Supplementary Material | RRM: Relightable assets using Radiance guided Material extraction

Diego Gomez[1][0009−0005−2847−8617], Julien Philip[2][0000−0003−3125−1614], Adrien Kaiser[2][0000−0002−5998−3932], and Élie Michel[2][0000−0002−2147−3427]

[1] École polytechnique
[2] Adobe Research

1 BSDF Model

The function f_r in Section "PBR module" of the main paper is our spatially varying BRDF model based on the Torrance–Sparrow model with a normal distribution function based on the Beckmann–Spizzichino model. We write it as,

$$f_r(\omega_o, \omega_i; \beta) = f_{diffuse}(\gamma) + f_{specular}(\omega_o, \omega_i; \beta) \tag{1}$$

Where, $\mathrm{f}_{diffuse}(\gamma) = \gamma\pi$
$f_{specular}(\omega_o, \omega_i; \beta) = \frac{D(h,\rho)G(\omega_o,\omega_i,\rho)F_{Schlick,\hat{x}}(\omega_o,h)}{4\langle n,\omega_i\rangle\langle n,\omega_o\rangle}$ where h is the half-vector between ω_o and ω_i, D the Beckmann–Spizzichino model [1],

$$G(\omega_o, \omega_i, \rho) = G1(\omega_o, \rho)G1(\omega_i, \rho)$$

with G_1 the Smith's masking-shadowing function, and

$$F_{Schlick,\hat{x}}(\omega_o, h) = F_0 + (1 - F_0)(1 - \langle\omega_o, h\rangle^5)$$

the Schlick. [5] approximation for the Fresnel term.

Multiple Importance Sampling. The BRDF sampling we implement involves the sampling of the normal distribution function $\mathrm{D}_\omega(\omega_h) = \frac{D(\omega_h)G_1(\omega_h)\langle\omega,\omega_h\rangle_+}{\cos\theta}$ To perform light sampling, we retrieve the environment(s) map(s) at the current iteration. For a given environment map we map it into a 2D probability distribution by considering for each cell the corresponding normalized light intensity (sum of the RGB channels). Each cell is associated to a light direction given our far away lighting assumption. We retrieve the pdf of a given direction in polar coordinates as, $\mathrm{p}_{light}(\theta, \phi) = \frac{p_{envmap}(u,v)N_l}{2\pi^2\sin(\theta)}$ where (θ, ϕ) is the polar coordinates of the direction that we query, (u, v) is the corresponding coordinates of the associated cell and N_l is the total number of pixels in our envmap ($N_l = 512 \times 1024$ in our case).

The MIS algorithm is thouroughly explored in [6], let us give a brief overview. We use this method to evaluate the function $\mathrm{g}(\omega) = L_i(\hat{x}, \omega) f_r(\omega_o, \omega; \beta) \langle\omega, n\rangle_+$ Consider n sampling strategies with corresponding density functions $\{p_i\}_{i=1}^n$.

Call $X_{i,j}$ the j-th sample from strategy i. The *multi-sample estimator* is given by,

$$F = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} w_i(X_{i,j}) \frac{g(X_{i,j})}{p_i(X_{i,j}} \tag{2}$$

With a w_i the associated weighing function of strategy i. Note that it must follow,

- $\sum_{i=1}^{n} w_i(x) = 1$ for any x s.t. $g(x) \neq 0$.
- $w_i(x) = 0$ whenever $p_i(x) = 0$

Finally, we apply the above we our two sampling strategies: light and BRDF sampling.

Sampling Light directions. The integral in the rendering equation can only be evaluated numerically in general. Previous radiance field relighting methods such as TensoIR [2] or NeRFactor [7] use a uniform discretization of the set Ω of incoming directions, but decades of light transport research point to the use of Monte Carlo methods to retrieve high-frequency glossy effects. We thus employ a Multiple Importance Sampling (MIS) scheme [6], using both a prior based on the environment light and one based on the BRDF in order to select good light rays. We compare qualitatively our method with and without MIS in figure 2 and give more details about our importance sampling in the supplementary material.

2 Radiance Decomposition details

In order to enforce a meaningful decomposition of the radiance into view-independent and view-dependent components, c_i and c_d, we introduce a simple yet effective strategy. During training, we apply a dropout layer to the output of the view-dependent network $\mathcal{D}_{c_d}$ with a probability of $p = 0.01$, meaning we drop the view-dependent component of the radiance for 1% of the samples. The 1% of the time where it is not present, all the gradient of this loss goes towards the diffuse component effectively pushing it to reconstruct the full signal and thus capture as much radiance as it can and helps to disambiguate the split. Without this dropout there are infinitely many solutions to $c_d + c_i = O$ with O the observed values for a given 3D point.

3 Extra-Properties

We inherit the ability to input multiple unknown lighting conditions from TensoIR [3], as shown in Fig. 1. We did not explore this feature in this paper since it is not part of our contributions. It is nonetheless to be expected that the aggregation of different light conditions will lead to better results since the ambiguity in the system is greatly reduced.

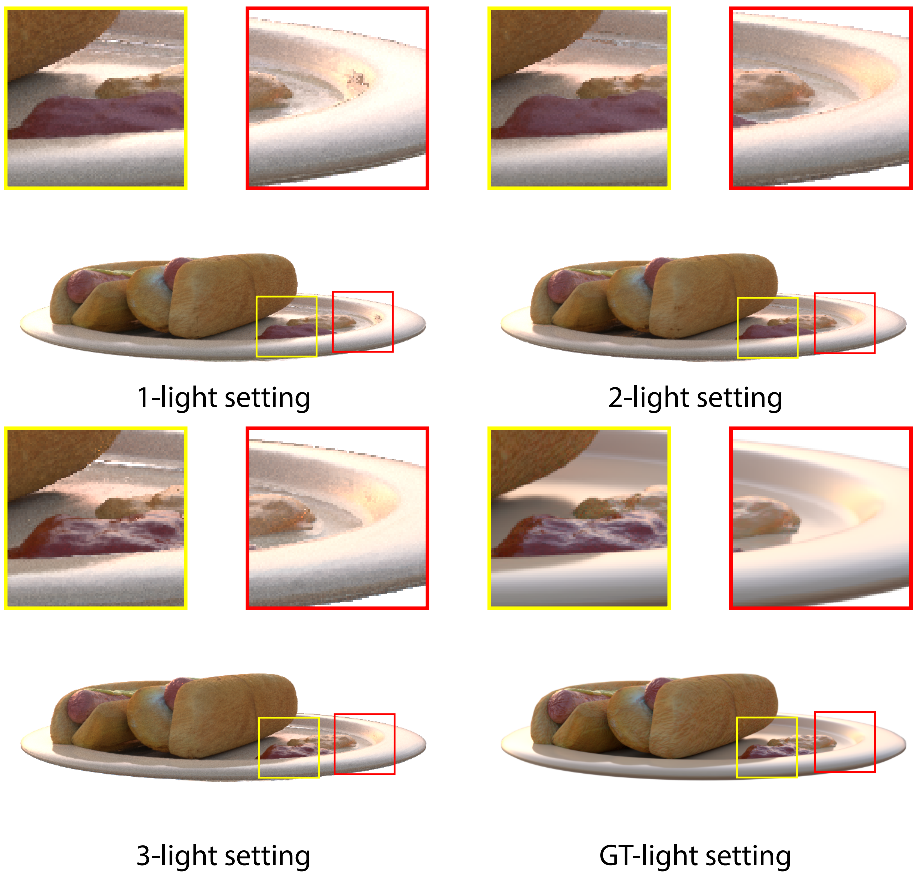

Fig. 1. Similarly to TensoIR [3], our method gradually improves as we increase the number of unknown light settings under which input images are taken.

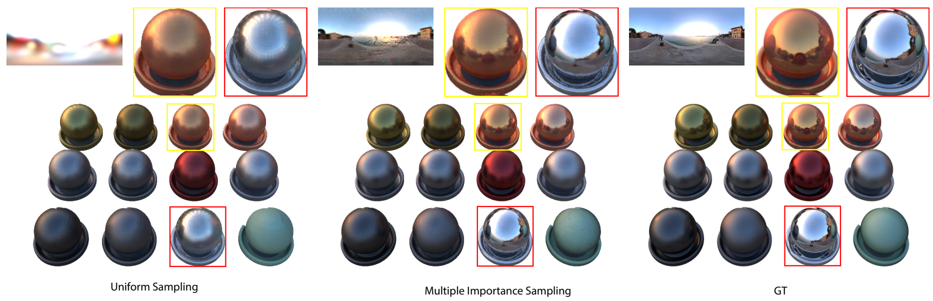

Fig. 2. Novel-View Synthesis and retrieved environment maps for models trained using Uniform (left) or Multiple Importance Sampling (middle). Uniform sampling limits the range of glossy effects the model can learn as seen when compared to the Ground Truth (right).

Fig. 3. Our model fails to retrieve the specular effects on the coffee scene, notably on the cup. This is a hard scene due to the strong near-lighting effects. We note that our model manages to "understand" the scene better than NMF [4] which results in more plausible normals.

Table 1. PSNR ↑ on novel view synthesis tasks (higher is better).

Method	Teapot	Toaster	Car	Ball	Coffee	Helmet	Materials	Lego	Hotdog	Ficus	Armadillo
Ours	41.75	26.33	29.77	36.07	29.71	30.74	29.04	31.90	25.22	29.57	37.89
Ours separate ρ	41.63	26.00	29.66	35.73	30.44	30.71	29.46	31.88	29.22	29.57	37.80
Ours No Decomposition	41.43	22.44	28.07	36.11	30.41	28.57	29.45	31.89	27.77	29.65	37.79
Ours No Supervision	41.65	26.26	29.65	35.93	29.55	31.16	29.39	31.95	OOM	29.68	37.67
TensoIR	42.44	19.69	26.52	N/A	31.22	25.94	26.80	34.700	36.820	29.780	39.050
NMF	45.29	27.52	30.28	38.41	31.47	34.38	31.19	32.98	35.23	29.24	N/A

4 Quantitative results detail

Tables 1,2, 3 and 4 give the results of the novel view synthesis test. A notable thing is that the no supervision model seems to yield heavier assets, unfortunately we did not perform a formal study to confirm this. We see however that in the hotdog scene this ablation fails with an OOM error message.

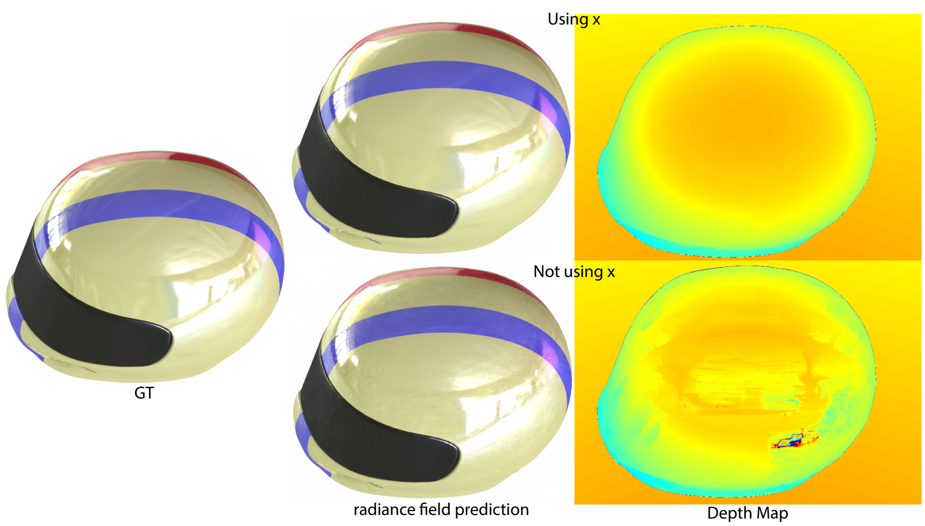

Fig. 4. If we do not give an encoding of the location x our radiance method fails in some scenes with complex lighting information.

References

1. Beckmann, P.: Spizzichino, the scattering of electromagnetic waves from rough surfaces (1963)
2. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. In: European Conference on Computer Vision. pp. 333–350. Springer (2022)

Table 2. SSIM ↑ on novel view synthesis tasks (higher is better).

Method	Teapot	Toaster	Car	Ball	Coffee	Helmet	Materials	Lego	Hotdog	Ficus	Armadillo
Ours	0.9843	0.9217	0.9429	0.9786	0.9100	0.9075	0.9452	0.9591	0.8957	0.9711	0.9822
Ours separate ρ	0.9844	0.9173	0.9401	0.9779	0.9171	0.9070	0.9481	0.9586	0.9352	0.9699	0.9820
Ours No Decomposition	0.9874	0.8635	0.9211	0.9798	0.9122	0.8875	0.9486	0.9560	0.9383	0.9702	0.9817
Ours No Supervision	0.9881	0.9196	0.9428	0.9784	0.9115	0.9128	0.9471	0.9577	OOM	0.9703	0.9807
TensoIR	0.9953	0.8222	0.9220	N/A	0.9656	0.9107	0.9274	0.968	0.976	0.973	0.985
NMF	0.996	0.917	0.951	0.983	0.960	0.969	0.959	0.963	0.964	0.952	N/A

Table 3. LPIPS ↓ on novel view synthesis tasks (lower is better).

Method	Teapot	Toaster	Car	Ball	Coffee	Helmet	Materials	Lego	Hotdog	Ficus	Armadillo
Ours	0.0449	0.1473	0.0704	0.1350	0.1672	0.1921	0.0675	0.0771	0.1498	0.0504	0.0329
Ours separate ρ	0.0452	0.1470	0.0730	0.1397	0.1622	0.1919	0.0649	0.0726	0.1241	0.0504	0.0333
Ours No Decomposition	0.0436	0.1994	0.0965	0.1311	0.1654	0.2079	0.0656	0.0797	0.1226	0.0540	0.0341
Ours No Supervision	0.0430	0.1459	0.0686	0.1346	0.1674	0.1850	0.0660	0.0762	OOM	0.0539	0.0355
TensoIR	0.0177	0.2166	0.0724	N/A	0.1360	0.1586	0.0791	0.037	0.045	0.041	0.039
NMF	0.010	0.104	0.034	0.046	0.069	0.055	0.026	0.024	0.046	0.044	N/A

3. Jin, H., Liu, I., Xu, P., Zhang, X., Han, S., Bi, S., Zhou, X., Xu, Z., Su, H.: Tensoir: Tensorial inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 165–174 (June 2023)
4. Mai, A., Verbin, D., Kuester, F., Fridovich-Keil, S.: Neural microfacet fields for inverse rendering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 408–418 (October 2023)
5. Schlick, C.: An inexpensive brdf model for physically-based rendering. In: Computer graphics forum. vol. 13, pp. 233–246. Wiley Online Library (1994)
6. Veach, E., Guibas, L.J.: Optimally combining sampling techniques for monte carlo rendering. In: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. pp. 419–428 (1995)
7. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (ToG) **40**(6), 1–18 (2021)

Table 4. MAE ↓ on geometry normal extraction (lower is better).

Method	Teapot	Toaster	Car	Ball	Coffee	Helmet	Materials	Lego	Hotdog	Ficus	Armadillo
Ours	0.38	2.45	1.51	0.47	3.62	1.19	1.40	5.97	3.16	3.78	1.85
Ours separate ρ	0.38	2.54	1.50	0.46	3.55	1.18	1.37	5.41	3.37	3.75	1.83
Ours No Decomposition	0.44	6.58	1.68	0.36	3.14	3.17	1.43	5.45	3.33	3.94	1.90
Ours No Supervision	0.40	2.50	1.49	0.44	3.75	1.23	1.41	5.43	OOM	3.90	2.05
TensoIR	1.05	8.17	2.90	N/A	4.57	7.45	3.02	5.980	4.050	4.420	1.950
NMF	0.752	4.474	2.598	1.563	5.352	1.924	2.868	8.452	3.546	4.949	N/A

Table 5. Tests done in with our method to compare against TensoIR in the relighting task

Ours	Armadillo	Ficus	Lego	Hotdog
W/ Forest Envmap				
PSNR	34.02	25.57	26.76	28.93
SSIM	0.9733	0.9541	0.9324	0.9249
LPIPS	0.0475	0.0657	0.0942	0.1204
W/ City Envmap				
PSNR	30.85	24.4	24.75	27.05
SSIM	0.9658	0.9385	0.9151	0.9369
LPIPS	0.0484	0.0688	0.0919	0.0935
W/ Fireplace Envmap				
PSNR	31.9	22.73	23.62	29.67
SSIM	0.9535	0.9394	0.8775	0.8894
LPIPS	0.0539	0.078	0.1188	0.1388
W/ Bridge Envmap				
PSNR	33.59	24.93	25.57	26.29
SSIM	0.9733	0.9487	0.9232	0.9138
LPIPS	0.0488	0.0679	0.0984	0.1237
W/ Night Envmap				
PSNR	33.49	24.49	26.96	31.78
SSIM	0.9745	0.9552	0.8506	0.8849
LPIPS	0.0519	0.0728	0.0826	0.1015

Table 6. Ablation of our method in the relighting task. Note that the “No Supervision" Ablation failed in the hotdog scene when relighting with an OOM error. This goes to show an aspect of our method that we did not explore, the resulting size of the asset obtained. For each ablation we tested on 2 different environment maps.

Relight Ablation Details										
Ours no supervision	Armadillo	Ficus	Lego	Hotdog	Materials	Teapot	Toaster	Coffee	Car	Helmet
W/ Snow Envmap PSNR	28.96	22.98	27.01	OOM	25.37	28.94	20.8	18.83	29.16	25.98
SSIM	0.9547	0.9338	0.9239	OOM	0.9383	0.9764	0.8951	0.8807	0.9497	0.9107
LPIPS	0.0564	0.0693	0.1011	OOM	0.0688	0.0392	0.1351	0.1512	0.0596	0.1535
W/ Courtyard Envmap PSNR	30.44	24.45	25.89	OOM	25.46	25.59	20.85	19.33	27.97	25.2
SSIM	0.9526	0.9431	0.9179	OOM	0.9308	0.9701	0.8823	0.8567	0.9469	0.898
LPIPS	0.0556	0.071	0.0929	OOM	0.0631	0.0402	0.1332	0.1327	0.0586	0.1571
Ours no decomp	Armadillo	Ficus	Lego	Hotdog	Materials	Teapot	Toaster	Coffee	Car	Helmet
W/ Snow Envmap PSNR	28.94	22.95	27.19	28.33	25.39	29.29	16.34	19.46	28.69	23.63
SSIM	0.9562	0.9337	0.9234	0.8857	0.9391	0.9763	0.8532	0.8763	0.938	0.8928
LPIPS	0.0546	0.0694	0.1038	0.1429	0.0688	0.0392	0.1834	0.1531	0.0782	0.1721
W/ Courtyard Envmap PSNR	30.38	24.49	26.06	26.53	25.51	25.79	19.21	19.96	27.16	22.41
SSIM	0.9543	0.9428	0.9178	0.8917	0.9316	0.9698	0.8488	0.8564	0.9345	0.8776
LPIPS	0.0536	0.071	0.0934	0.1423	0.0623	0.0409	0.1758	0.1325	0.0785	0.1763
Ours rho sep	Armadillo	Ficus	Lego	Hotdog	Materials	Teapot	Toaster	Coffee	Car	Helmet
W/ Snow Envmap PSNR	28.74	23.14	25.41	28.33	25.45	29.6	20.82	16.98	25.87	27.12
SSIM	0.9585	0.9334	0.9226	0.8862	0.9397	0.9755	0.8965	0.8737	0.9385	0.9124
LPIPS	0.0524	0.0679	0.0969	0.1429	0.0677	0.0401	0.1371	0.1563	0.0621	0.1575
W/ Courtyard Envmap PSNR	30.14	24.6	24.39	26.63	25.6	27.78	20.91	17.53	27.39	25.82
SSIM	0.9563	0.9421	0.9085	0.8919	0.9319	0.9716	0.8842	0.8525	0.9445	0.8993
LPIPS	0.0523	0.0702	0.0962	0.1409	0.0626	0.0389	0.1377	0.1356	0.0603	0.1603
Ours	Armadillo	Ficus	Lego	Hotdog	Materials	Teapot	Toaster	Coffee	Car	Helmet
W/ Snow Envmap PSNR	28.85	22.67	25.14	28.72	25.34	29.77	21.3	16.84	28.53	27.27
SSIM	0.9585	0.9329	0.9188	0.8878	0.9394	0.9803	0.8995	0.8727	0.949	0.9313
LPIPS	0.0524	0.0686	0.1019	0.1423	0.0679	0.0372	0.1327	0.1561	0.0594	0.1483
W/ Courtyard Envmap PSNR	30.24	24.02	23.72	27.16	25.41	27.91	21.08	17.41	27.93	25.91
SSIM	0.9565	0.9409	0.9041	0.8959	0.9309	0.9761	0.8876	0.8533	0.947	0.9192
LPIPS	0.0523	0.0709	0.1074	0.1393	0.0631	0.0356	0.1338	0.1348	0.0588	0.147