Photometric Depth Super-Resolution – Supplementary Material

Bjoern Haefner*, Songyou Peng*, Alok Verma*, Yvain Quéau, and Daniel Cremers

1 Ill-posedness in Depth Super-resolution and Shape-from-shading (Sect. 2)

Figure 1 illustrates the ambiguities of depth super-resolution, and Figure 2 those of shape-from-shading.

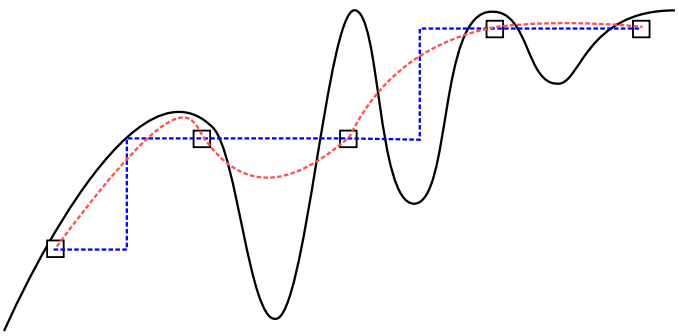

Fig. 1: There exist infinitely many ways (dashed lines) to interpolate between low-resolution depth samples (rectangles). Our disambiguation strategy builds upon shape-from-shading applied to the companion high-resolution color image (c.f. Figure 2), in order to resurrect the fine-scale geometric details of the genuine surface (solid line).

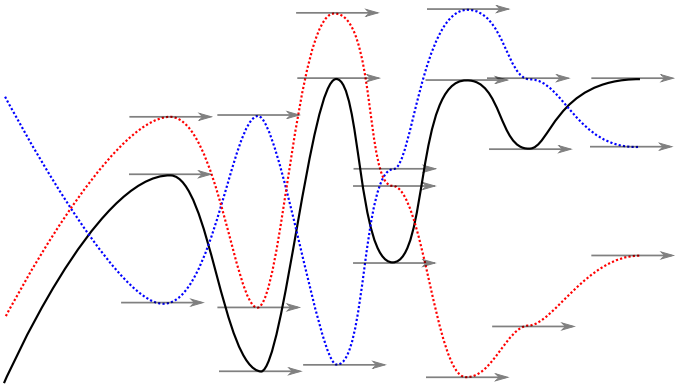

Fig. 2: Shape-from-shading suffers from the concave / convex ambiguity: the genuine surface (solid line) and both the surfaces depicted by dashed lines produce the same image, if lit and viewed from above. We put forward low-resolution depth clues (c.f. Figure 1) for disambiguation.

* *Those authors contributed equally*

2 Single Shot Depth Super-Resolution using Shape-from-Shading (Sect. 3)

Figure 3 illustrates the synthetic datasets used for evaluation, which were generated using four different 3D-shapes ("Lucy", "Thai Statue", "Armadillo" and "Joyful Yell"), each of them rendered using three different albedo maps ("voronoi", "rectcircle" and "bar") and three different scaling factors (2, 4 and 8) for the low-resolution depth image.

Figure 4 illustrates the effect of each hyper-parameter on shape and reflectance estimation (these experiments were conducted on the "Joyful Yell" dataset, with the three proposed albedo maps and three different scaling factors).

Table 1 presents the quantitative results on all synthetic datasets, in comparison with other state-of-the-art methods.

Figure 5 presents insightful qualitative comparisons on four synthetic datasets. Note that in this visualisation we only show super-resolution using a scaling factor of 4 to make comparisons fair, as [4] only provides code to perform super-resolution at such an upsampling rate.

Figure 6 shows the qualitative comparison on real-world data captured with a RealSense D415 Camera. The input images $\mathbf{I}$ are shown in the main paper in Figure 2. Note that [3] seems to give good depth estimates whereever the underlying assumption (an edge in the RGB image coincides with an edge in the depth ime) is met, cf. "Rucksack" dataset, but fails to result in detail preserving depth maps where reflectance is uniform or changes only slightly, as it only uses a sparse set of information from the RGB data to improve geometry, cf. "Android" and "Minion" dataset. [4] can not hallucinate surface details since it does not use the color image. [5] does a much better job at improving geometry, but it is largely overcome by shading-based super-resolution, as it uses information from a high-resolution RGB image.

Figure 7 shows four qualitative comparisons with state-of-the-art multi-view approaches on the publicly available datasets [6], [7]. "Augustus", "Lucy" and "Relief" in column four are the results of [8], where data is captured using a PrimeSense RGB-D camera. "Gate" in the fourth column is the result of [9], where data is acquired using a Structure Sensor for an iPad. "Augustus", "Relief" and "Gate" use an upsampling factor of 2, whereas "Lucy" provides RGB-D of $[640 \times 480 \text{ px}^2]$ for both $\mathbf{I}$ and z^0. Although our approach needs significantly less data compared to multi-view approaches, we are still able to recover fine geometry

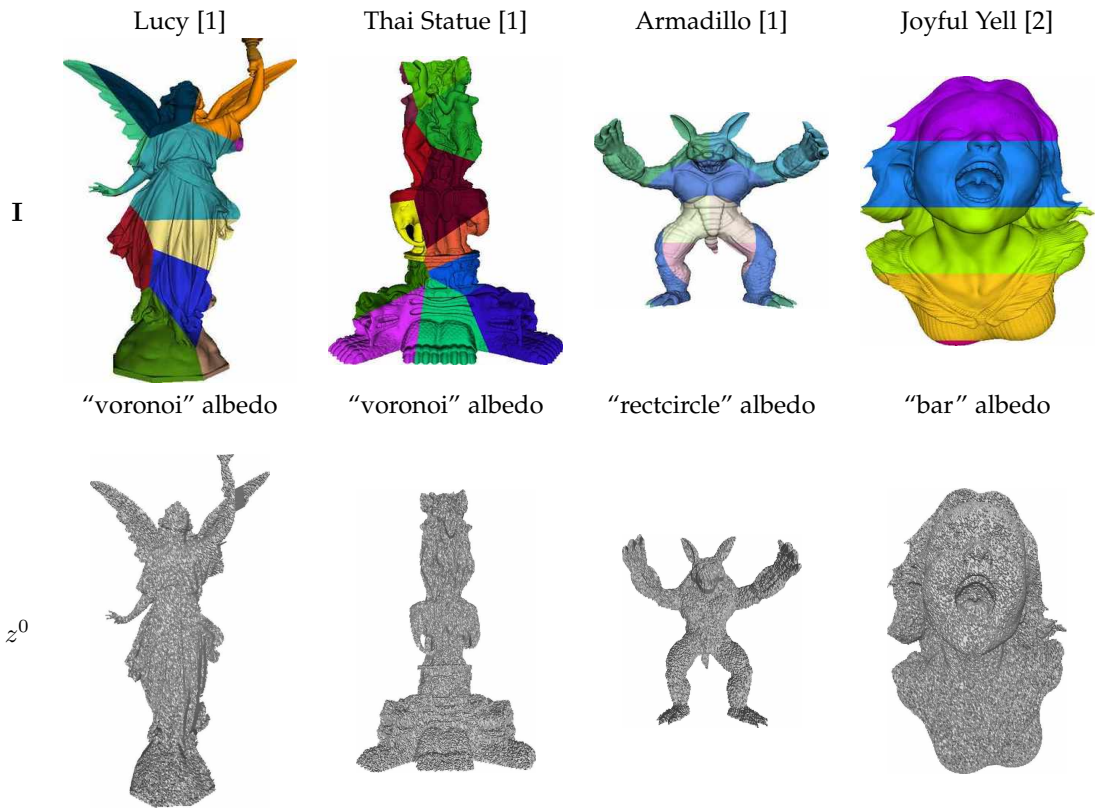

Fig. 3: Illustration of synthetic data used for quantitative evaluation. z^0 with a scaling factor of 2 is shown here.

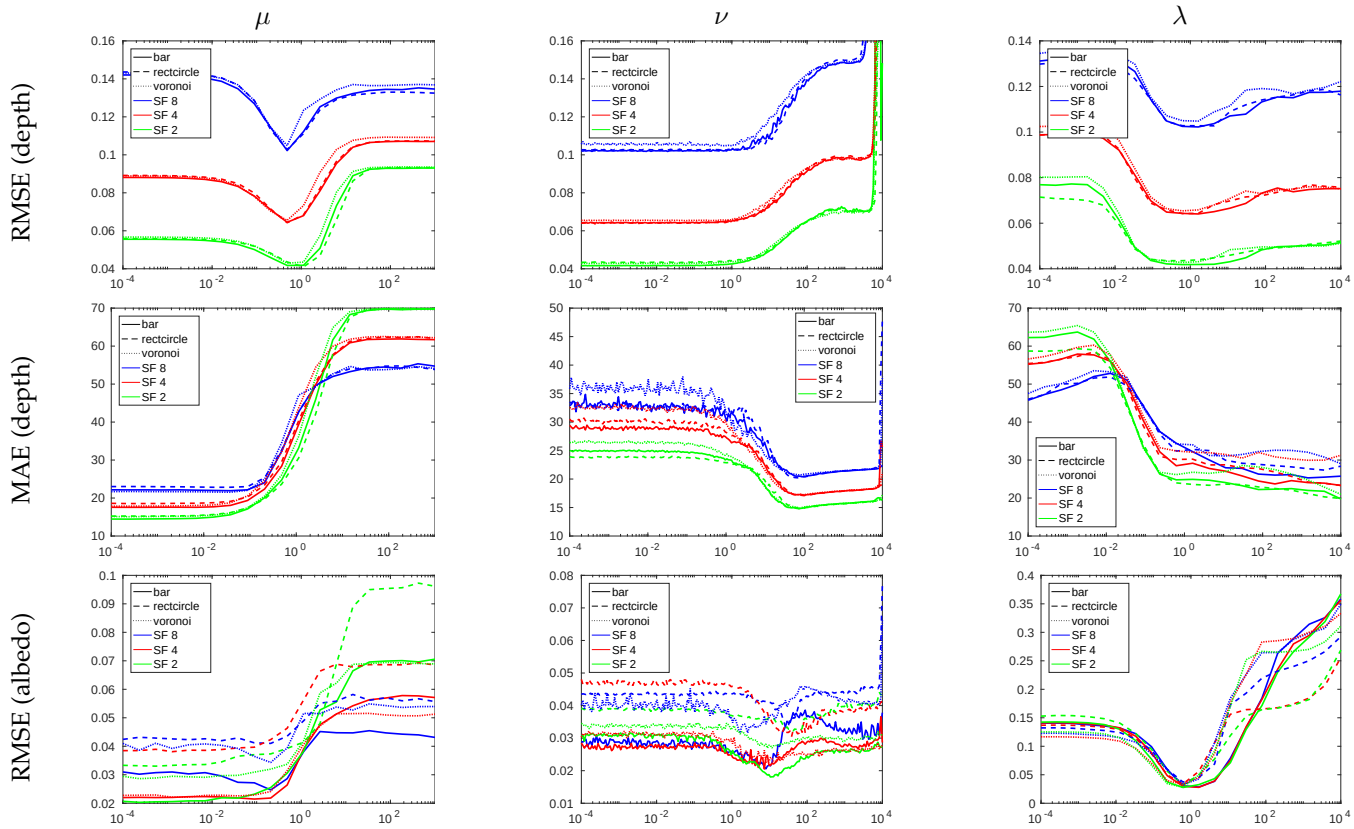

Fig. 4: Impact of the parameters (μ, ν, λ) on the accuracy of the albedo and depth estimates.

Albedo	3D-shape	SF	[3]		[4]		[5]		Ours	
			RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
bar	Armadillo	2	0.043643	38.6274	–	–	0.41993	67.2643	**0.034655**	**16.7496**
		4	**0.051558**	42.2277	0.17865	45.6972	0.45139	66.2117	0.054679	**19.0314**
		8	**0.072466**	43.5649	–	–	0.58837	69.3262	0.091263	**20.8836**
	Joyful Yell	2	0.05089	29.1719	–	–	0.1721	47.4836	**0.050694**	**16.7414**
		4	**0.066517**	33.0843	0.084094	42.611	0.22867	32.9784	0.079271	**19.0695**
		8	**0.10212**	36.565	–	–	0.37923	31.2894	0.128	**21.9886**
	Lucy	2	0.057987	39.4714	–	–	0.21309	66.5525	**0.053989**	**25.0955**
		4	**0.068502**	42.7169	0.50472	47.605	0.34091	69.2566	0.081005	**28.3044**
		8	**0.098713**	46.4775	–	–	0.43619	59.5434	0.1195	**30.1058**
	Thai Statue	2	0.040821	42.8976	–	–	0.12948	63.06	**0.035736**	**23.9147**
		4	**0.050296**	47.1017	0.22363	49.9553	0.15489	54.6139	0.057313	**28.492**
		8	**0.066515**	49.8604	–	–	0.22835	56.4247	0.087054	**31.65**
rectcircle	Armadillo	2	0.044026	39.108	–	–	0.34323	70.8526	**0.03494**	**18.4909**
		4	**0.052115**	43.3175	0.17782	45.6324	0.2338	50.6919	0.056727	**18.8487**
		8	**0.069467**	45.4735	–	–	0.61917	70.9363	0.09155	**21.9959**
	Joyful Yell	2	**0.051296**	30.7886	–	–	0.14841	41.5424	0.05226	**17.134**
		4	**0.066911**	33.3	0.10328	42.7531	0.28311	51.0665	0.080387	**19.8717**
		8	**0.10201**	36.2961	–	–	0.39518	35.4817	0.1281	**22.8027**
	Lucy	2	0.058495	39.7374	–	–	0.19546	64.8212	**0.054383**	**24.8427**
		4	**0.069893**	43.9016	0.50464	48.1068	0.23235	53.2901	0.082547	**28.7517**
		8	**0.099402**	46.3739	–	–	0.39583	64.3269	0.12283	**29.1531**
	Thai Statue	2	0.039821	40.6144	–	–	0.11355	58.2254	**0.036845**	**23.9036**
		4	**0.04973**	46.1154	0.20894	49.4124	0.16749	52.9663	0.05866	**28.155**
		8	**0.067799**	50.6515	–	–	0.21058	50.9074	0.094688	**33.5308**
voronoi	Armadillo	2	0.043635	38.9089	–	–	0.33005	69.3157	**0.034751**	**17.6873**
		4	**0.051989**	41.57	0.17182	45.5833	0.4407	65.5811	0.056032	**20.168**
		8	**0.07077**	43.1987	–	–	0.50548	63.8618	0.090708	**22.2767**
	Joyful Yell	2	**0.052002**	28.7903	–	–	0.16893	47.72	0.052429	**17.0453**
		4	**0.066557**	32.3448	0.086394	43.1744	0.24753	39.6569	0.079888	**19.6512**
		8	**0.10238**	35.8017	–	–	0.47694	47.4707	0.12916	**21.6663**
	Lucy	2	0.058222	36.2327	–	–	0.29164	72.9002	**0.054442**	**26.1333**
		4	**0.068253**	40.8878	0.5066	48.0387	0.32955	71.1042	0.079877	**28.4506**
		8	**0.099838**	43.7671	–	–	0.37839	57.6856	0.11877	**29.6331**
	Thai Statue	2	0.039872	39.6508	–	–	0.13261	65.8352	**0.037607**	**25.6126**
		4	**0.049783**	45.7178	0.22688	49.4132	0.16533	58.3933	0.058957	**28.6314**
		8	**0.065577**	48.7962	–	–	0.21927	49.6711	0.091959	**32.0347**
Median		2	0.047458	39.0085	–	–	0.18378	65.3282	**0.044151**	**21.1973**
		4	**0.059316**	42.4723	0.19379	46.6511	0.24067	53.952	0.069114	**24.1615**
		8	**0.085589**	44.6203	–	–	0.3955	57.0551	0.10673	**25.9779**
Mean		2	0.048392	36.9999	–	–	0.22154	61.2978	**0.044394**	**21.1126**
		4	**0.059342**	41.0238	0.24812	46.4986	0.27298	55.4842	0.068779	**23.9521**
		8	**0.084754**	43.9022	–	–	0.40275	54.7438	0.1078	**26.4768**

TABLE 1: Quantitative comparison between our results and three state-of-the-art methods, on all the synthetic datasets.

close to the degree of detail of [6], [7]. Even with more complex lighting, cf. "Gate", our approach can result in high-resolution depth maps with fine scale details and the depth does not seem to deteriorate.

Figure 8 shows additional qualitative comparison on data we captured with an Asus Xtion Pro Live camera and a scaling factor of 4, i.e. depth maps were acquired in $[320 \times 240\ \mathrm{px}^2]$ resolution.

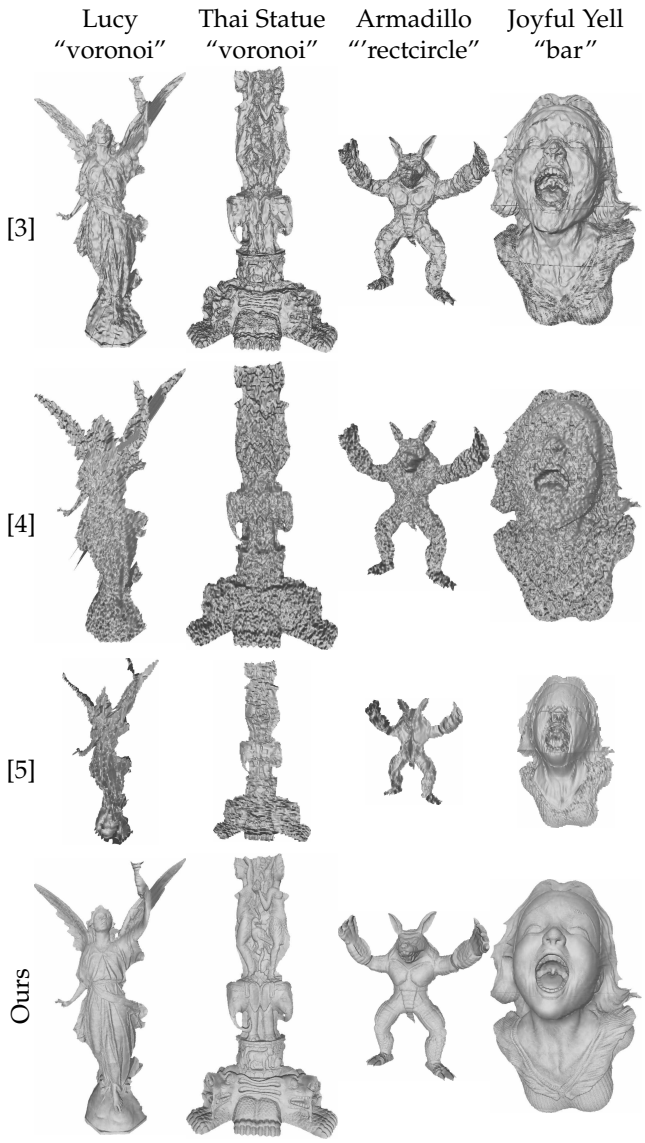

Fig. 5: Qualitative comparison of our results against state-of-the-art methods on four synthetic datasets using a scaling factor of 4.

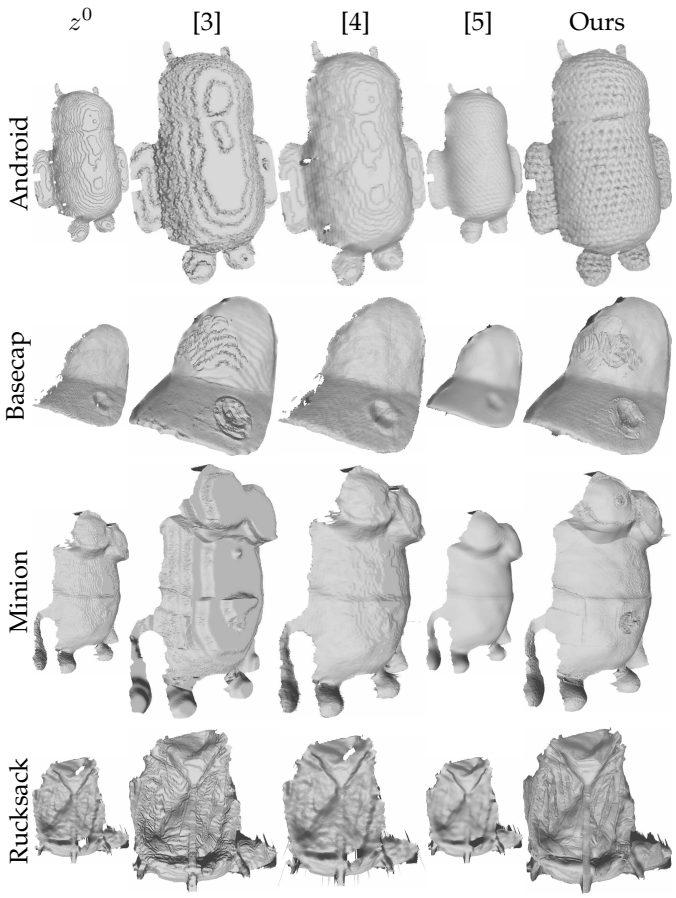

Fig. 6: Qualitative comparison with with other state-of-the-art methods on four real-world datasets captured with a RealSense D415 camera.

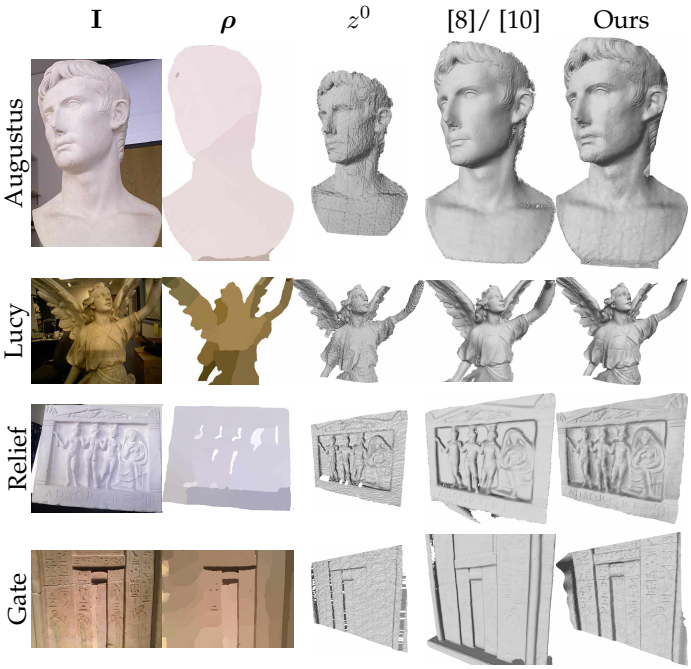

Fig. 7: Qualitative comparison against state-of-the-art multi-view approaches. The publicly available dataset [6] was captured with a PrimeSense camera, whereas [7] was acquired with a Structure Sensor for the iPad.

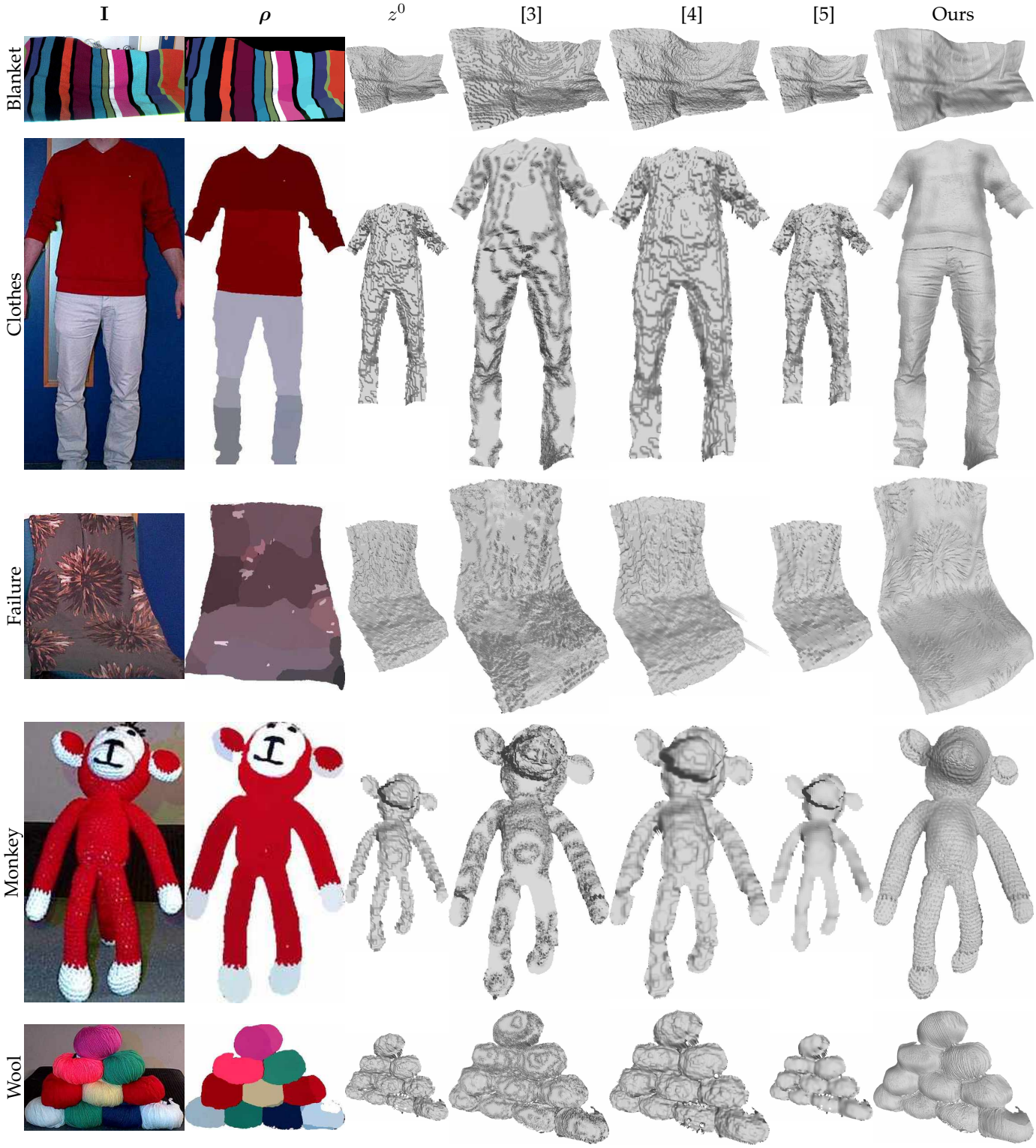

Fig. 8: Qualitative comparison of state-of-the-art single-view approaches on five real-world datasets captured with an Asus Xtion Pro Live camera at resolution $[1280 \times 960\ \text{px}^2]$ for the RGB images and $[320 \times 240\ \text{px}^2]$ for the low-resolution depth.

3 Depth Super-Resolution using Shape-from-Shading and Reflectance Learning (Sect. 4)

Figure 9 illustrates the lack of inter-class generalisation in learning-based methods: the approach of [11] (trained on Sintel [12] and MIT [13] datasets) performs poorly on the car image, because such an object was not present in the learning database. For the same reason, the alternative approach of [14] (trained on ShapeNet objects [15]) fails on the MIT object, and both approaches fail on the face image.

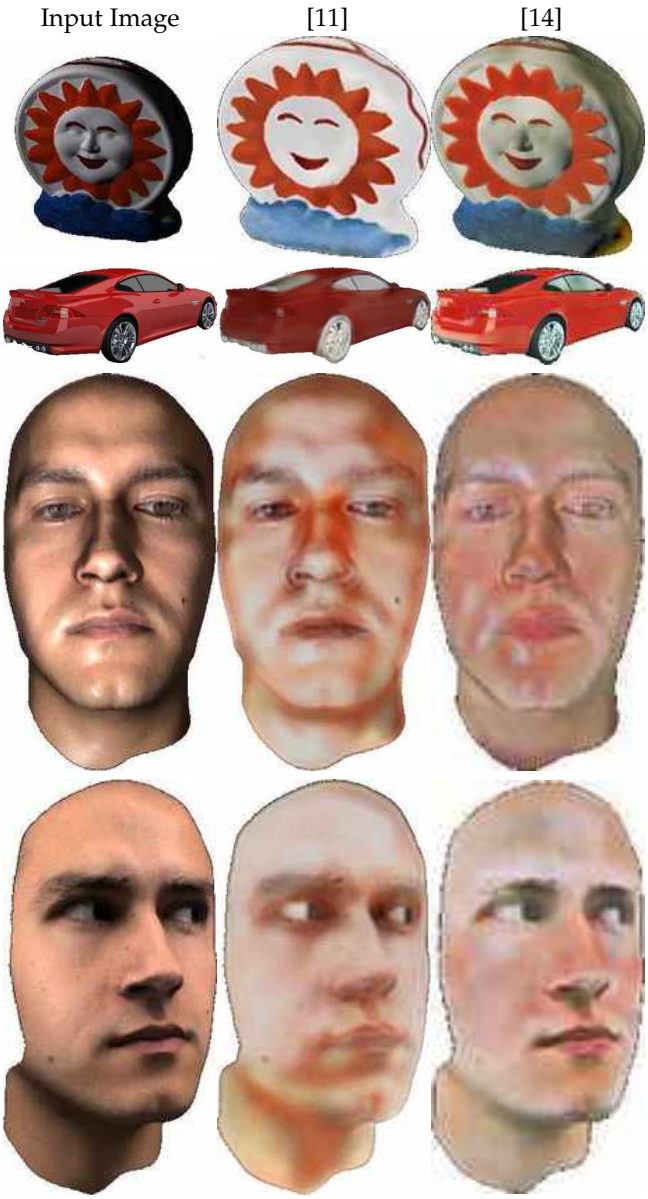

Fig. 9: CNN-based albedo estimation applied to an object from the MIT database (first row), a car from the ShapeNet dataset (second row), and two images of human faces we generated with a renderer using ICT-3DRFE [16] database

Figure 10 shows an example of a failure case for end-to-end learning approaches which simultaneously estimate reflectance and geometry. As soon as the scene to analyse contains unexpected deviation from the learned model, artifacts appear.

Input Image Estimated Normal from [17]

Fig. 10: Reconstruction results from SfSNet [17], which is an end-to-end deep learning based approach. It fails to account for small departures from usual face images, here fingers for example, and provides an erroneous normal estimation.

Figure 11 illustrates the rendering of synthetic human faces with extended non-directional light sources, emulating usual indoor light conditions. Geometry and reflectance are obtained from ICT-3DRFE database [16].

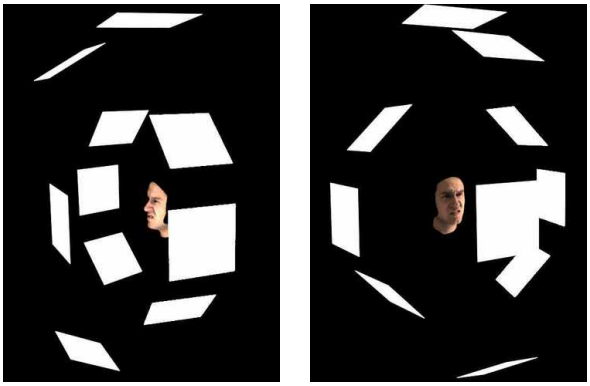

Fig. 11: Light sources used for rendering human faces.

Figure 12, illustrates the U-Net architecture used for albedo estimation. It essentially comprises of an initial convolution layer of kernel size 4, stride 2 and padding 1; after which there are repeated blocks of 8 ReLU-Conv-BatchNorm layers. This results in downsampling of a 512x512 resolution image to a 1x512 vector at the bottleneck of the "U". Further, the 1D array is upsampled to input resolution with multiple ReLU-Transpose Convolution-BatchNorm layers. Dropout is also used in a few layers to allow for randomness while learning the mapping from input images to albedo maps. Finally, for the loss function we use the L1 loss, which favors sharper output compared to the L2 loss.

Figure 13 and Table 2 show several results of our approach on synthetic datasets, in comparison with two other state-of-the-art methods. We choose to compare against SIRFS [18] and Pix2Vertex [19], because the former is a completely prior-based approach with minimal learning while the latter is a deep neural networks-based approach. Our approach, which stands inbetween, inherits the strengths of both approaches. It reconstructs fine-scale details without extensive smoothing, and it can also easily reconstruct new

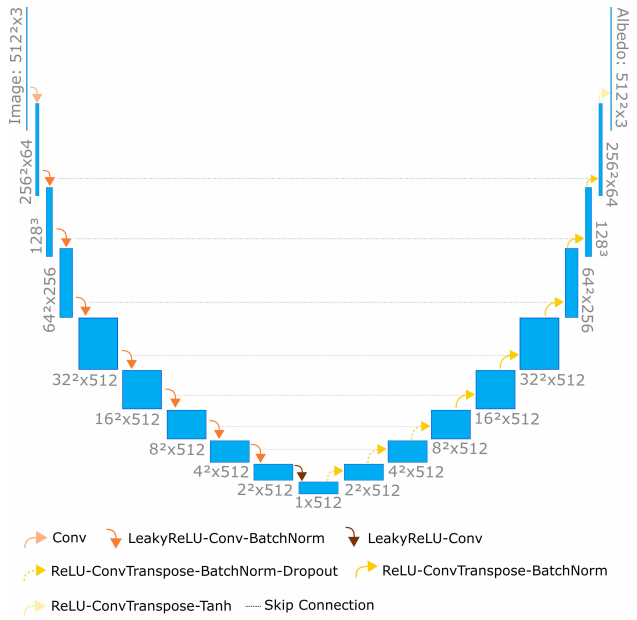

Fig. 12: The U-Net Architecture used for albedo estimation. The top two layers on the extreme left and right are the input and output respectively. The rest hidden layers are obtained by performing the operations mentioned for every color of arrow. The skip connections are shown as dotted lines which implies that the layers on the left are concatenated to the layers on the right.

geometries which were not present in the training database.

Figure 14 presents the qualitative comparison on real-world results with [18] and [17]. [18] attempts to provides reflectance which has minimal shading effects, but due to large number of priors on smoothness, parsimony and absolute color, the reflectance estimate is deceiving. [17] performs better than the pure prior-based approach, but is limited by the resolution $[128 \times 128 \text{ px}^2]$ and thus misses small-scale details. Our method provides high-resolution realistic albedo and depth maps directly out of the box.

Subject (S)	Expression (E)	SF	[18]		[17]		Ours	
			RMSE	MAE	RMSE	MAE	RMSE	MAE
11	0	2	**0.1069**	11.2811	-	-	0.1874	**5.8277**
		4	0.5148	41.7036	1.1023	12.2877	**0.1217**	**7.0191**
	1	2	**0.1110**	12.1757	-	-	0.1774	**5.9840**
		4	0.6776	51.5362	1.2368	11.3087	**0.1186**	**7.0827**
	2	2	**0.1371**	17.6559	-	-	0.1885	**7.4410**
		4	0.9092	58.2719	0.8175	19.8059	**0.1463**	**9.7861**
	3	2	**0.1200**	13.1908	-	-	0.1935	**6.2763**
		4	0.7650	53.3137	1.2194	13.6341	**0.1308**	**7.6617**
	4	2	**0.1106**	12.0300	-	-	0.1909	**6.7654**
		4	0.5584	49.2738	0.9065	13.1894	**0.1354**	**8.0273**
	5	2	**0.1964**	14.5500	-	-	0.2068	**7.3362**
		4	0.6844	53.0044	1.0657	15.0912	**0.1909**	**8.9047**
	6	2	**0.1206**	13.2598	-	-	0.1883	**6.8373**
		4	0.7318	57.4497	0.7211	14.3706	**0.1398**	**8.3742**
	7	2	0.3674	15.1031	-	-	**0.2287**	**7.2632**
		4	0.6998	47.2498	1.0601	18.9270	**0.2642**	**9.7440**
	8	2	**0.1094**	15.9824	-	-	0.1813	**9.3954**
		4	0.5888	55.8985	0.8822	13.4861	**0.1147**	**9.8121**
	9	2	**0.1118**	12.3690	-	-	0.1714	**6.7411**
		4	0.5300	47.9372	0.9714	13.0998	**0.1267**	**7.8888**
	10	2	**0.1225**	14.4960	-	-	0.1955	**7.9537**
		4	1.0062	55.9198	0.8631	17.9839	**0.1417**	**8.9901**
	11	2	**0.1159**	12.4566	-	-	0.1890	**6.6804**
		4	0.6296	49.2944	1.0929	12.4240	**0.1563**	**7.7767**
	12	2	**0.1092**	11.8244	-	-	0.1886	**6.0235**
		4	0.7968	51.1406	0.8824	11.0540	**0.1207**	**7.2522**
	13	2	**0.1113**	12.4305	-	-	0.1870	**6.1529**
		4	0.7014	52.5077	0.5021	10.0072	**0.1322**	**7.3263**
	14	2	**0.1091**	11.9522	-	-	0.1828	**6.2014**
		4	0.6463	49.8636	0.7394	11.3868	**0.1212**	**7.3132**
4	0	2	0.2281	23.4415	-	-	**0.1908**	**5.9752**
		4	0.6488	49.1959	0.4741	13.9368	**0.1117**	**6.8132**
	1	2	0.2854	30.5486	-	-	**0.1849**	**6.9702**
		4	0.5953	47.9548	1.0130	14.7240	**0.1127**	**7.6400**
	2	2	0.2528	29.8795	-	-	**0.1776**	**8.2012**
		4	0.6011	55.0417	1.0439	18.9314	**0.1284**	**9.7949**
	3	2	**0.1888**	19.0886	-	-	0.1927	**7.7149**
		4	0.6380	51.7959	0.6933	16.3201	**0.1223**	**8.9794**
	4	2	0.2837	30.3095	-	-	**0.1817**	**7.4706**
		4	0.5122	45.8693	0.5258	15.4585	**0.1138**	**8.5289**
	5	2	0.3609	35.5668	-	-	**0.2028**	**9.1059**
		4	0.6447	54.1181	0.9701	19.2865	**0.1970**	**11.2153**
	6	2	0.2205	27.0965	-	-	**0.1995**	**8.0220**
		4	0.5927	51.2974	1.0100	19.4058	**0.1542**	**9.6648**
	7	2	0.2369	24.9252	-	-	**0.2087**	**8.6590**
		4	0.6050	49.9729	0.8597	20.1422	**0.1941**	**9.5206**
	8	2	0.2145	26.4886	-	-	**0.1827**	**8.0033**
		4	0.5710	53.4401	0.7588	14.0183	**0.1075**	**8.2670**
	9	2	**0.1859**	22.4924	-	-	0.1892	**7.2511**
		4	0.5734	50.5616	0.5135	14.7545	**0.1169**	**8.5230**
	10	2	0.2744	28.8505	-	-	**0.1877**	**7.0955**
		4	0.6009	53.1664	0.5744	17.2112	**0.1117**	**7.7747**
	11	2	**0.1044**	12.7931	-	-	0.1961	**7.1938**
		4	0.6136	52.9554	1.0092	14.1251	**0.1115**	**7.6181**
	12	2	**0.1814**	21.9536	-	-	0.2034	**6.8525**
		4	0.6190	50.8171	0.6188	14.5414	**0.1293**	**7.7816**
	13	2	**0.1151**	13.6560	-	-	0.1866	**6.6517**
		4	0.5817	51.1208	1.1871	14.8568	**0.1115**	**7.7797**
	14	2	**0.1810**	21.5087	-	-	0.1906	**6.8284**
		4	0.5812	49.7192	0.7939	14.3981	**0.1111**	**7.6929**
Median		2	0.1590	15.5428	-	-	0.1888	7.0328
		4	0.6163	51.2190	0.8823	14.4698	0.1245	7.9581
Mean		2	0.1791	18.9786	-	-	0.1911	7.1625
		4	0.6473	51.3797	0.8703	15.0056	0.1365	8.3518

TABLE 2: Quantitative comparison between our method combining variational methods with machine learning, and two other state-of-the-art methods on two subjects from 3DRFE Dataset [20].

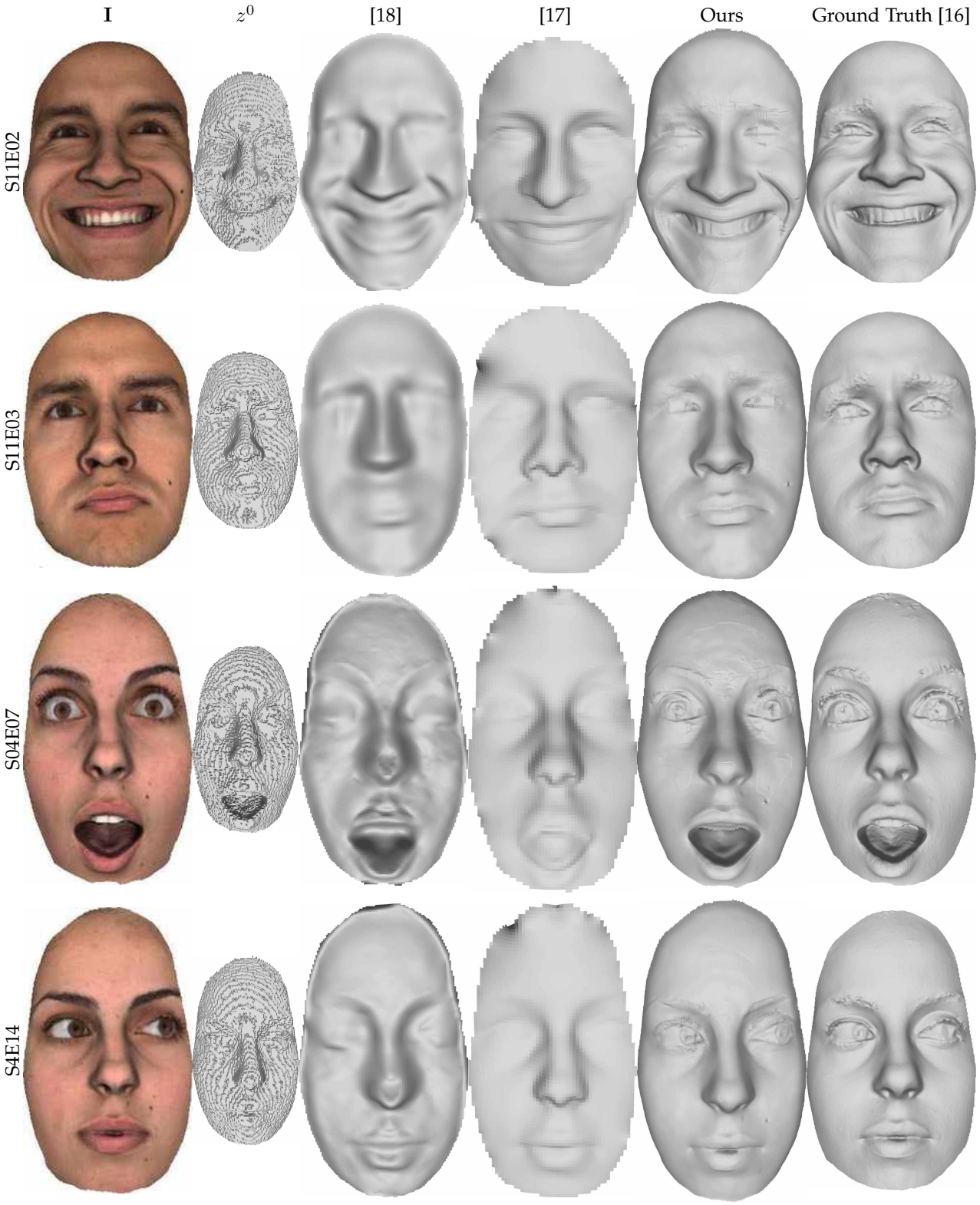

Fig. 13: Reconstruction results of the state-of-the-art and our combined variational and machine learning approach. SIRFS [18] (third column) provides smoothed out faces. Results of SfSNet [17] (fourth column) are shown as depth maps here, after integrating the output normals. Our method directly provides depth and is reliably able to do super-resolution and reconstruction of fine wrinkles of the face without any false enhancements.

Fig. 14: Qualitative comparison of our results against state-of-the-art methods, on four real-world subjects faces captured from Intel RealSense D415 camera.

4 Multi-Shot Depth Super-Resolution using Photometric Stereo (Sect. 5)

Figure 15 illustrates the synthetic datasets used for evaluation. Same as the the experiments for shape-from-shading, four objects ("Lucy", "Thai Statue", "Armadillo" and "Joyful Yell") are used to render with three different albedo maps ("ebsd"[1], "mandala"[2] and "rectcircle") and three different scaling factors (2,4 and 8).

Figure 16 shows the impact of the number of images n on the accuracy of the albedo and depth estimates, as well as the runtime using our multi-shot photometric approach ($\gamma = 0.01$). These experiments were conducted on the Joyful Yell dataset, with three different scaling factors and three different albedos.

Figure 17 illustrates the effect of the hyper-parameter γ on shape and reflectance estimation ($n = 10$). Same as Figure 16, these experiments were conducted on the Joyful Yell dataset, with three different scaling factors and three different albedos.

Table 3 quantitatively compares various methods including ours ($n = 20$). For [5], we randomly select one image out of 20. $\gamma = 0.01$ is used for ours in all experiments.

Figure 18 presents qualitative comparisons against three other methods on synthetic datasets shown in Figure 15.

Figure 19 shows four qualitative comparisons on real-world data captured with an Asus Xtion Pro Live camera against three other state-of-the-art methods. It can be seen that image-based depth super-resolution approach hallucinates reflectance information as geometric information, since the underlying concept assumes to allow for larger depth variations where strong image gradients are present. Clearly, [21] suffers from the GBR problem, as geometry deteriorates in the uncalibrated photometric stereo setup with a data-free depth prior, cf. "Tablet Case" and "Vase". [5] provides better depth estimates, as it takes into account depth images from a depth sensor, but it mistakenly hallucinates albedo information, as it uses only a single image. This clearly shows the advantages of acquiring multiple images under different illumination to separate reflectance and geometry in a regularisation free-manner.

References

[1] M. Levoy, J. Gerth, B. Curless, and K. Pull, "The stanford 3d scanning repository," 2005, http://www-graphics.stanford.edu/data/3dscanrep.

[2] "The joyful yell," 2015, https://www.thingiverse.com/thing:897412.

[3] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[4] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, 2016.

[5] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, and A. Bruckstein, "RGBD-Fusion: Real-Time High Precision Depth Recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5407–5416.

[6] M. Zollhöfer, A. Dai, M. Innman, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based Refinement on Volumetric Signed Distance Functions," 2015, http://graphics.stanford.edu/projects/vsfs/.

[7] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner, "Intrinsic3D Dataset," 2017, http://vision.in.tum.de/data/datasets/intrinsic3d.

[8] M. Zollhöfer, A. Dai, M. Innman, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based refinement on volumetric signed distance functions," *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 96:1–96:14, 2015.

[9] R. Maier, J. Stückler, and D. Cremers, "Super-resolution keyframe fusion for 3D modeling with high-quality textures," in *Proceedings of the International Conference on 3D Vision (3DV)*, 2015, pp. 536–544.

[10] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner, "Intrinsic3d: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3114–3122.

[11] M. M. Takuya Narihira and S. X. Yu, "Direct intrinsics: Learning albedo-shading decomposition by convolutional regression," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[12] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012, pp. 611–625.

[13] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2335–2342.

[14] J. Shi, Y. Dong, H. Su, and S. X. Yu, "Learning non-lambertian object intrinsics across shapenet categories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5844–5853.

[15] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An Information-Rich 3D Model Repository," Stanford University — Princeton University — Toyota Technological Institute at Chicago, Tech. Rep. arXiv:1512.03012 [cs.GR], 2015.

[16] G. Stratou, A. Ghosh, P. Debevec, and L. Morency, "Effect of illumination on automatic expression recognition: A novel 3d relightable facial database," in *Face and Gesture*, 2011, pp. 611–618.

[17] S. Sengupta, A. Kanazawa, C. D. Castillo, and D. W. Jacobs, "SfSNet: Learning Shape, Refectance and Illuminance of Faces in the Wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6296–6305.

[18] J. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1670–1687, 2015.

[19] M. Sela, E. Richardson, and R. Kimmel, "Unrestricted facial geometry reconstruction using image-to-image translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[20] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec, "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination," in *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, 2007, pp. 183–194.

[21] P. Favaro and T. Papadhimitri, "A closed-form solution to uncalibrated photometric stereo via diffuse maxima," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 821–828.

1. https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html
2. http://www.cleverpedia.com/mandala-coloring-books-20-coloring-books-with-brilliant-kaleidoscope-designs/

Fig. 15: Illustration of synthetic data used for quantitative evaluation in multi-shot depth super-resolution setup.

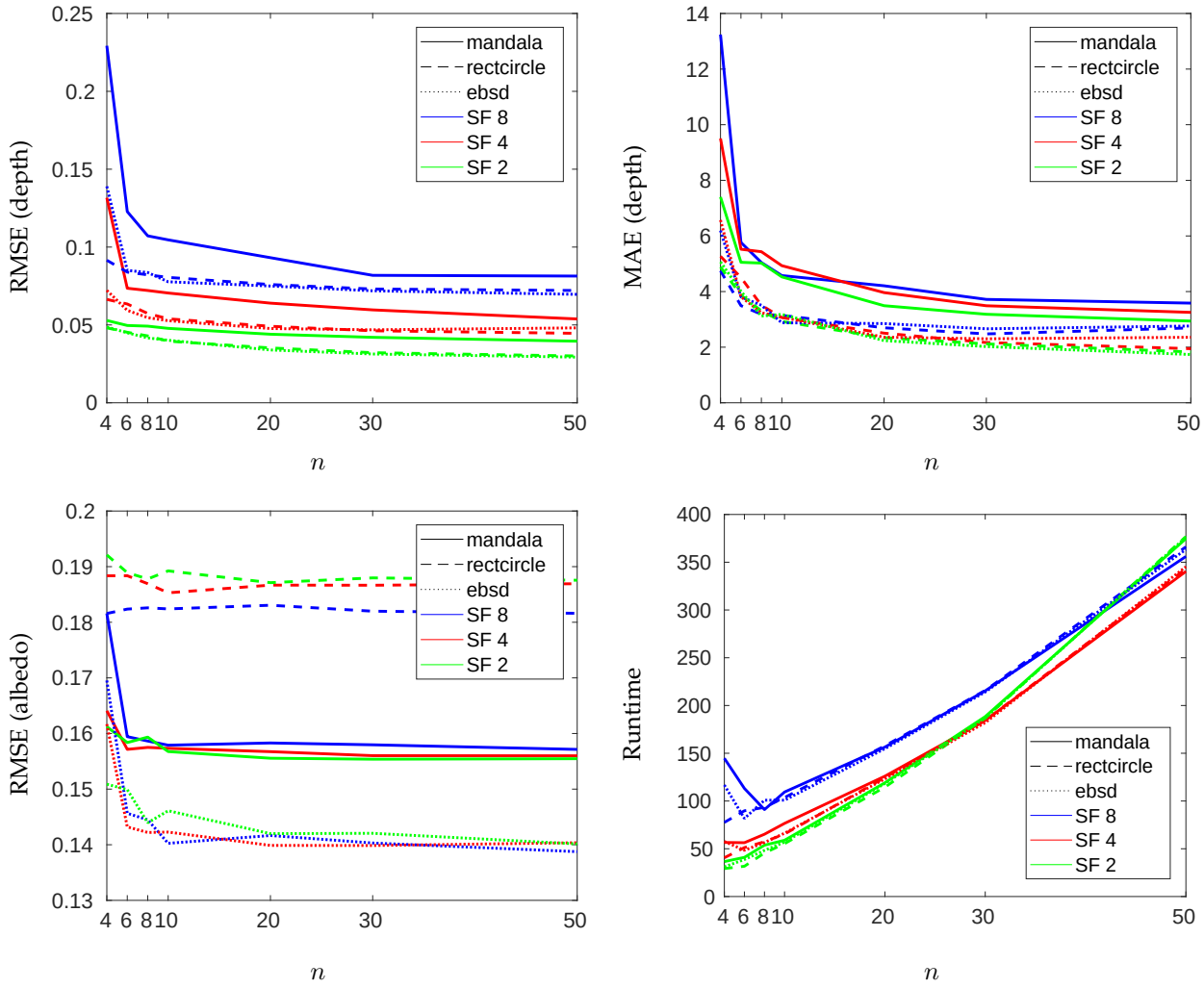

Fig. 16: Impact of the number of images n on the accuracy of the albedo and depth estimates using our multi-shot photometric approach ($\gamma = 0.01$).

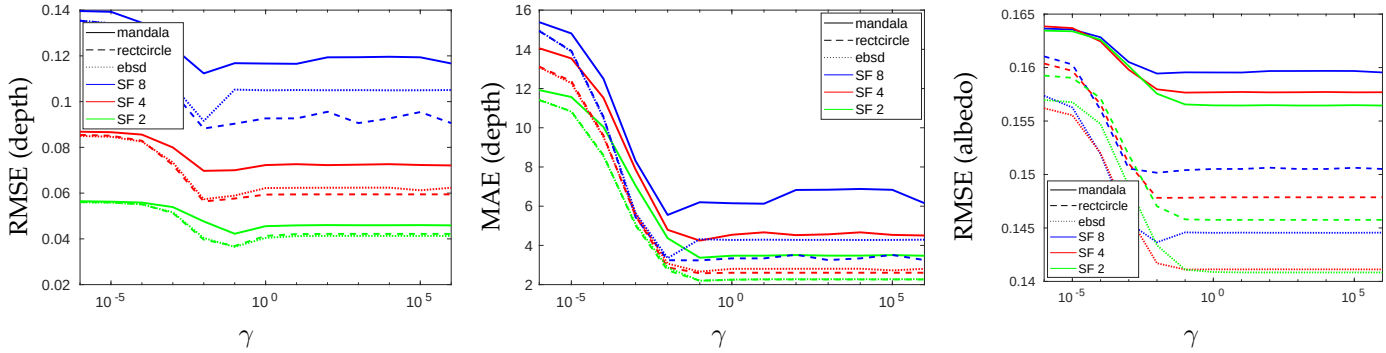

Fig. 17: Impact of the parameter γ on the accuracy of the albedo and depth estimates using our multi-shot photometric approach ($n = 10$).

Albedo	3D-shape	SF	Image Based depth SR		[21]		[5]		Ours	
			RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
mandala	Armadillo	2	0.031468	46.4149	0.51996	17.4225	0.4320	69.7311	**0.023266**	**2.883**
		4	0.042467	43.5403	0.5472	18.5244	0.3948	63.7382	**0.037789**	**2.8391**
		8	0.088849	42.6184	0.52598	17.6553	0.5961	83.7853	**0.073928**	**2.9196**
	Lucy	2	0.043889	46.4903	0.37197	15.7192	0.4755	84.6189	**0.036334**	**3.5842**
		4	0.065857	44.0677	1.9302	79.8752	0.4951	82.0516	**0.051316**	**3.6142**
		8	0.12668	42.8905	0.93938	15.9988	0.5231	64.7317	**0.084713**	**4.7864**
	Joyful Yell	2	0.048887	45.0552	1.0735	14.2243	0.3757	70.2724	**0.044198**	**3.3143**
		4	0.069088	42.644	1.0524	13.9289	0.2985	55.6927	**0.063392**	**3.6407**
		8	0.13103	40.0426	1.0625	14.0809	0.4240	44.2549	**0.1046**	**3.753**
	Thai Statue	2	0.032432	47.8575	0.37738	13.372	0.4615	70.3271	**0.022446**	**3.579**
		4	0.053061	45.5618	0.37775	13.6733	0.4211	90.2134	**0.036245**	**3.6985**
		8	0.094911	43.838	0.39766	14.0079	0.3371	53.2791	**0.049733**	**4.1133**
rectcircle	Armadillo	2	0.028459	41.506	0.52582	18.0902	0.2844	55.3096	**0.020885**	**2.0047**
		4	0.038966	38.7345	0.5264	18.0104	0.3031	48.1000	**0.035145**	**1.9458**
		8	0.11182	36.3801	0.52504	17.8218	0.5805	80.4625	**0.073139**	**2.1436**
	Lucy	2	0.040635	42.3051	0.32285	13.6126	0.4868	85.9076	**0.026858**	**1.8617**
		4	0.062747	39.0783	0.32295	13.5222	0.4685	75.9166	**0.041968**	**2.2851**
		8	0.12325	37.956	0.32509	13.7086	0.3767	56.5020	**0.075311**	**3.8793**
	Joyful Yell	2	0.045765	39.9946	0.84162	11.4847	0.2012	41.3053	**0.038698**	**2.7879**
		4	0.064537	37.1175	0.84386	11.4906	0.3189	37.2107	**0.053871**	**3.1022**
		8	0.09492	34.7218	0.83033	11.424	0.4432	36.3990	**0.084381**	**3.2463**
	Thai Statue	2	0.030859	44.4276	0.38981	13.3935	0.2625	66.0562	**0.018374**	**2.1086**
		4	0.045516	41.7235	0.36671	12.8741	0.3151	85.4734	**0.028457**	**2.2876**
		8	0.10507	39.7697	0.37632	12.9615	0.2389	55.0568	**0.041552**	**3.0519**
ebsd	Armadillo	2	0.031939	46.9515	0.49466	16.3427	0.3473	65.4823	**0.021037**	**2.0398**
		4	0.04424	44.2571	0.50255	16.2739	0.5933	58.6932	**0.036102**	**2.0035**
		8	0.10062	42.2539	0.57469	17.7183	0.6453	81.5187	**0.073138**	**1.8159**
	Lucy	2	0.04299	47.5844	0.32989	13.0463	0.4141	84.9623	**0.028555**	**1.9483**
		4	0.072388	44.5851	0.32774	12.9568	0.4541	75.3771	**0.04325**	**2.1771**
		8	0.16385	42.4252	0.33182	13.1555	0.6460	74.8618	**0.079427**	**3.6839**
	Joyful Yell	2	0.049515	46.0065	1.0052	13.1767	0.2645	55.3462	**0.034162**	**2.1722**
		4	0.069491	43.4654	0.99844	13.0798	0.2770	42.4242	**0.04818**	**2.3335**
		8	0.11255	40.9818	1.0032	13.1334	0.4589	38.8507	**0.073515**	**2.5774**
	Thai Statue	2	0.03307	48.7666	0.30254	12.0112	0.2371	69.6653	**0.019305**	**2.3639**
		4	0.046843	45.6104	0.30597	12.0833	0.2792	77.7622	**0.029185**	**2.4529**
		8	0.089646	43.7591	0.31316	12.4814	0.2847	64.3520	**0.041307**	**2.9642**
Median		2	0.036853	46.2107	0.44223	13.503	0.12186	45.0229	**0.025062**	**2.2681**
		4	0.057904	43.5029	0.51448	13.5977	0.18929	41.3767	**0.039879**	**2.3932**
		8	0.10844	41.6178	0.52551	13.8582	0.31159	41.3102	**0.073722**	**3.1491**
Mean		2	0.038326	45.28	0.54626	14.3246	0.11516	42.7392	**0.027843**	**2.554**
		4	0.056267	42.5321	0.67519	19.6911	0.18488	40.6331	**0.042075**	**2.6984**
		8	0.11193	40.6364	0.60043	14.5123	0.29819	40.1205	**0.071228**	**3.2446**

TABLE 3: Comparison results on the various multi-shot depth super-resolution methods. $n = 20$ are used for this task. For [5], we randomly select one image out of 20. $\gamma = 0.01$ is used for Ours in all experiments.

Fig. 18: Qualitative comparison of our UPS results against state-of-the-art methods on four synthetic datasets using a scaling factor of 4.

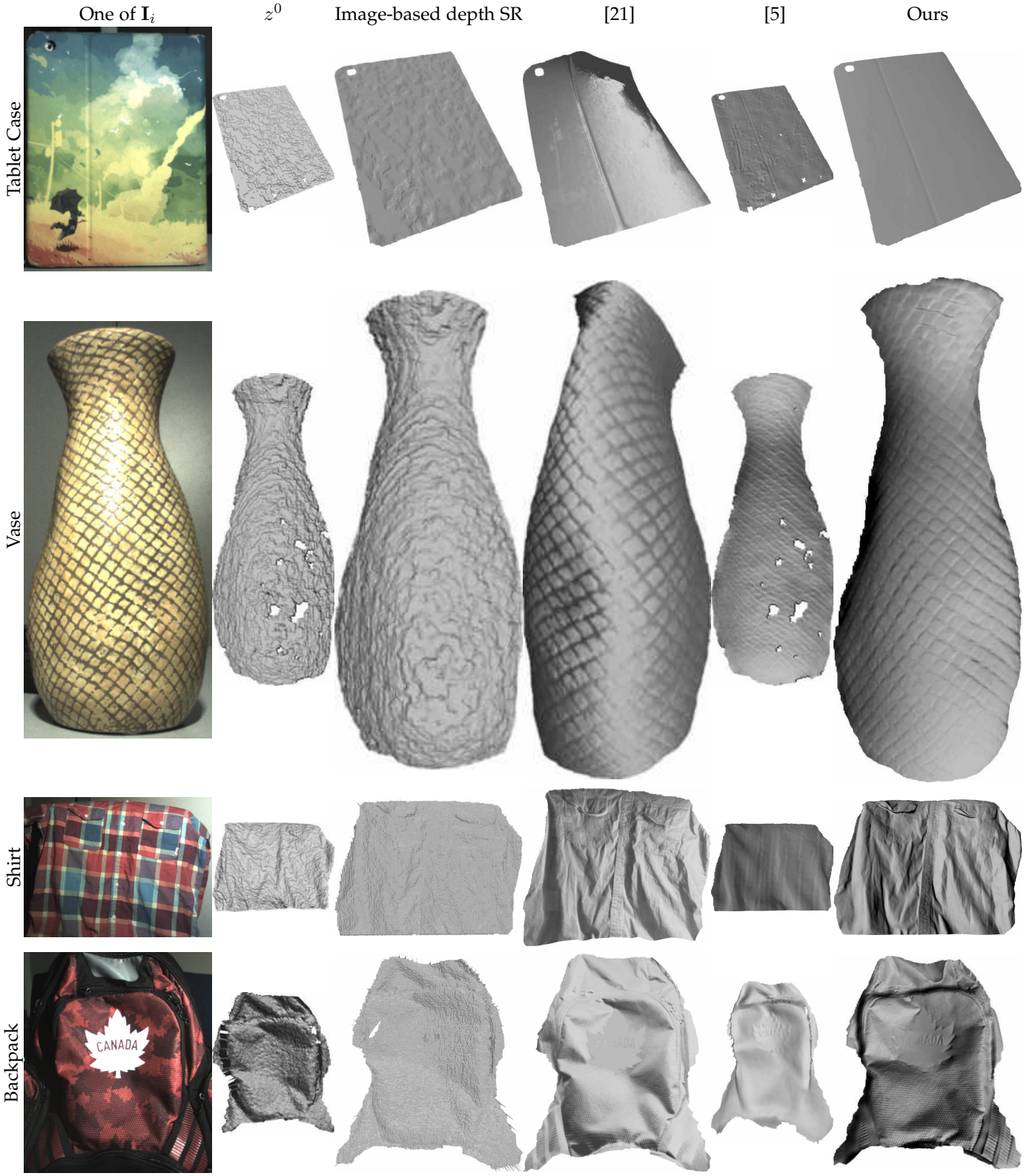

Fig. 19: Comparison between the proposed multi-shot method and 3 state-of-the-art methods, on real-world datasets. These results confirm the conclusion of the synthetic experiments in Figure 19.