Shading-Based Surface Detail Recovery Under General Unknown Illumination

Di Xu, Qi Duan, Jianmin Zheng, Juyong Zhang, Jianfei Cai, *Senior Member, IEEE*, and Tat-Jen Cham

Abstract—Reconstructing the shape of a 3D object from multi-view images under unknown, general illumination is a fundamental problem in computer vision. High quality reconstruction is usually challenging especially when fine detail is needed and the albedo of the object is non-uniform. This paper introduces vertex overall illumination vectors to model the illumination effect and presents a total variation (TV) based approach for recovering surface details using shading and multi-view stereo (MVS). Behind the approach are the two important observations: (1) the illumination over the surface of an object often appears to be piecewise smooth and (2) the recovery of surface orientation is not sufficient for reconstructing the surface, which was often overlooked previously. Thus we propose to use TV to regularize the overall illumination vectors and use visual hull to constrain partial vertices. The reconstruction is formulated as a constrained TV-minimization problem that simultaneously treats the shape and illumination vectors as unknowns. An augmented Lagrangian method is proposed to quickly solve the TV-minimization problem. As a result, our approach is robust, stable and is able to efficiently recover high-quality surface details even when starting with a coarse model obtained using MVS. These advantages are demonstrated by extensive experiments on the state-of-the-art MVS database, which includes challenging objects with varying albedo.

Index Terms—Shape from shading, 3D reconstruction, shape refinement, general unknown illumination, spatially varying albedo

1 Introduction

This paper considers the problem of recovering the detailed 3D surface shape of an object, which may have spatially varying albedo, from multi-view images. With advance and wide availability of various acquisition devices, 3D shape recovery under general unknown illumination conditions is of significant interest in practice. Extensive research has been done in this area and many techniques have been developed. In particular, multi-view stereo (MVS) methods [2] compute depth from corresponding views of the same point in multiple images and reconstruct the overall shape of an object. Shape-from-shading (SfS) [3] and photometric stereo (PS) methods [4], [5], [6] use shading information to estimate shapes and recover surface details. Recently much work aims to combine different techniques to improve the performance of reconstruction [5], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18].

Although the state-of-the-art methods often work well, many of them still have various underlying requirements that limit their application scope in practice. For example, traditional SfS and PS methods require the surface reflectance properties to be known in advance or assume varying controlled lighting conditions (in the studio lighting environment) in order to compute the normals of surfaces accurately [12], [19]. More often, simple albedo is assumed in the process of reconstruction [11], [20], [21], [22], which may not be accurate in practice. Some works [23], [24] relax this assumption and explicitly estimate the albedo for surface recovery or other purposes, but their computational complexity is high. To model scenes with complex lighting, spherical harmonics (SH) are used to parameterize incident illumination [11], [22], [24], [25] and the visibility of objects has to be pre-calculated before estimating the SH lighting. In Wu et al.' s methods [11], [24], multi-view stereo and shading-based optimization are combined to provide a nice high quality reconstruction, but the methods assume a good initial MVS mesh. Usually such a mesh is densely tessellated, e.g., in their "angel" example, the mesh has 500,000 vertices.

This paper presents a new approach to recover surface details using shading and multi-view stereo, which overcomes the above mentioned issues like the assumption of constant albedo and the requirement of a good initial 3D mesh. We restrict the reconstruction to a Lambertian object from multi-view images captured under general, unknown illumination conditions. Different from previous methods, we do not assume constant albedo or compute the visibility explicitly. This makes our approach more widely applicable. Moreover, we aim at producing a high resolution model with fine surface details from a coarse initial MVS mesh, which will make our work insensitive to the initial MVS result as compared to previous methods such as [11]. To this end, we introduce the concept of overall illumination vectors where an incident illumination vector is stored at each vertex, which models the overall illumination effect including general/unknown lighting conditions, self-shadowing, occlusion and albedo. We also propose a total variation (TV) based formulation to integrate multi-view stereo and shading cues for surface reconstruction. In particular,

- *D. Xu, Q. Duan, J. Zheng, J. Cai, and T.-J. Cham are with Nanyang Technological University, 50 Nanyang Ave, Singapore 639798. E-mail: {xudi, duan0013, asjmzheng, asjfcai, astjcham}@ntu.edu.sg.*
- *J. Zhang is with the University of Science and Technology of China, Hefei, Anhui 230026, China. E-mail: juyong@ustc.edu.cn.*

Manuscript received 8 Oct. 2015; revised 7 Dec. 2016; accepted 31 Jan. 2017. Date of publication 16 Feb. 2017; date of current version 5 Jan. 2018. Recommended for acceptance by Y. Matsushita. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TPAMI.2017.2671458

we observe that the illumination over the surface of an object often tends to be piecewise smooth, so a TV term is imposed to the overall illumination map for regularization.

Moreover, we notice that the recovery of surface orientation is not sufficient for reconstructing the surface of the object. Even with perfect surface normal recovery, the geometry might still be wrong. Considering a uniformly shrunk ball, its surface normal is exactly the same as that of the original ball. We believe that this is a problem that has been largely overlooked by the previous shading based methods, e.g., [11], [20], [21], [22]. Although there is usually a fidelity term in the existing shading based methods, which can prevent the recovered mesh from drifting away from the initial mesh, its significance often limits the capability of the shading term in surface detail recovery. Thus, in this paper we propose to use a visual hull reconstruction to constrain a small portion of vertices of the initial MVS mesh, which can avoid the drifting problem without sacrificing the surface detail recovery.

Our algorithm starts with a coarse initial MVS mesh, refines it using classic subdivision scheme, and uses the proposed TV-based minimization with visual hull constraint to optimize the mesh and estimate the illumination as well. We have validated the algorithm using various models from the latest MVS dataset [26]. Over five categories of models including challenging objects with multiple albedos have been tested and the results confirm that the algorithm is robust, stable and is able to efficiently recover surface details even when starting with a coarse MVS mesh. The major contributions of the paper include:

- We introduce the concept of vertex overall illumination vectors to simulate the overall illumination effect which includes lighting, occlusion and albedo. As a result, our technique is able to handle objects with spatially varying albedo, general lighting, and vertex visibility without the need of extra processes, which greatly facilitates the reconstruction.
- We propose a new formulation of shape optimization based on TV-minimization, which enables us to generate high fidelity 3D models from multi-images captured under general lighting conditions. Different from previous works [11], [21], [23], [24] which alternate between lighting estimation and geometry optimization, our formulation estimates both geometry and illumination using the same objective function. This helps speed up the computation and reduce the chance of converging to a wrong solution or diverging. Moreover, we propose to use a visual hull reconstruction to constrain a small portion of vertices of the initial mesh in order to remove possible shape ambiguity.
- We develop an augmented Lagrangian solver to effectively solve the TV-minimization problem.

The paper is organized as follows: Section 2 briefly reviews some related work. Section 3 introduces the concept of the vertex overall illumination vector, which plays an important role in our approach. Then in Section 4 we describe our variational model for surface recovery using shading and coarse multi-view-stereo, and in Section 5 we present an Augmented Lagrangian-based approach for solving the proposed model. Section 6 provides experimental results on the state-of-the-art MVS database and Section 7 concludes the paper by summarizing the results and discussing topics for future work.

2 Related Work

An excellent survey on MVS is given in [2] where the state-of-the-art MVS algorithms are compared and evaluated. These algorithms are classified into four categories: volumetric methods [27], [28], [29], surface evolution methods [30], [31], depth map fusion methods [32], [33], [34] and surface region growing methods [35], [36]. One essential part of MVS methods is to find point matches from multi-view images and then to calculate depth through triangulation. If good correspondence cannot be found, high-frequency shape detail may not be well recovered. The MVS evaluation benchmark [2] shows that even the-state-of-the-art MVS algorithms may miss high-frequency surface details though they can recover overall shapes well.

In contrast to MVS, SfS estimates surface normals from shading cues using multiple images taken with a fixed viewpoint and variable lighting conditions [3]. The shape is then reconstructed based on the normal field. Since the variances of surface normals intrinsically reflect high-frequency shape changes, SfS can better recover fine surface details. However, SfS usually imposes some assumptions on the lighting condition and surface material properties. For instance, the lighting is assumed to be known as a prior [3] or distant light sources [37]. The objects to be reconstructed are assumed to have simple albedo [20], [21], [23], [24]. Meanwhile, some simplification strategies are adopted for the ease of computation. For example, the computation of visibility map for SH lighting is omitted [24] and the updating of vertices of the reconstructed mesh is performed only along the surface normal direction [23].

To handle complicated lighting conditions and reflectance properties, various sophisticated methods have also been developed. Hertzmann and Seitz proposed an example-based approach to recovering the geometry of objects with general reflectance properties in arbitrary distant and unknown lighting environments [38]. The method assumes that the camera viewpoint is fixed but the illumination varies over the input images. Moreover, there must be one or more examples with similar materials and known geometry that are captured under the same illumination conditions. Goldman et al. proposed a photometric stereo approach for recovering shape and spatially varying BRDF from multiple images captured at the same viewpoint but with different illumination [39]. The basic idea of the approach is the assumption that the materials of most objects can be represented as a convex combination of a small number of fundamental materials. Alldrin et al. proposed a data-driven approach for simultaneously recovering shape and spatially varying reflectance of a surface from photometric stereo images [40]. In [21], Han et al. proposed to combine a global lighting model and a local lighting model. The global lighting model is the quadratic function from [22] describing natural illumination and the local lighting model considers photo-consistency and smoothness of local lighting parameters. Recently Richter and Roth proposed a discriminative learning approach to estimate surface normals from shading

in uncalibrated illumination, which used a large-scale dataset for training and analysis [41].

Due to the nature of MVS and SfS, it is of interest to combine them to maximize the strengths of both techniques [5], [7], [8], [9], [10], [11], [12]. For example, Jin et al. [7], [9] introduced variational frameworks that combine various MVS and shading constraints to estimate the shape of Lambertian objects, surface albedo and lighting condition through surface evolution and variational minimization. Joshi et al. [8] proposed to merge the depth map and surface normal field to reconstruct objects and Hernández et al. [5] first recovered the surface normal and then refined the MVS mesh. Although these two methods can produce good 3D reconstruction, they require a very large number of sample images in different viewpoints and vary lighting conditions in each viewpoint. Moreover, they both assume a single point light source in a dark room and cannot handle general lighting environments. Yoshiyasu et al. [12] introduced a topology adaptive mesh evolution method by evolving the object model to match MVS boundary and shading constraints. Compared with [5], [8], this method reduces the number of required sample images but still assumes a single light source with known position and direction in a dark environment. The work most related to ours is Wu et al.'s method [11] that uses spherical harmonics to model general lighting conditions and refines the geometric model through shading. The method can efficiently recover surface details of constant albedo objects. While our method jointly estimates the geometry and lighting and just requires a coarse initial MVS mesh, their method performs the lighting estimation and geometry refinement separately and assumes a good initial guess of the geometry to ensure convergence of the iterations.

3 Vertex Overall Illumination Vectors

For a Lambertian surface, the image formation can be approximately described by the Lambertian reflectance model [42]

$$I_o(v) = \int_{\Omega(v)} \rho(v) I_i(v,\omega) \max(\omega \cdot \mathbf{n}(v), 0) V(v,\omega) d\omega, \quad (1)$$

where $I_o(v)$ is the reflected radiance of the object at vertex v, $\rho(v)$ is the albedo or an approximation of the bidirectional reflectance distribution function (BRDF), ω is the incident direction, $I_i(v,\omega)$ is the incident radiance along ω, $\mathbf{n}(v)$ is the unit surface normal at v, $\Omega(v)$ represents a hemisphere of incident directions at v, and $V(v,\omega)$ stands for a binary visibility function of vertex v to direction ω. This is the fundamental model used by most SfS methods. While many SfS methods attempt to recover both geometry and lighting, the task involves the estimation of underlying lighting conditions, visibility functions, and surface albedo, which is very difficult in general as shown in formula (1). Even though the use of spherical harmonics can simplify the formulation of incident illumination [25], it still requires the computation of visibility map and the estimation of albedo, which actually take a considerable amount of time especially for a dense mesh [11]. Taking the bunny model with 30,000 vertices as an example, rendering the visibility map takes 33 minutes, which is around 25 percent of the whole processing time.

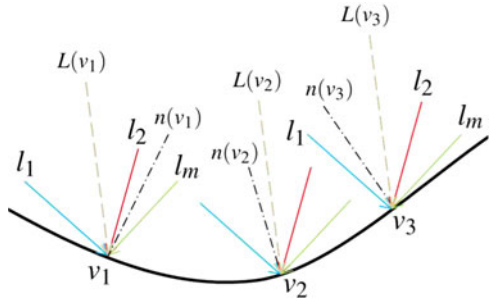


Fig. 1. At each point v_i on the surface, the vertex overall illumination vector $L(v_i)$ represents the overall effect of all incident lights such as $l_1, l_2, \ldots, l_m$ from different directions and $n(v_i)$ represents the normal of the surface at v_i.

Our goal is to recover geometry. There is no need to estimate the exact lighting conditions. In fact, we let $\Omega'(v)$ denote the subset of Ω for which $\omega \cdot \mathbf{n}(v) > 0$ and $V(v,\omega) = 1$. Then model (1) can be rewritten as

$$I_o(v) = \left(\int_{\Omega'(v)} \rho(v) I_i(v,\omega) \omega d\omega \right) \cdot \mathbf{n}(v). \quad (2)$$

Let

$$L(v) = \int_{\Omega'(v)} \rho(v) I_i(v,\omega) \omega d\omega. \quad (3)$$

We call $L(v)$ the *vertex overall illumination vector* at v. It represents the overall effect of all incident lights at v (see Fig. 1 for an illustration). If $L(v)$ is known, the reflected radiance at v can be easily computed: $I_o(v) = L(v) \cdot \mathbf{n}(v)$. That is, shading cues only depend on the surface normals and the overall illumination vectors. This implies that as long as the overall illumination vectors are sufficiently accurate, whether underlying lighting conditions can be recovered or not is not important. Therefore instead of estimating individual incident lights and computing the visibility maps of each vertex to each possible incident light direction, we propose to estimate the overall illumination vector for each vertex in our shading-based surface recovery process. In this way, concave surfaces and self-occlusion can be automatically handled in estimating the overall illumination vectors. There is no need to compute visibility maps explicitly, which greatly simplifies the surface reconstruction process. Moreover, note that the overall illumination vector has implicitly included the albedo and each vertex has such a vector. Hence no matter whether an object has constant albedo or spatially varying albedo, the formula for computing the reflected radiance is the same. This enables us to handle objects with constant albedo or spatially varying albedo in the same manner.

It can be seen from (3) that the vertex overall illumination vector $L(v)$ depends on surface normal $\mathbf{n}(v)$, visibility function $V(v,\omega)$, albedo $\rho(v)$ and the lighting condition. In most practical scenarios, the surface of the object is (piecewise) smooth, $\rho(v)$ is constant or continuous globally or locally, and the lighting condition varies smoothly. Then $L(v)$ is likely to be piecewise constant or piecewise smooth over the surface of the object.

We validate this piecewise smoothness property by computing the distribution of $L(v)$ on two triangular mesh

models (bunny and budda) under general and natural lighting conditions using light probe images of [43]. The two models are typical graphics models and have complicated geometric shapes. A light probe image is an omnidirectional, high dynamic range image that records the incident illumination conditions at a particular point in space. Such images are usually captured under general and natural illumination conditions. While in theory we can compute the overall illumination vectors from the light probe images, the computation time would be huge if we exhaustively apply each pixel of the images due to the high resolution of the images and the huge number of vertices of the meshes. For efficiency purposes, we adopt the MedianCut method of [44]. For each light probe image, we generate 32 light sources to approximate the original global illumination and then use them to illuminate the two models. As shown in [44], the MedianCut result gives a very good approximation to the global illumination. After $L(v)$ at each vertex is computed, we map the values of three Cartesian components of $L(v)$ to RGB channels and use the RGB color to render the models. Fig. 2 shows the visualization of $L(v)$ under four different lighting conditions. The visualization indicates that the distribution of $L(v)$ is piecewise smooth over the surfaces in both indoor and outdoor scenes.

4 TV-Based Formulation

This section describes a variational formulation for our surface reconstruction algorithm. The input to the algorithm is a set of multiple images captured at different viewpoints under general, unknown illumination, and the output is a triangular mesh for the reconstructed surface. The algorithm consists of two stages. The first stage is to create an MVS mesh. We just use existing MVS methods to generate a coarse mesh. The camera parameters are also recovered using the existing MVS methods if they are unknown. The mapping between the mesh and the multi-view images is then established so that for each mesh vertex, its corresponding intensity value captured in each of multi-view images can be found. It is worth pointing out that while a dense MVS mesh with a good guess of the geometry is usually needed for previous work such as [11], a coarse MVS mesh is sufficient for our method. This makes our method simple and fast in building the initial MVS mesh. Once we obtain a coarse initial mesh, we use classic mesh subdivision schemes such as the Butterfly subdivision to refine the mesh [45], generating a dense mesh.

The second stage is our contribution that uses the shading cues from the multi-view images to optimize the position of the mesh vertices in order to deliver a high-resolution triangular mesh that recovers high-frequency surface details in addition to the overall geometry shape. This is done by iteratively minimizing a TV-based objective function formulated using the geometry and overall illumination vectors (see Section 4.1). Specifically, the total variation is applied to constrain the overall illumination vectors. The optimal geometry is obtained by solving the minimization problem with visual hull constraint (see Section 4.2), which enables the constructed surface details to reflect the shading variations in the multi-view images. The rest of this section elaborates the formulation of the constrained optimization problem and in Section 5 we present an augmented Lagrangian method for effectively solving the minimization problem.

4.1 Iterative TV-Minimization

Assume that the obtained initial MVS mesh is composed of N vertices $\{v_i^{in}\}$ and a set of triangles. Denote by $\mathbf{n}_i^{in}$ the normal of the mesh at v_i^{in}. We fix the connectivity of the mesh, but move vertices v_i^{in} to new positions v_i such that the new mesh better matches the intensity captured in the multi-view images. Our basic approach is to treat vertices v_i and their respective overall illumination vectors $L(v_i)$ as variables and find them as a solution to the minimization problem

$$\min\left\{\frac{\alpha}{2}E_f+\frac{\beta}{2}E_{sh}+\frac{\eta}{2}E_{lap}+E_{tv}\right\}, \tag{4}$$

where α, β and η are weights. Unless the difference between new positions v_i and v_i^{in} is within a prescribed threshold, we replace v_i^{in} by v_i and repeat solving the minimization problem. The objective function of (4) consists of four terms accounting for position, shading, smoothness and lighting constraints, which are explained below.

(1) *Fidelity term*:

$$E_f=\sum_{i=1}^{N}(\|v_i-v_i^{in}\|^2+\|\mathbf{n}(v_i)-\mathbf{n}_i^{in}\|^2).$$

This term is introduced to prevent the refined vertices v_i and their normals $\mathbf{n}(v_i)$ from deviating from their counterparts of the initial MVS mesh too much. The normal $\mathbf{n}(v_i)$ is calculated as the area-weighted average of the normals of the neighboring triangle faces surrounding v_i. If the vertices on the one-ring neighbors of v_i are $v_{i,1}, v_{i,2}, \ldots, v_{i,a}$, it can be verified that the normal can be computed by

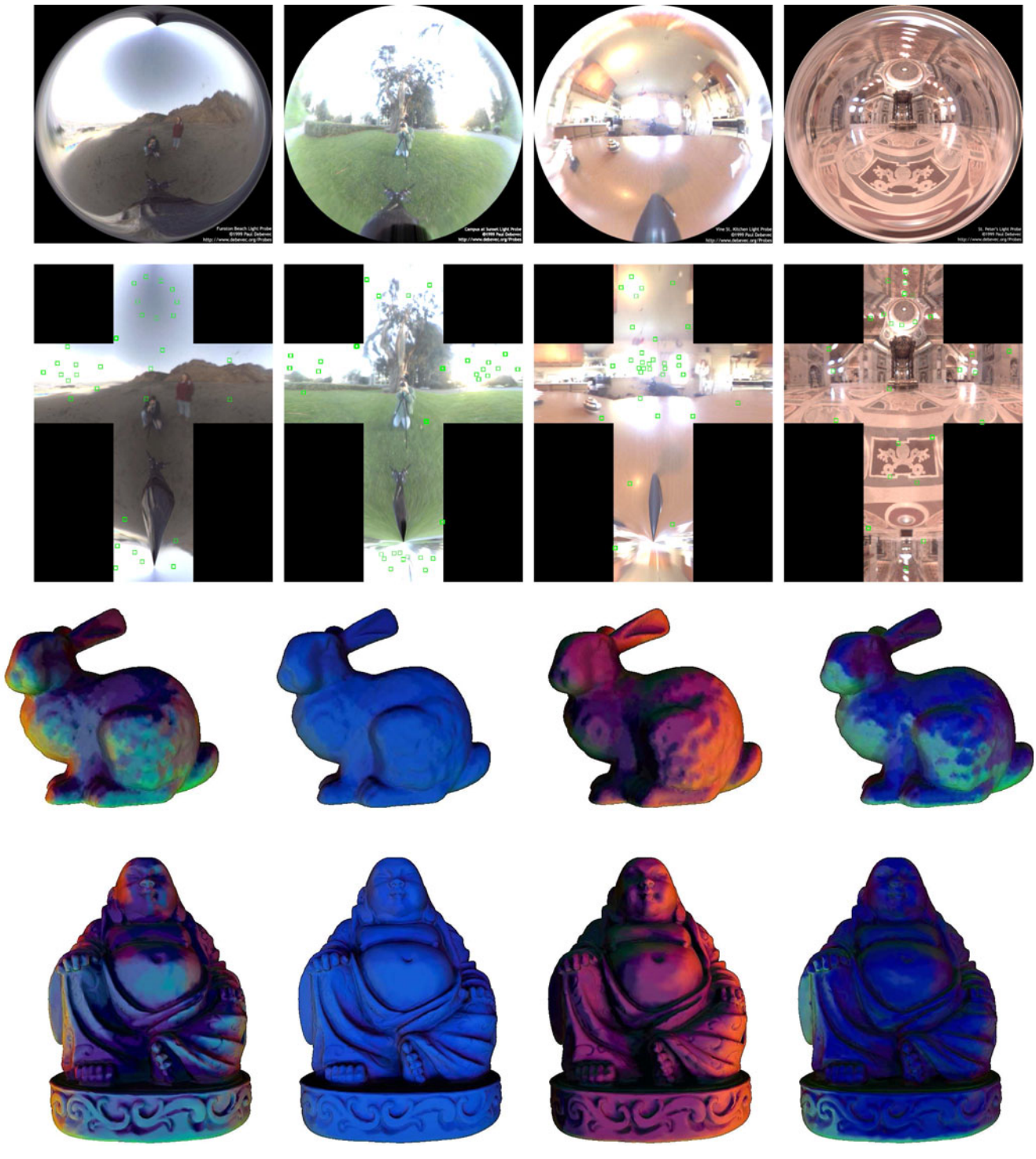
$$\mathbf{N}(v_i)=v_{i,1}\times v_{i,2}+v_{i,2}\times v_{i,3}+\cdots+v_{i,a}\times v_{i,1},$$

and $\mathbf{n}(v_i)=\frac{\mathbf{N}(v_i)}{\|\mathbf{N}(v_i)\|}$. E_f is a function of vertices v_i.

(2) *Shading term*:

$$E_{sh}=\sum_{i=1}^{N}\|L(v_i)\cdot\mathbf{n}(v_i)-c_i\|^2+\sum_{(v_i,v_j)\in E}\|(L(v_i)\cdot\mathbf{n}(v_i)-L(v_j)\cdot\mathbf{n}(v_j))-(c_i-c_j)\|^2,$$

where E is the set of all edges of the mesh and c_i is the average of the intensity values in all the multi-view images corresponding to vertex v_i. The first term of E_{sh} is the intensity error measuring the difference between the computed reflected radiance and the average of the captured intensities, which is the key of SfS. The second term of E_{sh} is the gradient error measuring the difference between the gradients of the computed reflected radiance and the average of the captured intensities. As pointed out in [11], the gradient error term imposes another constraint on the normal changes that reflect the high frequency surface details, which the conventional MVS methods have difficulty to recover, and introducing the gradient error makes the computation more stable than just using the intensity error. E_{sh} is



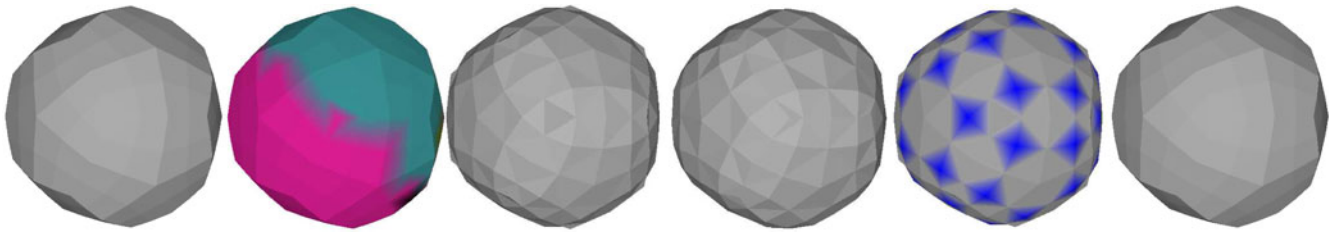
(a) beach (b) campus (c) kitchen (d) stperters

Fig. 2. Visualization of $L(v)$, which is mapped to RGB, over the 3D models lit by different light probes. First row: the original light probe map from [43]. Second row: the distribution of light sources in the cube map, where each green box indicates the location of one light source. Third and fourth rows: bunny and budda models with color mapped $L(v)$.

a function of both vertices v_i and vertex overall illumination vectors $L(v_i)$.

(3) *Laplacian term*: $E_{lap} = \sum_{i=1}^{N} \|v_i - \overline{v}_i\|^2$ where $\overline{v}_i$ is the average of all the 1-ring neighboring vertices $v_{i,j}$ of v_i, $j = 1, \ldots, a$, and thus $v_i - \overline{v}_i = \frac{1}{a}\sum_{i=1}^{a}(v_i - v_{i,j})$ is the Laplacian of v_i. This term is computed as the squared sum of the Laplacian of all vertices. The Laplacian is used to regularize the mesh to produce smooth model and help avoid generating singular or invalid triangles in the mesh updating process [46].

(4) *TV term*: $E_{tv} = \sum_{i=1}^{N} \|\nabla L(v_i)\|$ where ∇ is the discrete operator of the intrinsic gradient on the triangular mesh. The computation of ∇ can refer to [47]. E_{tv} is a total variation regulation term applied to the overall illumination vectors, which enables non-smoothness preservation while removing noise [48]. It has been



(a) Ground truth (b) Lighting condition (c) Initial mesh (d) Output mesh without constraints (e) Constrained candidates (in blue) (f) Output mesh with constraints

Fig. 3. Reconstruction with and without visual hull constraints.

observed that the overall illumination vectors tend to be piecewise constant or piecewise smooth. The TV term E_{tv} is introduced to constrain the overall illumination vectors and preserve this property in the optimization process.

4.2 Visual Hull Constraint

In the proposed TV-minimization, the purpose of the TV term is to regularize the lighting and the purpose of the shading term is to constrain the interaction of the lighting and the surface normal. Like all the existing shading methods, our approach aims at recovering better surface normal maps. However, it is pointed out in [11] that generating 3D geometry from recovered normal fields for general surfaces is non-trivial. In fact, the situation is even worse than that. Sometimes it is impossible to uniquely determine the geometry from the normal fields. For example, for two Lambertian surface meshes lit by directional light sources, if one is a scaled version of the other, they could have the same normal fields and the same intensities. This observation implies that the formulation only relying on shading cues in SfS methods is likely under-constrained for 3D geometry. Thus it may produce artifacts of prick-shapes in reconstructed geometry or it may not deliver accurate vertex positions even if the recovered normal field is correct. Fig. 3 shows such an example, where (a) is a ground truth, (b) depicts the lighting condition, (c) is a mesh obtained by perturbing the ground truth and is used as the initial mesh for reconstruction, and (d) is the reconstruction result using the TV-minimization. Note that in this case the objective function has actually reached its minimum value 0 which suggests that Fig. 3d is an optimal solution, but obviously the shape in Fig. 3d is different from the ground truth.

The above example implies that we need to add further constraints in the minimization process in order to precisely recover the surface shape. Here we propose a heuristic approach based on visual hull. The visual hull is created by the intersection of silhouette cones and it contains the reconstructed object. We compare the initial MVS mesh with the visual hull and identify those vertices of the MVS mesh, which are close to locally convex regions of the visual hull and at which the MVS mesh is approximately tangential to the visual hull. The blue region in Fig. 3e shows such vertices. These vertices are likely on the true surface and we thus fix them during the optimization. In this way, the TV-minimization problem becomes a constrained one and the variables are the overall illumination vectors and a subset of vertices, which usually makes the minimization problem converge to the correct solution and avoids producing artifacts of prick-shapes. Fig. 3f is the reconstruction result with those vertices in the blue region being fixed, which reproduces the ground truth. Note that visual hull constraints have been used to help MVS based 3D reconstruction in [49], [50], [51], [52]. Here, we use it to constrain our shading based surface detail recovery.

5 Augmented Lagrangian-Based Solver

Due to the non-differentiability of the total variation term and the non-linearity of the unit normal vectors, solving the minimization problem (4) is difficult. In this section we propose an augmented Lagrangian (ALM)-based method to solve (4). Augmented Lagrangian methods are known as an effective alternative to penalty methods for solving constrained optimization problems in that they replace a constrained optimization problem by a series of unconstrained problems.

Let $\mathcal{V} = [v_1, v_2, \ldots, v_N]$, $\mathcal{L} = [L(v_1), \ldots, L(v_N)]$ and denote $[\nabla L(v_1), \ldots, \nabla L(v_N)]$ by $\nabla\mathcal{L}$. We introduce a new variable $\mathcal{P} = [P_1, P_2, \ldots, P_N]$ and reformulate the TV-based minimization problem (4) to the following constrained problem

$$\begin{aligned} &\min\left\{\frac{\alpha}{2}E_f + \frac{\beta}{2}E_{sh} + \frac{\eta}{2}E_{lap} + R(\mathcal{P})\right\} \\ &s.t. \quad \mathcal{P} = \nabla L(v), \end{aligned}$$

where $R(\mathcal{P}) = \sum_{i=1}^{N} \|P_i\|$.

To eliminate the constraint, we define the augmented Lagrangian functional

$$\begin{aligned} \mathcal{G}(\mathcal{V}, \mathcal{L}, \mathcal{P}; \lambda) = &\frac{\alpha}{2}E_f + \frac{\beta}{2}E_{sh} + \frac{\eta}{2}E_{lap} + R(\mathcal{P}) \\ &+ \lambda \cdot (\mathcal{P} - \nabla\mathcal{L}) + \frac{r}{2}\|\mathcal{P} - \nabla\mathcal{L}\|^2, \end{aligned} \tag{5}$$

where $\lambda = [\lambda_1, \ldots, \lambda_N]$ are Lagrange multipliers, r is the penalty coefficient which is a positive constant, and $\frac{r}{2}\|\mathcal{P} - \nabla\mathcal{L}\|^2$ is the augmented term. Consider the following saddle-point problem: finding $(\mathcal{V}^*, \mathcal{L}^*, \mathcal{P}^*; \lambda^*)$ such that

$$\mathcal{G}(\mathcal{V}^*, \mathcal{L}^*, \mathcal{P}^*; \lambda) \le \mathcal{G}(\mathcal{V}^*, \mathcal{L}^*, \mathcal{P}^*; \lambda^*) \le \mathcal{G}(\mathcal{V}, \mathcal{L}, \mathcal{P}; \lambda^*),$$

for all $(\mathcal{V}, \mathcal{L}, \mathcal{P}; \lambda)$. According to [53], the saddle-point problem has at least one solution and all the saddle-points $(\mathcal{V}^*, \mathcal{L}^*, \mathcal{P}^*; \lambda^*)$ have the same $\mathcal{V}^*$ and $\mathcal{L}^*$ which are the solution to the original problem (4). Thus we solve (5) by iteratively solving two subproblems: the $\mathcal{VL}$-subproblem and the $\mathcal{P}$-subproblem, and updating the Lagrange multipliers.

- $\mathcal{VL}$-subproblem: Given $\mathcal{P}$, solve

$$\min_{\mathcal{V},\mathcal{L}}\left\{\frac{\alpha}{2}E_f + \frac{\beta}{2}E_{sh} + \frac{\eta}{2}E_{lap} + \frac{r}{2}\left\|\left(\mathcal{P}^{(k)} + \frac{\lambda^{(k)}}{r}\right) - \nabla\mathcal{L}\right\|^2\right\},$$

for $\mathcal{V}$ and $\mathcal{L}$ where k is the previous iteration number. This problem refines both the vertices and the vertex overall illumination vectors. It is a non-linear least squares optimization problem. To solve the non-linear least squares problem, we use the Levenberg-Marquardt algorithm, which is one of the most widely used nonlinear optimization methods [54], [55].

To describe the LM algorithm, we rewrite the objective function as

$$\frac{\alpha}{2}E_f+\frac{\beta}{2}E_{sh}+\frac{\eta}{2}E_{lap}+\frac{r}{2}\|\left(\mathcal{P}^{(k)}+\frac{\lambda^{(k)}}{r}\right)-\nabla\mathcal{L}\|^2 = \sum_{i=1}^{n}\delta_i^2(\mathcal{V},\mathcal{L}),$$

where $\delta_i(\mathcal{V},\mathcal{L})$ are some functions in $\mathcal{V}$ and $\mathcal{L}$. The algorithm works in an iterative way. Starting from an initial guess for $(\mathcal{V},\mathcal{L})$, each iteration step replaces the current $(\mathcal{V},\mathcal{L})$ by a new estimate $(\mathcal{V},\mathcal{L})+(\Delta\mathcal{V},\Delta\mathcal{L})$, where $(\Delta\mathcal{V},\Delta\mathcal{L})$ are the solution to a sparse linear system

$$(\mathbf{J}^T\mathbf{J}+\mu\,\mathrm{diag}(\mathbf{J}^T\mathbf{J}))\begin{pmatrix}\Delta\mathcal{V}\\ \Delta\mathcal{L}\end{pmatrix}=-\mathbf{J}^T\mathbf{F}(\mathcal{V},\mathcal{L}),$$

with the Jacobian matrix

$$\mathbf{J}=\begin{pmatrix}\frac{\partial\delta_1}{\partial\mathcal{V}} & \frac{\partial\delta_1}{\partial\mathcal{L}}\\ \vdots & \vdots\\ \frac{\partial\delta_n}{\partial\mathcal{V}} & \frac{\partial\delta_n}{\partial\mathcal{L}}\end{pmatrix},$$

the diagonal matrix $\mathrm{diag}(\mathbf{J}^T\mathbf{J})$ consisting of the diagonal elements of $\mathbf{J}^T\mathbf{J}$, the damping factor μ and $\mathbf{F}(\mathcal{V},\mathcal{L})=(\delta_1(\mathcal{V},\mathcal{L}),\ldots,\delta_n(\mathcal{V},\mathcal{L}))^T$. The damping factor is adjusted at each iteration. If the reduction of the objective function is rapid, a smaller value is used for μ; Otherwise, if an iteration gives insufficient reduction, μ is increased. This damping strategy makes the iterative procedure converge quickly from a wider range of initial guesses.

- $\mathcal{P}$-subproblem: Given $\mathcal{V}$ and $\mathcal{L}$, solve

$$\min_{\mathcal{P}}\Big\{R(\mathcal{P})+\lambda^{(k)}\cdot\mathcal{P}+\frac{r}{2}\|\mathcal{P}-\nabla\mathcal{L}\|^2\Big\},$$

for $\mathcal{P}$. This problem is decomposable and thus can be solved for each P_i independently. That is, for each i, we solve

$$\min_{P_i}\Big\{\|P_i\|+\lambda_i^{(k)}\cdot P_i+\frac{r}{2}\|P_i-\nabla L(v_i)\|^2\Big\}.$$

By a simple geometric analysis, it can be found that the above problem has a closed form solution

$$P_i=\max\left(0,1-\frac{1}{r\|w_r\|}\right)w_r$$

where $w_r=\nabla L(v_i)-\lambda_i^{(k)}/r$.

The λ is updated along the direction of $\mathcal{P}^{(k+1)}-\nabla\mathcal{L}^{(k+1)}$ in order to increase $\mathcal{G}(\mathcal{V}^{(k+1)},\mathcal{L}^{(k+1)},\mathcal{P}^{(k+1)};\lambda)$. The whole augmented Lagrangian-based solver is outlined in Algorithm 1.

Algorithm 1. Augmented Lagrangian-Based Solver

Input: Initial MVS mesh $\{v_i^{in}\in\mathbb{R}^3|i=1,\ldots,N\}$ and the average intensity $\{c_i\in\mathbb{R}^1|i=1,\ldots,N\}$; r,ϵ;
Output: Optimal $\{v_i\in\mathbb{R}^3|i=1,\ldots,N\}$ and overall illumination vector $\{L(v_i)\in\mathbb{R}^3|i=1,\ldots,N\}$;
1: initialization: $v_i^{(0)}=v_i^{in}$, $\mathcal{L}^{(0)}=0$, $\mathcal{P}^{(0)}=0$, $\lambda^{(0)}=0$;
2: **repeat**
3: Solve $\mathcal{VL}$-subproblem:

$$\min_{\mathcal{V},\mathcal{L}}\left\{\frac{\alpha}{2}E_f+\frac{\beta}{2}E_{sh}+\frac{\eta}{2}E_{lap}+\frac{r}{2}\left\|\left(\mathcal{P}^{(k)}+\frac{\lambda^{(k)}}{r}\right)-\nabla\mathcal{L}\right\|^2\right\}$$

4: Solve $\mathcal{P}$-subproblem:

$$\min_{\mathcal{P}}\Big\{R(\mathcal{P})+\lambda^{(k)}\cdot\mathcal{P}+\frac{r}{2}\|\mathcal{P}-\nabla\mathcal{L}^{(k+1)}\|^2\Big\}$$

5: Update Lagrange multipliers λ:

$$\lambda^{(k+1)}=\lambda^{(k)}+r(\mathcal{P}^{(k+1)}-\nabla\mathcal{L}^{(k+1)})$$

6: **until** $\sum_{i=1}^{N}\|v_i^{(k+1)}-v_i^{(k)}\|^2<\epsilon$

6 Experiments

In this section we test our algorithm on several datasets including the latest MVS dataset (DTU Robot Image Data Sets [26]), models from [11], synthetic models, and Middlebury [56] as well. The algorithm is run on a PC of a 3.20 GHz Intel(R) Core(TM) i7 processor with 16 GB memory. The algorithm has a few parameters: penalty coefficient r, threshold ϵ, and three weights α, β and η. r and ϵ are set to 0.5 and $1.0e-10$ for all examples. The weights α, β and η are empirically set to 1.0, 0.01, 0.01 for DTU Robot Image Data Sets, $1.0e+6$, $5.0e+5$, 100 for the angel model from [11], $1.0e+5$, $1.0e+5$, 100 for the synthetic models, and $1.0e+6$, $1.0e+5$, 100 for Middlebury data sets. Note that the values of the parameters for the DTU data sets is quite different from others. This is because the vertex coordinates in the DTU data sets are of large values. The experiments are conducted to evaluate the algorithm and compare it with the spherical harmonics lighting based reconstruction [11] in several aspects: recovering surface details on DTU data sets which contain objects with uniform or non-uniform albedos, the requirement of the initial MVS mesh, and the effect of visual hull constraints. The computational cost and limitations of the algorithm are also discussed.

Recovering Surface Details on DTU Data Sets. DTU Robot Image Data Sets [26] include various types of models such as groceries, vegetables, building material and stuffed animals. We use DTU Data Sets to evaluate and compare our methods with the spherical harmonics lighting based approach [11] which is a representative shading based method for 3D reconstruction under general unknown illumination. since a publicly available implementation of [11] is unavailable, we re-implement their method by replacing our overall illumination model by the 4th order SH formulation for the incident illumination. The lighting estimation and the geometry refinement are performed alternately. The visibility maps are computed before optimization. Unless specified, the results of [11] shown in this paper are based on our implementation.

(a) Ground truth generated by structured light scanning (b) MVS results of [57] (c) Results of [11] (d) Our results

Fig. 4. Reconstruction of four objects (bird, hands, owl and budda) with uniform albedo.

First we choose four models from DTU data sets, which have uniform albedo. For each model, the MVS method [57] is used to generate a coarse mesh, which is then subdivided to generate the initial MVS mesh as the input to our optimization procedure. The reconstructed results by our method and [11] are shown in Fig. 4.

(a) Ground truth generated by structured light scanning (b) MVS results from [57] (c) Results of [11] (d) Our results

Fig. 5. Reconstruction of four objects (pilot, elephant, bunny, vegetables) with non-uniform albedo.

Second, we choose four models which have non-uniform albedos. We run our algorithm without any prior knowledge about the lighting conditions and surface albedos. The visual results are given in Fig. 5.

The leftmost columns of Figs. 4 and 5 show point cloud models provided by DTU Robot Image Data Sets and generated by structured light scanning. These point clouds can serve as the ground truth for us to perform quantitative evaluation. We project each point in the point clouds to the nearest face of the reconstructed mesh and compute the distance of the point to the face. The mean of the distances of all points to the mesh is defined to be the reconstruction

TABLE 1
Reconstruction Errors for the Input Meshes Generated by [57], the Reconstruction Meshes of [11], and Our Optimized Meshes

Reconstruction error (‰)	pilot	elephant	bunny	veg	bird	hands	owl	budda
input meshes	4.16	3.38	5.89	2.43	5.65	6.17	6.09	5.89
results of [11]	3.91	2.96	5.63	2.45	4.72	5.49	4.68	5.14
our results	**2.93**	**2.54**	**4.68**	**1.75**	**4.29**	**4.96**	**4.43**	**4.07**

The errors are normalized by the bounding box dimension of the meshes.

error. Table 1 reports the errors for the meshes generated by [57], [11] and our approach. It can be seen that our optimized meshes have much higher accuracy.

From Figs. 4 and 5, it can be seen that our overall illumination vectors based TV-minimization performs better than the SH lighting based shape refinement method [11]. Note that [11] is not meant to be applied on non-uniform albedo objects. Therefore some of the quantitative results are even worse than the input MVS models as shown in Table 1.

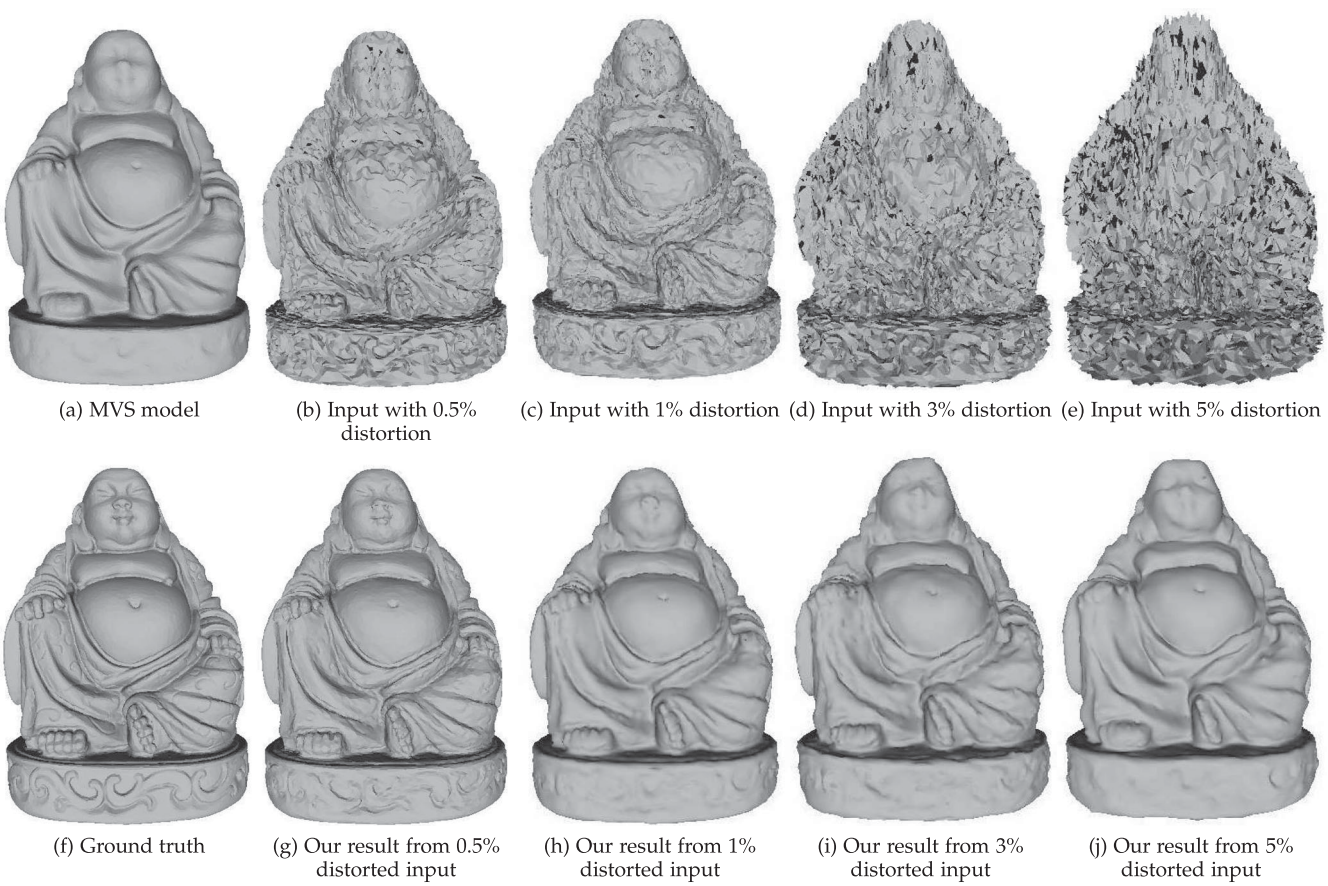
Requirement of the Initial MVS Mesh. While some previous methods such as [11] require a good initial MVS mesh, our method can accept a modest mesh as input. To test this, we generate the synthetic multi-view images by rendering a 3D model from 24 viewpoints. The MVS method of [57] and the subdivision are used to generate an initial model. Then we purposely perturb the initial model by giving random vertex displacement with magnitude up to 0.5 percent of the bounding box dimension (see Fig. 6b). The perturbed mesh is used as the input to our algorithm. The output and the ground truth are shown in Figs. 6g and 6f.

Furthermore, we increase perturbation ratios from 1 to 5 percent. As shown in Figs. 6c, 6d, and 6e, the quality of the input meshes decreases with higher perturbation ratio, but our method is still able to produce the reconstruction results Figs. 6h, 6i, and 6j with reasonable amount of surface details. Our experiment also shows that with 1 percent perturbation, the results of Wu et al.'s method [11] have already become poor.

Effect of Visual Hull Constraints. Fig. 7 shows two examples comparing the reconstruction without and with the visual hull constraints. Without the visual hull constraints, the reconstructed meshes on the left of Fig. 7 contain the artifacts of some prick-shape vertices. This is caused by the fact that the reconstruction system is partially under-constrained. With the help of visual hull constraints, the problem can be overcome and the artifacts disappear, as shown in Fig. 7(right).

Note that Wu et al.'s method [11] does not have visual hull constraints. It is not surprising that it sometimes suffers from the artifacts of prick-shape vertices, too. Fig. 8(left) shows such artifacts in a reconstructed model that is provided by one author of [11]. Using the visual hull constraints, with the same input our method can produce the reconstruction result that does not have such artifaces as shown in Fig. 8(right).

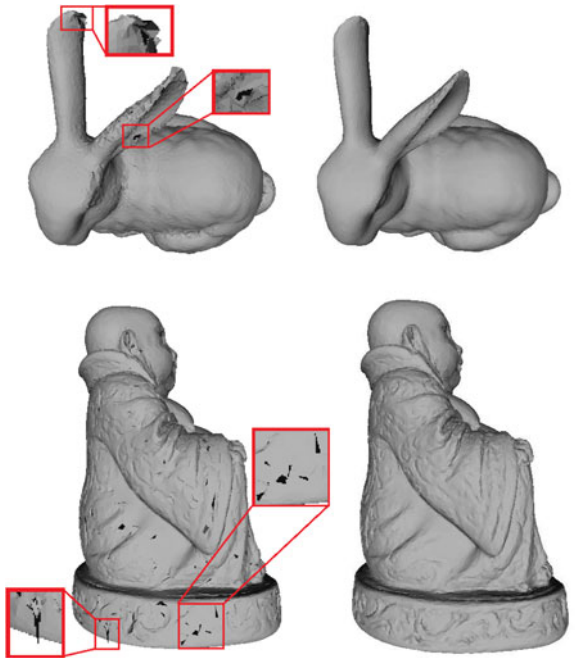
Test on Middlebury Data Sets. We also test our algorithm on datasets from [56]. We compare our results with those



(a) MVS model (b) Input with 0.5% distortion (c) Input with 1% distortion (d) Input with 3% distortion (e) Input with 5% distortion

(f) Ground truth (g) Our result from 0.5% distorted input (h) Our result from 1% distorted input (i) Our result from 3% distorted input (j) Our result from 5% distorted input

Fig. 6. Reconstruction of synthetic budda model using heavily distorted initial meshes.



(c) Results without visual hull constraints (d) Results with visual hull constraints

Fig. 7. The reconstructions without (left) and with (right) visual hull constraints.

generated from the volumetric MVS algorithm [29], which is one of the high-performance MVS methods. We use the MVS method of [57] and the subdivision to generate an initial MVS meshes and then optimize them using our variational procedure. We seek the standard Middlebury MVS evaluation and obtain the accuracy scores of 0.62 and 0.44 mm and the completeness scores of 98.1 and 99.3 percent for templeRing and dinoRing, respectively. These scores are reasonable but not at the top. We argue that the standard Middlebury MVS evaluation criteria are less in favor of recovering surface details, but focus more on the density of meshes. Note that our reconstructed models have much lower density than those produced by high score MVS methods, which is the main reason that we cannot obtain a higher completeness score in Middlebury benchmark.

Recovering Vertex Overall Illumination Vectors. Given the light probe images [43] and the synthetic models "Bunny" and "Budda", we can compute the ground truth vertex overall illumination vectors $L(v)$, as illustrated in Fig. 2. From the compute surface intensity values, we then recover the corresponding vertex overall illumination vectors $L(v)$ using the proposed framework. Table 2 shows the

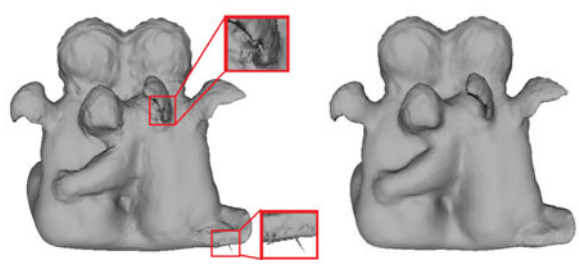


Fig. 8. Left: the reconstruction result of [11] without visual hull constraints; Right: the reconstruction result of our method with visual hull constraints.

TABLE 2
Quantitative Evaluation of Our Recovered $L(v)$ for the Two Synthetic Models Under the Four Light Probe Images Against the Ground Truth

	Bunny		Budda	
	angle (deg.)	mag. (%)	angle (deg.)	mag. (%)
beach	16.79	7.71	13.43	4.66
campus	15.98	6.86	12.04	6.05
kitchen	19.19	9.70	15.11	5.85
stpeters	17.14	3.89	13.99	3.39

First column: mean angle errors of $L(v)$ in degree. Second column: mean magnitude errors of $L(v)$.

quantitative results of the recovered $L(v)$, compared with the ground truth. It can be seen that our method can recover $L(v)$ at a certain degree of accuracy in terms of angles and magnitudes.

Computational Cost. In Algorithm 1, the vertices and their overall illumination vectors are optimized iteratively until convergence. In each iteration, the computational cost is mainly for the $\mathcal{VL}$-sub problem while the $\mathcal{P}$-subproblem and the λ updating can be solved very quickly. The Levenberg-Marquardt algorithm is used to solve the $\mathcal{VL}$-subproblem, which takes about $5 \sim 6$ minutes for the bunny dataset. The iteration is usually repeated about $2 \sim 3$ times. Thus, the time needed to solve the entire minimization problem is around 45 minutes to 2 hours with our unoptimized code. The runtime for DTU Robot Image Data Sets is listed in Table 3.

Limitations. Since our method treats the vertices and overall illumination vectors as variables at the same time, the memory cost is very high, which limits the number of vertices we can process. Another cause of high memory cost is due to the nonlinear representation of surface normal by vertices, which makes the Jacobian matrix for solving the nonlinear least squares problem huge. In our current setup, our unoptimized code can only handle meshes up to 300,000 vertices. One way to overcome the limitation is to use mesh face-based representation instead of the current vertex-based representation, where we can treat surface normals as unknowns directly, instead of converting them into a nonlinear combination of neighboring vertices. Another possible solution is to use a hierarchical approach with the divide-and-conquer strategy. These new approaches are worth further investigation.

7 Conclusions

This paper has presented a new algorithm for recovering surface details of an object from multi-view images

TABLE 3
Statistics of DTU Robot Image Data Sets in the Experiments: Model Sizes, Percentage of the Constrained Vertices from Visual Hull, and Runtime

# of vertices	pilot	elephant	bunny	veg	bird	hands	owl	budda
coarse model	26 K	28 K	50 K	40 K	50 K	40 K	50 K	40 K
refined model	90 K	100 K	70 K	120 K	160 K	120 K	160 K	140 K
VH constraint	8%	8%	6%	7%	5%	7%	5%	6%
runtime	50 min	1 h	45 min	1 h	2 h	2 h	2 h	2 h

captured under general unknown illuminations. The algorithm is based on a TV-minimization formulation which integrates MVS, shading cues and visual hull. The shape refinement and the overall illumination estimation are obtained simultaneously by solving the minimization problem using the proposed augmented Lagrangian method. It is shown that the algorithm can efficiently reconstruct geometric models with high-frequency surface details. Compared to the existing methods, our algorithm has less requirement on the initial MVS mesh, lighting conditions and object material properties such as albedos.

This work also suggests a few problems for future research. First, the paper introduces the concept of overall illumination vectors and has shown its advantage in SfS. It would be interesting to explore applications of this concept in other tasks. Second, in this paper the overall illumination vectors are regularized by total variation, which is not a strong regularizer. We would like to investigate how to impose further constraints on the overall illumination vectors to further enhance the reconstruction quality and speed. Finally, this paper just considers the reconstruction of Lambertian surfaces. It would be interesting to extend the proposed variational approach to handle objects with specular reflection, texture, etc., which are more common in practice [58], [59].

Acknowledgments

This research is partially supported by Singapore MoE AcRF Tier-1 Grants RG138/14, RG26/15, and the BeingTogether Centre, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The BeingTogether Centre is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative. We would also like to thank Dr. Chenglei Wu for providing some data of [11]. This work was presented in part in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 2014 [1]. Qi Duan and Jianfei Cai are with the corresponding authors.

References

[1] D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai, and T.-J. Cham, "Recovering surface details under general unknown illumination using shading and coarse multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1526–1533.

[2] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 519–528.

[3] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 690–706, Aug. 1999.

[4] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," in *Shape from Shading*, B. K. P. Horn and M. J. Brooks, Eds. Cambridge, MA, USA: MIT Press, 1989, pp. 513–531.

[5] C. H. Esteban, G. Vogiatzis, and R. Cipolla, "Multiview photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 548–554, Mar. 2008.

[6] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. S. Kweon, "Multiview photometric stereo using planar mesh parameterization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1161–1168.

[7] H. Jin, S. Soatto, and A. J. Yezzi, "Multi-view stereo reconstruction of dense shape and complex appearance," *Int. J. Comput. Vis.*, vol. 63, no. 3, pp. 175–189, 2005.

[8] N. Joshi and D. J. Kriegman, "Shape from varying illumination and viewpoint," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–7.

[9] H. Jin, D. Cremers, D. Wang, E. Prados, A. J. Yezzi, and S. Soatto, "3-D reconstruction of shaded objects from multiple images under unknown illumination," *Int. J. Comput. Vis.*, vol. 76, no. 3, pp. 245–256, 2008.

[10] C. Wu, Y. Liu, Q. Dai, and B. Wilburn, "Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 8, pp. 1082–1095, Aug. 2011.

[11] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt, "High-quality shape from multi-view stereo and shading under general illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 969–976.

[12] Y. Yoshiyasu and N. Yamazaki, "Topology-adaptive multi-view photometric stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1001–1008.

[13] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3D geometry," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 536–543, 2005.

[14] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, "High-quality single-shot capture of facial geometry," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 40:1–40:9, 2010.

[15] T. Beeler, et al., "High-quality passive facial performance capture using anchor frames," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 75.

[16] L. Valgaerts, C. Wu, A. Bruhn, H.-P. Seidel, and C. Theobalt, "Lightweight binocular facial performance capture under uncontrolled lighting," *ACM Trans. Graph.*, vol. 31, no. 6, 2012, Art. no. 187.

[17] D. Xu, J. Cai, T. J. Cham, P. Fu, and J. Zhang, "Kinect-based easy 3D object reconstruction," in *Proc. 13th Pacific-Rim Conf. Advances Multimedia Inf. Process.*, 2012, pp. 476–483.

[18] S. Xiong, J. Zhang, J. Zheng, J. Cai, and L. Liu, "Robust surface reconstruction via dictionary learning," *ACM Trans. Graph.*, vol. 33, 2014, Art. no. 201.

[19] K.-J. Yoon, E. Prados, and P. Sturm, "Joint estimation of shape and reflectance using multiple images with known illumination conditions," *Int. J. Comput. Vis.*, vol. 86, pp. 192–210, 2010.

[20] S. Haque, A. Chatterjee, and V. M. Govindu, "High quality photometric reconstruction using a depth camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2283–2290.

[21] Y. Han, J.-Y. Lee, and I. S. Kweon, "High quality shape from a single RGB-D image under uncalibrated natural illumination," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1617–1624.

[22] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2553–2560.

[23] G. Choe, J. Park, Y.-W. Tai, and I. S. Kweon, "Exploiting shading cues in kinect IR images for geometry refinement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3922–3929.

[24] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt, "Real-time shading-based refinement for consumer depth cameras," *ACM Trans. Graph.*, vol. 33, 2014, Art. no. 200.

[25] P.-P. Sloan, J. Kautz, and J. Snyder, "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 527–536, 2002.

[26] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, "Large scale multi-view stereopsis evaluation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 406–413.

[27] G. Vogiatzis, C. H. Esteban, P. H. S. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2241–2246, Dec. 2007.

[28] S. N. Sinha, P. Mordohai, and M. Pollefeys, "Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[29] P. Song, X. Wu, and M. Y. Wang, "Volumetric stereo and silhouette fusion for image-based modeling," *Vis. Comput.*, vol. 26, no. 12, pp. 1435–1450, 2010.

[30] A. Zaharescu, E. Boyer, and R. Horaud, "TransforMesh: A topology-adaptive mesh-based approach to surface evolution," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 166–175.

[31] R. Tylecek and R. Sara, "Refinement of surface mesh for accurate multi-view reconstruction," *Int. J. Virtual Reality*, vol. 9, no. 1, pp. 45–54, 2010.

[32] P. Merrell, et al., "Real-time visibility-based fusion of depth maps," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[33] N. D. F. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, "Using multiple hypotheses to improve depth-maps for multi-view stereo," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 766–779.
[34] J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang, "Bundled depth-map merging for multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2769–2776.
[35] M. Habbecke and L. Kobbelt, "A surface-growing approach to multi-view stereo reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
[36] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, Aug. 2010.
[37] R. Basri, D. W. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *Int. J. Comput. Vis.*, vol. 72, no. 3, pp. 239–257, 2007.
[38] A. Hertzmann and S. M. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying BRDFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1254–1264, Aug. 2005.
[39] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and spatially-varying BRDFs from photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1060–1071, Jun. 2010.
[40] N. G. Alldrin, T. Zickler, and D. J. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[41] S. R. Richter and S. Roth, "Discriminative shape from shading in uncalibrated illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1128–1136.
[42] J. T. Kajiya, "The rendering equation," in *Proc. 13th Annu. Conf. Comput. Graph. Interactive Techn.*, 1986, pp. 143–150.
[43] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. 25th Annu. Conf. Comput. Graph. Interactive Techn.*, 1998, pp. 189–198.
[44] P. Debevec, "A median cut algorithm for light probe sampling," in *Proc. ACM SIGGRAPH Classes*, 2008, Art. no. 33.
[45] N. Dyn, D. Levin, and J. A. Gregory, "A butterfly subdivision scheme for surface interpolation with tension control," *ACM Trans. Graph.*, vol. 9, no. 2, pp. 160–169, 1990.
[46] O. Sorkine, "Laplacian mesh processing," in *Eurographics—State of the Art Reports*, Y. Chrysanthou and M. Magnor, Eds. Lyon, France: Eurographics Assoc., 2005.
[47] C. Wu, J. Deng, F. Chen, and X.-C. Tai, "Scale-space analysis of discrete filtering over arbitrary triangulated surfaces," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 670–709, 2009.
[48] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
[49] C. H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," *Comput. Vis. Image Understanding*, vol. 96, no. 3, pp. 367–392, 2004.
[50] S. N. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, vol. 1, pp. 349–356.
[51] K. Kolev and D. Cremers, "Integration of multiview stereo and silhouettes via convex functionals on convex domains," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 752–765.
[52] Y. Furukawa and J. Ponce, "Carved visual hulls for image-based modeling," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 53–67, 2009.
[53] C. Wu, J. Zhang, Y. Duan, and X.-C. Tai, "Augmented lagrangian method for total variation based image restoration and segmentation over triangulated surfaces," *J. Sci. Comput.*, vol. 50, no. 1, pp. 145–166, 2012.
[54] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, no. 2, pp. 164–168, 1944.
[55] D. Marquardt, "An algorithm for the least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[56] "Middlebury multi-view stereo evaluation," (2006). [Online]. Available: http://vision.middlebury.edu/mview/eval/
[57] M. Jancosek and T. Pajdla, "Multi-view reconstruction preserving weakly-supported surfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 3121–3128.
[58] J. Sun, M. Smith, L. Smith, S. Midha, and J. Bamber, "Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities," *Image Vis. Comput.*, vol. 25, no. 7, pp. 1050–1057, 2007.
[59] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich, "Transparent and specular object reconstruction," *Comput. Graph. Forum*, vol. 29, pp. 2400–2426, 2010.

Di Xu received the BS and PhD degrees from Nanyang Technological University, Singapore, in 2011 and 2016, respectively. He is currently a research scientist in NetVirta Singapore, working on cutting-edge 3D vision technology. His research interests include computer vision, virtual reality, and computer graphics.

Qi Duan received the BS degree from National University of Defense Technology, in 2004, the MS degree from Shanghai Jiao Tong University, in 2008, and the PhD degree from Nanyang Technological University, in 2013. He is currently manager of the visualization technology team in Shanghai United-Imaging Healthcare Co. Ltd. His research interests include computer graphics and computer vision related area, especially in largescale image data visualization, and scene reconstruction.

Jianmin Zheng received the BS and PhD degrees from Zhejiang University, China. He is an associate professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His recent research focuses on T-spline technologies, digital geometric processing, reality computing, 3D vision, interactive digital media, and applications. He has published more than 130 technical papers in international conferences and journals. He was the conference co-chair of Geometric Modeling and Processing 2014 and has served on the program committee of many international conferences. He is an associate editor of the *Visual Computer*.

Juyong Zhang received the BS degree from the University of Science and Technology of China, in 2006 and the PhD degree from Nanyang Technological University, Singapore, in 2012. He is currently an associate professor in the School of Mathematical Sciences, University of Science and Technology of China. His research interests include computer graphics, computer vision, and image processing.

Jianfei Cai (S'98-M'02-SM'07) received the PhD degree from the University of Missouri-Columbia. He is currently an associate professor and has served as the Head of Visual & Interactive Computing Division and the head of Computer Communication Division, School of Computer Engineering, Nanyang Technological University, Singapore. His major research interests include computer vision, visual computing, and multimedia networking. He has published more than 180 technical papers in international journals and conferences. He has been actively participating in program committees of various conferences. He has served as the leading technical program chair for IEEE International Conference on Multimedia & Expo (ICME) 2012 and the leading general chair for Pacific-rim Conference on Multimedia (PCM) 2012. Since 2013, he has been serving as an associate editor of the *IEEE Transactions on Image Processing*. He has also served as an associate editor of the *IEEE Transactions on Circuits and Systems for Video Technology* from 2006 to 2013. He is a senior member of the IEEE.

Tat-Jen Cham received the BA degree in engineering and the PhD degree both from the University of Cambridge, in 1993 and 1996, respectively. He is an associate professor in the School of Computer Engineering, Nanyang Technological University and a principal investigator in the NRF BeingThere Centre for 3D Telepresence. His research interests include broadly in computer vision and machine learning. He received the Jesus College Research Fellowship in Cambridge 1996-97 and was a research scientist in the DEC / Compaq Research Lab in Boston 1998-2001. While in NTU, he was concurrently a Singapore-MIT Alliance Fellow 2003-2006, the director for the Centre of Multimedia and Network Technology 2007-2015 and an NTU Senator 2010-2014. He is on the editorial board of the *International Journal of Computer Vision*, and was a general co-chair of the Asian Conference on Computer Vision 2014.