

## Multiview Photometric Stereo

Carlos Hernández, *Member, IEEE*,  
George Vogiatzis, *Member, IEEE*, and  
Roberto Cipolla, *Member, IEEE*

**Abstract**—This paper addresses the problem of obtaining complete, detailed reconstructions of textureless shiny objects. We present an algorithm which uses silhouettes of the object, as well as images obtained under changing illumination conditions. In contrast with previous photometric stereo techniques, ours is not limited to a single viewpoint but produces accurate reconstructions in full 3D. A number of images of the object are obtained from multiple viewpoints, under varying lighting conditions. Starting from the silhouettes, the algorithm recovers camera motion and constructs the object's visual hull. This is then used to recover the illumination and initialize a multiview photometric stereo scheme to obtain a closed surface reconstruction. There are two main contributions in this paper: First, we describe a robust technique to estimate light directions and intensities and, second, we introduce a novel formulation of photometric stereo which combines multiple viewpoints and, hence, allows closed surface reconstructions. The algorithm has been implemented as a practical model acquisition system. Here, a quantitative evaluation of the algorithm on synthetic data is presented together with complete reconstructions of challenging real objects. Finally, we show experimentally how, even in the case of highly textured objects, this technique can greatly improve on correspondence-based multiview stereo results.

**Index Terms**—Photometric stereo, multiple views, light calibration, silhouette.

## 1 INTRODUCTION

DIGITAL archiving of 3D objects is a key area of interest in cultural heritage preservation. While laser range scanning is one of the most popular techniques, it has a number of drawbacks, namely, the need for specialized, expensive hardware and also the requirement of exclusive access to an object for significant periods of time. Also, for a large class of shiny objects such as porcelain or glazed ceramics, 3D scanning with lasers is challenging [1]. Recovering 3D shape from photographic images is an efficient, cost-effective way to generate accurate 3D scans of objects.

Several solutions have been proposed for this long-studied problem. When the object is well textured, its shape can be obtained by densely matching pixel locations across multiple images and triangulating [2], however, the results typically exhibit high frequency noise.

Alternatively, photometric stereo is a well-established technique which uses the shading cue and can provide very detailed, but partial 2.5D reconstructions [3].

In this paper, we propose an elegant and practical method for acquiring a *complete* and *accurate* 3D model from a number of images taken around the object, captured under changing light conditions (see Fig. 1). The changing (but otherwise unknown) illumination conditions uncover the fine geometric detail of the object surface which is obtained by a generalized photometric stereo scheme.

The object's reflectance is assumed to follow Lambert's law, i.e., points on the surface keep their appearance constant irrespective of

viewpoint. The method can, however, tolerate isolated specular highlights, typically observed in glazed surfaces such as porcelain. We also assume that a single, distant light source illuminates the object and that it can be changed arbitrarily between image captures. Finally, it is assumed that the object can be segmented from the background and silhouettes extracted automatically.

## 2 RELATED WORK

This paper addresses the problem of shape reconstruction from images and is therefore related to a vast body of computer vision research. We draw inspiration from the recent work of [4] where the authors explore the possibility of using photometric stereo with images from multiple views when correspondence between views is not initially known. Picking an arbitrary viewpoint as a reference image, a depth-map with respect to that view serves as the source of approximate correspondences between frames. This depth-map is initialized from a Delaunay triangulation of sparse 3D features located on the surface. Using this depth-map, their algorithm performs a photometric stereo computation obtaining normal directions for each depth-map location. When these normals are integrated, the resulting depth-map is closer to the true surface than the original. The paper presents high quality reconstructions and gives a theoretical argument justifying the convergence of the scheme. The method, however, relies on the existence of distinct features on the object surface which are tracked to obtain camera motion and initialize the depth-map. In the class of textureless objects we are considering, it may be impossible to locate such surface features and, indeed, our method has no such requirement. Also, the surface representation is still depth-map-based and, consequently, the models produced are 2.5D.

A similar approach of extending photometric stereo to multiple views and more complex BRDFs was presented in [5] with the limitation of almost planar 2.5D reconstructed surfaces. Our method is based on the same fundamental principle of bootstrapping photometric stereo with approximate correspondences, but we use a general volumetric framework which allows complete 3D reconstructions from multiple views.

Quite related to this idea is the work of [6] and [7] where photometric stereo information is combined with 3D range scan data. In [6], the photometric information is simply used as a normal map texture for visualization purposes. In [7], a very good initial approximation to the object surface is obtained using range scanning technology which, however, is shown to suffer from high-frequency noise. By applying a fully calibrated 2.5D photometric stereo technique, normal maps are estimated which are then integrated to produce an improved, almost noiseless surface geometry. Our acquisition technique is different from [7] in the following respects: 1) We only use standard photographic images and simple light sources, 2) our method is fully uncalibrated—all necessary information is extracted from the object's contours, and 3) we completely avoid the time consuming and error prone process of merging 2.5D range scans.

The use of the silhouette cue is inspired by the work of [8] where a scheme for the recovery of illumination information, surface reflectance and geometry is described. The algorithm described makes use of frontier points, a geometrical feature of the object obtained by the silhouettes. Frontier points are points of the visual hull where two contour generators intersect and, hence, are guaranteed to be on the object surface. Furthermore, the local surface orientation is known at these points, which makes them suitable for various photometric computations such as extraction of reflectance and illumination information. Our method generalizes the idea by examining a much richer superset of frontier points which is the set of contour generator points. We overcome the difficulty of localizing contour generators by a robust random sampling strategy. The price we pay is that a considerably simpler reflectance model must be used.

- C. Hernández and G. Vogiatzis are with the Computer Vision Group, Toshiba Research Europe, 208 Cambridge Science Park, Cambridge, CB4 0GZ, UK. E-mail: {carlos.hernandez, george}@crl.toshiba.co.uk.
- R. Cipolla is with the University of Cambridge, Cambridge, CB2 1PZ, UK. E-mail: cipolla@eng.cam.ac.uk.

Manuscript received 25 May 2007; revised 18 Sept. 2007; accepted 27 Sept. 2007; published online 26 Oct. 2007.

Recommended for acceptance by Y. Sato.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2007-05-0306.

Digital Object Identifier no. 10.1109/TPAMI.2007.70820.

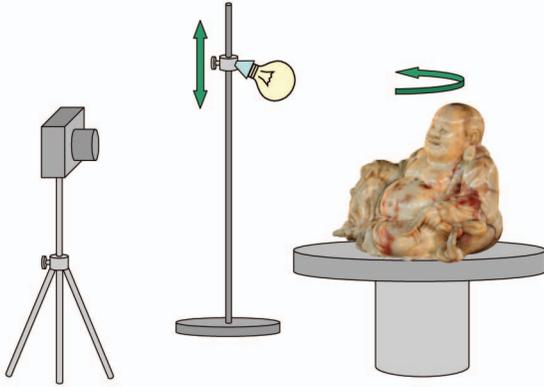


Fig. 1. Our acquisition setup. The object is rotated on a turntable in front of a camera and a point light source. A sequence of images is captured, while the light source changes position between consecutive frames. No knowledge of the camera or light source positions is assumed.

Although solving a different type of problem, the work of [9] is also highly related mainly because the class of objects addressed is similar to ours. While the energy term defined and optimized in their paper bears strong similarity to ours, their reconstruction setup keeps the lights fixed with respect to the object so, in fact, an entirely different problem is solved and, hence, a performance comparison between the two techniques is difficult. However, the results presented in [9] at first glance seem to be lacking in detail, especially in concavities, while our technique considerably improves on the visual hull. Finally, there is a growing volume of work on using specularities for calibrating photometric stereo (see [10] for a detailed literature survey). This is an example of a different cue used for performing uncalibrated photometric stereo on objects of the same class as the one considered here. However, methods proposed have so far only been concerned with the fixed view case.

### 3 ALGORITHM

In this paper, we reconstruct the complete geometry of 3D objects by exploiting the powerful silhouette and shading cues. We modify classic photometric stereo and cast it in a multiview framework where the camera is allowed to circumnavigate the object and illumination is allowed to vary. First, the object's silhouettes are used to recover camera motion using the technique presented in [11] and, via a novel robust estimation scheme, they allow us to accurately estimate the light directions and intensities in every image.

Second, the object surface, which is parameterized by a mesh and initialized from the visual hull, is evolved until its predicted appearance matches the captured images. The advantages of our approach are the following:

- It is fully uncalibrated: No light or camera pose calibration object needs to be present in the scene. Both camera pose and illumination are estimated from the object's silhouettes.
- The full 3D geometry of a complex, textureless multi-albedo object is accurately recovered, something not previously possible by any other method.
- It is practical and efficient as evidenced by our simple acquisition setup.

#### 3.1 Robust Estimation of Light-Sources from the Visual Hull

For an image of a Lambertian object with varying albedo, under a single distant light source, and assuming no self-occlusion, each surface point projects to a point of intensity given by

$$i = \lambda l^T \mathbf{n}, \quad (1)$$

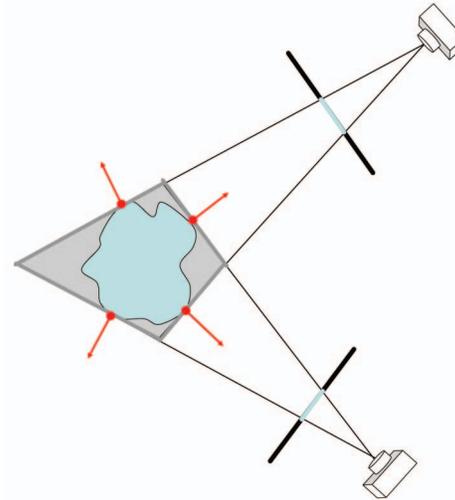


Fig. 2. The visual hull for light estimation. The figure shows a 2D example of an object which is photographed from two viewpoints. The visual hull (gray quadrilateral) is the largest volume that projects inside the silhouettes of the object. While the surface of the visual hull is generally quite far from the true object surface, there is a set of points where the two surfaces are tangent and, moreover, share the same local orientation (these points are denoted here with the four dots and arrows). In the full 3D case, three points, with their surface normals, are enough to fix an illumination hypothesis against which all other points can be tested for agreement. This suggests a robust random sampling scheme, described in the main text, via which the correct illumination can be obtained.

where  $l$  is a 3D vector directed toward the light source and scaled by the light source intensity,  $\mathbf{n}$  is the surface unit normal at the object location, and  $\lambda$  is the albedo at that location. Equation (1) provides a single constraint on the three coordinates of the product  $\lambda l$ . Then, given three points  $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c$  with an unknown but equal albedo  $\lambda$ , their normals (noncoplanar)  $\mathbf{n}_a, \mathbf{n}_b, \mathbf{n}_c$ , and the corresponding three image intensities  $i_a, i_b, i_c$ , we can construct three such equations that can uniquely determine  $\lambda l$  as

$$\lambda l = [\mathbf{n}_a \ \mathbf{n}_b \ \mathbf{n}_c]^{-1} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix}. \quad (2)$$

For multiple images, these same three points can provide the light directions and intensities in each image up to a global unknown scale factor  $\lambda$ . The problem is then how to obtain three such points.

Our approach is to use the powerful silhouette cue. The observation on which this is based is the following: When the images have been calibrated for camera motion, the object's silhouettes allow the construction of the *visual hull* [12], which is defined as the maximal volume that projects inside the silhouettes (see Fig. 2). A fundamental property of the visual hull is that its surface coincides with the real surface of the object along a set of 3D curves, one for each silhouette, known as *contour generators* [13]. Furthermore, for all points on those curves, the surface orientation of the visual hull surface is equal to the orientation of the object surface. Therefore, if we could detect points on the visual hull that belong to contour generators and have equal albedo, we could use their surface normal directions and projected intensities to estimate lighting. Unfortunately, contour generator points with equal albedo cannot be directly identified within the set of all points of the visual hull. Light estimation, however, can be viewed as robust model fitting where the inliers are the contour generator points of some constant albedo and the outliers are the rest of the visual hull points. The albedo of the inliers will be the *dominant* albedo, i.e., the color of the majority of the contour generator points. One can expect that the outliers do not generate consensus in favor of any particular illumination model, while the inliers do so in favor of the correct model. This observation motivates us to use a robust RANSAC scheme [14] to separate inliers

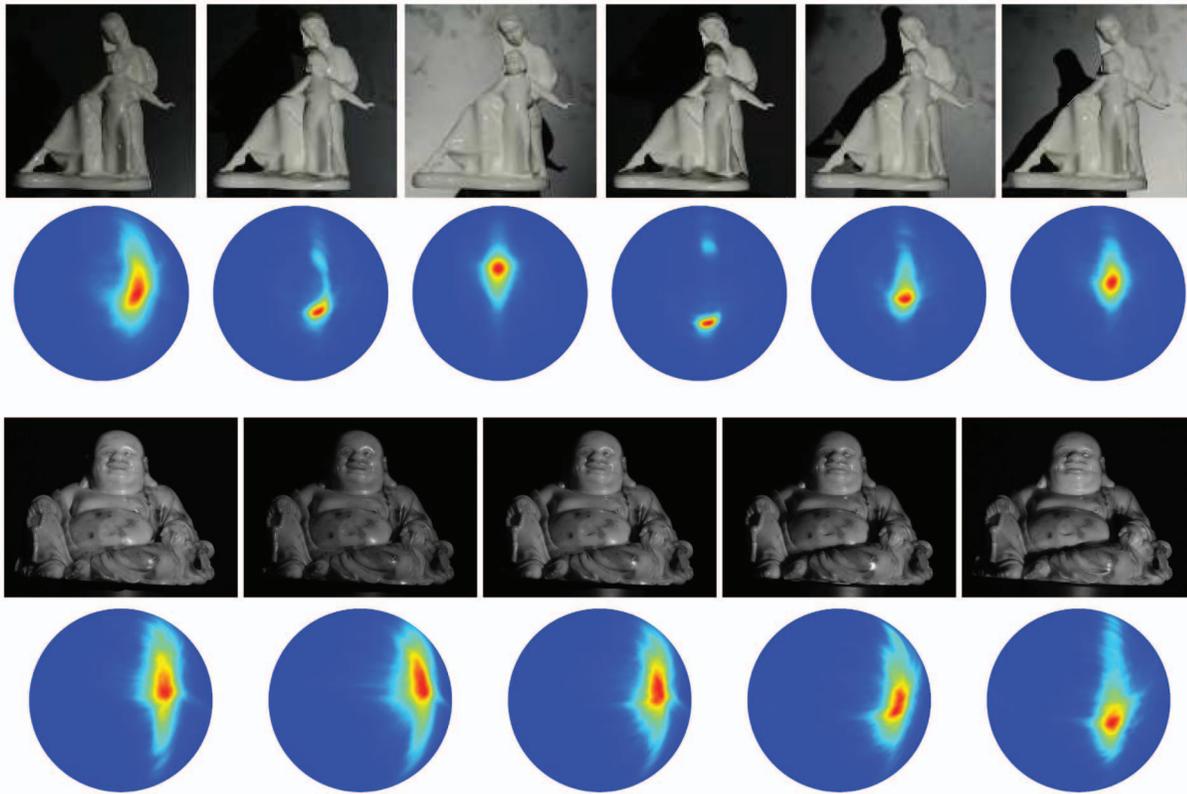


Fig. 3. Shape of illumination consensus. For different illumination configurations, we have plotted the consensus as a function of light direction. For each direction, consensus has been maximized with respect to light intensity. Red values denote big consensus. The shape of the maxima of this cost function as well as the lack of local optima implies a stable optimization problem. Top: Six different illuminations of a single albedo object. Bottom: Four different illuminations of a multialbedo object. Although the presence of multiple albedos degrades the quality of the light estimation (the peak is broader), it is still a clear single optimum.

from outliers and estimate illumination direction and intensity. The scheme can be summarized as follows:

1. Pick three points on the visual hull and from their image intensities and normals estimate an illumination hypothesis for  $\lambda$ .
2. Every point on the visual hull  $\mathbf{x}_m$  will now vote for this hypothesis if its predicted image intensity is within a given threshold  $\tau$  of the observed image intensity  $i_m$ , i.e.,

$$|\lambda \mathbf{l}^T \cdot \mathbf{n}_m - i_m| < \tau, \quad (3)$$

where  $\tau$  allows for quantization errors, image noise, etc.

3. Repeat 1 and 2 a set number of times always keeping the illumination hypothesis with the largest number of votes.

The shape of the actual function being optimized by the RANSAC scheme described above was explored graphically for a porcelain object in Fig. 3. The number of points voting for a light direction (maximized with respect to light intensity) was plotted as a 2D function of latitude and longitude of the light direction. These graphical representations, obtained for six different illuminations, show the lack of local optima and the presence of clearly defined maxima.

This simple method can also be extended in the case where the illumination is kept fixed with respect to the camera for  $K$  frames. This corresponds to  $K$  illumination vectors  $R_1 \mathbf{l}, \dots, R_K \mathbf{l}$ , where  $R_k$  are  $3 \times 3$  rotation matrices that rotate the fixed illumination vector  $\mathbf{l}$  with respect to the object. In that case, a point on the visual hull  $\mathbf{x}_m$  with normal  $\mathbf{n}_m$  will vote for  $\mathbf{l}$  if it is visible in the  $k$ th image where its intensity is  $i_{m,k}$  and

$$|\lambda (R_k \mathbf{l})^T \cdot \mathbf{n}_m - i_{m,k}| < \tau. \quad (4)$$

A point is allowed to vote more than once if it is visible in more than one image.

Even though, in theory, the single image case suffices for independently recovering illumination in each image, in our acquisition setup, light can be kept fixed over more than one frame. This allows us to use the extended scheme in order to further improve our estimates. A performance comparison between the single view and the multiple view case is provided through simulations with synthetic data in Section 4.

An interesting and very useful byproduct of the robust RANSAC scheme is that any deviations from our assumptions of a Lambertian surface of uniform albedo are rejected as outliers. This provides the light estimation algorithm with a degree of tolerance to sources of error such as highlights or local albedo variations. The next section describes the second part of the algorithm, which uses the estimated illumination directions and intensities to recover the object surface.

### 3.2 Multiview Photometric Stereo

Having estimated the distant light-source directions and intensities for each image, our goal is to find a closed 3D surface that is photometrically consistent with the images and the estimated illumination, i.e., its predicted appearance by the Lambertian model and the estimated illumination matches the images captured. To achieve this, we use an optimization approach where a cost function penalizing the discrepancy between images and predicted appearance is minimized.

Our algorithm optimizes a surface  $S$  that is represented as a mesh with vertices  $\mathbf{x}_1 \dots \mathbf{x}_M$ , triangular faces  $f = 1 \dots F$ , and corresponding albedo  $\lambda_1, \dots, \lambda_F$ . We denote by  $\mathbf{n}_f$  and  $A_f$  the mesh normal and the surface area at face  $f$ . Also, let  $i_{f,k}$  be the intensity of face  $f$  on image  $k$  and let the set  $\mathcal{V}_f$  be the set of images (subset of  $\{1, \dots, K\}$ ) from which face  $f$  is visible. The light direction and intensity of the  $k$ th image will be denoted by  $\mathbf{l}_k$ .

We use a scheme similar to the ones used in [9], [15], where the authors introduce a decoupling between the mesh normals  $\mathbf{n}_1 \dots \mathbf{n}_F$  and the direction vectors used in the Lambertian model equation. We call these new direction vectors  $\mathbf{v}_1 \dots \mathbf{v}_F$  *photometric normals* and they are independent of the mesh normals. The minimization cost is then composed of two terms, where the first term  $E_v$  links the photometric normals to the observed image intensities

$$E_v(\mathbf{v}_{1,\dots,F}, \lambda_{1,\dots,F}; \mathbf{x}_{1,\dots,M}) = \sum_{f=1}^F \sum_{k \in \mathcal{V}_f} (\mathbf{1}_k^T \lambda_f \mathbf{v}_f - i_{f,k})^2 \quad (5)$$

and the second term  $E_m$  brings the mesh normals close to the photometric normals through the following equation:

$$E_m(\mathbf{x}_{1,\dots,M}; \mathbf{v}_{1,\dots,F}) = \sum_{f=1}^F \|\mathbf{n}_f - \mathbf{v}_f\|^2 A_f. \quad (6)$$

This decoupled energy function is optimized by iterating the following two steps:

1. **Photometric normal optimization.** The vertex locations are kept fixed while  $E_v$  is optimized with respect to the photometric normals and albedos. This is achieved by solving the following independent minimization problems for each face  $f$ :

$$\mathbf{v}_f, \lambda_f = \arg \min_{\mathbf{v}, \lambda} \sum_{k \in \mathcal{V}_f} (\mathbf{1}_k^T \lambda \mathbf{v} - i_{f,k})^2 \text{ s.t. } \|\mathbf{v}\| = 1. \quad (7)$$

2. **Vertex optimization.** The photometric normals are kept fixed while  $E_m$  is optimized with respect to the vertex locations using gradient descent.

These two steps are interleaved until convergence, which takes about 20 steps for the sequences we experimented with. Typically, each iteration phase takes about 100 gradient descent iterations. Note that, for the first step described above, i.e., evolving the mesh until the surface normals converge to some set of *target* orientations, a variety of solutions is possible. A slightly different solution to the same geometric optimization problem has recently been proposed in [7], where the target orientations are assigned to each vertex rather than each face as we do here. That formulation lends itself to a closed-form solution with respect to the position of a single vertex. An iteration of these local vertex displacements yields the desired convergence. As both formulations offer similar performance, the choice between them should be made depending on whether the target orientations are given on a per vertex or per facet basis.

The visibility map  $\mathcal{V}_f$  is a set of images in which we can measure the intensity of face  $f$ . It excludes images in which face  $f$  is occluded using the current surface estimate as the occluding volume as well as images where face  $f$  lies in shadow. Shadows are detected by a simple thresholding mechanism, i.e., face  $f$  is assumed to be in shadow in image  $k$  if  $i_{f,k} < \tau_{shadow}$ , where  $\tau_{shadow}$  is a sufficiently low intensity threshold. Due to the inclusion of a significant number of viewpoints in  $\mathcal{V}_f$ , (normally at least four) the system is quite robust to the choice of  $\tau_{shadow}$ . For all of the experiments presented here, the value  $\tau_{shadow} = 5$  was used (for intensities in the range 0-255). As for the highlights, we also define a threshold  $\tau_{highlight}$  such as a face  $f$  is assumed to be on a highlight in image  $k$  if  $i_{f,k} > \tau_{highlight}$ . In order to compute  $\tau_{highlight}$ , we need to distinguish between single albedo objects and multi-albedo objects. Single albedo objects are easily handled since the light calibration step gives us the light intensity. Hence, under the Lambertian assumption, no point on the surface can produce an intensity higher than the light intensity, i.e.,  $\tau_{highlight} = \|\lambda\|$ . In the multi-albedo case,  $\lambda$  can also vary and it is likely that the albedo picked by the robust light estimation algorithm is not the brightest one present on the object. As a result, we prefer to use a global threshold to segment the highlights on the images. It is worth noting that this approach works for the porcelain objects

```

Capture images of object.
Extract silhouettes.
Recover camera motion and compute visual hull.
Estimate light directions and intensities in every image
(Section III-A).
Initialise a mesh with vertices  $\mathbf{x}_1 \dots \mathbf{x}_M$  and faces  $f = 1 \dots F$  to the object's visual hull.
while mesh-not-converged do
  Optimise  $E_v$  with respect to  $\mathbf{v}_1 \dots \mathbf{v}_F$  (5).
  Optimise  $E_m$  with respect to  $\mathbf{x}_1 \dots \mathbf{x}_M$  (6).
end while

```

Fig. 4. The multiview reconstruction algorithm.

because highlights are very strong and localized, so just a simple sensor saturation test is enough to find them, i.e.,  $\tau_{highlight} = 254$ .

## 4 EXPERIMENTS

The setup used to acquire the 3D model of the object is quite simple (see Fig. 1). It consists of a turntable onto which the object is mounted, a 60 W halogen lamp, and a digital camera. The object rotates on the turntable and 36 images (i.e., a constant angle step of 10 degrees) of the object are captured by the camera while the position of the lamp is changed. In our experiments, we have used three different light positions, which means that the position of the lamp was changed after 12 and again after 24 frames. The distant light source assumptions are satisfied if an object of 15 cm extent is placed 3-4 m away from the light.

The algorithm (see Fig. 4) was tested on five challenging shiny objects, two porcelain figurines shown in Fig. 5, two fine relief Chinese Qing-dynasty porcelain vases shown in Fig. 6, and one textured Jade Buddha figurine in Fig. 7. Thirty-six  $3,456 \times 2,304$  images of each of the objects were captured under three different illuminations. The object silhouettes were extracted by intensity thresholding and were used to estimate camera motion and construct the visual hull (second row of Fig. 5). The visual hull was processed by the robust light estimation scheme of Section 3.1 to recover the distance light-source directions and intensities in each image. The photometric stereo scheme of Section 3.2 was then applied. The results in Fig. 6 show reconstructions of porcelain vases with very fine relief. The reconstructed relief (especially for the vase on the right) is less than a millimeter while their height is approximately 15-20 cm. Fig. 7 shows a detailed reconstruction of a Buddha figurine made of polished Jade. This object is actually textured, which implies classic stereo algorithms could be applied. Using the camera motion information and the captured images, a state-of-the-art multiview stereo algorithm [16] was executed. The results are shown in the second row of Fig. 7. It is evident that, while the low frequency component of the geometry of the figurine is correctly recovered, the high frequency detail obtained by [16] is noisy. The reconstructed model appears bumpy even though the actual object is quite smooth. Our results do not exhibit surface noise while capturing very fine details such as surface cracks.

To quantitatively analyze the performance of the multiview photometric stereo scheme presented here with ground truth, an experiment on a synthetic scene was performed (Fig. 8). A 3D model of a sculpture (digitized via a different technique) was rendered from 36 viewpoints with uniform albedo and using the Lambertian reflectance model. The 36 frames were split into three sets of 12 and, within each set, the single distant illumination source was held constant. Silhouettes were extracted from the images and the visual hull was constructed. This was then used to estimate the illumination direction and intensity as described in Section 3.1. In 1,000 runs of the illumination estimation method for the synthetic scene, the mean light direction estimate was 0.75 degrees away from the true direction with a standard deviation of 0.41 degrees. The model

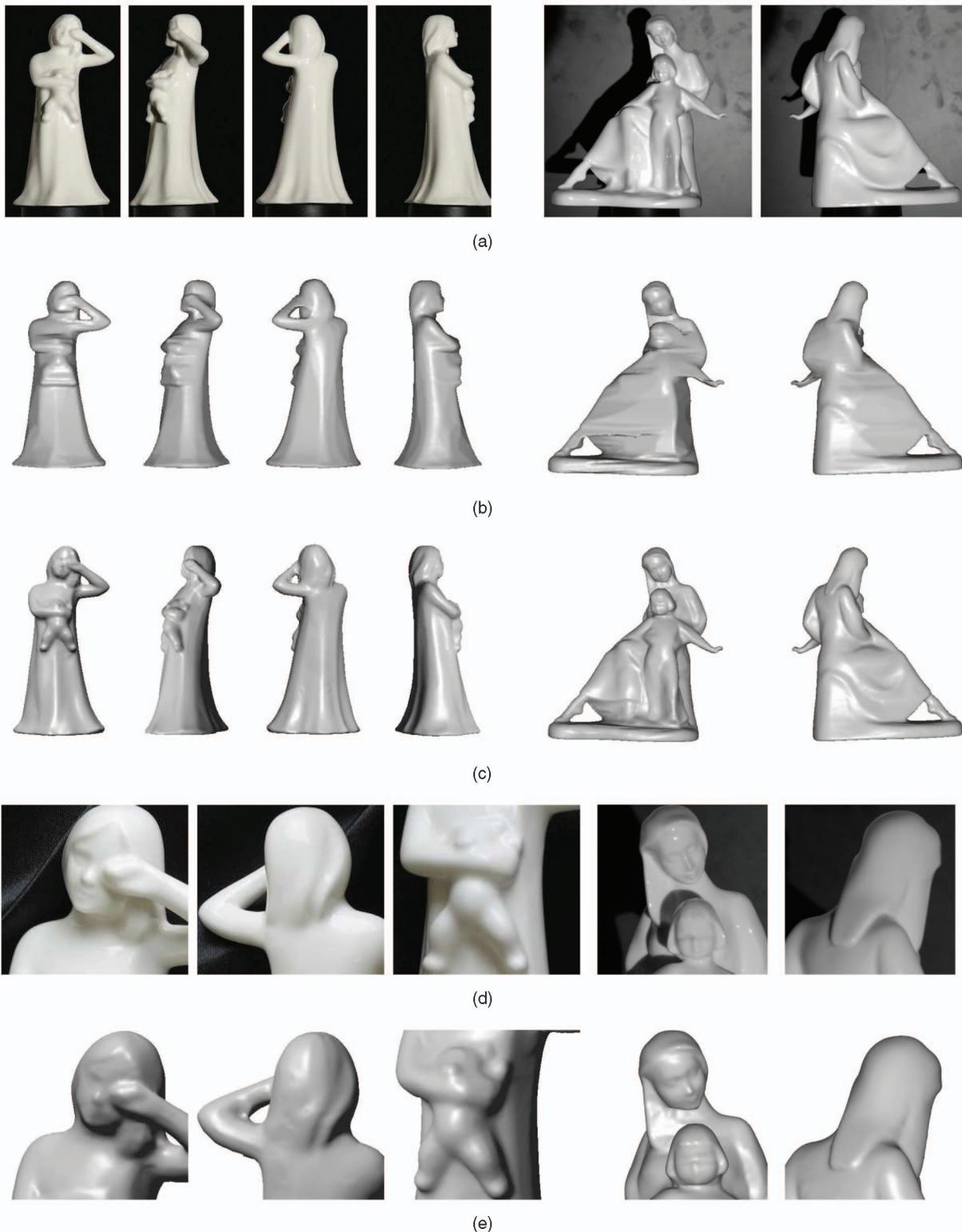


Fig. 5. Reconstructing porcelain figurines. Two porcelain figurines reconstructed from a sequence of 36 images each (some of the input images are shown in (a)). The object moves in front of the camera and illumination (a 60 W halogen lamp) changes direction twice during the image capture process. (b) shows the results of a visual hull reconstruction, while (c) shows the results of our algorithm. (d) and (e) show detailed views of the figurines and the reconstructed models respectively. (a) Input images. (b) Visual hull reconstruction. (c) Our results. (d) Close up views of porcelains. (e) Close up views of reconstructed models.

obtained by our algorithm was compared to the ground truth surface by measuring the distance of each point on our model from the closest point in the ground truth model. This distance was found to be about 0.5 mm when the length of the biggest diagonal of the bounding box volume was defined to be 1 m. Even though this result

was obtained from perfect noiseless images, it is quite significant since it implies that any loss of accuracy can only be attributed to the violations of our assumptions rather than the optimization methods themselves. Many traditional multiview stereo methods would not be able to achieve this due to the strong regularization that must be

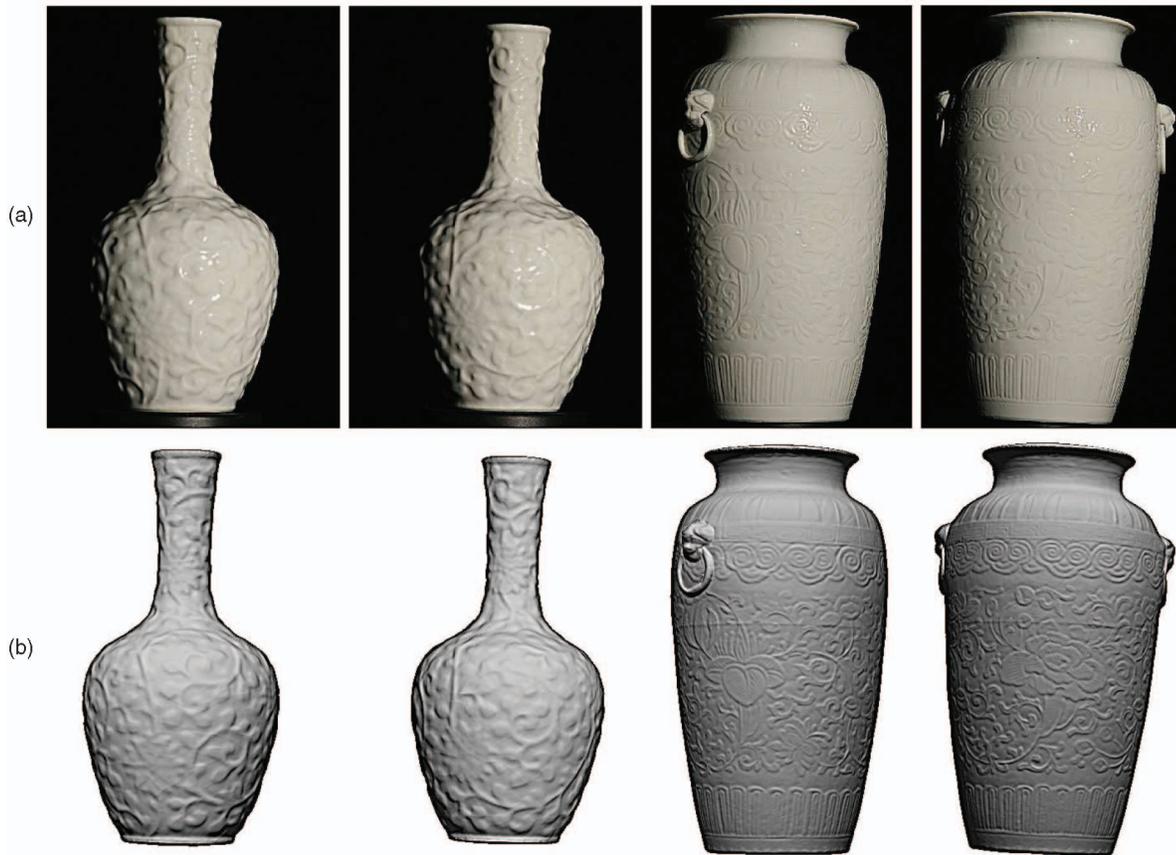


Fig. 6. Reconstructing Chinese Qing dynasty porcelain vases. (a) Sample of input images. (b) Proposed method. The resulting surface captures all of the fine details present in the images, even in the presence of strong highlights.



Fig. 7. Reconstructing colored jade. (a) Two input images. (b) Model obtained by multiview stereo method from [16]. (c) Proposed method. The resulting surface is filtered from noise, while new high frequency geometry is revealed (note the reconstructed surface cracks in the middle of the figurine's back).

imposed on the surface. By contrast, our method requires no regularization when faced with perfect noiseless images.

Finally, we investigated the effect of the number of frames during which illumination is held constant with respect to the camera frame. Our algorithm can, in theory, obtain the illumination direction and intensity in every image independently.

However, keeping the lighting fixed over two or more frames and supplying that knowledge to the algorithm can significantly improve estimates. The next experiment was designed to test this improvement by performing a light estimation over  $K$  images where the light has been kept fixed with respect to the camera. The results are plotted in Fig. 8b and show the improvement of the

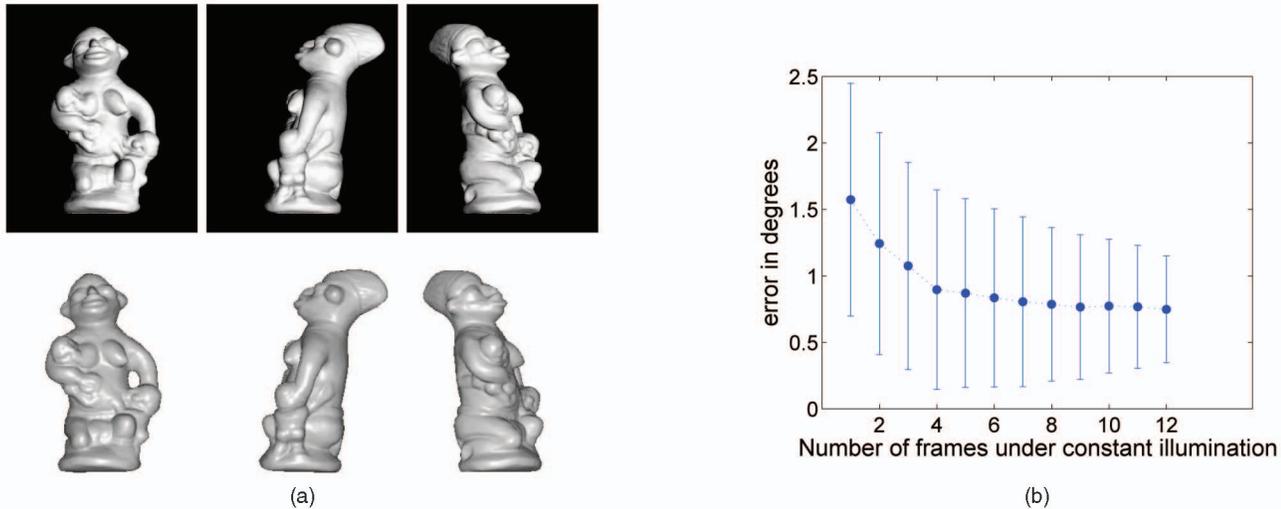


Fig. 8. Synthetic evaluation. (a) The accuracy of the algorithm was evaluated using an image sequence synthetically generated from a 3D computer model of a sculpture. This allowed us to compare the quality of the reconstructed model against the original 3D model as well as to measure the accuracy of the light estimation. The figure shows the reconstruction results obtained below the images of the synthetic object. The mean distance of all points of the reconstructed model from the ground truth was found to be about 0.5 mm if the bounding volume's diagonal is 1 m. (b) The figure shows the effect of varying the length of the frame subsequences that have constant light. The angle between the recovered light direction and ground truth has been measured for 1,000 runs of the RANSAC scheme for each number of frames under constant lighting. With just a single frame per illumination, the algorithm achieves a mean error of 1.57 degrees with a standard deviation of 0.88 degrees. With 12 frames sharing the same illumination, the mean error drops to 0.75 degrees with a standard deviation of 0.41 degrees.

accuracy of the recovered lighting directions as  $K$  increases from 1 to 12. The metric used was the angle between the ground truth light direction and the estimated light direction over 1,000 runs of the robust estimation scheme. For  $K = 1$ , the algorithm achieves a mean error of 1.57 degrees with a standard deviation of 0.88, while, for  $K = 12$ , it achieves 0.75 degrees with a standard deviation of 0.41 degrees. The decision for selecting a value for  $K$  should be a consideration of the trade-off between practicality and maximizing the total number of different illuminations in the sequence which is  $M/K$ , where  $M$  is the total number of frames.

## 5 CONCLUSION

This paper has presented a novel reconstruction technique using silhouettes and the shading cue to reconstruct Lambertian objects in the presence of highlights. The main contribution of the paper is a robust, fully self-calibrating, efficient setup for the reconstruction of such objects, which allows the recovery of a detailed 3D model viewable from 360 degrees. We have demonstrated that the powerful silhouette cue, previously known to give camera motion information, can also be used to extract photometric information. In particular, we have shown how the silhouettes of a Lambertian object are sufficient to recover an unknown illumination direction and intensity in every image. Apart from the theoretical importance of this fact, it also has a practical significance for a variety of techniques which assume a precalibrated light-source and which could use the silhouettes for this purpose, thus eliminating the need for special calibration objects and the time-consuming manual calibration process.

## REFERENCES

- [1] M. Levoy, "Why Is 3D Scanning Hard?" *Proc. 3D Processing, Visualization, Transmission*, invited address, 2002.
- [2] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 519-528, 2006.
- [3] R. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Eng.*, vol. 19, no. 1, pp. 139-144, 1980.
- [4] J. Lim, J. Ho, M. Yang, and D. Kriegman, "Passive Photometric Stereo from Motion," *Proc. IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1635-1642, Oct. 2005.
- [5] J. Paterson, D. Claus, and A. Fitzgibbon, "BRDF and Geometry Capture from Extended Inhomogeneous Samples Using Flash Photography," *Proc. Eurographics '05*, vol. 24, no. 3, pp. 383-391, 2005.
- [6] F. Bernardini, H. Rushmeier, I. Martin, J. Mittleman, and G. Taubin, "Building a Digital Model of Michelangelo's Florentine Pieta," *IEEE Computer Graphics and Applications*, vol. 22, no. 1, pp. 59-67, Jan./Feb. 2002.
- [7] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently Combining Positions and Normals for Precise 3D Geometry," *Proc. ACM SIGGRAPH*, pp. 536-543, 2005.
- [8] G. Vogiatzis, P. Favaro, and R. Cipolla, "Using Frontier Points to Recover Shape, Reflectance and Illumination," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 228-235, 2005.
- [9] H. Jin, D. Cremers, A. Yezzi, and S. Soatto, "Shedding Light in Stereoscopic Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 36-42, 2004.
- [10] O. Dbrohlav and M. Chandler, "Can Two Specular Pixels Calibrate Photometric Stereo?" *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1850-1857, 2005.
- [11] C. Hernández, F. Schmitt, and R. Cipolla, "Silhouette Coherence for Camera Calibration under Circular Motion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 343-349, Feb. 2007.
- [12] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150-162, Feb. 1994.
- [13] R. Cipolla and P. Giblin, *Visual Motion of Curves and Surfaces*. Cambridge Univ. Press, 1999.
- [14] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model-Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381-395, 1981.
- [15] G. Vogiatzis, C. Hernández, and R. Cipolla, "Reconstruction in the Round Using Photometric Normals and Silhouettes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 1847-1854, 2006.
- [16] C. Hernández and F. Schmitt, "Silhouette and Stereo Fusion for 3D Object Modeling," *Computer Vision and Image Understanding*, vol. 96, no. 3, pp. 367-392, Dec. 2004.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).