MINDFLOW BASED DENSE MATCHING BETWEEN TIR AND RGB IMAGES

J. Zhu[1], Z. Ye[1,2],[*] Y. Xu[1], L. Hoegner[1], U. Stilla[1]

[1] Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80290 Munich, Germany -
(jingwei.zhu, z.ye, yusheng.xu, ludwig.hoegner, stilla)@tum.de
[2] College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China

Commission II, ICWG II/III

KEY WORDS: TIR image, optical image, image registration, dense matching, MINDflow

ABSTRACT:

Image registration is a fundamental issue in photogrammetry and remote sensing, which targets to find the alignment between different images. Recently, registration of images from difference sensors become the hot topic. The registered images from different sensors are able to offer additional information, which help with different tasks like segmentation, classification, and even emergency analysis. In this paper, we proposed a registration strategy to calculate the dominant orientation difference and then achieve the dense alignment of Thermal Infrared (TIR) image and RGB image with MINDflow. Firstly, the orientation difference of TIR images and RGB images is calculated by finding the dominant image orientations based on phase congruency. Then, the modality independent neighborhood descriptor (MIND) together with global optical flow algorithm are adopted as MINDflow for dense matching. Our method is tested in the image sets containing TIR images and RGB images captured separately but in the same construction site areas. The results show that it is able to achieve the optimal results with features of significance even for dramatically radiometric differences between TIR images and RGB images. By comparing the results with other descriptor, our method is more robust and keep the features of objects in the images.

1. INTRODUCTION

Image registration is a fundamental issue in photogrammetry and remote sensing, which targets to find the alignment between different images. Since images acquired by different sensors or platforms provide information on various properties of objects in the scene, combining information from different platforms and sensors will provide a new perspective of information delivery for better visualization and analysis of scenes and objects (Tong et al., 2019, Ye et al., 2020). Thermal infrared (TIR) images, acquired by thermographic sensors, depict temperature and emission properties of objects. Different from camera in the visible spectrum, thermographic sensors detect radiation in a long-infrared range of the electromagnetic spectrum. Compared to the near infrared images, thermal infrared images have lower resolutions and different features due to the long wavelength. Since all the objects with a temperature above absolute zeros emit infrared radiation to the environment, thermal infrared images enable us to observe the geometry of objects, their moving process, inside structures, and their thermal properties without sufficient visible illumination (Zin et al., 2007, Weinmann et al., 2014, Christiansen et al., 2014). This property helps with thermal inspection of objects and recognition of hidden structures. Though thermal images provide thermal indicators for analysis, it is hard to achieve results of high accuracy for tasks like classification, recognition and abnormal phenomenon analysis due to low spatial resolution and strong distortion compared to RGB cameras. Besides, prior knowledge of the study area and the thermal images are required for analysis. As supplementary, RGB images are typical data sources in the field of photogrammetry, which provide accurate and detailed geometry and texture information of the scene. Considering

the difficulties in analyzing thermal images, combining RGB images and thermal images could improve the geometry and texture information to compensate the disadvantages of thermal images.

Generally, the challenges in the registration of thermal images (Fig. 1a) and RGB images (Fig. 1b) from different datasets lie in several aspects, including different geometric properties caused by various image resolutions, different intensity information resulting from various radiometric characteristics, and differences of sensor pose. Firstly, the thermal infrared images usually have lower spatial resolution compared to RGB images. The different in resolution cause different features in RGB and TIR images, which makes it hard to find the correspondences. Besides, due to different wavelengths, thermal infrared images and RGB images have different intensity information. The radiation in RGB images depicts the texture information of the objects, which is similar to the visible information observed by human eyes. The thermal images reflect the temperature or emission properties of the objects. Though a similar shape could be observed, the radiation information dramatically varied. Besides, the RGB images and TIR images may acqired seperately, or mounted on the same plateform with different viewing angles. The different extrior parameters lead to difficulties in finding corresponding points.

Considering the issues in the registration of thermal images and RGB images, in this study, we propose a robust method to deal with the dense matching of TIR and RGB images with large radiometric differences. Besides, our method can be applied for asynchronous devices with images that have small overlapping areas and orientation differences.

The structure of this paper is as follows. The introduction part
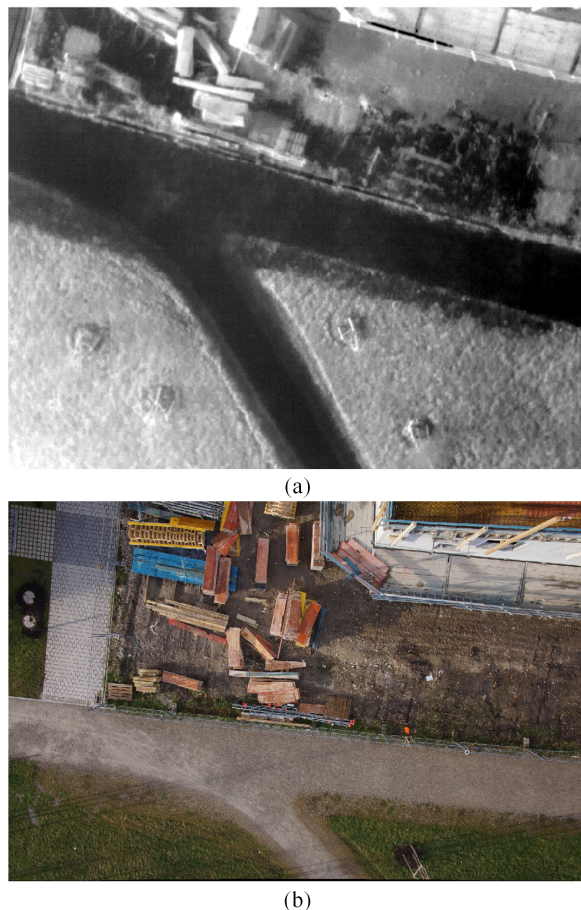
---

* Corresponding author

(a)

(b)

Figure 1. Illustration of (a) TIR image and (b) RGB image
acquired from the same scene.

in Section 1 describes the motivation and problems in the registration of thermal infrared images and RGB images. The literature review part in Section 2 summarises the state of art methods for image matching. Then, the proposed strategy will be described in detail with each step of our proposed method elaborated on in Section 3. After that, the data and result based on the proposed method will be demonstrated in Section 4. The result will be compared with several current methods. Finally, we will draw a conclusion regarding the results in Section 5.

2. LITERATURE REVIEW

The typical automatic image matching methods can be roughly divided into two categories: feature-based methods and area-based methods. The feature-based methods use a feature symbolic descriptor to describe the characteristics of elements in the image. By finding a similar descriptor, the corresponding elements in other images can be found. The feature-based image matching process includes the following steps: feature detection, feature description, feature matching, transformation model estimation, and image resampling (Zitová, Flusser, 2003, Chen et al., 2019a). Here, reliable and effective feature detection and feature description normally deem to be the preconditions for image matching. Feature detection detects distinctive features, such as points, lines, blobs between images. Typical algorithms are i.e. Harris detector (Harris, Stephens, 1988), Laplace of Gaussian (Lindeberg, 1994),Harris-Laplace (Mikolajczyk, Schmid, 2004), Hessian-Laplace (Mikolajczyk,

Schmid, 2005), Difference-of-Gaussian (Lowe, 2004) and so on.

After the feature detection, feature descriptors will be constructed to depict the characteristics of detected features. The scale-invariant feature transform (SIFT) (Lowe, 2004) is widely used due to its robustness to image rotation and scale changes. After SIFT, a series of descriptors have been proposed to improve the performance of SIFT, such as speed up robust feature (SURF) (Bay et al., 2008), fast local descriptor for dense matching (DAISY) (Tola et al., 2009), gradient location and orientation histogram (GLOH) (Mikolajczyk et al., 2005), Binary Robust Independent Element Feature (BRIEF) (Calonder et al., 2010) and Fast Retina Keypoint (FREAK) (Alahi et al., 2012). The feature-based method is usually more robust to geometric and radiometric differences, but ignores the positional relation between neighboring pixels in the image. Compared to the feature-based methods, the area-based methods make use of intensity information of the images. Feature detection and feature matching are combined by using some similarity measures, such as mutual information (MI), normalized cross-correlation (NCC), sum of square diference (SSD) and matching by tone mapping. Though this method is sensitive to geometric distortion, it is optimal for dense matching and subpixel matching.

Due to the increasing focus on registration of images from different sensor platforms, researchers have done a lot of work on multi-modal image matching, especially in the medical imaging and remote sensing domains. The challenge mainly lies in the geometric deformation and radiometric variation, leading to mis-correlated features. (Wang et al., 2012) proposed the bilateral filter SIFT to find the feature matched for optical images and SAR image. (Xiong et al., 2019) put forward the Rank-based Local self Similarity descriptor describing the local shape properties for SAR to optical images registration. (Ye et al., 2017) come up with the shape descriptor DLSC to form a similarity metric (named DLSC). By using the normalized cross correlation (NCC) of the DLSS descriptors, a template matching strategy is used to register the optical images and SAR images. (Xiang et al., 2020) incorporate the dense feature representations into the 3-D phase congruency scheme to estimate the translation of the Optical images and SAR images in sub-pixel level. Due to the outstanding performance in feature detection in intensity variant situations, phase congruency is widely used in illumination-invariant image matching. (Xiang et al., 2019) present the global and local frames for matching of opticla images and SAR images with gradiant location and orientation histogram (GLOH) descriptor. (Ye et al., 2018b) adopts the magnitude and orientation of phase congruency feature as structural image representation for subpixel image correlation. (Ye, Shen, 2016) proposes the histogram of oriented phase congruency (HOPC) and integrates a similarity matrix to represent the geometric structure features of images, which achieve great performance in the registration of SAR, optical and map image. Based on the HOPC descriptor combined with normalized correlation coefficient, a similarity measure called HOPCncc is used for image registration. (Ye et al., 2018a) proposes the minimum moment of phase congruency with LoG to detect feature points and local histograms of phase congruency for feature descriptors.

In addition to phase congruency, methods have been applied for infrared and RGB image registration. For example, (Chen et al., 2019b) utilizes the multi-scale and multi-orientation Gabor filter to encode the edge information as a descriptor to match the

infrared and visible image. (Yu et al., 2019) proposes a gray-scale weight with window algorithm together with normalized mutual information to register infrared images and RGB images. (Rahaghi et al., 2019) utilizes mutual information and particle swarm optimization. (Istenic et al., 2007) detects the line segments in visible and infrared images, and find the corresponding lines in Hough parameter space to recover the translation and rotation. The scale problem is not properly solved here. (Hrkać et al., 2007) proposed the method to co-register the IR images and RGB images based on the assumption that the images are taken at close viewpoints to simplify the transformation modal and reduce computation, which limit the application of this method. In (Shen et al., 2014), a robust selective normalized cross correlation (RSNCC) is used as the matching cost applying to a coarse-to-fine registration between the RGB/NIR and RGB/range images. (Turner et al., 2014) demonstrates a workflow to co-registrater visible, multi-spectral and thermal images aquired with a micor-UAV. The registration utilize the ground control points and is conducted by Photoscan. (Sanchez et al., 2015) register thermal and visible light images based on a novel multi-scale method that employs the stationary wavelet transform. The silhouettes of diseased plants extracted can be used to register thermal and visible light images with high accuracy. However, how the method can be used to registrate the IR images and RGB images captures in the urban areas without positioning information still remains to be solved.

3. PROPOSED METHOD

Considering the radiometric differences between thermal infrared images and RGB images, this contribution proposes a strategy to achieve a registration which is orientation-invariant and illumination-invariant. The scale was estimated by forming a pyrimad space of both images. The the best scale is chosen when the peak of normalized correlation of two edge strech images by phase congruency reaches the highest. Orientation of phase congruency and modality independent neighborhood descriptor (MIND) are adopted as a similarity descriptor. After that, the MINDflow is adopted for dense matching of thermal infrared images and RGB images. Finally, corresponding points from the dense matching are used for the transformation of the image. In Fig. 2, the workflow of the proposed method is given.
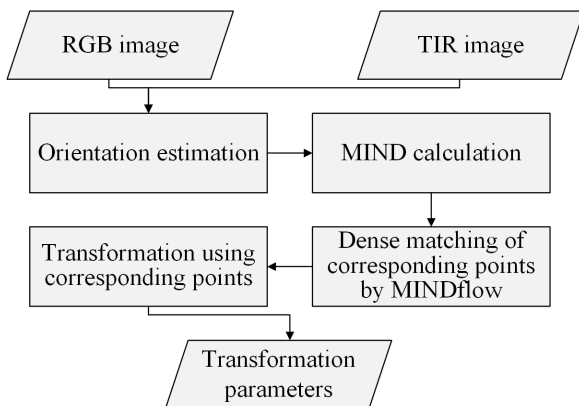

Figure 2. Workflow of the proposed method.

3.1 Orientation estimation

Phase congruency is calculated via local frequency analysis which originated from local energy modal. It assumes that per-

ceived features are located at points where the Fourier components are maximal in phase (Morrone, Owens, 1987). The phase congruency function is defined as the ratio of local energy as position $x[E(x)]$ to the sum of local Fourier components.

$$PC_1(x) = \frac{E(x)}{\sum_n A_n} = \frac{\sum_n A_n \cos(\phi_n(x) - \hat{\phi}_n(x))}{\sum_n A_n} \quad (1)$$

$\sum_n A_n \cos(\phi_n(x))$ is the Fourier series expansion of the function where A_n represents amplitude and $\phi_n(x)$ represents the local phase. Since it is not robust to noise and does not consider the frequency spread, a refined calculation model using log Gabor wavelets is adopted from Kovesi (Kovesi et al., 1999). The transfer function of a log Gabor filter is show as Eq. 2:

$$g(\omega) = exp(\frac{log(\omega/\omega_0)}{2(log\sigma_\omega/\omega_0)}) \quad (2)$$

ω_0 is the center of the filter and σ_ω/ω_0 is the constant parameters for bandwidth. If M_n^e and M_n^o represent the e even-symmetric and o odd-symmetric components of log Gabor filter at scale n, the response vector of a quadratic pair can be expressed as:

$$[e_n(x), o_n(x)] = [f(x) \times M_n^e, f(x) \times M_n^o] \quad (3)$$

Then, the amplitude $A_n(x)$ and phase $\phi_n(x)$ at the scale n can be expressed as:

$$A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$$
$$\phi_n(x) = atan2(e_n(x), o_n(x)) \quad (4)$$

Based on the magnitude of phase congruency, the orientation of phase congruency Θ can be achieved by the log Gabor odd-symmetric wavelet of multiple directions.

$$a = \sum_\theta (o_{no}(\theta)cos(\theta))$$
$$b = \sum_\theta (o_{no}(\theta)sin(\theta)) \quad (5)$$
$$\Theta = atan2(b, a)$$

θ is the orientation of odd-symmetric wavelet. Therefore, the orientation of the image can be calculated by the accumulated orientation.

3.2 Modality independent neighborhood descriptor

MIND descriptor (Heinrich et al., 2012) uses the principle of local self-similarity. The patch distance is first calculated as the sum of square distances between two voxels x_1 and x_2 as $D_p(x_1, x_2)$, between the patches P of size $(2p + 1)^d$.

$$D_p(I, x_1, x_2) = \sum_{p \in P} (I(x_1 + p) - I(x_2 + p))^2 \quad (6)$$

The exact patch-distance can be efficiently calculated using convolution filter C of size $(2p + 1)^d$ with point-wise squared difference between image I and transformed image $I'(r)$ (r is the transformed distance.

Moreover, in order to highlight the response of MIND with similar patches, a Gaussian function is used. The variance of the image can be estimated via mean of the patch distances them-

selves within a six-neighborhood $n \in \chi$.

$$V(I, x) = \frac{1}{k} \sum_{n \in \chi} D_p(I, x, x + n) \quad (7)$$

where k is the number of neighbors in the defined neighborhood, in our case it equals to six.

Then, the output of MIND can be defined by a distance D_p, a variation estimate V and a spatial search region R.

$$MIND(I, x, r) = \frac{1}{n} exp(-\frac{D_p(I, x, x + r)}{V(I, x)}), r \in R \quad (8)$$

3.3 Dense matching of corresponding points by MINDflow

MINDflow is adapted from the renowned Siftflow (Liu et al., 2010) and here it is adopted for dense matching in feature space. Considering the SIFT descriptor can hardly describe the similarity between the images from different spectral bands, especially between the thermal infrared images and RGB images, here MIND descriptor is applied.

Specifically, let $s1$ and $s2$ be two images represented by MIND descriptor, and ε contains all the spatial neighborhoods. $w(p)$ is the flow vector at grid coordinate $p = (x, y)$. The energy function is defined as:

$$E(w) = \sum_p min(||s_1(p) - s_2(p + w(p))||1, t) +$$
$$\sum_p \eta(|u(p) + \mu(p)|) + \quad (9)$$
$$\sum_{(p,q) \in \varepsilon} min(\alpha|w(p) - w(q)|, d)$$

which contains a data term (i.e., the first term of Eq. 9), a small displacement term (i.e., the second term of Eq. 9) and a smoothness term (i.e., the third term of Eq. 9). The data term contains the MIND descriptor to be matched along with the flow vector $w(p)$. The small displacement term constrains the flow vector to be as small as possible when no information is available. The smoothness term constrains the flow vector of adjacent pixels to be similar. L_1 norms are used in the data term and the smoothness term to account for matching outliers and flow discontinuities. This energy function can be then solved by the use of sequential belief propagation (BP-S) (Szeliski et al., 2008).

3.4 Image transformation using corresponding points

Once the corresponding points are obtained, the dense alignment between TIR and RGB images can be achieved. Based on the dense matching points, a non-parametric transformation based on a flow map can be achieved. Each pixel in the transformed image can be thus resampled.

4. EXPERIMENTAL DATA AND RESULT

The experimental data used consist of two datasets: TIR images (512×640 pixels) and RGB images (2680×4019 pixels). Two UAV flights with the same predefined flight path have been performed: one with the RGB camera and one with the TIR camera. Three image sets are utilized for validation. However, due

to limited positioning accuracy, the exterior orientation of the same waypoints for both flight are slightly different. In order to evaluate the co-registration result, we select two images (Fig. 3) and manually selected matched keypoints for evaluation. The RGB images lay on the left, while the TIR images are on the right. In image 1, 46 point pairs are selected and in image 2, 40 point pairs are selected. In Fig. 3, we found the images we use here have two characteristics. Firstly, the scale differences exist between the two datasets. The RGB images are much larger than the corresponding TIR images. Secondly, the overlap between the corresponding images are not always the same. For the second image pairs, the selected points in the RGB images distributed on the bottom, while the corresponding points in the TIR images spread close to the top. Compared to the first image pairs, the overlap between the second image pairs is smaller.
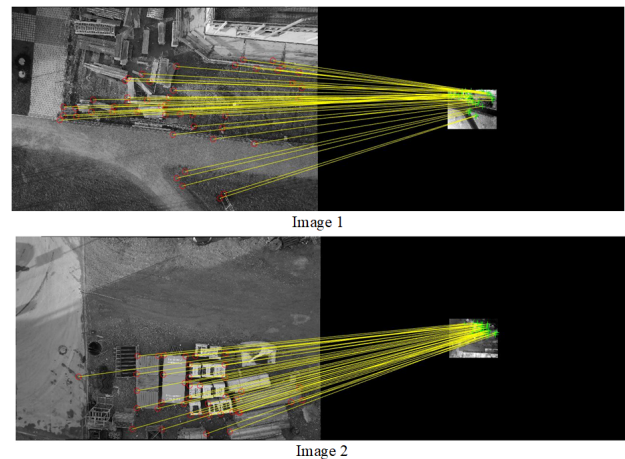

Image 1

Image 2

Figure 3. Manually select keypoint pairs for evaluation.

Data	Image Set 1	Image Set 2
Rotation (degree)	-6	-5
MIND($error_x$)	18.52	8.06
MIND($error_y$)	5.12	11.99
SIFT($error_x$)	14.90	23.38
SIFT($error_y$)	58.68	41.19

Table 1. Orientation difference between the TIR image and RGB image.

Fig. 4 shows the results of the proposed method. The first column are the original RGB images, the second column are the TIR images, the third column plots the flow map results of each pixel, the third column shows how phase congruency edge strength image by the RGB image are warped to that of the TIR image, and the last column show the checkerboard image from the TIR image and the RGB image. In the flow map, the color demonstrates the flow vector of each pixel with respect to the RGB images. Generally speaking, although the overlap between the two images are not exactly the same, the proposed method is able to find the corresponding areas with significant features. Compared to other data, Image 3 didn't achieve an optimal results due to the repeated texture information and similar boundary structures inner and outer the building roof.

Fig. 5 presents the comparision results between the proposed MINDflow and SIFTflow. The first two rows are the results for image 1, the last two rows are results for the Image 2. Note that the images applied to the SIFT flow has already adopted the scale parameters generated by the proposed method. Bedside, Table lists the estimated orientation differences between
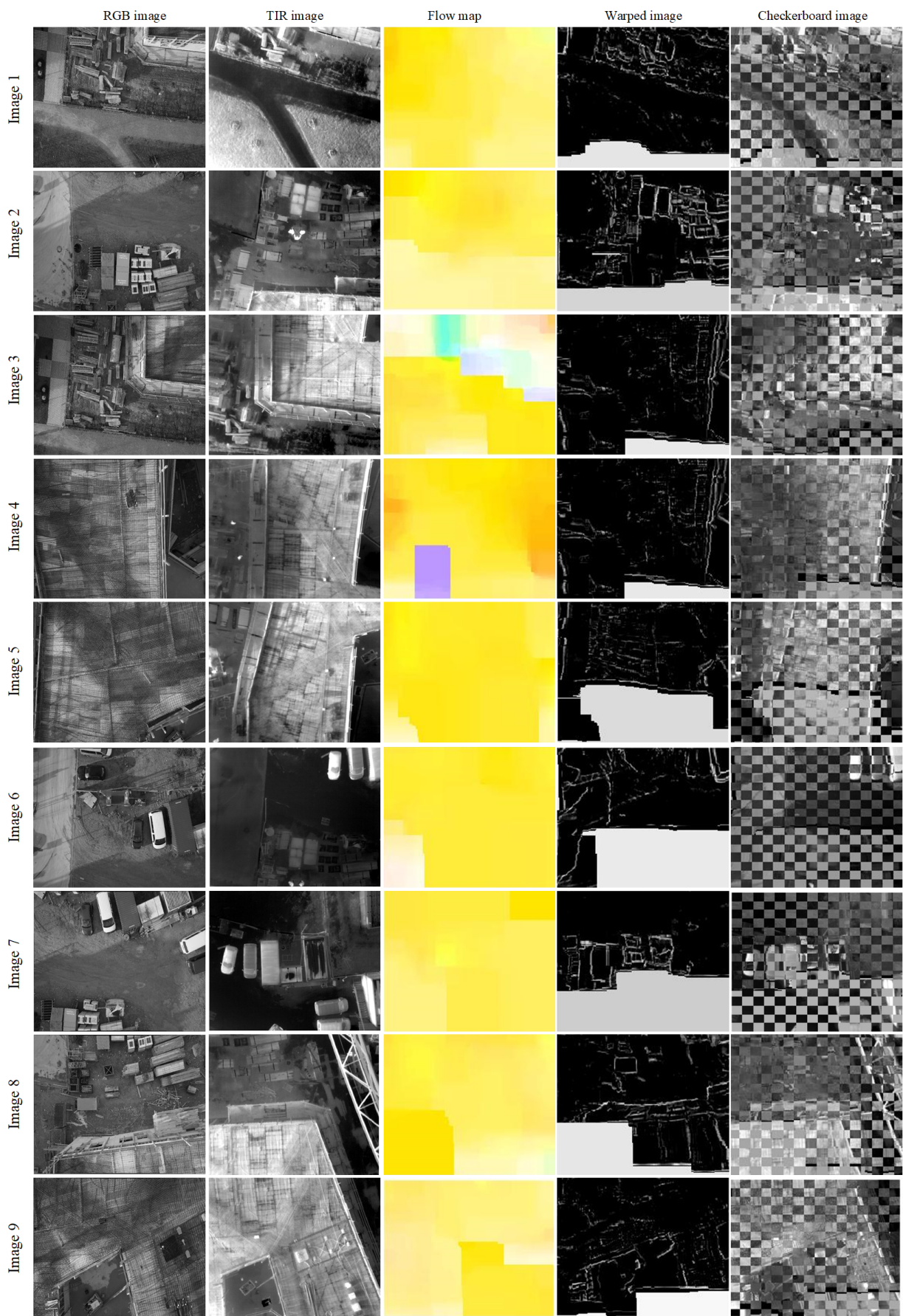
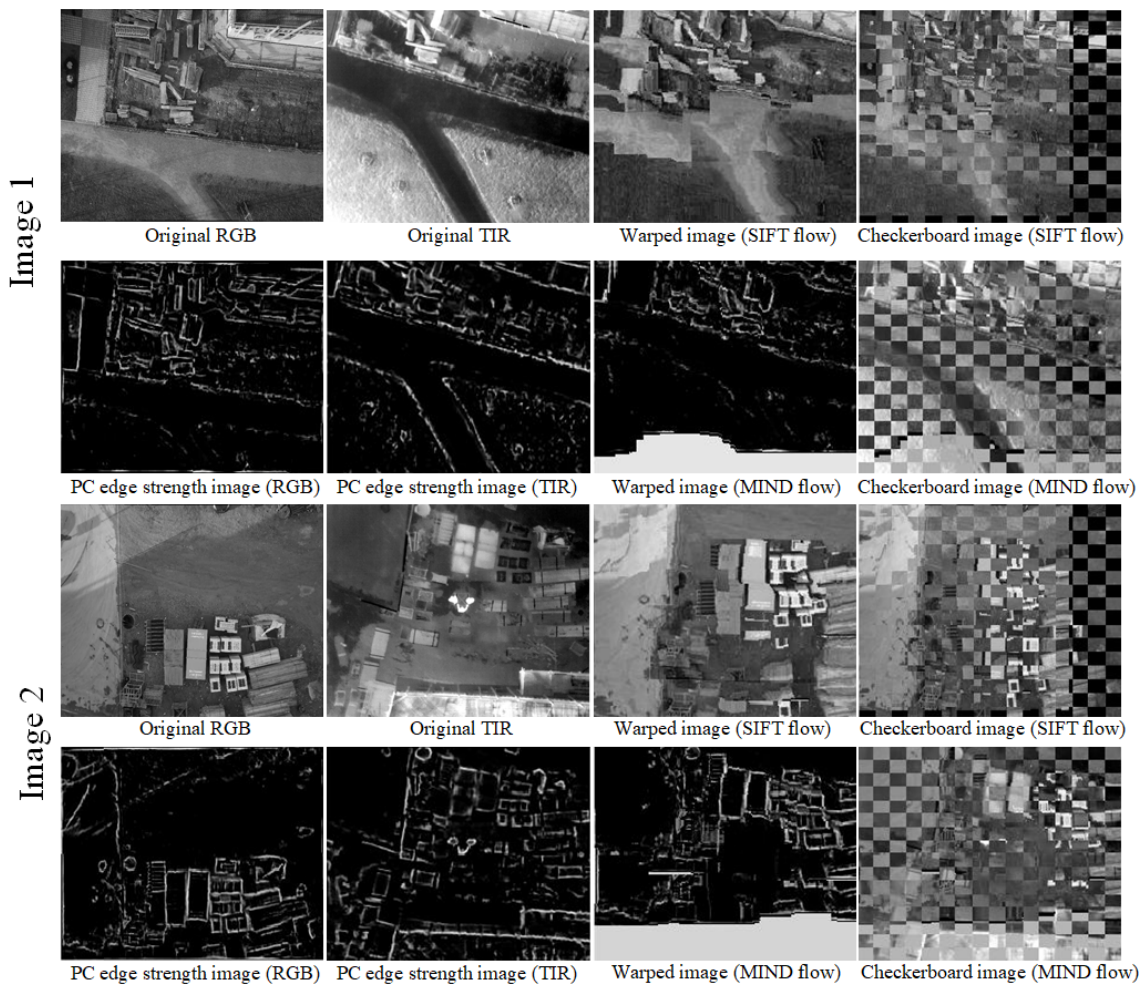Figure 4. Results by proposed method.

Figure 5. Result by proposed method using MINDflow and SIFTflow.

image pairs based on phase congruency and the error in x and y direction. The average error for MINDflow is less than that of SIFTflow. From the table, MINDflow gets better result in y-direction while SIFtflow receives better result in x-direction for Image 1. Besides, the errors for Image 2 are always less than that in Image 1. However, when we analysis the results in Fig. 3, especially comparing the warped images in the third column, results by SIFTflow have larger distortions compared to the results by MINDflow. By checking the boundary of checkerboard image by Image 1, the MINDflow can achieve the approximate aligned road along the upper boundary. The white area on the bottom of warped image indicating a large offsets in y-direction. Comparing the original RGB image and the TIR image, a large offset on y direction is reasonable due to the overlapping areas spanned positions in different images. The phenomenon could be better observed in the Image 2. The overlap between the two images is about half of each image. In the Checkerboard image by MINDflow, the shape of the boxes are kept despite the large offsets and the radiation differences in the RGB and the TIR images.

5. CONCLUSION AND DISCUSSION

In this paper, we proposed a strategy to register thermal infrared images and RGB images based on MINDflow. Firstly, the orientation difference of TIR images and RGB images is solved by finding the dominant image orientations with phase congru-

ency. Then, the MIND descriptor and flow matching are adopted as MINDflow for dense matching. Our method is tested in 9 image sets in construction areas. The result shows that it is able to achieve an optimal result in the images with significant features but can hardly deal with the area with repeat patterns or insufficient structure differences. Comparing the result with SIFTflow, MINDflow combined with Phasecongruency can better extract the features for dense matching.

Though our proposed method is able to preliminary accomplish the mission of the registration of thermal infrared images and RGB images, there are some challenges we need to face. First, we need to find a better way to realize optimal orientation and scale estimation for multi-modal images. For both area-based and feature-based methods, the orientation and scale will highly affect the feature descriptor or similarity descriptor. Though we are able to offer the coarse scale and orientation, an optimal result will improve the result. Besides, the method is limited by a certain overlap between the images, which is the typical problem when the data are acquire separately. The overlap areas influence the estimation of rotation and scale. Last but not least, we need to find the optimal solution for the texture repetition areas.

The next challenges is the detection of overlapping areas in RGB and TIR images to optimal the image orientations and scale. How to match images with less textures or repetitive patterns is also our concern.

REFERENCES

Alahi, A., Ortiz, R., Vandergheynst, P., 2012. Freak: Fast retina keypoint. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Ieee, 510–517.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), 346–359.

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. Brief: Binary robust independent elementary features. *European conference on computer vision*, Springer, 778–792.

Chen, H., Xue, N., Zhang, Y., Lu, Q., Xia, G.-S., 2019a. Robust visible-infrared image matching by exploiting dominant edge orientations. *Pattern Recognition Letters*, 127, 3 - 10. Advances in Visual Correspondence: Models, Algorithms and Applications (AVC-MAA).

Chen, H., Xue, N., Zhang, Y., Lu, Q., Xia, G.-S., 2019b. Robust visible-infrared image matching by exploiting dominant edge orientations. *Pattern Recognition Letters*, 127, 3–10.

Christiansen, P., Steen, K. A., Jørgensen, R. N., Karstoft, H., 2014. Automated detection and recognition of wildlife using thermal cameras. *Sensors (Switzerland)*, 14(8), 13778–13793.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, pp.

Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, M., Schnabel, J. A., 2012. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis*, 16(7), 1423–1435.

Hrkać, T., Kalafatić, Z., Krapac, J., 2007. Infrared-visual image registration based on corners and hausdorff distance. *Scandinavian Conference on Image Analysis*, Springer, 383–392.

Istenic, R., Heric, D., Ribaric, S., Zazula, D., 2007. Thermal and visual image registration in hough parameter space. *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, IEEE, 106–109.

Kovesi, P. et al., 1999. Image features from phase congruency. *Videre: Journal of computer vision research*, 1(3), 1–26.

Lindeberg, T., 1994. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2), 225–270.

Liu, C., Yuen, J., Torralba, A., 2010. Sift flow: Dense correspondence across scenes and its applications. *IEEE transactions on pattern analysis and machine intelligence*, 33(5), 978–994.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.

Mikolajczyk, K., Schmid, C., 2004. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1), 63–86.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10), 1615–1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2), 43–72.

Morrone, M. C., Owens, R. A., 1987. Feature detection from local energy. *Pattern recognition letters*, 6(5), 303–313.

Rahaghi, A. I., Lemmin, U., Sage, D., Barry, D. A., 2019. Achieving high-resolution thermal imagery in low-contrast lake surface waters by aerial remote sensing and image registration. *Remote sensing of environment*, 221, 773–783.

Sanchez, V., Prince, G., Clarkson, J. P., Rajpoot, N. M. et al., 2015. Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. *Pattern Recognition*, 48(7), 2119–2128.

Shen, X., Xu, L., Zhang, Q., Jia, J., 2014. Multi-modal and multi-spectral registration for natural images. *European Conference on Computer Vision*, Springer, 309–324.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C., 2008. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE transactions on pattern analysis and machine intelligence*, 30(6), 1068–1080.

Tola, E., Lepetit, V., Fua, P., 2009. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32(5), 815–830.

Tong, X., Ye, Z., Xu, Y., Gao, S., Xie, H., Du, Q., Liu, S., Xu, X., Liu, S., Luan, K. et al., 2019. Image Registration With Fourier-Based Image Correlation: A Comprehensive Review of Developments and Applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), 4062–4081.

Turner, D., Lucieer, A., Malenovskỳ, Z., King, D. H., Robinson, S. A., 2014. Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds. *Remote Sensing*, 6(5), 4003–4024.

Wang, S., You, H., Fu, K., 2012. BFSIFT: A Novel Method to Find Feature Matches for SAR Image Registration. *IEEE Geoscience and Remote Sensing Letters*, 9(4), 649-653.

Weinmann, M., Leitloff, J., Hoegner, L., Jutzi, B., Stilla, U., Hinz, S., 2014. Thermal 3D mapping for object detection in dynamic scenes. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-1(November), 53–60.

Xiang, Y., Tao, R., Wan, L., Wang, F., You, H., 2020. OS-PC: Combining Feature Representation and 3-D Phase Correlation for Subpixel Optical and SAR Image Registration. *IEEE Transactions on Geoscience and Remote Sensing*.

Xiang, Y., Wang, F., Wan, L., Jiao, N., You, H., 2019. OS-Flow: A Robust Algorithm for Dense Optical and SAR Image Registration. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9), 6335-6354.

Xiong, X., Xu, Q., Jin, G., Zhang, H., Gao, X., 2019. Rank-Based Local Self-Similarity Descriptor for Optical-to-SAR Image Matching. *IEEE Geoscience and Remote Sensing Letters*, 1-5.

Ye, Y., Shan, J., Hao, S., Bruzzone, L., Qin, Y., 2018a. A local phase based invariant feature for remote sensing image matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, 142, 205–221.

Ye, Y., Shen, L., 2016. Hopc: A novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 9.

Ye, Y., Shen, L., Hao, M., Wang, J., Xu, Z., 2017. Robust Optical-to-SAR Image Matching Based on Shape Properties. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 564-568.

Ye, Z., Tong, X., Zheng, S., Guo, C., Gao, S., Liu, S., Xu, X., Jin, Y., Xie, H., Liu, S. et al., 2018b. Illumination-robust subpixel Fourier-based image correlation methods based on phase congruency. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 1995–2008.

Ye, Z., Xu, Y., Chen, H., Zhu, J., Tong, X., Stilla, U., 2020. Area-Based Dense Image Matching with Subpixel Accuracy for Remote Sensing Applications: Practical Analysis and Comparative Study. *Remote Sensing*, 12(4), 696.

Yu, K., Ma, J., Hu, F., Ma, T., Quan, S., Fang, B., 2019. A grayscale weight with window algorithm for infrared and visible image registration. *Infrared Physics & Technology*, 99, 178–186.

Zin, T. T., Takahashi, H., Toriu, T., Hama, H., 2007. Fusion of Infrared and Visible Images for Robust Person Detection. *Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on*, 1, 310–310.

Zitová, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing*, 21(11), 977 - 1000.