Discriminative Shape from Shading in Uncalibrated Illumination

Stephan R. Richter, Stefan Roth
Department of Computer Science, TU Darmstadt

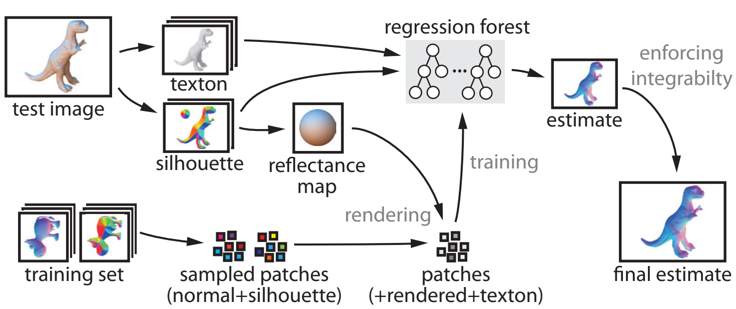

Figure 1: *(left)* Pipeline for training and testing. For each test image, we estimate a reflectance map to train the regression forests on synthetically generated data. We optionally enforce integrability on their pixel-wise predictions. *(right)* Qualitative results for shape and reflectance estimation from a single internet image: input image [9], estimated normals and reflectance map, and novel view (from left to right).

Estimating surface normals from just a single image is a severely under-constrained problem. Previous work has thus made a number of simplifying assumptions, *e.g.*, presuming smooth surfaces, uniform albedo, and directional lighting from a single light source of known direction. Yet, relying on a single directional light, or assuming a given reflectance map renders shape from shading impractical, however. To go beyond the lab, some of these assumptions need to be relaxed. We thus investigate estimating the reflectance map of a diffuse object with uniform albedo together with its surface under uncontrolled illumination, given only a single image (Fig. 1, right). Moreover, we aim to avoid strong spatial regularization to recover fine surface detail. To address this challenging setting, shading cues need to be generalized to more realistic lighting and also combined due to their complementary strengths. This increases the model and computational complexity, as well as the number of parameters.

We approach these challenges with a *discriminative learning approach* to shape from shading. This notably makes it easy to combine several shading cues. We begin by considering *(1) color*, which helps under hued illumination [4], and can be exploited with a second order approximation to Lambertian shading [7]. While powerful, our experimental investigation shows that correlated color channels (*e.g.*, in near white light) or the presence of noise severely impair accuracy. We thus add *(2) local context*, which aids disambiguation [10], but until now has been limited to directional lighting. Instead of using colors from a neighborhood directly, we choose a Texton filter bank [8] for capturing local context. Employing these filters in a learning framework allows for automatic adaptation to uncontrolled lighting and fine surface detail. We finally exploit *(3) silhouette features*. Previous work has constrained surface normals at the occluding contour [3, 6], and propagated this information to the surface interior during global reasoning. We generalize the occluding contour constraint to the surface interior and provide (spatial) contour information at every pixel. These silhouette features yield a coarse estimate of the surface from just the silhouette, which we additionally exploit for estimating the reflectance map.

Adopting a learning approach to uncalibrated shape from shading poses several challenges: First, example-based approaches require a database of surfaces imaged under the same conditions as the object in question. Capturing all possible combinations of surfaces and lighting conditions is next to impossible, and placing known objects in the scene [2] often impractical. Recent learning approaches [1, 5] thus created a database on-the-fly by rendering synthetic shapes under a known lighting condition. We follow this avenue, but capture the variation of realistic surfaces with a significantly larger database of 3D models created by artists. Second and unlike [1, 5], we aim to cope with unknown illumination at test time. We address this by estimating the reflectance map from silhouette features, and only then training the discriminative approach on the estimated reflectance. Third, (re-)training a model for a lighting condition at test time requires efficient learning and inference. Adapting regression forests to store von Mises-Fisher distributions in the leaves, as well as leveraging the set of cues discussed above enables us to perform efficient pixel-independent surface normal prediction. To further refine the estimate, we optionally enforce integrability on the predicted surface. Fig. 1 depicts the entire pipeline.

To assess the contribution of the different cues, we first evaluate their importance statistically across a range of scenarios. Then, we measure their effect as well as that of the integrability approach on the entire pipeline using the synthetic MIT intrinsic image dataset, which also provides a comparison to the state-of-the-art. Finally, to assess the suitability for realistic settings, we capture a novel real-world dataset with ground truth and use it for quantitative comparisons. We find our approach to outperform several state-of-the-art algorithms.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Forrester Cole, Phillip Isola, William T. Freeman, Frédo Durand, and Edward H. Adelson. ShapeCollage: Occlusion-aware, example-based shape interpretation. In *ECCV*, 2012.

[2] Aaron Hertzmann and Steven M. Seitz. Shape and materials by example: A photometric stereo approach. In *CVPR*, 2003.

[3] Katsushi Ikeuchi and Berthold K.P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(1):141–184, 1981.

[4] Micah K. Johnson and Edward H. Adelson. Shape estimation in natural illumination. In *CVPR*, 2011.

[5] Nazar Khan, Lam Tran, and Marshall Tappen. Training many-parameter shape-from-shading models using a surface database. In *ICCV Workshops*, 2009.

[6] Jan J. Koenderink. What does the occluding boundary tell us about solid shape. *Perception*, 13(3), 1984.

[7] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH*, 2001.

[8] Jamie Shotton, John M. Winn, Carsten Rother, and Antonio Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *IJCV*, 81(1):2–23, 2009.

[9] Daniel Stockman. Vienna 2010 35. Flickr, 2010. licensed under `https://creativecommons.org/licenses/by-sa/2.0/`.

[10] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *TPAMI*, 37(1):67–79, 2014.