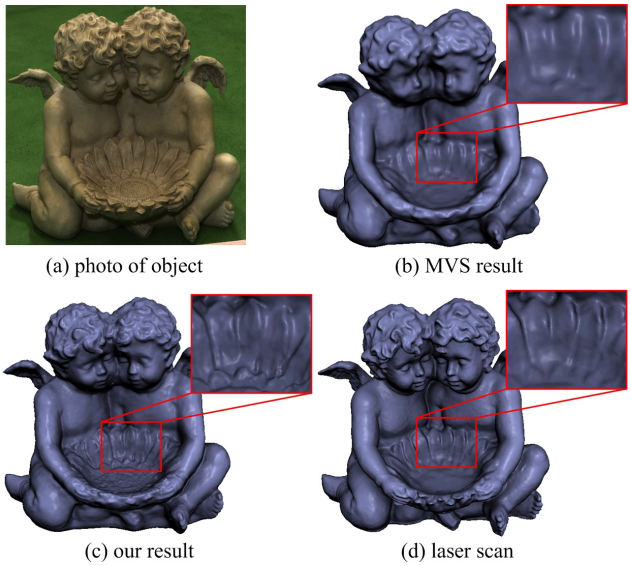
High-quality shape from multi-view stereo and shading under general illumination

Chenglei Wu[1,3*] Bennett Wilburn[2] Yasuyuki Matsushita[2] Christian Theobalt[1]
[1]MPI Informatik [2]Microsoft Research Asia [3]Intel Visual Computing Institute

Abstract

Multi-view stereo methods reconstruct 3D geometry from images well for sufficiently textured scenes, but often fail to recover high-frequency surface detail, particularly for smoothly shaded surfaces. On the other hand, shape-from-shading methods can recover fine detail from shading variations. Unfortunately, it is non-trivial to apply shape-from-shading alone to multi-view data, and most shading-based estimation methods only succeed under very restricted or controlled illumination. We present a new algorithm that combines multi-view stereo and shading-based refinement for high-quality reconstruction of 3D geometry models from images taken under constant but otherwise arbitrary illumination. We have tested our algorithm on several scenes that were captured under several general and unknown lighting conditions, and we show that our final reconstructions rival laser range scans.



(a) photo of object (b) MVS result (c) our result (d) laser scan

Figure 1. Our approach reconstructs models of much higher detail than state-of-the-art stereo approaches.

1. Introduction

Multi-view stereo (MVS) methods compute depth by triangulation from corresponding views of the same scene point in multiple images. Establishing correspondence is difficult within smoothly shaded regions, so MVS methods compute accurate depth for a sparse set of well-localized points and must interpolate elsewhere. Seitz *et al*. present a taxonomy and evaluation of MVS algorithms [26]. Results posted on the benchmark website [1] accompanying that work show that today's best-performing methods capture the rough shape of the scene well, but generally cannot recover the high-frequency shape detail well. In contrast to MVS, shape-from-shading (SfS) methods compute per-pixel surface orientation instead of sparse depth. SfS techniques use shading cues to estimate shape from a single image, usually taken under illumination from a single direction [33]. It was shown that SfS approaches are able to recover high-frequency shape detail, also if surfaces are smoothly shaded. SfS reconstruction can therefore often shine where stereo fails, and vice versa. Generalizing this shading-based reconstruction to the multi-view case is not easy, though. Recovered normal fields usually need to be integrated to obtain 3D geometry, which is non-trivial for general surfaces seen from multiple viewpoints [22]. Furthermore, most SfS algorithms make strong assumptions about the incident illumination, which effectively restricts most of them to studio lighting conditions.

In this paper, we therefore propose a new multi-view reconstruction approach that combines the strengths of MVS and SfS. It enables us to capture high-quality 3D geometry of Lambertian objects from images recorded under fixed but otherwise arbitrary unknown illumination. In the design of our algorithm, we capitalize on recent progress in the field of real-time rendering. By parameterizing incident illumination and visibility in the spherical harmonic (SH) domain, the image formation model - the rendering equation - can be efficiently evaluated, and even complex lighting effects un-

*Part of this work was done while the first author was visiting Microsoft Research Asia as a research intern.

der general illumination can be displayed in real-time [27]. In our work, we use the spherical harmonic formulation of the imaging equation to solve the opposite problem, namely to passively reconstruct high-quality geometry under general illumination.

Contributions. In this paper, we propose a shape reconstruction method that uses stereo for initial geometry estimation and shading-based shape refinement under *general* and *uncalibrated* illumination. Our method estimates high-fidelity shapes that include subtle geometric details that cannot be captured by triangulation-based approaches. We develop a new multi-view shading constraint for achieving this goal. For efficient computation, we use spherical harmonics to estimate and encode arbitrary lighting conditions and local visibility. We also develop an adaptive anisotropic smoothness term for preserving high-frequency details while filtering out noise. In addition, we show an adaptive computation approach that takes the complexity of lighting and visibility estimates into account at each surface point for efficient computation.

We have tested and validated our approach on a variety of real and synthetic scenes captured under several illumination conditions. The level of detail in our results rivals and sometimes even goes beyond the quality of data obtained with laser range scanners as shown in Figs. 1 and 10.

2. Related work

The complementary nature of MVS and SfS has long been known [5], and much work has been done on efficiently combining the two techniques. Leclerc *et al*. [19] use stereo to provide initialization and boundary constraints for SfS. Cryer *et al*. [6] combine depth maps from SfS and stereo in the frequency domain using filtering. Rather than fusing MVS and SfS results, Fua *et al*. [8] start with a coarse mesh computed from binocular or tri-view stereo, then minimize an error function with stereo, shading, and smoothness components. They handle slowly varying albedo of Lambertian surfaces. Samaras *et al*. [25] iteratively estimate both shape and illumination given multiple views taken under fixed illumination. They assume piecewise constant albedo. Jin *et al*. [13, 14, 15] have proposed a series of variational algorithms that combine MVS and SfS. Their recent work [12] focuses on 3D reconstruction of Lambertian objects with piece-wise constant albedo.

The methods discussed above all neglect self-shadowing (shadows cast by the scene onto itself). They also all assume either a single distant light source, or a distant point light source plus uniform ambient illumination. Even for the case of multiple distant light sources, we are not aware of any methods that do not require the number of light sources to be known in advance. For more general capture scenarios, the illumination is usually neither known nor simply a sum of ambient lighting and a few distant point light sources.

Beeler *et al*. [4] recently proposed a high-quality stereo method that uses shading-based refinement. Their refinement embosses or extrudes the geometry at locations where high-frequency shading variations are visible, producing qualitatively pleasing results that are not intended to be metrically correct. Although we consider only Lambertian reflectance, prior work has also focused on reconstructing non-Lambertian objects. Yu *et al*. [32, 31] propose two algorithms for modeling non-Lambertian objects illuminated by distant light sources. Both explicitly model the reflectance using either a Phong or Torrence-Sparrow model, then iteratively optimize the estimated shape and reflectance. One of the methods [31] can be tailored to handle unknown lighting directions, but only for Lambertian objects. Yoon *et al*. [30] reconstruct geometry by combining stereo and shading cues in a variational framework. Their method handles general dichromatic surfaces using the Blinn-Phong shading model and assumes known lighting directions. Because the algorithm is variational, it is susceptible to local minima and is generally overly smooth.

The main difference between our method and the prior art is that we combine MVS and SfS for general, unknown illumination. Our approach is motivated by the work of Basri *et al*. [3] and Ramamoorthi *et al*. [23], who observe that light reflection is a convolution of a reflectance kernel over the sphere of incoming light directions, and that the Lambertian reflectance kernel acts as a low-pass filter which preserves only the lowest frequency components of the illumination. Thus, the illumination can be modeled well using low-frequency spherical harmonics. Frolova *et al*. [7] further analyze the accuracy of the spherical harmonic approximation for far and near illumination. The spherical harmonic lighting approximation has been used for photometric stereo [2] and multi-view photometric stereo [29]. Photometric stereo [28], however, assumes images taken under different lighting conditions to fully constrain the surface normal. Multi-view photometric stereo [11, 16, 10] combines MVS and photometric stereo to achieve high-quality reconstruction, but still requires multi-illumination images to be captured. These methods also do not account for self-shadowing. By contrast, our method assumes only a single lighting condition and explicitly handles self-shadowing. Besides spherical harmonics, wavelet is also employed to represent the general illumination for relighting applications [10].

3. Algorithm

Our goal is to compute high-quality shape of a static object based on given multiple images taken from different viewpoints by combining MVS and SfS. The illumination is assumed to be fixed and distant, but is otherwise general

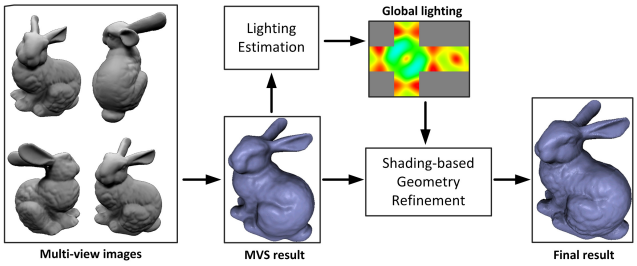


Figure 2. Outline of our processing pipeline.

and unknown. Cameras are assumed to be calibrated both geometrically and radiometrically. We represent the geometry using a high-resolution mesh model and the illumination using spherical harmonics. In order to keep the problem tractable, we henceforth assume that the albedo of the object is constant. We will see in our results, though, that this assumption does not prevent us from reconstructing detailed shape models even in the presence of small albedo variations. We also neglect inter-reflections on the object.

The workflow of our method is shown in Fig. 2. It has three steps. First, we use existing MVS methods to create an initial closed, 3D triangle mesh model of the object. Next, we use this model to estimate the spherical harmonic coefficients for the incident illumination (Sec. 3.2). Finally, we refine the MVS geometry so that shading variations in the input images are properly explained by our image formation model and the estimated geometry (Sec. 3.3). The next subsections review image formation using the SH illumination model and explain the illumination estimation and geometry refinement in detail. As we will describe, we handle concave surfaces and self-occlusion by computing the visibility of each vertex from all directions, and we adaptively tune the order of the SH approximation for higher accuracy in areas with higher ambient occlusion (*i.e.*, more self-occlusion).

3.1. Image Formation Model

Assuming all objects in the scene are non-emitters and the light sources are infinitely distant, the image irradiance equation can be defined as [17]

$$B(x,\omega_o) = \int_\Omega L(\omega_i)V(x,\omega_i)\rho(\omega_i,\omega_o)\max(\omega_i\cdot\boldsymbol{n}(x),0)d\omega_i, \tag{1}$$

where $B(x,\omega_o)$ is the reflected radiance, and the variables x, $\boldsymbol{n}$, ω_i and ω_o are the spatial location, the surface normal, and the incident and outgoing angles, respectively. The symbol Ω represents the domain of all possible directions, $L(\omega_i)$ represents the incident lighting, $V(x,\omega_i)$ is a binary visibility function, and $\rho(\omega_i,\omega_o)$ is the bidirectional reflectance distribution function (BRDF) of the surface. For convenience, we scale the incident illumination by the albedo, letting $L_a(\omega_i) = \rho L(\omega_i)$.

3.2. Lighting Estimation

An initial mesh model of the object is reconstructed via MVS [9, 20]. Based on this model, our method first estimates the SH coefficients for the incident illumination. Here, we explicitly take the visibility function into account because we want to reconstruct non-convex objects. For convenience, we define $T(x,\omega_i) = V(x,\omega_i)\max(\omega_i\cdot\boldsymbol{n}(x),0)$. Representing L_a and T in Eq. (1) using low order spherical harmonics, and due to the orthogonality of the SH basis, the image irradiance equation becomes

$$B(x) = \int_\Omega L_a(\omega_i)T(x,\omega_i)d\omega_i = \sum_{k=1}^{n^2} l_k t_k, \tag{2}$$

where $n-1$ is the order of the SH, and l_k and t_k are, respectively, the SH coefficients of lighting L_a and visibility T. The irradiance B is known from the images, and the MVS geometry gives us an approximation for the visibility coefficient t_k. First, we use the model to compute the visibility of each vertex as a function of incident light direction. For each vertex, the coefficients t_k are the projection of the product of the visibility function and the clamped cosine function onto the SH basis functions. We calculate the coefficients $\boldsymbol{l} = \{l_1,\dots l_{n^2}\}$ by minimizing the ℓ_1 norm of the difference between the measured and computed image irradiances at each mesh vertex:

$$\hat{\boldsymbol{l}} = \underset{\boldsymbol{l}}{\operatorname{argmin}} \sum_i \sum_{c\in Q(i)} |\sum_{k=1}^{n^2} l_k t_k - I_c(P_c(\boldsymbol{x}_i))|. \tag{3}$$

Here, i is the vertex index, c is the camera index, $Q(i)$ is the set of cameras that can see the i-th vertex $\boldsymbol{x}_i$, P_c is the projection matrix for camera c, and $I_c(P_c(\boldsymbol{x}_i))$ represents the image intensity corresponding to vertex i and captured by camera c. The ℓ_1 norm makes this estimation robust in the presence of outliers like interreflections, specularities, and errors in the MVS geometry.

We are estimating the low order SH coefficients for the illumination here. The specified order number is automatically decided by local occlusion situation on the surface, see Sec. 3.4.

3.3. Shading-based Geometry Refinement

Given the current estimated geometry and illumination, the final step is to refine the geometry using shading information. For this step, we compute visibility for each vertex using the current geometry, and assume that it does not change during the refinement. If we define $L_v(x,\omega_i) = L_a(\omega_i)V(x,\omega_i)$, the image irradiance equation can be rewritten

$$B(x) = \int_\Omega L_v(x,\omega_i)\max(\omega_i\cdot\boldsymbol{n}(x),0)d\omega_i. \tag{4}$$

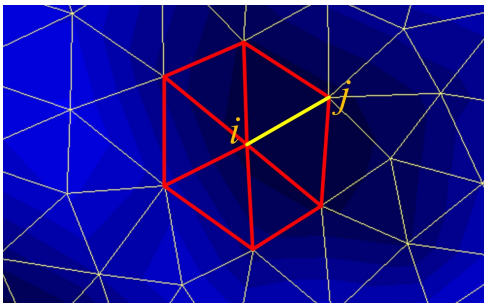


Figure 3. Anisotropic smoothness constraint: the smoothness weight for each edge is determined by the image gradient along the edge.

This is a convolution of L_v with the clamped cosine kernel, which is determined by the surface normal. Letting g_{km} be the SH coefficients for L_v and according to the Funk-Hecke theorem [3], the convolution results in the same harmonic scaled by g_{km} and another scalar α_k. Thus, the image irradiance equation can be expressed as

$$B(x) = \sum_{k=0}^{n-1} \sum_{m=-k}^{k} \alpha_k g_{km} Y_{km}, \tag{5}$$

where Y_{km} is the SH function. The scalar α_k is defined as

$$\alpha_k = \sqrt{\frac{4\pi}{2k+1}} h_k, \tag{6}$$

where h_k are the SH coefficients for the clamped cosine function. Here, we have to allow the use of higher order spherical harmonic approximations when necessary (see Sec. 3.4). The function Y_{km} depends only on the surface normal $\boldsymbol{n}$.

We run an optimization for each vertex position that attempts to minimize shading errors in all visible views. The computed irradiance is unlikely to match the observed irradiance, for many reasons: interreflections, radiometric calibration errors, approximation errors for the spherical harmonic illumination representation, and so on. Rather than directly comparing irradiance values, we compare the gradients of the observed and computed irradiances at each vertex. This is natural, because shading is expected to be more accurate for higher frequency shape components. Mathematically, we define the multi-view shading gradient error E_0 as

$$E_0 = \sum_i \sum_{j \in N(i)} \sum_{c \in Q(i,j)} \big(g_c(i,j) - s(i,j)\big)^2, \tag{7}$$

where i and j are vertex indices, $N(i)$ is the set of the neighbors of the i-th vertex, c is the camera index, $Q(i,j)$ is the set of cameras which see vertex i and j, and g and s are the measured image gradient and predicted shading gradient, respectively. We compute the gradients g and s with direct differences, namely,

$$\begin{aligned} g_c(i,j) &= I_c(P_c(\boldsymbol{x}_i)) - I_c(P_c(\boldsymbol{x}_j)), \quad \text{and} \\ s(i,j) &= B(\boldsymbol{x}_i) - B(\boldsymbol{x}_j). \end{aligned}$$

The shading value B is calculated according to the Eq. (5). With the estimated illumination, the only remaining undefined variable in Eq. (5) is the normal $\boldsymbol{n}$, which we can compute from the vertices' positions. We limit vertex displacements to 3D locations that project into the object's silhouettes in all input views. Combining the silhouette and shading constraints gives the following new objective function E_1 for the multi-view shading gradient:

$$E_1 = \sum_i \sum_{j \in N(i)} \sum_{c \in Q(i,j)} d(i,j,c), \tag{8}$$

where i, j, $N(i)$, c, $Q(i,j)$ are the same as in Eq. (7). The function $d(i,j,c)$ has the following form:

$$d(i,j,c) = \begin{cases} (g_c(i,j) - s(i,j))^2, & M(\boldsymbol{x}_i) \cdot M(\boldsymbol{x}_j) \neq 0 \\ \infty, & \text{otherwise,} \end{cases} \tag{9}$$

where ∞ is a large constant that imposes a severe penalty if a vertex leaves the silhouettes, and M is a mask image which is non-zero inside the silhouettes and zero outside.

Smoothness constraint In practice, we have found that the shading gradient error alone leads to noisy reconstructions in areas where the normal is not sufficiently constrained or where errors in our image irradiance approximation are significant. Traditional smoothness terms might erroneously remove fine shape detail. We thus use an anisotropic smoothness constraint based on the image gradient that filters noise while preserving details captured by the shading gradient constraint. We observe that for objects of uniform albedo, the image gradient can be used to infer geometric smoothness. We use a small smoothness weight in regions with large image gradients, allowing the shading constraint to capture fine detail. In areas where the image gradient is small, the shape is most likely smooth, so we use a larger smoothness weight. Fig. 3 shows this idea.

The smoothness constraint is imposed between vertex i and its 1-ring neighbors, with the weight being assigned to the corresponding edges. An isotropic smoothness constraint would require the geometric differences between vertex i and its neighbors to be as small as possible, with the same weight for each edge. Our anisotropic smoothness term, on the other hand, assigns different weights based on the image gradient between neighboring vertices. The weight of e_{ij}, for example, is determined by the corresponding image gradient in the camera most directly facing the vertex i. The weight for each edge is defined as

$$w_{ij}^s = 1 - \min(\hat{g}(i,j), C)/C, \tag{10}$$

where $\hat{g}(i,j)$ is the image gradient and C is a constant setting a lower bound on the smoothness weight when the gradient is large.

Combining the anisotropic weights with traditional mean curvature flow [21], the smoothness term E_2 becomes

$$E_2 = \sum_i \| \sum_{j \in N(i)} w_{ij}^s w_{ij}^m (\boldsymbol{x}_i - \boldsymbol{x}_j) \|_2^2, \qquad (11)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the positions of vertex i and j, and w_{ij}^m is the common cotangent weight. The cotangent weight w_{ij}^m is defined as

$$w_{ij}^m = \frac{1}{2A_i}(\cot \alpha_{ij} + \cot \beta_{ij}), \qquad (12)$$

where α_{ij} and β_{ij} are the two angles opposite to the edge (v_i, v_j), and A_i is the Voronoi area of vertex v_i.

We optimize a cost function summing the shading gradient E_1 and smoothness constraints E_2, defined as

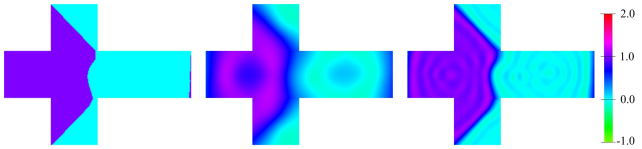
$$E = \lambda E_1 + (1 - \lambda) E_2, \qquad (13)$$

where λ is a weighting factor. Optimizing all the vertex positions simultaneously is computationally intractable because of the non-linear SH function. Optimizing vertices one at a time, however, does not afford enough flexibility to adjust the local surface shape. Our algorithm visits each vertex in turn in a fixed order, optimizing the positions of a patch comprising the vertex and its 1-ring neighbors in each step. To avoid self-intersections as far as possible, we restrict vertex motion to be along the initial surface normal direction.

We could iterate by recomputing visibility using the refined geometric model, re-estimating lighting, refining the geometric model, and so on. In practice, however, we find that one pass suffices for an accurate reconstruction.

3.4. Adaptive Geometry Refinement

For convex Lambertian objects, low-order spherical harmonics suffice to approximate the irradiance well. For more complex objects, however, we must use high-order approximations, which are slower to compute. We use the local ambient occlusion [18] to adapt the order of the SH approximation to the geometry. Ambient occlusion corresponds roughly to an integral over the local visibility hemisphere, so it is high for vertices with more local self-occlusion. We segment the mesh into two sets based on whether the ambient occlusion at each vertex is over a threshold, and use high-order and low-order SH approximations for vertices with high and low ambient occlusion, respectively (Fig. 6 (d)(e)). Although the SH approximation error depends on the specific visibility function at each vertex, not just its integral, we have found that the ambient occlusion gives a good balance between reconstruction accuracy and computational complexity.



(a) visibility map (b) representation by 4th order SH (c) representation by 14th order SH

Figure 4. An example visibility map and its SH representations of different order.

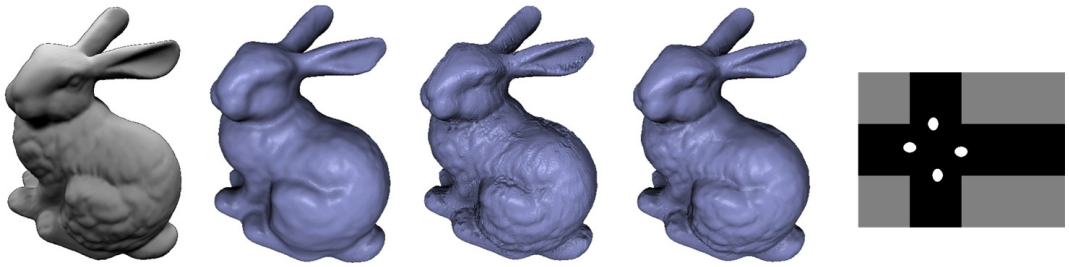
	Position[‰]		Normal[deg.]		Runtime
	mean	std	mean	std	
MVS result	1.44	1.24	8.66	6.93	
adaptive, no smoothing	1.17	1.13	8.53	9.99	2 hours
adaptive + smoothing	1.15	1.07	7.05	**6.03**	2 hours
4th order + smoothing	1.19	1.13	7.28	6.28	**1 hour**
16th order + smoothing	**1.13**	**1.06**	**6.91**	6.17	4 hours

Table 1. Quantitative evaluation on synthetic data. First column: position error (in ‰ of bounding box dimension). Second row: error in surface normal direction in degrees. Third row: run time.

4. Results

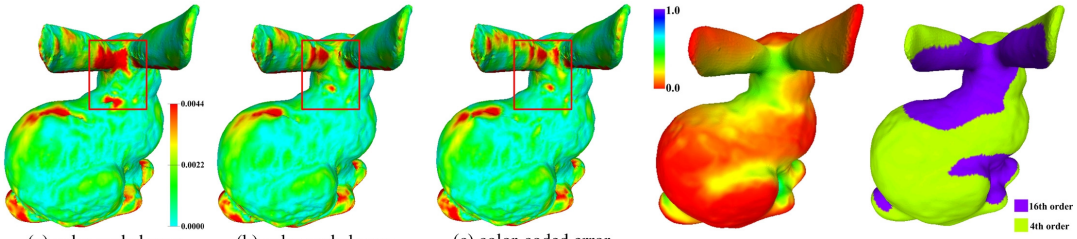
We validated our algorithm using a synthetic bunny model, shown in Fig. 5, and four real world data sets: an angel statue (Fig. 1), a sculpture of a fish (Fig. 7), a plaster cast of a face (Fig. 10), and a crumpled sheet of paper (Fig. 9). For the real world models, we took between 22 and 33 photos with a Canon $5D$ Mark II from calibrated positions. We captured images at the full camera resolution and cut out the region of interest containing the object, yielding images of around 800×600 pixels. For some models we also captured laser range scans with a Minolta Vivid 910. We used Furukawa's method [9] to generate the initial MVS models for the angel and the paper, and Liu's method [20] for the bunny, the fish and the face. These MVS results are re-meshed to get a uniform triangulation, resulting in 30000 vertices for the bunny and 200000 vertices for the real scenes. We use DirectX to render a cube-map for the visibility function at each vertex in the re-meshed result. Fig. 4 shows an example visibility map and its SH representations at different orders. For the synthetic model, we used 4 simulated area light-sources (Fig. 5 (e)). The real objects were captured in two different environments: a large indoor atrium environment with a variety of light sources at different locations and distances (lighting I), and a room with several rows of standard office lighting on the ceiling (lighting II), Fig. 8. For the lighting estimation, we used conjugate gradient to solve the ℓ_1 minimization problem in Eq. (3). The shape is then refined by minimizing Eq. (13) using the Lebvenberg-Marquardt algorithm.

Parameters There are two tunable parameters in our method, λ in Eq. (13), and C in Eq. (10). Experimentally, we determined $\lambda = 0.3$ for all data sets. C was set to 20 for the bunny model, 100 for the angel model, and 50 for the



(a) image synthetically generated | (b) MVS result | (c) the reconstructed model without anisotropic smoothing | (d) the final optimized model with anisotropic smoothing | (e)environment map for synthetic data

Figure 5. Evaluation on synthetic data.



(a) color-coded error rendering for 4th order | (b) color-coded error rendering for 16th order | (c) color-coded error rendering for adaptive order | (d) ambient occlusion | (e) SH order adjustment

Figure 6. Adaptive geometry refinement: Our reconstruction using adaptive SH order (c) is almost as accurate as the high order case (b), which is obviously better than the low order case (a). The SH order used for every vertex (e) depends on its ambient occlusion value (d).

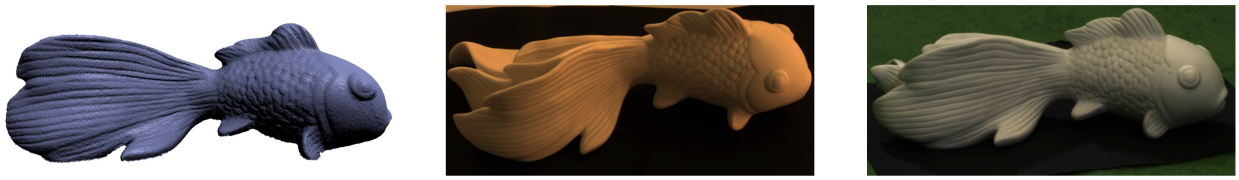
other real-world models. Generally, the selection of C depends on the level of image noise and uniformity of the albedos. For instance, less uniform albedos require a higher C. The per-vertex ambient occlusion threshold value (Sec. 3.4) was set to 0.1. 4-th order SH approximations were used for vertices with low ambient occlusion. Vertices over the threshold used 14-th and 16-th order approximations for the real and synthetic sets, respectively.

Runtime performance The algorithm's run time depends on the mesh density, the SH order, and the cube map dimensions for rendering and SH projection. The bunny mesh has 30000 vertices and was computed using visibility cube maps with 64×64 facets. Using unoptimized code on a standard PC with a 2.66 GHz Core 2 Quad processor, rendering the visibility map takes 33 minutes, and optimizing the shape takes roughly 1 hour and 30 minutes. Higher SH orders improve reconstruction quality, particularly in starkly occluded areas, Fig. 6 (a), (b). Reconstruction of the bunny with 4-th order SH coefficients (Fig. 6 (a)) takes roughly 1 hour, but produces less accurate results than a full reconstruction with 16-th order (Fig. 6 (b)) which takes 4 hours to compute. Adaptive refinement reduces the runtime to only 2 hours with accuracy comparable to using high order coefficients throughout (Fig. 6 (c)).

Synthetic scene Our synthetic dataset was generated by rendering 20 images of the "bunny" model at 800×600 pixel resolution. Fig. 5 (a) shows an example image. The MVS result (Fig. 5 (b)) lacks fine scale detail. Our refinement without anisotropic smoothing (Fig. 5 (c)) brings out more detail, but also has artifacts on the surface. In contrast, our complete reconstruction approach (Fig. 5 (d)) shows the high-frequency shape detail nicely with no disturbing artifacts. Table 1, a numerical evaluation of the reconstruction error w.r.t. the ground truth model, confirms the accuracy of our results.

Real-world scenes Our algorithm produces results of similarly high quality for the real objects as shown in Figs. 1, 7, 9, and 10. While the MVS reconstruction consistently fails to capture high-frequency details, our algorithm produces results with an accuracy that rivals and sometimes exceeds the quality of a laser range scan. For instance, in Fig. 10 our approach not only brings out the birth marks and pimples in the skin, but also extracts ridges on the rubber cap that are completely masked by measurement noise in the laser scan. Although the angel statue in Fig. 1 has a slightly varying albedo, our algorithm achieves high-quality results. Thus, in practice the constant albedo assumption is not a strict limitation. Fig. 7 shows reconstructions of a fish figurine captured under two very different lighting conditions (lighting I and II). In both cases, our final model is very accurate and close to the laser scan.

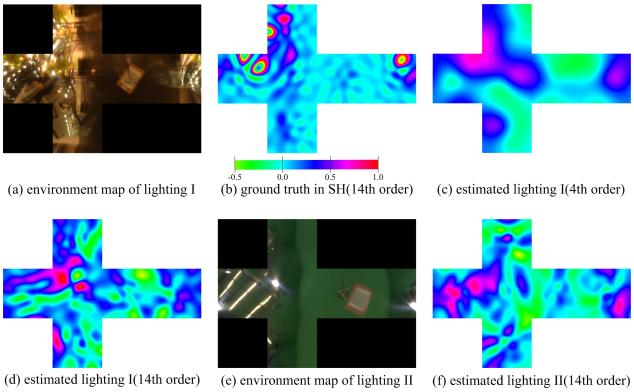
Limitations The approach is subject to a few limitations. The constant albedo assumption limits the application range. In future, we intend to modify the approach to



(a) laser scan (b) photo of object under lighting I (c) photo of object under lighting II

(d) MVS result for lighting II (e) our result for lighting I (f) our result for lighting II

Figure 7. Fish reconstructed under two lighting conditions (cf Fig. 8) - in both cases, our final result (e),(f) is much more detailed than the MVS results ((d) is only shown for lighting II and is better than MVS results for lighting I) and close to the laser scan (a).



(a) environment map of lighting I (b) ground truth in SH(14th order) (c) estimated lighting I(4th order)

(d) estimated lighting I(14th order) (e) environment map of lighting II (f) estimated lighting II(14th order)

Figure 8. Comparing estimated lightings (c), (d), (f) to the captured environment map (a), (e) and the ground-truth SH representation (b).

handle clearly varying surface albedo. Another limitation comes from the assumption of Lambertian reflectance. We would like to amend the approach to be applied to more general materials. Besides, as we use low-order spherical harmonics to represent the lighting, our method may lose effectiveness for band-limited illumination. Also, our method assumes a good initial guess of the geometry and would suffer from a failure of the MVS. In future, we intend to start from a mesh obtained by active sensing methods [24].

5. Conclusion

We presented a new approach for purely image-based reconstruction of 3D models with extremely high surface detail. The core of the method is a shading-based refinement strategy for stereo reconstructions that succeeds under general unconstrained illumination. An efficient representation of visibility and lighting in the spherical harmonic domain enables the method to reliably estimate incident illumination and exploit it for high-quality shape improvement. Both visual and quantitative analysis show that our purely image-based results even rival laser range scans.

Acknowledgment

We thank Minmin Gong for help with the SH implementation, Yebin Liu for his MVS method, Carsten Stoll and Samir Hammann for discussions and the face model.

References

[1] Middlebury multi-view stereo evaluation. http://vision.middlebury.edu/mview/.

[2] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 72(3):239–257, 2006.

[3] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE TPAMI*, 25(2):218–233, 2003.

[4] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-quality single-shot capture of facial geometry. *ACM TOG (Proc. SIGGRAPH)*, 29(3), 2010.

[5] A. Blake, A. Zimmerman, and G. Knowles. Surface descriptions from stereo and shading. *Image Vision Comput.*, 3(4):183–191, 1986.

[6] J. E. Cryer, P. S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, 1995.

[7] D. Frolova, D. Simakov, and R. Basri. Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. *Proc. of ECCV*, 2004.

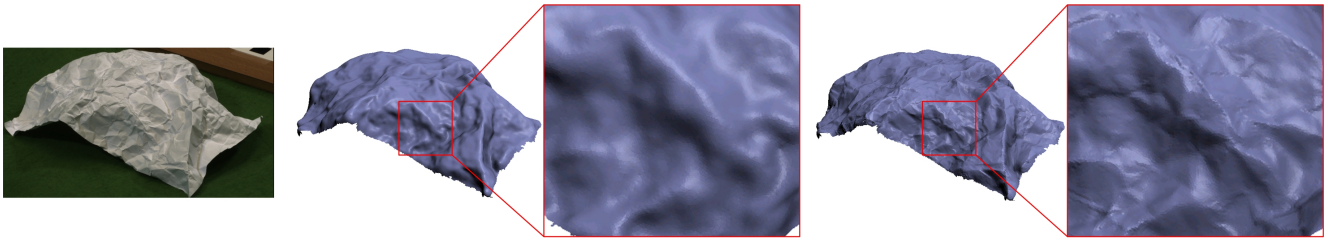
[8] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *IJCV*, 16:35–56, 1995.

[9] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE TPAMI*, 2010.

[10] T. Haber, C. Fuchs, P. Bekaer, H.-P. Seidel, M. Goesele, and H. Lensch. Relighting objects from image collections. In *Proc. of CVPR*, pages 627 –634, 2009.

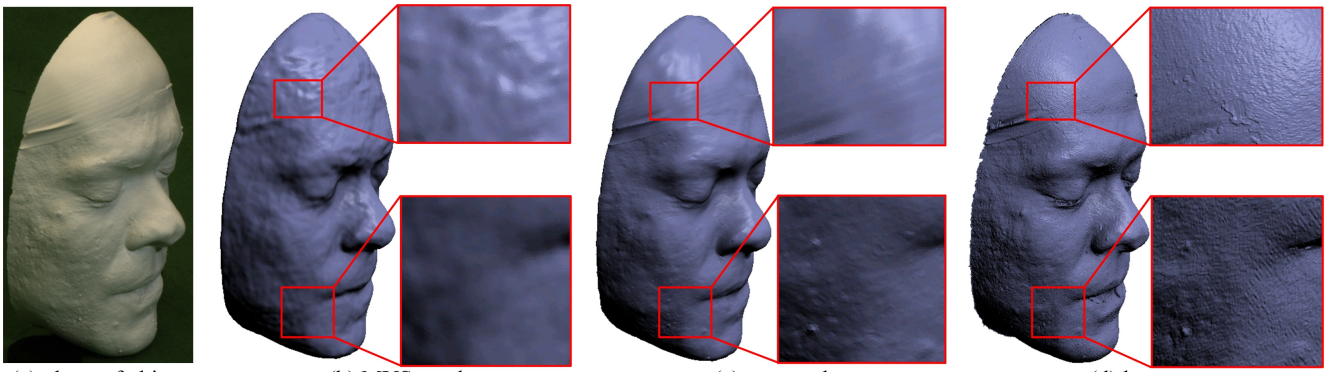
[11] C. Hernndez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE TPAMI*, 30:548–554, 2008.

[12] H. Jin, D. Cremers, D. Wang, E. Prados, A. Yezzi, and S. Soatto. 3-d reconstruction of shaded objects from multiple images under unknown illumination. *IJCV*, 76(3):245–256, 2008.



(a) photo of object (b) MVS result (c) our result

Figure 9. Our reconstruction of the crumpled paper recovers high frequency shape (c), while it is absent in the stereo result (b).



(a) photo of object (b) MVS result (c) our result (d) laser scan

Figure 10. Reconstruction of a face plaster cast - the stereo result (b) lacks a lot of detail. Our reconstruction (c) captures even small scale detail that, in the laser scan (d), is hidden by noise.

[13] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. *CVPR*, 1:36–42, 2004.

[14] H. Jin, A. Yezzi, and S. Soatto. Stereoscopic shading: integrating multi-frame shape cues in a variational framework. In *Proc. of CVPR*, volume 1, pages 169–176, 2000.

[15] H. Jin, A. J. Yezzi, and S. Soatto. Region-based segmentation on evolving surfaces with application to 3d reconstruction of shape and piecewise constant radiance. In *Proc. of ECCV*, pages 114–125, 2004.

[16] N. Joshi and D. Kriegman. Shape from varying illumination and viewpoint. In *Proc. of ICCV*, pages 1 –7, 2007.

[17] J. T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, 1986.

[18] M. Langer and H. Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 29:649–660, 2000.

[19] Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *CVPR*, pages 552–558, 1991.

[20] Y. Liu, Q. Dai, and W. Xu. A point cloud based multi-view stereo algorithm for free-viewpoint video. *IEEE TVCG*, 16(3):407–418, 2010.

[21] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Proceedings of VisMath*, pages 35–57, 2002.

[22] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM TOG (Proc. SIGGRAPH)*, 24(3), 2005.

[23] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of convex lambertian object. *Journal of the Optical Society of America*, pages 2448–2459, 2001.

[24] M. Reynolds, J. Doboš, L. Peel, T. Weyrich, and G. J. Brostow. Capturing time-of-flight data with confidence. In *CVPR*, 2011.

[25] D. Samaras, D. Metaxas, P. Ascalfua, and Y. G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *Proc. of CVPR*, pages 480–487, 2000.

[26] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. *CVPR*, 1:519 – 528, 2006.

[27] P.-P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *ACM TOG (Proc. SIGGRAPH)*, pages 527–536, New York, NY, USA, 2002.

[28] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[29] C. Wu, Y. Liu, Q. Dai, and B. Wilburn. Fusing multi-view and photometric stereo for 3d reconstruction under uncalibrated illumination. *IEEE TVCG*, 2010.

[30] K.-J. Yoon, E. Prados, and P. Sturm. Joint estimation of shape and reflectance using multiple images with known illumination conditions. *IJCV*, 86(2-3):192–210, 2010.

[31] T. Yu, N. Xu, and N. Ahuja. Recovering shape and reflectance model of non-lambertian objects from multiple views. In *CVPR*, volume 2, pages 226–233, 2004.

[32] T. Yu, N. Xu, and N. Ahuja. Shape and view independent reflectance map from multiple views. *IJCV*, 73:123–138, 2007.

[33] R. Zhang, P. Tsai, J. Cryer, and M. Shah. Shape from shading: A survey. *IEEE TPAMI*, 21(8):690–706, 1999.