Towards Practical and Efficient High-Resolution HDR Deghosting with CNN

Anonymous ECCV supplementary submission

Paper ID 3792

Abstract. In this supplementary file, we provide the additional ablation on different downsample factors and qualitative results. We also provide the detailed description of the N_g network model used in our method.

1 Notations

In Table 1, we provide a brief description of notations used in this supplementary and in the main paper.

Table 1. Description of notations used in this supplementary file and in the main paper.

Notation	Description
$\square^h$	Original full resolution image
$\square^l$	Image downsampled by factor of d
N_{mc}	Motion correction network
N_g	Guide image generation network
N_f	Final HDR generation network
$\{I^h_-, I^h_0, I^h_+\}$	Full resolution input images
$\{S^h_-, S^h_0, S^h_+\}$	Ground truth motion segmentation maps as mentioned in Section 4.1 of the main paper
$\{\widetilde{M}^h_-, \widetilde{M}^h_+\}$	Combined ground truth motion segmentation maps as described in Section 4.1 of the main paper
$\{M^h_-, M^h_+\}$	Motion segmentation maps predicted by N_{mc} (equation 1 and 2 of the main paper)
$\{I^h_{mc,-}, I^h_0, I^h_{mc,+}\}$	Motion corrected images
$\{w^h_-, w^h_0, w^h_+\}$	N_g output weight maps (equation 6 of the main paper)
H_g	Guide image generated using equation 7 of the main paper
H_f	Final HDR output from N_f
$\widetilde{H}$	Ground truth HDR image

2 N_g model

We use U-net model for both N_{mc} and N_g network. The architecture details are given in Table 2. All the convolutional layers use 3×3 kernel with LeakyRelu activation func-

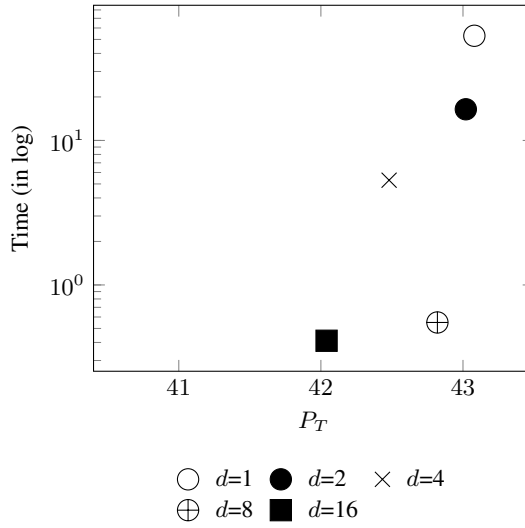


Fig. 1. Tonemapped PSNR vs Time plot for d = 1,2,4,8 and 16. The scores are obtained on test set of UCSD dataset [1]. The images in the UCSD dataset has resolution of 1000×1500.

tion. The last layer has one channel output for N_{mc} model and three channel output for N_g model with sigmoid activation function.

Table 2. The network architecture used for N_{mc} and N_g. Conv2d(n) denotes convolutional layer with n filters. And, Maxpool(k) denotes max pooling with pool size of (k, k). Upsample2d(m) denotes upsampling by size (m, m).

Tag	x_1		x_2		x_3		x_4																									
Input	Conv2d(32)	Conv2d(32)	MaxPool(2)	Conv2d(64)	Conv2d(64)	MaxPool(2)	Conv2d(128)	Conv2d(128)	MaxPool(2)	Conv2d(256)	Conv2d(256)	MaxPool(2)	Conv2d(512)	Upsample2d(2)	Concat(x_4)	Conv2d(512)	Conv2d(512)	Upsample2d(2)	Concat(x_3)	Conv2d(256)	Conv2d(256)	Upsample2d(2)	Concat(x_2)	Conv2d(128)	Conv2d(128)	Upsample2d(2)	Concat(x_1)	Conv2d(64)	Conv2d	Output

3 Ablation on d

In Figure 1, we show the performance of our proposed model on various downsampling factors (d). While d=1 method achieves the best performance, the time taken for processing in full resolution is in the range of 54 seconds for 1000×1500 resolution images. Comparatively, downsampling by a factor of 8 has equal balance between speed and accuracy. Downsampling further to 16, reduces the accuracy significantly. It should

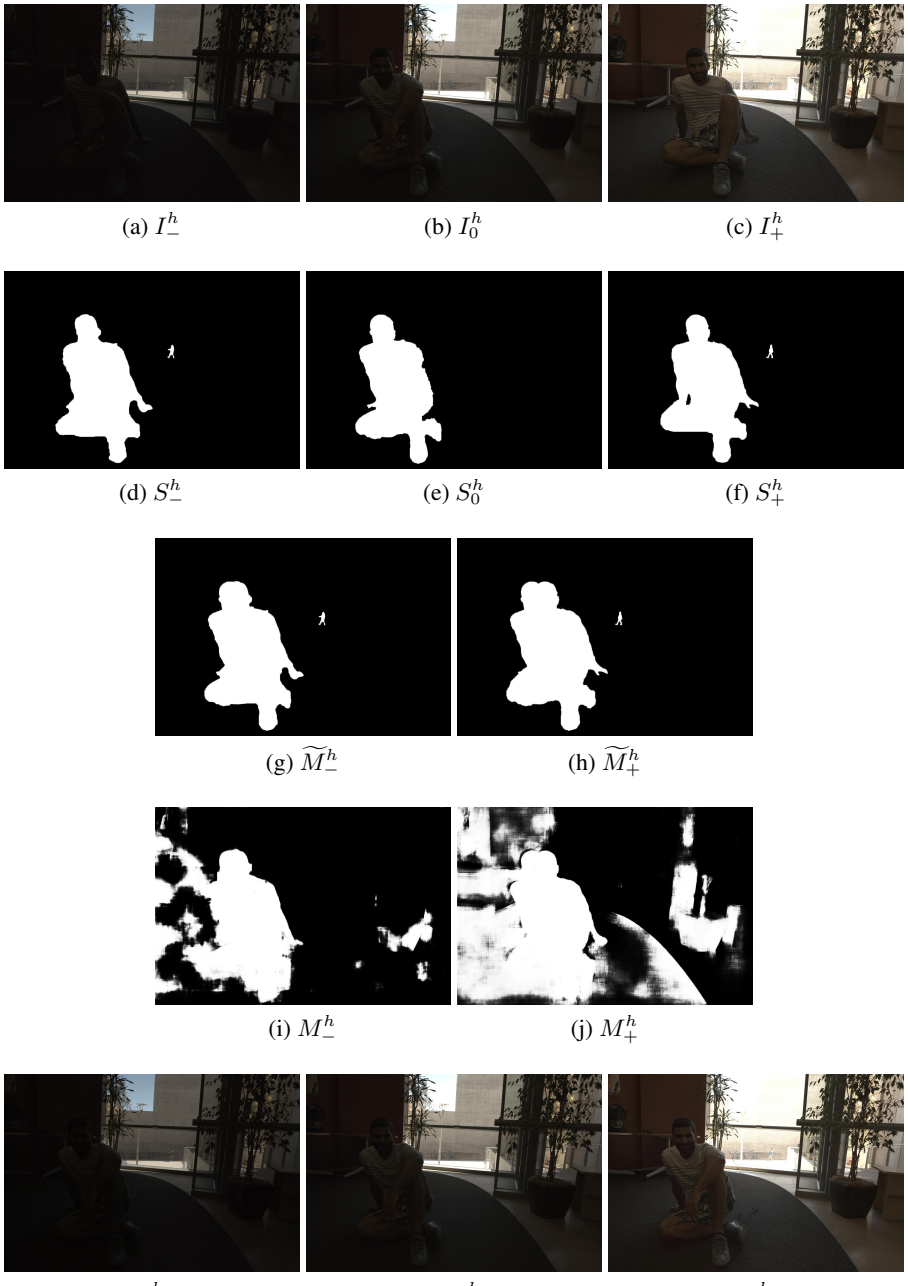
be noted that, major time consuming part for d=2 and 4 is the camera alignment and optical flow registration.

4 Motion segmentation dataset

In Figures 2, 3, 4, 5 and 6, we show examples of annotated ground truth segmentation maps, predictions from trained N_{mc} model and the motion corrected sequence. As seen in Figures 2 and 5 (i) and (j), the network has predicted more regions in addition to moving objects. This is due to the fact, the input images are not perfectly aligned with I_0. Due to small camera motion misalignment, the network identified those regions as moving. From these examples, we observe that N_{mc} model can successfully handle small alignment errors as well.

5 Additional results

In Figures 7 to 12, we show the qualitative comparison between proposed method against state-of-the-art methods.



(a) I^h_- (b) I^h_0 (c) I^h_+

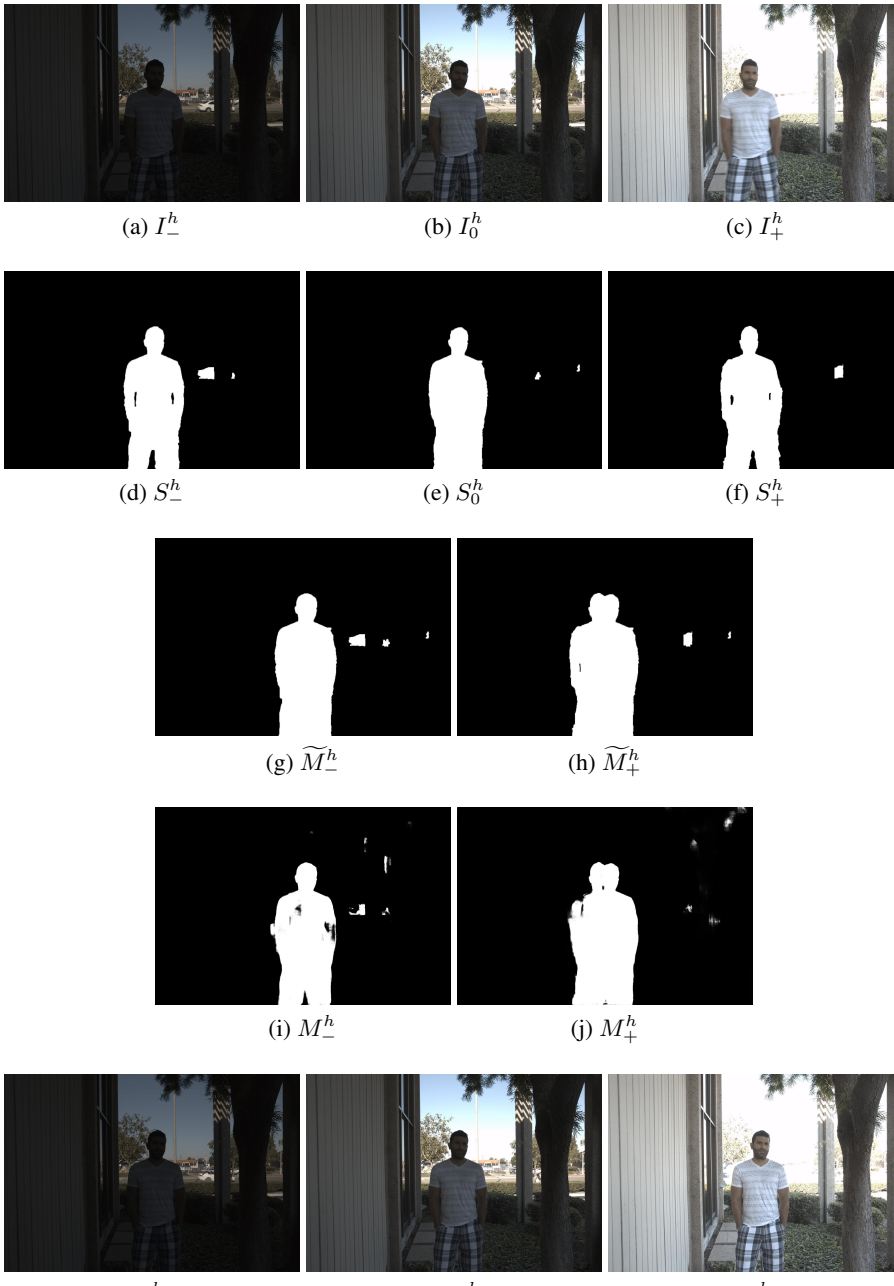
(d) S^h_- (e) S^h_0 (f) S^h_+

(g) $\widetilde{M}^h_-$ (h) $\widetilde{M}^h_+$

(i) M^h_- (j) M^h_+

(k) $I^h_{mc,-}$ (l) I^h_0 (m) $I^h_{mc,+}$

Fig. 2. An example of annotated motion segmentation maps for images from UCSD dataset [1].



(a) I^h_- (b) I^h_0 (c) I^h_+

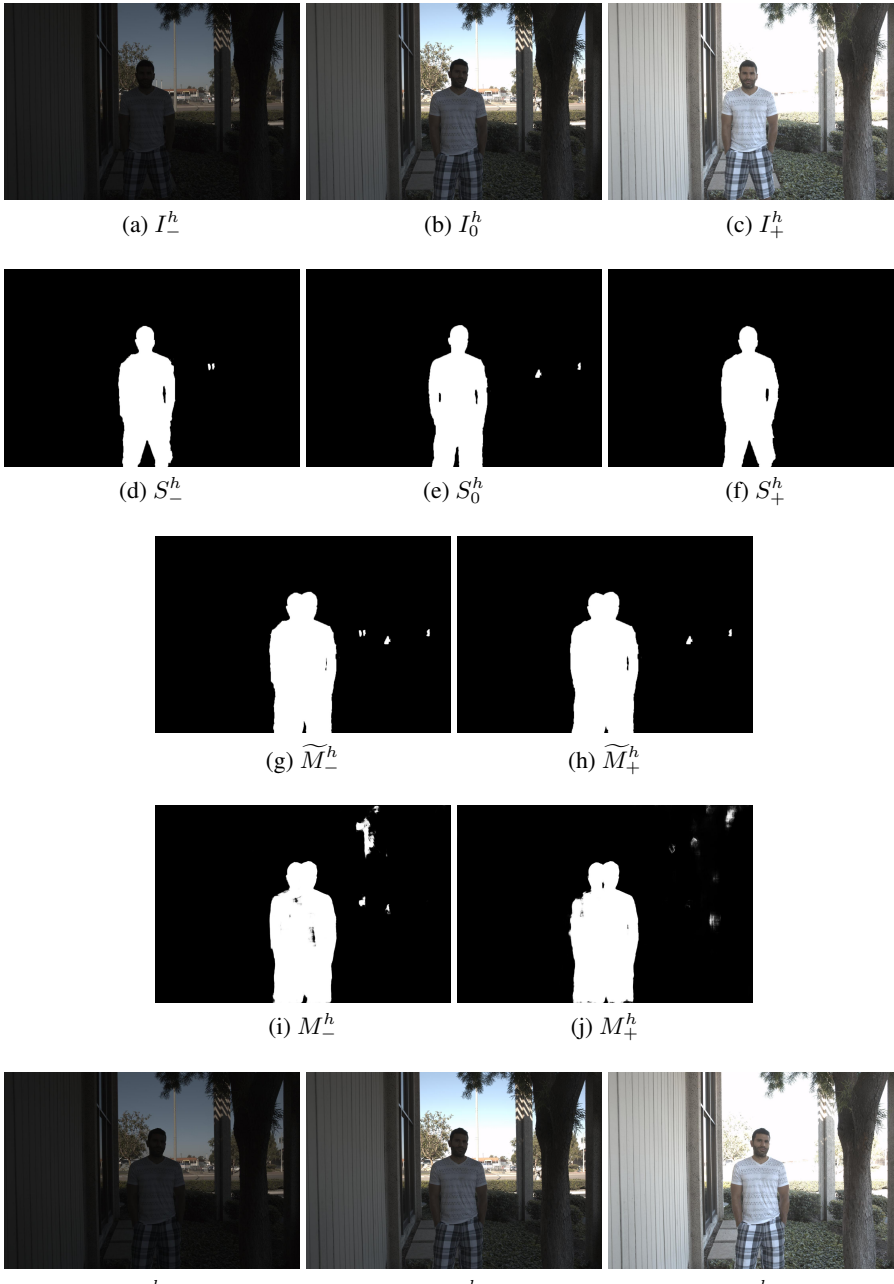
(d) S^h_- (e) S^h_0 (f) S^h_+

(g) $\widetilde{M}^h_-$ (h) $\widetilde{M}^h_+$

(i) M^h_- (j) M^h_+

(k) $I^h_{mc,-}$ (l) I^h_0 (m) $I^h_{mc,+}$

Fig. 3. An example of annotated motion segmentation maps for images from UCSD dataset [1].



(a) I^h_- (b) I^h_0 (c) I^h_+

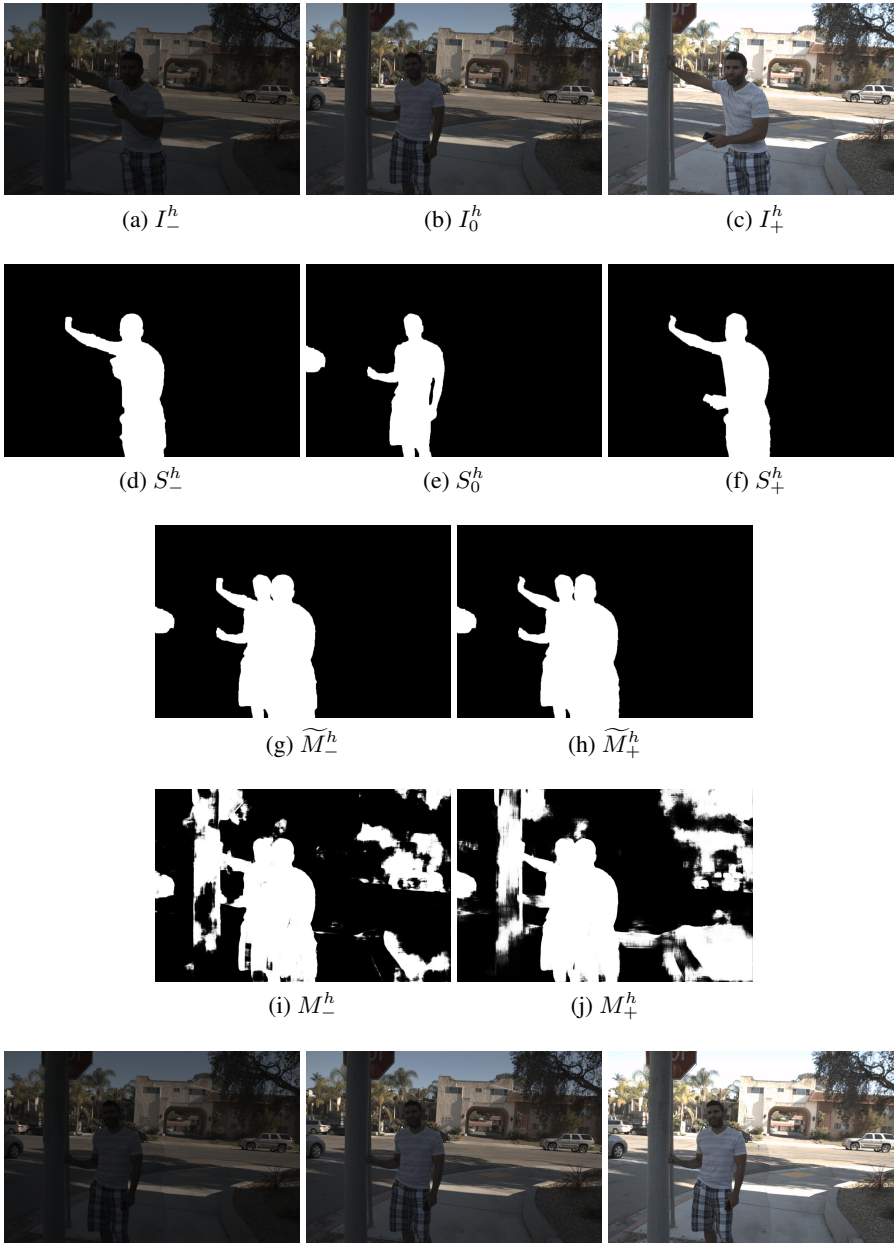
(d) S^h_- (e) S^h_0 (f) S^h_+

(g) $\widetilde{M}^h_-$ (h) $\widetilde{M}^h_+$

(i) M^h_- (j) M^h_+

(k) $I^h_{mc,-}$ (l) I^h_0 (m) $I^h_{mc,+}$

Fig. 4. An example of annotated motion segmentation maps for images from UCSD dataset [1].



(a) I_{-}^{h} (b) I_{0}^{h} (c) I_{+}^{h}

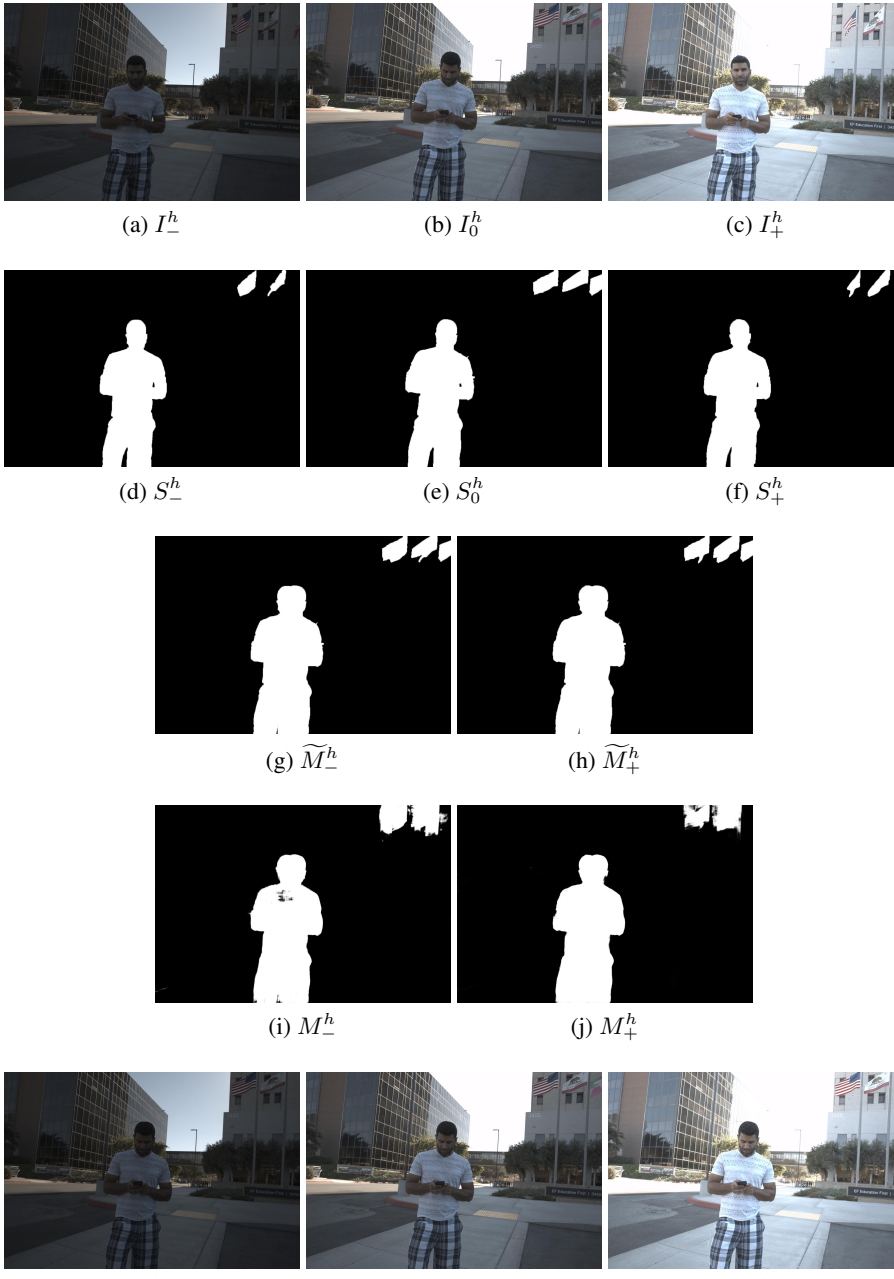
(d) S_{-}^{h} (e) S_{0}^{h} (f) S_{+}^{h}

(g) $\widetilde{M}_{-}^{h}$ (h) $\widetilde{M}_{+}^{h}$

(i) M_{-}^{h} (j) M_{+}^{h}

(k) $I_{mc,-}^{h}$ (l) I_{0}^{h} (m) $I_{mc,+}^{h}$

Fig. 5. An example of annotated motion segmentation maps for images from UCSD dataset [1].



(a) I^h_- (b) I^h_0 (c) I^h_+

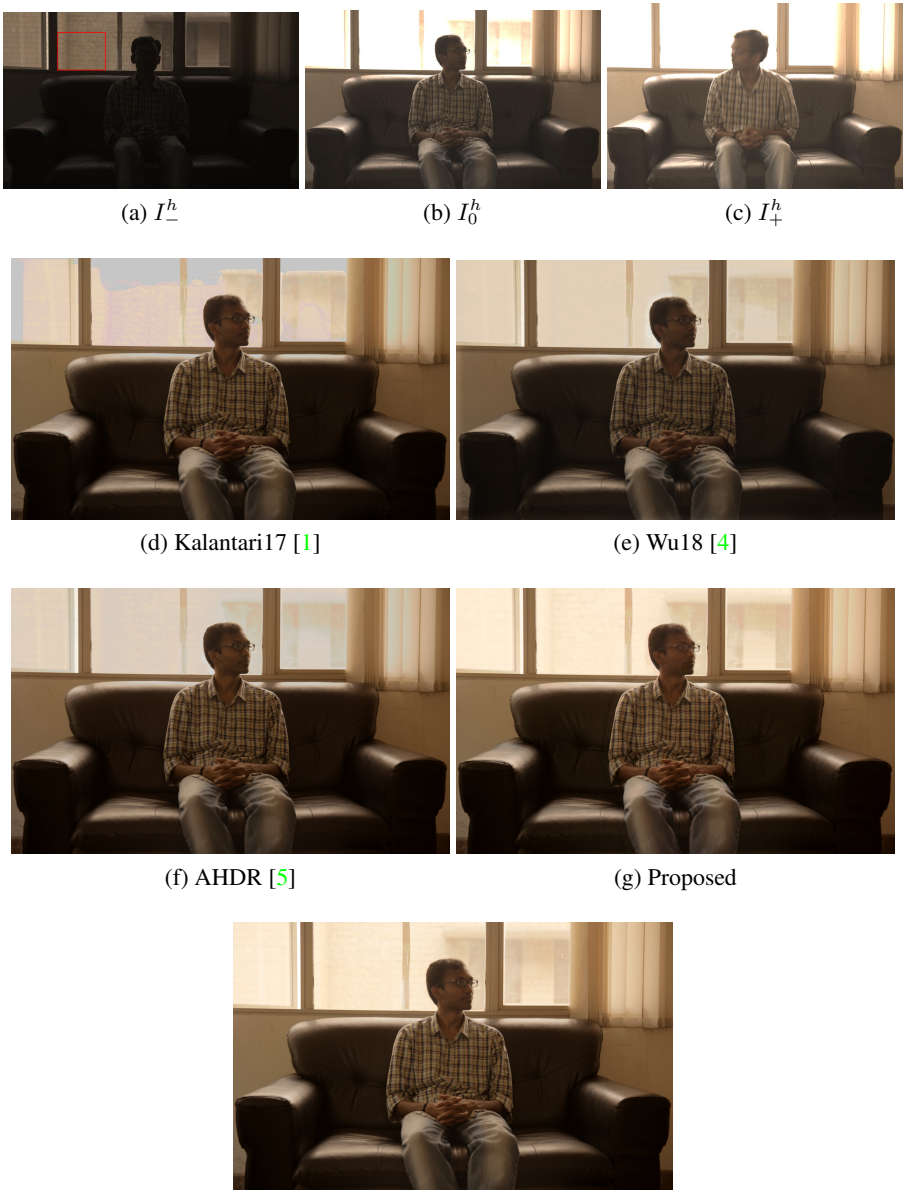
(d) S^h_- (e) S^h_0 (f) S^h_+

(g) $\widetilde{M}^h_-$ (h) $\widetilde{M}^h_+$

(i) M^h_- (j) M^h_+

(k) $I^h_{mc,-}$ (l) I^h_0 (m) $I^h_{mc,+}$

Fig. 6. An example of annotated motion segmentation maps for images from UCSD dataset [1].

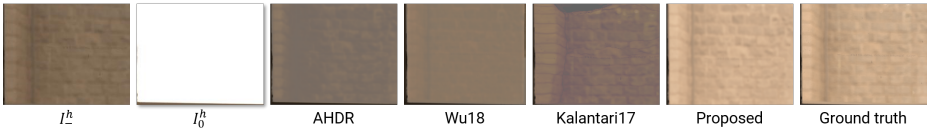


(a) I_{-}^{h} (b) I_{0}^{h} (c) I_{+}^{h}

(d) Kalantari17 [1] (e) Wu18 [4]

(f) AHDR [5] (g) Proposed

(h) Ground truth



(i) cropped regions in linear domain

Fig. 7. Qualitative comparison on an example from [2] dataset.

(a) I_{-}^{h}

(b) I_{0}^{h}

(c) I_{+}^{h}

(d) Kalantari17 [1]

(e) Sen12 [3]

(f) SCHDR [2]

(g) Proposed

(h) Ground truth

Fig. 8. Qualitative comparison on an example from [1] dataset.

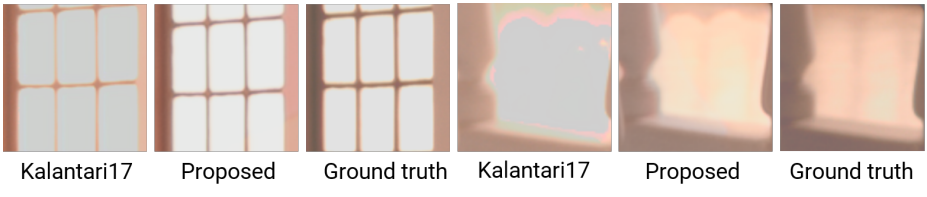
(a) I^h_-

(b) I^h_0

(c) I^h_+

(d) Kalantari17 [1]

(e) Proposed



(f) Cropped regions

Fig. 9. Qualitative comparison on an example from [2] dataset.

(a) I_{-}^{h}

(b) I_{0}^{h}

(c) I_{+}^{h}

(d) Kalantari17 [1]

(e) AHDR [5]

(f) SCHDR [2]

(g) Wu18 [4]

(h) Proposed

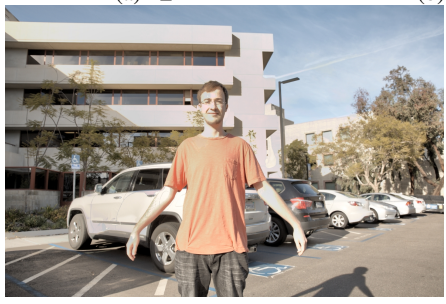
(i) Ground truth

Fig. 10. Qualitative comparison on an example from [1] dataset.

(a) I^h_-

(b) I^h_0

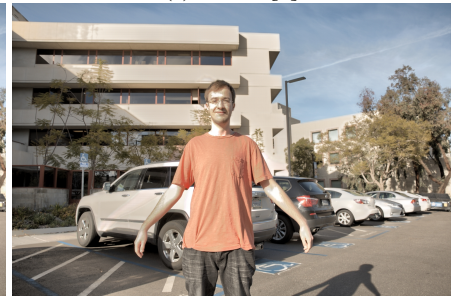
(c) I^h_+



(d) Kalantari17 [1]

(e) AHDR [5]

(f) SCHDR [2]



(g) Wu18 [4]

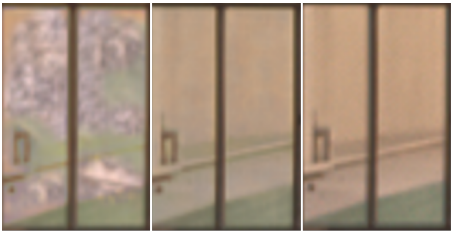
(h) Proposed

(i) Ground truth

(j) Cropped regions in linear domain. Stripped texture is preserved in Kalantari17 and proposed method, while Wu18 and AHDR has smoothed the details.

Fig. 11. Qualitative comparison on an example from [1] dataset.

(a) Sen12 [3] (b) Proposed (c) Ground truth

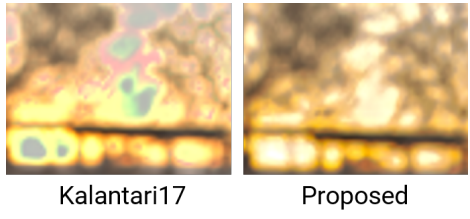


(d) Cropped regions

Fig. 12. Qualitative comparison on an example from [2] dataset.

(a) Kalantari17 [1]

(b) Proposed



(c) Cropped regions

Fig. 13. Qualitative comparison on an example from Sen *et al.* [3] dataset.

References

1. Kalantari, N.K., Ramamoorthi, R.: Deep high dynamic range imaging of dynamic scenes. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017) **36**(4) (2017) 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15
2. Prabhakar, K.R., Arora, R., Swaminathan, A., Singh, K.P., Babu, R.V.: A fast, scalable, and reliable deghosting method for extreme exposure fusion. In: 2019 IEEE International Conference on Computational Photography (ICCP). pp. 1–8. IEEE (2019) 9, 10, 11, 12, 13, 14
3. Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E.: Robust patch-based HDR reconstruction of dynamic scenes. ACM Trans. Graph. **31**(6), 203 (2012) 10, 14, 15
4. Wu, S., Xu, J., Tai, Y.W., Tang, C.K.: Deep high dynamic range imaging with large foreground motions. In: European Conference on Computer Vision. pp. 120–135 (2018) 9, 12, 13
5. Yan, Q., Gong, D., Shi, Q., van den Hengel, A., Shen, C., Reid, I., Zhang, Y.: Attention-guided network for ghost-free high dynamic range imaging. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1751–1760 (2019) 9, 12, 13