Embedding Novel Views in a Single JPEG Image
Supplementary Material

Yue Wu* HKUST
Guotao Meng* HKUST
Qifeng Chen HKUST

1. Experiment details.

During training, our network is supervised by ground-truth MPI generated from stereo images.

2. Evaluation details.

When comparing with ground-truth images, we use 10^{th} 12^{th} as source frames. And render using poses of 6^{th}, 8^{th}, 14^{th}, 16^{th} 18^{th} frames for evaluation.

3. Comparison to Facebook App

The Facebook App can not be compared quantitatively since they only release part of the code. Our model produces visually better novel views than the Facebook App **as shown in Figure**

4. Comparison to LDI and "Two views"

We compare our method with LDI [2] and "Two views." "Two views" embeds another view into the reference image using Mono3D [1] and then renders novel views from the restored images. Its results are in the PNG format since Mono3D only releases its model in the PNG format.

	SSIM/PSNR	Speed (s)	Model (M)
LDI	0.8426/25.735	42.3	438
Two views (PNG)	0.9016/27.639	0.448	264
Ours (PNG)	0.8953/27.198	0.017	6.3

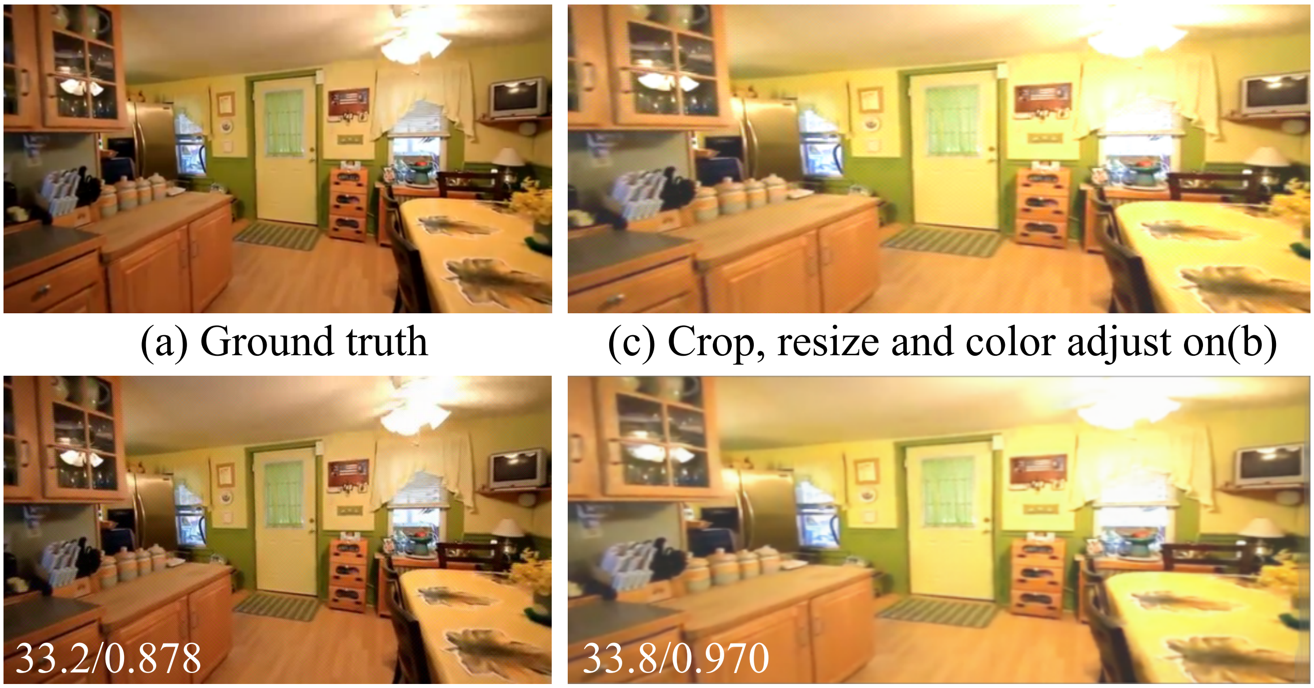
Table 1. The Comparison between our method and LDI [2] in terms of render quality

Our model renders fast and uses a small network while maintaining high rendering quality. (1) The advantage of embedding MPI is that MPI stores content in occlusion areas. In contrast, occlusion areas are inpainted by the neural network by LDI and thus often less realistic. Our model is generally more stable, while the performance of LDI depends on the image content. (2) The rendering speed of "Two views" is slow, and its model size is large since "Two views" method needs to compute MPI first for rendering.

5. Robustness against image manipulations

We conducted the ablation study against various image manipulations.

Then, we demonstrate the robustness of our model image manipulations, which means MPI can be restored from a retouched embedding image. As shown in Figure 1, we first randomly crop and resize the embedding image, then adjust the hue, brightness, contrast, and saturation of the image. The decoder can still restore the MPIs from the embedding image stably.



(a) Ground truth (c) Crop, resize and color adjust on(b)

(b) Embedding Image (d) A novel view rendered from (c)

Figure 1. The demonstration of the robustness of our method. An embedding image that is cropped, resized, and color adjusted still can be decoded into MPIs for novel view synthesis. The color of novel views (d) is changed adaptively according to the embedding image after retouching (c).

JPEG Compression. We apply JPEG compression within different JPEG quality factors to the embedding image, then use our restoration network to render novel views. Quality 100 means no compression is applied. As shown in Table 2, we compare the performance of rendered novel views from restored MPI and that from ground-truth MPI, the performance decrease as the quality factor decreases. This is because, as the factor decrease, more information is lost. However, our rendering performance is still stable and visually pleasing when quality is 75.

Quality (%)	100	90	80	75
SSIM/PSNR	0.980/38.0	0.975/36.7	0.925/30.8	0.887/28.4

Table 2. The effect of JPEG compression

Cropping and color adjusting. We apply cropping and random color adjusting to the embedding image then compare the rendering quality. After the color adjustment, the rendering quality drops from 0.975/36.68 to 0.958/34.73. After the cropping, the rendering quality keeps as 0.975/36.7. This indicates that our method is robust.

6. The influence of number and the number of MPI planes.

Influence of number of MPI planes. As shown in the Table 3, using 32 MPI planes is optimal. Increasing the number of MPI planes will make the network difficult to extract important information.

The influence of resolution of MPI planes. As shown in Table 4, when the MPI resolution is changed to 1X, 1.25X, 1.5X, the SSIM of the embedding images is always around 0.89, and that of rendered novel views is always around 0.975. Our method is not sensitive to image resolution.

Number	Embedding			Render		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
128	0.8452	28.568	0.2676	0.9163	28.813	0.2163
64	0.8950	29.831	0.208	0.9533	31.987	0.1411
32	0.9593	36.736	0.0951	0.9533	32.840	0.0951

Table 3. The influence of number of MPI planes. Experiment conducted on PB-MPI dataset.

Resolution	Embedding			Render		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
1.5X	0.8894	35.461	0.2127	0.975	37.300	0.07022
1.25X	0.8919	35.144	0.1933	0.974	37.149	0.0635
1.0X	0.8941	34.616	0.1736	0.975	36.683	0.0535

Table 4. The influence of resolution of MPI planes. Experiment conducted on Stereo Dataset.

7. Comparison to single image MPI method.

This comparison is to illustrate the strengths of our method. Given a single image (i.e., embedding image in our framework and reference image in single view framework),our model can produce better novel views.

During training, round-truth MPIs used for training are generated from S-MPI [3], the embedding performance is0.89/34.3, the rendering quality is 0.80/23.6 which is close to the performance of directly run S-MPI, 0.81/23.8.

References

[1] Wenbo Hu, Menghan Xia, Chi-Wing Fu, and Tien-Tsin Wong. Mononizing binocular videos. *ACM Transactions on Graphics*, 39(6):228:1–228:16, December 2020. 1

[2] Meng-Li Shih, Shih-Yang Su, Johannes Kopf, and Jia-Bin Huang. 3D photography using context-aware layered depth inpainting. In *CVPR*, 2020. 1

[3] Richard Tucker and Noah Snavely. Single-view view synthesis with multiplane images. In *CVPR*, 2020. 3