High-Performance Graphics (2021)
N. Binder and T. Ritschel (Editors)

Supplementary Material for Multi-Resolution Shared Representative Filtering for Real-Time Depth Completion

Yu-Ting Wu[1] Tzu-Mao Li[2] I-Chao Shen[3] Hong-Shiang Lin[4] Yung-Yu Chuang[1]

[1] National Taiwan University, Taiwan
[2] Massachusetts Institute of Technology CSAIL, United States
[3] The University of Tokyo, Japan
[4] FIH Mobile Limited, Taiwan

In this document, we first present the detailed process for generating synthetic data in Section 1. Next, Section 2 showcases the visualization of mean absolute errors of four synthetic cases. In Section 3, we show detailed comparisons on Middlebury dataset [SHK*14]. Finally, we provide the parameters of our method and all compared approaches used for the images shown in the main paper and this document in Section 4.

1. Generation of Synthetic Data

We describe the implementation details of our synthetic data generation process using the Unity game engine. Given a 3D virtual scene, our goal is to generate depth maps that look like the ones captured by Intel RealSense RGB-D cameras. Figure 1 shows our setting. The virtual camera used to capture depth images is placed with D offset to the camera that captures color images. We use $D = 6$ centimeters.

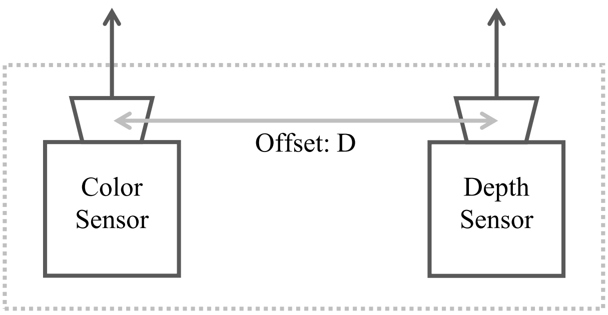

Figure 1: *Setting of our virtual cameras for generating the RGB-D pair of synthetic input. The depth camera is placed with D offset to the color camera.*

To generate input data, we render the color and reference depth images with the camera pose of the color sensor. For the raw depth map, we first render the 3D world-space position of the surfaces into a floating buffer with the depth sensor's camera pose, then we reproject the 3D data to the view of the color sensor for the aligned depth map.

To simulate the holes due to specular or glossy reflections and depth ranges, for each pixel, we first compute a *glossiness* term and a distance term. Shiny materials and larger distances result in larger glossiness and distance terms, respectively. We then multiply the two terms with a random value sampled from a Perlin noise map. Finally, if the multiplied value is larger than a predefined threshold, we invalidate that pixel's depth value.

2. Error Visualization of Synthetic Data

Figure 2 visualizes the mean absolute errors of the four synthetic cases. For all images, we map small errors to blue color and large errors to red color. We clamp the maximum error to 0.1 times the maximum depth in the scene for better visualization. As shown in the figure, our methods (SRF and M-SRF) significantly reduce the errors compared to all alternatives.

3. Comparisons on Middlebury 2014 Data

We compare the multi-resolution joint bilateral upsampling (M-JBU) [RSD*12], fast bilateral solver (FBS) [BP16], and our method (M-SRF) on the Middlebury 2014 dataset [SHK*14]. Middlebury 2014 dataset provides high-quality RGB-D image pairs with ground truth depth values; however, their depth maps do not contain large holes. To evaluate the performance of depth completion algorithms, we generate depth holes by using Perlin noise map. That is, we remove the depth values in depth maps if the pixel values at the same image coordinate in Perlin noise map are larger than a predefined threshold. To cover more possible variations of missing depth values, we use two different Perlin noise maps to generate the input depth maps. Figure 3, Figure 4, Table 1 and Table 2 show the mean absolute error (MAE) and peak signal-to-noise ratio (PSNR) of the 23 cases. The lowest MAE and the highest PSNR values are shown with italics. As shown from the table, our method

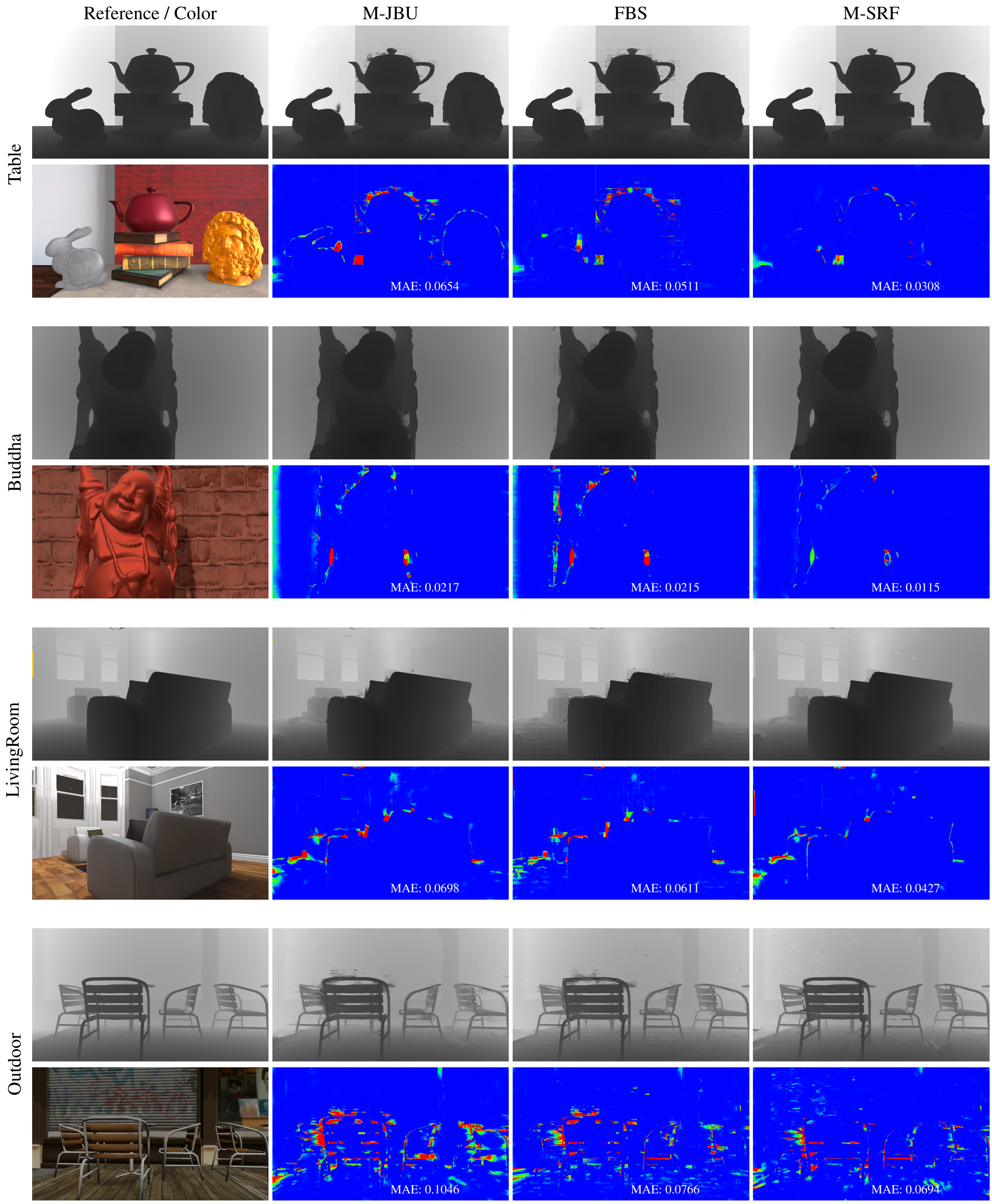

Figure 2: *We show the mean absolute error (MAE) visualization of the four synthetic cases. We allocate 7.5 ms for M-JBU and M-SRF. The Python version fast bilateral solver (FBS) [BP16] takes about 0.3 to 0.5 second, running on CPU. Its results are still listed for quality comparison. The mean absolute error (MAE, or L_1 error) is denoted for each depth map. Our method dramatically reduces the error over previous methods (the last column).*

(M-SRF) produces the lowest MAE for all 23 cases and the highest PSNR in 20 cases.

Figure 5 shows the comparisons of eight cases. Our method (M-SRF) can better preserve depth discontinuities, such as the boundaries of recycling bin in *Recycle* and monitor in *Vintage*. The structures of fine geometry are also better recovered, as shown in the grids in *Cable* and *Sword2*, the plants in *Jadeplant* and *Sword1*, and the thin objects in *Couch* and *Playroom*. For other cases, please refer to the attached files.

4. Parameters

For Joint bilateral filtering (JBF) [KCLU07] and M-JBU, we first tune the spatial standard deviation (σ_s) to meet a given time budget, then we tune the color standard deviation (σ_c) for optimizing the error. For our SRF and M-SRF, we fixed most parameters as follows: the spatial standard deviation for representative searching is set to 0.12 times the image width of the coarsest level. The distance between adjacent non-local samples along a sampling line is set to 0.05 times the image width of the coarsest level. The patch standard deviation (σ_p) is fixed to 0.1. The spatial standard deviation (σ_s) for the depth reconstruction is fixed to 1.5. We tuned the number of sampling lines K to meet a given time budget and the color standard deviation (σ_c) for optimizing the error. We additionally optimized the number of image pyramid levels (level) for multi-resolution methods (M-JBU and M-SRF). For FBS, we fixed the maximum iteration of the conjugate gradient solver to 25. We then obtained the best set of parameters including the chroma support (σ_c), spatial support (σ_{xy}), luma support (σ_l), and the smoothness multiplier (λ) by running 200 iterations of Bayesian optimization [Nog]. Table 3 and Table 4 list the final parameters of all methods for generating the results shown in the main paper and this document, respectively.

References

[BP16] BARRON J. T., POOLE B.: The fast bilateral solver. In *Europeon Conference on Computer Vision* (2016), pp. 617–632. 1, 2

[KCLU07] KOPF J., COHEN M. F., LISCHINSKI D., UYTTENDAELE M.: Joint bilateral upsampling. *ACM Trans. Graph. (Proc. SIGGRAPH) 26*, 3 (2007). 3

[Nog] NOGUEIRA F.: Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–. URL: https://github.com/fmfn/BayesianOptimization. 3

[RSD*12] RICHARDT C., STOLL C., DODGSON N. A., SEIDEL H.-P., THEOBALT C.: Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos. *Comput. Graph. Forum (Proc. Eurographics) 31*, 2 (2012), 247–256. 1

[SHK*14] SCHARSTEIN D., HIRSCHMÜLLER H., KITAJIMA Y., KRATHWOHL G., NEŠIĆ N., WANG X., WESTLING P.: High-resolution stereo datasets with subpixel-accurate ground truth. In *German conference on pattern recognition* (2014), pp. 31–42. 1, 4, 7

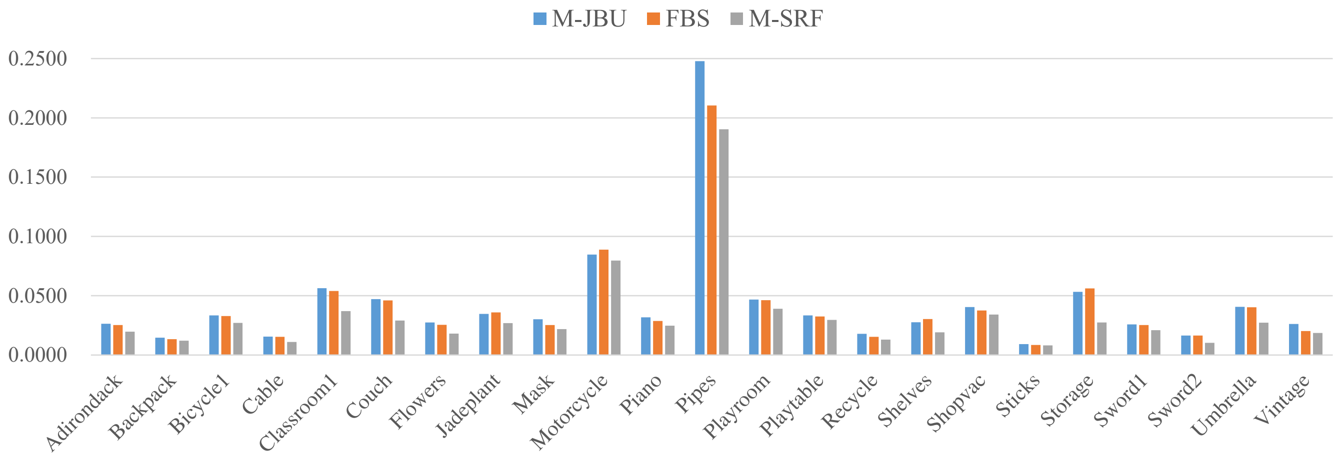

Figure 3: ***Mean absolute error (MAE) comparisons on Middlebury 2014 dataset [SHK*14].*** *We allocate 20 milliseconds for M-JBU and M-SRF. The unoptimized FBS takes about 1.2 to 2.0 seconds. Our method (M-SRF) outperforms M-JBU and FBS in all 23 cases. Our method's average MAE is 0.753× and 0.800× of M-JBU and FBS, respectively.*

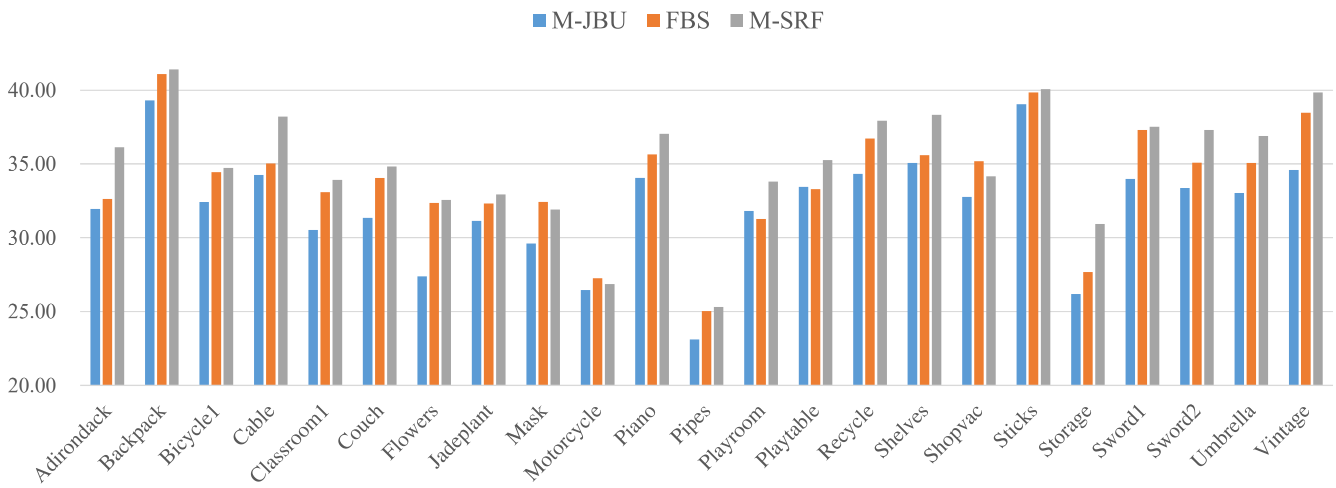

Figure 4: ***Peak signal-to-noise ratio (PSNR) comparisons on Middlebury 2014 dataset [SHK*14].*** *We allocate 20 milliseconds for M-JBU and M-SRF. The unoptimized FBS takes about 1.2 to 2.0 seconds. Our method (M-SRF) outperforms M-JBU in all 23 cases and achieves 2.99 dB improvement on average. Compared to FBS, our method has higher PSNR in 20 cases and obtains 1.18 dB gain on average.*

Case		M-JBU MAE	M-JBU PSNR	FBS MAE	FBS PSNR	M-SRF (ours) MAE	M-SRF (ours) PSNR
Adirondack	input A	0.0260	31.93	0.0258	33.92	*0.0204*	*36.13*
	input B	0.0265	31.99	0.0244	31.33	*0.0188*	*36.13*
	average	0.0263	31.96	0.0252	32.63	*0.0196*	*36.13*
Backpack	input A	0.0147	38.52	0.0150	40.08	*0.0119*	*41.27*
	input B	0.0145	40.12	*0.0116*	*42.10*	0.0123	41.54
	average	0.0146	39.32	0.0133	41.09	*0.0121*	*41.41*
Bicycle1	input A	0.0329	33.80	0.0345	34.26	*0.0282*	*34.40*
	input B	0.0338	31.01	0.0310	34.61	*0.0259*	*35.07*
	average	0.0333	32.41	0.0327	34.43	*0.0271*	*34.74*
Cable	input A	0.0158	35.01	0.0122	37.27	*0.0111*	*38.71*
	input B	0.0150	33.49	0.0184	32.81	*0.0110*	*37.71*
	average	0.0154	34.25	0.0153	35.04	*0.0110*	*38.21*
Classroom1	input A	0.0513	32.06	0.0538	33.70	*0.0337*	*35.07*
	input B	0.0613	29.04	0.0539	32.47	*0.0401*	*32.80*
	average	0.0563	30.55	0.0539	33.09	*0.0369*	*33.93*
Couch	input A	0.0464	31.87	0.0532	32.63	*0.0318*	*33.83*
	input B	0.0476	30.84	0.0387	35.47	*0.0261*	*35.82*
	average	0.0470	31.35	0.0459	34.05	*0.0289*	*34.83*
Flowers	input A	0.0302	26.48	0.0290	31.22	*0.0205*	*31.28*
	input B	0.0245	28.29	0.0217	33.52	*0.0155*	*33.86*
	average	0.0273	27.39	0.0254	32.37	*0.0180*	*32.57*
Jadeplant	input A	0.0352	31.70	0.0352	*32.49*	*0.0291*	32.20
	input B	0.0340	30.60	0.0367	32.14	*0.0245*	*33.66*
	average	0.0346	31.15	0.0359	32.32	*0.0268*	*32.93*
Mask	input A	0.0304	29.65	0.0273	31.75	*0.0213*	*31.83*
	input B	0.0297	29.58	0.0231	*33.13*	*0.0222*	32.00
	average	0.0300	29.61	0.0252	*32.44*	*0.0218*	31.91
Motorcycle	input A	0.0870	26.40	0.0887	*27.41*	*0.0805*	26.93
	input B	0.0822	26.54	0.0891	*27.10*	*0.0788*	26.79
	average	0.0846	26.47	0.0889	*27.25*	*0.0797*	26.86
Piano	input A	0.0306	34.59	0.0262	*36.77*	*0.0252*	36.54
	input B	0.0329	33.54	0.0310	34.52	*0.0241*	*37.55*
	average	0.0318	34.07	0.0286	35.64	*0.0246*	*37.04*
Pipes	input A	0.2438	22.85	0.2072	24.94	*0.1831*	*25.46*
	input B	0.2519	23.35	0.2136	25.14	*0.1976*	*25.18*
	average	0.2479	23.10	0.2104	25.04	*0.1904*	*25.32*
Playroom	input A	0.0426	32.32	0.0427	29.91	*0.0354*	*34.54*
	input B	0.0507	31.32	0.0497	32.63	*0.0424*	*33.07*
	average	0.0466	31.82	0.0462	31.27	*0.0389*	*33.81*
Playtable	input A	0.0307	34.27	0.0315	32.84	*0.0269*	*36.20*
	input B	0.0359	32.65	0.0333	33.75	*0.0321*	*34.29*
	average	0.0333	33.46	0.0324	33.29	*0.0295*	*35.25*
Recycle	input A	0.0182	33.18	0.0167	34.39	*0.0134*	*36.41*
	input B	0.0173	35.50	0.0140	39.05	*0.0126*	*39.46*
	average	0.0178	34.34	0.0154	36.72	*0.0130*	*37.93*
Shelves	input A	0.0259	35.15	0.0272	36.62	*0.0199*	*36.90*
	input B	0.0292	34.97	0.0334	34.57	*0.0180*	*39.77*
	average	0.0275	35.06	0.0303	35.59	*0.0190*	*38.34*
Shopvac	input A	0.0344	33.35	0.0345	*35.38*	*0.0313*	34.77
	input B	0.0463	32.18	0.0407	*34.99*	*0.0367*	33.55
	average	0.0403	32.77	0.0376	*35.18*	*0.0340*	34.16
Sticks	input A	0.0096	38.34	0.0084	39.76	*0.0083*	*39.78*
	input B	0.0086	39.75	0.0085	39.95	*0.0076*	*40.35*
	average	0.0091	39.04	0.0084	39.85	*0.0080*	*40.06*

Table 1: *MAE and PSNR Comparisons on Middlebury 2014 dataset (Part I). Please continue to Table 2 for more cases.*

		M-JBU		FBS		M-SRF (ours)	
Case		MAE	PSNR	MAE	PSNR	MAE	PSNR
Storage	input A	0.0385	28.42	0.0443	30.60	*0.0234*	*31.71*
	input B	0.0680	23.97	0.0681	24.74	*0.0312*	*30.15*
	average	0.0532	26.20	0.0562	27.67	*0.0273*	*30.93*
Sword1	input A	0.0274	32.46	0.0281	36.65	*0.0213*	*37.26*
	input B	0.0242	35.52	0.0221	*37.96*	*0.0206*	37.79
	average	0.0258	33.99	0.0251	37.30	*0.0209*	*37.52*
Sword2	input A	0.0172	32.42	0.0140	37.16	*0.0095*	*37.81*
	input B	0.0154	34.29	0.0188	33.03	*0.0111*	*36.78*
	average	0.0163	33.35	0.0164	35.09	*0.0103*	*37.30*
Umbrella	input A	0.0530	31.00	0.0524	32.61	*0.0310*	*35.90*
	input B	0.0282	35.05	0.0282	37.52	*0.0234*	*37.87*
	average	0.0406	33.02	0.0403	35.06	*0.0272*	*36.88*
Vintage	input A	0.0221	36.01	0.0171	40.24	*0.0167*	*41.45*
	input B	0.0300	33.16	0.0233	36.71	*0.0204*	*38.25*
	average	0.0261	34.58	0.0202	38.47	*0.0185*	*39.85*
Total	input A	0.0419	32.25	0.0402	34.03	*0.0319*	*35.06*
	input B	0.0438	32.03	0.0406	33.88	*0.0327*	*35.19*
	average	0.0429	32.14	0.0404	33.95	*0.0323*	*35.13*

Table 2: *MAE and PSNR Comparisons on Middlebury 2014 dataset (Part II). We allocate 20 milliseconds for M-JBU and M-SRF. The unoptimized FBS takes 1.2 - 2.0 seconds. For each test case, the methods producing the lowest MAE and highest PSNR values are shown with italics. Our method (M-SRF) produces the lowest MAE for all 23 cases and the highest PSNR in 20 cases.*

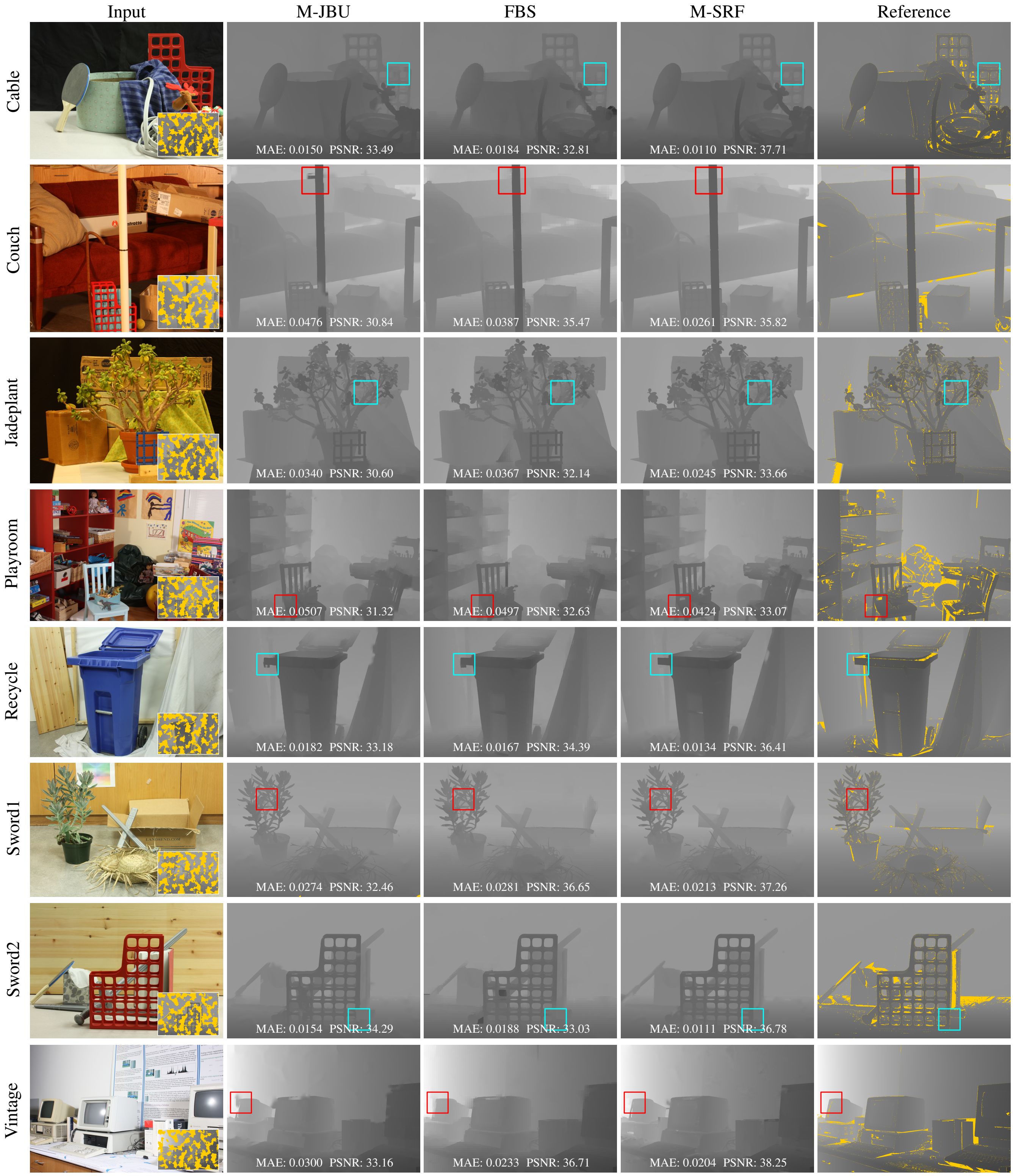

Figure 5: ***Qualitative comparisons of four scenes in Middlebury 2014 data set [SHK*14].*** *We allocate 20 milliseconds for M-JBU and M-SRF. The Python-version FBS takes about 1.2 to 2.0 seconds, running on CPU. Our method can better preserve the depth discontinuities and achieve lower MAE and higher PSNR than M-JBU and FBS.*

Scene	Place of Figure	JBF	M-JBU	FBS	SRF	M-SRF
Puppies	Fig. 1 in the main paper	σ_c: 0.02 σ_s: 8.50 (7.5 ms)	σ_c: 0.02 σ_s: 5.50 level: 4	σ_{uv}: 10.00 σ_{xy}: 20.00 σ_l: 10.00 λ: 128.00	σ_c: 0.02 K: 24	σ_c: 0.02 K: 48 level: 4
Table	Fig. 9 in the main paper	σ_c: 0.01 σ_s: 13.00	σ_c: 0.01 σ_s: 5.00 level: 3	σ_{uv}: 5.12 σ_{xy}: 31.30 σ_l: 18.84 λ: 239.67	σ_c: 0.07 K: 38	σ_c: 0.04 K: 62 level: 4
Buddha	Fig. 9 in the main paper	σ_c: 0.01 σ_s: 13.00	σ_c: 0.01 σ_s: 6.50 level: 4	σ_{uv}: 2.85 σ_{xy}: 27.34 σ_l: 10.03 λ: 228.38	σ_c: 0.02 K: 58	σ_c: 0.05 K: 81 level: 3
LivingRoom	Fig. 9 in the main paper	σ_c: 0.03 σ_s: 12.50	σ_c: 0.02 σ_s: 6.00 level: 3	σ_{uv}: 25.75 σ_{xy}: 27.20 σ_l: 20.69 λ: 246.33	σ_c: 0.15 K: 44	σ_c: 0.12 K: 75 level: 3
Outdoor	Fig. 9 in the main paper	σ_c: 0.01 σ_s: 12.50	σ_c: 0.01 σ_s: 6.00 level: 3	σ_{uv}: 2.63 σ_{xy}: 22.35 σ_l: 14.07 λ: 39.37	σ_c: 0.09 K: 34	σ_c: 0.02 K: 57 level: 3
Adirondack	Fig. 12 in the main paper	N/A	σ_c: 0.01 σ_s: 4.00 level: 4	σ_{uv}: 28.16 σ_{xy}: 30.12 σ_l: 20.82 λ: 31.47	N/A	σ_c: 0.05 K: 31 level: 4
Umbrella	Fig. 12 in the main paper	N/A	σ_c: 0.04 σ_s: 4.00 level: 4	σ_{uv}: 29.65 σ_{xy}: 31.12 σ_l: 30.45 λ: 11.59	N/A	σ_c: 0.20 K: 37 level: 4
Man	Fig. 13 in the main paper	σ_c: 0.02 σ_s: 13.50	σ_c: 0.01 σ_s: 6.00 level: 4	σ_{uv}: 8.00 σ_{xy}: 24.00 σ_l: 8.00 λ: 64.00	σ_c: 0.04 K: 54	σ_c: 0.03 K: 66 level: 4
Work	Fig. 13 in the main paper	σ_c: 0.01 σ_s: 12.50	σ_c: 0.02 σ_s: 5.50 level: 4	σ_{uv}: 16.00 σ_{xy}: 24.00 σ_l: 12.00 λ: 64.00	σ_c: 0.03 K: 20	σ_c: 0.04 K: 34 level: 4
Office	Fig. 13 in the main paper	σ_c: 0.05 σ_s: 12.50	σ_c: 0.05 σ_s: 5.00 level: 4	σ_{uv}: 20.00 σ_{xy}: 24.00 σ_l: 20.00 λ: 100.00	σ_c: 0.05 K: 34	σ_c: 0.08 K: 32 level: 4

Table 3: *Parameters for the results shown in the main paper.*

Scene	Place of Figure	JBF	M-JBU	FBS	SRF	M-SRF
Cable	Fig. 5 in this document	N/A	σ_c: 0.02 σ_s: 4.00 level: 4	σ_{uv}: 37.99 σ_{xy}: 31.82 σ_l: 38.10 λ: 54.03	N/A	σ_c: 0.20 K: 47 level: 4
Couch	Fig. 5 in this document	N/A	σ_c: 0.02 σ_s: 4.50 level: 4	σ_{uv}: 38.25 σ_{xy}: 32.00 σ_l: 38.25 λ: 37.99	N/A	σ_c: 0.05 K: 48 level: 4
Jadeplant	Fig. 5 in this document	N/A	σ_c: 0.01 σ_s: 4.00 level: 4	σ_{uv}: 12.53 σ_{xy}: 31.83 σ_l: 28.24 λ: 54.68	N/A	σ_c: 0.03 K: 24 level: 4
Playroom	Fig. 5 in this document	N/A	σ_c: 0.03 σ_s: 4.00 level: 4	σ_{uv}: 38.25 σ_{xy}: 32.00 σ_l: 38.25 λ: 29.53	N/A	σ_c: 0.11 K: 39 level: 4
Recycle	Fig. 5 in this document	N/A	σ_c: 0.02 σ_s: 4.00 level: 4	σ_{uv}: 17.35 σ_{xy}: 31.79 σ_l: 27.40 λ: 45.79	N/A	σ_c: 0.07 K: 42 level: 4
Sword1	Fig. 5 in this document	N/A	σ_c: 0.01 σ_s: 4.00 level: 4	σ_{uv}: 17.99 σ_{xy}: 31.97 σ_l: 34.18 λ: 18.07	N/A	σ_c: 0.05 K: 35 level: 4
Sword2	Fig. 5 in this document	N/A	σ_c: 0.02 σ_s: 4.00 level: 4	σ_{uv}: 33.52 σ_{xy}: 32.00 σ_l: 38.25 λ: 157.53	N/A	σ_c: 0.18 K: 35 level: 4
Vintage	Fig. 5 in this document	N/A	σ_c: 0.03 σ_s: 4.00 level: 4	σ_{uv}: 25.36 σ_{xy}: 31.07 σ_l: 32.73 λ: 13.17	N/A	σ_c: 0.03 K: 28 level: 4

Table 4: *Parameters for the results shown in this document.*