University, South Korea. His research interests include image registration, image fusion, depth estimation, and medical image processing.

* * *

VOLUME 8, 2020 78053