Supplemental for : Multi-view Relighting using a Geometry-Aware Network

JULIEN PHILIP, Université Côte d'Azur and Inria
MICHAËL GHARBI, Adobe
TINGHUI ZHOU, UC Berkeley
ALEXEI A. EFROS, UC Berkeley
GEORGE DRETTAKIS, Université Côte d'Azur and Inria

CCS Concepts: • **Computing methodologies → Image manipulation**.

ACM Reference Format:
Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A. Efros, and George Drettakis. 2019. Supplemental for : Multi-view Relighting using a Geometry-Aware Network. *ACM Trans. Graph.* 38, 4, Article 78 (July 2019), 2 pages. https://doi.org/10.1145/3306346.3323013

We present additional details of the various steps of our algorithm, in particular on compositing and data augmentation, on the implementation and on the RGB shadow images generation.

1 COMPOSITING AND DATA AUGMENTATION DETAILS

Online compositing. We randomly scale the linear sun images by 2^u with u in $[-0.1, 0.1]$, separately for each color channel, allowing a small shift in sun color, and randomly scale the sky image by 2^u with u in $[-1.0, 1.0]$ (giving a uniform sampling between 0.5 and 2, centered at 1), with the same value for each channel. The latter is a simple approximation for sky turbidity since higher turbidity results in higher sky emission. We then sum sky and sun images.

Data-augmentation. To be robust to different exposures in the input, we apply a random exposure operation by multiplying the image by 2^e where e is in $[-2, 2]$. Similary, to handle white balance differences, we multiply each channel by 2^w with w randomly selected in $[-0.1, 0.1]$, separately for each channel.

We now have two linear images: one for source and one for target lighting condition. We perform the standard gamma correction operation to convert to sRGB with gamma randomly selected in $[2.0, 2.8]$ We clip to 1 for the source image, since all input images will be in the $[0, 1]$ range, but we clip the output to 2, to avoid zeroing the gradients and thus adversly affecting training for values slightly above 1.

Authors' addresses: Julien Philip, Université Côte d'Azur and Inria, julien.philip@inria.fr; Michaël Gharbi, Adobe, mgharbi@adobe.com; Tinghui Zhou, UC Berkeley, tinghuiz@eecs.berkeley.edu; Alexei A. Efros, UC Berkeley, efros@eecs.berkeley.edu; George Drettakis, Université Côte d'Azur and Inria, George.Drettakis@inria.fr.

0730-0301/2019/7-ART78 $15.00
https://doi.org/10.1145/3306346.3323013

2 GENERATING THE RGB SHADOW IMAGES FOR TRAINING

RGB shadow images depend both on the source lighting condition, as we need to sample pixels from the source images, and on the target lighting condition as shadows are cast from the target sun direction. To generate the complete set of RGBD shadow images for training, we would have to render 49×(5+49) images for each viewpoint. Our RGBD shadow images depend on the target sun position for the shape of the shadows, and also on the source images for the color reprojection. These source images are composited sun and sky renderings. Given the number of viewpoints in each scene, the cost of rendering and storing these images would be too high. Instead, we only compute the colors for 5 source sun positions, and thus compute 49×(5+5) images for each viewpoint. During training, we use the closest source sun position rendering stored for the sun layer in the compositing. We can then apply the same image processing transformations on the fly as for the source and target image to the color values in the RGBD shadow images using the same random variables for consistency.

3 INPUT IMAGES NEEDED FOR SYNTHETIC SCENES RECONSTRUCTION

We list the number of input images used for multi-view reconstruction for all our training scenes in Table 1.

Table 1. List of training scenes and their statistics. R_{rc} is the number of images used for reconstruction.

Scene	Old Street	Oriental House	Palm House	Hacienda	House w/ trees
R_{rc}	75	150	75	60	190
Scene	Manor	Arches	Nvidia Street	Nvidia Square	Village
R_{rc}	480	75	1180	1000	179

4 ABLATION: NO SHADOW REFINEMENT, NO RGB SHADOW IMAGES

The main paper shows ablation with *either* the RGB shadow images or the refinement networks disabled. For completeness, Fig. 1 illustrates what happens when disabling both. If we completely remove shadow refinement from our solution, shadow removal is slightly worse, and shadow re-synthesis can be problematic (Fig. 1). For this test we remove the shadow refinement subnetworks and we do not provide the RGB shadow images to the network, instead we provide gray-scale shadow masks.

Fig. 1. Ablation study for shadow refinement. Without the shadow refinement sub-network nor RGB shadow images shadow reconstruction is less effective. Please see supplemental for more examples.

5 IMPLEMENTATION DETAILS

- We do not apply transformations such as tanh or sigmoid at the end of the network allowing to produce images with a higher range than [0,1].
- We export our scenes using a 3DSMax to Mitsuba format exporter that we have developed in the 3DSMax scripting language. This gives flexibility in rendering thanks to mitsuba but we loose a bit of quality of the materials as our exporter only partially handles "baked" materials. The specificities of the V-ray materials in the scenes used poses some issues of compatibility with Mitsuba.
- Generating path traced image for reconstruction would be prohibitively expensive as we require high resolution and little noise. Instead we generate lower resolution ambient occlusion images, upscale and multiply them with full resolution albedo images, which give a coarse approximation to global illumination (see Fig. 2). This approximation is sufficient for reconstruction in most cases. For some scenes, the repetitive nature of the synthetic geometry and lack of texture detail, requires additional processing to allow SfM and MVS to succeed, e.g., by adding calibration targets in very uniform textures.

Fig. 2. Top: From left to right, the rendered ambient occlusion, the albedo , and the multiplication of both. The right most image is used along other point of views to reconstruct the scene and get MVS-like data. Bottom: The obtained reconstruction.

- Some ground truth scenes we used had very repetitive and/or flat textures which led to poor quality or failure in reconstruction when using Sfm + MVS. To allow for a reconstruction with the same level of quality as with real pictures we had to modify some textures in two of our 10 scenes by adding targets on them. These modified textures were only used for reconstruction.
- Reconstructed scenes are not aligned with the original ground truth ones. To align them we used pairs of 3D points of the original ground truth scene and the proxy using the respective original and calibrated cameras. We then used iteratively reweighted least squares on those pairs to compute a transformation from proxy space to original (3DSMax) space.