Misalignment-Robust Joint Filter for Cross-Modal Image Pairs (Supplemental materials)

1. Supplemental video

A supplemental video is available on **https://youtu.be/jS7oVqqx6cY**

2. Joint-filter cost volume

As described in our paper, the recent cross-modal joint filters are designed based on the cost functions. In this section, we describe the joint-filter cost volume for these filters. Let $\mathbf{t} = (t_1, \cdots, t_i, \cdots, t_N)^T$ and $\mathbf{g} = (g_1, \cdots, g_i, \cdots, g_N)^T$ be the vectorized target image and the vectorized original guidance image, where i is the pixel index. The set of the translated guidances are given by $\{\mathbf{H}_1\mathbf{g}, \cdots, \mathbf{H}_k\mathbf{g}, \cdots, \mathbf{H}_K\mathbf{g}\}$, where $\mathbf{H}_k$ is the matrix operator for k-th labeled translation vector.

Guided Filter (GF) [3]: As described in our paper, the joint-filter cost volume $\mathbf{C}^{jf}$ for GF [3] is designed based on the residual term in Eq. (8) in our original paper as

$$c^{\text{jf}}_{i,k} = \sum_{l\in\mathcal{N}_i}\Big(a_{k,i}(\mathbf{H}_k\mathbf{g})_l + b_{k,i} - t_l\Big)^2, \tag{1}$$

where $c^{\text{jf}}_{i,k}$ is the joint-filter cost volume at k-th label and i-th pixel, $a_{k,i}$ and $b_{k,i}$ are the linear transformation coefficients for the k-th labeled translated guidance at the pixel i. Note that the proposed method is the general framework for the cross-modal joint filters.

Mutual Structure for Joint Filtering (MSJF) [8]: The cost function for the original MSJF [8] is given by

$$\begin{aligned} E_{MSJF}(\tilde{\mathbf{t}}, \tilde{\mathbf{g}}, a, b) &= \lambda_t||\tilde{\mathbf{t}} - \mathbf{t}|| + \lambda_g||\tilde{\mathbf{g}} - \mathbf{g}|| \\ &+ \sum_{l\in\mathcal{N}_i}\Big(a^t_i g_l + b^t_i - t_l\Big)^2 \\ &+ \sum_{l\in\mathcal{N}_i}\Big(a^g_i t_l + b^g_i - g_l\Big)^2 \\ &+ \varepsilon_g {a^g_i}^2 + \varepsilon_t {a^t_i}^2, \end{aligned}$$

where λ_t, λ_g, a^g_i, a^t_i, b^g_i and b^t_i are the parameters for the data term and the linear coefficient corresponding to MSJF [8], $\tilde{\mathbf{t}}$ and $\tilde{\mathbf{g}}$ are the filtered results by MSJF [8] , ε_g and ε_t are the smoothing parameters. Here, the first and the second terms are the data term, the third and the fourth terms are the structure-similarity term, and the last two terms are the smoothing term, respectively. In the proposed method, the joint-filter cost volume for MSJF [8] with the k-th translated guidance image $\mathbf{H}_k\mathbf{g}$ is designed based on the data term and the structure-similarity term as

$$\begin{aligned} c^{\text{jf}}_{i,k} &= \lambda_t||\tilde{\mathbf{t}}_k - \mathbf{t}|| + \lambda_g||\tilde{\mathbf{g}}_k - (\mathbf{H}_k\mathbf{g})_l|| \\ &+ \sum_{l\in\mathcal{N}_i}\Big(a^t_{i,k}(\mathbf{H}_k\mathbf{g})_l + b^t_{i,k} - t_l\Big)^2 \\ &+ \sum_{l\in\mathcal{N}_i}\Big(a^g_{i,k} t_l + b^g_{i,k} - (\mathbf{H}_k\mathbf{g})_l\Big)^2 \end{aligned} \tag{2}$$

where λ_t, λ_g, $a^g_{k,i}$, $a^t_{k,,i}$, $b^t_{k,i}$ and $b^g_{k,i}$ are the parameters for the data term and the linear coefficient corresponding to MSJF [8], $\tilde{\mathbf{t}}_k$ and $\tilde{\mathbf{g}}_k$ are the filtered results for the target $\mathbf{t}$ and the k-th translated guidance $\mathbf{H}_k\mathbf{g}$ obtained by MSJF [8] .

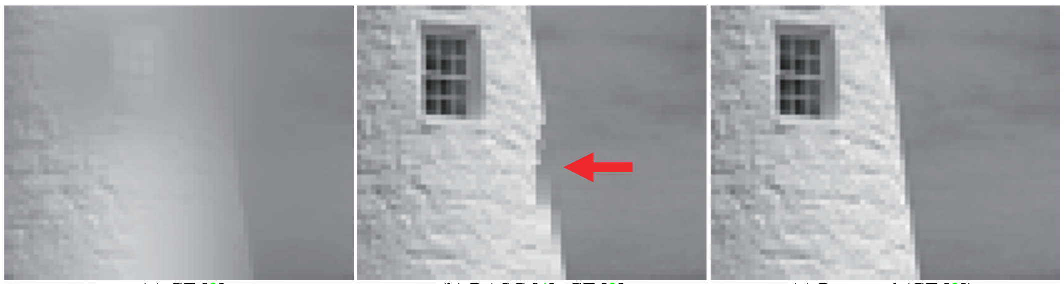
Joint Bilateral Filter (JBF) [7]: The recent study of the image filter [6, 2] also showed that the classical guidance filter, *e.g.* JBF [7], can be formulated as the cost function minimization by the kernel function. Based on these studies, we can also define the joint-filter cost volume for these classical joint filters. For example, the joint-filter cost volume for JBF [7] is given by

$$\begin{aligned} c^{\mathrm{jf}}_{i,k} &= \sum_l ((\mathbf{H}_k\mathbf{g})_i - t_l)^2 K_{i.l,(\mathbf{H}_k\mathbf{g})_i,(\mathbf{H}_k\mathbf{g})_l}, \\ K_{i.l,(\mathbf{H}_k\mathbf{g})_i,(\mathbf{H}_k\mathbf{g})_l} &= \exp\left[-\frac{(\mathbf{x}_i - \mathbf{x}_l)^2}{2\sigma_s^2} - \frac{((\mathbf{H}_k\mathbf{g})_i - (\mathbf{H}_k\mathbf{g})_l)^2}{2\sigma_i^2}\right], \end{aligned} \tag{3}$$

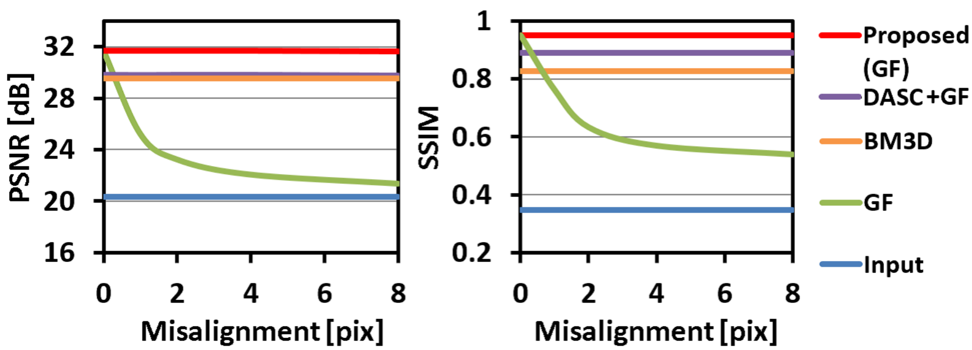
where σ_s and σ_i are the spatial similarity and the range (intensity/color) similarity, $\mathbf{x}_i$ and $\mathbf{x}_l$ are the position of i-th and l-th pixel, respectively. Note that we can apply the proposed framework to the existing joint filters such as the scale map image restoration [9] and the spatial-spectral objective function for the dark flash photography [5], if the cost function of the joint filter is defined pixel-by-pixel.

(a) Target (R channel) (b) Guidance (G channel)

Figure 1. Robustness to misalignment between target and guidance images. (It's same to Fig. 5 (a) and (b) in our paper.)



(a) GF [3] (b) DASC [4]+GF [3] (c) Proposed (GF [3])



(d) Quantitative evaluation by PSNR (left) and SSIM (right).

Figure 2. Robustness of the proposed method with GF [3]. The results (a), (b), and (c) are generated from the input image obtained by adding the 8-pixel horizontal shift. (σ=25)

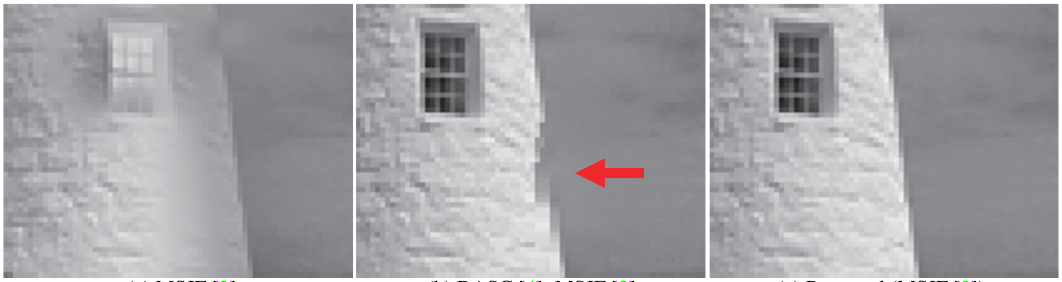
3. Additional results and whole images in our paper

In this section, we show the additional results and the whole images in our paper.

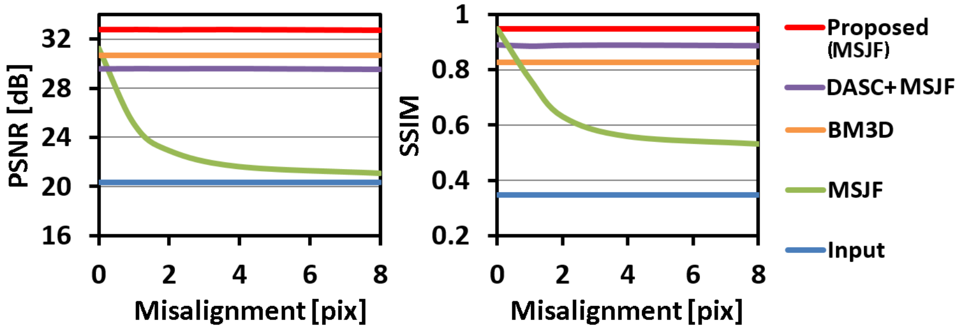
3.1. Additional results for image denosing

First, we show the additional results described in Sec. 4. 1 in our paper. To evaluate the performance of the existing joint filters with the cross-modal registration, we also used the aligned guidance image by applying DASC [4], which is a state-of-the-art registration method for cross-modal images. Note that DASC [4] is sensitive to noise, the target images are denoised [1] before applying DASC. As shown in Fig. 2 (a), the blurred output is generated by naively applying GF [3]. Although the blur artifacts can be reduced by applying DASC [4], the discontinuity artifacts due to the registration error are generated (Fig. 2 (b)). On the other hand, Fig. 2 (c) shows that the proposed method can remove the noise without the blur and the discontinuity artifacts.

Next, we also evaluated the performance by measuring PSNR and SSIM. As shown in Fig. 2 (d), PSNR and SSIM are dramatically decreased by naively applying GF [3] due to the blur artifacts. The proposed method can maintain the PSNR and SSIM while the misalignment exists between the target and the guidance images. Note that, although the PSNR and SSIM become to be stable by applying DASC [4], this approach cannot outperform the proposed method in terms of PSNR and SSIM.



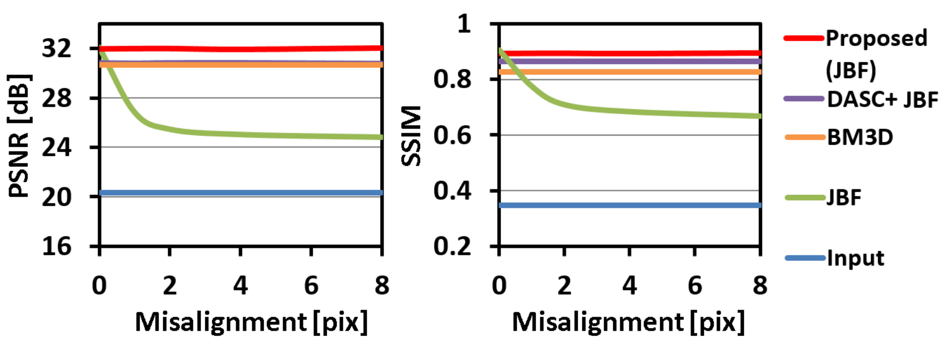
(a) MSJF [3] (b) DASC [4]+MSJF [3] (c) Proposed (MSJF [3])



(d) Quantitative evaluation by PSNR (left) and SSIM (right) .

Figure 3. Robustness of the proposed method with MSJF [3]. The results (a), (b), and (c) are generated from the input image obtained by adding the 8-pixel horizontal shift. (σ=25)

(a) JBF [3] (b) DASC [4]+JBF [3] (c) Proposed (JBF [3])



(d) Quantitative evaluation by PSNR (left) and SSIM (right) .

Figure 4. Robustness of the proposed method with JBF [3]. The results (a), (b), and (c) are generated from the input image obtained by adding the 8-pixel horizontal shift. (σ=25)

Finally, we evaluated the performance of the proposed framework with JBF [7] and MSJF [8] using the joint-filter cost volume described in Eq. (2) and Eq. (3). As shown in Fig. 3 and 4, the proposed framework is effective for JBF [7] and MSJF [8].

3.2. Whole images in our paper

In our original paper, almost figures are close-up images due to the space limitations. We show the whole images of these figures in this section.

Figure 5 shows the whole image shown in Fig. 6 in our original paper. As shown in Fig. 5, the proposed method can remove the noise effectively while reducing the artifacts due to the misalignment. Similarly, Fig. 6, Fig. 7, and Fig. 8 shows the whole image shown in Fig. 7, Fig. 8, and Fig. 9 in our original paper.

References

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. *In Proc. of Int. Conf. on Image Processing (ICIP)*, 2007. 3, 6

[2] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Trans. on Image Processing (TIP)*, 11(10):1141–1151, 2002. 2

[3] K. He, J. Sun, and X. Tang. Guided image filtering. *In Proc. of European Conf. on Computer Vision (ECCV)*, 2010. 1, 3, 4, 6, 7, 8

[4] S. Kim, D. Min, B. Ham, S. Ryu, M. N. Do, and K. Sohn. Dasc: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence. *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015. 3, 4, 6, 7

[5] D. Krishnan and R. Fergus. Dark flash photography. *ACM Trans. on Graphics (TOG)*, 28(3)(96), 2009. 2

[6] P. Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30(1):106–128, 2013. 2

[7] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. on Graphics (TOG)*, 23(3):664–672, 2004. 2, 5, 6, 7

[8] X. Shen, C. Zhou, L. Xu, and J. Jia. Mutual-structure for joint filtering. *In Proc. of IEEE Int. Conf. on Computer Vision(ICCV)*, 2015. 1, 2, 5, 6, 7

[9] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia. Cross-field joint image restoration via scale map. *In Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, 2013. 2

(a)Target(VIS) (b)Guidance(NIR)

(c) Ground truth (d) BM3D [1]

(e) JBF [7] (f) DASC [4]+JBF [7] (g)Proposed (JBF [7])

(h) MSJF[8] (i) DASC [4]+MSJF[8] (j)Proposed(MSJF[8])

(k) GF [3] (l) DASC [4]+GF [3] (m) Proposed(GF [3])

Figure 5. Visible image denosing guided by NIR image. ($\sigma = 50$)

(a)Target (FIR)

(b)Guidance(VIS)

(c)Ground truth

(d)Bicubic

(e) JBF[7]

(f) DASC [4]+JBF[7]

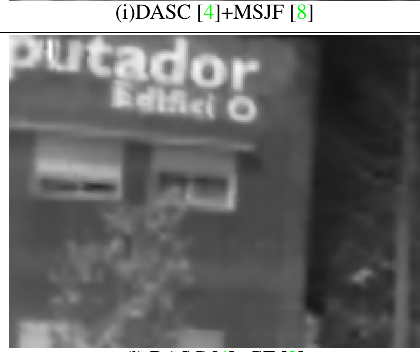
(g) Proposed(JBF[7])

(h) MSJF [8]

(i)DASC [4]+MSJF [8]

(j)Proposed(MSJF [8])

(k) GF [3]



(l) DASC [4]+GF [3]

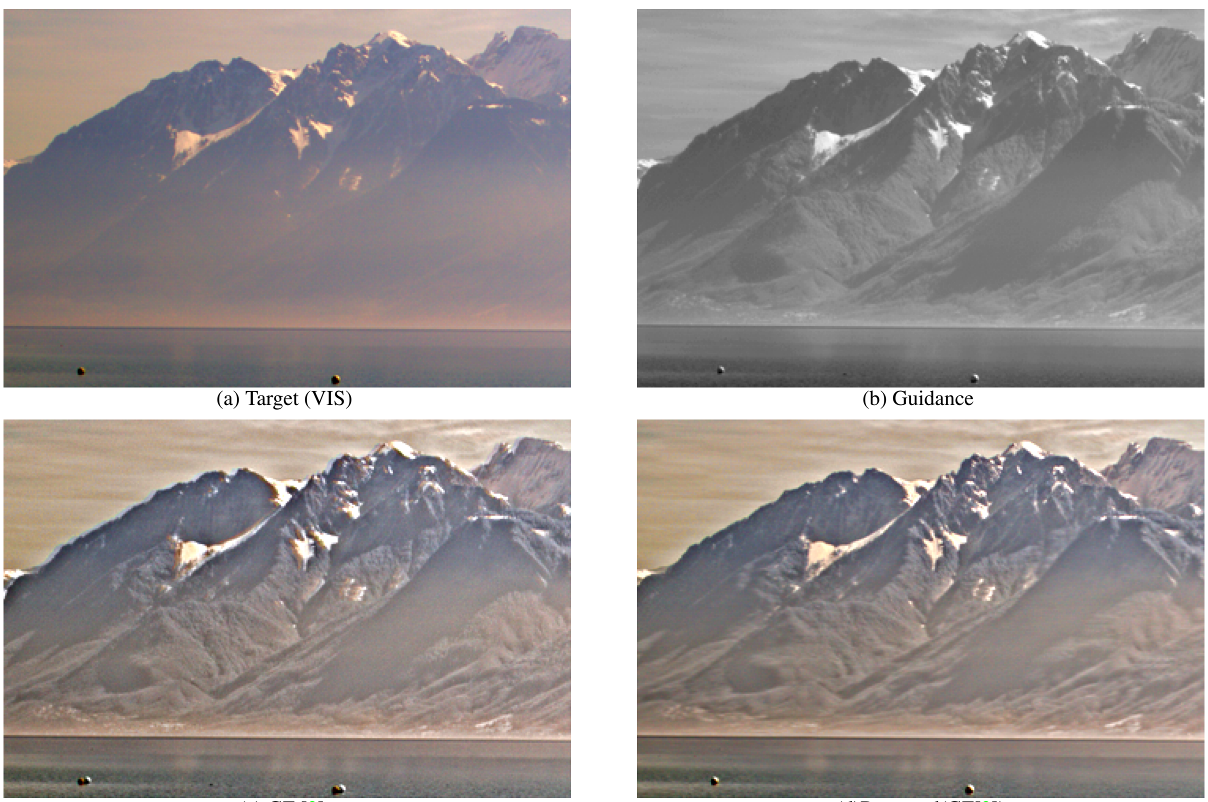
(m) Proposed (GF [3])

Figure 6. FIR image up-sampling guided by visible image. ($\times 4$)

(a) Target (b) Guidance

(c) GF [3] (d)Proposed(GF[3])

Figure 7. Flash/no-flash photography .



(a) Target (VIS) (b) Guidance

(c) GF [3] (d)Proposed(GF[3])

Figure 8. Results for haze removal using visible and NIR images.