

40 years of Computer Graphics in Darmstadt

A discriminative approach to perspective shape from shading in uncalibrated illumination



Stephan R. Richter*, Stefan Roth

Department of Computer Science, TU Darmstadt, Germany

ARTICLE INFO

Article history:

Received 21 August 2015

Received in revised form

23 August 2015

Accepted 23 August 2015

Available online 9 September 2015

Keywords:

Shape from shading

Discriminative prediction

Silhouette cues

Synthetic training data

ABSTRACT

Estimating surface normals from a single image alone is a challenging problem. Previous work made various simplifications and focused on special cases, such as having directional lighting, known reflectance maps, etc. This is problematic, however, as shape from shading becomes impractical outside the lab. We argue that addressing more realistic settings requires multiple shading cues to be combined as well as generalized to natural illumination. However, this requires coping with an increased complexity of the approach and more parameters to be adjusted. Starting from a novel large-scale dataset for training and analysis, we pursue a discriminative learning approach to shape from shading. Regression forests enable efficient pixel-independent prediction and fast learning. The regression trees are adapted to predicting surface normals by using von Mises–Fisher distributions in the leaves. Spatial regularity of the normals is achieved through a combination of spatial features, including textron as well as novel silhouette features. The proposed silhouette features leverage the occluding contours of the surface and yield scale-invariant context. Their benefits include computational efficiency and good generalization to unseen data. Importantly, they allow estimating the reflectance map robustly, thus addressing the uncalibrated setting. Our method can also be extended to handle perspective projection. Experiments show that our discriminative approach outperforms the state of the art on various synthetic and real-world datasets.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Shape from shading – the problem of estimating surface normals from just a single image – is a heavily ill-posed problem. For this reason many simplifying assumptions have been made, such as assuming smooth surfaces, uniform albedo, a known reflectance map, or even light coming from a single directional light source in known direction. Such strong assumptions strongly limit the applicability in practice, however. Outside of controlled lab settings, less restrictive assumptions are needed. In this paper, we estimate the surface of a diffuse object with uniform albedo together with its reflectance map in uncontrolled illumination, given only a single image (Fig. 1). To recover fine surface detail, our goal is to avoid strong spatial regularization. To that end, we generalize shading cues to more realistic lighting, as well as combine them owing to their complementary strengths. While this affects the model and computational complexity, and leads to an increased number of parameters, we show how to address

these challenges with a discriminative learning approach to shape from shading.

A key property of our approach is that it allows to combine several shading cues. We consider (1) *the color* of the pixel itself, which is a strong cue in hued illumination [2], and is often exploited by using a second order approximation of Lambertian shading [3]. Our experiments (Section 9.1) show, however, that the cue becomes less reliable in the presence of correlated color channels (e.g. in near white light) or noise. We aid disambiguation by adding (2) *local context* [4], which to date has been limited to the case of directional lighting. We capture the local appearance context using a textron filter bank [5], instead of using the colors in the neighborhood directly. Through cue combination in our learning framework, we achieve automatic adaptation to uncontrolled lighting and reconstruct fine surface detail. Finally, we introduce novel (3) *silhouette features*. While the use of silhouette information in shape from shading dates back to foundational work by Ikeuchi, Horn, and Koenderink [6,7], previous work has only constrained surface normals at the occluding contour and employed global reasoning to propagate the information to the interior [8]. We show how to generalize the occluding contour constraint to the surface interior, which yields (spatial) contour information at every pixel that is furthermore invariant to the local

* Corresponding author.

E-mail address: stephan.richter@visinf.tu-darmstadt.de (S.R. Richter).



Fig. 1. Shape and reflectance estimation from a single internet image: input image [1], estimated normals and reflectance map from our method, and rendering from a novel view (from left to right).

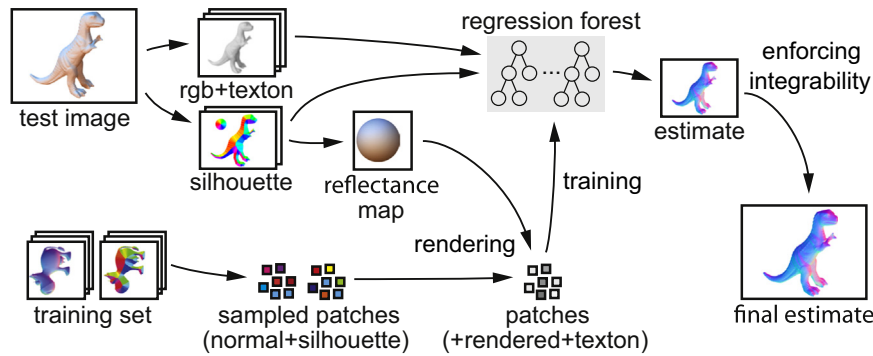


Fig. 2. Pipeline for both training and testing. For each test image, we estimate a reflectance map to train the regression forest on synthetically generated data. Pixel-independent surface normal predictions are made using the trained regression forest. Integrability of the normal field can be enforced optionally.

scale of the object. These silhouette features are applicable to both orthographic and perspective cameras. Moreover, our novel silhouette features also give a coarse estimate of the surface by themselves, which allows us to estimate the unknown reflectance map.

A number of challenges arise in discriminative learning for uncalibrated shape from shading: First, we require a training database of surfaces captured in the same conditions as the object to be reconstructed. It seems infeasible to capture all possible combinations of surfaces and lighting conditions, and inserting known reference objects in the scene [9] is typically impractical. For this reason, some learning approaches [10,11] create databases on-the-fly by rendering synthetic shapes once the lighting condition is known. Our approach adopts this strategy, but relies on a significantly larger database of 3D models than previous work in order to capture the variation of realistic surfaces. Second, and in contrast to [10,11], we cope with unknown illumination at test time. To that end we estimate the reflectance map from our silhouette features, and train the discriminative approach once the reflectance has been estimated. Third, (re-)training for the specific lighting condition at test time requires efficient learning and inference. Enabled by the diverse cues discussed above, we adopt regression forests for efficient pixel-independent surface normal prediction by storing von Mises–Fisher distributions in the leaves. Finally, an optional refinement step enforces integrability of the predicted surface normals. Fig. 2 depicts the entire pipeline. Note that this work is based on a previous conference publication [12], which we generalize here to the perspective case. Moreover, we provide additional detail and illustrations.

After introducing our approach, we assess the contribution of the different cues using a statistical evaluation and as components of our

pipeline. Moreover, we evaluate our method both qualitatively and quantitatively on synthetic data as well as a novel real-world dataset, where it outperforms several state-of-the-art algorithms.

2. Related work

As shape from shading has an extensive literature, we only review the most relevant, recent work here and refer the reader to [13,14]. Lambertian shape from shading has historically assumed a single white point light source, presuming this to simplify the problem. Recently, it became apparent [15,2], however, that chromatic illumination not only resembles real-world environments more closely, but also yields additional constraints on shape estimation, thus substantially increasing accuracy. Nevertheless, these methods focus on the case of favorable illumination and do not address nearly monochromatic lighting. Moreover, assuming the illumination to be known limits their practical applicability.

Recent work also aimed to infer material properties or illumination alongside the shape. Oxholm and Nishino [8] exploit orientation cues present in the lighting environment to estimate the object's bidirectional reflectance function (BRDF) together with its shape. They require a high-quality environment map to be captured, however. Barron and Malik [16,17] integrate shape estimation into the decomposition of a single image into its intrinsic components. Training and inference in their generative model take significant time; moreover, extending the model with additional cues is not necessarily straightforward. Since their formulation requires strong regularity assumptions, the amount of fine surface detail recovered is quite limited.

Download English Version:

<https://daneshyari.com/en/article/10336904>

Download Persian Version:

<https://daneshyari.com/article/10336904>

[Daneshyari.com](https://daneshyari.com)