A Differential Volumetric Approach to Multi-View Photometric Stereo

Fotios Logothetis[1,2],   Roberto Mecca[2],   Roberto Cipolla[1]

fl302@cam.ac.uk, roberto.mecca@crl.toshiba.co.uk, rc10001@cam.ac.uk

[1]Department of Engineering, University of Cambridge, United Kingdom
[2]Toshiba Research, Cambridge, United Kingdom

Abstract

Highly accurate 3D volumetric reconstruction is still an open research topic where the main difficulty is usually related to merging some rough estimations with high frequency details. One of the most promising methods is the fusion between multi-view stereo and photometric stereo images. Beside the intrinsic difficulties that multi-view stereo and photometric stereo in order to work reliably, supplementary problems arise when considered together.

In this work, we present a volumetric approach to the multi-view photometric stereo problem. The key point of our method is the signed distance field parameterisation and its relation to the surface normal. This is exploited in order to obtain a linear partial differential equation which is solved in a variational framework, that combines multiple images from multiple points of view in a single system. In addition, the volumetric approach is naturally implemented on an octree, which allows for fast ray-tracing that reliably alleviates occlusions and cast shadows.

Our approach is evaluated on synthetic and real datasets and achieves state-of-the-art results.

1. Introduction

Recovering the 3D geometry of an object is still a quite open challenge in computer vision as most of the techniques provide good results in specific frameworks only. In particular, two well-known approaches namely multi-view (MVS) and photometric stereo (PS) have been developed to produce great results considering some key complementary assumptions. Indeed, while MVS is assumed to provide rough 3D volumetric reconstructions of textured objects, PS is supposed to retrieve highly detailed surfaces from a single view. High quality volumetric reconstruction of objects achieved by refining coarse multi-view reconstruction [10, 40] with shading information [14, 47, 42] is a classical way [6] of combining complementary information.

Multi-View Photometric Stereo (MVPS) approaches have been developed so as to overcome constraints coming from both sides, in order to deal with: specular highlights
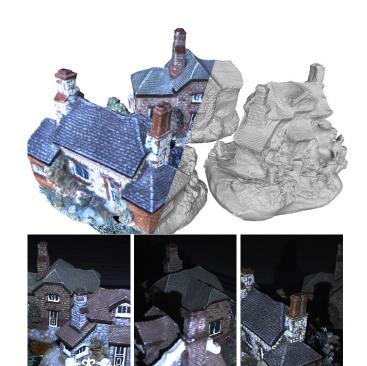

Figure 1. Top: 3D reconstruction of a three (toy) house scene with and without albedo, respectively left and right side. Bottom row: sample images. The scene scale is indicated by the coin on left bottom image. Our approach is capable of dealing with the near-field effects (point light sources, out-of-focus shapes (*middle*)), occlusions and cast shadows.

[16, 1], dynamic scenes [44], visibility and occlusions [7] and mapping of the PS views onto the volume [37, 35].

Since implicit parameterisation of volumes has been developed using level-set approaches [27, 34], recent advances in parameterising volumes with signed distance functions (SDFs) [53, 32] have made the multi-view approach prone to be merged with differential formulation of irradiance equation providing shading information [26]. On the other hand, recent photometric stereo approaches have moved towards more realistic assumptions considering point light sources [29, 36] that make the acquisition

process easier by using LEDs in a calibrated setting.

In this work, our aim is to propose a new MVPS approach capable of dealing with scene having complex geometry when acquiring images in the near-field. This increases the difficulty of the reconstruction problem due to the severe visibility issues given by occlusions, cast shadows, out-of-focus regions, etc. (see Figure 1).

Contribution To do so, our novel approach is based on

- A differential parameterisation of the volume, based on the signed distance function that allows irradiance equation ratios to deal with near-field Photometric Stereo modelling.

- A variational optimisation that fuses information from multiple viewpoints into a single system.

- An octree implementation allowing for quick raytracing so as highly accurate volumetric reconstructions can be obtained in scenes with multiple objects.

2. Related Works

Reconstructing accurate 3D geometry of a volume has been a challenging area in computer vision. Most of the research trying to solve this problem has been developed by merging multi-view methods for coarse reconstruction [10], with techniques based on shading information for providing high frequency details of the surface [30, 46, 48, 31, 5, 8] rather than based on a topological evolution of the surface [16]. However, regarding the refinement, several methods take inspiration from Shape from Shading [14] to extract 3D geometry from a single image (MVSfS) and consider shape refinement resulting from single shading cues [50, 49, 3]. With the aim to improve the quality of the details and make the reconstruction more robust to outliers, multiple shading images from a single view point are considered. A number of MVPS approaches have been presented [12, 35, 52].

Merging shading information with multi-view images becomes a more complicated problem when considering specular surfaces. Drastic changes in both shading under different lighting and the viewing point modify the appearance of the 3D geometry so that specific approaches have been developed to deal with irradiance equations with not negligible specular component. Jin *et al.* [16] exploit a rank constraint on the radiance tensor field of the surface in space with the aim to fit the Ward reflectance model [45]. Other approaches instead reconstructed an unknown object by using a radiance basis inferred from reference objects [43, 1]. Zhou *et al.* [52] developed a camera and a handheld moving light system for firstly capturing sparse 3D points and then refining the depth along iso-depth contours [2]. A similar handheld system has been developed by Higo *et al.* [13]

where multi-view images were acquired under varying illumination by a handled camera with a single movable LED point light source for reconstructing a static scene.

In order to make the MVPS solvable, additional assumptions have been considered. In particular, with the aim to compute the camera positions so as to map accurately the photometric stereo views, the relative motion of the camera and the object can be constrained. Hernandez *et al.* [12] captured multi-view images for a moving object under varying illuminations by combining shading and silhouettes assuming circular motion in order to compute the visual hull. Zhang *et al.* [51] generalised optical flow, photometric stereo, multiview-stereo and structure from motion techniques assuming rigid motion of the object under orthographic viewing geometry and directional lighting. Furthermore, shadows and occlusions are not considered.

When photometric stereo (as well as SfS) has to be integrated with multi-view techniques, the problem of finding the correspondence of pixels with shading information onto the 3D surface is crucial. Geometric distortions produced by changes in pose have to be combined with varying illumination. One way to do so is by region tracking considering brightness variations using optical flow [11], parametric models of geometry and illumination [9], or outlier rejection [15]. Okatani and Deguchi [33] proposed a photometric method for estimating the second derivatives of the surface shape of an object when only inaccurate knowledge of the surface reflectance and illumination is given by assuming represented in a probabilistic fashion.

Other approaches instead align the shading images with the coarse 3D in order to map the photometric stereo data onto the 3D shape [21, 17]. Delaunoy and Prados [7] use a gradient flow approach whereas Sabzevari *et. al* [37] firstly computes a 3D mesh with structure from motion with a low percentage of missing point and then the mesh is reprojected onto a plan using a mapping scheme [22]. Recently, Park *et al.* [35] proposed a refinement method by computing an optical displacement map in the same 2D planar domain of the photometric stereo images. So, they transformed the coarse 3D mesh into a parameterised 2D space using a parameterization technique that reduces distortions [41].

In this work, with the aim to avoid the mapping procedure, we present a differential approach for MVPS. Being inspired by the signed distance function parameterisation used by Maier *et al.* [26] for the MVSfS problem, we derive a volumetric parameterisation which handles the differential irradiance equation ratio presented in [29] for near-field photometric stereo. The problem is posed in a 3D domain which in practice is implemented in a octree, which allows for fast ray-tracing. This accelerates the computation of cast shadows (that are similarly conceived as in [23]) and occlusions from different views and makes it possible to generate sub-milimeter precision models for scenes occupying a vol-

ume of several liters.

3. Signed Distance Function Parameterisation

With the aim to provide suitable mathematical characterisation of a collection of solid objects, we consider the implicit surface parameterisation in terms of the SDF $d(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3$. This parameterisation turns out to be suitable for our aim due to its practical way of describing the outgoing normal vector to a surface. In fact, the SDF allows to describe the volumetric surface as the zeroth level-set of $d, d(\mathbf{x}) = 0$. The essence of our differential approach is the observation that the surface normal n equals to the gradient of the SDF d as follows

$$\mathbf{n}(\mathbf{x}) = \nabla d(\mathbf{x}). \qquad (1)$$

Similarly to [53] that used the SDF for single image shading refinement, we consider the SDF for the irradiance equation to derive a differential multi-view photometric stereo formulation where we assume to have N_{ps} images (i.e. light sources) for each known camera position C_q (that is $N_{ps}(C_q), q = 1, \ldots N_{views}$).

To exploit the monocular aspect of the photometric stereo problem, we consider image ratios for the Lambertian shading model [20] assuming calibrated nearby LED light sources

$$i_k(\mathbf{u}(\mathbf{x})) = \rho(\mathbf{x})a_k(\mathbf{x})\overline{\mathbf{n}}(\mathbf{x}) \cdot \bar{\mathbf{l}}_k(\mathbf{x}) \qquad (2)$$

where $\mathbf{u} \in \mathbb{R}^2$ is the image-plane projection of the 3D point $\mathbf{x}$ and $\rho(\mathbf{x})$ indicates the albedo. Note that as we are following a volumetric approach, the irradiance equation is considered for a set of 3D points $\mathbf{x}$. The bar over a vector means that it is normalized (i.e. $\overline{\mathbf{n}} = \frac{\mathbf{n}}{|\mathbf{n}|}$). We model point light sources by considering the following $\mathbf{l}_k(\mathbf{x}) = \mathbf{x} - \mathbf{p}_k$ from [29], where $\mathbf{p}_k$ is the known position of the point light source with respect to the global coordinate system. We model the light attenuation considering the following nonlinear radial model of dissipation

$$a_k(\mathbf{x}) = \phi_k \frac{(\mathbf{l}_k(\mathbf{x}) \cdot \mathbf{s}_k)^{\mu_k}}{||\mathbf{l}_k(\mathbf{x})||^2} \qquad (3)$$

where ϕ_k is the intrinsic brightness of the light source, $\mathbf{s}_k$ is the principal direction (i.e. the orientation of the LED point light source) and μ_k is an angular dissipation factor.

Modeling with image ratios As in [29], we follow the ratio method that significantly simplifies the PS problem by eliminating the dependence on the albedo as well as the non-linear normalisation of the normal.

Indeed, dividing equations for images i_h and i_k (from the same point of view C_q) as in (2), we have

$$\frac{i_h(\mathbf{x})}{i_k(\mathbf{x})} = \frac{a_h(\mathbf{x})\mathbf{n}(\mathbf{x}) \cdot \bar{\mathbf{l}}_h(\mathbf{x})}{a_k(\mathbf{x})\mathbf{n}(\mathbf{x}) \cdot \bar{\mathbf{l}}_k(\mathbf{x})} \qquad (4)$$

which leads to

$$\mathbf{n}(\mathbf{x}) \cdot (i_h(\mathbf{x})a_k(\mathbf{x})\bar{\mathbf{l}}_k(\mathbf{x}) - i_k(\mathbf{x})a_h(\mathbf{x})\bar{\mathbf{l}}_h(\mathbf{x})) = 0. \qquad (5)$$

By substituting the parametrisation of the normal from (1), we get the following albedo independent, homogeneous linear PDE

$$\mathbf{b}_{hk}(\mathbf{x}) \cdot \nabla d(\mathbf{x}) = 0 \qquad (6)$$

where

$$\mathbf{b}_{hk}(\mathbf{x}) = i_h(\mathbf{x})a_k(\mathbf{x})\bar{\mathbf{l}}_k(\mathbf{x}) - i_k(\mathbf{x})a_h(\mathbf{x})\bar{\mathbf{l}}_h(\mathbf{x}). \qquad (7)$$

The geometrical meaning of (6) is the extension to the 3D volumetric reconstruction of the PDE approach presented in [29] and the proposed PS model still consists of a homogeneous linear PDE. However, an important difference with [29] is that $\mathbf{b}_{hk}(\mathbf{x})$ does not depend on d (i.e. (6) is linear and not quasi-linear as proposed in [29]) due to the fact that the relevant quantities are expressed on a global coordinate system independent of the existence of a surface. Crucially, this allows to use the nonlinear lighting model of (3) without linearising approximations (e.g. spherical harmonics [4]) or dependence on any initial surface estimates.

An interesting observation is that (6) is conceptually similar with the iso-depth curves in the work of [52]. Nonetheless, the SDF formulation is a more natural 'object centered' depth and this allows for a unified optimisation as we describe in the next section.

In order to simplify the notation, we will rename the pair hk as p and we will call the set of all the combinations of pairs of images (with no repetition).

MVPS as a weighted least squares problem With the aim to consider photometric stereo images coming from different views into a single mathematical framework, we stack in a single system the following weighted version of (6)

$$w_p(C_q, \mathbf{x})\mathbf{b}_p(\mathbf{x}) \cdot \nabla d(\mathbf{x}) = 0 \qquad (8)$$

where $w_p(C_q, \mathbf{x}) = \max(\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}_q(\mathbf{x}), 0)$. This weight term w_p is essentially a measure of visibility. The resulting system then counts $\sum_{q=1}^{N_{views}} \binom{N_{ps}(C_q)}{2}$ equations as shown in (9)

$$\begin{bmatrix} [w_1(C_1, \mathbf{x})\mathbf{b}_1(\mathbf{x})]^t \\ [w_2(C_2, \mathbf{x})\mathbf{b}_2(\mathbf{x})]^t \\ \vdots \end{bmatrix} \nabla d(\mathbf{x}) = \mathbf{0}. \qquad (9)$$

With the aim to solve it as a least square problem, we consider the normal equations

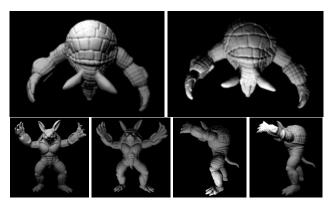$$B(\mathbf{x})\nabla d(\mathbf{x}) = 0 \qquad (10)$$

Figure 2. Synthetic data samples where we show per pairs differences in near lighting, perspective deformation and self occlusions.

with

$$B(\mathbf{x}) = [w_1(C_1, \mathbf{x})\mathbf{b}_1(\mathbf{x}), w_2(C_2, \mathbf{x})\mathbf{b}_2(\mathbf{x}), \ldots] \cdot$$
$$\begin{bmatrix} [w_1(C_1, \mathbf{x})\mathbf{b}_1(\mathbf{x})]^t \\ [w_2(C_2, \mathbf{x})\mathbf{b}_2(\mathbf{x})]^t \\ \vdots \end{bmatrix}$$

$B(\mathbf{x})$ is now a positive, semi-definite, 3x3 matrix.

B adjustment The geometrical constraint coming from (6) ensures that all the vector fields $\mathbf{b}_p(\mathbf{x}) \in \mathbb{R}^3$ span the same bi-dimensional space $\forall \mathbf{x}$ of the volume as they define the level-set of the SDF. This means that under ideal circumstances, the rank of B in (10) should be exactly 2. However, due to numerical approximations this is never exactly true; we enforce this constraint by using eigenvalue decomposition of B hence

$$B = Q\Lambda Q^{\mathrm{t}} = Q \begin{bmatrix} \Lambda_1 & 0 & 0 \\ 0 & \Lambda_2 & 0 \\ 0 & 0 & \Lambda_3 \end{bmatrix} Q^{\mathrm{t}} \qquad (11)$$

with $\Lambda_1 \geq \Lambda_2 \geq \Lambda_3$ and setting $\Lambda_3 = 0$.

We note that this rank correction is a sanity check step. Indeed if $B\nabla d = 0$ with B full rank, then $\nabla d = 0$ which can never be true as $|\nabla d| = 1$ (Eikonal equation) and so d cannot be the SDF of any real surface. In practice, this would lead to over-smoothing of the SDF and loss of details. In addition, the Eikonal equation can be implicitly enforced by demanding that $\nabla d = \mathbf{q}_3$, where $\mathbf{q}_3$ is the third collumn of Q and corresponds to the nullspace of B ($\mathbf{q}_3$ is a unit vector hence enforcing the Eikonal equation). Hence (10) is updated to the following full rank system ($I_{3\times3}$ is the identity matrix)

$$\left(B(\mathbf{x}) + I_{3\times3}\right)\nabla d(\mathbf{x}) = B'(\mathbf{x})\nabla d(\mathbf{x}) = \mathbf{q}_3. \qquad (12)$$
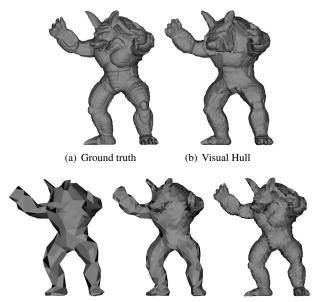
4. Variational resolution

In this section, we describe how we build the variational solver to compute the signed distance field based parameterisation introduced in the previous section.

Discretisation To avoid excessive computation, we note that the photometric stereo equations do not need to be computed in the whole volume but rather only to a subset of voxels $\Omega \subset \mathbb{R}^3$, which are close to the surface. In fact, (1) is only true in the vicinity of the surface. We discretise the variational problem (12) by using first order forward finite differences $\nabla\mathbf{d} = G \cdot \mathbf{d}$, with $\mathbf{d}$ being the vector stacking $d(\mathbf{x})$ in all of Ω. G is the sparse kernel matrix describing the connectivity in Ω.

We note that the differential approach is inevitably rank deficient (needing an integration constant for the theoretical minimal solution). Hence, we follow the standard of most modern variational approaches (e.g. [28]) and adopt a Tikhonov regulariser of the form $\mathbf{d} = \mathbf{d}_0$ where $\mathbf{d}_0$ is some initial estimate of the SDF obtained from the distance transform of the an initial surface estimate. Also note that differential approaches are guaranteed to recover smooth surfaces without the need of a smoothness regularisers (like the one of [8]). Thus the problem becomes (using $\lambda = 0.05$)

$$\min_{\mathbf{d}}(||B_\Omega G\mathbf{d} - \mathbf{q}_\Omega|| + \lambda||\mathbf{d} - \mathbf{d}_0||) \qquad (13)$$

where B_Ω and $\mathbf{q}_\Omega$ are obtained by stacking the relevant quantities from (12) for all voxels in Ω. The resulting linear system is solved with the conjugate gradients method.

(a) Ground truth   (b) Visual Hull

(c) Initial estimates: 500,1500,10K triangles

Figure 3. Synthetic data experiment - initial estimates used for initialising the MVPS optimisation. The variable quality initial estimates of the bottom row are generated by subsampling the ground truth (3(a)) using Meshlab's edge collapse decimation function.

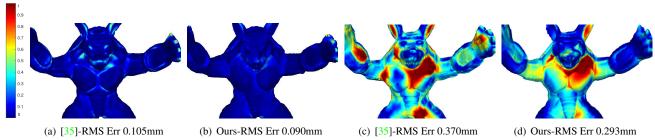(a) [35]-RMS Err 0.105mm   (b) Ours-RMS Err 0.090mm   (c) [35]-RMS Err 0.370mm   (d) Ours-RMS Err 0.293mm

Figure 4. Evaluation using the 1500 triangles mesh initial estimate (a),(b) and the visual hull initial estimate (c),(d). The colour coding shows the error in millimeters compared to the ground truth (computed using Meshlab's Hausdorff distance function).

Jacobi preconditioner is used as the system is too large ($10^7 - 10^9$ elements) to use a more sophisticated one.

4.1. Octree Implementation

To manage the required set of voxels $\in \Omega$ described above we use an octree structure. Ω is defined at the leafs of the tree and Voxel neighbors for computing finite differences are found by bottom up traversal of the tree.

We perform an iterative procedure of solving (13) on the leafs of the tree and then subsequently subdividing those leafs where the absolute value of SDF is smaller than 2 voxel sizes. The procedure repeats until the voxels are small enough so as their projection on the image planes is smaller than the pixel size and thus the maximum obtainable resolution has been reached. As a result, only a small fraction of the volume is considered for calculations and the hierarchy of voxels is densely packed around the surface. Finally, the reconstructed surface is computed with the Marching cubes variant of [19].

It is important to note however that this iterative procedure is only needed for computational reasons. If the whole volume could be filled with voxels, solving (13) would recover the whole surface in a single step.

Visibility Estimation In order to deal with scenes with a complex geometry and multiple objects, occlusions and cast shadows need to be addressed. This is performed by raytracing lines from each voxel to each light source and camera and using the current estimate of the geometry. As it is well accepted in the graphics community (e.g [25]), octree

structures allows for very quick visibility checks and whenever an occlusion/shadow is detected, the relevant weight in (9) is set to 0.

The details of the tree evolution and raytracing operations are elaborated in the supplementary §A.

The method was implemented in Matlab with the octree code using c++ in a mex file.

5. Experiments

With the aim to prove the capability of our approach to reconstruct complex scenes, we considered both synthetic and real data. We compared against [35] using the code from their website. It is worth to mention that differently from our method, their state-of-the-art approach for MVPS is based on a fully un-calibrated PS model.

For the synthetic case, we used the Armadillo model from the Stanford 3D Scanning Repository[1]. The virtual object was scaled to have approximate radius 20mm and the virtual camera of focal length 6mm was placed in several locations on a sphere of 45mm around the object. We rendered 12 views with 8 images each of resolution 1200x800x24bits per pixel (see Figure 2).

In order to quantitatively describe the dependency of the accuracy of the volumetric reconstruction to the initial estimate, we subsampled the initial mesh[2] to 5 different meshes

---

[1]http://graphics.stanford.edu/data/3Dscanrep/
[2]Using the quadric edge collapse decimation function of Meshlab.

Table 1. Quantitative evaluation based on the initial estimate quality. Errors are in mm. Noise added to vertex positions and the magnitude is relative to the average triangle size.

Experiment		Triangle Number					Visual Hull
Method	Noise	250	500	1500	10K	30K	14K
[35]	0%	0.245	0.141	0.105	0.029	0.025	0.370
	5%	0.290	0.172	0.119	0.036	0.029	-
	10%	0.393	0.250	0.153	0.046	0.031	-
Proposed	0%	0.203	0.114	0.090	0.026	0.023	0.293
	5%	0.234	0.137	0.104	0.033	0.024	-
	10%	0.321	0.193	0.131	0.043	0.028	-

Figure 5. Hardware setup used for the acquisition. Only 9/52 LEDs are used for the acquisitions. The stereo configuration allows to get a wider view from the camera on the right whereas the camera in the center of the PCB is able to acquire near-field views of the objects.
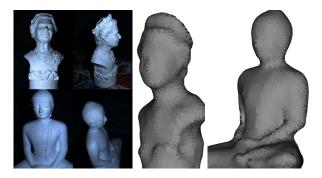
Figure 6. Real data: 2/108 photometric stereo images (we used 12 views with 9 lights in each view) and initial geometry estimate obtained with MVS. This initial estimates are only 8k and 11k triangles for the Queen and Buddha datasets respectively.

with number of triangle ranging from 250 to 30K (the original mesh was 150k triangles). For each of these meshes we added Gaussian noise to the vertex coordinates with std 0, 5, 10% of the average triangle size. The visual hull (computed with voxel carving) was also used for a final experiment.

The evaluation metric is the RMS Hausdorff distance to the ground truth (computed with Meshlab). Results are shown in Figures 4 and Table 1. The proposed approach outperforms [35] in all experiments with the difference being more significant on the low quality initial estimate experiments.

5.1. Real Data

For acquiring real world data we used an active light system (see Figure 5) consisting of two FLIR cameras FL3-U3-32S2C-CS. One camera mounted an 8mm lens and was surrounded by OSRAM ultra bright LEDs for capturing data in the near-field. The second camera had a 4mm lens for acquiring a wider area helping to track the trajectories of both cameras. The stereo pair was calibrated using Matlab's computer vision toolbox and a checkrboard, in order to be able to capture the absolute scale of the world, which is needed for the near light modeling (Equation (3)).

The images have been acquired while moving the setup around the scene. We used COLMAP-SFM [38, 39] to process multi-view data to get camera rotation and translation

Figure 7. Qualitative evaluation on real data set of Figure 6. The proposed approach outperforms [35] and generates more detailed reconstructions.
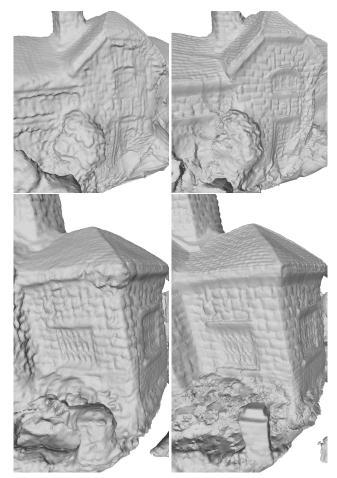

Figure 8. Closeup rednerings showing comparison of [35] (*left*) to the proposed method (*right*). Note that the proposed method is clearly outperforming the competition, especially in the top row which corresponds to a view in the middle of the scene which is particularly challenging due to cast shadows (see Figure 1).

between the photometric stereo views as well as a low quality reconstruction to use as initial estimate. In addition, a few more images were captured in between the photometric stereo sequences (with neutral illumination) in order to make SFM more robust with respect to a too small overlap between images. To make the models obtained through MVS have less noise, we remove some noisy regions and background points far away from the scenes of interest. Then, we performed Poisson reconstruction [18] with a low level setting so as the initial estimate contains continues surfaces (and not point clouds). As Table 1 suggests, our method does not need a very accurate initial estimate. Finally, the initial SFD d_0 is computed as the distance transform of the initial surface.

We performed minimal prepossessing to remove saturated and almost saturated pixels as they likely correspond to specular highlights, which are inconsistent with the Lambertian shading model assumed.

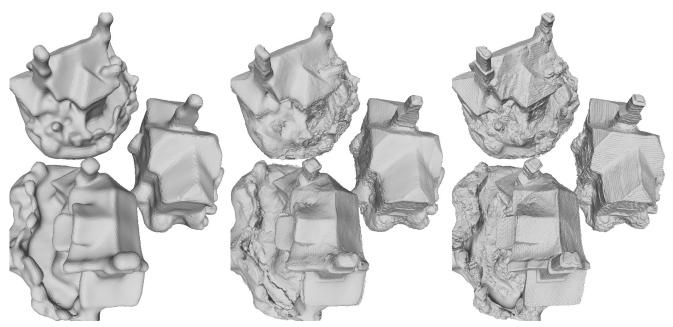Our real-world datasets include a marble Buddha statue,

Figure 9. Top view of the reconstruction of the 'village' dataset presented in Figure 1. This includes the initial geometry estimate obtained with SFM([38, 39] (*left*), [35] (*middle*) and the proposed method (*right*).

plaster bust of Queen Elisabeth (Figure 6), a plastic 3D printed version of the Armadillo next to an owl statue (Figure 10) as well as toy village scene (Figure 1).

The experiments were run on a server machine. The computational requirements were 15-20GB of RAM and 10-20 minutes of CPU time for the synthetic experiments and the single object datasets showed in Figure 6. Figure 1 and 10 correspond to much bigger volumes and thus they needed around 2 hours and 150GB of RAM; this cost is justified by the fact that the octree was allowed to grow to around 30M voxels of size 0.2mm which is required to cover the few litters volume.

The proposed approach outperforms [35] in all datasets. In the simple objects (see Figures 7) the difference is that our method is able to get more detailed surfaces as it is not limited by the initial estimate and iteratively refines the surface to the limit of precision. This is in contrast to [35] which is limited by the initial estimates as it is used to create a 2D domain and the calculate a displacement map. Note

that using [35] in an iterative fashion is explicitly discouraged by the authors as the 2D parameterisation gets very expensive if the initial estimate has a high triangle count. In addition, their 2D domain is set to a fixed resolution and thus cannot generate surfaces with arbitrary level of details. This is in contrast to our octree implementation which naturally allows continues refinement and different level of precision in different parts of the scene.

In addition, our methods performs very well on a challenging datasets where the existence of cast shadows highly degrades the quality of the reconstructions generated by [35] (see Figures 8, 9, 11 and 12). This is expected as

Figure 10. Sample images (3/90) from the Armadillo-owl dataset. The height of the armadillo is around 15cms.
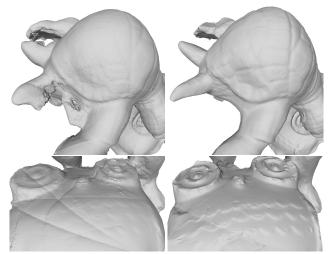

Figure 11. Closeup rednerings showing comparison of [35] (*left*) to the proposed method (*right*) for the dataset of Figure 10.

Figure 12. Top view of the reconstruction of the armadillo dataset. This includes the initial geometry estimate obtained with SFM([38, 39] (*left*), [35] (*middle*) and the proposed method (*right*).

our method ray-traces cast shadows whereas [35] only indirectly handles them using a robust estimation. Finally, albedo colourised reconstructions can be seen in Figure 13.

6. Conclusion

We presented the first volumetric parameterisation based on the signed distance function for the MVPS problem. Very high accuracy is achieved by using an octree implementation for processing and ray-tracing the volume on a tree. While considering photometric stereo images, our fully differential formulation is albedo independent as it uses the irradiance equation ratio approach for the near-field photometric stereo presented in [29].

The main limitation of the proposed approach is the inability to cope with missing big portions of scene (this also true for most competing approaches e.g. [35, 52, 50]). This is because Equation (1) only applies in the vicinity of a sur-

Figure 13. Albedo-mapped reconstruction of the 2 multi-object datasets. The albedo is calculated using simple least squares on Equation (2) using the final geometry estimate (along with occlusion/shadow maps). Black albedo regions signify that this particular part of the surface is not visible in any image.

face so the geometry cannot move very far away from initial estimate. In addition, the initial estimate is used for calculating the initial occlusion/shadow maps hence if an object that casts a shadow on the scene is not included in the initial estimate, the reconstruction of the shaded part will be problematic.

The main drawback of our method compared to mesh parameterisation techniques (e.g. [35]) is the elevated memory requirements. Even though the octree implementation minimises the number of voxels required, it is inevitable to need a few voxels per each potential surface point. This is due to the fact that the surface is the zero crossings of the SDF and at least a pair of opposite signed values are required per surface point. In addition, the use of the variational optimisation is also memory expensive as the matrix encoding the neighbouring information about voxels (G in Equation (13)) needs to be stored in memory as well.

As future work, the image ratio based modeling can be extended in order to handle specular highlights using the model presented in [28]. This requires to enhance the variational solver with the inclusion of a shininess parameter, as an additional unknown per voxel. Additional realistic effects such as ambient light ([24]) can also be included in the proposed model.

References

[1] Jens Ackermann, Fabian Langguth, Simon Fuhrmann, Arjan Kuijper, and Michael Goesele. Multi-view photometric stereo by example. In *2014 2nd International Conference on 3D Vision (3DV)*, volume 1, pages 259–266, 2014. 1, 2

[2] Neil G Alldrin and David J Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified

photometric stereo approach. In *2007 IEEE International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. 2

[3] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(8):1670–1687, 2015. 2

[4] Ronen Basri and David W. Jacobs. Photometric stereo with general, unknown lighting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 374–381, Kauai, USA, 2001. 3

[5] Mate Beljan, Jens Ackermann, and Michael Goesele. Consensus multi-view photometric stereo. In *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*, pages 287–296, 2012. 2

[6] Andrew Blake, Andrew Zisserman, and Greg Knowles. Surface descriptions from stereo and shading. *Image Vision Comput.*, 1985. 1

[7] Amael Delaunoy and Emmanuel Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3D reconstruction problems dealing with visibility. *International Journal of Computer Vision (IJCV)*, 95(2):100–123, 2011. 1, 2

[8] Kaiwen Guo, Feng Xu, Tao Yu, Xiaoyang Liu, Qionghai Dai, and Yebin Liu. Real-time geometry, albedo, and motion reconstruction using a single rgb-d camera. *ACM Transactions on Graphics (ToG)*, 36(3), June 2017. 2, 4

[9] Gregory D. Hager and Peter N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(10):1025–1039, 1998. 2

[10] Richard Harltey and Aandrew Zisserman. *Multiple view geometry in computer vision (2. ed.)*. Cambridge University Press, 2006. 1, 2

[11] Horst W. Haussecker and David J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):661–673, 2001. 2

[12] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(3):548–554, 2008. 2

[13] Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. A hand-held photometric stereo camera for 3-D modeling. In *2009 IEEE International Conference on Computer Vision (ICCV)*, pages 1234–1241, 2009. 2

[14] Berthold K. P. Horn. Obtaining shape from shading information. *The Psychology of Computer Vision, Winston, P. H. (Ed.)*, pages 115–155, 1975. 1, 2

[15] Hailin Jin, Paolo Favaro, and Stefano Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV*, 2001. 2

[16] Hailin Jin, Stefano Soatto, and Anthony J. Yezzi. Multi-view stereo beyond Lambert. In *2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–I, 2003. 1, 2

[17] Neel Joshi and David J. Kriegman. Shape from varying illumination and viewpoint. In *2007 IEEE International Conference on Computer Vision (ICCV)*, pages 1–7, 2007. 2

[18] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 61–70, 2006. 6

[19] Michael Kazhdan, Allison Klein, Ketan Dalal, and Hugues Hoppe. Unconstrained isosurface extraction on arbitrary octrees. In *Eurographics Symposium on Geometry Processing*, 2007. 5

[20] Johann Heinrich Lambert. *Photometria sive De mensura et gradibus luminis, colorum et umbrae*. Sumptibus viduae Eberhardi Klett, typis Christophori Petri Detleffsen, 1760. 3

[21] Jongwoo Lim, Jeffrey Ho, Ming-Hsuan Yang, and David J. Kriegman. Passive photometric stereo from motion. In *2005 IEEE International Conference on Computer Vision (ICCV)*, pages 1635–1642, 2005. 2

[22] Ligang Liu, Lei Zhang, Yin Xu, Craig Gotsman, and Steven J. Gortler. A local/global approach to mesh parameterization. *Comput. Graph. Forum*, 2008. 2

[23] Fotios Logothetis, Roberto Mecca, and Roberto Cipolla. Semi-calibrated near field photometric stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 3, page 8, 2017. 2

[24] Fotios Logothetis, Roberto Mecca, Yvain Quéau, and Roberto Cipolla. Near-field photometric stereo in ambient light. In *British Machine Vision Conference (BMVC)*, 2016. 8

[25] David Luebke, Martin Reddy, Jonathan D. Cohen, Amitabh Varshney, Benjamin Watson, and Robert Huebner. *Level of Detail for 3D Graphics*. Morgan Kaufmann, 2003. 5

[26] Robert Maier, Kihwan Kim, Daniel Cremers, Jan Kautz, and Matthias Niessner. Intrinsic3D: High-quality 3D Reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 1, 2

[27] Ravikanth Malladi, James A. Sethian, and Baba C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(2):158–175, 1995. 1

[28] Roberto Mecca, Yvain Quéau, Fotios Logothetis, and Roberto Cipolla. A single lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4):1858–1888, 2016. 4, 8

[29] Roberto Mecca, Aaron Wetzler, Alfred M. Bruckstein, and Ron Kimmel. Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770, 2014. 1, 2, 3, 8

[30] Yael Moses and Ilan Shimshoni. 3d shape recovery of smooth surfaces: Dropping the fixed-viewpoint assumption. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(7):1310–1324, 2009. 2

[31] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3):536–543, 2005. 2

[32] Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. Real-time 3d reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32(6):169, 2013. 1

[33] Takayuki Okatani and Kaichiro Deguchi. Optimal integration of photometric and geometric surface measurements using inaccurate reflectance/illumination knowledge. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 254–261, 2012. 2

[34] Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer New York, 2006. 1

[35] Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Robust multiview photometric stereo using planar mesh parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 39(8):1591–1604, 2017. 1, 2, 5, 6, 7, 8

[36] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, Francois Lauze, and Jean-Denis Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(3):313–340, 2018. 1

[37] Reza Sabzevari, Alessio Del Bue, and Vittorio Murino. Multi-view photometric stereo using semi-isometric mappings. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, 2012. 1, 2

[38] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 6, 7, 8

[39] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016. 6, 7, 8

[40] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 519–528, 2006. 1

[41] Alla Sheffer, Emil Praun, and Kenneth Rose. Mesh parameterization methods and their applications. *Foundations and Trends in Computer Graphics and Vision*, 2006. 2

[42] Boxin Shi, Zhe Mo, Zhipeng Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 1

[43] Adrien Treuille, Aaron Hertzmann, and Steven M. Seitz. Example-based stereo with general BRDFs. In *8th European Conference on Computer Vision (ECCV)*, pages 457–469, 2004. 2

[44] Daniel Vlasic, Pieter Peers, Ilya Baran, Paul E. Debevec, Jovan Popovic, Szymon Rusinkiewicz, and Wojciech Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics*, 28(5), 2009. 1

[45] Gregory J. Ward. Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics*, 26(2):265–272, 1992. 2

[46] Martin Weber, Andrew Blake, and Roberto Cipolla. Towards a complete dense geometric and photometric reconstruction under varying pose and illumination. In *British Machine Vision Conference (BMVC)*, 2002. 2

[47] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):134–144, 1980. 1

[48] Chenglei Wu, Yebin Liu, Qionghai Dai, and Bennet Wilburn. Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. *IEEE transactions on visualization and computer graphics*, 17(8):1082–1095, 2011. 2

[49] Chenglei Wu, Kiran Varanasi, Yebin Liu, Hans-Peter Seidel, and Christian Theobalt. Shading-based dynamic shape refinement from multi-view video under general illumination. In *2011 IEEE International Conference on Computer Vision (ICCV)*, pages 1108–1115, 2011. 2

[50] Chenglei Wu, Michael Zollhöfer, Matthias Nieß ner, Marc Stamminger, Shahram Izadi, and Christian Theobalt. Real-time shading-based refinement for consumer depth cameras. *ACM Transactions on Graphics (TOG)*, 33(6):200, 2014. 2, 8

[51] Li Zhang, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multiview stereo. In *2003 IEEE International Conference on Computer Vision (ICCV)*, pages 618–625, 2003. 2

[52] Zhenglong Zhou, Zhe Wu, and Ping Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *2013 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 1482–1489, 2013. 2, 3, 8

[53] Michael Zollhöfer, Angela Dai, Matthias Innmann, Chenglei Wu, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics (TOG)*, 34(4):96, 2015. 1, 3