

Supplementary Material for “3D Object Manipulation in a Single Photograph using Stock 3D Models”

Natasha Kholgade¹

Tomas Simon¹

Alexei Efros²

Yaser Sheikh¹

¹Carnegie Mellon University

²University of California, Berkeley



Figure 1: (a) Stock 3D models used in this paper (b) Alignments provided by four users for different models.

1 Stock 3D Models

Figure 1 shows the 3D models used in this paper. All models were obtained from TurboSquid, with the exception of the top hat which was obtained from 3D Warehouse.

2 Focal Length Extraction

To extract focal length using vanishing points, we assume that the photograph contains rectangular structures (e.g., floor tiles, rugs, carpet striations). The user marks two sets of parallel lines, the one set being perpendicular to the other. We compute the vanishing points of these two lines, and use them to extract the focal length through the absolute conic as described by Hartley and Zissermann [2004].

3 Ground Plane Estimation

We assume that objects in the photograph are at rest on a ground plane π_g . We estimate π_g by one of two standard methods, depending upon the nature of the photograph and the object model. For object models with good contact support, the user marks three points $\bar{\mathbf{X}}_0$, $\bar{\mathbf{X}}_1$, and $\bar{\mathbf{X}}_2$ on the base in counter-clockwise order (as observed from the bottom of the model), and we compute the normal of π_g as $\mathbf{n}_g = \mathbf{R}^T \left(\frac{(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0) \times (\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_0)}{\|(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0) \times (\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_0)\|^2} - \mathbf{t} \right)$. For photographs which have rectangular structures on the ground plane, we use the method described in Section 1 to estimate vanishing points, and we use these points to obtain the horizon, \mathbf{v}_l . We calculate the ground plane normal as the line perpendicular to the horizon, $\mathbf{n}_g = \frac{\mathbf{K}^{-T} \mathbf{v}_l}{\|\mathbf{K}^{-T} \mathbf{v}_l\|^2}$. In either case, we compute the distance d_g of the plane from the origin as $d_g = \min_i \mathbf{n}_g^T (\mathbf{R} \bar{\mathbf{X}}_i + \mathbf{t})$, $i \in \{1, \dots, N\}$. We back-project the 2D points in the ground mask provided by the user on to π_g to obtain the corresponding ground points in 3D.

4 User Alignment of Geometry

Figure 2 shows alignments performed by four users for various models used in this paper. Alignment timings are included in the paper.

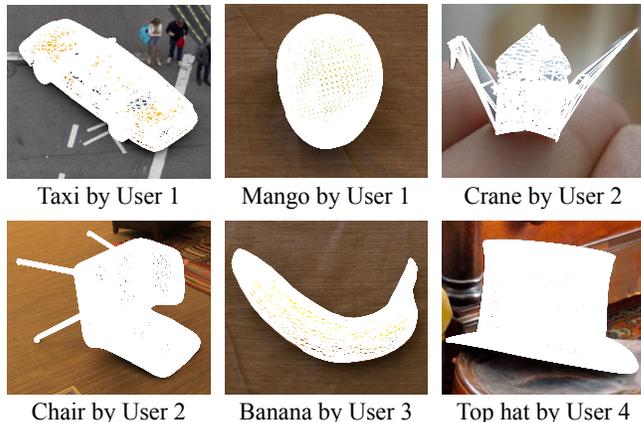


Figure 2: Alignments provided by four users for different models.

5 User Study Images

Figure 3 shows the images shown to users in our forced choice user study. For each object, we created an original-photograph/final-result pair (“Condition 1” in the paper), i.e. a pair with the left and right columns of Figure 3, and an original-photograph/intermediate-result pair (“Condition 2” in the paper), i.e., a pair with the left and center columns of Figure 3. Each subject was shown one of the two pairs per object, and not both. For the chair, cab, and crane, we produced the results in the center and right columns of Figure 3 using the photographs in the left column. For the mango in the last row, we used an alternate photograph to create the results in the center and right. We randomized the assignment of image pair to each subject.

6 Geometry Evaluation

Figure 4 shows the results of alignments using our geometry correction approach compared to Xu et al. [2011] for a variety of photographs used in this paper. As the results demonstrate, Xu et al. approximate the form of the objects well, but do not achieve exact alignments. This happens either because the original rigid alignment is incorrect (as for the taxi or the mango), or because their deformation model does not capture significant deformations as for the crane.

7 Illumination Evaluation

Figure 5 shows the spherical environment maps generated using a light probe, projecting the background of the photograph onto the environment map as in Khan et al. [2006] (Background), and using the approaches of Haber et al. [2009] (Haar basis), Mei et al. [2009] (Spherical harmonics with L_1 prior), and our approach (von Mises-Fisher or vMF basis) for increasing numbers of basis components K . Figures 6 through 9 show ground truth photographs used in our illumination experiments (Photographs 1 and 2 are shown in the paper), outputs synthesized using the light probe, and outputs

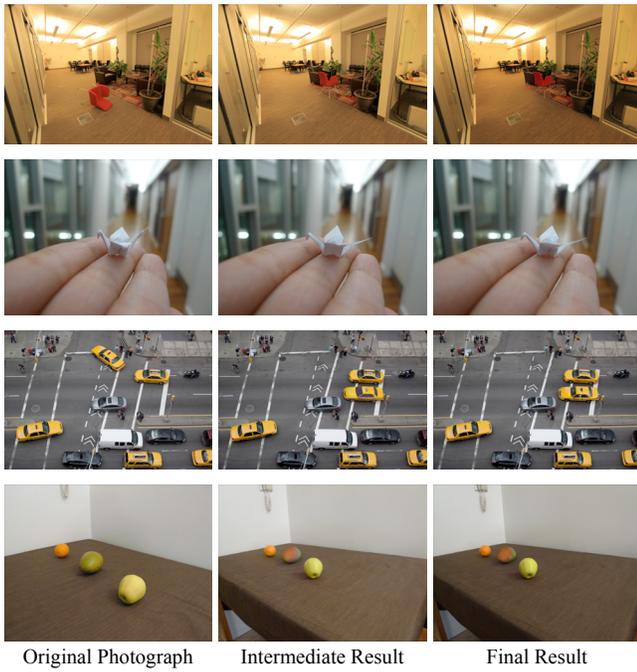


Figure 3: Images shown in our user study.

generated using the background projected as an environment map, and the three illumination estimation approaches for $K = 256$ and $K = 4096$. The environment maps for the three methods were computed using Photograph 1. One issue is that our geometry correction approach currently allows alignment to a single viewpoint, and does not simultaneously include multiple viewpoints and ground contact constraints. As such the ground plane computed using Photograph 1 may not be accurate for the rest of the photographs, and the chair may appear to float in some of the synthesized images. Our approach generates non-negative illumination and soft cast shadows, and represents light sources that are not visible in the photograph.

References

- HABER, T., FUCHS, C., BEKAERT, P., SEIDEL, H.-P., GOESELE, M., AND LENSCH, H. P. A. 2009. Relighting objects from image collections. In *CVPR*, IEEE, 627–634.
- HARTLEY, R. I., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518.
- KHAN, E. A., REINHARD, E., FLEMING, R. W., AND BÜLTHOFF, H. H. 2006. Image-based material editing. In *Proc. ACM SIGGRAPH*, 654–663.
- MEI, X., LING, H., AND JACOBS, D. 2009. Sparse representation of cast shadows via l_1 -regularized least squares. In *ICCV*.
- XU, K., ZHENG, H., ZHANG, H., COHEN-OR, D., LIU, L., AND XIONG, Y. 2011. Photo-inspired model-driven 3d object modeling. *ACM Transactions on Graphics* 30, 4.

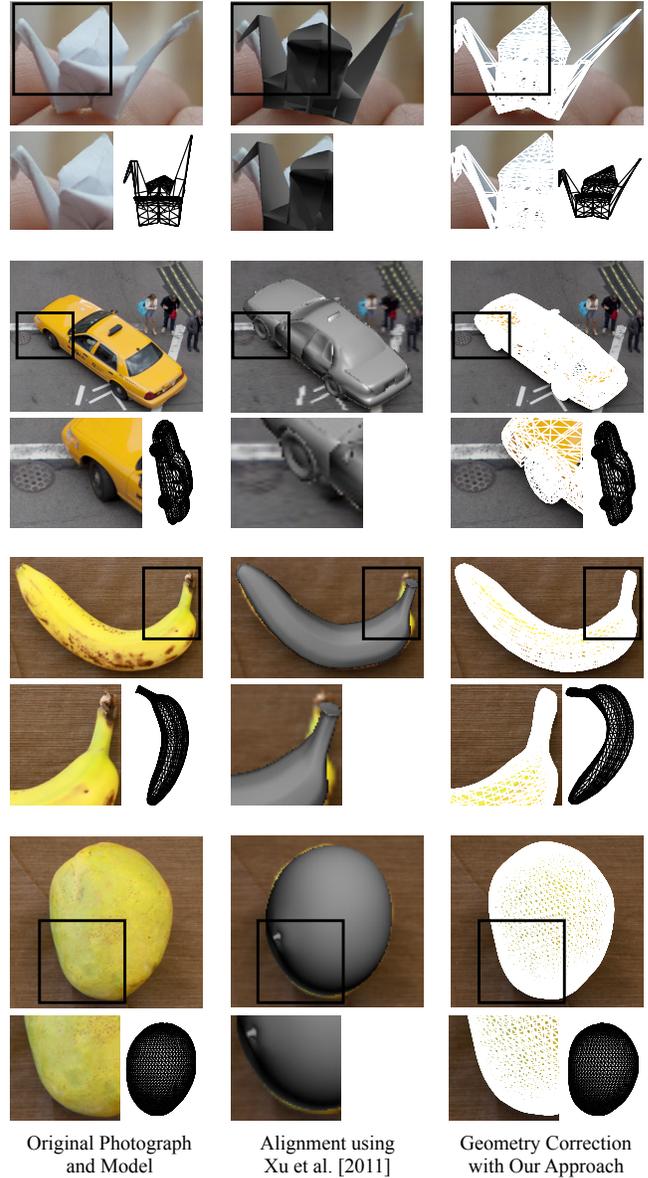


Figure 4: Correction of the stock model geometry using our approach (right) compared to the approach of Xu et al. [2011]. As shown by the insets, the approach Xu et al. achieves close but not perfect alignment, either due to incorrect rigid alignment, or due to failure of the deformation model to capture large deformations. Through our user-guided geometry correction approach, users can provide accurate alignments of the geometry to the photograph.

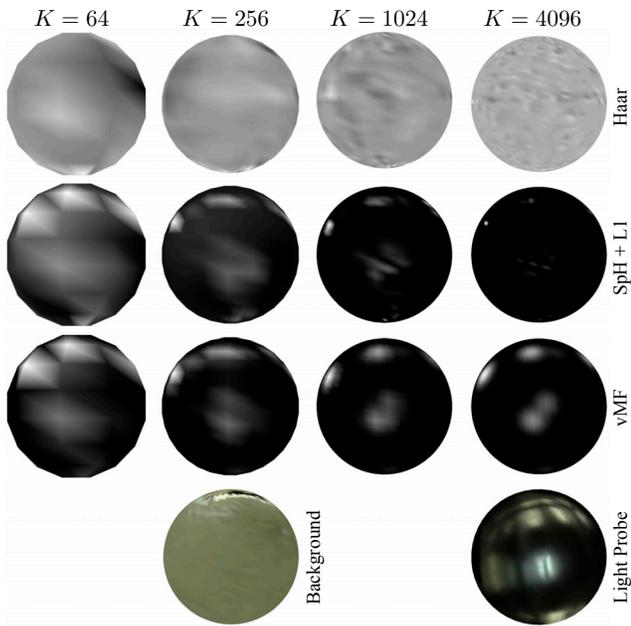


Figure 5: Environment maps obtained using Haber et al. [2009] (Haar), Mei et al. [2009] (SpH+L1) and our approach (vMF) for various values of K . The lowest rows show the background image projected out as in Khan et al. [2006] and the light probe for comparison. As the number of components increases, the environment map using vMFs approximates the light sources in the scene underlying the photograph.

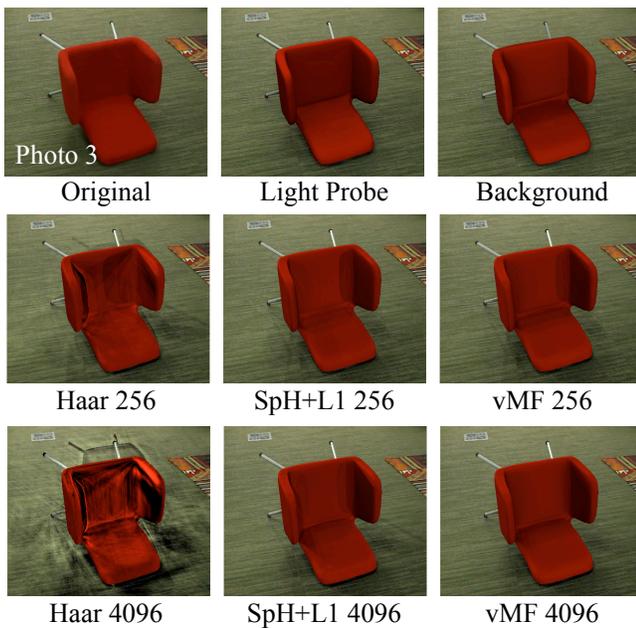


Figure 6: Original photograph 3, and images synthesized using a light probe, and using Haber et al. [2009] (Haar), Mei et al. [2009] (SpH+L1) and our approach (vMF) for $K = 256$ and $K = 4096$.

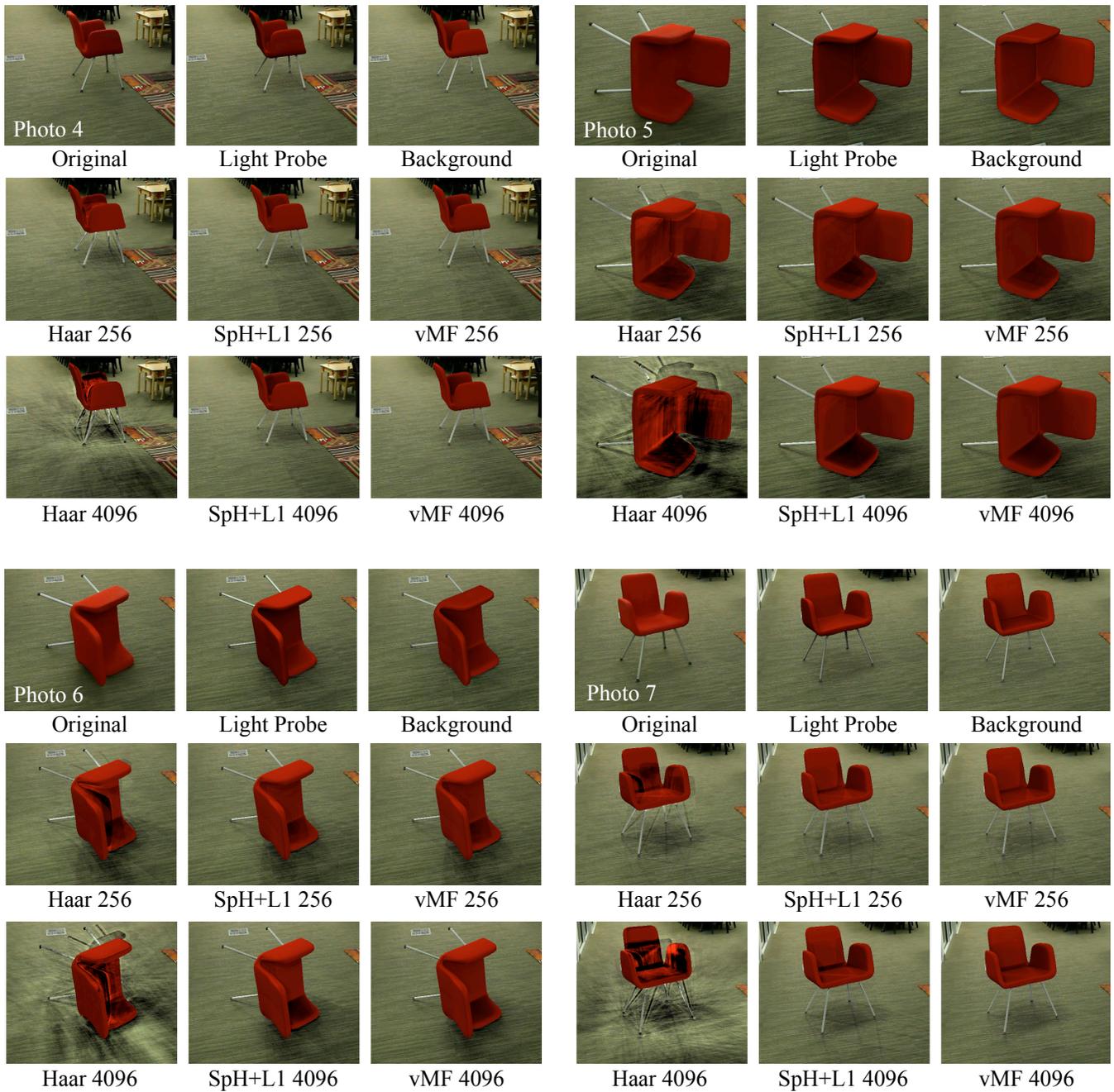


Figure 7: Original photographs 4 to 7, and images synthesized using a light probe, and using Haber et al. [2009] (Haar), Mei et al. [2009] (SpH+L1) and our approach (vMF) for $K = 256$ and $K = 4096$.

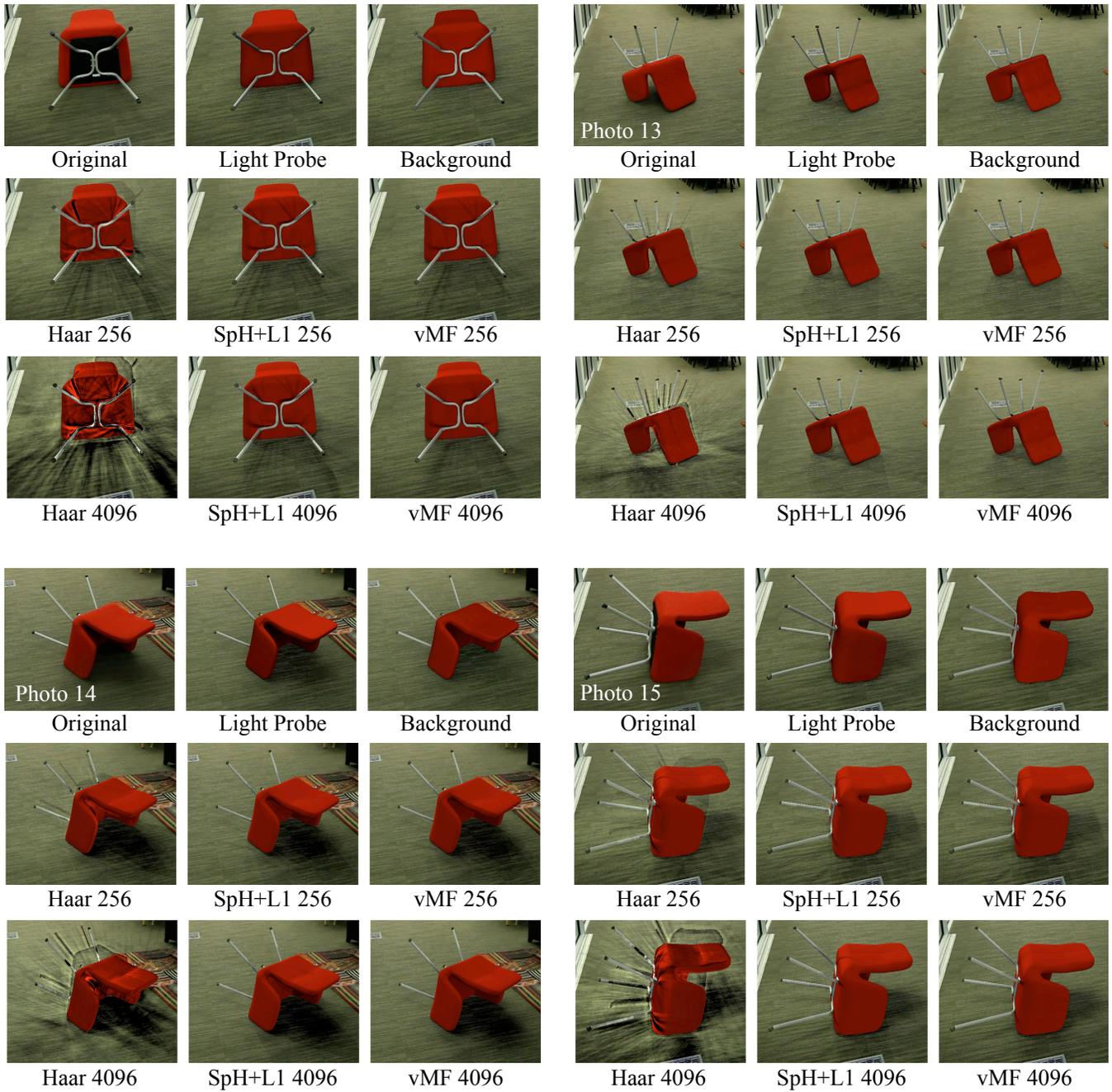


Figure 8: Original photographs 8 to 11, and images synthesized using a light probe, and using Haber et al. [2009] (Haar), Mei et al. [2009] (SpH+L1) and our approach (vMF) for $K = 256$ and $K = 4096$.

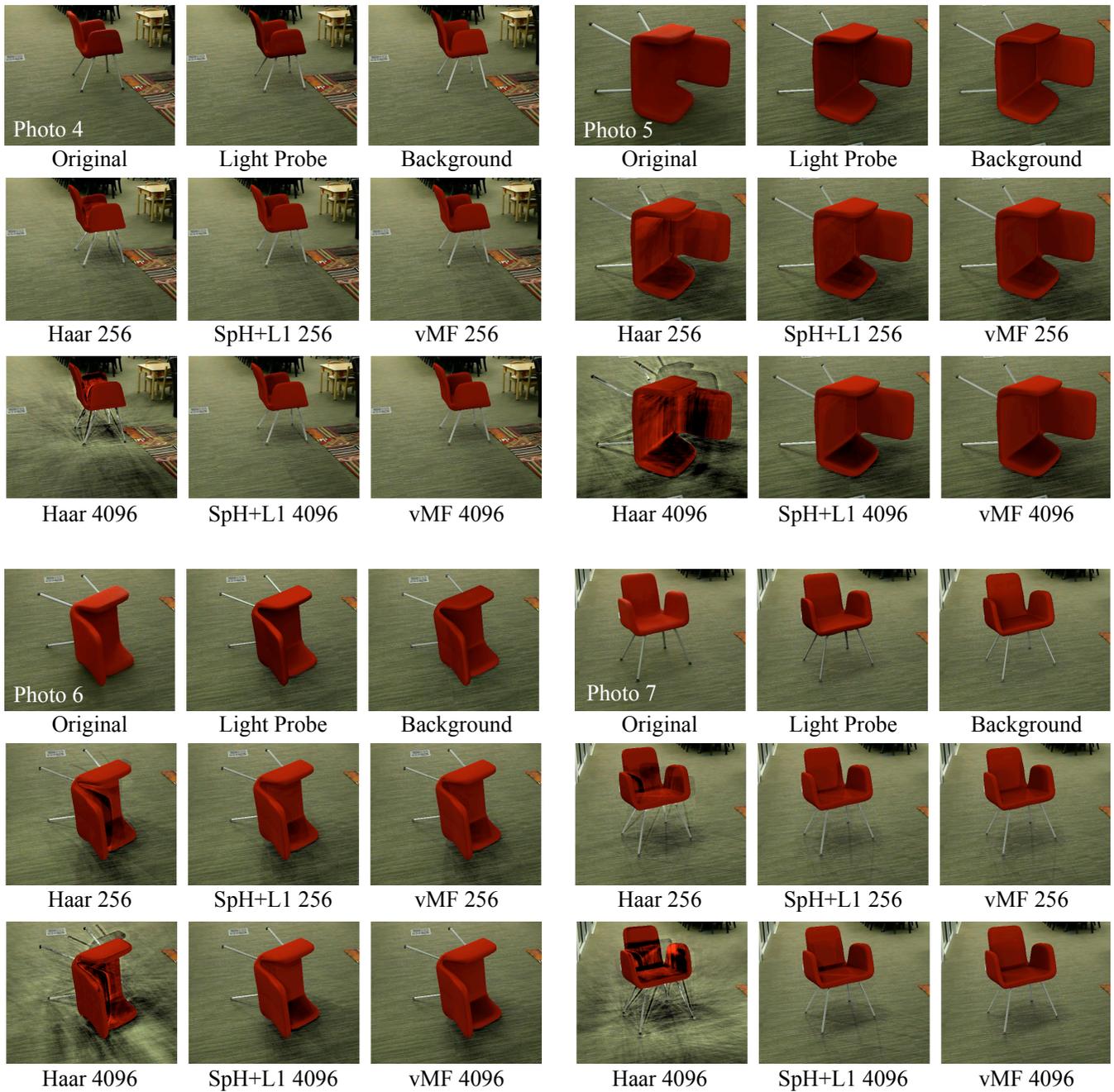


Figure 9: Original photographs 12 to 15, and images synthesized using a light probe, and using Haber et al. [2009] (Haar), Mei et al. [2009] (SpH+L1) and our approach (vMF) for $K = 256$ and $K = 4096$.