DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244

Register Multimodal Images of Large Scene Depth Variation with Global Information

Hongbin Jin*[1] *, *Yong Li*[1] , *Chunxiao Fan*[1] , *Robert Stevenson*[2] ; *1. Beijing Key Laboratory of Work Safety and Intelligent Monitoring, School of Electronic Engineering, Beijing University of Posts and Teles., Beijing, China, 2. Dept. of Electrical Engineering, Univ. of Notre Dame, Indiana, USA.

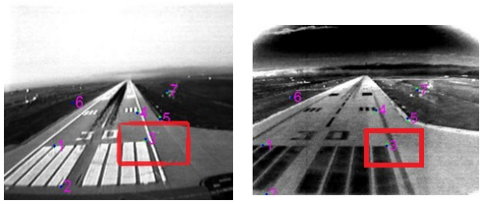
Abstract

This paper addresses the problem of registering multimodal images of scene depth variation. The existing methods typically build matches of keypoints with descriptors and then apply consensus/consistency check to rule out incorrect matches. However, the consistency check often fails to work when there are a large number of wrong matches. Given a set of initial matches built with descriptors, we seek to search the best or correctly matched keypoints. To this end, this work employs the global information over entire images to assess the quality of keypoint matches. Since the image content has depth variation, projection transformations are needed to account for the misalignment and hence quadruples of keypoint matches are considered. In order to search the correctly matched keypoints, an iterative process is used that considers all preserved quadruples passing the spatial coherence constraint. Extensive experimental results on various image data show that the proposed method outperforms the state-of-the-art methods.

Introduction

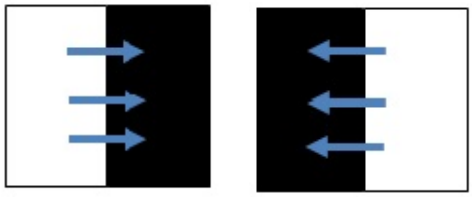
The goal of registering multimodal images containing largely varying scene depth is to align images acquired by different sensors and/or from different views. The challenge comes from two-fold aspects, one is the scene depth, and the other is the multimodality. As to the scene depth, if images contain no depth variation, the misalignment can be exactly represented by an affine or projective transformation. When scene depth varies largely, the geometric transformation of two images can not be exactly formulated to be a linear model, e.g., affine or projective. One feasible approach is to employ projective transformations to approximately account for the misalignment. The second challenge is about the multimodality that significantly decreases the repeatability and distinctiveness of descriptors. SIFT [1] and SURF [2] have been designed to build keypoint mappings between two images based on the assumption that corresponding keypoints have similar gradient pattern around them. Alahi *et al.* [3] propose fast retina keypoint(FREAK) which is faster to compute and more robust than SIFT and SURF. Yu [4] propose ASIFT, which is fully affine invariant. It simulates all image views by varying the two camera axis parameters and covers the other four parameters by applying SIFT. Cai *et al.* [5] then further investigate a perspective scale invariant feature transform (PSIFT) by using homographic transformation to simulate perspective distortion. To increase the number of keypoints, Park *et al.* [6] proposed using higher-order scale space derivatives, $\partial^2 L(x,y,\sigma)/\partial\sigma^2$, $\partial^3 L(x,y,\sigma)/\partial\sigma^3$, $\partial^4 L(x,y,\sigma)/\partial\sigma^4$, and then extracted the extrema in the high-order scale space. Since gradient information in multimodal images will change[7], it is very difficult for these methods to achieve highly accurate registration on multimodal images[8].

The gradient changes are caused by the non linear response of the scene contents to the wavelength channels used in the mutlimodal imaging[9]. That is to say, the same scene contents appear differently within channel images and result in different orientation of gradient at the correspondence keypoint between mutlimodal images as illustrated in Fig. 1 for a visible and an infrared image regions.



(a) An example of a visible image

(b) An example of an infrared image



(c) The orientation of gradient in red pane on the visible image

(d) The orientation of gradient in red pane on the infrared image

Figure 1. *(a) shows a visible image region of airstrip and keypoints calculated by SIFT, (b) shows the corresponding IR channel image region and SIFT keypoints, (c) shows the orientation of gradient at zebra crossing of the visible airstrip, whereas in (d) the orientation of the corresponding region is reversed which makes the SIFT keypoints unable to be matched.*

In order to enable the descriptors designed for single-mode images (e.g., SIFT) to mutlimodal images, many algorithms have been proposed, including NG_SIFT (NG, normalized gradient) [7], SAR_SIFT [10] and MIND [11]. NG_SIFT utilizes the normalized gradients around keypoints for describing the local pattern to achieve the invariance against non-linear intensity changes between mutlimodal images. It outperforms the original SIFT on the mutlimodal images of a structured scene. SAR_SIFT proposes a new local gradient pattern around keypoints, in which the orientation and magnitude are robust against the speckle noise. SAR_SIFT gives a higher ratio of correct keypoint mappings than the original SIFT on mutlimodal images. Mainal-

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244

i *et al.* [12] proposed the D-SIFER scale-invariant feature detection algorithm using the 10th order scale-space optimal Gaussian derivative filter. D-SIFER was validated on hyperspectral images and was shown to perform better than SIFT and SUR-F. Hossain *et al.* [13] propose improved symmetric-SIFT (ISS) to address the gradient reversal, region reversal, and the descriptor merging problem. Chen *et al.*[14] propose partial intensity invariant feature descriptor(PIIFD) for multimodal retinal image registration. PIIFD assigns the main orientation of keypoints to a number range $[0,\pi)$ and extracts feature descriptors characterizing the outlines (contours) around keypoints. They perform better than registration methods for mono-modal images. However, when the ratio of initial correct keypoint mappings is small, which often occurs on multimodal image pairs, the correct mappings cannot be effectively obtained by utilizing random sample consensus (RANSAC)[15] or other outlier removing techniques. To establish more correct matched keypoints robustly, a projective transformation is used to model the misalignment between multimodal images of depth information in this letter. We utilize the number of overlapped edge pixels over the whole images as similarity metric, incorporating global information in the evaluation of keypoint mappings. Quadruples of keypoint mappings are chosen with the spatial constraints and evaluated with similarity metric by using global information. An iterative updating process is designed to find the best matched reference keypoint for each test keypoint. The main contribution of this letter is that we exploit global information to build keypoint mappings in conjunction with local descriptors. Keypoint mappings are evaluated not only with local information descriptors around keypoints but with whether the keypoint mappings can bring entire images into alignment (i.e., global information). The incorporation of global information ensures that the built keypoint mappings are consistent well with the content of entire images in the sense that these keypoint mappings can bring two multimodal images into alignment.

Similarity metric

The number of overlapped pixels is defined as the similarity metric. Edges are extracted by the Canny [16] operator in which the high and low thresholds are set based on the image content [17] as Fig. 2 shows. In this paper, we use the same steps as the original Canny operator to detect edges on test and reference image. However, the high and low thresholds are set locally, i.e., they are determined within a moving window centered on the current pixel rather than over the entire image. The window size is set to be 20*20 like the configuration parameters in Ref. There are 64 bins being used to calculate the histogram of the window so that the high thresold is set to the gradient magnitude that is ranked top 30% in the window and the low thresold is set to 40% of high thresold.

Let $I_r(x,y)$ and $I_t(x,y)$ denote the reference and test multimodal images to be registered, $I_t^T(x,y)$ denote the transformed version of $I_t(x,y)$ by a projective transformation T. Then the similarity metric is defined as

$$N_{op}\left(I_r(x,y),I_t^T(x,y)\right)=\sum_{x,y}E_r(x,y)\cdot E_t^T(x,y), \tag{1}$$

where $E_r(x,y)$ and $E_t^T(x,y)$ are the edge maps of $I_r(x,y)$ and $I_t^T(x,y)$, respectively. As Fig. 2 shows that extracted edge pixels are evenly distributed on the entire image due to locally setting the high and low threshold.

(a) Original image (b) Image edge

Figure 2. *An image and its edge map detected with the Canny operator.*

Distance constraints for keypoint correspondences

We use the following projective transformation for registering images,

$$u_i=\frac{a_1\cdot x_i+a_2\cdot y_i+a_3}{a_7\cdot x_i+a_8\cdot y_i+1},\quad v_i=\frac{a_4\cdot x_i+a_5\cdot y_i+a_6}{a_7\cdot x_i+a_8\cdot y_i+1}, \tag{2}$$

where (x_i,y_i), $i=1,2,3,4$, denote the coordinates of four keypoints $\left(P_t^{i_1},P_t^{i_2},P_t^{i_3},P_t^{i_4}\right)$ on test image and (u_i,v_i), $i=1,2,3,4$, denote the coordinates of four keypoints $\left(P_r^{k_{i_1}},P_r^{k_{i_2}},P_r^{k_{i_3}},P_r^{k_{i_4}}\right)$ on reference image which are matched to $\left(P_t^{i_1},P_t^{i_2},P_t^{i_3},P_t^{i_4}\right)$, a_i, $i=1,2,3,4,5,6,7,8$ denote the coefficients of the projective transformation determined by the four pairs of keypoints. To apply projective transformations for registering images, the four keypoints in (2) are required to be located in a plane. For this purpose, a distance constraint is enforced on them since the scene depth varies over pixel locations. Intuitively, the smaller the distance between keypoints in the quadruple, the more likely it will lie in a plane. Motivated by the idea in [18], we limit the distance of keypoints in a quadruple with the following equation,

$$\|P_t^i-P_t^j\|_2\le\alpha\cdot\max(H,W),\quad \|P_r^i-P_r^j\|_2\le\alpha\cdot\max(H,W), \tag{3}$$

where P_t^i and P_t^j are any two points in the test quadruple, P_r^i and P_r^j are any two points in the reference quadruple,H and W are the height and width of the test and reference image, α is a coefficient. (3) ensure that the area enclosed is more likely a plane scene in test(reference) image. In this letter, the limited coefficient α is set to $\frac{1}{4}\sqrt{2}$, the $\frac{1}{4}$ times the diagonal length of an image.

Search best keypoint mappings with global information

In order to search best keypoint mappings, we attempt to find the test keypoints which have the maximum similarity metric N_{op}. This searching process comprises the following four steps. Firstly, we extract the keypoints from $I_r(x,y)$ and $I_t(x,y)$ by applying SURF. Let K_t^i, $i=1,2,...,N_t$, denote the i_{th} keypoint on $I_t(x,y)$, and K_r^j,$i=1,2,...,N_r$, denote the j_{th} keypoint on $I_r(x,y)$. Let f_t^i, $i=1,2,...,N_t$, denote the descriptor of K_t^i and let f_r^j, $j=1,2,...,N_r$, denote the descriptor of K_r^j. In the original SURF method, a reference keypoint $K_r^{j_0}$ is matched to be a test keypoint $K_t^{i_0}$, if

$$d(f_t^{i_0},f_r^{j_0})<0.8\cdot d(f_t^{i_0},f_r^{j_1}),$$

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244

where d() is the Euclidean distance and $f_r^{j_1}$ is the 2nd-closest neighbor to $f_t^{i_0}$. The parameter 0.8 is the default value of SUFR method which can be replaced by a smaller value in practice. The smaller value means a tighter matching criterion giving fewer matched keypoints. However, because of the gradient reversal and region reversal as illustrated in Fig. 1 , the repeatability and distinctiveness decrease significantly on mutlimodal images[19], and hence the initial keypoint mappings built by SURF include a high ratio of incorrect ones. The result of initially built keypoint mappings is used in next steps for searching the best matched reference keypoint for every test keypoint.

Secondly, since the SURF descriptor often yields incorrect mappings for multimodal images, we assign multiple putative mapping reference keypoints for each test keypoint to improve the probability of yielding correct mappings. The putative mapping reference keypoints of K_t^i are obtained by sorting $d(f_t^i, f_r^j)$ and finding N_c reference keypoints of the smallest distances to f_t^i. In this letter, N_c is set to 3.

Thirdly, an iterative process is designed that considers all preserved quadruples passing the spatial coherence constraint as indicated in (3) to search best keypoint mappings. In the iterative process, consider a quadruple of the test keypoints $\left(K_t^{i_1}, K_t^{i_2}, K_t^{i_3}, K_t^{i_4}\right)$, $i_1 < i_2 < i_3 < i_4$, for each test keypoint of the quadruple pick a candidate reference keypoint. Then we get a quadruple of reference keypoints $\left(K_r^{k_{i_1}}, K_r^{k_{i_2}}, K_r^{k_{i_3}}, K_r^{k_{i_4}}\right)$, where $K_r^{k_{i_1}}$ is one of the putative mapping reference keypoints to $K_t^{i_1}$, and similar meaning applies to $K_r^{k_{i_2}}$, $K_r^{k_{i_3}}$, and $K_r^{k_{i_4}}$. Next, the quadruple of mappings $\left(K_t^{i_1}, K_t^{i_2}, K_t^{i_3}, K_t^{i_4}\right) \sim \left(K_r^{k_{i_1}}, K_r^{k_{i_2}}, K_r^{k_{i_3}}, K_r^{k_{i_4}}\right)$ is used to determine a projection transformation T with (2), and then the similarity metric N_{op} is calculated with (1). The iterative process considers all quadruples of keypoint mappings and stores the maximum N_{op} for each test keypoint K_t^i in a vector $N_{op}[i]$. As a result, the best mapped reference keypoint for every test keypoint can be achieved by sorting through the value of $N_{op}[i]$. Note, $N_{op}[i]$ is different from one to another. This step is summarized in Algorithm 1.

Finally, of all test keypoints those whose maximum N_{op} ranked top 10% are selected for computing the final projective transformation.

Algorithm 1: Iteratively processing pairs of keypoint mappings

input : $I_r(x,y)$, $I_t(x,y)$.
output: Keypoint mappings whose maximum N_{op} ranked top 10%.

1 **Extract image features**:
- Detect keypoints K_r^i and descriptors $f_r^i, i \in [1, N_r]$, from $I_r(x,y)$, and K_t^i and $f_t^i, i \in [1, N_t]$, from $I_t(x,y)$.
- Generate edge maps $E_r(x,y)$ and $E_t(x,y)$ from $I_r(x,y)$ and $I_t(x,y)$.

Iteratively searching out best pairs of keypoint mappings:
for $i_1, i_2, i_3, i_4 \in [1, N_t]$ **do**

1. Require $i_1 < i_2 < i_3 < i_4$.
2. Find the matched reference keypoint to $K_t^{i_1}$, $K_r^{k_{i_1}}$, the matched reference keypoint to $K_t^{i_2}$, $K_r^{k_{i_2}}$, the matched reference keypoint to $K_t^{i_3}$, $K_r^{k_{i_3}}$, the matched reference keypoint to $K_t^{i_4}$, $K_r^{k_{i_4}}$.
3. Require $k_{i_1} \neq k_{i_2} \neq k_{i_3} \neq k_{i_4}$.
4. Require any two points from $(P_t^{i_1}, P_t^{i_2}, P_t^{i_3}, P_t^{i_4})$, satisfying the geometrical constraint in Equation (3).
5. Require any two points from $(P_r^{k_{i_1}}, P_r^{k_{i_2}}, P_r^{k_{i_3}}, P_r^{k_{i_4}})$, satisfying the geometrical constraint in Equation (3).
6. Determine T between $(P_t^{i_1}, P_t^{i_2}, P_t^{i_3}, P_t^{i_4})$ and $(P_r^{k_{i_1}}, P_r^{k_{i_2}}, P_r^{k_{i_3}}, P_r^{k_{i_4}})$ by equation (2).
7. Transform edge points of $I_t(x,y)$ by the determined T.
8. Compute similarity metric $N_{op}\left(I_r(x,y), I_t^T(x,y)\right)$ by equation (1).
9. Updates the maximum N_{op} for each test keypoint K_t^i in a vector $N_{op}[i]$.

end

Remove outlier with ransac

At the last step, we apply the RANSAC algorithm because there are still some outlier keypoint mappings after searching best keypoint mappings with global information, although most of incorrect keypoint mappings are expected to have been removed. Random sample consensus (RANSAC) is an iterative approach to estimating the parameters of a mathematical model from a set of observed data containing outliers. RANSAC performs well in removing outliers of keypoint mappings if the correct ratio is high. However, the performance of RANSAC decreases dramatically especially when the correct ratio is low, e.g., 20% or less. Due to this, not all the keypoint mappings built global information are used as input of this step, rather, only the keypoint matches who have high similarity metric are fed into this step since these keypoint matches have a greater probability of being correct. Affine or projective transformations are utilized with RANSAC to remove outliers. When the distance of real scene content to the camera is all the same, an affine transformation would be enough to account for the misalignment. When the distance varies from point to point, a projective transformation or polynomial transformation is necessitated. Polynomial transformations require at least 6 keypoint mappings, which significantly increases the possibility that a sample composed of 6 mappings contains incorrect ones. Consequently, projective transformations are utilized to address images of scene depth, and the proposed method can build correct mappings.

Experimental results

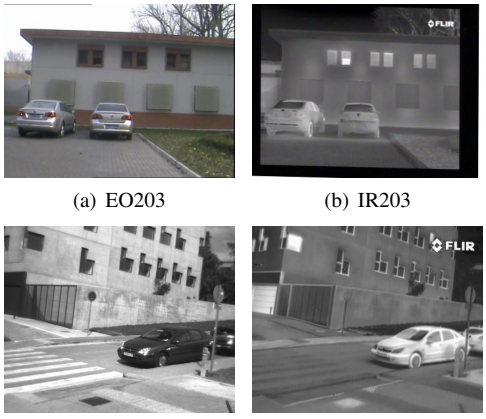
We evaluate the performance of the proposed algorithm using three datasets of mutlimodal images. Dataset EOIR includes 87 image pairs covering outdoor depth scenes captured by ourselves, 12 Landsat image pairs from NASA, four remote sens-

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244

ing image pairs of the 2008 Sichuan earthquake and two image pairs from the OSU Color and Thermal Database. The 87 image pairs cover outdoor depth scenes, with one image taken with visible light and the other taken with middle-wave infrared (MWIR) light. In addition to the spectral distance, they are taken at different times, so the content of one image may be slightly different from that of the other. The 12 Landsat image pairs are downloaded from http://landsat.usgs.gov/ with one taken with the visible band, e.g., Landsat 8 Band 3 Visible (0.53–0.59 μm), and the other taken with middle-wave light or the Thermal Infrared Sensor (TIRS), e.g., Landsat 8 Band 10 TIRS 1 (10.6–11.19 μm). The four remote sensing image pairs were taken over Wenchuan county (Sichuan Province, China) during the 2008 Sichuan earthquake. They were acquired by the Formosat-2 satellite. One image is a mutlimodal image (1960×1683) before the earthquake, and the other is a panchromatic image (1968×1705) after the earthquake of the same area. In order to further verify the performance of the proposed method for mutlimodal images taken at different times, we take two image pairs from the OSU Color and Thermal Database (Data 03, http://www.vcipl.okstate.edu/otcbvs/bench/). The two image pairs are captured by a thermal sensor (Raytheon PalmIR 250D, 25-mm lens) and a color sensor (Sony TRV87 Handycam). Dataset RWHI includes Real-World Hyperspectral Image from [9] containing 50 scenes. Most of scenes have large variations in depth. The images of RWHI were acquired by sequentially tuning a filter through a series of thirty-one narrow wavelength bands, each with approximately 10nm bandwidth and centered at steps of 10nm from 420nm to 720nm. We use 50 image pairs of 420nm and 560nm to test and the dimensions of the images are 464×346 pixels. Dataset VS-LWIR is from [20] containing 100 image pairs, one image taken with the visible bandwidth (0.4–0.7 μm) and the other taken with the long-wave infrared bandwidth (LWIR, 8–14 μm). Fig. 3 gives some examples from dataset 1 and dataset 3.

We compared the registration results of the proposed method with SIFT+RANSAC[1], FREAK[3], ISS[13], and PIIFD[14]. Firstly visual matching results are shown, followed by the quantitative analysis on matching results. These methods are implemented with OpenCV.

Fig. 4 shows the keypoint matching results given by proposed method, SIFT+RANSAC, PIIFD, ISS, and FREAK on an image pair from dataset EOIR. Due to the variation of gradient pattern between multimodal image pairs, the SIFT+RANSAC and FREAK method can not build correct keypoint mappings. There are 5 pairs of matching points built by ISS in fig. 4(c) and 4 pairs built by PIIFD in fig. 4(d), none of which is correct. Compared with SIFT+RANSAC, PIIFD, ISS, and FREAK, there are 3 correct pairs of matching points built in Fig. 4(e). The reason is that the proposed method utilizes the global information over entire images to assess the quality of keypoint matches and an updating process is applied to find the best matched keypoint for every test keypoint. In Fig. 5 the visual results for keypoint matching on a pair of images on dataset VS-LWIR are shown. Because of large spectral difference between the two images in a pair of images on dataset VS-LWIR, it is more difficult to build reliable keypoint matching in such images. The visual results in Fig. 5 indicated that PIIFD and ISS can hardly build one correct keypoint matching, while the proposed method can build four correct keypoint matchings in Fig. 5(e).



(a) EO203 (b) IR203

(c) VS0081 (d) IR0081

Figure 3. *Two image pairs from dataset EOIR and Dataset VS-LWIR. (a), (c) are outdoor scenes captured by visible band sensor and (b), (d) are taken by infrared sensor respectively.*

To quantitatively evaluate the performance of proposed method, the histogram of the distance between mapped keypoints is employed. The bins are set to [0, 2], [2, 5], [5, 10], [10, 20], and [20,∞). Table 1 gives the histogram of distances between mapped keypoints for different methods. On dataset EOIR and dataset VS-LWIR, the proposed method significantly performs better than SIFT+RANSAC, PIIFD, ISS, and FREAK. For example, on dataset EOIR, the proposed method has 23% of keypoint mappings with a distance falling in $[0,2]$ and 24% falling in $[2,5]$, while SIFT + RANSAC has 9% in $[0,2]$ and 7% in $[2,5]$, and ISS only has 9% and 4% falling in $[0,2]$ and $[2,5]$. On dataset VS-LWIR, the proposed method has 76 pairs of keypoint mappings with a distance falling in $[0,2]$ and 117 pairs falling in $[2,5]$, while PIIFD has 2 pairs in $[0,2]$ and 5 pairsin $[2,5]$, and Freak only has 2 and 3 pairs falling in $[0,2]$ and $[2,5]$. An interesting result shown in table 1 is that all of the methods could achieve good performance on dataset RWHI. The reason maybe that the spectral difference between the two images of a pair on dataset RWHI is small, so that the gradient changes not so much. In conclusion, the results show that there are two advantages of the presented method over other methods. The first is that the proposed method provides a higher ratio of keypoint mappings. The second is that the presented method provides a lower ratio of keypoint mappings that have a distance greater than 20.

Conclusion and future work

This paper proposes a method to register multimodal images of depth information. Quadruples of keypoint mappings are chosen with the spatial constraints and evaluated with similarity metric by using global information. An iterative updating process is designed to find the best matched reference keypoint for each test keypoint. Experimental results show that the presented method can provide more reliable keypoint mappings and achieve a better matching performance than the state-of-the-art.

There are several future research directions that can be done to improve the capability of the proposed method. The first is to design reliable descriptors based on the common information between mutlimodal images. Although it is not the focus of this

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244

Table 1: the distribution of the distances between matched keypoints

Dataset	method	[0-2]	[2-5]	[5-10]	[10-20]	>20
EOIR	Proposed	143	147	105	173	64
	SIFT+RANSAC	75	54	263	205	231
	PIIFD	14	32	40	16	167
	ISS	31	16	5	12	295
	FREAK	11	11	5	10	358
RWHI	Proposed	602	1	2	5	1
	SIFT+RANSAC	597	2	1	3	0
	PIIFD	1093	16	5	12	0
	ISS	1082	18	3	7	0
	FREAK	863	2	4	11	0
VS-LWIR	Proposed	76	117	108	105	20
	SIFT+RANSAC	15	29	33	105	495
	PIIFD	2	5	6	11	198
	ISS	2	16	8	7	106
	FREAK	2	3	4	6	489

paper, enhancing the matching ability of descriptors can improve the overall registration accuracy. The second is on the similarity metric that has been used for evaluating the quality of keypoint mappings. It uses the global information of entire image consuming large computational cost, which needs to be optimized in the future.

Acknowlegement

This work was supported by the National Natural Science Foundation of China (Grants No., NSFC-61170176, NSFC-61471067), Fund for the Doctoral Program of Higher Education of China (Grants No., 20120005110002), Fund for Beijing University of Posts and Telecommunications (Grants No., 2013XD-04, 2015XD-02), Fund for National Great Science Specific Project (Grants No. 2014ZX03002002-004), Fund for Beijing Municipal Administration of Hospitals Clinical medicine Development of special funding support (code: XMLX201406).

References

[1] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), 91–110 (2004).

[2] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V., "Speeded up robust features (surf)," *Computer Vision and Image Understanding* **110**(3), 346–359 (2008).

[3] Alahi, A., Ortiz, R., and Vandergheynst, P., "Freak: Fast retina keypoint," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], 510–517 (2012).

[4] Morel, J.M.; Yu, G. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Imag. Sci.* **2009**, *2*, 438–469.

[5] Cai, G.R.; Jodoin, P.M.; Li, S.Z.; Wu, Y.D.; Song-Zhi.; Su.; Huang, Z.K. Perspective-SIFT: An Efficient Tool for Low-Altitude Remote Sensing Image Registration. *Signal Process.* **2013**, *93*, 3088–3110.

[6] Park, U.; Park, J.; Jain, A.K. Robust Keypoint Detection Using Higher-Order Scale Space Derivatives: Application to Image Retrieval. *IEEE Signal Process. Lett.* **2014**, *21*, 962–965.

[7] Saleem, S.; Sablatnig, R. A Robust SIFT Descriptor for mutlimodal Images. *IEEE Signal Process. Lett.* **2014**, *21*, 400–403.

[8] M. Vural, Y. Yardimci, and A. Temizel, Registration of mutlimodal satellite images with orientation-restricted SIFT. *IEEE Int. Geoscience and Remote Sensing Symp.* **2009**, *3*, 243–246.

[9] Chakrabarti, A. and Zickler, T., "Statistics of real-world hyperspectral images," in [*IEEE Conference on Computer Vision and Pattern Recognition*], 193 – 200 (2011).

[10] Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 453–466.

[11] Heinrich, M.P.; Jenkinson, M.; Bhushan, M.; Matin, T.; Gleeson, F.V.; Brady, S.M.; Schnabel, J.A. MIND: Modality Independent Neighbourhood Descriptor for Multi-modal Deformable Registration. *Med. Image Anal.* **2012**, *16*, 1423–1435.

[12] Mainali, P.; Lafruit, G.; Tack, K.; Gool, L.V.; Lauwereins, R. Derivative-Based Scale Invariant Image Feature Detector with Error Resilience. *IEEE Trans. Image Process.* **2014**, *23*, 2380–2391.

[13] Hossain, M. T., Lv, G., Teng, S. W., Lu, G., and Lackmann, M., "Improved symmetric-sift for multi-modal image registration," in [*International Conference on Digital Image Computing: Techniques and Applications (DICTA)*], 197–202 (2011).

[14] Chen, J.; Tian, J.; Lee, N.; Zheng, J.; Smith, R.T.; Laine, A.F. A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1707–1718.

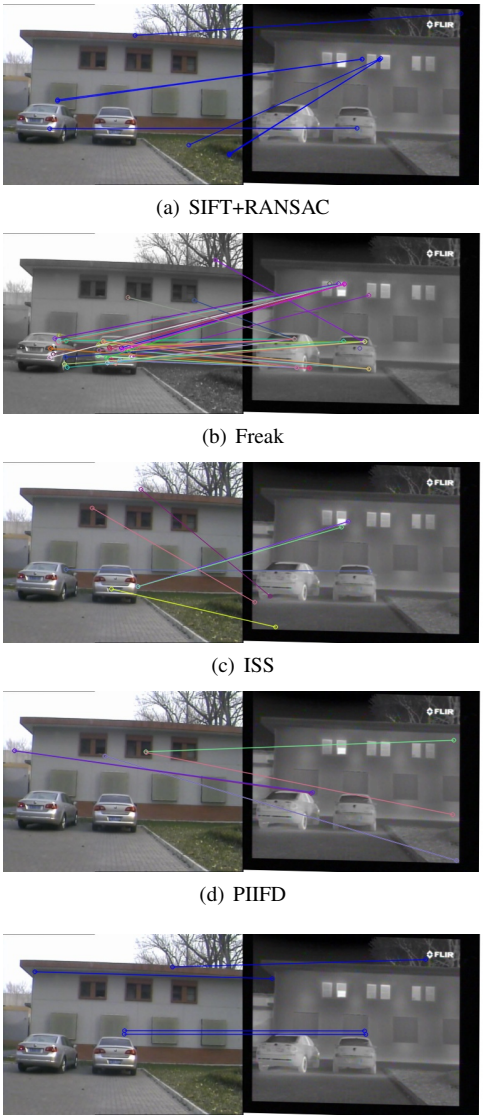
[15] Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395.

[16] Canny, J., "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698 (Nov. 1986).

[17] Simonson, K. M., Jr., S. M. D., and Tanner, F. R., "A statistics-based approach to binary image registration with uncertainty analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 112–125 (Jan. 2007).

[18] Evangelidis, G. D. and Bauckhage, C., "Efficient subframe video alignment using short descriptors," *IEEE Trans. Pattern Anal. Mach.*

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-244



(a) SIFT+RANSAC

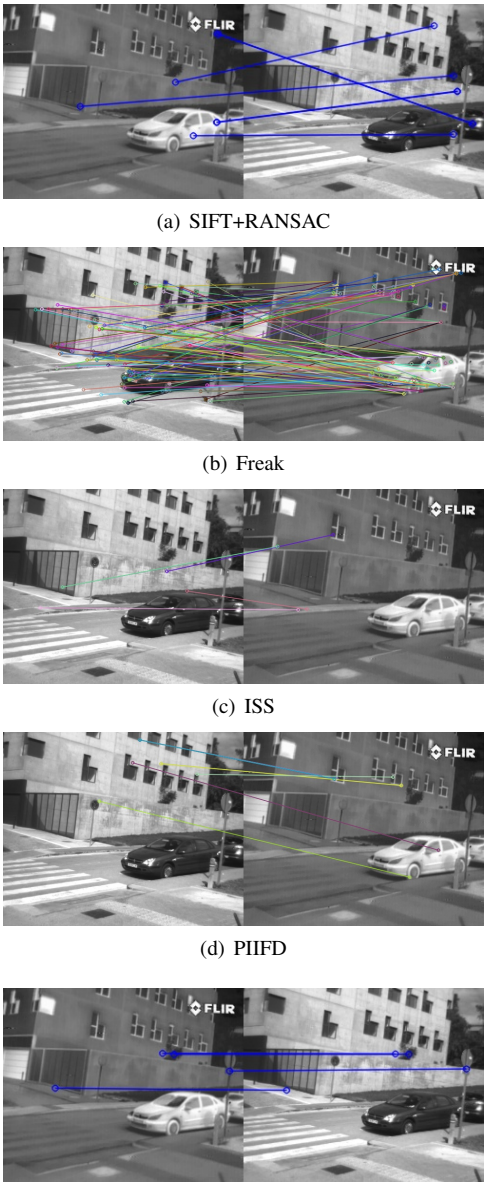
(b) Freak

(c) ISS

(d) PIIFD

(e) Proposed

Figure 4. *Keypoint mapping results of one pair images on dataset EOIR by different methods. (a) shows the result by SIFT+RANSAC, (b) shows the result by FREAK, (c) shows the result by ISS, (d) shows the result by PIIFD and (e) shows the result by the proposed method.*



(a) SIFT+RANSAC

(b) Freak

(c) ISS

(d) PIIFD

(e) Proposed

Figure 5. *Keypoint mapping results of one pair images on dataset VS-LWIR by different methods. (a) shows the result by SIFT+RANSAC, (b) shows the result by FREAK, (c) shows the result by ISS, (d) shows the result by PIIFD and (e) shows the result by the proposed method.*

Intell. **35**, 2371–2386 (Oct. 2013).

[19] MT Hossain, SW Teng, G Lu, "Achieving High Multi-Modal Registration Performance Using Simplified Hough-Transform with Improved Symmetric-SIFT," in [*International Conference on Digital Image Computing: Techniques and Applications (DICTA)*], 1 – 7 (2012).

[20] Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. mutlimodal image feature points. *Sensors* **2012**, *12*, 12661–12672.

Author Biography

Hongbin Jin is a Ph.D candidate of Beijing University of Posts and Telecommunications. His research is focused on computer vision and image processing.

Yong Li received his Master of Science in applied mathematics with Prof. Gerald Misiolek, and his PhD with Prof. Robert L. Stevenson, both from the University of Notre Dame. His research is focused on computer vision, image processing, and differential geometry.

Chunxiao Fan is currently a professor at the school of electronic engineering, Beijing University of Posts and Teles. Her research is focused on computer vision and big data.

Robert Stevenson received PhD in electrical engineering from Purdue University, West Lafayette, Indiana, in 1990. He joined the Faculty of the Department of Electrical Engineering at the University of Notre Dame, Indiana, in 1990, where he is currently a professor.