Supplementary Material of
Recovering Real-World Reflectance Properties and Shading From HDR Imagery

Bjoern Haefner[1,2] Simon Green[2] Alan Oursland[2] Daniel Andersen[2]
Michael Goesele[2] Daniel Cremers[1] Richard Newcombe[2] Thomas Whelan[2]
[1]Technical University of Munich [2]Facebook Reality Labs Research
{bjoern.haefner, cremers}@tum.de,
{simongreen, ours, andersed, goesele, newcombe, twhelan}@fb.com

A. Details on importance sampling (Section 3.4.2)

This section discusses the technical implementation details on an efficient sampling strategy to evaluate

$$I_{\text{nd}}\left(\boldsymbol{p};\varphi,\psi\right) := \int_{\mathcal{H}^2} f_{\text{nd}}\left(\boldsymbol{x},\omega,\omega_{\text{o}};\varphi,\psi\right) L\left(\boldsymbol{x},\omega\right) \left\langle \omega,\boldsymbol{n}\right\rangle \mathrm{d}\omega. \tag{1}$$

Importance sampling is a powerful tool to estimate the integral in (1) and we refer the interested reader to [9], Chapter 13 for the mathematical reasoning behind it. The Monte-Carlo estimator used for Eqn. (1) using the non-diffuse part of the simplified Disney BRDF is the finite sum of the form,

$$\mathcal{I}_{\text{nd}}(\boldsymbol{p};\varphi,\psi)=\frac{1}{N}\sum_{j=1}^{N}\frac{f_{\text{nd}}(\boldsymbol{x},\boldsymbol{\Omega}_j,\omega_{\text{o}};\varphi,\psi)L(\boldsymbol{x},\boldsymbol{\Omega}_j)\langle\boldsymbol{\Omega}_j,\boldsymbol{n}\rangle}{p(\boldsymbol{\Omega}_j)}, \tag{2}$$

where the random variables $\boldsymbol{\Omega}_j$ are samples drawn from the probability density function $p(\omega)$. We expect, i.e. given enough samples N,

$$\mathbb{E}\left[\mathcal{I}_{\text{nd}}\left(\boldsymbol{p};\varphi,\psi\right)\right] = I_{\text{nd}}\left(\boldsymbol{p};\varphi,\psi\right), \quad \forall \boldsymbol{p},\varphi,\psi. \tag{3}$$

The probability density function used in our approach is

$$p\left(\omega\right) = \frac{1}{2}\frac{\left|\langle\omega,\boldsymbol{n}\rangle\right|}{\pi} + \frac{1}{2}\frac{D\left(\varphi\right)\left|\langle h,\boldsymbol{n}\rangle\right|}{4\left|\langle\omega_{\text{o}},h\rangle\right|}. \tag{4}$$

To evaluate (2), we need to be able to sample random variables $\boldsymbol{\Omega}_j$ from $p(\omega)$ and we realize this the following way: Given the j-th observation of random variables following a uniform distribution over $[0,1)$, $X_0^j, X_1^j, X_2^j \sim \mathcal{U}(0,1)$, we calculate a sample of incident direction as

$$\boldsymbol{\Omega}_j = \begin{cases} T s_{\mathcal{H}^2}(X_1^j, X_2^j), & X_0^j < \frac{1}{2}, \\ R(\omega_{\text{o}}, T h_s(X_1^j, X_2^j)), & \text{else}, \end{cases} \tag{5}$$

where $R(\omega_{\text{o}},h) = 2\left\langle\omega_{\text{o}},h\right\rangle h - \omega_{\text{o}}$ resembles the reflection of ω_{o} on h, and $T := (\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_3) \in \mathbb{R}^{3\times3}$ is an orthonormal basis transform in the normal's coordinate system, aligning the north pole of $\mathcal{H}^2$ with the normal $\boldsymbol{n}$,

$$\mathbf{t}_1 = \mathbf{t}_2 \times \mathbf{t}_3 \tag{6}$$

$$\mathbf{t}_2 = \begin{cases} \frac{(-\boldsymbol{n}_y,\boldsymbol{n}_x,0)^\top}{\|(-\boldsymbol{n}_y,\boldsymbol{n}_x,0)\|}, & \left|\boldsymbol{n}_x\right| > \left|\boldsymbol{n}_z\right|, \\ \frac{(0,-\boldsymbol{n}_z,\boldsymbol{n}_y)^\top}{\|(0,-\boldsymbol{n}_z,\boldsymbol{n}_y)\|}, & \text{else}, \end{cases} \tag{7}$$

$$\mathbf{t}_3 = \boldsymbol{n}. \tag{8}$$

To sample the diffuse lobe of the BRDF (the case when $X_0^j < \frac{1}{2}$ in (5)), we generate random samples on the upper hemisphere $\mathcal{H}^2$ using $s_{\mathcal{H}^2} : [0,1)^2 \to \mathcal{H}^2$,

$$s_{\mathcal{H}^2}(x_1,x_2) = \begin{pmatrix} s_1 \\ s_2 \\ \sqrt{\max\left(0, 1-s_1^2-s_2^2\right)} \end{pmatrix}, \tag{9}$$

with $s_1 := \sqrt{x_1}\cos\left(2\pi x_2\right)$ and $s_2 := \sqrt{x_1}\sin\left(2\pi x_2\right)$. The non-diffuse lobe of the BRDF (the case when $X_0^j \geq \frac{1}{2}$ in (5)) is sampled using $h_s : [0,1)^2 \to \mathcal{H}^2$,

$$h_s(x_1,x_2) = \begin{pmatrix} \sin\left(\theta\right)\cos\left(2\pi x_1\right) \\ \sin\left(\theta\right)\sin\left(2\pi x_1\right) \\ \cos\left(\theta\right) \end{pmatrix}, \tag{10}$$

with $\theta := \cos^{-1}\left(\sqrt{\frac{1-x_2}{1+(\varphi^2-1)x_2}}\right)$.

B. Details on capturing process (Section 4)

We perform two full scans of a room sized environment, where camera poses are recovered using SLAM [2, 7], geometry is reconstructed with [8]. In a post-processing step we fill larges holes manually or using Poisson reconstruction [4, 5] and repair any remaining issues automatically using [3]. Object segmentation is carried out in a manual step.

C. Further quantitative results on albedo and shading estimation validation (Section 4.1)

Additional qualitative results of the albedo and shading estimation applied to real-world data sets are shown in Figure 1.

D. Further quantitative results on specular appearance estimation validation (Section 4.2)

This section discusses **further quantitative results** of the specular appearance estimation for novel views. The main paper depicts quantitative results as well as a qualitative visualization of notable peaks in the corresponding graph on the "Office 1" sequence of [11]. For full insight, we show the results on the remaining sequences of the Replica datasets [11], cp. Figure 2 for insight in the "Office" sequences using the L_2 metric, Figure 3 for insight in the "Office" sequences using the FLIP evaluator [1], Figure 4 for insight in the "Room" sequences using the L_2 metric, and Figure 5 for insight in the "Room" sequences using the FLIP evaluator [1].

Figures 2 and 3 show results on the "Office" sequences of [11], they consist of 1293, 2117, 2459, and 2101 frames, which include 24, 38, 43, and 31 target frames, respectively, thus incorporating 1269, 2079, 2416, and 2070 novel, unseen viewpoints. The "Office 0", "Office 2", "Office 3", and "Office 4" sequences have their largest improvement and deterioration for the L_2 error at frames $(1264, 243)$, $(629, 910)$, $(1799, 2319)$, and $(1899, 1903)$, respectively and are visualized for qualitative inspection in Figure 2. The same sequences have their largest improvement and deterioration for the FLIP evaluator [1] at frames $(1263, 163)$, $(1801, 1107)$, $(1799, 637)$, and $(1899, 1929)$, respectively and are visualized for qualitative inspection in Figure 3.

Figure 4 shows results on the "Room" sequences of [11], they consist of 2642, 1828, and 1789 frames, which include 51, 33, and 34 target frames, respectively, thus incorporating 2591, 1795, and 1755 novel, unseen viewpoints. The "Room 0", "Room 1", and "Room 2" sequences have their largest improvement and deterioration for the L_2 error at frames $(1856, 1203)$, $(31, 114)$, and $(604, 118)$, respectively and are visualized for qualitative inspection in Figure 4. The same sequences have their largest improvement and deterioration for the FLIP evaluator [1] at frames $(1552, 1204)$, $(31, 83)$, and $(656, 118)$, respectively and are visualized for qualitative inspection in Figure 5.

Overall, the average reconstruction error decreases for all experiments and validates our findings described in Section 4.2.1 on a larger scale. This can also be seen qualitatively; note the overall increase of realism, for the improvements, due to view-dependent effects, while the deteriorations seem to be only slightly worse than the baseline, but still visually pleasing to the human eye – an effect desired in AR/VR/MR applications.

Further quantitative results on the Room sequences of [11] are shown in Figure 6. For specular highlights that seem too wide such as the vase in "Room 0" our reconstructions still look more faithful compared to a purely diffuse one. Notice that the anisotropic BRDF of the window blinds in "Room 2" is difficult to recover with our approach as we do not model this effect. Instead, we estimate an isotropic approximation of it, which still looks realistic.

Robustness against inaccurate geometry can affect the final reconstruction in accuracy and realism. Figure 3 "Office 3" shows how specularities are misplaced and BRDF estimates too rough, if the geometry (clock) is inaccurate at the location of reflection. Figure 4 and 5 "Room 1" and Figure 6 "Room 0" and "Room 1" show results were different levels of deteriorated geometry affects the non-diffuse BRDF estimate. While the vase in "Room 1" is almost diffuse, the vase in "Room 0" shows specular reflections, although not as strong as the capture. The reason for both failures are caused by an estimated specular highlight having no overlap with the genuine reflection, cp. the error maps in Figure 6 "Room 0", the specular reflections are not perfectly superimposed.

E. Further Relighting results (Section 4.3)

Further renderings under novel lighting with artificially placed objects are shown in Figure 7.

F. Attached video file

The video file attached to the supplementary material shows a number of video renderings of our results as well as comparisons to the baseline. This video is encoded with the H.264 codec in an MP4 container. Some of the images shown in the video will have a somewhat grainy appearance - this is caused by the relatively simple path tracer we implemented for visualizing the results of our approach, rather than being an intrinsic part of the estimated appearance.

References

[1] P. Andersson, J. Nilsson, T. Akenine-Möller, M. Oskarsson, K. Åström, and M. D. Fairchild. Flip: a difference evaluator for alternating images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques (HPG 2020)*, 3(2), 2020. 2, 5, 7, 8

[2] J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625, 2017. 1

[3] W. Jakob, M. Tarini, D. Panozzo, and O. Sorkine-Hornung. Instant field-aligned meshes. *ACM Trans. Graph.*, 34(6):189–1, 2015. 1

Figure 1. We deploy our albedo and shading estimation on challenging real-world "Room" data sets of the Replica data set [11] and are able to estimate per-texel albedo and shading information, using the reconstructed mesh and lit diffuse HDR texture only.

[4] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006. 1

[5] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 1

[6] M. McGuire. Computer graphics archive. https://casual-effects.com/data, July 2017. 8

[7] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015. 1

[8] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, 2011. 1

[9] M. Pharr, W. Jakob, and G. Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016. 1

[10] The Stanford 3D Scanning Repository. http://graphics.stanford.edu/data/3Dscanrep/. 8

[11] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma, A. Clarkson, M. Yan, B. Budge, Y. Yan, X. Pan, J. Yon, Y. Zou, K. Leon, N. Carter, J. Briales, T. Gillingham, E. Mueggler, L. Pesqueira, M. Savva, D. Batra, H. M. Strasdat, R. D. Nardi, M. Goesele, S. Lovegrove, and R. Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 2, 3, 4, 5, 6, 7, 8

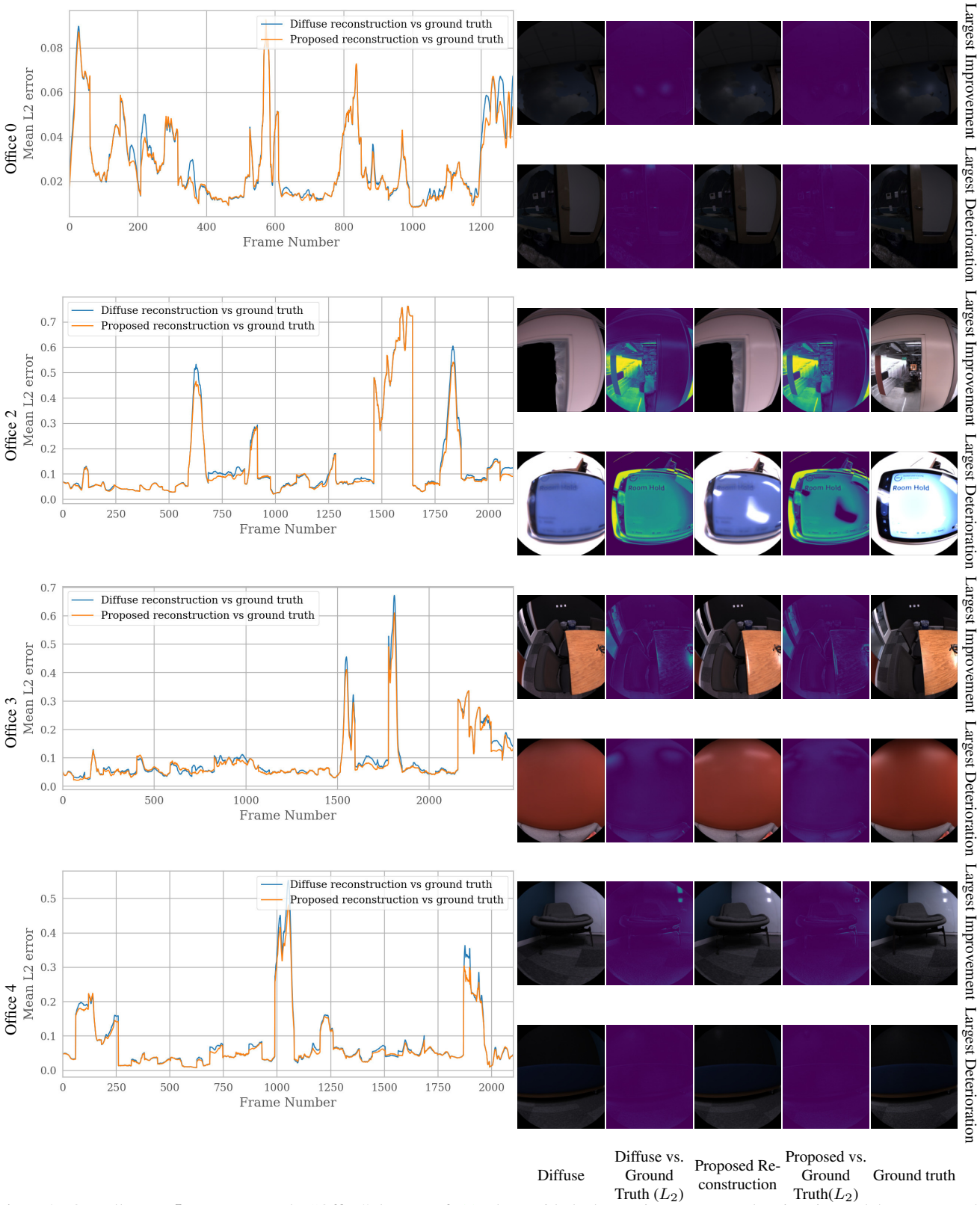

Figure 2. Overall Mean L_2 error across the "Office" datasets of [11] along with the largest improvement, deterioration and the corresponding L_2 error maps.

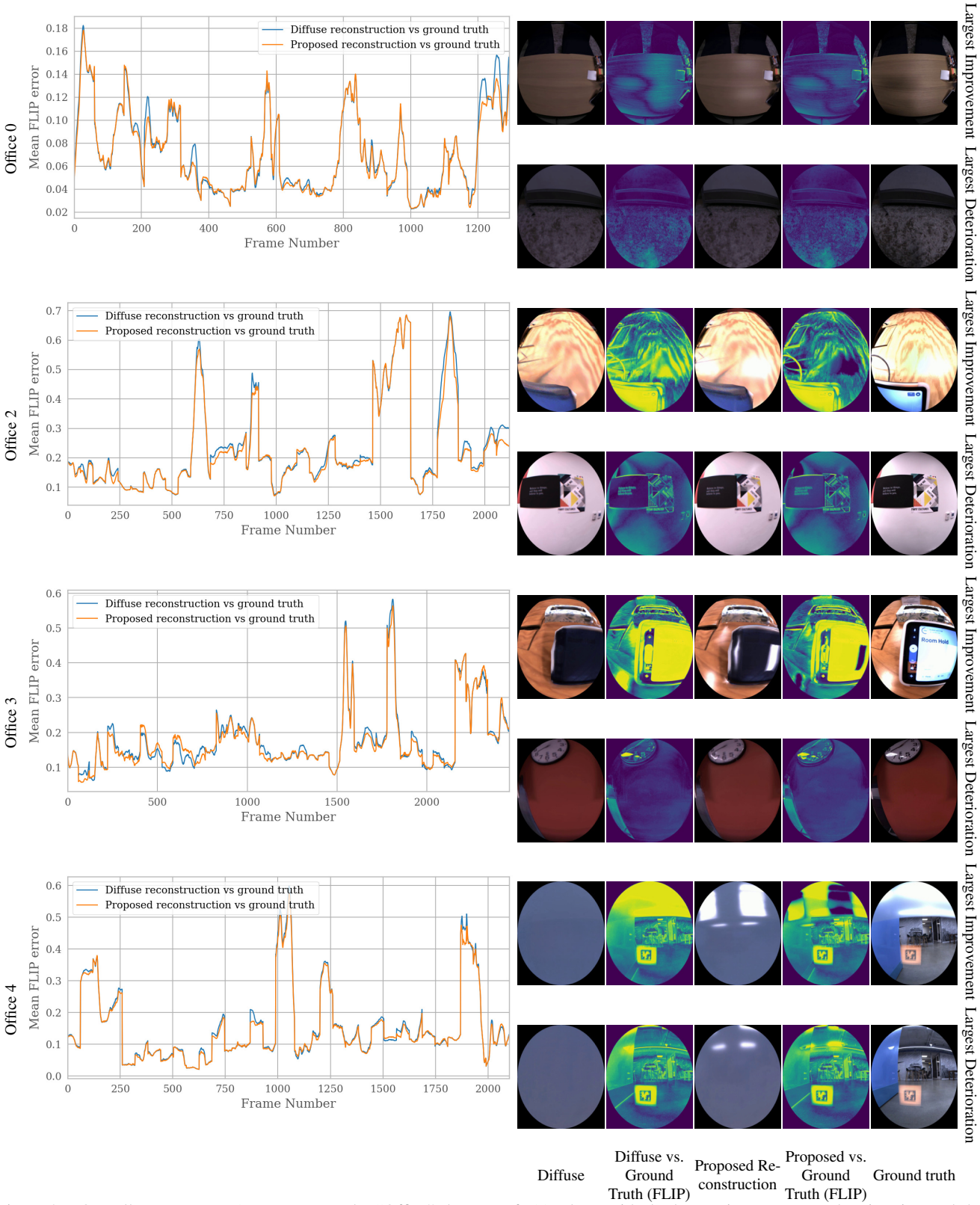

Figure 3. Overall Mean FLIP [1] error across the "Office" datasets of [11] along with the largest improvement, deterioration and the corresponding FLIP error maps.

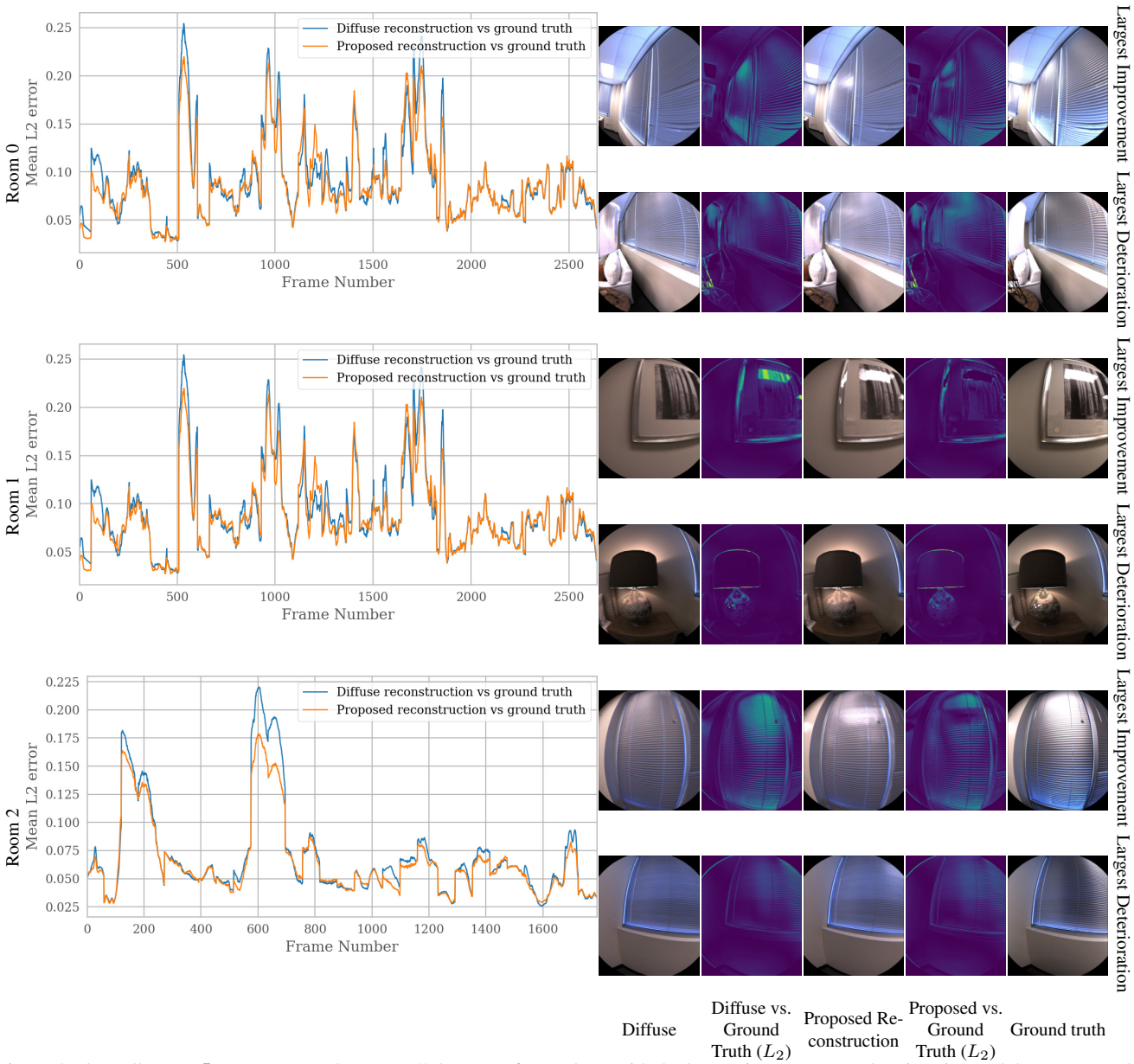

Figure 4. Overall Mean L_2 error across the "Room" datasets of [11] along with the largest improvement, deterioration and the corresponding L_2 error maps.

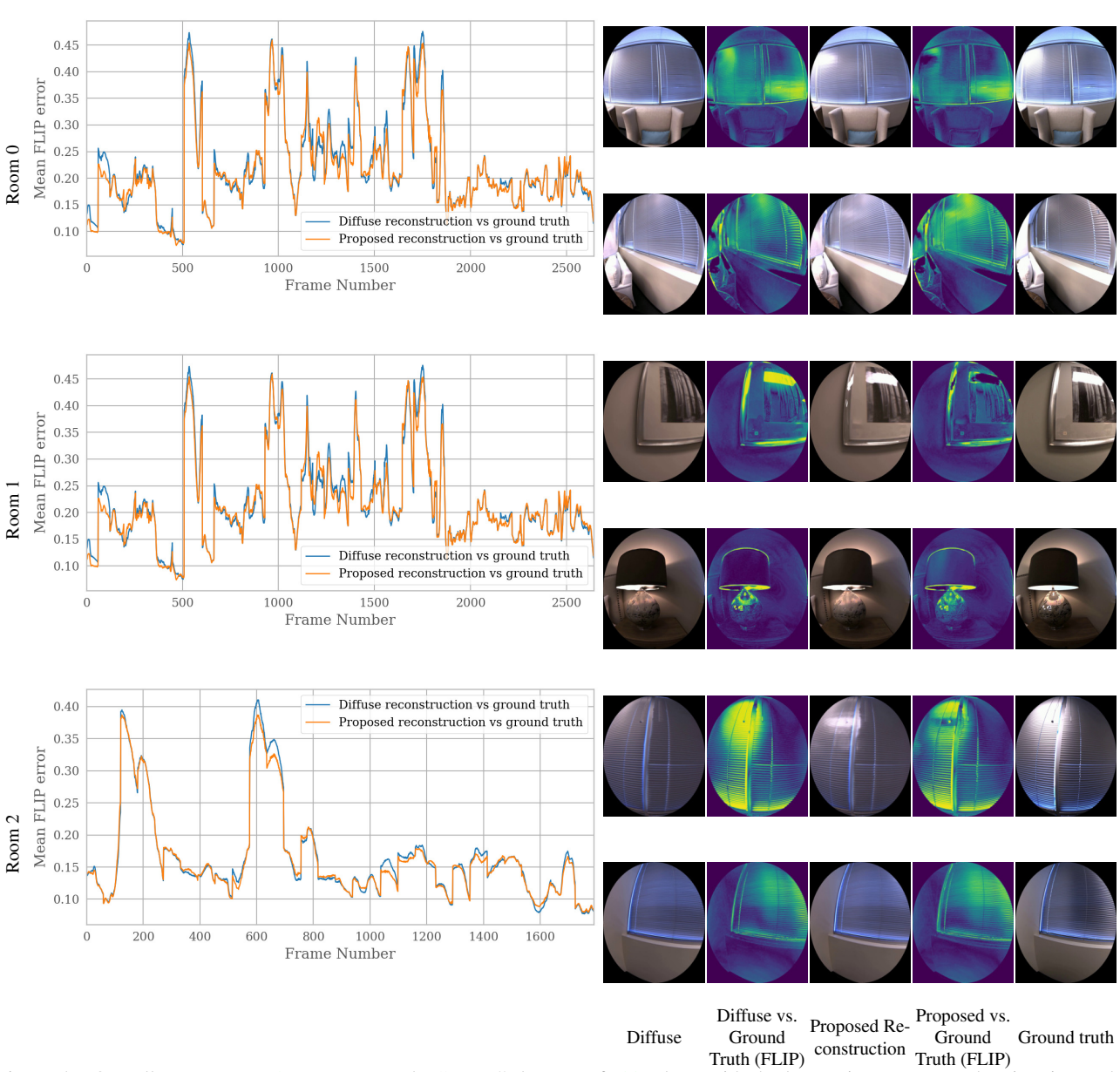

Figure 5. Overall Mean FLIP [1] error across the "Room" datasets of [11] along with the largest improvement, deterioration and the corresponding FLIP error maps.

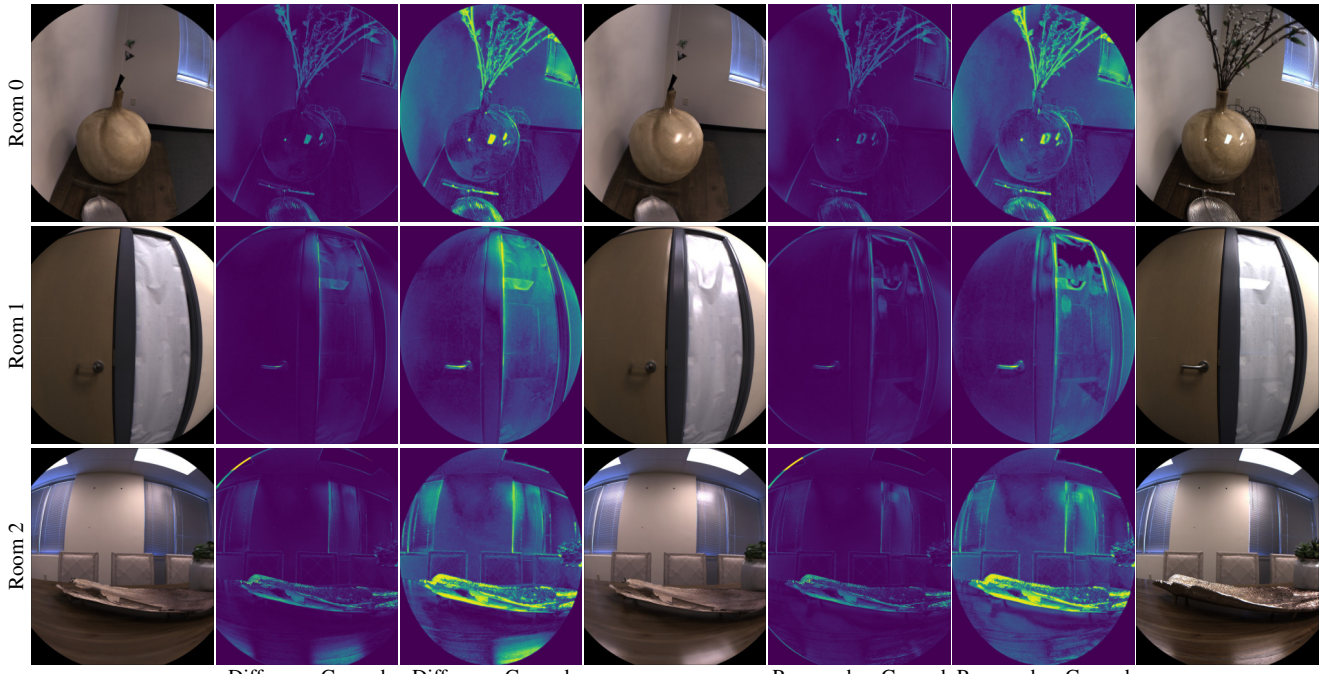

Figure 6. Side-by-side comparisons between the diffuse and the proposed reconstruction along with the ground truth and the corresponding L_2 errors and FLIP evaluator [1]. Adding the proposed specular appearance estimate makes reconstructions more realistic.

Figure 7. Complete synthetic relighting of different data sets (Office 0, Room 0 and Room 1 [11]) with additional virtually placed objects [10, 6].