

Alignment and Fusion of Visible and Infrared Images Based on Gradient-Domain Processing

Ayaka Tanihata, Masayuki Tanaka, and Masatoshi Okutomi; Tokyo Institute of Technology, Japan

Abstract

An image fusion of different modal images, such as visible and long wavelength infrared (LWIR) images, is an important image processing technique because different modal images have compensate information for each other. Many existing image fusion algorithms assume that different modal images are perfectly aligned. However, that assumption is not satisfied in many practical situations.

In this paper, we propose an image alignment and fusion algorithm with gradient-domain processing. First, we extract the gradient information from both modality images. Then, assuming disparities between the two gradient maps, candidate gradient maps for the target fused image are generated by selecting the gradient having larger power from different modality images pixel-by-pixel. A key observation is as follows. If the assumed disparity is wrong, the fused image includes ghost edges. If the assumed disparity is correct, the single edge is preserved without the ghost edge in the fused image. Therefore, we evaluate the gradient power in the region-of-interest of the fused image with different disparities. Then, we can align images based on the disparity associated with the minimum gradient power. Finally, we apply gradient-based image fusion with the aligned image pairs.

We experimentally validate that the proposed approach can effectively align and fuse the visible and long wavelength infrared images.

Introduction

Multi-modal image sensing is becoming common in practice. Particularly, a combination of visible and long wavelength infrared (LWIR) images has great potential [17]. The visible camera is a very good tool for bright and clear scenes, while it requires lighting for a night scene, and it's difficult to see through the deep fog. In contrast, the LWIR camera is robust to a low-light condition and a foggy scene. In order to obtain the benefits of both visible and LWIR images, image fusion is highly demanded.

Figure 1 shows an example of image fusion of visible and LWIR images. The visible camera cannot capture the environment for the foggy scene as shown in Fig. 1-(a), while the LWIR camera can observe the environment through the fog as shown in Fig.1-(b). Figure 1-(c) is the fusion result which is good for clear bright scenes, foggy scenes, night scenes, and other scenes.

Many image fusion algorithms have been proposed in the literature [6]. Those existing algorithms assume that input two images are perfectly aligned. Although a coaxial camera system has been proposed [18], the alignment is not perfect in many practical situations because the viewpoints of the visible and the LWIR cameras are usually different. Therefore, even after applying the intrinsic and extrinsic camera calibration, there are disparities between the visible and the LWIR image depending on the depth of

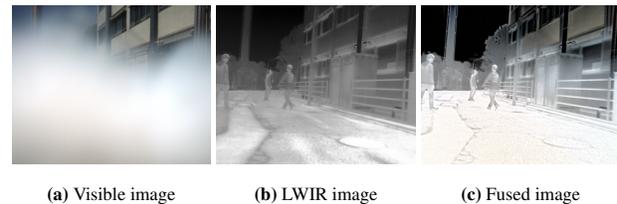


Figure 1: Example of visible and infrared image pair and its fusion result.

objects. The misalignment often generates severe artifacts such as a ghost, a halo, and a discontinuity in the fused image [23]. The alignment between different modal images is a challenging task because the local properties of the different modal images are different.

In this paper, we assume the camera calibration for the visible and thermal cameras has already been performed [22]. Then, the intrinsic and extrinsic parameters are known. In that situation, the image alignment equals to estimate depth of target objects. Even if there is one degree of freedom for the image registration, the image alignment is still a challenging problem because image properties between the visible and the LWIR images are very different. In order to solve that difficulty, we propose a gradient-domain-based approach. The image fused by our algorithm robustly provides rich information for clear bright scenes, low-light night scenes, and foggy scenes.

We use the gradient-domain based approach. There are many applications of gradient-domain approach in computer graphics and computer vision. Perez et al. [2] presented the gradient-domain interpolation framework called Poisson image editing. Image fusion techniques based on the gradient-domain were also proposed [3][4][5]. In gradient-domain image processing, first, the gradients are extracted from the input image. Then, the extracted gradients are manipulated to obtain expected effects. Finally, the image is reconstructed based on the manipulated gradients.

We propose a visible and infrared image alignment and fusion method based with gradient-domain processing. The results fused by the proposed method contain important information from both images. Experiments showed that the proposed alignment effectively works. We also demonstrated that the proposed alignment and fusion algorithm superiors to existing algorithms.

Alignment and Fusion Based on Gradient-Domain Processing

The objective of this paper is image alignment and image fusion algorithms for multi-modal images. We propose gradient-domain-based image alignment and fusion algorithms. Figure 2 shows the pipeline of the proposed method. The input of the pro-

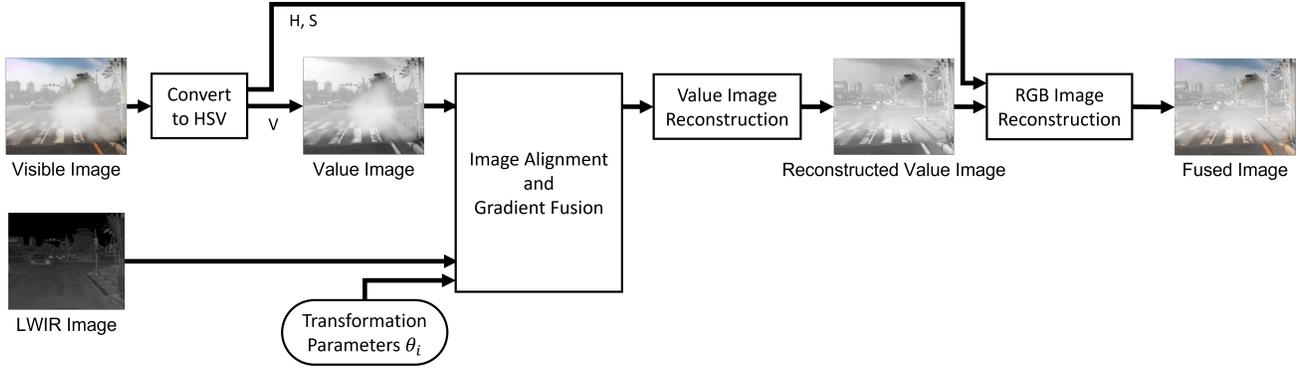


Figure 2: Processing pipeline of fusion method.

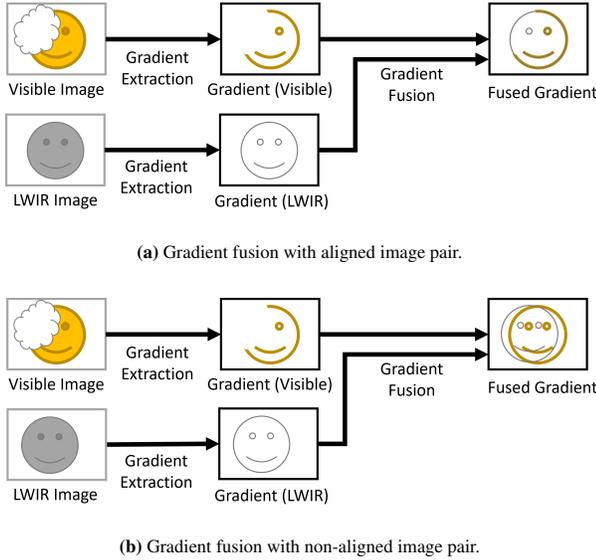


Figure 3: Schematic image of gradient fusion.

posed algorithm is visible and LWIR image pair without alignment. The output is the fused image without ghost artifacts.

First, we convert the visible image into HSV color space. We will only use the value image (V) for fusion. Then, we assume candidate transformation parameters θ_i . The LWIR image is warped with those transformation parameters. The gradient of visible image and gradient of warped LWIR images will be fused. This gradient fusion is performed by choosing the larger gradient power of the gradient pixel-by-pixel manner. For each fused gradient map, we evaluate an average gradient power in the region of interest (ROI). Then, we select the fused gradient map with the minimum average gradient power of ROI. We reconstruct the value image from the selected gradient map to obtain the final fused image. Finally, we reconstruct the RGB image by using H and S from the input visible image and V from the reconstructed value image.

Gradient Fusion

Figure 3 shows the schematic image of the proposed gradient fusion method. Gradient power $\xi_g(\mathbf{x})$ of an image $\xi(\mathbf{x})$ is given by

$$\xi_g(\mathbf{x}) = \xi_h^2(\mathbf{x}) + \xi_v^2(\mathbf{x}), \quad (1)$$

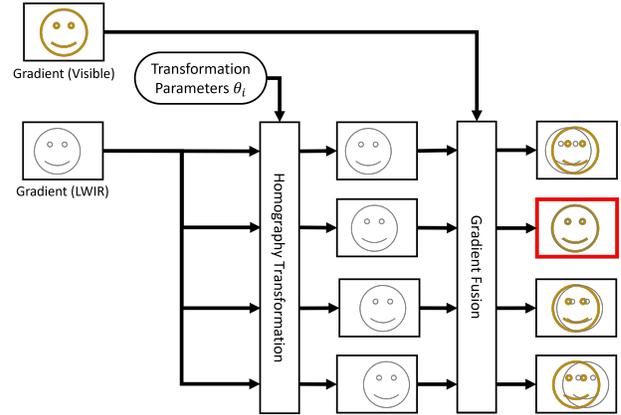


Figure 4: Processing pipeline of image alignment based on gradient power of fused gradient map, where the red box represents the selected gradient map.

where $\xi_h(\mathbf{x})$ represents the horizontal derivative of $\xi(\mathbf{x})$, and $\xi_v(\mathbf{x})$ represents the vertical derivative of $\xi(\mathbf{x})$.

Let $u^{vis}(\mathbf{x})$ and $u^{lr}(\mathbf{x})$ be the intensity of the visible and LWIR images at position \mathbf{x} . The fused gradient map $q_d^\theta(\mathbf{x})$ with the transformation parameter θ is given as

$$q_d^\theta(\mathbf{x}) = \begin{cases} u_d^{vis}(\mathbf{x}) & u_g^{vis}(\mathbf{x}) \geq u_g^{lr}(T(\mathbf{x}; \theta)) \\ u_d^{lr}(T(\mathbf{x}; \theta)) & u_g^{vis}(\mathbf{x}) < u_g^{lr}(T(\mathbf{x}; \theta)) \end{cases}. \quad (2)$$

where $u_g^{vis}(\mathbf{x})$ represents gradient power of the visible image, $u_g^{lr}(\mathbf{x})$ represents gradient power of the LWIR image, $T(\mathbf{x}; \theta)$ represents the warping function with the parameter θ , and d represents the direction of the horizontal or vertical gradient. This gradient fusion is performed by choosing the larger gradient power of the gradient pixel-by-pixel manner. This enables to preserve important edges.

Image Alignment

The image alignment in this paper is to find the suitable transformation parameter θ for the region-of-interests (ROI). The transformation parameter θ^* is given as

$$\theta^* = \arg \min_{\theta} \int_{\Omega} \left([q_h^\theta(\mathbf{x})]^2 + [q_v^\theta(\mathbf{x})]^2 \right) d\mathbf{x}, \quad (3)$$

where Ω represents the region of interest (ROI).

The main idea behind the proposed algorithm is as follows. We will obtain ghost artifacts, or duplicated edges, in the fused gradient map if the assumed depth is wrong as shown in Fig. 3-(b). On the other hand, if the assumed depth is correct, there are no ghost artifacts as shown in Fig. 3-(a). Our observation is that the ghost artifacts increase the gradient power of the ROI. Based on that observation, we choose the gradient map by evaluating the average gradient power of the fused gradient map in the ROI. Figure 4 shows the schematic of the image alignment based on the gradient power of the fused images.

Image Reconstruction

There are many image fusion techniques based on gradient-domain. In this paper, we use image reconstruction based on [1]. Image reconstruction is performed by optimizing the energy functional $E[u(\mathbf{x})]$ given as

$$E[u(\mathbf{x})] = F[u(\mathbf{x})] + R[u(\mathbf{x})], \quad (4)$$

where $F[u(\mathbf{x})]$ is the the gradient fidelity term, and $R[u(\mathbf{x})]$ is the the intensity-range constraint term.

The gradient fidelity term is defined as

$$F[u(\mathbf{x})] = \int \sum_{d=h,v} |u_d(\mathbf{x}) - q_d^{\theta^*}(\mathbf{x})|^2 d\mathbf{x}. \quad (5)$$

This term minimizes the gradient residual between the target and the reconstructed gradients in the same manner as the existing gradient-domain approach.

The gradient intensity-range constraint is defined as

$$R[u(\mathbf{x})] = \int \gamma(u(\mathbf{x})) d\mathbf{x}, \quad (6)$$

$$\gamma(\eta) = \begin{cases} \infty & \eta < R_{\min} \\ 0 & R_{\min} \leq \eta \leq R_{\max} \\ \infty & \eta > R_{\max} \end{cases}, \quad (7)$$

where R_{\max} and R_{\min} is the upper and lower bound defined by a fixed target range. This term prevents the output image being overexposed or underexposed.

Experiments

We have taken several scenes with the visible and LWIR cameras. The intrinsic and extrinsic calibrations were performed beforehand [22]. We will show qualitative and quantitative comparisons.

Qualitative comparison

Figure 5 shows a comparison of fused images with and without alignment. Figure 5 (a) and (b) are observed visible and LWIR images. From those images, we can find the disparity between those images. Then, we apply the proposed gradient-based image alignment. Figure 5 (c) is the LWIR image aligned by the proposed method. We can find the position of the fence in the zoomed region of the aligned LWIR image is moved to that position of the LWIR image. Figure 5 (d) and (e) are fusion results without alignment and with the proposed alignment. We can see the ghost artifact in Fig. 5 (d), while there are no ghost artifacts in Fig. 5 (e) of the image fused by the proposed method.

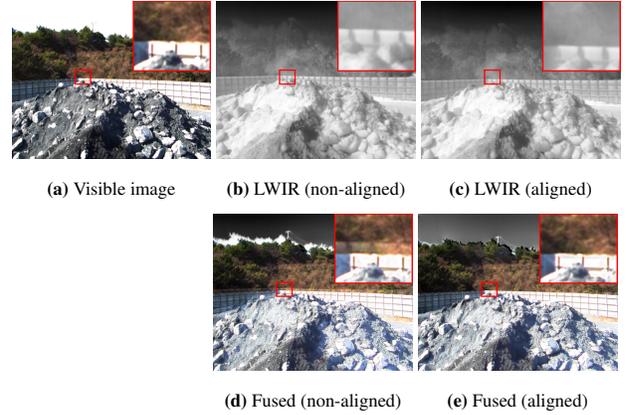


Figure 5: Input images and fusion results by proposed method. Red boxes represent zoomed regions.

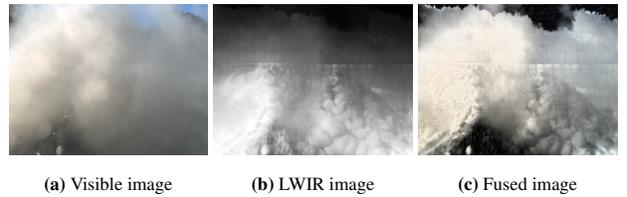


Figure 6: Example of visible and infrared image pair and its fusion result at a hazy scene.

We applied the proposed algorithm for several practical scenes. Figure 6 shows an example of a hazy scene, which has a sand storm at a construction site in a mountain. The visible image of Fig. 6 (a) shows a whiteout in which we can see nothing. Even the human eye cannot see anything. However, we can recognize pit gravel through the haze from the LWIR image of Fig. 6 (b). Fig. 6 (c) shows the image fused by the proposed algorithm. From that fused image, we can also observe pit gravel through the haze. Figure 7 shows an example of a night scene.

The proposed alignment and fusion algorithm outputs the image for the bright and clear scene as in Fig. 1 (e). For the dense haze scene, the result of the proposed algorithm is Fig. 6 (c). For the night scene, the result of the proposed algorithm is Fig. 9 (c). Those results demonstrate that the proposed alignment and fusion algorithm robustly provide good information for clear, haze and night scenes.

Figure 8 shows the visual result comparison with other methods. The input images are not aligned. The visible camera and LWIR camera were placed in a row. Camera calibration was done and results are used in our method. We compared with 3 other methods, CNN [19], GTF [20] and MST_SR [21]. We can see that our method effectively align images.

Quantitative comparison

In the proposed algorithm, we fuse the gradient map with assuming disparity candidates. If the assumed disparity candidate is wrong, the resultant fused gradient map includes ghost edges. Then, those ghost edges increase the average gradient power of the ROI. Figure 9 shows examples of visible and LWIR images warped with wrong and correct disparities and their associated gradient maps. We can find ghost edges from Fig. 9 (h) of the gradient map fused with wrong disparity, while ghost edges are

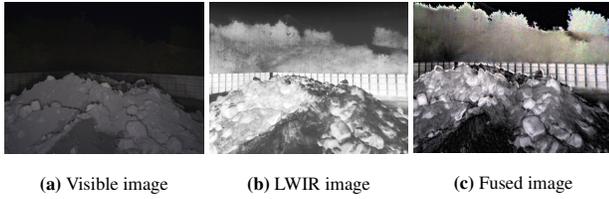


Figure 7: Example of visible and infrared image pair and its fusion result at a night scene.

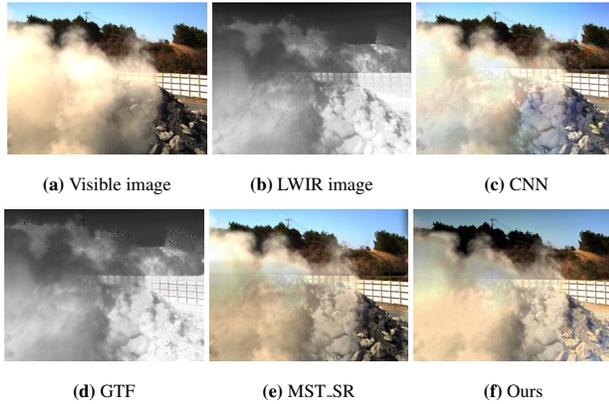


Figure 8: Visual results comparison between different methods.

suppressed in Fig. 9 (g) of the gradient map fused with correct disparity. We also quantitatively evaluate the average gradient power of Fig. 9 (h) and (g). Those values are 0.0407 and 0.0365, respectively. Based on this simple observation, we can estimate the disparity by finding the minimum gradient power of the fused gradient map.

Table 1 shows the comparison of fusion results by non-aligned image pairs and aligned image pairs using proposed method. We used images from [7], which contains 21 pairs of aligned visible and LWIR images [8][9][10]. In order to make non-aligned images, we assumed that the optic axes of two cameras are aligned as parallel, and shifted all LWIR images $[-10, 0]$.

We employed 6 evaluation methods, cross entropy (CE) [11], mutual information (MI) [12], PSNR [13], $Q^{AB/F}$ [14], Q_{CV} [15], and SSIM [16]. We used [7] to calculate these methods. Non-shifted LWIR images are used for source images. Cross entropy (CE) shows the similarity between source images and fused image. Mutual information (MI) are used to measure the amount of information that is transferred from source images to the fused image. PSNR indicates the ratio of peak value power and noise power in the fused image. $Q^{AB/F}$ shows the amount of edge information that is transferred from source images to fused image. Q_{CV} is a quality factor obtained based on human vision system.

Table 1: Comparison of fusion results on non-aligned images and aligned images with proposed method. The better results are highlighted in bold.

Method	CE	MI	PSNR	$Q^{AB/F}$	Q_{CV}	SSIM
Non-Aligned	1.28	1.92	57.30	0.54	1052.92	1.17
Aligned	1.25	2.05	57.48	0.59	852.90	1.30

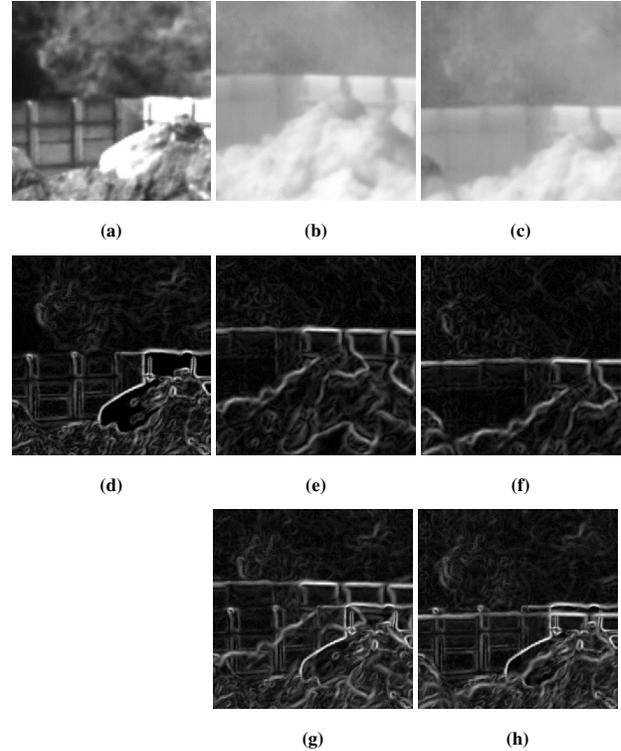


Figure 9: Examples of visible and LWIR images warped with wrong and correct disparities, and their associated gradient maps, where the brighter (white) represents the larger gradient power.

SSIM models image loss and distortion. Table 1 shows that the results with alignment are better than those without alignment. Especially, $Q^{AB/F}$ got better after alignment. This result means that edge information are transferred from the source images to the fused result, which suggests that alignment is effectively done.

Conclusion

In this paper, we proposed image alignment and image fusion algorithm based on gradient-domain processing. The output fused images are easy to recognize and have fewer ghost artifacts. The proposed algorithm is robust to many scenes such as bright scenes, low-light night scenes, and hazy scenes.

References

- [1] Shibata, Takashi and Tanaka, Masayuki and Okutomi, Masatoshi, Gradient-Domain Image Reconstruction Framework With Intensity-Range and Base-Structure Constraints, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [2] P. Perez, M. Gangnet, and A. Blake, Poisson image editing, ACM Trans. on Graphics (TOG), 22(3), pg. 313–318, 2003.
- [3] R. Raskar, A. Ilie, and J. Yu, Image fusion for context enhancement and video surrealism, Proc. of Int. Symp. on NonPhotorealistic Animation and Rendering (NPAR), 7, pg. 85–152, 2004.
- [4] D. Connah, M. S. Drew, and G. D. Finlayson, Spectral edge image fusion: Theory and applications, Proc. of European Conf. on Computer Vision (ECCV), pg. 65–80, 2014.
- [5] J. Sun, H. Zhu, Z. Xu, and C. Han, Poisson image fusion based on markov random field fusion model, Information Fusion, 14(3), pg. 241–254, 2013.

- [6] Jiayi Ma, Yong Ma, Chang Li, Infrared and visible image fusion methods and applications: A survey, *Information Fusion*, 45, pg. 153-178, 2019.
- [7] X. Zhang, P. Ye, G. Xiao, VIFB: A visible and infrared image fusion benchmark, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, 2020.
- [8] C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang, Rgb-t object tracking: benchmark and baseline, *Pattern Recognition*, p106977, 2019.
- [9] J. W. Davis and V. Sharma, Background-subtraction using contour-based fusion of thermal and visible imagery, *Computer vision and image understanding*, vol. 106, no. 2-3, pp. 162–182, 2007.
- [10] C. O’Conaire, N. E. O’Connor, E. Cooke, and A. F. Smeaton, Comparison of fusion methods for thermo-visual surveillance tracking, *9th International Conference on Information Fusion*, IEEE, pp. 1–7, 2006.
- [11] D. M. Bulanon, T. Burks, and V. Alchanatis, Image fusion of visible and thermal images for fruit detection, *Biosystems Engineering*, vol. 103, no. 1, pp. 12–22, 2009.
- [12] G. Qu, D. Zhang, and P. Yan, Information measure for performance of image fusion, *Electronics letters*, vol. 38, no. 7, p. 313–315, 2002.
- [13] P. Jagalingam and A. V. Hegde, A review of quality metrics for fused image, *Aquatic Procedia*, vol. 4, no. Icwrcoc, pp. 133–142, 2015.
- [14] C. S. Xydeas and P. V. V., Objective image fusion performance measure, *Military Technical Courier*, vol. 36, no. 4, pp. 308–309, 2000.
- [15] H. Chen and P. K. Varshney, A human perception inspired quality metric for image fusion based on regional information, *Information fusion*, vol. 8, no. 2, pp. 193–207, 2007.
- [16] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli et al., Image quality assessment: from error visibility to structural similarity, *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [17] Soonmin Hwang and Jaesik Park and Namil Kim and Yukyung Choi and In So Kweon, Multispectral Pedestrian Detection: Benchmark Dataset and Baselines, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [18] Yuka Ogino, Takashi Shibata, Masayuki Tanaka and Masatoshi Okutomi, Coaxial visible and FIR camera system with accurate geometric calibration, *Proceedings of SPIE Defense + Commercial Sensing (DCS2017)*, Vol.10214, pp.1021415-1-6, April, 2017.
- [19] Yu Liu, Xun Chen, Juan Cheng, Hu Peng, and Zengfu Wang, Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(03):1850018, 2018
- [20] Jiayi Ma, Chen Chen, Chang Li, and Jun Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, *Information Fusion*, 31:100–109, 2016.
- [21] Yu Liu, Shuping Liu, and Zengfu Wang, A general framework for image fusion based on multi-scale transform and sparse representation, *Information Fusion*, 24:147–164, 2015.
- [22] T. Shibata, M. Tanaka, and M. Okutomi, Accurate Joint Geometric Camera Calibration of Visible and Far-Infrared Cameras, *Electronic Imaging, Image Sensors and Imaging Systems*, pp. 7-13(7), 2017.
- [23] T. Shibata, M. Tanaka, and M. Okutomi, Misalignment-Robust Joint Filter for Cross-Modal Image Pairs *IEEE International Conference on Computer Vision (ICCV)*, 2017.

Author Biography

Ayaka Tanihata received her bachelor’s degree from Tokyo University of Science, Tokyo, Japan, in 2020. She is currently a master’s student at Tokyo Institute of Technology.

Masayuki Tanaka received his Ph.D. degree from Tokyo Institute of Technology, Tokyo, Japan in 2003 and joined Agilent Technology. He was a research scientist at Tokyo Institute of Technology from 2004 to 2008, a visiting scholar at Stanford University from 2013 to 2014. He is currently an associate professor at Tokyo Institute of Technology.

Masatoshi Okutomi received his master’s degree from Tokyo Institute of Technology in 1983 and joined Canon Inc., Tokyo, Japan. He was a visiting research scientist at Carnegie Mellon University, PA, USA from 1987 to 1990. He received his Ph.D. degree from Tokyo Institute of Technology in 1993. He is currently a Professor at Tokyo Institute of Technology.