Supplementary Material: Real-Time Denoising of Volumetric Path Tracing for Direct Volume Rendering

Jose A. Iglesias-Guitian,Prajita Mane and Bochang Moon *Member, IEEE*

Contents—This is the supplementary material for the publication titled *Real-Time Denoising of Volumetric Path Tracing for Direct Volume Rendering* [3]. The supplementary material includes:

- Supplementary results suite. An interactive website that allows a deep analysis of the evaluation scenarios.
- Supplementary video 1. Interactive sessions recorded live and demonstrating the capabilities of our denoising.
- Supplementary video 2. Detailed comparisons with state-of-the-art RAE [1] and SVGF [7] methods. Comparisons with RLS [5].
- The present document. We provide some DVR background useful to understand the VRI formulation used the main paper. Then we explain some details about how we trained the RAE network. Lastly, we include the whole set of time-wise plot comparisons for each experiment, having used both a distortion metric (i.e. PSNR) and a perceptual distance metric (LPIPS [8]).

1 Direct Volume Rendering in a Nutshell

Volume rendering for scientific visualization is commonly used to display a 2D projection of a 3D discretely sampled data set. Monte Carlo Direct Volume Rendering (MC-DVR) requires the computation of multiple light paths traveling from different light sources through the volume and finally reaching the virtual camera.

1.1 Optical properties of rectangular volume grids.

Each discrete element of an heterogeneous volume grid is often referred to as a *voxel*. A scalar value ρ, indicates the *density* of matter per unit volume. *Transfer functions* are used in volume rendering to map voxel densities with different appearance properties. The default transfer functions used for each data set appearing in our paper are shown in Figure 1. The *absorption* and *scattering* coefficients, respectively μ_a and μ_s, describe the probability of either absorbing or scattering radiance energy. The extinction coefficient $\mu_t = \mu_a + \mu_s$ indicates the combined probability of either event happening per unit distance. The behavior of these events might also be affected by the incident or outgoing direction of radiance. Moreover, voxels can contribute by adding new energy to the light path.

1.2 The Beer-Lambert law.

The *optical depth* or *optical thickness* τ along a line between two points in the volume $\mathbf{x}$ and $\mathbf{y}$ is defined as:

$$\tau(\mathbf{x},\mathbf{y}) = \int_0^y \mu_t(\mathbf{x} - s\omega)\,\mathrm{d}s \tag{1}$$

where ω represents the direction of that line delimited by $\mathbf{x}$ and $\mathbf{y} = \mathbf{x} - y\omega$. A physical interpretation of this measure can be how long light might travel before extinction. The *transmittance* $T(\mathbf{x},\mathbf{y})$ along the same path describes the attenuation of light in the volume between the same two points and it is defined by the following equation known as the Beer-Lambert law:

$$T(\mathbf{x},\mathbf{y}) = e^{-\tau(\mathbf{x},\mathbf{y})} \tag{2}$$

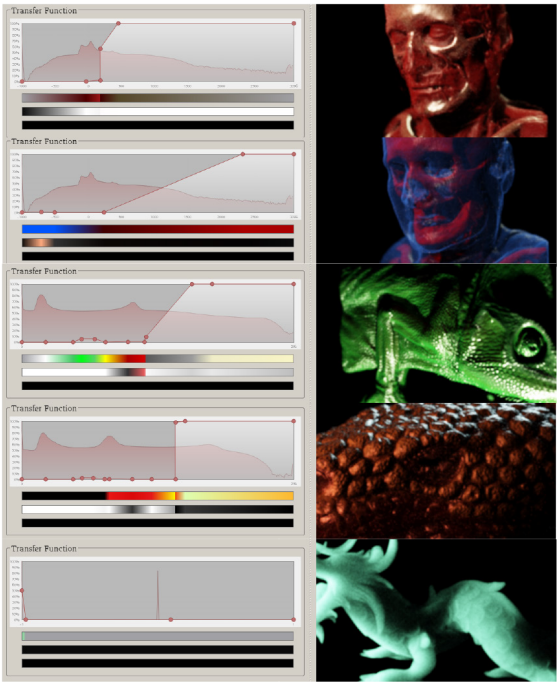
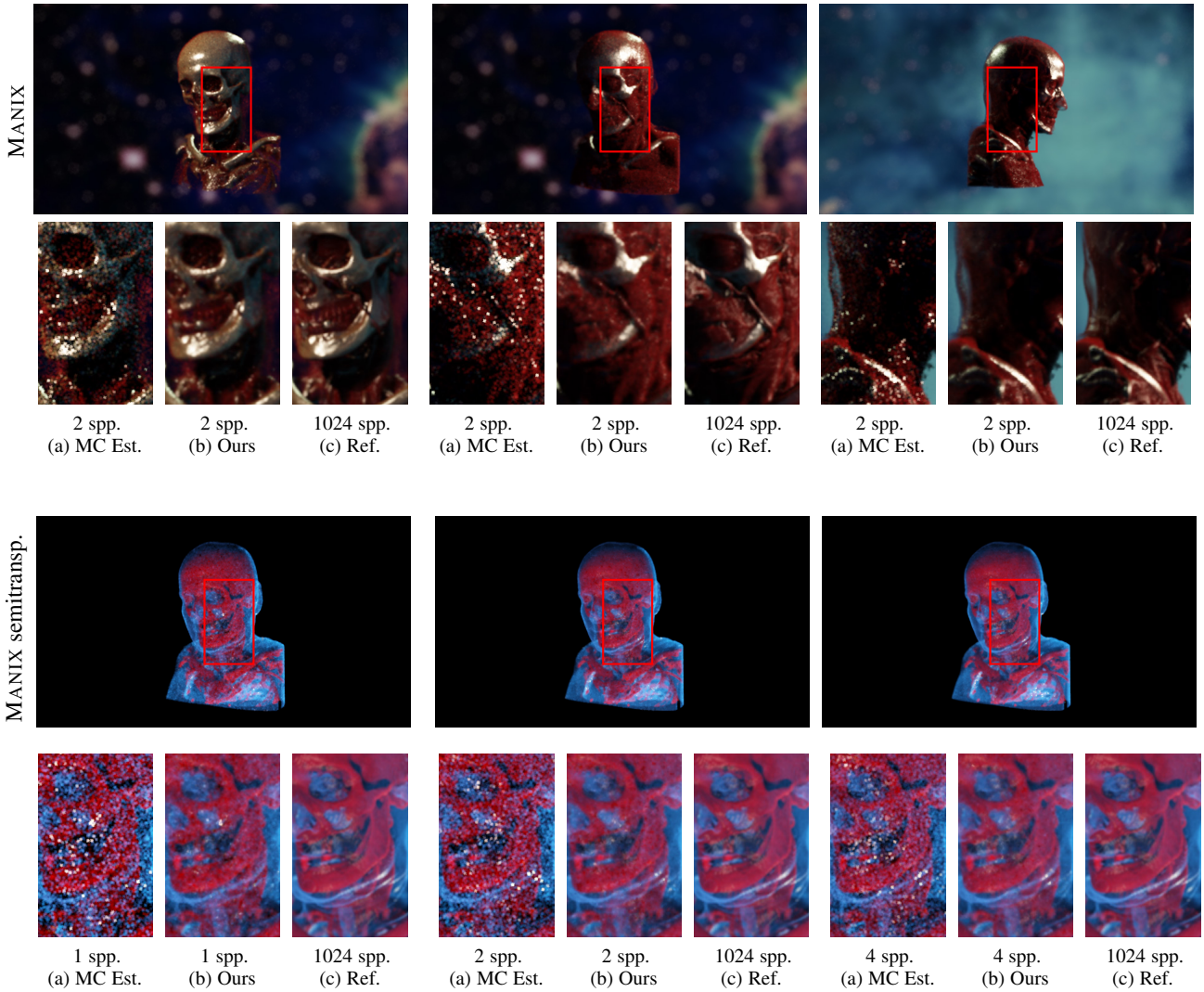


Fig. 1: Default transfer functions used in our experiments. The three colormaps shown for each dataset correspond with the diffuse, specular and emissive properties of the materials. From top to bottom we shown MANIX, MANIX with highly semitransparent isosurfaces, CHAMELEON, HELODERMA and semitransparent homogeneous volumetric DRAGON.



1.3 Transmittance estimation.

In order to estimate the transmittance $T(\mathbf{x}_i, \mathbf{x}_{i+1})$, the optical depth τ can be approximated using a Riemann summation (RS):

$$\langle \tau(t) \rangle_{RS} = \sum_{i=1}^{k} \mu_t(t_i) \Delta t, \tag{3}$$

where t_i is the i-th distance between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$ path vertices. Similarly to Kroes et al. [4], we utilize this approximation for our MC volume rendering.

1.4 The radiative transfer equation (RTE) framework.

The physically accurate simulation of all possible interactions of light with a participating medium remains a high dimensionality problem thoroughly described in [6]. The variation of the radiance energy traveling along a given direction ω through a differential volume element at point $\mathbf{x}$ is governed by the radiative transfer equation (RTE):

$$\begin{aligned}(\omega \cdot \nabla) L(\mathbf{x}, \omega) = &- \overbrace{\mu_a(\mathbf{x}) L(\mathbf{x}, \omega)}^{absorption} - \overbrace{\mu_s(\mathbf{x}) L(\mathbf{x}, \omega)}^{out-scattering} \\ &+ \underbrace{\mu_a(\mathbf{x}) L_e(\mathbf{x}, \omega)}_{emission} + \underbrace{\mu_s(\mathbf{x}) L_s(\mathbf{x}, \omega)}_{in-scattering}\end{aligned} \tag{4}$$

The first two terms in Equation 4 correspond to energy absorption and out-scattering, while the third and the fourth term represent energy gains due to emission and in-scattering properties, respectively. The in-scattered radiance L_s collects incident radiance L_i from all directions on the unit sphere S^2:

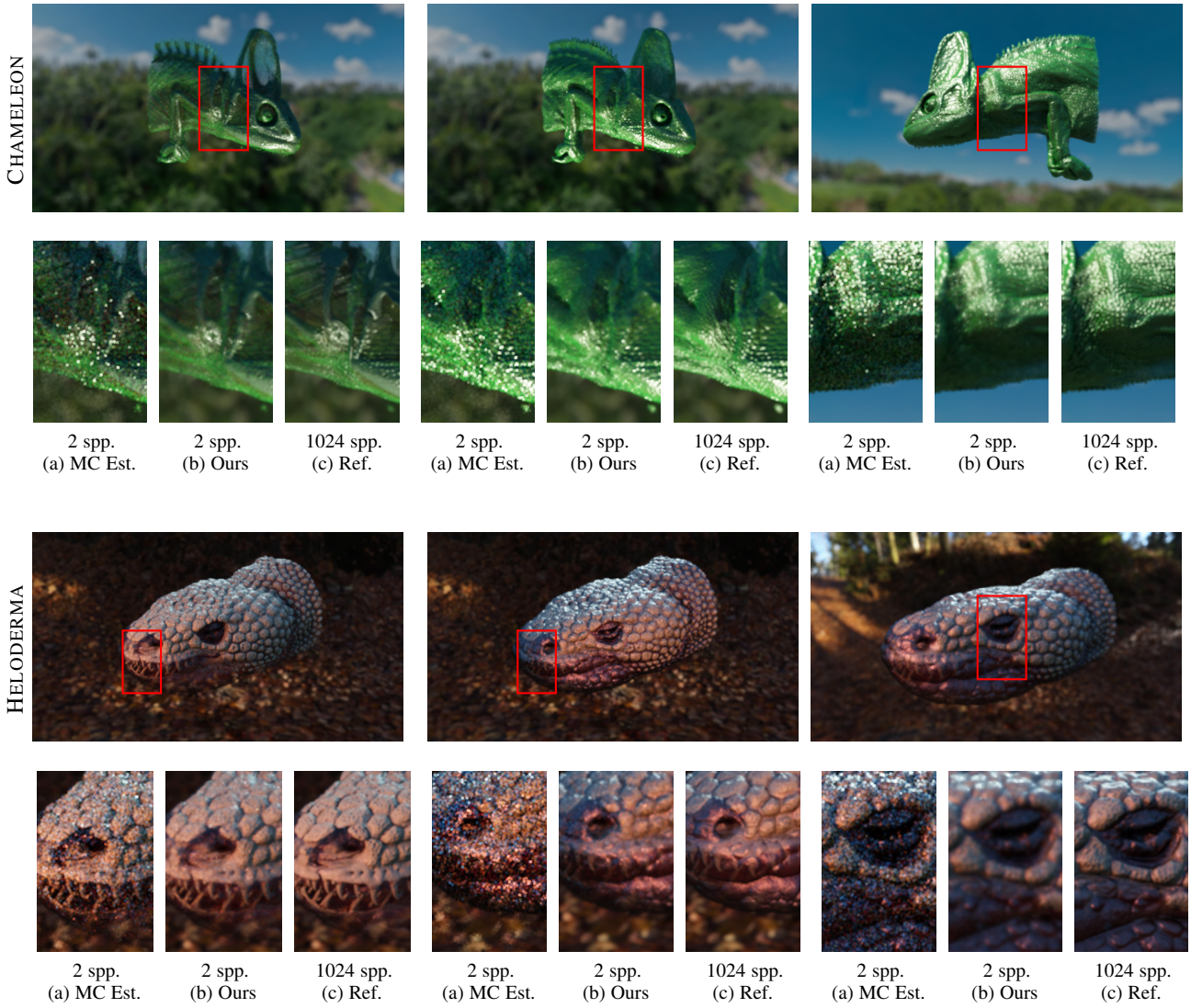
$$L_s(\mathbf{x}, \omega) = \int_{S^2} f_p(\omega, \overline{\omega}) L_i(\mathbf{x}, \overline{\omega}) \, \mathrm{d}\overline{\omega} \tag{5}$$

1.5 The volume rendering integral (VRI).

The formulation of the RTE in Equation 4 corresponds with the differential absorption, in-scattering, out-scattering and emission events affecting the light transport in a differential volume element. Integrating both sides of Equation 4 along direction ω yields the integral form of the RTE:

$$L(\mathbf{x}, \omega) = \int_0^{\infty} T(\mathbf{x}, \mathbf{y}) \left[\mu_a(\mathbf{y}) L_e(\mathbf{y}, \omega) + \mu_s(\mathbf{y}) L_s(\mathbf{y}, \omega) \right] \mathrm{d}y \tag{6}$$

The integration bounds of Equation 6 can be truncated to consider the endpoint $\mathbf{z}$ as the corresponding nearest surface model along the ω direction, where $\mathbf{z} = \mathbf{x} - z\omega$. In the absence of surface models, $L(\mathbf{z}, \omega)$ will simply correspond with the background emitted radiance. This truncated integral version of the RTE is often known as the *volume rendering integral* (VRI):



$$L(\mathbf{x},\omega) = \int_0^z T(\mathbf{x},\mathbf{y})\left[\mu_a(\mathbf{y})L_e(\mathbf{y},\omega) + \mu_s(\mathbf{y})L_s(\mathbf{y},\omega)\right] \mathrm{d}y + \underbrace{T(\mathbf{x},\mathbf{z})L(\mathbf{z},\omega)}_{background} \quad (7)$$

This equation is often taken as starting point in order to develop further efficient approximations for interactive volume rendering scenarios. This is the equation also used in our main paper to develop a progressive volumetric path tracing framework that computes only a small number of light transport trajectories before the denoising stage.

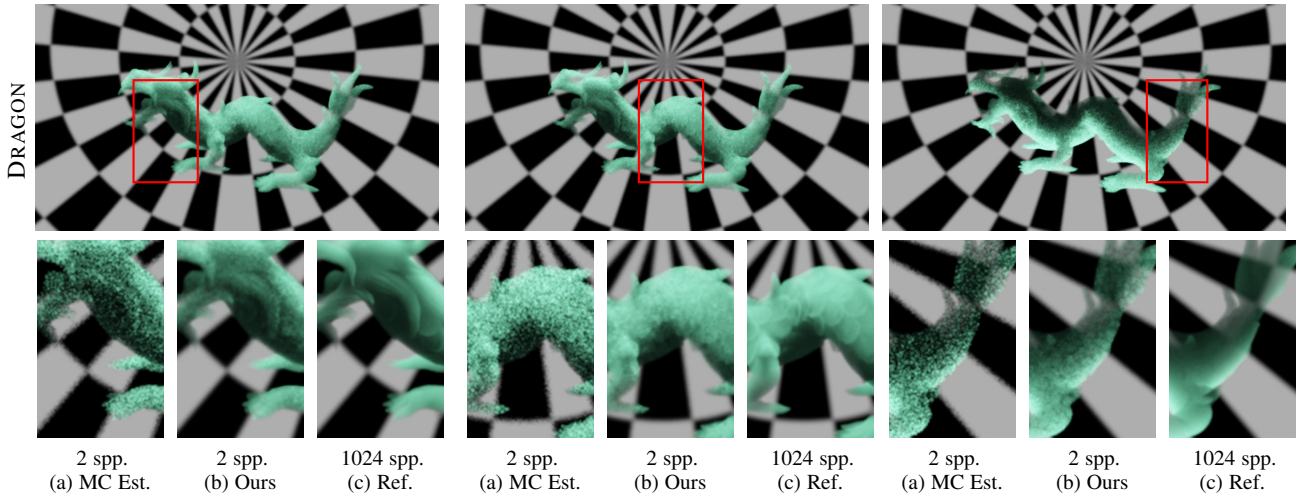
1.6 Emission-absorption models for the VRI.

The solution of the full VRI might be computationally prohibitive for interactive volume rendering. Hence, a common strategy to accelerate its computation is to remove or simplify some of the involved terms. In classic Direct Volume Rendering (DVR) the volume data is often considered to represent a light-emitting medium, resulting in a simplification of the RTE into an emission-absorption model [2]. This simplified model has been widely used for volume rendering because it provides a good compromise between generality and efficiency of computation. The emission-absorption model has been often implemented using ray-casting algorithms and ray marching approaches simulating light transport along straight rays but neglecting scattering and indirect illumination. However, in our paper the goal is to actually consider as many indirect global illumination effects as possible, thus we can not use this classic simplified model .

2 Training Data used for the RAE

Since we compared our method with a deep learning denoiser we provide here more details about how the training data looks like. Generating variations of the original data sets is achieved through scripted variations of the camera point-of-view (POV), the light types and their parameters and finally by modifying the transfer functions in order to allow more and less transparent materials.

It is important to note that training a network can be considered as a preprocessing requirement of the RAE approach. Networks trained for long on a given training set can easily overfit, which means that they will produce very good results for images that are like those used as part of the training set. However, we can interpret this overfitting as an upper bound of the performance of such networks. For example in Figure 2 we show an overly fair



comparison of a network trained on exactly the same data used for testing. Despite our approach does not require any preprocessing (i.e. training), our method remains comparable in terms of visual quality even when tested on the same training data.

In order to deal with temporal sequences, the RAE is trained using temporal clip patches pairs containing the generated noisy input and its corresponding ground truth (i.e. reference patch). We show an example for a training clip in Figure 3.

We report our numerical analysis with the help of a distortion metric (PSNR) and a perceptual distance metric (LPIPS). We report cases using various types of animations (i.e. camera, light and transfer function parameters) on at least three different data sets. We illustrate some of the downside effects of using RAE networks for camera animations (See Figure 4). Moreover, we have also tested how our denoising method performs when using exactly the same data the RAE network saw at training time, something which obviously could benefit the result for the RAE approach.

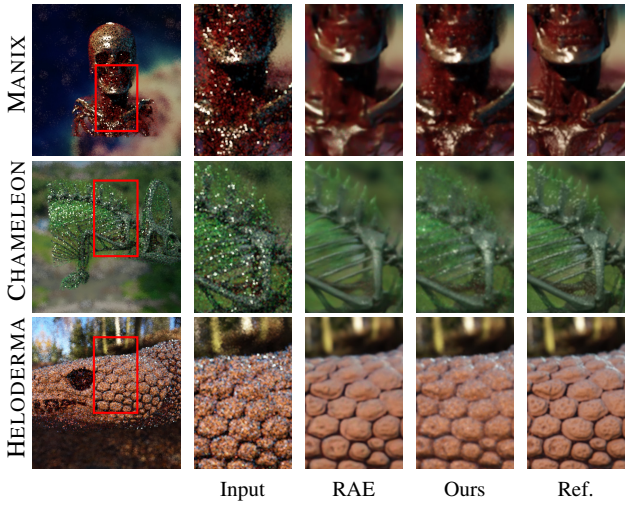


Fig. 2: Results for MANIX, CHAMELEON and HELODERMA data sets on same training data distributions. On the comparison we also show the result produced by our technique and the corresponding reference.

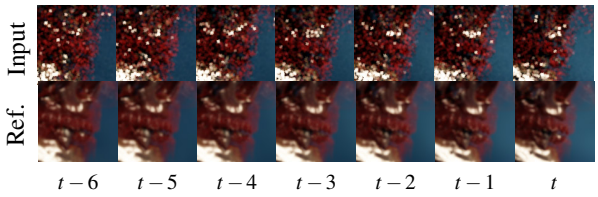


Fig. 3: Training clip example consisting of 128×128 cropped patches at 2 spp and their corresponding reference patches at 1024 spp. Our RAE network implementation used a temporal window of 7 frames as suggested by Chaitanya et al. [1] to learn context information.

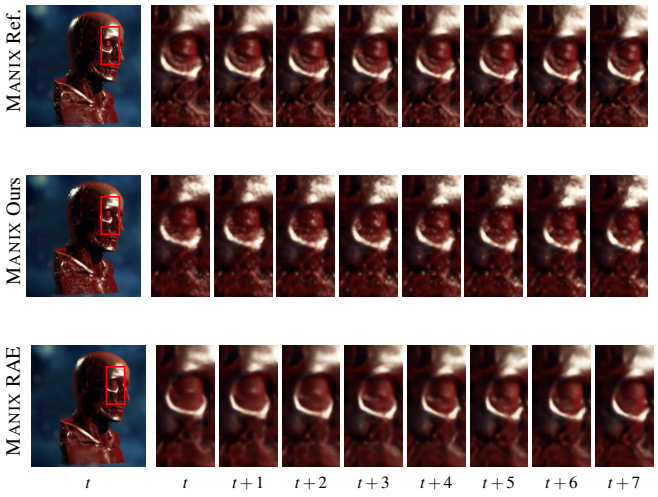
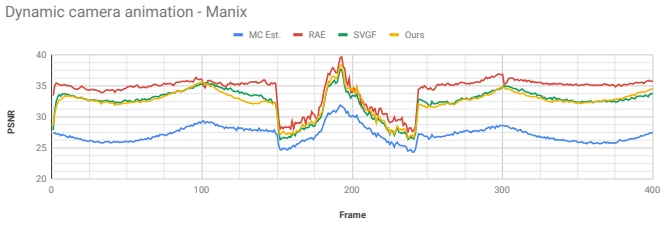
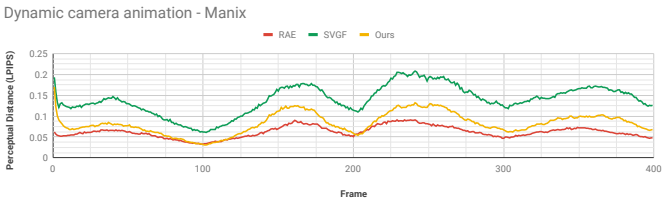


Fig. 4: Example of consecutive frames rendered during camera translation and rotation movements. The selected frames correspond to frames for MANIX sequence starting at frame $t = 62$. Note how for instance how the edge in between both eyelids is clearly shown for all frames for reference and ours, instead it gets washed away for some frames for the RAE. A similar effect also happens with the two thin vessels on the cheek which appear and disappear for some frames due to an overblurring effect.

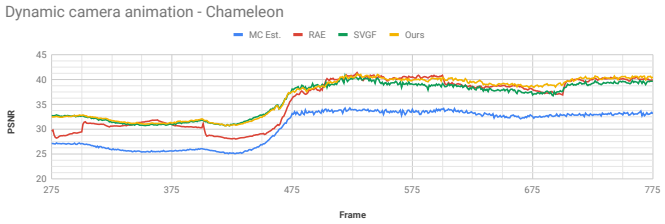
3 Dynamic camera animation

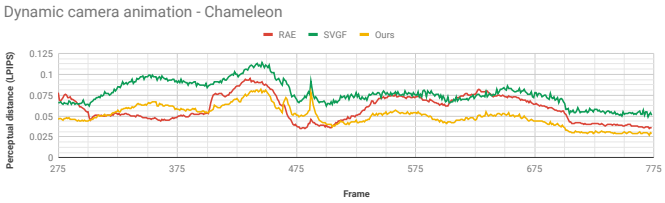
3.1 Manix dataset



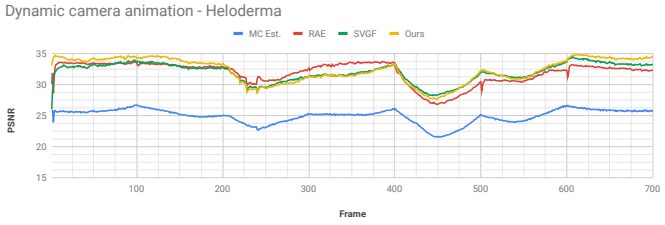


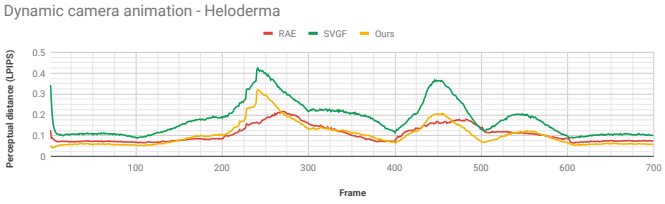
3.2 Chameleon dataset



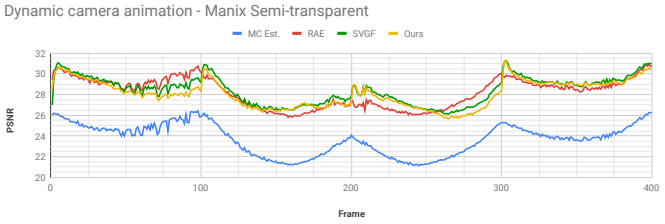


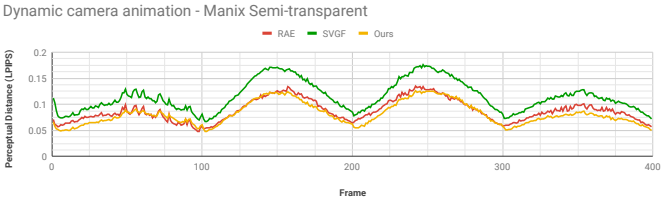
3.3 Heloderma dataset





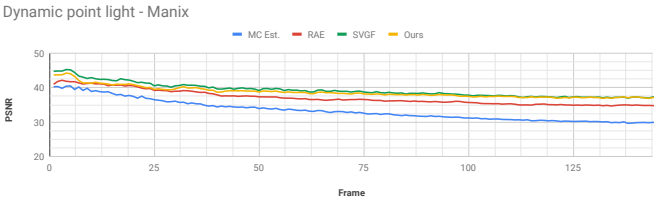
3.4 Manix dataset: semi-transparent isosurfaces.

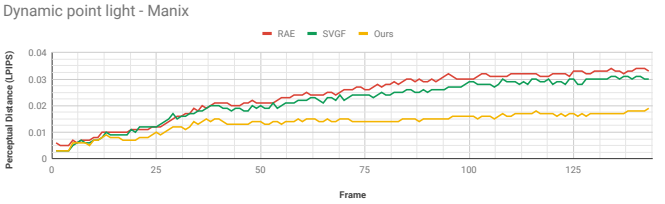




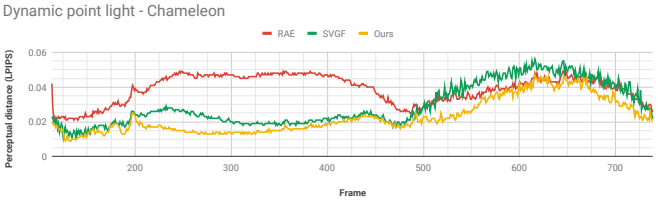
4 Dynamic lighting: point light

4.1 Manix dataset



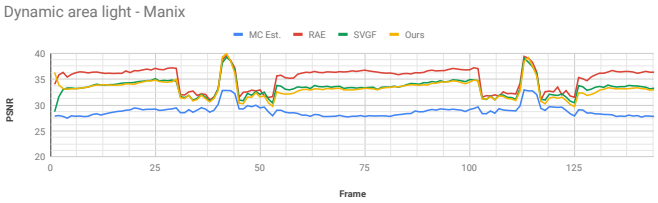


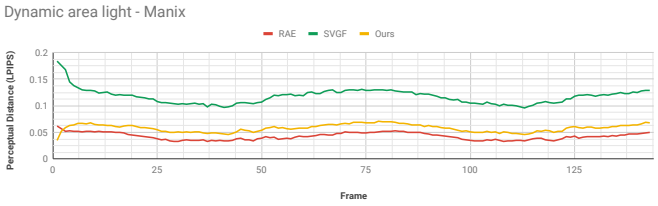
4.2 Chameleon dataset



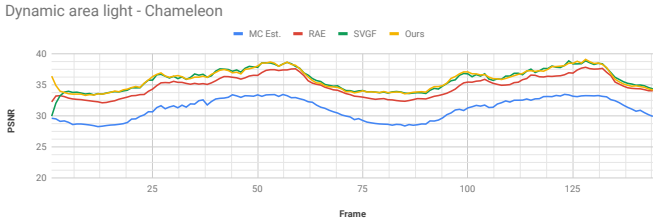
5 Dynamic lighting: area light

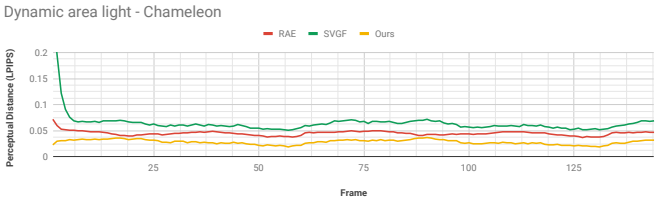
5.1 Manix dataset



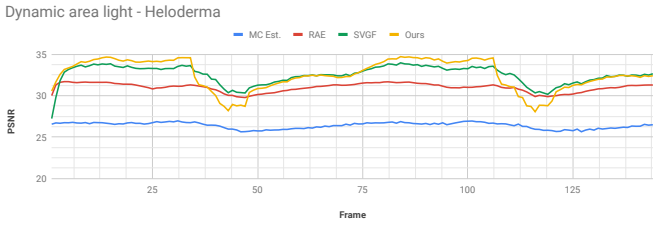


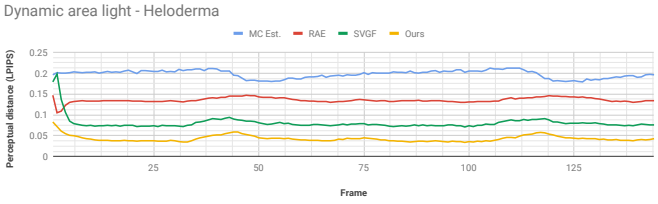
5.2 Chameleon dataset



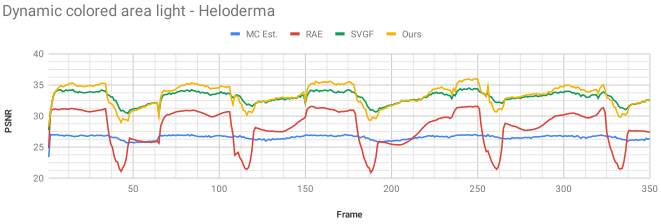


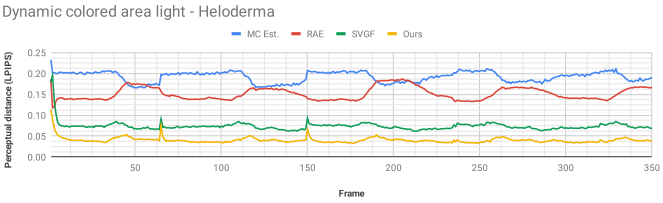
5.3 Heloderma dataset



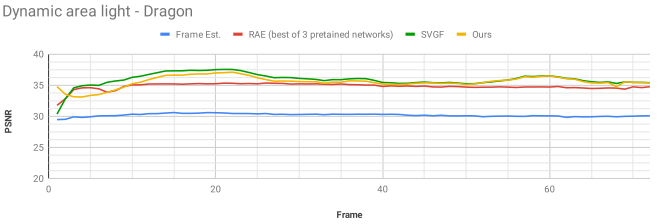


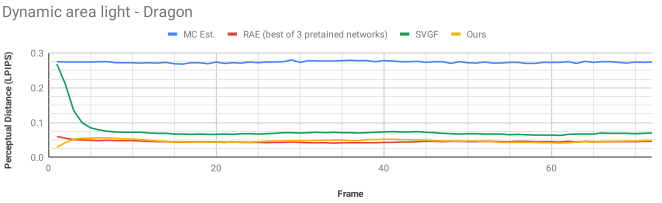
5.4 Heloderma dataset (dynamic light color)





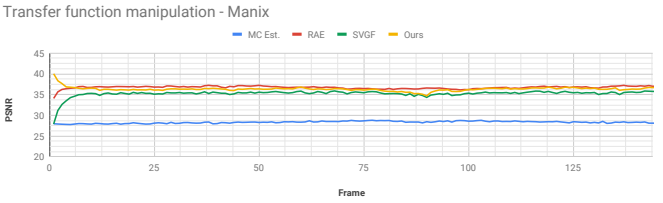
5.5 Dragon dataset

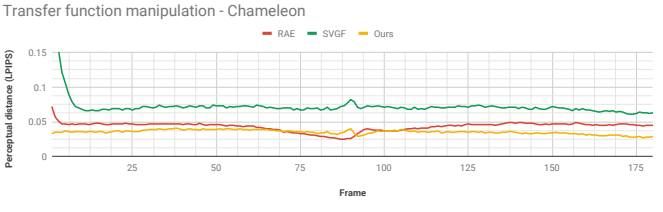




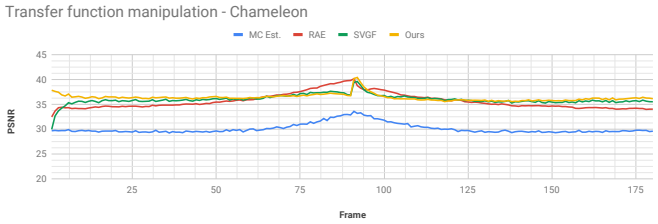
6 Transfer function manipulation

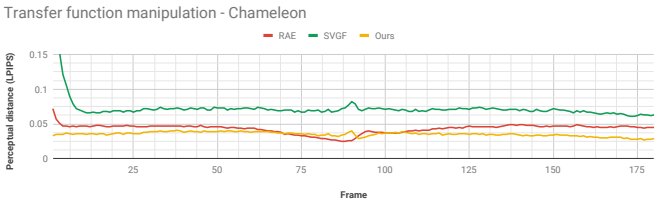
6.1 Manix dataset



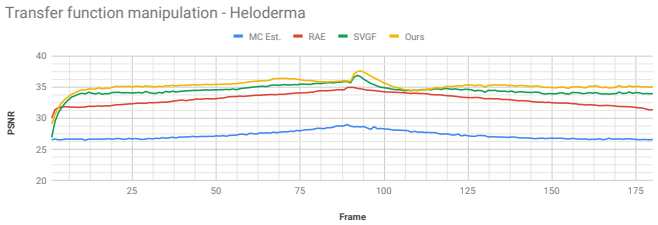


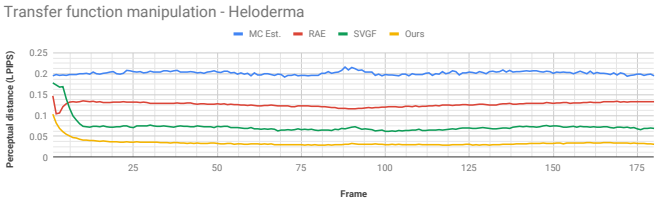
6.2 CHAMELEON dataset





6.3 HELODERMA dataset





REFERENCES

[1] C. R. A. Chaitanya, A. S. Kaplanyan, C. Schied, M. Salvi, A. Lefohn, D. Nowrouzezahrai, and T. Aila. Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder. *ACM Trans. on Graphics (TOG)*, 36(4):98, 2017.

[2] K. Engel, M. Hadwiger, J. Kniss, C. Rezk-Salama, and D. Weiskopf. *Real-time volume graphics*. AK Peters/CRC Press, 2006.

[3] J. Iglesias Guitián, P. Mane, and B. Moon. Real-time denoising of volumetric path tracing for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: 10.1109/TVCG.2020.3037680

[4] T. Kroes, F. H. Post, and C. P. Botha. Exposure render: An interactive photo-realistic volume rendering framework. *PloS one*, 7, 2012.

[5] B. Moon, J. A. Iglesias-Guitian, S.-E. Yoon, and K. Mitchell. Adaptive rendering with linear predictions. *ACM Trans. on Graphics (TOG)*, 34(4):121, 2015.

[6] J. Novák, I. Georgiev, J. Hanika, and W. Jarosz. Monte carlo methods for volumetric light transport simulation. In *Computer Graphics Forum*, vol. 37, pp. 551–576. Wiley Online Library, 2018.

[7] C. Schied, A. Kaplanyan, C. Wyman, A. Patney, C. R. A. Chaitanya, J. Burgess, S. Liu, C. Dachsbacher, A. Lefohn, and M. Salvi. Spatiotemporal variance-guided filtering: real-time reconstruction for path-traced global illumination. In *High Performance Graphics*, p. 2. ACM, 2017.

[8] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.