Pacific Graphics 2020
E. Eisemann, A. Jacobson, and F.-L Zhang
(Guest Editors)

Volume 39 (2020), Number 7

Deep Separation of Direct and Global Components from a Single Photograph under Structured Lighting – Supplemental

Z. Duan J. Bieron P. Peers

College of William & Mary

1. Setup Calibration Details

This section describes the geometric and radiometric calibration of the co-axial projector-camera setup used for deep separation of direct and global components.

Camera-Projector Alignment To calibrate the relative positioning of the camera-projector pair, we project a grid pattern on a non-planar scene with depth discontinuities and adjust the x and y coordinates of the camera until the grid-lines in the image are continuous. We adjust the camera's z-coordinate until a projected checkerboard pattern produces the same image when projected on two planar surfaces at different depths.

Pixel Alignment The basic form of our network assumes that the lighting **L** and captured image **I** are expressed "at the resolution and from the view of the projector". In other words, we would like to resample the captured image such that each pixel corresponds to a projector pixel. Instead of directly calibrating the intrinsic properties of the camera and projector (including lens distortion), we again follow a data-driven approach. We project a 2×2 checkerboard pattern and detect the corners. This allows us to interpolate for each camera pixel, the corresponding projector pixel. We resample the camera image by computing the integral over the projected projector pixels in the camera image, taking in account that a camera pixel might only partially overlap with the projected projector pixel.

Radiometric Calibration Furthermore, we radiometrically calibrate the projector by capturing 256 photographs of a spectralon sample lit by a full-on projector pattern for each possible intensity value. We then normalize and tabulate the average observed radiance, and inversely apply it to each projector pattern before projection. We do not perform explicit radiometrically calibration of the camera, but directly use radiometrically linear RAW camera images.

2. Uniform Lighting

In the paper we compare the results from our method with the deep learning method of Nie et al. [NGS*18]. For each method we used their ideal lighting (i.e., uniform white for Nie et al. and high frequency lighting for ours). For completeness we also trained our network with uniform white lighting. Unsurprisingly, our method does not fare well (Figure 1).

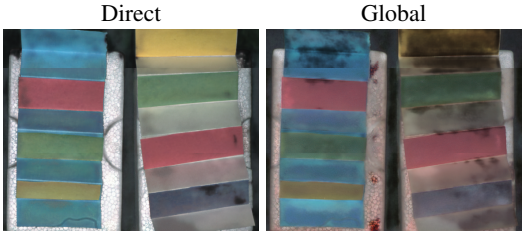

Figure 1: *Our deep separation network fails to produce plausible decompositions under full-on white illumination, even if trained on full-on white lit training data.*

The key reason is that both networks solve essentially a different problem. Nie et al. treats the decomposition problem as an image-translation problem. Consequently, they require a large training set of exemplars to inform the network on the relation between the observed photograph, and the resulting decomposition. In contrast, our method learns the relation of the structured lighting response, and learns to decompose small patches and reconstructs and removes the pattern introduced by the lighting. This is more akin (but not identical) to a super-resolution problem. Consequently, we can train our network with a much smaller training set because we only learn local relations, instead of whole-image relations as in Nie et al. An interesting avenue for future research is to investigate whether our method would perform better under uniform white lighting if an as large training set is used as that of Nie et al.

3. Loss Function Comparison

Figure 2 compares the L_1 and L_2 loss on RGB as well as on LAB. We can see that both losses on RGB exhibits artifacts at the top edge of the middle planter as well as in the highlights. Furthermore, the L_1 LAB decomposition exhibits artifacts in the global decomposition of the orange ball. In contrast the L_2 LAB loss produces the cleanest results with the least artifacts.

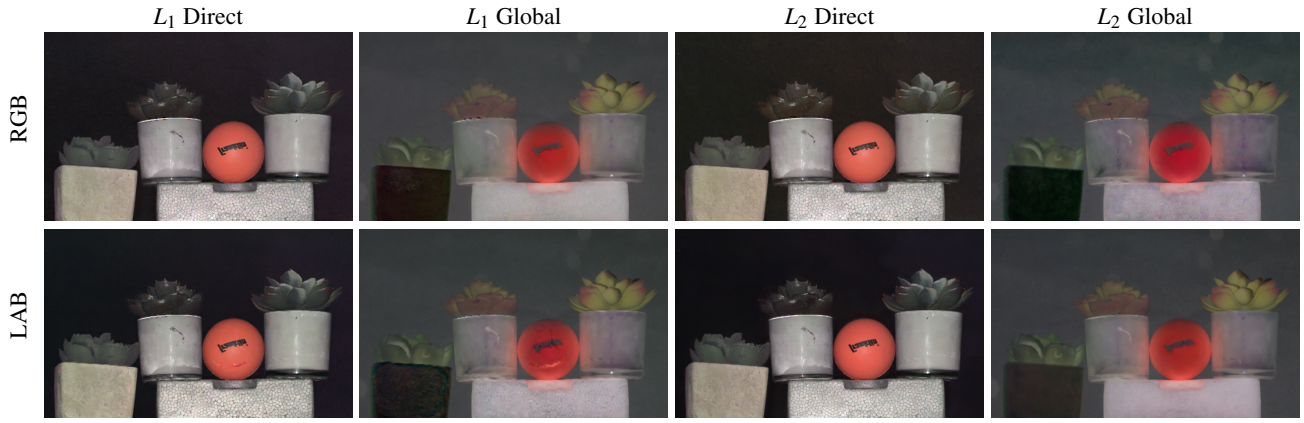

Figure 2: *Comparisons of different metrics and color spaces for the training loss. The RGB color space losses result in artifacts at the top-edge of the middle planter. The L_1 LAB loss results in an incorrect decomposition of the orange ball.*

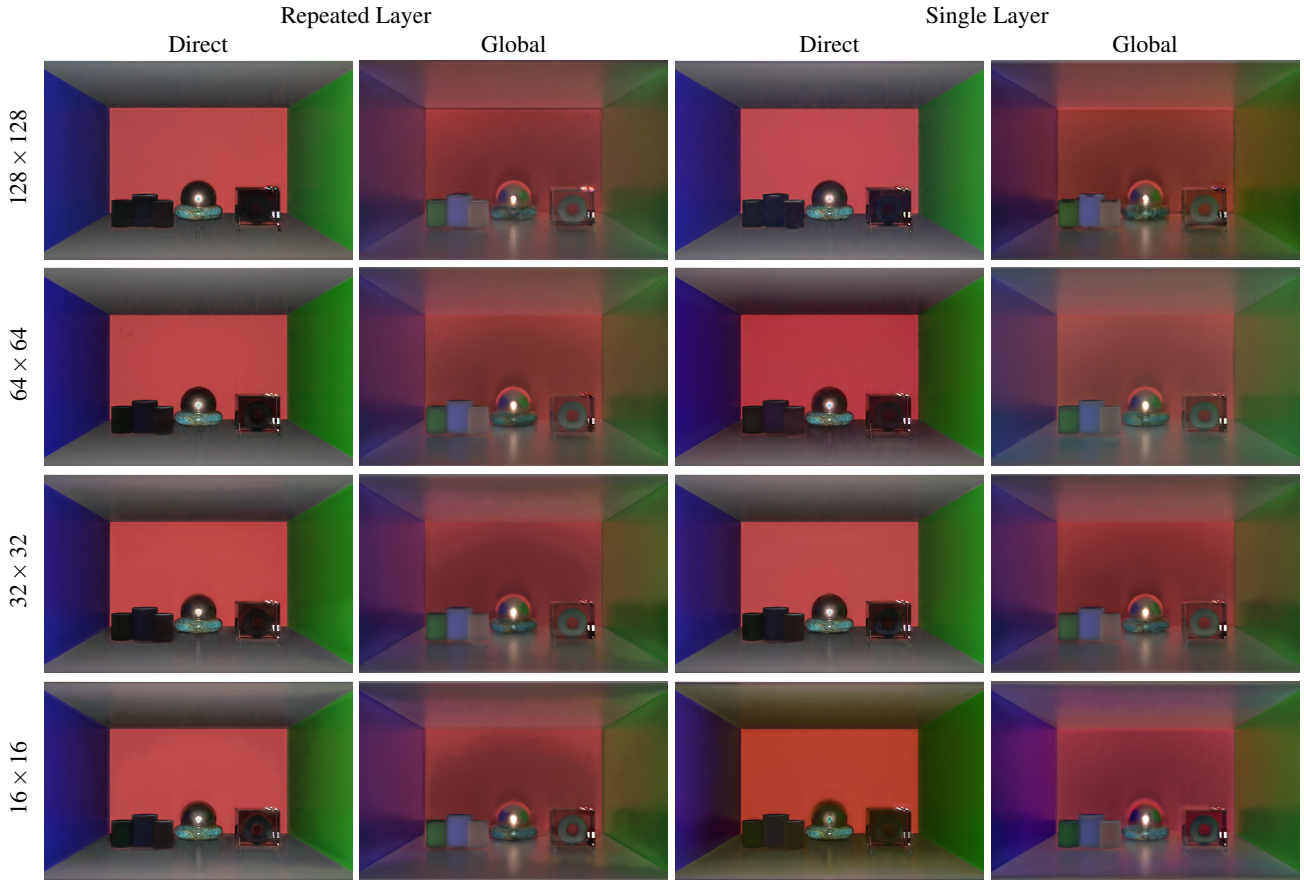

Figure 3: *A comparison between decompositions obtained from networks trained with different training patch sizes ranging from* 128×128 *to* 16×16*, and with or without doubling the number of layers at each resolution of the network.*

4. Training Patch Size

In Figure 3 we compare the decomposition quality on the *Cornell Box* scene for different networks trained with different patch sizes ranging from 128×128 to 16×16 and with or without doubling the number of layers at each resolution of the network. For each network we keep the bottleneck size and essentially remove the outer layers. We refer to the paper for a quantitative comparison.

Visually we can see that the decompositions for 128×128 and 64×64 are of similar quality. The most obvious difference lies in the separation of the diffuse shape inside the glass box. Doubling of the layers produces slightly better results. At 32×32, the decomposition is still reasonable, but some artifacts can be seen (e.g., the reflection on the ceiling). At 16×16 we can see quantization artifacts in the shadow on the back wall, and without doubling color artifacts appear in the indirect component. From this we conclude, that doubling the layers per resolution helps to obtain better results, and that a patch of 64×64 or larger is needed for high quality results.

References

[NGS*18] NIE S., GU L., SUBPA-ASA A., KACHER I., NISHINO K., SATO I.: A data-driven approach for direct and global component separation from a single image. In *ACCV* (2018), pp. 133–148. 1