Reconstruction in the round using photometric normals.

George Vogiatzis[1] Carlos Hernández[2] Roberto Cipolla[3]

Dept. of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK

{gv215[1], ch394[2], cipolla[3]}@eng.cam.ac.uk

Abstract

This paper addresses the problem of obtaining complete, detailed reconstructions of shiny textureless objects. We present an algorithm which uses silhouettes of the object, as well as images obtained under varying illumination conditions. In contrast with previous photometric stereo techniques, ours is not limited to a single viewpoint and 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 initialise a multi-view photometric stereo scheme to obtain a closed surface reconstruction. The contributions of the paper are twofold: Firstly we describe a robust technique to estimate light directions and intensities and secondly, 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 a complete reconstruction of a challenging real object.

1. Introduction

We propose a method for acquiring a complete 3D model of a uniform untextured object from a number of images captured under varying light conditions. The object's reflectance is assumed to follow Lambert's law but a significant number of highlights are present. A sequence of images of such an object is given, where the object moves in front of a fixed camera and a single distant light-source moves arbitrarily between each image capture. It is also assumed that the object can be segmented from the background and silhouettes extracted automatically.

Shape recovery from images is a well established computer vision task with two families of techniques offering the most accurate results, multi-view stereo and photometric stereo (see [6] and [5] for some of the best quality results
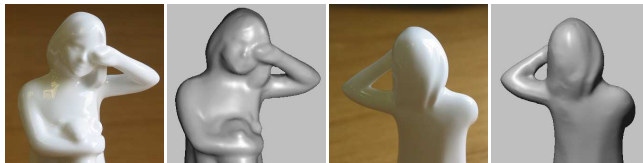

Figure 1. **Reconstructing textureless shiny objects.** Objects from textureless shiny materials such as the porcelain figurine shown, present a challenge for shape reconstruction algorithms. The lack of surface features makes traditional multi-view stereo very difficult to apply while photometric stereo has so far only been able to produce 2.5D reconstructions. Our algorithm is able to produce closed-surface, full 3D reconstructions of many-sided objects, from a sequence of uncalibrated images and an arbitrarily moving light-source. Here, two views of the reconstructed model are shown next to views of the porcelain object.

from each method). While correspondence based multi-view stereo techniques offer detailed full 3D reconstructions, they rely on richly textured objects to obtain correspondences between locations in multiple images which are triangulated to obtain shape. As a result these methods are not directly applicable to the class of objects we are considering due to the lack of detectable surface features. An attempt was made on reconstructing such objects in [7] but the reconstructed models shown lack surface detail which is due to the regularisation enforced on the reconstructed surface. On the other hand, photometric stereo works by observing the changes in image intensity of points on the object surface as illumination varies. These changes reveal the local surface orientations at those points that, when integrated, provide the 3D shape. Because photometric stereo performs integration to recover depth, much less regularisation is needed and results are generally more detailed. Furthermore, photometric stereo makes fewer assumptions about surface texture and reflectance, which can be almost completely arbitrary as demonstrated in [5]. However, the simplest way to collect intensities of the *same* point of the surface in multiple images is if the camera viewpoint is held constant, in which case every pixel always corresponds to the same point of the surface. This is a major limiting factor

of the method because it does not allow the recovery of the full 3D geometry of a complex many-sided object such as a sculpture. Due to this limitation existing photometric stereo techniques have so far only been able to extract depth-maps (e.g. [13]) with the notable recent exceptions of [15, 9], where the authors present techniques for recovering 2.5D reconstructions from multiple viewpoints. The full reconstruction of many-sided objects is however still not possible by these methods. While in theory one could apply photometric stereo from multiple viewpoints and then merge the multiple depth-maps of the object into a single 3D representation, in practice this procedure can be complicated and error-prone.

1.1. Our approach

In this paper a different solution is sought by exploiting the powerful silhouette cue. We modify classic photometric stereo and cast it in a multi-view framework where the camera is allowed to circumnavigate the object and illumination is allowed to vary. Firstly, the object's silhouettes are used to recover camera motion using a technique similar to [10], and via a novel robust estimation scheme they allow us to accurately estimate the light directions and intensities in every image.

Secondly, the object surface, which is parameterised by a mesh and initialised from the visual hull, is evolved until its predicted appearance matches the captured images. Each face of the mesh is projected in the images where it is visible and the intensities are collected. From these intensities and the illumination computed previously, a normal direction is assigned to each face by solving a local least squares problem. The mesh is then iteratively evolved until these directions converge to the actual surface normals of the mesh. These two phases are then repeated until the mesh converges to the true surface. 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.

- The full 3D geometry of a complex, shiny, textureless object is accurately recovered, something not previously possible by any other method.

- It is practical and efficient as evidenced by our simple acquisition setup.

The next section discusses work related to the ideas presented in this paper. Section 2 presents the part of the algorithm that deals with estimating light directions and intensities while Section 3 explains how a photometrically consistent closed 3D surface is recovered. Section 4 discusses the acquisition setup we use while 5 describes a set of experiments carried out on real and synthetic objects. Section

6 concludes with a discussion of our key contributions and future research directions.

1.2. 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 [9] where the authors explore the possibility of using photometric stereo with images from multiple views, when correspondence between views is not initially known. Their method however relies on the existence of distinct features on the object surface which are tracked to obtain camera motion and initialise the depth-map while 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 [12] 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 reconstruction *in the round*.

Quite related to this idea is the recent work of [11] where photometric stereo information is combined with 3D range scan data. In that paper, using range scanning technology a very good initial approximation to the object surface is obtained, 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 [11] 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 [14] 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 paper generalises the idea by examining a much richer superset of frontier points which is the set of contour generator points. We overcome the difficulty of localising contour generators by a robust random sampling strategy. The price we pay is

that a considerably simpler reflectance model must be used.

Although solving a different type of problem, the work of [7] is also highly related mainly because the class of objects addressed is similar to ours. While the energy term defined and optimised 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 [7] at first glance seem to be lacking in detail especially in concavities, while our technique considerably improves on the visual hull (see Figures 4(c) vs (b)).

Finally, there is a growing volume of work on using specularities for calibrating photometric stereo (see [3] 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.

2. Robust estimation of light-sources from the visual hull

When illumination directions and surface reflectance are completely unknown it is only possible to reconstruct the surface up to an unknown Generalised Bas-Relief ambiguity by enforcing the integrability of the recovered surface normals [1]. When surface albedo is known or constant however, as in the case we are considering, the ambiguity is removed. Unfortunately as mentioned in Section 1 when the viewpoint is not fixed, image intensities of the same surface point cannot be collected since correspondence between pixels is unknown; this is in fact what we seek to estimate.

A different approach is to estimate illumination independently and then focus solely on the task of reconstructing the object surface. For an image of a Lambertian object of constant albedo under a single distant light source and ignoring self-cast shadows, each surface point projects to a point of intensity given by:

$$i = \mathbf{l^T n} \tag{1}$$

where $\mathbf{l}$ is a 3D vector directed towards the light-source and scaled by the light-source intensity and $\mathbf{n}$ is the surface unit normal at the object location. It is obvious that given three known normals and the corresponding three image intensities we can construct three such equations that can uniquely determine $\mathbf{l}$.

Our approach is to estimate illumination using 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* [8], which is defined as the maximal volume that projects inside the silhouettes. 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* [2]. Furthermore, for all points on those curves, the normal direction of the visual hull surface is equal to the normal direction of the object surface. Therefore if we could detect points on the visual hull that belong to contour generators we could use their normals and projected intensities to estimate lighting. Unfortunately contour generator points 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 and the outliers are the rest of the visual hull points. One can expect that the outliers do not generate consensus in favour of any particular illumination model while the inliers do so in favour of the correct model. This observation motivates us to use a robust RANSAC scheme [4] to separate inliers from outliers and estimate illumination direction and intensity. The scheme is now described in detail.

Consider firstly the case of estimating the distant light source direction and intensity in a single image. Assume we are given a dense but discrete set of locations on the visual hull $\mathbf{x_1}, \ldots, \mathbf{x_M}$ which are visible in the image and whose corresponding visual hull surface normals are $\mathbf{n_1}, \ldots, \mathbf{n_M}$. Let the observed image intensities of those points be $i_1, \ldots, i_M$. At each RANSAC iteration we pick three points at random, say $\mathbf{x_a}, \mathbf{x_b}, \mathbf{x_c}$ and estimate a tentative illumination vector

$$\mathbf{l} = [\mathbf{n_a} \, \mathbf{n_b} \, \mathbf{n_c}]^{-1} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix}. \tag{2}$$

Every visual hull point $\mathbf{x_m}$ will now vote for this hypothesis if the discrepancy between its observed intensity and its predicted intensity is less than a threshold τ i.e.

$$|\mathbf{l} \cdot \mathbf{n_m} - i_m| < \tau. \tag{3}$$

where τ allows for quantisation errors, image noise, etc. During the entire process, the illumination vector that gathers the maximum consensus (number of votes) is kept and after convergence its voters are used to estimate the optimal illumination vector as the solution of a linear least squares problem.

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 $R_1\mathbf{l}, \ldots, R_K\mathbf{l}$ where R_k are 3×3 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

$$|(R_k\mathbf{l}) \cdot \mathbf{n_m} - i_{m,k}| < \tau. \tag{4}$$

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 frames. 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 the experiments section.

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, local albedo variations or self-cast shadows.. The next section describes the second part of the algorithm which uses the estimated illumination directions and intensities to recover the object surface.

3. Multi-view 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 develop a mesh-based optimisation approach where a cost function penalising the discrepancy between images and predicted appearance is minimised.

Our algorithm optimises a surface S that is represented as a mesh with vertices $\mathbf{x_1} \ldots \mathbf{x_M}$ and triangular faces $f = 1 \ldots F$. 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 denote by $\mathcal{V}_f$ the set of images (subset of $\{1, \ldots, K\}$) in which the intensity of face f can be measured. We will describe $\mathcal{V}_f$ in more detail in section 3.1. The light direction and intensity of the k-th image will be denoted by a 3D vector $\mathbf{l_k}$.

The simplest possible formulation of the photometric consistency cost is

$$E\left(\mathbf{x_1}, \ldots, \mathbf{x_M}\right) = \sum_{f=1}^{F} \sum_{k \in \mathcal{V}_f} \left(\mathbf{l_k}^T \mathbf{n_f} - i_{f,k}\right)^2 A_f \quad (5)$$

Unfortunately, as we verified experimentally, this scheme fails to converge to the right solution. This was also noted in [7] where the authors investigated a similar equation for a multi-view shape-from-shading algorithm. Following their intuition albeit for a different problem, we introduce a decoupling between the mesh normals, which depend on $\mathbf{x_1} \ldots \mathbf{x_M}$, and the direction vectors used in the Lambertian

model equation which become a set of independent variables $\mathbf{v_1} \ldots \mathbf{v_F}$ which we call *photometric normals*. The new term becomes

$$
\begin{aligned}
E\left(\mathbf{x_1}, \ldots, \mathbf{x_M}, \mathbf{v_1}, \ldots, \mathbf{v_F}\right) = \\
E_m\left(\mathbf{x_1}, \ldots, \mathbf{x_M}; \mathbf{v_1}, \ldots, \mathbf{v_F}\right) \\
+ E_v\left(\mathbf{v_1}, \ldots, \mathbf{v_F}; \mathbf{x_1}, \ldots, \mathbf{x_M}\right) \quad (6)
\end{aligned}
$$

where the first term E_m brings the mesh normals close to the photometric normals through the following equation:

$$E_m\left(\mathbf{x_1}, \ldots, \mathbf{x_M}; \mathbf{v_1}, \ldots, \mathbf{v_F}\right) = \sum_{f=1}^{F} \|\mathbf{n_f} - \mathbf{v_f}\|^2 A_f \quad (7)$$

and the second term E_v links the photometric normals to the observed image intensities through:

$$E_v\left(\mathbf{v_1}, \ldots, \mathbf{v_F}; \mathbf{x_1}, \ldots, \mathbf{x_M}\right) = \sum_{f=1}^{F} \sum_{k \in \mathcal{V}_f} \left(\mathbf{l_k}^T \mathbf{v_f} - i_{f,k}\right)^2 . \quad (8)$$

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

1. **Vertex optimisation.** The photometric normals are kept fixed while E_m is optimised with respect to the vertex locations using gradient descent.

2. **Photometric normal update.** The vertex locations are kept fixed while E_v is optimised with respect to the photometric normals. This is achieved by solving the following independent linear least squares problems for each face f:

$$\mathbf{v_f} = \arg \min_{\mathbf{v}} \sum_{k \in \mathcal{V}_f} \left(\mathbf{l_k}^T \mathbf{v} - i_{f,k}\right)^2 \text{ s.t. } \|\mathbf{v}\| = 1 \quad (9)$$

These two steps are interleaved until convergence which takes about 20 steps for the sequences we experimented with. Typically each integration phase takes about 100 gradient descent iterations.

3.1. Visibility map

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 τ_{shadow} is a sufficiently low intensity threshold. Due to the inclusion of a significant number of viewpoints in $\mathcal{V}_f$, (normally at least 4) the system is quite robust to the choice of τ_{shadow}. For all the experiments presented here, the value $\tau_{shadow} = 5$ was used (for intensities in the range 0-255).

```
Capture images of object.
Extract silhouettes.
Recover camera motion and compute visual hull.
Estimate light directions and intensities in every image
(Section 2).
Initialise a mesh with vertices x₁ ... x_M and faces f =
1 ... F to the object's visual hull.
while mesh-not-converged do
    Optimise E_v with respect to v₁ ... v_F.
    Optimise E_m with respect to x₁ ... x_M.
end while
```

Figure 2. **The multi-view reconstruction algorithm.**

4. Acquisition setup

The setup used to acquire the 3D model of the object is quite simple. It consists of a turntable, onto which the object is mounted, a 60W halogen lamp and a digital camera. The object rotates on the turntable and 36 images of the object are captured by the camera while the position of the lamp is changed. Even though our method is able to recover the light direction and intensity independently in each image, estimation accuracy is improved if light can be held constant for more than one frame, as shown in Figure 8. In our experiments we have used three different light positions which means that the position of the lamp was changed after twelve, and again after twenty-four frames. The distant light source assumptions are satisfied if an object of about 15cm is placed 3-4m away from the light.

5. Experiments

In this section we present an experimental evaluation of our algorithm, first on a real challenging object and then on a synthetic scene for which ground truth is known.

5.1. Real objects

The algorithm was tested on two challenging shiny porcelain figurines shown in figures 4 and 5. Thirty-six 922×1158 images of each of the porcelain objects were captured under three different illuminations (twelve images per illumination). The object silhouettes were extracted by intensity thresholding and were used to estimate camera motion and construct the visual hull (second row of figures 4 and 5). The visual hull was then processed by the robust light estimation scheme of Section 2 to recover the distance light-source directions and intensities in each image. Figure 3 qualitatively demonstrates the stability of the light estimation algorithm for the first and simplest porcelain object. The light directions obtained by 20 independent runs of the robust scheme are shown to be within 1.4 degrees off the mean light directions obtained.
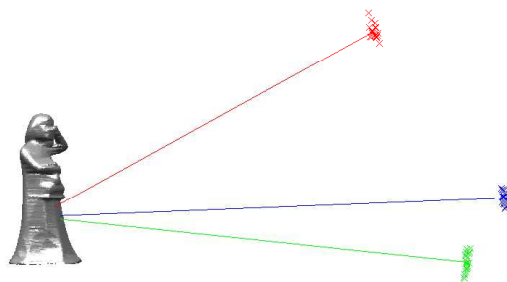
Figure 3. **Light estimation is stable.** The figure shows the visual hull and the three recovered light directions. The point clouds at the end of each of the three light direction vectors show the results of individual RANSAC runs which are on average 0.8 degrees away from the mean estimate with a standard deviation of 0.5 degrees.

In the next part of the algorithm, as described in Section 3, a mesh surface is initialised from the visual hull and iteratively deformed until it becomes photometrically consistent with the images. This is achieved in 20 iterations of the photometric normal update and vertex optimisation phases. 100 gradient descent steps were used for each vertex optimisation phase. Figures 4(c) and 5(c) show the reconstruction results obtained by our algorithm for the two porcelain figurines. Most of the surface details of the two reconstructed objects have been captured. Due to the fact that the objects are completely textureless, standard correspondence-based multi-view stereo algorithms will fail because of the inability to establish correspondences between different images. As a result, the only method able to produce a multi-sided, closed surface reconstruction is Shape from Silhouettes, which generates the visual hull shown in the Figures 4(b) and 5(b). Note that the visual hull reconstruction lacks all shape concavities, which are however correctly recovered by our method.

To better understand the effect of the quality of the visual hull on our algorithm, we performed the same experiment of reconstructing the first porcelain figurine from thirty-six images, but this time the visual hull was constructed from just four silhouettes, generating the shape shown in Figure 7. Both light estimation and the initialisation were performed using this volume, and the results demonstrated the robustness of the algorithm against visual hulls that are far away from the true surface. Figure 7 provides an illustration of the voting process of the light estimation algorithm. The visual hull from four views is a simple shape with four facets, each traversed by a contour generator. Two views of this volume are shown, on which have been marked the positions of the voters for the estimated light direction. The voters are forming curves along the facets which coincide with the regions where the porcelain figurine would be tangent
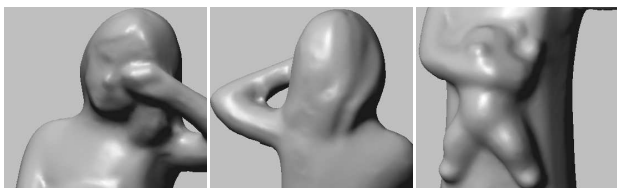
(a) Input images.
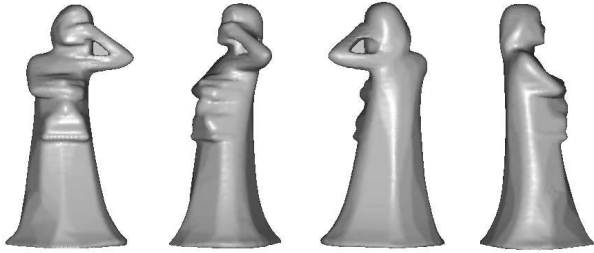
(b) Visual hull reconstruction.

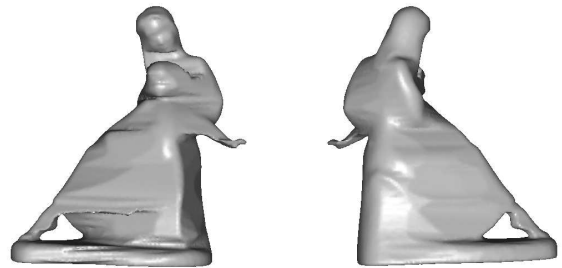(c) Our results.

(d) Close up images of porcelain.

(e) Close up views of reconstructed model.

Figure 4. **Reconstructing porcelain.** A porcelain figurine is re-constructed from a sequence of 36 images (four of which are shown in (a)). The object moves in front of the camera and illumination (a 60W 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. Details of the reconstruction are shown below photographic images of the actual object from similar views in (d) and (e).

(a) Input images.

(b) Visual hull reconstruction.
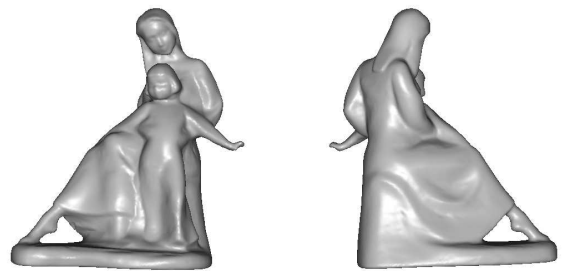
(c) Our results.

Figure 5. **Reconstructing porcelain figurines.** A more complex porcelain object is reconstructed from 36 images. Experimental setup is identical to that of Figure 4.

to the visual hull volume, i.e. the contour generators. Even using such a coarse shape for the visual hull, the estimation algorithm is able to obtain a light direction estimate which is just 1.5 degrees away from the estimate obtained from the full visual hull of thirty-six views.

Finally, we initialised the mesh using this coarse visual hull, and the evolution, six snapshots of which are shown in Figure 6, once again converged to the true surface after the same number of iterations as previously. This implies that the mesh evolution algorithm is tolerant to poor initialisations such as the one provided.

5.2. Synthetic object

To quantitatively analyze the performance of the multi-view photometric stereo scheme presented here with ground
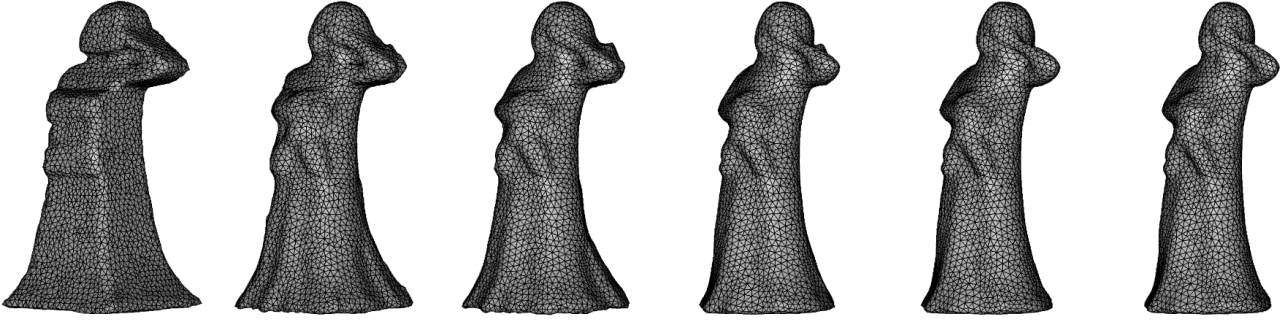
Figure 6. **Mesh evolution converges even with poor initialisation.** To test the radius of convergence of the iterative mesh evolution algorithm the mesh was initialised from the visual hull generated by four silhouettes (leftmost image). The figure shows several snapshots of the evolution. The mesh gradually evolves to the correct volume after 20 iterations of the two phases, the photometric normal update phase followed by 100 gradient descent minimisation steps for the vertex optimisation phase.
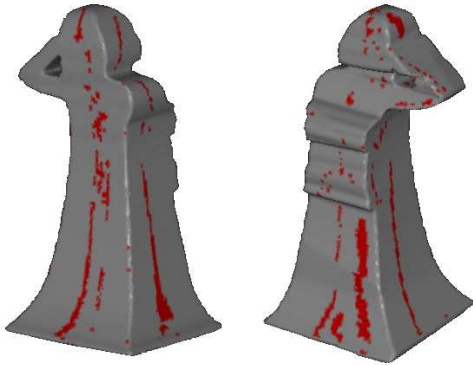

Figure 7. **Light recovery.** To better illustrate the illumination estimation algorithm here we performed the estimation using the visual hull generated from four silhouettes. The figure shows two views of the visual hull, on which the points that vote for the final illumination direction are marked in red. The voters are forming long 3D curves along the sides of the visual hull which coincide with the contour generators. Even with such a coarse approximation to the original geometry the RANSAC estimation scheme recovers light directions less than 1.5 degrees away from the estimates recovered using the full visual hull computed from 36 silhouettes.

truth, an experiment on a synthetic scene was performed (Figure 8). A 3D model of a sculpture (digitised 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 2. In 1000 runs of the illumination estimation method for the synthetic scene, the mean light direction es-

timate was 0.75 degrees away from the true direction with a standard deviation of 0.41 degrees. The model 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 0.5mm when the length of the biggest diagonal of the bounding box volume was defined to be 1m. 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 optimisation methods themselves. Many traditional multi-view stereo methods cannot achieve this due to the strong regularisation that must be imposed on the surface. By contrast our method requires no regularisation when faced with 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 by 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 Figure 8 and show the improvement of the 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 1000 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 tradeoff between practicality and max-
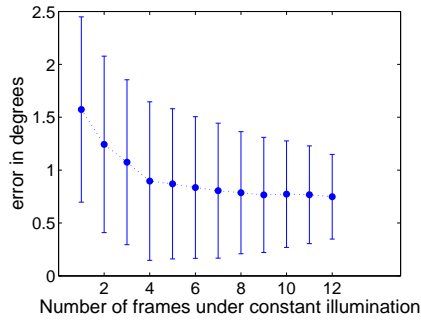
Figure 8. **Evaluation of illumination estimation.** The figure shows the effect of varying the length of the frame subsequences that have constant light. The angle between the recovered light direction and ground truth has been measured for 1000 runs of the RANSAC scheme for each number of frames under 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.

imising the total number of different illuminations in the sequence which is M/K where M is the total number of frames. We qualitatively found that the greater the number of different illuminations used, the better the surface reconstruction accuracy, but further investigation of this is necessary.

6. Conclusion

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 uniform 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 pre-calibrated 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.

This paper has presented a novel reconstruction technique using silhouettes and the cue of photometric stereo to reconstruct uniform featureless 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. This is, to our knowledge, the first photometric stereo based method to achieve this.

References

[1] P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. *Intl. Journal of Computer Vision*, 35(1):33–44, 1999. 3

[2] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Intl. Journal of Computer Vision*, 9(2):83–112, November 1992. 3

[3] O. Dbrohlav and M. Chandler. Can two specular pixels calibrate photometric stereo ? In *Proc. 10ᵗʰ Intl. Conf. on Computer Vision*, 2005. 3

[4] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model-fitting with applications to image analysis and automated cartography. *CACM*, 24(6):381–395, 1981. 3

[5] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Proc. 10ᵗʰ Intl. Conf. on Computer Vision*, 2005. 1

[6] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, December 2004. 1

[7] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light in stereoscopic segmentation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 36–42, 2004. 1, 3, 4

[8] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2), 1994. 3

[9] J. Lim, J. Ho, M. Yang, and D. Kriegman. Passive photometric stereo from motion. In *Proc. 10ᵗʰ Intl. Conf. on Computer Vision*, 2005. 2

[10] P. Mendonça, K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):604–616, 2001. 2

[11] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *Proc. of the ACM SIGGRAPH*, pages 536–543, 2005. 2

[12] J. Paterson, D. Claus, and A. Fitzgibbon. Brdf and geometry capture from extended inhomogeneous samples using flash photography. In *Proc. of Eurographics 2005*, 2005. 2

[13] A. Treuille, A. Hertzmann, and S. Seitz. Example-based stereo with general brdfs. In *Proc. 8ᵗʰ Europ. Conf. on Computer Vision*, may 2004. 2

[14] G. Vogiatzis, P. Favaro, and R. Cipolla. Using frontier points to recover shape, reflectance and illumination. In *Proc. 10ᵗʰ Intl. Conf. on Computer Vision*, pages 228–235, 2005. 2

[15] L. Zhang, B. Curless, A. Hertzmann, and S. Seitz. Shape and motion under varying illumniation: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proc. 9ᵗʰ Intl. Conf. on Computer Vision*, 2003. 2