Fusing Multiview and Photometric Stereo for 3D Reconstruction under Uncalibrated Illumination

Chenglei Wu, Yebin Liu, Qionghai Dai, *Senior Member*, *IEEE*, and Bennett Wilburn

Abstract—We propose a method to obtain a complete and accurate 3D model from multiview images captured under a variety of unknown illuminations. Based on recent results showing that for Lambertian objects, general illumination can be approximated well using low-order spherical harmonics, we develop a robust alternating approach to recover surface normals. Surface normals are initialized using a multi-illumination multiview stereo algorithm, then refined using a robust alternating optimization method based on the ℓ_1 metric. Erroneous normal estimates are detected using a shape prior. Finally, the computed normals are used to improve the preliminary 3D model. The reconstruction system achieves watertight and robust 3D reconstruction while neither requiring manual interactions nor imposing any constraints on the illumination. Experimental results on both real world and synthetic data show that the technique can acquire accurate 3D models for Lambertian surfaces, and even tolerates small violations of the Lambertian assumption.

Index Terms—Multiview stereo, photometric stereo, Lambertian reflectance, ℓ_1 minimization.

1 Introduction

Automatic and accurate modeling of objects and scenes from multiple photographs or video clips is an important goal in vision and graphics research. Applications range from creating realistic models for film, television, and computer games, to recovering metric information for scientific data analysis. Although still a challenging and fundamental problem, image-based modeling techniques have improved greatly in the last decade.

Multiview stereo (MVS) techniques acquire solid 3D models from multiple calibrated photographs. Most MVS approaches use stereo matching techniques [6], which work well for textured Lambertian surfaces but often fail in the presence of specular highlights or uniform textures. Alternatively, photometric stereo is a well-established shape recovery technique that creates partial 2.5D reconstructions from a single viewpoint. Photometric stereo measures surface orientations, so the estimated surface normals must be integrated to produce the final shape [18]. These two techniques are complementary, and recent work has shown the benefits of combining both methods to produce very accurate surface models [35].

In this paper, we combine MVS and photometric stereo for a scene imaged under multiple unknown illuminations to produce watertight 3D reconstructions. State-of-the-art multiview photometric stereo algorithms (see Section 2 for an overview) either require accurate light calibration or assume distant point light sources for any image. Calibrating many light sources can be troublesome, and ensuring that all lights can be represented as distant point lights is not always practical. Unlike the prior art, we impose no constraints on the lighting conditions except for distant illumination. In particular, we represent general illumination using spherical harmonics. Basri et al. [21] use spherical harmonics lighting representation for single-view photometric stereo method and recover surface normals up to a transformation. Additional information or manual intervention can resolve this ambiguity; we resolve it using multiple views and a coarse-to-fine approach. We propose a multi-illumination MVS method, which we use to initialize the surface normals. These normals are optimized by a robust alternating method according to the ℓ_1 metric. Our method computes the normals, the reflectance, and the illumination. We explicitly handle outliers (e.g., non-Lambertian reflectance, cast shadows, and interreflections) using an ℓ_1 error metric.

The specific contributions of this work are:

- a shape reconstruction method that combines multiview stereo and photometric stereo method, and works with uncalibrated and relatively unconstrained (distant) active lighting,
- a robust normal recovery method based on an ℓ_1 metric and a low-order spherical harmonic representation for general illuminations, and
- an automatic outlier detection and correction approach based on a low-frequency shape prior.

We have developed a multicamera multilight 3D acquisition system to test the proposed method on extensive data sets. The experimental results validate the effectiveness of the proposed method on both static objects and quasistatic human actors. The multiview multi-illumination data sets captured by our system and the corresponding reconstruction results in this work are all available on our webpage.[1] Recent high-quality free-viewpoint performance capture systems [1], [2] use laser scanners to capture very

- *C. Wu, Y. Liu, and Q. Dai are with the TNList and the Department of Automation, Tsinghua University, Main Building, Beijing 100086, China. E-mail: wucl07@mails.tsinghua.edu.cn, liuyebin@mail.tsinghua.edu.cn, qhdai@tsinghua.edu.cn.*
- *B. Wilburn is with the Visual Computing Group, Microsoft Research Asia, 5F, Beijing Sigma Center, No. 49, Zichun Road, Haidian District, Beijing 100190, China. E-mail: bennett.wilburn@microsoft.com.*

Manuscript received 2 Aug. 2009; revised 20 Mar. 2010; accepted 11 Aug. 2010; published online 13 Oct. 2010.
Recommended for acceptance by B. Guo.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2009-08-0164.
Digital Object Identifier no. 10.1109/TPDS.2010.224.

1. http://media.au.tsinghua.edu.cn/mvml.jsp.

accurate key models. By combining MVS and photometric stereo, we hope to not only generate accurate key models, but also capture reflectance information for high-quality 3D relighting applications.

The paper is organized as follows: Section 2 reviews related work in multiview and photometric stereo, and Section 3 gives an overview of our method. Next, we explain the three steps of the method: multi-illumination MVS (Section 4), normal recovery (Section 5), and normal-based geometry improvement (Section 6). We conclude with results and discussions in Sections 7 and 8, respectively.

2 Related Work

Our work draws ideas primarily from the fields of MVS and photometric stereo. Both of these methods aim at obtaining high-fidelity geometric models of complex 3D shapes from multiple photographs. Here, we describe prior art in each field individually and methods for combining both approaches.

2.1 Multiview Stereo

MVS has achieved great success in recent years. The relative reconstruction accuracies of the most advanced MVS methods are about 1/400 (0.5 mm for a 20-cm wide object) [6]. MVS techniques can be divided into multistage local approaches and global optimization approaches. Multistage local methods include depth map-based algorithms [11], [12], [13] and feature growing approaches [14], while global methods include surface volume extraction [7], [8], [9] and surface evolution [10]. According to the Middlebury multi-view stereo evaluation web cite [15], multistage local MVS algorithms generally outperform global optimization methods. The higher accuracy of these methods are due to advances in image matching and point-based graphics techniques.

Bradley et al. [4] use multiview stereo to create a temporally consistent parameterization of the geometry of deforming garments. Popa et al. [5] capture smooth geometry of garments using MVS, and then wrinkle the surface based on the fold estimates to obtain a realistic looking virtual garment modeling. Although traditional MVS is designed for reconstructing Lambertian objects, some researchers focus on explicitly handling specular surfaces [16]. Jin et al. model unknown, fixed illumination using ambient and distant point lights to improve multiview reconstruction for Lambertian objects [17]. Due to the limited reconstruction cues, recovering high-frequency detail for 3D surfaces using fixed illumination is still challenging for MVS algorithms. The space-time stereo methods of Zhang et al. [29] and Davis et al. [30] use stereo images of objects under multiple illuminations for more reliable feature matching.

2.2 Photometric Stereo

Photometric stereo recovers surface orientations using images from a single viewpoint taken under different illumination. For Lambertian materials, known illumination from three nonplanar lighting directions suffices to recover surface orientations [53]. Uncalibrated photometric stereo methods estimate both the surface orientation and lighting, but suffer from a generalized bas-relief (GBR) ambiguity [24]. For single point light sources, approaches to resolving this ambiguity include iteratively estimating light positions [27], [28], or using a rank constraint on the observation matrix [23].

To model more general illumination conditions, Basri and Jacobs [19], and Ramamoorthi and Hanrahan [20] treat light reflection as a convolution and observe 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. [26] further analyze the accuracy of spherical harmonics approximation for far and near illumination. Basri et al. [21] use a spherical harmonic illumination model and a factorization approach to recover surface orientation given many images take under unknown, distant illumination. Chen and Chen [22] use an iterative approach that requires only four captured images. Ambiguities still exist for both of these algorithms, requiring the normals for key pixels to be set beforehand. Our work is motivated by the success of spherical harmonic representations for general illumination and focuses on combining it with multiview stereo. We build on the work of Basri et al. [21] and Chen and Chen [22] by incorporating multiple views and by explicitly accounting for outliers, e.g., cast shadows.

2.3 Hybrid Techniques for Image-Based 3D Modeling

Even with relatively accurate surface orientations, reliably computing surface shapes for orientation alone is still challenging. A major thrust of research has been to combine multiview and photometric stereo. Some of this work focuses on adding point light sources to traditional stereo [31], [32], [33], [38], yielding higher quality but only partial reconstructions. Other recent work combines MVS with photometric stereo. For objects with piece-wise smooth surfaces, Weber et al. [34] simultaneously estimate geometry and reflectance. The multiview photometric stereo method of Hernandez et al. [36], [37] uses RANSAC to estimate the positions of point light sources and then reconstruct Lambertian objects. For calibrated light sources, Birkbeck et al. [39] employ a variational method to evolve the surface and handle specular reflections using a Phong reflectance model. The marker-less human motion capture system of Theobalt et al. [40] estimates surface reflectance and time-varying normal fields of moving people using a smooth shape template. Ahmed et al. [41] employ calibrated illumination and multiview video to capture normal fields and improve the geometry templates. Vlasic et al. [3] develop a system for high-resolution capture of moving objects using multiview photometric stereo. They use carefully designed illumination to estimate the surface normals. All of these algorithms require either accurate light-source calibration or careful illumination design. In contrast, our method imposes none of these constraints and explicitly takes outliers into account.

In the context of shape recovery using spherical harmonic representations for general illumination, Simakov et al. [25] propose a correspondence metric for stereo matching under unknown illuminations. Because the GBR ambiguity prevents estimation of the surface normals,

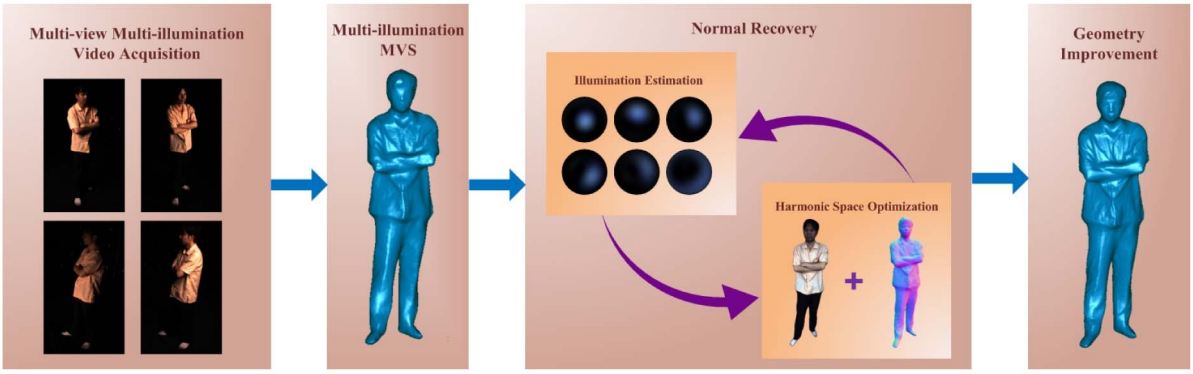

Fig. 1. **Overview of the method.** We first capture multiview, multi-illumination video. An initial model is computed using a multi-illumination, multiview stereo method. After that, the surface orientations are recovered by alternatingly estimating the surface normals and the illumination. Finally, the computed normals are used to refine the initial geometry.

textureless planar regions can not be resolved by this algorithm. By contrast, our method explicitly recovers surface normals in order to refine the geometry, enabling good performance for textured as well as textureless regions. Moreover, while Simakov et al. assume purely Lambertian objects, our method can handle objects with some non-Lambertian surface regions.

Other hybrid approaches acquiring high-quality geometry and reflectance, include that of Lensch et al. [43], who apply structured lights and photometric stereo to estimate spatial appearance and geometric detail. Ma et al. [42] use polarized light to obtain diffuse and specular normals, which are used to improve the geometry obtained from triangulation structured lights. These methods require special equipment to capture images under precisely specified lighting, and scale poorly to large environments because of the structured lights.

3 Overview

Fig. 1 illustrates the workflow of our reconstruction approach. The input of our method consists of multiview image sequences recorded under different illuminations. Unlike previous methods, our method imposes no strict constraints on the illuminations. We use the multiview multi-illumination image sequences to compute an initial 3D model using a multi-illumination MVS algorithm (Section 4). Based on this initial 3D model, we estimate surface normals and a spherical harmonic representation for the unknown illumination (Section 5). As the low-dimensional subspace approximation is valid only for Lambertian reflectance, we introduce a robust optimization method based on the ℓ_1 metric to estimate normals in the presence of non-Lambertian reflectance, cast shadows, and interreflections. Normals estimated using photometric stereo are generally accurate for high-frequency detail but may contain low-frequency errors. We use these normals to refine the 3D model using a variant of the method of Nehab et al. [35] (Section 6).

4 Multi-Illumination MVS

The space-time stereo methods of Zhang et al. [29] and Davis et al. [30] both consider space-time matching windows for stereo imagery of a fixed object under varying illumination. Davis et al. performed experiments changing illumination using a laser pointer and shadows cast by a hand in front of a fixed light source. They found that the best stereo matching results were obtained with a purely temporal matching vector: a window with a 1×1 pixel spatial extent and a temporal extent that includes all input frames. They did not analyze this result in detail, other than to say they believe it to be true in general for static scene geometry and variable illumination. In this work, we also use a purely temporal matching vector. Here, we pause briefly to analyze the significance of the purely temporal matching vector for photometric stereo.

4.1 Multi-Illumination Radiance Vector

We define the multi-illumination radiance vector as a vector containing the radiances under multiple illuminations for a surface point. Assuming orthographic projection under distant illumination, and no shadows or interreflections, the kth element in the multi-illumination radiance vector at a surface point p can be described as

$$I_p^k = \eta_k f_p(\theta_k, \phi_k; \theta_v, \phi_v), \tag{1}$$

where η_k is the intensity of the kth illumination, and the function f_p is the bidirectional reflectance distribution function (BRDF) of the surface at point p, specifying the reflectance as a function of the incidence direction (θ_k, ϕ_k) and the reflection direction (θ_v, ϕ_v).

For view-independent BRDFs (e.g., Lambertian materials), the reflection direction parameters (θ_v, ϕ_v) can be omitted. The incidence angle (θ_k, ϕ_k) can be also computed from the surface normal and the incident light direction, so (1) can be rewritten as

$$I_p^k = \eta_k f_p(\boldsymbol{n}_p, \boldsymbol{l}_k), \tag{2}$$

where $\boldsymbol{n}_p$ and $\boldsymbol{l}_k$ are the surface normal and the incident direction of the kth illumination, respectively. Therefore, the reflected radiance under the kth illumination at point p is determined by the reflectance function f_p and the surface normal $\boldsymbol{n}_p$.

Let $\boldsymbol{I}_p$ and $\boldsymbol{I}_q$ be the multi-illumination radiance vectors for two points p and q. If the number of the illumination conditions is large enough, then $\boldsymbol{I}_p = \boldsymbol{I}_q$ implies $\boldsymbol{n}_p = \boldsymbol{n}_q$ and

$f_p = f_q$. As an example, the Lambertian BRDF for gray scale images with fixed illumination has three degrees of freedom: one for the surface albedo, and two for the surface normal. Therefore, for Lambertian objects, three nonplanar illumination directions suffice to discriminate points with different normals or albedos. For multiview photometric stereo, even without casting high-frequency information into the scene (as in the space-time stereo work), a matching window using a multiradiance vector is more discriminative precisely because it encodes information about the surface orientation as well as the albedo.

4.2 Multi-Illumination Photoconsistency Metric

Although the space-time stereo work used a sum of squared distances (SSD) matching error, we can obtain better results with a more robust error measure. We choose to measure the angle between the multi-illumination radiance vectors in homogeneous coordinates to account for differences in both direction and magnitude of the original vectors. To define the new photoconsistency metric, the multi-illumination radiance vector is extended into homogeneous coordinates by appending a predetermined constant c, giving $\boldsymbol{I}' = (\boldsymbol{I}, c)$. The distance between two multi-illumination radiance vectors is measured using the angular difference between the extended vectors

$$d(\boldsymbol{I}_s, \boldsymbol{I}_t) = \frac{\boldsymbol{I}'_s \cdot \boldsymbol{I}'_t}{|\boldsymbol{I}'_s||\boldsymbol{I}'_t|} = \cos\theta, \tag{3}$$

where θ is the angle between the extended vectors $\boldsymbol{I}'_s$ and $\boldsymbol{I}'_t$. Incorporating a spatial neighborhood constraint, the multi-illumination photoconsistency metric is defined as

$$E(s,t) = \sum_{x \in W_s, y \in W_t} d(\boldsymbol{I}_x, \boldsymbol{I}_y), \tag{4}$$

where W_s and W_t are the spatial neighborhood window around pixels s and t. Our experience has shown that for nine or more different illumination conditions, a 1×1 spatial window works best. For fewer conditions, a larger spatial window should be used to increase robustness.

4.3 Creating and Merging Multiview Depth Maps

We create a watertight mesh model using the multiview MVS algorithm presented in [12], modified to take advantage of the multiple illumination conditions. Depth maps from each view are created using the photoconsistency metric just described, then merged to create a single point cloud. This cloud is down-sampled and refined using an error and conflict cleaning procedure, then converted to a mesh using Poisson surface reconstruction [44]. Vertices that lie outside the visual hull are projected back to the visual hull along their inward normal directions. See Liu et al. [12] for more details about merging multiview depth maps to obtain a watertight 3D mesh.

Fig. 2 shows example results from the multi-illumination multiview stereo step for our system. 29 illumination conditions were used with a 1×1 spatial matching window. Our proposed angular multi-illumination photoconsistency measure produces results with much better detail than an SSD multi-illumination photoconsistency measure. Reconstructions such as the one in this figure are used as an initial 3D model for our method. In the next section, we describe how we estimate the lighting and surface normals to further improve the mesh quality.

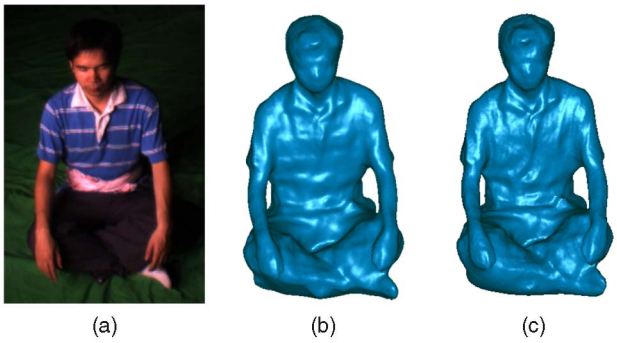

Fig. 2. Experimental comparison of the effects of SSD and angular photoconsistency metric on MVS reconstruction. (a) shows one of the multiview multi-illumination images. The surface computed using SSD multi-illumination photoconsistency (b) shows much less detail than the result using our proposed angular multi-illumination photoconsistency measure (c).

5 Normal Recovery

We present a new multiview photometric stereo algorithm to recover surface orientations under general unknown illuminations. We use a coarse to fine strategy, with coarse normals initialized using the mesh produced by our MVS algorithm. These estimated normals are refined by minimizing the shading errors under multiple illuminations. Because the illumination is also unknown, we must iteratively estimate both normals and lighting.

This section is organized as follows: we briefly review the spherical harmonic representation for general illumination in Section 5.1. In Section 5.2, we describe our alternating constrained reweighted least absolute values (ACRLAV) method for iteratively estimating illumination and normals. ACRLAV assumes Lambertian reflectance but uses an ℓ_1 metric, so surface normals can often be accurately estimated even in the presence of specularities or cast shadows. However, some erroneous estimates for surface normals in non-Lambertian regions could corrupt the geometry in the next step. Section 5.3 explains how we detect and correct erroneous orientation estimates using a shape prior from the MVS algorithm.

5.1 Spherical Harmonic Illumination Representation

For Lambertian reflectance and convex objects, irradiance can be approximated well by convolving a clamped cosine with a low-order spherical harmonic representation of incident illumination [19], [20]. Computing radiance using second-order spherical harmonic representations for distant, isotropic illumination accurately models more than 98 percent of the reflected light. Using a first order approximation, the accuracy still exceeds 75 percent. Moreover, empirical evidence shows that this approximation remains valid even for fairly near illumination [26].

We use the second-order approximation for greater accuracy. Our MVS algorithm ensures that the projected size of the facets is roughly one pixel in all views, and we

use a single orientation and reflectance for each facet. The reflected radiance at a mesh facet is described by

$$I_i = \boldsymbol{l}_9^T \boldsymbol{S}_9, \tag{5}$$

where $\boldsymbol{l}_9$ is the 9D vector of spherical harmonic coefficients describing the illumination, and $\boldsymbol{S}_9$ contains the irradiance of that facet when illuminated by the nine spherical harmonic lighting bases.

To estimate the surface normal on an object's surface, we first project multiview images onto the 3D model to obtain the radiance of each facet for each captured illumination condition. Similar to [47], the reflected radiance at each facet is the weighted average of multiview radiances. The weight is defined as

$$\omega_i = \frac{1}{(\max_i(1/\Theta_i) + 1 - 1/\Theta_i)^\alpha}, \tag{6}$$

where ω_i is the weight assigned to the camera view i and Θ_i is the angle between the facet normal and the viewing vector toward the camera i. The weights are normalized to sum to 1. The sharpness value α controls the degree to which the largest weight is exaggerated.

Using the second-order spherical harmonic lighting representation, the matrix I of measured surface radiances for each illumination is

$$I_{N\times M} = L_{N\times 9} S_{9\times M}. \tag{7}$$

Each row of I is the radiance on the object's surface corresponding to one of the illumination conditions. N is the number of illuminations, and M is the number of facets. Each row of L is the second-order spherical harmonic coefficients for a certain illumination. Each row of S is a harmonic image—the radiance on the object's surface corresponding to illumination by one of the spherical harmonic basis functions. For distant illumination and convex Lambertian objects, the nine harmonic images have these forms:

$$\begin{aligned} &s_1 = \rho, \quad s_2 = \rho n_x, \quad s_3 = \rho n_y, \\ &s_4 = \rho n_z, \quad s_5 = \rho\left(3n_z^2 - 1\right), \quad s_6 = \rho n_x n_y, \\ &s_7 = \rho n_x n_z, \quad s_8 = \rho n_y n_z, \quad s_9 = \rho\left(n_x^2 - n_y^2\right), \end{aligned} \tag{8}$$

where $[n_x, n_y, n_z]^T$ are the coordinates of the normal and ρ is the albedo.

These harmonic images form a 9D linear space called the 9D *harmonic space*. This subspace is described fully by the reflectance and the normal. Therefore, if we can recover the harmonic space S from the observation matrix I, the surface normal and the reflectance can be recovered as well.

5.2 Alternating Constrained Reweighted Least Absolute Values

Matrix factorization is one approach to obtaining the harmonic space from the reflected radiance. Traditionally, one would use the SVD to extract the 9D subspace. This method, however, can only recover the subspace up to a linear ambiguity. Moreover, the SVD is optimal only when the additional noise is Gaussian. The low-order spherical harmonic approximation is only valid for Lambertian surfaces with no cast shadows or interreflections. Violations of these assumptions cause outliers in the data, which we will call outlier reflections. SVD solutions can be arbitrarily far from the correct answer in the presence of these outliers.

Handling outlier reflections requires a new matrix factorization method. We introduce an ACRLAV approach for estimating the normals robustly. ACRLAV, inspired by the nonparametric material representation in [48], is a robust version of the alternating constrained least squares (ACLS) algorithm. ACRLAV retains the benefits of ACLS, such as its flexibility for including additional constraints, and extends it by robustly handling outliers using an ℓ_1 minimization. Recent research demonstrates that the linear equation can be solved exactly when the additional error is sparse [49]. Section 5.2.1 describes how we use this to estimate the illumination conditions.

Our matrix factorization algorithm uses a two-step alternating optimization on (7). The first step optimizes the illumination L while holding the 9D harmonic space S fixed; the second step optimizes the harmonic space S while holding the illumination L fixed. In practice, two to four iterations are sufficient to significantly improve the estimated normals. The following sections describe the illumination estimation and the harmonic space optimization in detail.

5.2.1 Illumination Estimation

Normals initialized from the coarse MVS geometric model give a roughly correct approximation to the harmonic space. We compute the second-order spherical harmonic coefficients for each illumination by solving the following linear equation:

$$\boldsymbol{I}_i = \boldsymbol{l}_i^T S + \boldsymbol{e}. \tag{9}$$

Here, S is the harmonic space, $\boldsymbol{I}_i$ represents the ith row of the observation matrix I, $\boldsymbol{l}_i$ is the unknown illumination, and $\boldsymbol{e}$ is an error.

Conventionally, this equation would be solved in the least squares sense, assuming that the noise $\boldsymbol{e}$ is gaussian. Given outlier reflections, ℓ_2 minimization can lead to a solution arbitrarily far from the correct answer. ℓ_1 minimization is more robust in the presence of outliers. Moreover, Candes and Tao [49] show that when e is sparse but arbitrary, the illumination $\boldsymbol{l}_i$ can be recovered exactly by solving the following ℓ_1 minimization problem:

$$\min_{\boldsymbol{l}_i \in R^9} \| \boldsymbol{I}_i - \boldsymbol{l}_i^T S \|_1 . \tag{10}$$

The globally optimal solution of (10) can be found using linear programming [46]. We assume that the outlier reflections only account for a small portion of the whole surface, so the error $\boldsymbol{e}$ in (9) should be sparse. Therefore, the illumination conditions can be robustly estimated by linear programming.

After obtaining a good initial estimation of the illumination, we calculate the weight of each facet and use weighted least absolute values to reestimate the illumination coefficients. The weighted ℓ_1 minimization problem is

$$\min_{\boldsymbol{l}_i \in R^9} \| \left(\boldsymbol{I}_i - \boldsymbol{l}_i^T S\right) W \|_1 . \tag{11}$$

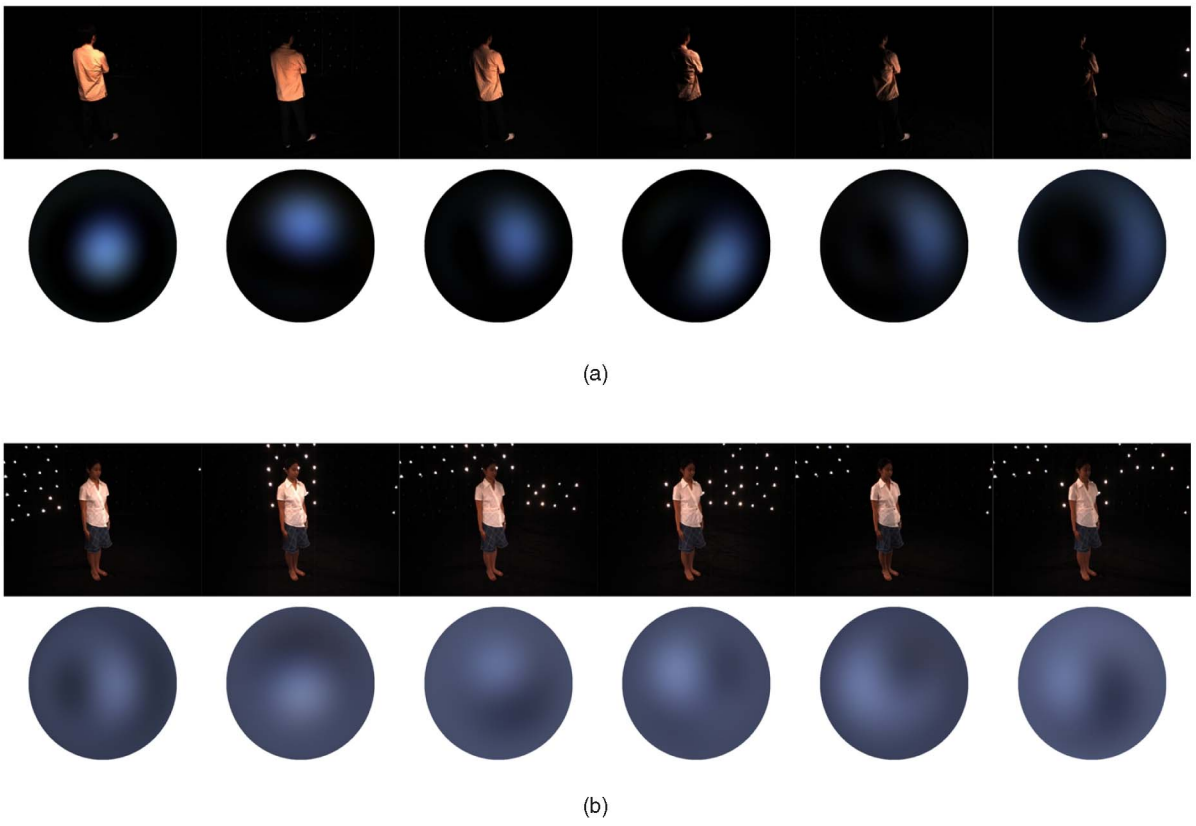

Fig. 3. **Estimating illumination.** Here we show examples of actors imaged under different lighting conditions and the results of our lighting estimation. We visualize the computed lighting by using it to render a uniform Lambertian sphere, then remapping the surface of the sphere to 2D polar coordinates relative to a zenith pointing out of the page. The top row used directional illumination, and the bottom used multiplexed illumination.

Here, W is a diagonal matrix with each diagonal element equal to

$$w_{jj} = \frac{1}{\gamma} exp\left(-\frac{(I_{ij} - \boldsymbol{l}_i^T \boldsymbol{S}_j)^2}{2\sigma^2}\right), \tag{12}$$

where γ is a normalization constant such that $\sum_j w_{jj} = 1$, and $\boldsymbol{S}_j$ is the jth column of the matrix S. The value σ is estimated using median absolute deviation (MAD) [45], defined as

$$\sigma = 1.4826\ median_j \parallel I_{ij} - \boldsymbol{l}_i^T \boldsymbol{S}_j \parallel_1 . \tag{13}$$

We solve (11) using linear programming.

To review, we obtain an initial estimate of the lighting by solving (10), then use a weighted least absolute values algorithm to refine the initial estimate by solving (11). Equation (10) can be seen as a weighted least absolute values problem with uniform weights, so our illumination acquisition approach can be viewed as a two-step reweighted least absolute values approach. In this way, we robustly and accurately estimate each illumination. Fig. 3 shows examples of illumination conditions estimated using this method.

5.2.2 Harmonic Space Optimization

Given the estimated illumination conditions, the normal and albedo for each facet can be computed from the following equation:

$$\boldsymbol{I}_j = L\boldsymbol{s}_j, \tag{14}$$

where $\boldsymbol{I}_j$ is a column vector containing the multi-illumination radiance of each facet, L is the estimated illuminations, and $\boldsymbol{s}_j$ is the radiance of the facet under the nine spherical harmonic illumination bases. We would prefer to use an ℓ_1 minimization to solve this equation robustly in the presence of outlier reflections due to cast shadows and interreflections. The quantities $\boldsymbol{s}_j$, however, are not linear with respect to the normal and the albedo (8), so the ℓ_1 norm minimization problem cannot be recast into an efficient linear programming problem. Instead, we use a reweighted ℓ_2 minimization. Similar to the illumination estimation approach, we formulate a generalized weighted least squares problem as

$$\begin{aligned} &\min_{\boldsymbol{n},\rho} \parallel (\boldsymbol{I}_j - L\boldsymbol{s}_j)W \parallel_2^2 \\ &subject\ to \parallel \boldsymbol{n} \parallel_2 = 1, \end{aligned} \tag{15}$$

where $\boldsymbol{n}$ is the normal, ρ is the albedo, $\boldsymbol{s}_j$ has the form in (8), and W is a diagonal matrix defining the weight of each illumination.

We solve this constrained nonlinear least squares problem using the Levenberg-Marquardt algorithm to efficiently acquire the normal and the albedo for each facet. We use normals computed from the multi-illumination MVS geometry to initialize the optimization of (15). Similar to the illumination estimation, a two-stage reweighted strategy is applied to estimate the normal. In the first stage, an initial estimate of the normal and the albedo is obtained

by setting the diagonal of W to be all ones. Then, the new weight for each illumination is recomputed using (12). Equation (15) is optimized again using the new weights to derive more accurate normals and albedos.

5.3 Shape-Prior-Based Outlier Detection and Correction

The spherical harmonic illumination estimate is quite robust due to the ℓ_1 metric, but the computed normals and reflectances are more susceptible to errors. Our method does suppress effects of cast shadows and interreflections on Lambertian surfaces, but for very non-Lambertian regions, the low-order spherical harmonic approximation is invalid. Trying to refine the object geometry using erroneous orientation estimates in these regions will corrupt rather than improve our results.

We use the MVS geometry as a prior to detect and correct erroneous orientation estimates. Although the fine detail may be suspect, the low-frequency shape from the multi-illumination MVS should be quite accurate. The low-frequency orientation calculated by differentiating the geometry (we call this the "geometric normal") should also be accurate. We assume that the low-frequency component of the surface orientations recovered by ACRLAV (the "photometric normals") should match the low-frequency component of the geometric normals. To filter out erroneous normal estimates, we define the low-frequency orientation consistency as the angle between the low-frequency components of the geometric and photometric normals normal:

$$\delta_i = \arccos\left(\phi\left(\boldsymbol{n}_i^g\right)^\top \phi\left(\boldsymbol{n}_i^p\right)\right), \tag{16}$$

where $\boldsymbol{n}_i^g$ and $\boldsymbol{n}_i^p$ represent the geometric and photometric normal, respectively, and $\phi(\cdot)$ is a function that extracts the low frequency of the normal (we use a Gaussian filter).

We assume that this angle follows a Gaussian distribution with the standard deviation computed median absolute deviation, and consider estimated normals to be outliers if their low-frequency consistency δ_i is β times larger than the standard deviation. To obtain the final surface normal, we fuse the photometric normal with the geometric normal according to the following metric:

$$\boldsymbol{n}_i = \begin{cases} \boldsymbol{n}_i^p, & \delta_i \leq \beta\, \sigma_a, \\ \boldsymbol{n}_i^g, & \delta_i > \beta\, \sigma_a, \end{cases} \tag{17}$$

where σ_a is the standard deviation of δ_i. Fig. 4 gives an example of detecting and correcting outliers based on the shape prior.

6 Normal-Based Geometry Improvement

The high-frequency content of the photometric normal is generally much more accurate than that of the geometric normal from the original MVS shape estimate. Nehab et al. [35] propose an elegant method to fuse normals and position to improve the geometry. We use a variant of their full-mesh optimization approach to refine our MVS geometry. In their algorithm, each vertex is assigned a normal, and the polygon formed by the vertex's neighbors is used as an approximation of its tangent space when imposing a surface normal constraint. By contrast, we assign a normal to each facet of the mesh and optimize the geometry by enforcing the normal constraint on each facet.

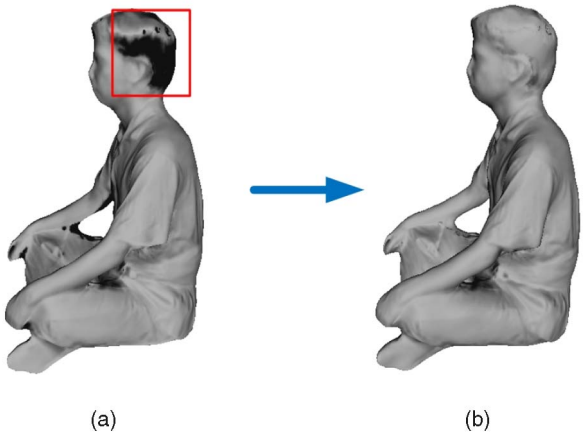

Fig. 4. **Outlier detection and correction using a shape-based prior**. (a) An actor rendered with uniform albedo and normals computed using ACRLAV. The estimated surface orientation for much of the hair is incorrect. We detect these outliers based on the low-frequency prior of the MVS geometry, and replace them with normals from the MVS shape. (b) The result after outlier correction.

To refine the geometry, we optimize an error function consisting of a position error E^p and a surface normal error E^n:

$$E = \lambda E^p + (1-\lambda)E^n. \tag{18}$$

The parameter $\lambda \in [0,1]$ controls the relative influence of the positions and normals in the optimization. E^p is defined the same with [35]. E^n in our algorithm is defined on each facet and has the following form:

$$E^n = \sum_{u,w \in F_f} \left[\boldsymbol{n}_f \cdot (\boldsymbol{p}_u - \boldsymbol{p}_w)\right]^2, \tag{19}$$

where u and w are the vertices belonging to the facet F_f, and $\boldsymbol{n}_f$ is its normal. Based on the normals estimated in Section 5, we optimize the vertices' positions by minimizing the error function. In this way, high-frequency details can be added to the original MVS geometry. Fig. 5 shows an example of the geometry improvement using our approach.

7 Experimental Results

In this section, we test our algorithm on synthetic and real-world data. The real-world experiments allow a subjective evaluation of our algorithm, e.g., visual quality, rendering, and relighting quality. Due to the differing acquisition setups and input data, it is difficult to compare our method with other state-of-the-art approaches. Instead, we use synthetic data to validate our algorithm quantitatively against a ground truth model. The parameters used for all experiments are listed in Table 1.

7.1 Multicamera Multilight Dome System

Testing our algorithm on real-world data requires a multi-view acquisition system. As we impose no constraints on the illumination type or light position, we can use

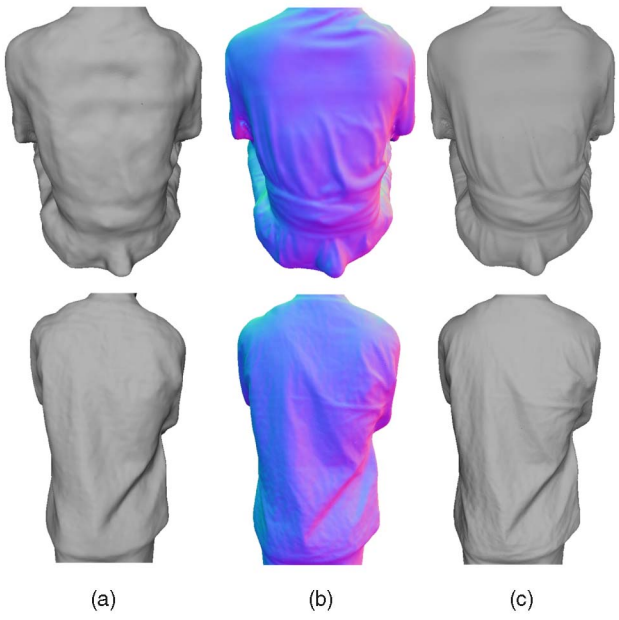

Fig. 5. **Refining geometry using the photometric normals.** (a) the initial mesh, (b) the estimated normal field, and (c) the improved mesh.

traditional multicamera systems with varying illuminations to capture the multicamera multi-illumination images. Liu et al. [12], [13] have established a multicamera dome system for free-viewpoint video capture. To test our algorithm under a variety of illuminations, we added hundreds of LEDs on the surface of the dome. The multicamera and multilight acquisition system is shown in Fig. 6. Our algorithm is not limited to this system and can be applied to any multicamera system with several light sources. This system can produce a wide range of illuminations by controlling the LEDs, allowing us to comprehensively test our algorithm.

For completeness, we briefly describe our acquisition platform. The schematic figure of our system is shown in Fig. 6. The dome system is a hemisphere 6 m in diameter, with 20 low-speed FLEA2 cameras (30 fps) mounted on a ring roughly at head height. 290 Luxeon K2 LEDs are spread

TABLE 1
Parameters Used for All Experiments

Parameters	Value	Section
Const c in the photo-consistency metric	25	4.2
Other MVS parameters	the same with [12]	4.3
Sharpness value α	2	5.1
Threshold β for outlier detection	2.5	5.3
Weighting parameter λ	0.1	6

evenly over the whole hemisphere. The camera resolution is set to 1024×768. Neighboring sets of 10 LEDs are configured as a single area light source to provide enough lumens to illuminate the scene. Custom control circuitry synchronously triggers the lights and camera. Individually calibrating 290 LEDs is nontrivial, making our alternative approach using uncalibrated illumination appealing.

We tested our algorithm using two types of illumination patterns. The first turns on each individual area light source one at a time. The dome system provides 29 different illuminations. We refer to this illumination pattern as *directional illumination*. The other illumination pattern uses *multiplexed illumination*, simultaneously activating several area lights according to a Hadamard sequence with 29 patterns [54].

7.2 Experiments on Real-World Data

The first experiment reconstructs an actor sitting with legs crossed. Some of the captured images are shown in Fig. 7a. Although the resolution of our camera is $1{,}024 \times 768$, to capture the entire actor from all views, the resolution of the bounding area of the actor had to be limited to 300×400. We used 29 different directional illuminations. The frame rate of the camera is 30 fps, so the multiview multi-illumination image sequences can be captured in only one second. It is possible that even during this short time interval, a cooperative actor may not be able to hold absolutely still. Empirically, we saw no ill effects from any small motion of the actor.

The initial 3D geometric model produced by our MVS algorithm is shown in Fig. 7b. The normals recovered after decomposing the observation matrix into illumination coefficients and the harmonic space are shown in Fig. 7c.

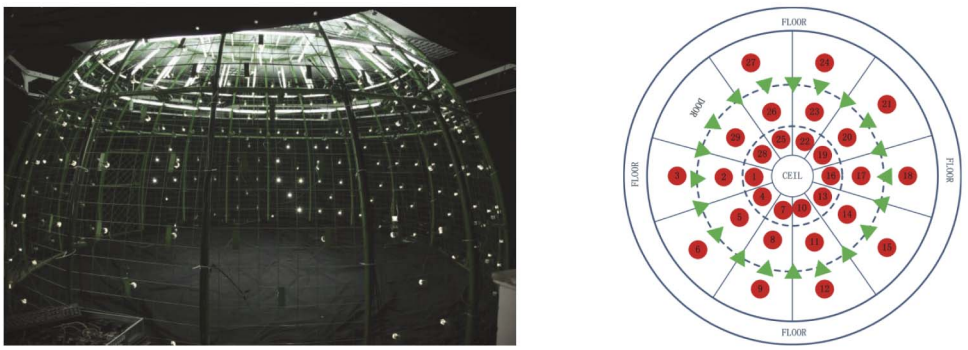

Fig. 6. **The multicamera, multilight dome.** Our capture system, shown on the left, uses 20 low-speed FLEA2 cameras mounted in a ring. The hemisphere is divided into 29 lighting areas, each comprised of 10 evenly spaced LEDs. The schematic on the right shows the distribution of the cameras (green triangles) and lights (red circles).

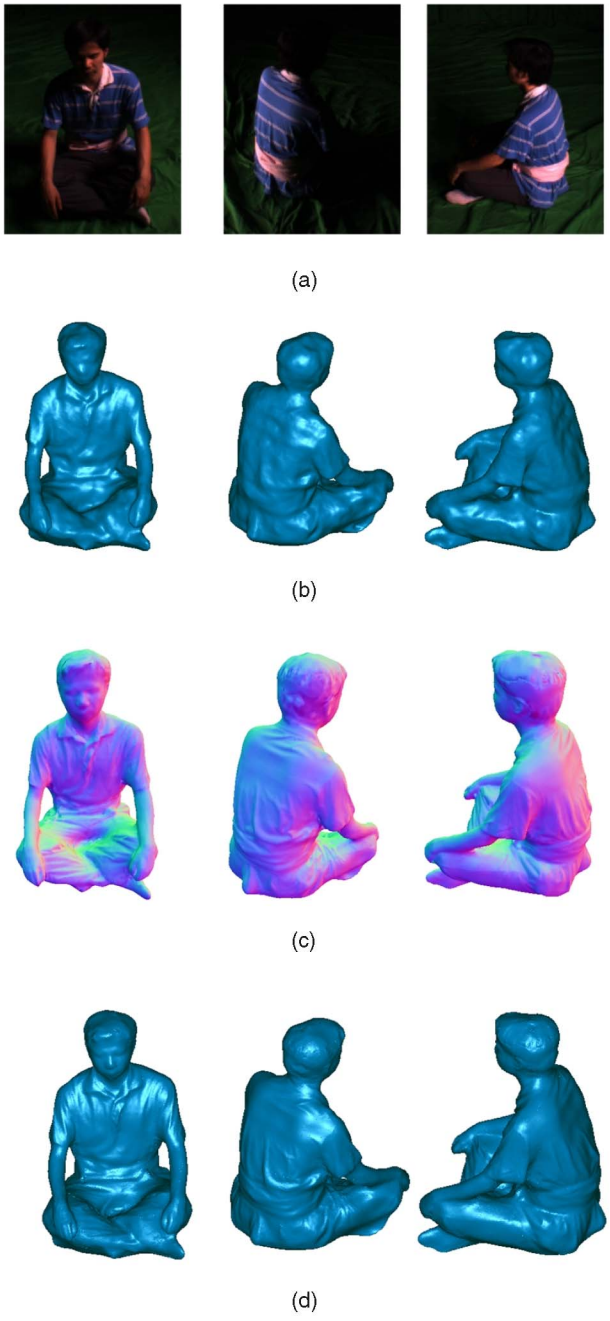

Fig. 7. **Reconstructing a sitting actor.** This actor is reconstructed using 20 views with 29 different illuminations. Some of the input images are shown in (a). (b) The multi-illumination MVS results. (c) The recovered normal field. (d) The geometry of the improved mesh.

The final 3D model refined using the derived photometric normals is shown in Fig. 7d. Relative to the multi-illumination MVS results, the final result does not exhibit surface noise yet captures fine details such as wrinkles.

Our second experiment tests our algorithm on extensive data sets including both static objects and quasistatic human actors. The captured images and the reconstructed models are shown in Fig. 8. We reconstructed three different static scenes. One is of the objects on a desk covered by a ruffled tablecloth, captured by 20 cameras under 29 directional illuminations. The other two are replicas of the statues "Dying Slave" and "Brutus". These statues have approximately uniform albedo without salient texture, which is challenging for traditional MVS algorithms. We reconstruct the two statues only using 12 cameras and 31 multiplexed illumination conditions. These results show that our method can give satisfactory reconstructions with only 12 cameras. The images with human actors demonstrate that we can reconstruct people with a variety of clothing and poses.

7.3 Ground Truth Evaluation Using Synthetic Data

We analyzed the performance of our method quantitatively using a 3D model of the stanford bunny [52] with uniform albedo and Lambertian reflectance, rendered from 20 viewpoints under 20 different illuminations. The generated images are of 800×600 pixel resolution. To create 20 different illuminations, we used 20 distant point light sources, illuminated one at a time. The renderings contain cast shadows, which are interpreted as outliers by our algorithm. We also added zero-mean Gaussian noise with standard deviation $\sigma = 0.01$ to each pixel to simulate sensor noise. Several of the synthetic images are shown in Fig. 9a. The surface normals reconstructed by our method are shown in Fig. 9b, and the final improved geometric model of the bunny is shown in Fig. 9c. For comparison, Fig. 9d shows the ground truth model.

We also compared the result of our algorithm to the ground truth surface by measuring the distance from each point on our model to the closest point in the ground truth model. Table 2 lists the mean, median, and standard deviation of the error. They are normalized by the longest diagonal of the bounding box volume, so the results in Table 2 mean that if the biggest diagonal of the bounding box is 1 m, the mean error is 0.85 mm, and the median error is only 0.65 mm. Traditional MVS methods cannot achieve this accuracy.

7.4 Reconstruction under Varying Numbers of Illuminations

Intuitively, the reconstruction quality of our method depends on the number and distribution of cameras and illuminations. We explored this relation by testing our algorithm using from three to 28 directional illuminations. We use directional instead of multiplexed illumination to facilitate analyzing the dependence of our results on the illumination conditions. Although we refined the geometry using varying numbers of illuminations, we used the same initial MVS geometry (obtained under 14 illuminations) for each trial to reduce the influence of the starting geometry model. The difference in reconstruction quality is determined solely by the normal recovery.

Fig. 10 shows the reconstruction results and the illumination distributions. The 3D model detail improves with more illuminations, but the 10-illumination case appears to be an outlier. In the illumination pattern shown in Fig. 10b, we see that in the 10-illumination case the different illuminations were generated by moving the lights in a 2D plane, while in the six-illumination case the light positions are not coplanar and are distributed better in 3D space. This particular six-illumination case leads to better results than the 10-illumination one, demonstrating that the illumination distribution is a key factor for the normal estimation and that the illumination position should be well distributed in 3D space. This experiment also shows that as few as six illuminations are

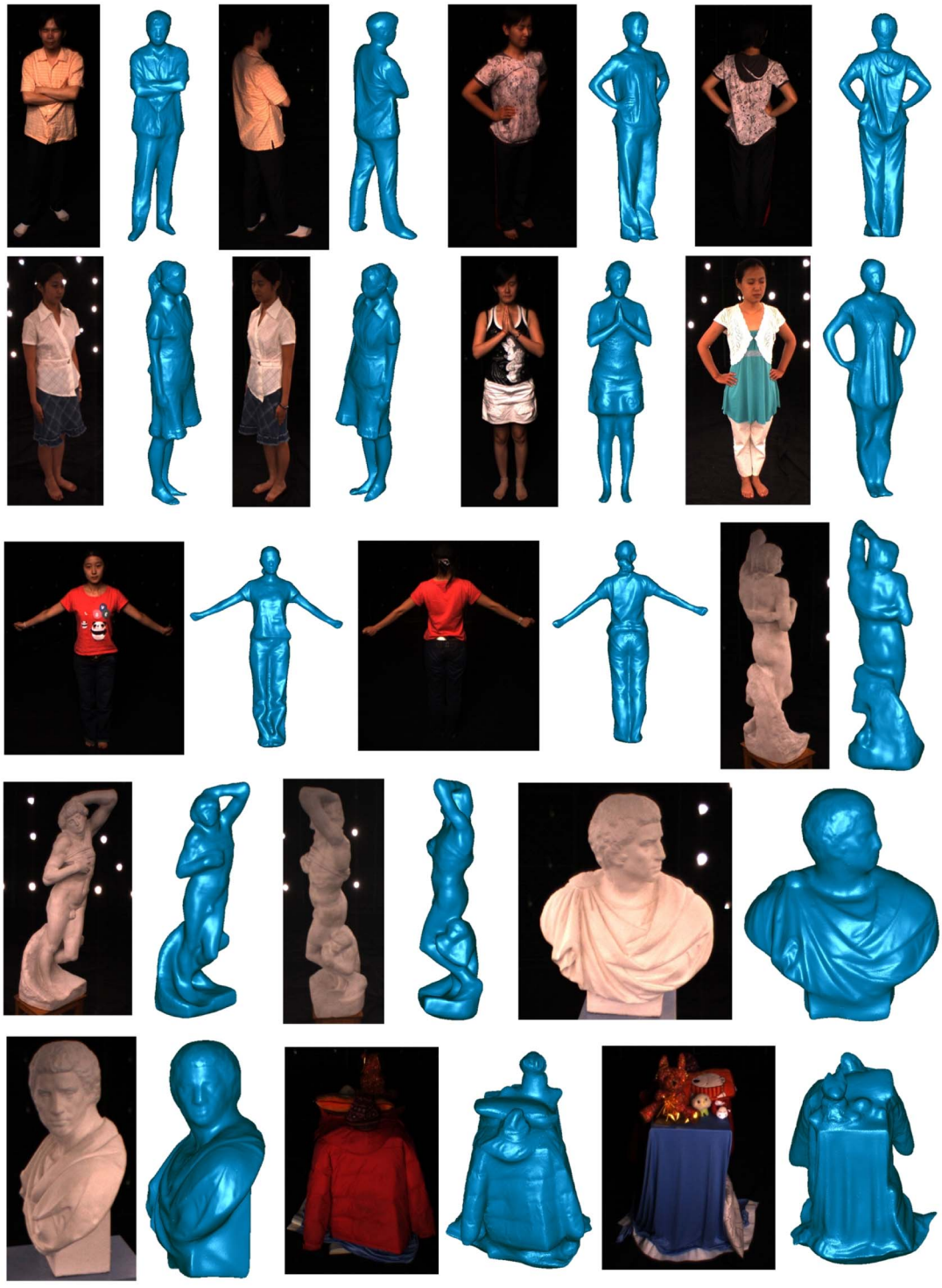

Fig. 8. **Results using our method for a variety of data sets.** These images show the quality of the models our system produces for a variety of people with different outfits and poses. We also show three reconstructions of completely static objects.

needed to obtain the normal field on the whole surface if the illumination distribution is adequate.

7.5 Reconstructions with Varying Numbers of Views

In this experiment, we investigated how the final reconstruction quality is affected by the number of camera views. We used 14 directional illuminations distributed as shown in Fig. 10b, and ran our algorithm using 8, 10, 12, 16, and 20 camera views. The results are shown in Fig. 11. We see that more views lead to a better initial multiview stereo reconstruction. We also see, however, that the final models are very similar, indicating that the photometric stereo refinement produces good quality results even with

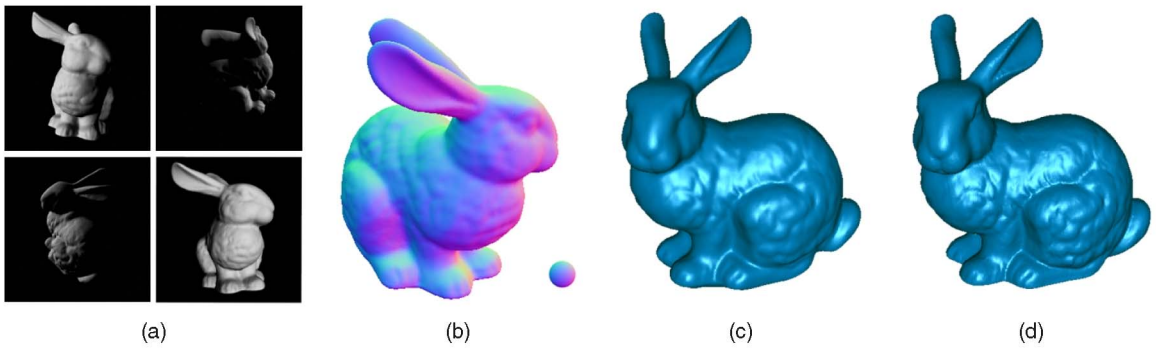

Fig. 9. **Evaluation using synthetic ground truth data.** The images on (a) show example multiview multi-illumination input images synthetically generated using a 3D computer model. (b) The estimated normal map for the bunny, along with a ground truth sphere for reference. The final reconstructed model (c) and the ground truth (d) show that our algorithm correctly reproduces much of the fine detail of the bunny.

a rough initial mesh. Combined with the experimental results in Section 7.4, this experiment leads us to conclude that as long as the initial mesh is roughly correct, the performance of our method is more strongly related to the number and distribution of the illuminations than the number of camera views.

7.6 Rendering and Relighting

We experimented with relighting one of the captured human 3D models shown in Fig. 8. We relight the 3D model with our computed albedo maps using the method of Ramamoorthi and Hanrahan [50]. The irradiance environment maps are represented using the first nine spherical harmonic illumination coefficients. For the illumination environment maps, we used a light probe of the Grace Cathedral, Eucalyptus Grove, and St. Peter's Basilica [51]. Fig. 12 shows the relighted 3D model rendered from different viewpoints. Although the method is only valid for diffuse objects, the relighting results in Fig. 12 are consistent with the illumination.

7.7 Complexity

The complexity of our algorithm depends on the number of cameras and illuminations. Using 12 cameras and 15 illuminations, and running on a 2.66 GHz PC with 2 GB of RAM, typical execution times of the steps in the pipeline are 8 minutes for the multi-illumination MVS, 30 minutes to run three iterations of the surface normal estimation, and 5 minutes for the geometry refinement. The code has not been optimized or parallelized, so this execution time could be significantly reduced.

7.8 Limitations

The low-order spherical harmonic approximation is only valid for Lambertian objects with no cast shadows. Thus, our method fails when the object's surface is totally non-Lambertian, when some parts of the object are constantly in cast shadows, and in deep concavities with strong interreflections. Fig. 13 illustrates the influence of cast shadows. There is a deep groove in the statue marked by a

TABLE 2
Quantitative Results of Synthetic Data

Accuracy[%]		
mean	med	std
0.085	0.065	0.072

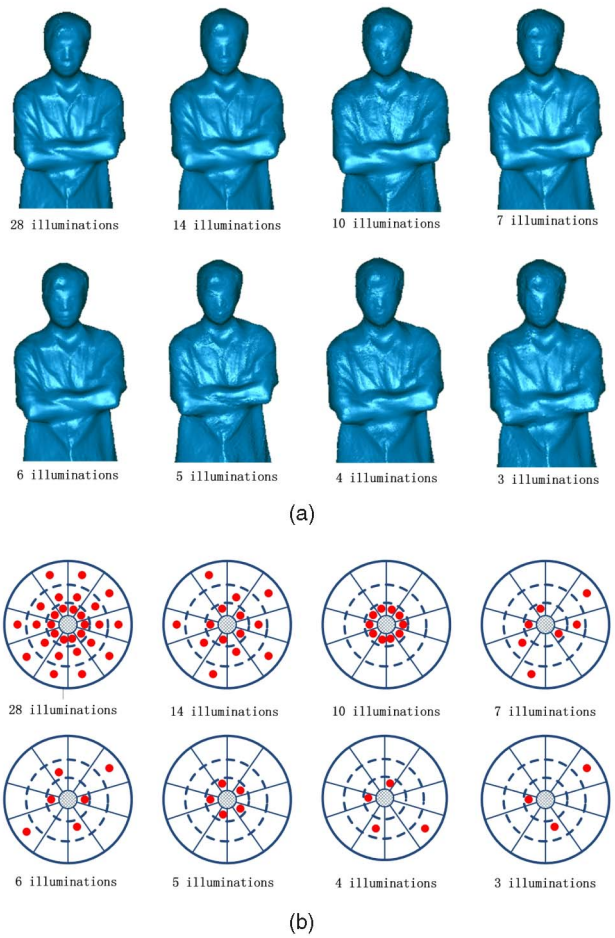

Fig. 10. **Reconstructions with varying numbers of illuminations.** (a) The reconstruction results using different numbers of illuminations (from 28 illuminations to three illuminations). (b) The illumination distributions in the dome system used for the different reconstructions. As in Fig. 6, red dots denote the illumination areas we used.

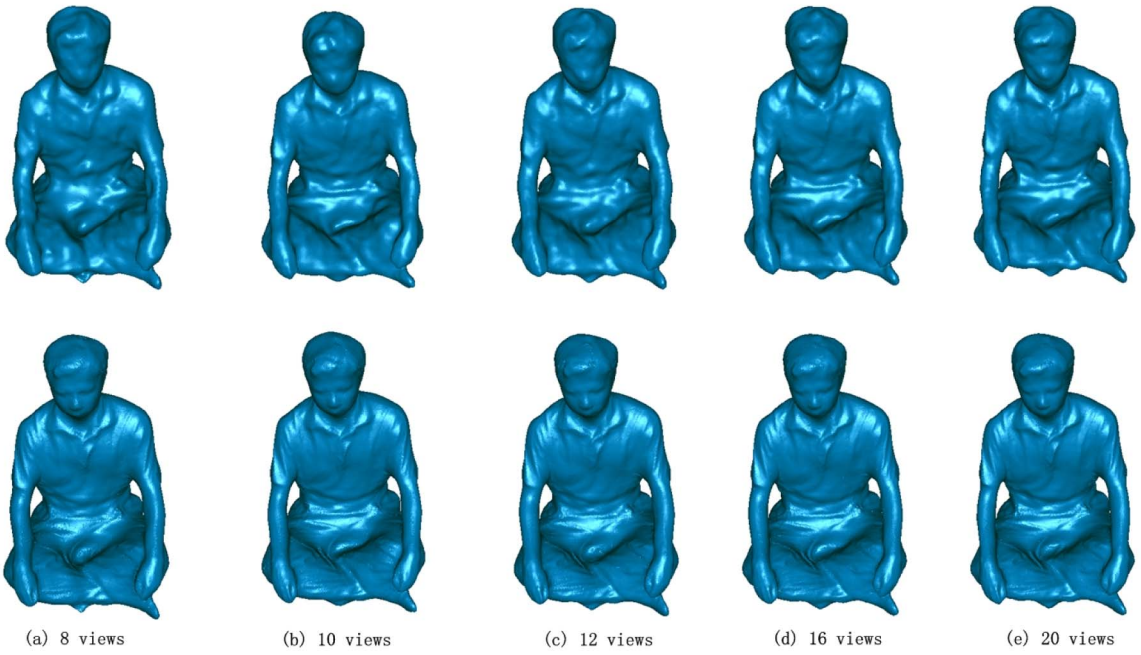

Fig. 11. **Reconstructions with varying numbers of camera views.** The first row gives the initial multi-illumination MVS models produced using different numbers of camera views. The second row shows the final models after refinement using photometric stereo. Although the quality of the initial model is poor for small numbers of views, the refined models are all very similar, implying that the method only requires a roughly accurate model to produce good results.

red rectangle in Fig. 13a. This groove is shadowed for most of the illumination conditions, and as shown in Fig. 13b, our algorithm does not accurately reconstruct the shadowed geometry.

Another limitation of our method is the reliance of the photometric stereo normal estimation on the initial MVS mesh. Although using the multi-illumination radiance vector leads to better MVS results compared to MVS with a single illumination condition, in the worst case where our system fails to obtain a reasonable initial mesh, the photometric stereo and geometry refinement will also fail. This may happen due to specularities or a limited number of cameras. Fig. 14 shows an example result applying our method to a synthetic non-Lambertian object. Due to

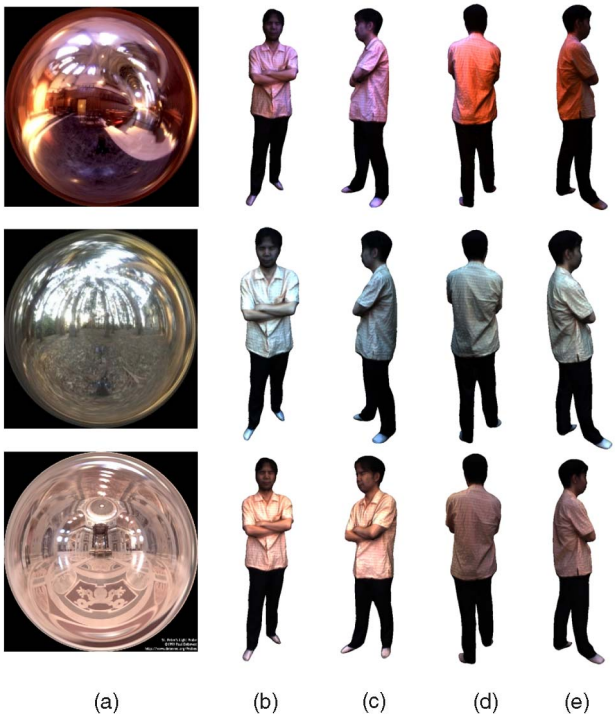

Fig. 12. **Relighting a human actor.** We represent the irradiance environment maps using the first nine spherical harmonic coefficients of the illumination as in [50]. The environment maps are shown in the first column of the figure, and the relighted models rendered from different viewpoints are shown on the right.

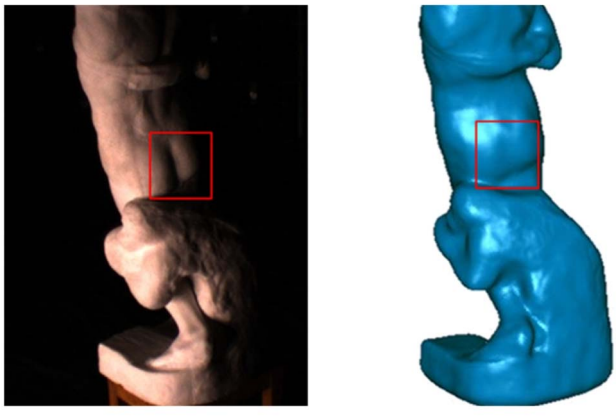

Fig. 13. **The influence of cast shadows.** (a) A deep groove, outlined in red. Due to the cast shadows, our method was unable to recover the geometry in this area, as shown in (b).

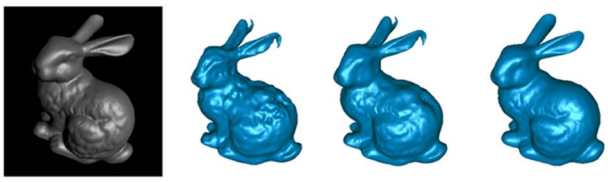

Fig. 14. **Experiment using a non-Lambertian synthetic object.** An example rendered input image is shown in (a). Because the initial MVS mesh (b) is inaccurate, the improved mesh (c) is not as good as the reconstruction (d) from Section 7.3.

specularities, the multi-illumination MVS computed a fairly inaccurate initial mesh. Because the initial mesh is corrupted, the model refined using photometric normals also shows little improvement.

8 Conclusion and Future Work

We present a 3D modeling algorithm to reconstruct objects filmed from multiple views, with each view captured under multiple unknown illuminations. The unique characteristic of our algorithm is that we assume neither calibrated illumination nor single point light sources. Instead, we use spherical harmonics to model general lighting conditions. We also present ACRLAV, an algorithm for iteratively estimating the illumination and surface normals. ACRLAV robustly computes the illumination conditions despite regions of non-Lambertian reflectance, cast shadows, and interreflections. It steadily refines surface normals using the estimated illumination, then the illumination using the improved normals. Experiments on a wide range of data sets demonstrate the ability of the algorithm to capture detailed geometry.

In the future, we plan to extend this method to performance capture. Aguiar et al. [1] deformed static laser-scanned key models to frames of a captured performance. We hope to use our system to capture these key models, removing the need for a laser scanner. We would also like to directly extend our algorithm to reconstructing dynamic scenes. Vlasic et al. recently demonstrated a system with high-speed cameras that film dynamic scenes with a different controlled illumination for each frame, then use optical flow to warp the images to key frames captured under uniform illumination. We could use similar techniques to apply our method to dynamic environments. The experiments in Sections 7.4 and 7.5 show that 12 cameras and 15 illumination conditions are enough to produce good quality reconstructions, and the execution time is tractable for performance capture.

Acknowledgments

This work is supported by the National Basic Research Project of China (973 Program), No. 2010CB731800, the Programs of National Science Foundation of China (NSFC), No. 60933006 and No. 61073072, and the National High Technology Research and Development Program of China (863 Program), No. 2009AA01Z329.

References

[1] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance Capture from Sparse Multi-View Video," *ACM Trans. Graphics,* vol. 27, no. 3, pp. 98:1-98:10, 2008.

[2] D. Vlasic, I. Baran, W. Matusik, and J. Popovic, "Articulated Mesh Animation from Multi-View Silhouettes," *ACM Trans. Graphics,* vol. 27, no. 1, pp. 97(1)-97(9), 2008.

[3] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popovic, S. Rusinkiewicz, and W. Matusik, "Dynamic Shape Capture Using Multi-View Photometric Stereo," *ACM Trans. Graphics,* vol. 28, no. 5, Dec. 2009.

[4] D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur, "Markerless Garment Capture," *ACM Trans. Graphics,* vol. 27, no. 3, pp. 99-106, 2008.

[5] T. Popa, Q. Zhou, D. Bradley, V. Kraevoy, H. Fu, A. Sheffer, and W. Heidrich, "Wrinkling Captured Garments Using Space-Time Data-Driven Deformation," *Computer Graphics Forum,* vol. 28, no. 2, pp. 427-435, 2009.

[6] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '06),* vol. 1, pp. 519-528, 2006.

[7] J. Starck, G. Miller, and A. Hilton, "Volumetric Stereo with Silhouette and Feature Constraints," *Proc. British Machine Vision Conf. (BMVC),* Sept. 2006.

[8] J. Starck and A. Hilton, "Surface Capture for Performance Based Animation," *IEEE Computer Graphics and Applications,* vol. 27, no. 3, pp. 21-31, May 2007.

[9] G. Vogiatzis, C.H. Esteban, P.H.S. Torr, and R. Cipolla, "Multi-View Stereo via Volumetric Graph-Cuts and Occlusion Robust Photoconsistency," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 12, pp. 2241-2246, Dec. 2007.

[10] J.-P. Pons, R. Keriven, and O. Faugeras, "Modelling Dynamic Scenes by Registering Multi-View Image Sequences," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '05),* vol. 2, pp. 822-827, 2005.

[11] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate Multi-View Reconstruction Using Robust Binocular Stereo and Surface Meshing," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '08),* 2008.

[12] Y. Liu, Q. Dai, and W. Xu, "A Point Cloud Based Multi-View Stereo Algorithm for Free-Viewpoint Video," *IEEE Trans. Visualization and Computer Graphics,* vol. 16, no. 3, pp. 407-418, May/June 2010.

[13] Y. Liu, X. Cao, Q. Dai, and W. Xu, "Continuous Depth Estimation for Multi-View Stereo," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '09),* 2009.

[14] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '07),* 2007.

[15] mview, http://vision.middlebury.edu/mview/, 2010.

[16] T. Yu, N. Xu, and N. Ahuja, "Shape and View Independent Reflectance Map from Multiple Views," *Proc. European Conf. Computer Vision (ECCV),* 2004.

[17] 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," *Int'l J. Computer Vision,* vol. 76, no. 3, pp. 245-256, 2008.

[18] R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah, "Shape from Shading: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 690-706, Aug. 1999.

[19] R. Basri and D.W. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 218-233, Feb. 2003.

[20] R. Ramamoorthi and P. Hanrahan, "On the Relationship between Radiance and Irradiance: Determining the Illumination from Images of Convex Lambertian Object," *J. Optical Soc. of America A,* vol. 18, pp. 2448-2459, 2001.

[21] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric Stereo with General, Unknown Lighting," *Int'l J. Computer Vision,* vol. 72, no. 3, pp. 239-257, 2006.

[22] C.P. Chen and C.S. Chen, "The 4-Source Photometric Stereo under General Unknown Lighting," *Proc. European Conf. Computer Vision (ECCV '06),* 2006.

[23] S.K. Zhou, G. Aggarwal, R. Chellappa, and D.W. Jacobs, "Appearance Characterization of Linear Lambertian Objects, Generalized Photometric Stereo, and Illumination-Invariant Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 2, pp. 230-245, Feb. 2007.

[24] P. Belhumeur, D. Kriegman, and A. Yuille, "The Bas-Relief Ambiquity," *Int'l J. Computer Vision,* vol. 35, no. 1, pp. 33-44, 1999.
[25] D. Simakov, D. Frolova, and R. Basri, "Dense Shape Reconstruction of a Moving Object under Arbitrary, Unknown Lighting," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '03),* pp. 1202-1209, 2003.
[26] D. Frolova, D. Simakov, and R. Basri, "Accuracy of Spherical Harmonic Approximations for Images of Lambertian Objects under Far and Near Lighting," *Proc. European Conf. Computer Vision (ECCV '04),* 2004.
[27] A. Georghiades, "Recovering 3-D Shape and Reflectance from a Small Number of Photographs," *Proc. Eurographics Symp. Rendering (EGSR '03),* pp. 230-240, 2003.
[28] T. Migita, S. Ogino, and T. Shakunaga, "Direct Bundle Estimation for Recovery of Shape, Reflectance Property and Light Position," *Proc. European Conf. Computer Vision (ECCV '08),* 2008.
[29] L. Zhang, B. Curless, and S.M. Seitz, "Spacetime Stereo: Shape Recovery for Dynamic Scenes," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '03),* pp. 367-374, 2003.
[30] J. Davis, R. Ramamoothi, and S. Rusinkiewicz, "Spacetime Stereo : A Unifying Framework for Depth from Triangulation," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '03),* 2003.
[31] L. Zhang, B. Curless, A. Hertzmann, and S.M. Seitz, "Shape and Motion under Varying Illumination: Unifying Stucture from Motion, Photometric Stereo, and Multi-View Stereo," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '03),* vol. 1, p. 618, 2003.
[32] J. Lim, J. Ho, M.H. Yang, and D. Kriegman, "Passive Photometric Stereo from Motion," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '05),* vol. 2, pp. 1635-1642, 2005.
[33] N. Joshi and D. Kriegman, "Shape from Varying Illumination and Viewpoint," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '07),* vol. 2, pp. 1-7, 2007.
[34] M. Weber, A. Blake, and R. Cipolla, "Towards a Complete Dense Geometric and Photometric Reconstruction under Varying Pose and Illumination," *Proc. British Machine Vision Conf. (BMVC '02),* 2002.
[35] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently Combining Positions and Normals for Precise 3D Geometry," *Proc. ACM SIGGRAPH '05,* pp. 536-543, 2005.
[36] G. Vogiatzis, C. Hernandez, and R. Cipolla, "Reconstruction in the Round Using Photometric Normals," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '06),* 2006.
[37] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Muti-View Photometric Stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 3, pp. 548-554, Mar. 2008.
[38] D. Samaras and D. Metaxas, "Incorporating Illumination Constraints in Deformable Models for Shape from Shading and Light Direction Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 247-264, Feb. 2003.
[39] N. Birkbeck, D. Cobzas, P. Sturm, and M. Jagersand, "Variational Shape and Reflectance Estimation under Changing Light and Viewpoints," *Proc. European Conf. Computer Vision (ECCV '06),* 2006.
[40] C. Theobalt, N. Ahmed, H. Lensch, M. Magnor, and H.P. Seidel, "Seeing People in Different Light-Joint Shape, Motion, and Reflectance Capture," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 4, pp. 663-674, July/Aug. 2007.
[41] N. Ahmed, C. Theobalt, P. Dobrev, H.-P. Seidel, and S. Thrun, "Robust Fusion of Dynamic Shape and Normal Capture for High-Quality Reconstruction of Time-Varying Geometry," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '08),* 2008.
[42] W.C. Ma, T. Hawkins, P. Peers, C.F. Chabert, M. Weiss, and P. Debevec, "Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination," *Proc. Eurographics Symp. Rendering (EGSR '07),* 2007.
[43] H.P.A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, "Image-Based Reconstruction of Spatial Appearance and Geometric Detail," *ACM Trans. Graphics,* vol. 22, pp. 234-257, 2003.
[44] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," *Proc. Eurographics Symp. Geometry Processing (EGSG '06),* pp. 61-70, 2006.
[45] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection.* John Wiley and Sons, 1987.
[46] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.
[47] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel, "Free-Viewpoint Video of Human Actors," *ACM Trans. Graphics,* vol. 22, no. 3, pp. 569-577, 2003.
[48] J. Lawrence, A. Ben-Artzi, C. DeCoro, W. Matusik, H. Pfister, R. Ramamoorthi, and S. Rusinkiewicz, "Inverse Shade Trees for Non-Parametric Material Representation and Editing," *ACM Trans. Graphics,* vol. 25, no. 3, pp. 735-745, 2006.
[49] E. Candes and T. Tao, "Decoding by Linear Programming," *IEEE Trans. Information Theory,* vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
[50] R. Ramamoorthi and P. Hanrahan, "An Efficient Representation for Irradiance Environment Maps," *Proc. ACM SIGGRAPH '01,* 2001.
[51] P. Debevec, "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography," *Proc. ACM SIGGRAPH '98,* pp. 189-198, July 1998.
[52] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," *Proc. ACM SIGGRAPH '96,* pp. 311-318, 1996.
[53] R.H. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Eng.,* vol. 19, no. 1, pp. 139-144, 1980.
[54] Y.Y. Schechner, S.K. Nayar, and P.N. Belhumeur, "A Theory of Multiplexed Illumination," *Proc. IEEE Int'l Conf. Computer Vision (ICCV),* vol. 2, pp. 808-815, 2003.

Chenglei Wu received the BS degree in electronic science and engineering from Nanjing University, China, in 2007, and the ME degree from the Automation Department, Tsinghua University, Beijing, China, in 2010. His research interests include image-based modeling and performance capture.

Yebin Liu received the BE degree from Beijing University of Posts and Telecommunications, China, in 2002, and the PhD degree from the Automation Department, Tsinghua University, Beijing, China, in 2009. He is currently a postdoctoral research fellow at the Automation Department, Tsinghua University. His research interests include light field, image-based modeling and rendering, and multicamera array techniques.

Qionghai Dai (SM '05) received the BS degree in mathematics from Shanxi Normal University, China, in 1987, and the ME and PhD degrees in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the faculty of Tsinghua University, Beijing, China, and is currently a professor and the director of the Broadband Networks and Digital Media Laboratory. His research areas include video communication, computer vision, and graphics. He is a senior member of the IEEE.

Bennett Wilburn received the PhD degree from Stanford University in 2005. He is currently a researcher in the Visual Computing Group at Microsoft Research Asia in Beijing, China. For the PhD thesis work, he designed custom CMOS cameras for a scalable video camera array, and devised high-performance imaging methods using the 100 camera system. He is particularly interested in multiple view video capture, processing, and modeling, and has been exploring photometric stereo for dynamic scenes.