Joint Material and Illumination Estimation from Photo Sets in the Wild

Tuanfeng Y. Wang    Tobias Ritschel    Niloy J. Mitra
University College London

1. Supplementary Material

Figure 1. Here, we show the effect of Gauss-sphere coverage: Even for non-round objects with flat areas that have a bad coverage of the Gauss sphere the reconstruction (left) is similar to the reference (right).

The last row shows a pair of novel views, under input illumination where the first image has unobserved, the second an observed material.

Applications. A typical application of our approach is photo-realistic manipulation of objects in Internet images as shown in Figure 2. Having estimated the material and illumination parameters from all the images, we can insert virtual replica into the image (Figure 2b, 2d), transfer reflectance estimated from other Internet images to new scenes (Figure 2b and Figure 2f), or introduce new object with material under the estimated illumination. Please note that the estimated environment maps were used to render object shadows on (manually added) ground planes (Figure 2b, 2c, and 2d).

Examples. In Figure 3, Figure 4, and Figure 5 we present more results using our framework on different datasets.

Prediction. Using the INTERNET-EAMES data sets, we are able to test the predictive ability of our approach by leaving out one image form the set and compare it to the acquired ground truth. This is seen in Figure 7 where starting from the first image, we predict the red chair in the second image and the yellow chair in the third image.

Progressive estimation. A key property of our approach is to consolidate information from multiple images to disambiguate material and illumination. This characteristic implies that adding more images to the photo set should reduce the error. Figure 8 confirms the rise in performance as more images are added to the linked set.

Matrix size. In Figure 9a, we show the effect of increasing matrix size on the error of predicting the entries for the SYNTHETIC data set. Here, the matrix $O$ is complete, i.e.,
all material-illumination pairs are observed. We see, that with increasing size, the estimation for all entries gets more correct while the task being solved in some sense is also bigger (more different illuminations). Note that the total error residue can go up, but the estimation gets more accurate (compared to the ground truth).

When the matrix is reduced to a single row or column (Figure 9b) our approach can still estimate illumination and material. For a $1 \times 5$ matrix, which estimates a single material from multiple illuminations, the approach does well; but slightly degrades for a $5 \times 1$ setting, where multiple objects are seen under the same illumination.

**Label quality.** We assume the input images to have per-pixel normal and material labels. In Figure 9d, we study the effect of incorrect normal estimates by adding a noise
Figure 5. Results on INTERNET-EAMES dataset of six images of a celebrated Eames chair with four materials in the protocol in LAPD-result in the main paper.

Figure 6. Prior effect: The top three rows show a $3 \times 3$ input. The next row shows illumination. Following the GT left, we increase the prior weight to the right. We note that illumination chromaticity decreases with increasing weight and that a good trade-off is likely at 0.1.

Figure 7. Estimating materials and illumination using all chairs except the rendered one. Left: reference image; Middle: the red chair is rendered; Right: the yellow chair is rendered.

Figure 8. Progression of quality from left to right. Every row shows, for a selected material what the additional images can add to the quality in terms of re-rendering, material, and illumination.

Figure 9. Effect of different input properties (horizontal) on the quality of our approach in terms of the DSSIM error (vertical, less is better): matrix size, structure, and alignment error.

to the normal. Such noise is due to alignment errors and errors in the normal estimation in practice. We see that good normal fair better with the error in the order of one percent, while larger errors produce an error that saturates still at a low total value. Also, note that for the INTERNET and the PHOTOS data sets, the geometry models are coarse and/or noisy. But in absence of ground truth,