Article

Not peer-reviewed version

IR-RGB Image Registration Using Deformation Field and Mask Loss

Munhyung Lee * and Jangwoo Kwon

Posted Date: 25 September 2023

doi: 10.20944/preprints202309.1577.v1

Keywords: computer vision; deep learning; multi-modality image registration

Article

IR-RGB Image Registration Using Deformation Field and Mask Loss

Munhyung Lee [1,*] and Jangwoo Kwon [2]

[1] Department of Electrical & Computer Engineering, Inha university ; 1; mun0659@inha.edu
[2] Department of Electrical & Computer Engineering, Inha university ; 2; jwkwon@inha.ac.kr
* Correspondence: mun0659@inha.edu;

Abstract: This study proposes a method for image matching between infrared(IR)-RGB images using a deep learning network to estimate the deformation field. We propose a deformation field generator (DFG) that estimates the deformation field of the transformation matrix to match each pixel or IR image to the RGB image. DFG is a network that receives IR and RGB images as input; the output size is two channels and has the sample resolution as the input image. By warping the IR image through a grid-sampler that warps the image according to the value of the deformation field, we can obtain a warped IR image that matches the RGB image. Additionally, to check whether the warped IR image matched the RGB image, the masking images detecting the segmentation of objects were photographed in two images. Without directly comparing IR and RGB images, we proposed mask loss that warps the IR mask image through the deformation field and grid sampler and then compares the warped IR mask image with the RGB mask image. Mask loss solves the spatial similarity comparison problem with multi-modality images, such as IR and RGB images, by comparing the mask image with the same modality image as the mask image.

Keywords: computer vision; deep learning; multi-modality image registration

1. Introduction

Recently, the number of cases for which data have been acquired using multiple sensors at industrial sites has increased. For example, in an environment where it is difficult to obtain an accurate image using an RGB camera alone, such as at night, in heavy rain, or in fog, an IR sensor can be used to compensate for the disadvantages of an RGB camera [1]. Integrating IR and RGB images increases the performance of tasks such as object detection[1] and object tracking [2,3] in computer vision(CV).

when data are obtained through multiple sensors, each sensor has external factors, such as the base curve of the lens, the position of the sensor, and the measurement angle; therefore, the images obtained using the sensor have different coordinates. To integrate the data of two images, different coordinates must be matched. In many fields utilizing multi-sensor data, image registration between data with different modalities, such as IR and RGB, called multi-modality image registration is a very important task.

Classical multi-modal image registration [23] was adapted to another image by transforming an image in a direction that maximizes the resulting value of the predefined similarity comparison method. To accurately match external parameters such as lens curvature, the camera position and distance between sensors must be known to make accurate matching, and it is difficult to achieve accurate matching without parameters. In addition, the computational cost is excessively high for classical methodologies, which is a fatal problem in task requiring real-time response, such as autonomous driving.

Because of these problems, it is difficult to perform image registration through classical methodologies only image without additional environmental parameters. To solve this problem, research on image registration methodology using deep learning has become increasingly active.

In the medical field, multi-modality image registration uses deep learning to perform multi-modality image registration such as MRI and CT images[5]. However, in the case of medical

 doi:10.20944/preprints202309.1577.v1

images, the shooting environment is limited, so there is almost no background noise. So, general supervised learning-based image registration methodology uses images that have already been registered as ground truth. After setting the imaging device parameters to obtain images with the same coordinates, an affine transformation was applied to one image to artificially create an image with a misaligned coordinate and use it as the input . Thus, this approach is very inefficient and contradictory to misaligned the matched image again to train the registration network.

In this study, we propose a network that matches IR-RGB images and a method for evaluating spatial matching by comparing the masking of the two images. Through a deformation field generator(DFG), a deformation filed representing the x-axis and the y-axis change rates were inferred for each pixel to match the IR image with the RGB image, and each IR pixel is warped into a coordinate corresponding to the deformation field through a grid sampler to obtain an IR image that matched the RGB image. In addition, after converting the IR-masking iamge using this deformation field and then comparing the image with RGB masking image, we solved the spatial similarity comparison problem between the cross-modality images by converting them into mono-modality. The main contribution of out work are:

- We proposed a multi-modality image registration methodology that performs image matching between IR-RGB images through deformation field inference.
- We proposed a mask loss that compares the masking image of the image to solve the cross-modality problem.
- Experiments show that the proposed method is effective not only areas where image masking information is provided but also in areas where image masking information is not given.

2. Related works

2.1. Feature matching based

The traditional image registration method consists of four steps. First is Feature detection, which finds the same features regardless of image conversion or degradation. Second is feature matching, which considers the relationship between the features of non-aligned and reference images obtained during the feature detection process. Third is transform mode assessment, which determines the transformation parameters to be applied to the non-aligned images. Last is image transformation step, during which registration is performed by applying transformation parameters to non-aligned images.In Feature matching-based image registration methodology, feature detection and feature matching are very important. Representative methodologies that focused on processes such as the Harris corner [24], SIFT [25], and AGAST [26]. It is difficult to achieve accurate registration using this methodology alone and many studies have been conducted recently to increase accuracy through deep learning.

2.2. Homography estimation based

Homography-based image registration is a method in which image transformation is applied after obtaining a homography matrix with eight parameters based on a 2D image, which is a transformation parameter in the transform mode assessment stage. In studies such as [8–10], deep learning networks perform feature detection and matching, and transform mode assessment steps to infer accurate homography matrices using only image pairs. Deep homography estimation [8] infers a homography matrix using 10 simple convolutional(conv) layers and shows that image matching is possible using it. In this study, a deep learning-based methodology for inferring homography was a deep learning-based methodology for inferring homography was not accurate when two images are obtained using different cameras, such as RGB-IR, and the types of transformations that can occur through homography are limited. Owing to these limitations, optical flow-based image registrations have appeared, which yield the conversion matrix itself for each image pixel. The estimation of the movement of pixels or features in an image from a continuous image is called optical flow estimation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

2.3. Optical flow based

Estimating the movement of pixels or features in an image from a continuous image yielded optical flow estimation. These are as follows [11,15]:. as end-end methods that infer optical flow between two images for an image provided as input. Some studies have used optical flow for image registration [12–14]. FlowReg[13] uses FlowNet[11] to match medical images. FlowReg consists of FlowRegA, which estimates a 3D affine matrix, and FlowReg-O, which estimates optical flow. After converting the images using the affine matrix obtained from FlowRegA, fine-tuning was performed using FlowReg-O. In the case of PWC-Net [17] and pointPWC [18], the matching performance improved by pyramid-structured architecture. RAFT [19] improved performance by adding the GRU structure to the Iteration method. However, these studies performed image registration on mono-modality.

3. Method

In this study, the previously mentioned feature detection, feature matching, and IR-RGB image matching proceed in two stages: the DFG network, which replaces the transform mode analysis process, and a grid sampler that performs image transformation. The first step is a DFG, ResUnet[20]-based registration network that receives IR images(I_{ir}) and RGB images(I_{rgb}) as input and generates a deformation field(ϕ). ϕ represents the x-axis and y-axis translation of each pixel and has the same resolution as the input image. In the second step, ϕ and I_{ir} obtained through DFG are received as inputs, and a warped IR image(I_{ir}) is obtained using a grid sampler that performs an image transform that converts each pixel based on the deformation field on the IR image. Figure 1 shows the inference structure of the multi-modality image registration network proposed in this paper.

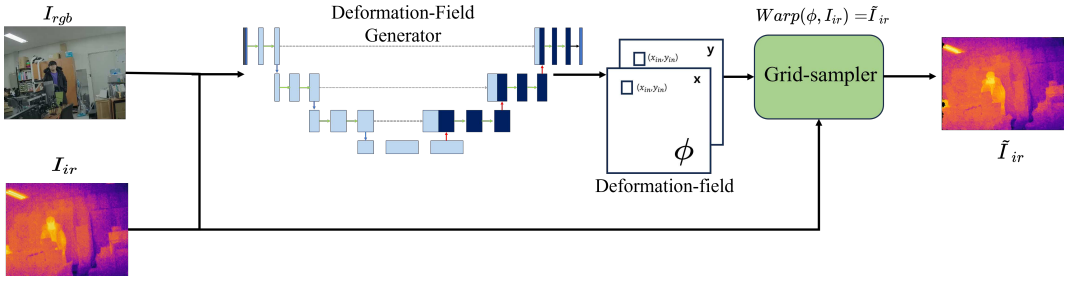

Figure 1. Inference process The network structure for obtaining the matched image is divided into two steps as follows. Deformation filed generator (DFG), which takes thermal images and real-life images as inputs and generates a deformation-field (ϕ) for mapping images from thermal images to real-life images. The second is a Grid sampler, which receives ϕ and thermal images as input and applies transformations corresponding to ϕ to thermal images. The first DFG is combined to obtain ϕ, and then the thermal image and ϕ are put into the Grid-sampler to obtain the warped IR image.

3.1. Deformation Field Generator

If input image's resolution $H \times W$, the deformation field generator (DFG) infers a 2-channel deformation field with resolution $H \times W$ and two channels representing the transformation in the x- and y-axes for image registration. DFG process sequentially receives data including tomcat I_{ir} and I_{rgb} as input and then repeats the following process: passing through a block composed of $3 \times 3Conv + leakeyrelu$, and performing down-sampling. Before performing down-sampling, the feature map was stored at each step. After down-sampling, the residual block is passed N times. In this study, N was set to 3, 5, or 7. Next, the up-sampling process was performed by combining the feature map and residual block results for the same input resolution previously saved in the down-sampling process. Finally, up-sampling was repeated until $H \times W$, which is the same resolution as the input image, and then $1 \times 1conv$ was applied and finally output a 2-channel result. Figure 2 shows the DFG structure.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

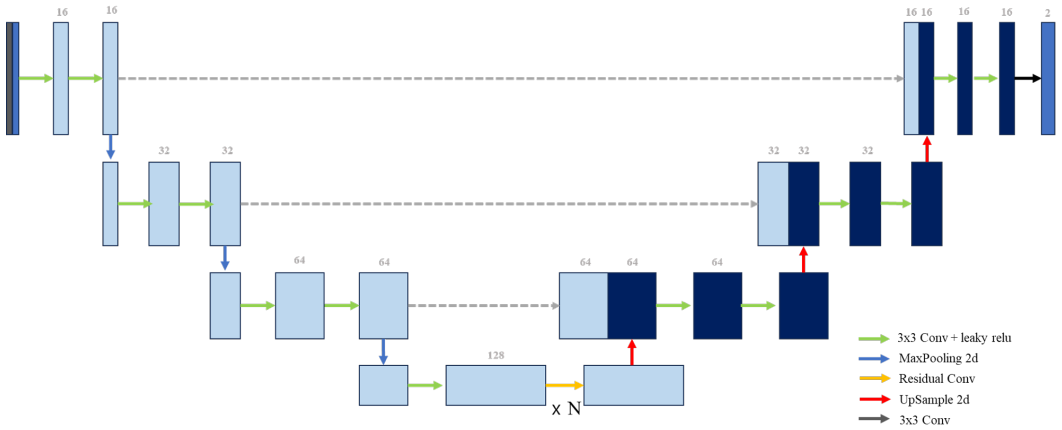

Figure 2. Deformation-Field generator architecture.

However, if the output of the network (ϕ_G) is directly transformed through a grid sampler, there is no information on the spatial feature of the original data in the early stages of training, which results in severe spatial distortion and difficulty in training. Therefore, we utilize the identical deformation field (ϕ_I) to allow the network to learn the process of maintaining and transforming the spatial characteristics of the original data to some extent. The identical deformation field satisfies Equation 2 as a deformation field representing the identical transform that allows the input image to emerge without transforming the image when the image and ϕ_I are inputted to the grid-sampler. As shown in Equation 3, the structure transforms the image into a deformation field(ϕ) created by adding ϕ_I and ϕ_G so it can maintain spatial features of the original image rather than using ϕ_G directly.

$$\phi_G = DFG(I_{rgb}, I_{ir}) \tag{1}$$

$$I = Warp(I, \phi_I) \tag{2}$$

As shown in Equation 3, the structure transforms the image into ϕ created by adding ϕ_I and ϕ_G so it can maintain spatial features of the original image rather than obtaining a matrix of $H \times W$ that directly matches RGB-IR.

$$\phi = \phi_G + \phi_I \tag{3}$$

3.2. Grid sampler

Figure 3 shows how the grid sampler changes the input image according to the deformation field. When any coordinate (i, j) of I_{ir} is v, when a value located in v of ϕ is defined as $\phi(v)$, when the two-dimensional value is defined as $\phi(v)$, and each pixel in I_{ir} is transformed into an x-axis given as input, as in Equation 4. As shown in Equation 5, all coordinates in I_{ir} are converted to obtain $\tilde{I}_{ir}$, which is an image warped by the gird-sampler. In addition, the grid sampler did not participate in training, there were no learning parameters, and only the image translation step was performed.

$$Warp(I_{ir}, \phi)[v] = I_{ir}[v + \phi(v)] \tag{4}$$

$$\tilde{I}_{ir} = Warp(I_{ir}, \phi) \tag{5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

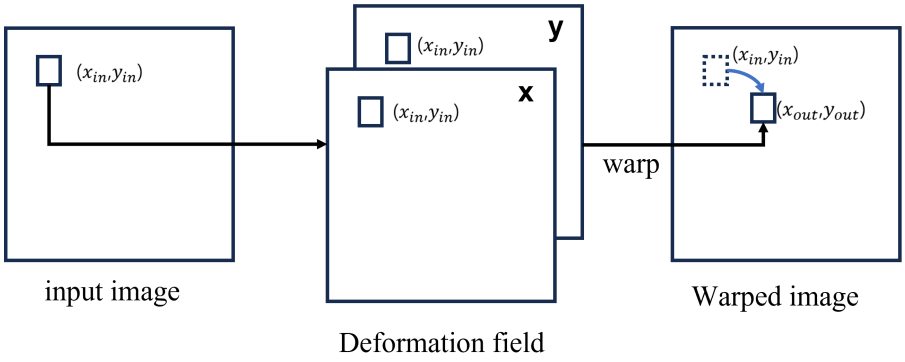

Figure 3. Grid sampler process.

3.3. Object function

3.3.1. Mask loss

To train the image registration process in the model, it is necessary to quantitatively indicate how the spatial match rate was calculated. In this study, this problem was solved by comparing the masking images of each RGB-IR image. Insert I_{ir} and I_{rgb} into DFG to obtain the deformation field and then insert M_{ir} and the deformation field into the Grid-sampler to obtain warped masking image $\tilde{M}_{ir}$. Using L1loss, $\tilde{M}_{ir}$ and RGB masking image(M_{rgb}) were compared and used as the object function. Equation 7 represents the mask loss.

$$\tilde{M}_{Ir} = Warp(M_{ir}, \phi) \tag{6}$$

$$L_m = \sum |\tilde{M}_{ir} - M_{rgb}| \tag{7}$$

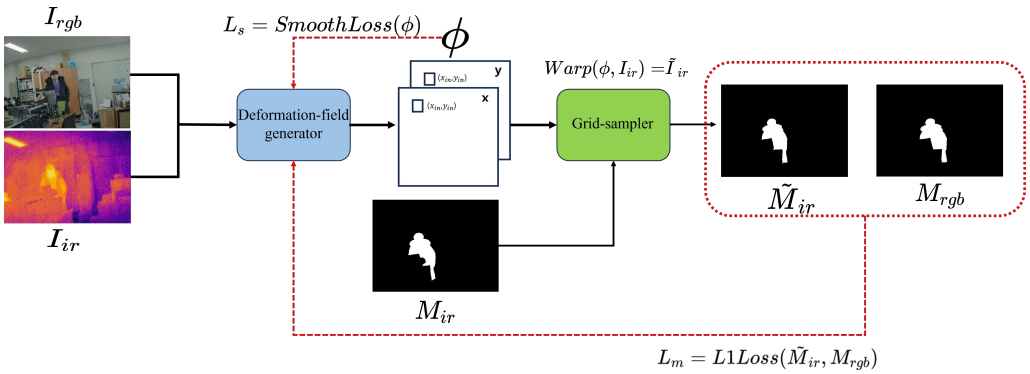

Figure 4. Model train architecture.When the network training, it does not utilize the converted image. A IR image is placed in a field generator to obtain ϕ. ϕ and thermal image masking images (M_{ir}) are placed in a Grid-sampler to compare the transformed masking images with the masking images of RGB images.

3.3.2. Smooth Loss

When converting an image using a grid sampler, it was necessary to limit the DFG to prevent the value of ϕ from becoming too large to damage the spatial properties of existing images and to ensure the spatial locality. Smooth loss applying L1 regularization allows the deformation field (ϕ) to represent a natural transformation. The smooth loss compares the similarity between each pixel and surrounding pixels in ϕ obtained from DFG. When difference values in x-axis and y-axis between each pixel and surrounding pixels defined as $\partial_x\phi$,$\partial_y\phi$, Equation 8 shows the smooth loss.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

$$L_{smooth} = \sum |\partial_x \phi| + |\partial_y \phi| \tag{8}$$

Finally, the total loss used for model training is shown using Equation 9.

$$L_{total} = \lambda_m * L_m + \lambda_{smooth} * L_{smooth} \tag{9}$$

4. Experiment

To obtain the dataset used in the experiment in this study, we used Ozray HK380,which has a thermal imaging sensor and an RGB sensor used at industrial sites. The specifications of each sensor are listed in Table 1. The IR-RGB images measured from the camera are defined in units of one data according to the fps of each sensor of one data according to the fps of each sensor. In addition, annotation was performed on one object taken in the same IR image and RGB images of each data pair to generate a masking image for loss calculation and evaluation. A total of 3,296 images were captured by changing the shooting angle at three different locations. Figure 5 shows a sample of the acquired dataset.

Table 1. The device specifications used in the experiment.

Thermal Sensor Information	
Technology	Amorphous Silicon (α-Si) microbolometer
Spectral Response	8 $\sim$ 14um (LWIR)
Pixel Pitch	17um
Focal Plane Array	384 x 288
Frame Rate	10fps
Color CMOS Sensor Information	
Maker	Omnivision
Pixel Pitch	2um
Optical Format	1/2.7"
Resolution	1920 x 1080
Frame Rate	30fps

HK380 device spec

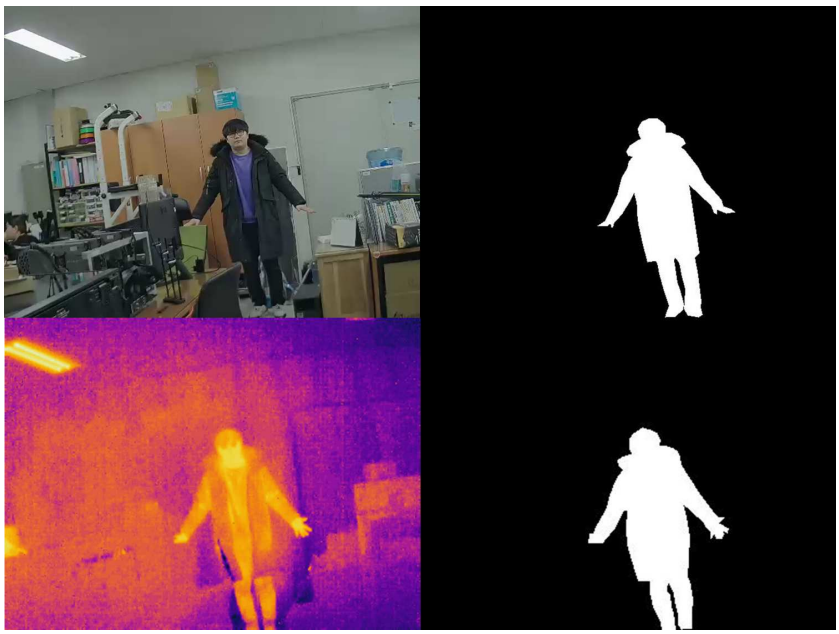

Figure 5. Dataset sample. Examples of train datasets are up-left real-life images, up-right real-life images, down-left, thermal images, and down-right thermal images.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

4.1. Evaluation

Intersection over union (IoU), a metric used in segmentation, was used as a metric to evaluate the proposed matching method. The deformation field (ϕ) and IR masking image (M_{ir}) obtained through DFG are input into a grid sampler to produce a matching IR masking image ($\tilde{M}_{ir}$). Then, the IoU between $\tilde{M}_{ir}$ and M_{rgb} masks is calculated using Equation 10, which represents IoU.

$$IoU(\tilde{M}_{ir}, M_{rgb}) = \frac{\sum(\tilde{M}_{ir} \cap M_{rgb})}{\sum(\tilde{M}_{ir} \cup M_{rgb})} \tag{10}$$

4.2. Result

The dataset used in this experiment consists of 3,296 pairs of image frames of IR and RGB images and the corresponding masking images, of which 660 were used as test datasets, 20 percent, and the remaining 2,576 were used as train datasets. To determine whether our methodology is effective for image matching; we compute the inter-IoU pair ir-mask (M_{ir}) and RGB-mask image (M_{rgb}) and then transform the IR-mask image(M_{ir}). After applying this methodology, we compared the IoU between $\tilde{M}_{ir}$ and M_{rgb} to compare IoU scores before and after applying the methodology. Comparison of the input size and performance according to the network depth of the model, the remaining layers of the DFG were changed to 3, 5, and 7 and the size of the input image was set to 640 × 480,512 × 512, and 256 × 256, and the application results were compared.

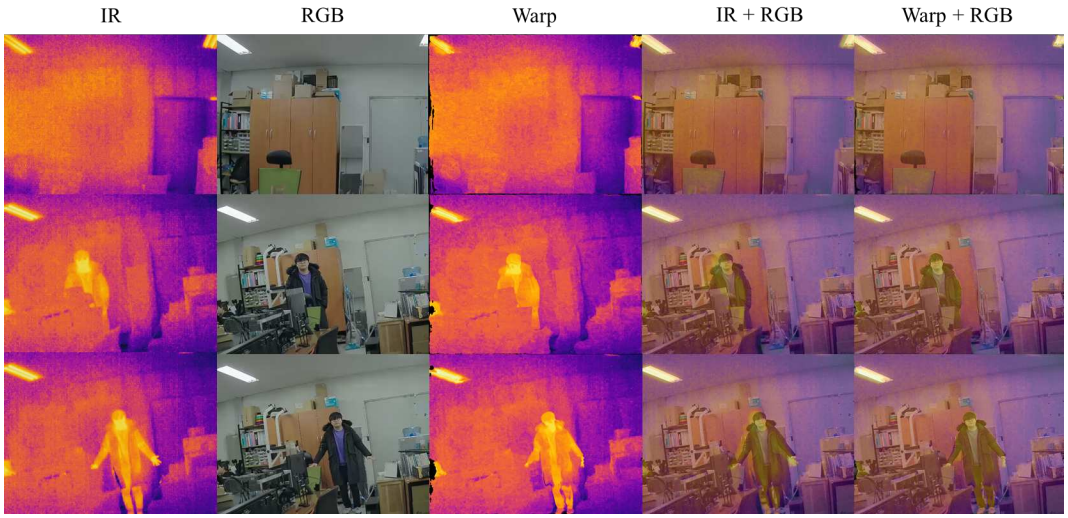

Figure 6. Registration result. Result of the matching. This is the result of applying alpha blending to the original thermal image, the corrected thermal image, and the actual image by applying alpha blending to the thermal image, the actual image converted from the left.

Table 2 shows that when the input size was 640 × 480, and the number of residual layers was 5, IoU 0.9083 was the highest, and when the residual layer was increased to 7, it was 0.8831, which was lower than when the number of layers was 5. When the input size was set to 256 × 256, the performance was highest at IoU 0.8983 when the number of residual layers was 3, and the performance was degraded at 5 when the number of layers was further increased. The larger the input size, the higher the matching performance and the higher the IoU score when the number of residual layers was properly adjusted depending on the input size. Figure 7 shows the results of matching the IR image using the proposed method. From left are IR images, RGB images, warped IR images, and alpha-blending results between I_{ir} and I_{rgb}, and alpha-blending results between $\tilde{I}_{ir}$ and I_{rgb}.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 September 2023 doi:10.20944/preprints202309.1577.v1

Table 2. Experiment result

$input_s$ \ $layer_d$		Unregistered	3-Layers	5-Layers	7-Layers
256×256	L_1	6.3451	2.616	2.715	2.912
	IoU	0.7757	**0.8983**	0.8924	0.8870
512×512	L_1	25.567	11.494	9.799	9.619
	IoU	0.7744	0.8911	**0.9045**	0.9031
640×480	L_1	29.057	13.707	11.191	13.867
	IoU	0.7804	0.8875	**0.9083**	0.8831

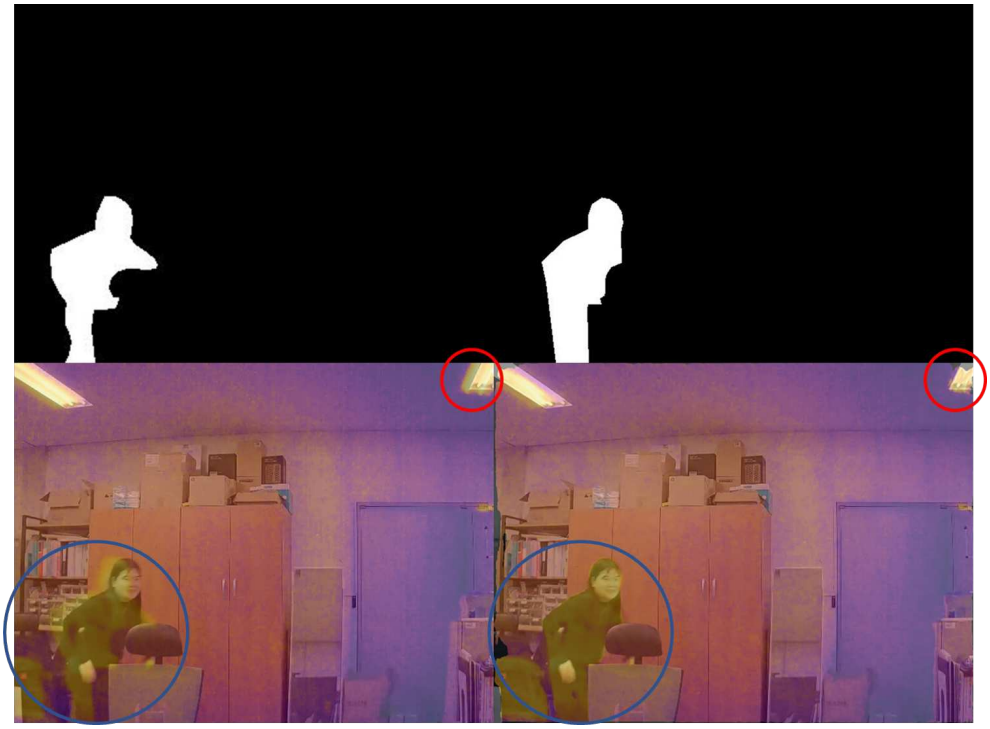

Figure 7. Unpair frame image result..This is a result of attempting to match images of frames that are not the same. The bottom-left is the result before calibration and after the bottom-right calibration. Looking at the part indicated by the blue circle, it can be seen that the correction is good even when the difference between the two objects is large. Looking at the part marked with a red circle, it can be seen that it is also corrected for fluorescent lights that do not give mask information during training.

In addition, instead of using the same frame, matching was performed on an image with a frame difference of four between the RGB and IR images. Figure 7 shows the result, and if you look at the part marked with a blue circle, it can be seen that the two images are matched, even though the human posture in the IR image and the posture in the RGB image are different. In addition, if we look at the part indicated by the orange circle, the image was corrected for lights that had not been segmented. This result indicates that the model learned the process of matching an entire image, even though a mask image segmented only one object common in the entire image was assigned for training.

5. Conclusion

In this study, a multi-modality image registration was conducted. The deformation field between the two images was inferred by receiving the IR and RGB images as input through the DFG network with two different modalities, and the IR image was corrected to match the RGB image using a grid sampler that converted the IR image based on inferred deformation fields. In addition, for the IR and RGB images, it was difficult to calculate the correction rate between the two images because they have

different modalities. However, this study solved the problem by comparing the mask images rather than directly comparing the two images. Additionally, to prevent the spatial characteristics of the original data from being excessively distorted owing to the deformation file, locality was guaranteed through an identical field, which performs the same transformation, and Smooth loss, which compares the difference values with the surrounding pixels. To show that our proposed methodology is effective for multi-image registration, we conducted an experiment. As a result of the experiment, the IoU score of the original before matching increased by 0.1279 from 0.7804 to 0.9083 based on an input size of 640×480, and matching appeared in the entire image, including areas where masking information was not provided.

Author Contributions: Conceptualization, M.L.; methodology, M.L.; software, M.L.; validation, M.L.; formal analysis, M.L.; investigation, M.L.; resources, M.L; data curation, M.L.; writing—original draft preparation, M.L.; writing—review and editing, M.L. and J.K.; visualization, M.L.; supervision, M.L.,J.K.; project administration, J.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data collected for this work is not available.

Acknowledgments: This work was supported by Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE)(20224B10100060, Development of Artificial Intelligence Vibration Monitoring System for Rotating Machinery).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, L., Ma, R., & Zakhor, A. (2022). Drone object detection using rgb/ir fusion. arXiv preprint arXiv:2201.03786.
2. Lan, X., Ye, M., Zhang, S., Zhou, H., & Yuen, P. C. (2020). Modality-correlation-aware sparse representation for RGB-infrared object tracking. Pattern Recognition Letters, 130, 12-20.
3. Zhang, X., Ye, P., Leung, H., Gong, K., & Xiao, G. (2020). Object fusion tracking based on visible and infrared images: A comprehensive review. Information Fusion, 63, 166-187.
4. Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., & Shen, D. (2018). Deep learning based inter-modality image registration supervised by intra-modality similarity. In Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9 (pp. 55-63). Springer International Publishing.
5. Wei, W., Haishan, X., Alpers, J., Rak, M., & Hansen, C. (2021). A deep learning approach for 2D ultrasound and 3D CT/MR image registration in liver tumor ablation. Computer Methods and Programs in Biomedicine, 206, 106117.
6. Lei, Y., Fu, Y., Wang, T., Liu, Y., Patel, P., Curran, W. J., ... & Yang, X. (2020). 4D-CT deformable image registration using multiscale unsupervised deep learning. Physics in Medicine & Biology, 65(8), 085003.
7. Kuppala, K., Banda, S., & Barige, T. R. (2020). An overview of deep learning methods for image registration with focus on feature-based approaches. International Journal of Image and Data Fusion, 11(2), 113-135.
8. DeTone, D., Malisiewicz, T., & Rabinovich, A. (2016). Deep image homography estimation. arXiv preprint arXiv:1606.03798.
9. Zhang, J., Wang, C., Liu, S., Jia, L., Ye, N., Wang, J., ... & Sun, J. (2020). Content-aware unsupervised deep homography estimation. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 653-669). Springer International Publishing.
10. Nie, L., Lin, C., Liao, K., Liu, S., & Zhao, Y. (2021). Depth-aware multi-grid deep homography estimation with contextual correlation. IEEE Transactions on Circuits and Systems for Video Technology, 32(7), 4460-4472.
11. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., ... & Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 2758-2766)

12. Lefébure, M., & Cohen, L. D. (2001). Image registration, optical flow and local rigidity. Journal of Mathematical Imaging and Vision, 14, 131-147.
13. Mocanu, S., Moody, A. R., & Khademi, A. (2021). Flowreg: Fast deformable unsupervised medical image registration using optical flow. arXiv preprint arXiv:2101.09639.
14. Yu, Q., Jiang, Y., Zhao, W., & Sun, T. (2022). High-Precision Pixelwise SAR–Optical Image Registration via Flow Fusion Estimation Based on an Attention Mechanism. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 3958-3971.
15. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T. (2017). Flownet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2462-2470).
16. Uzunova, H., Wilms, M., Handels, H., & Ehrhardt, J. (2017). Training CNNs for image registration from few samples with model-based data augmentation. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20 (pp. 223-231). Springer International Publishing.
17. Sun, D., Yang, X., Liu, M. Y., & Kautz, J. (2018). Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8934-8943).
18. Wu, W., Wang, Z., Li, Z., Liu, W., & Fuxin, L. (2019). Pointpwc-net: A coarse-to-fine network for supervised and self-supervised scene flow estimation on 3d point clouds. arXiv preprint arXiv:1911.12408.
19. Teed, Z., & Deng, J. (2020). Raft: Recurrent all-pairs field transforms for optical flow. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16 (pp. 402-419). Springer International Publishing.
20. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
21. Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters, 15(5), 749-753.
22. Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., & Johansen, H. D. (2019, December). Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE international symposium on multimedia (ISM) (pp. 225-2255). IEEE.
23. Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. Image and vision computing, 21(11), 977-1000.
24. Derpanis, K. G. (2004). The harris corner detector. York University, 2, 1-2.
25. Goncalves, H., Corte-Real, L., & Goncalves, J. A. (2011). Automatic image registration through image segmentation and SIFT. IEEE Transactions on Geoscience and Remote Sensing, 49(7), 2589-2600.
26. Mair, E., Hager, G. D., Burschka, D., Suppa, M., & Hirzinger, G. (2010). Adaptive and generic corner detection based on the accelerated segment test. In Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part II 11 (pp. 183-196). Springer Berlin Heidelberg.
27. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
28. Arar, M., Ginger, Y., Danon, D., Bermano, A. H., & Cohen-Or, D. (2020). Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13410-13419).