Using Illumination Estimated from Silhouettes to Carve Surface Details on Visual Hull

Shuda Li, Kwan-Yee K. Wong and Dirk Schnieders
Department of Computer Science, University of Hong Kong,
Hong Kong SAR, China
{sdli,kykwong,sdirk}@cs.hku.hk

Abstract

This paper deals with the problems of scene illumination estimation and shape recovery from an image sequence of a smooth textureless object. A novel method that exploits the surface points estimated from the silhouettes for recovering the scene illumination is introduced. Those surface points are acquired by a dual space approach and filtered according to their rank errors. Selected surface points allow a direct closed-form solution of illumination. In the mesh evolution step, an algorithm for optimizing the visual hull mesh is developed. It evolves the mesh by iteratively estimating both the surface normal and depth that maximize the photometric consistency across the sequence. Compared with previous work which optimizes the mesh by estimating the surface normal only, the proposed method shows better convergence and can recover better surface details, especially when concavities are deep and sharp.

1 Introduction

The problem of shape recovery from an image sequence has been well studied in the computer vision community over the past two decades. For texture-rich objects, techniques based on matching correspondences across images [12, 10], can be applied to recover the shape by triangulation, mesh evolution[3], surface growing[5], or graph cut[6]. For smooth textureless objects, however, correspondences cannot be established easily. A common approach to handle this class of objects is to make use of the silhouette information. Each silhouette, together with the camera center, defines a viewing cone in which the object surface is inscribed. The intersection of the viewing cones generated from a set of silhouettes observed at different viewpoints then produces the visual hull [13, 4] of the object. Given a large number of silhouettes observed at distributed viewpoints, the visual hull often gives a very good approximation to the object surface. However, surface details such as concavities can not be observed from the silhouettes and therefore can not be recovered.

As opposed to silhouette-based methods, methods based on shadings, like photometric stereo [11, 1], are able to recover the surface details of a complex object with concavities. For Lambertian objects, the shading of a surface point depends on both the surface orientation and the illumination direction. Theoretically, given the shadings of a point

observed under varying but known illumination, its surface normal vector can be estimated. Alternatively, the illumination direction can be recovered from the shadings of at least three points with known but unparalleled orientations. The main challenge in using photometric stereo is how to identify the shadings of the same surface point in different images, which is analogous to the *correspondence problem* in multi-view stereo. Most existing algorithms avoid such a problem by limiting their considerations to the case where the object and camera are relatively fixed while the illumination is varying. This, however, limits their methods to producing only a 2.5D depth map instead of a complete 3D reconstruction.

In this paper, we consider the problem of recovering a complete and detailed 3D model of a smooth textureless object from an image sequence. We address this problem by proposing a multi-step method that combines the strengths of both shape from silhouettes and shape from shadings. The proposed approach begins by generating a coarse model (i.e., the visual hull) from the object silhouettes, and then optimizes the model using shading information to recover concavities as well as other surface details. In order to make use of shading for the surface optimization, it is necessary to recover the scene illumination, which is equivalent to solving the GBR ambiguity. This is achieved by first recovering the surface points of the object from its silhouettes using a dual space method [14], and then measure the rank 1 error inspired by [8] to drop out those surface points whose errors are larger than a predefined threshold. The rest surface points are used to estimate the illumination direction and magnitude by a closed-form solution. Besides, a novel algorithm for optimizing the visual hull mesh is also developed, which evolves the mesh by iteratively estimating both the surface normal and depth that maximize the photometric consistency across the sequence. Compared with previous work, experimental results show that the proposed method gives better convergence and can recover better surface details, especially when concavities are deep and sharp.

2   Related Work

In the literature, there is a plethora of works concerning the recovery of shape from image sequences. For smooth textureless objects, Jin et al. propose two variational algorithms to jointly estimate the shape and surface reflection.[8, 7] As noted in [17], however, their methods tend to produce models lacking of surface details due to the insensitivity to the depth and the strong regularization term. In [16], Vogiatzis et al. propose to decouple the recovery of shape and illumination by exploiting frontier points[2]. Frontier points and their surface normals can be conveniently estimated from the epipolar tangents to the silhouettes, and provide local surface geometry for illumination estimation. However, such an approach is not always applicable to real sequences acquired around a target object (e.g., turntable sequences), since the number of available frontier points are usually small and their projections in the images are close to the boundaries of the silhouettes. This may lead to a low accuracy in the illumination estimation. In [17], Vogiatzis et al. proposed a RANSAC approach to estimate the illumination from a visual hull mesh. In the RANSAC process, vertices in the mesh which are lying on the object are defined as inliers, whereas the rest of the vertices are treated as outliers. This approach avoids the problem of not having sufficient frontier points with stable shadings. In practice, however, only a very small portion of the vertices are actually lying on the object surface, and this

will result in a large number of samplings required by the RANSAC algorithm.

This work reconstructs the objects by optimizing the photometric consistency of their Lambertian components. The *radiance tensor rank proposition* proofs that for non-Lambertian objects, the specularity of surface reflection can be either absent or washed out [8] under a point light source with the camera far away from the object. Hence glassy objects with no specularity obey Lambert's reflection Law almost perfectly in the real world. Our experiments on real data show high convergence under the assumption of Lambertian surfaces, which reconfirm the above proposition. The proposed method follows the same strategy of [17], which first constructs a coarse mesh model by a shape from silhouettes method, followed by illumination estimation, and finally optimizes the mesh using photometric consistency constraint. It differs from [17] in 3 main aspects:

- In the shape from silhouettes, instead of using just a volumetric approach to construct a visual hull mesh, we also adopt a previously developed dual space method [14] to recover surface points, with normals.

- In the illumination estimation step, instead of using a RANSAC approach, we use a rank constraint to drop out the outliers, which allows a direct closed-form solution.

- In the mesh optimization step, instead of evolving the mesh by iteratively estimating the photometric normal, we estimate the photometric depth along with the normal by maximizing the photometric consistency.

The rest of this paper is organized as follows: Section 3 introduces a novel way for precise estimation of the scene illumination by exploiting the recovered surface points. Section 4 derives a new functional based on the photometric normal and photometric depth for optimizing the visual hull mesh. Experimental results and analysis are presented in Section 5, followed by conclusions in Section 6.

3   Illumination Estimation Using a Closed-Form Solution

Consider an image of a Lambertian object with constant albedo and under a single distant light source. The shading (or intensity) of a visible surface point in the image is given by

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

where $\mathbf{n}$ is the outward-pointing unit normal vector at the surface point, and $\mathbf{l}$ is the 3D lighting vector directed towards the light source and scaled by the intensity of the light source. Equation (1) gives a linear constraint on the 3 unknowns of $\mathbf{l}$. It follows that $\mathbf{l}$ can be uniquely determined from the shadings of 3 visible surface points with known but distinct surface normals. In this paper, the surface normals and positions are provided by surface points recovered using a previously developed dual space method [14].The visual hull mesh constructed from the silhoettes is used to infer the visibility of these points. The shading of a surface point can be obtained by projecting it onto the image using the projection matrix of the camera.

However, due to noise and insufficient sampling rates, it may produce poor estimations over bumpy surface, surface concavities and near frontier points. To overcome this

problem, we employ a rank constraint to classify the surface points into two groups.[9] To exclude those surface points whose positions are far from the true objects or disturbed by shadow or specularities, we remove those who have large rank 1 errors. Figure 1 shows effect of the selection according to the rank 1 errors.
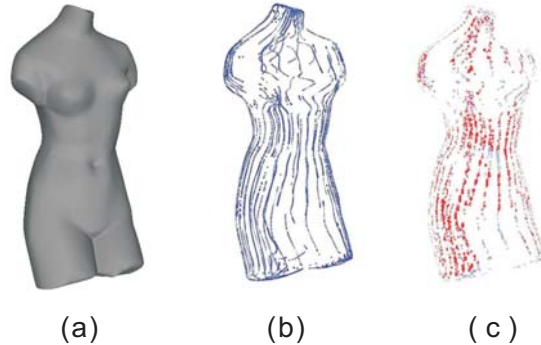

(a)   (b)   (c)

Figure 1: (a) Original model. (b) Surface points recovered using dual space method. (c) Selected surface points (red) used in illumination estimation. The light blue points over the concavities are dropped out as outliers.

4 Mesh Optimization

Having the scene illumination estimated from the surface points, we next describe a method to optimize the initial visual hull mesh. Our goal is to evolve the mesh so that its appearance predicted using the estimated scene illumination matches its actual appearance observed in the image sequence.

4.1 Photometric depth and photometric normal

Given more than 3 shadings under linearly independent lights, the surface normal can be easily estimated by solving a equation system. To make the difference from the original normal, the estimated one is termed as *photometric normal*, the same as [17]. For mesh triangles near the true object surface, the photometric normals usually can be robustly estimated and the vertex optimization step can drive those triangles towards the true surface. For mesh triangles in the visual hull over a concave section of the object, however, their projections on different images correspond to the projections of different surface points on the true surface which are far apart from each other (see Figure 2). It results in erroneous photometric normals and the vertex optimization step may drive those triangles to crazy positions. In order to make the optimization converge, the regularization term has to be enforced, resulting in loss of fine details.

In this paper, we introduce the *photometric depth* to solve the problem of erroneous photometric normal. The key idea is to estimate both the surface normal and depth for each mesh triangle that maximize its photometric consistency. To compute the photometric depth, we first offset the triangle along its normal direction, then compute the
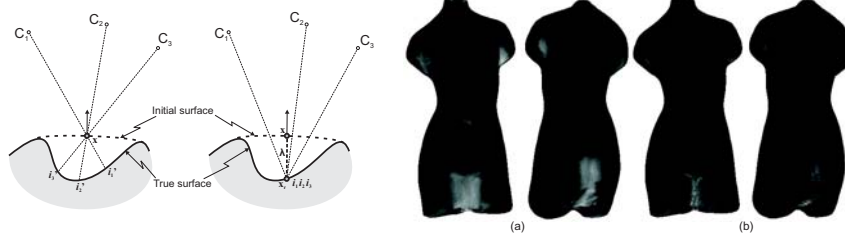
Figure 2: Illustration of the error at concave section. (left) $\mathbf{C_1}, \mathbf{C_2}, \mathbf{C_3}$ correspond to camera centers under different illuminations. i_1', i_2', i_3' in the left figure are the wrong intensities used for calculating the photometric normal at point $\mathbf{x}$. The right figure shows the photometric normal computation with correct intensity i_1, i_2, i_3 collected at real surface point $\mathbf{x_r}$. The offset from $\mathbf{x}$ to $\mathbf{x_r}$ is the *photometric depth* and is denoted as λ. Comparison between the photometric consistency achieved by optimizing photometric normal only, and by optimizing photometric normal and depth together. (right) Brighter color corresponds to lower photometric consistency. (a) optimizing photometric normal only. (b) optimizing photometric normal and depth.

photometric normal of the triangle for that particular offset value. The offset value that maximizes the photometric consistency of the triangle is taken as its photometric depth λ_f, and the corresponding normal is taken as the photometric normal $\mathbf{v}_f$. This is shown:

$$\mathbf{v}_f, \lambda_f = arg \min_{\mathbf{v}, \lambda} \sum_{k=1}^{K} \Phi_f(k) \Psi_f(k) \|\mathbf{l}_k^{\mathrm{T}} \mathbf{v} - i_{f,k}(\lambda)\|^2, \tag{2}$$

where $\Phi_f(k) \in [0,1]$ gives the visibility of the triangle in the k-th image, $\Psi_f(k) \in [0,1]$ gives the visibility of the triangle from the light source $\mathbf{l}_k$, λ is the offset of the triangle, $\mathbf{v}$ is the unit photometric normal computed at the offset position, and $i_{f,k}(\lambda_f)$ is the average of the shading values under the projection of the offset triangle in the k-th image.

In our current implementation, Equation (2) is only applied to those triangles visible from at least four images to compute their photometric depths and photometric normals. For triangles visible in less than four images, their photometric depths are set to zero and their surface normals are used as the photometric normals. Figure 2 shows the improvement in photometric consistency achieved by optimizing both photometric depth and photometric normals. It can be seen that the improvement is particularly significant in the concave sections.

4.2 Evolving the mesh

The estimated photometric depths and photometric normals of the mesh triangles are used to drive the mesh vertices towards the true object surface. This is achieved by optimizing the following cost function

$$\mathbf{x}_1, ..., \mathbf{x}_M = arg \min_{\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_M} \sum_{f=1}^{F} (\|\kappa_f - \lambda_f\|^2 + \alpha \|n_f - v_f\|^2) A_f \tag{3}$$

where $\mathbf{n}_f$ is the unit surface normal of the f-th triangle, κ_f is the distance between the centroid of the f-th triangle and the centroid of the f-th triangle in the previous evolution step, A_f is the area of the f-th triangle, and α is a weighting factor used to balance the effect of the depth energy and the normal energy. Note that $\mathbf{n}_f$, κ_f and A_f all depend on the f-th triangle formed by the mesh vertices $\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_M$.

In our implementation, we minimize (3) in two separate steps. The overall algorithm is shown in algorithm 1.

Algorithm 1 Multi-view photometric reconstruction

1: Extract silhouettes from the image sequence;
2: Estimate the camera projection matrices;
3: Construction the visual hull and contour generators from the silhouettes;
4: Estimate the illumination for each image;
5: **while** not converged **do**
6:   Estimate photometric normals and photometric depth;
7:   Optimized the mesh vertices to converge to the photometric depths and photometric normals;
8: **end while**

---

5 Experiments

Figure 3 shows the final reconstruction results of the synthetic model Venus and Bunny, using the proposed method. Compared with the initial visual hulls, it can be observed that surface details in the concave sections are recovered.

Figure 3: The reconstructions by the proposed method. (left) The initial visual hull models. (right) The final surfaces after evolutions have converged.

We also compare the performance of the previous mesh optimization method in [17] with the proposed method. Both methods begin with the same initial surface, and the same amount of regularization energy is assured. The quality of the reconstructed model is measured by a standard evaluation algorithm introduced in [15] in which the distances of all vertices to their nearest points on the ground truth are calculated.

For the Venus model, the mean distance to the ground truth with respect to evolution iterations is plotted and compared in Figure 5 (left). We can see that the mean distance of the model reconstructed using Vogiatzis et al.'s method gradually increases, which

indicates their method tends to produce a larger model. On the other hand, the mean distance of the proposed method gradually decreases to a very small value, showing the proposed mesh converges toward the ground truth. Figure 5 (right) shows the result for the Bunny model which shows similar trends. This can be explained as Vogiatzis' energy function does not incorporate the depth information.
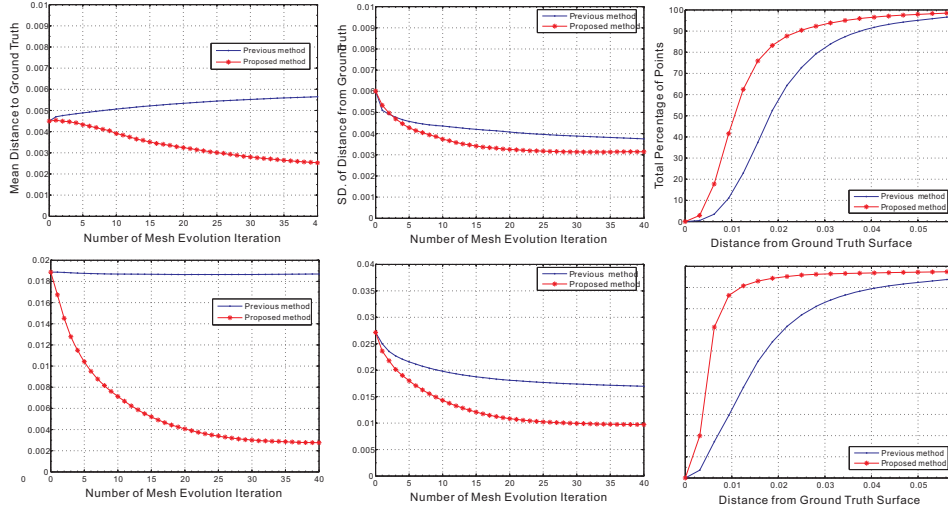

Figure 4: Comparison between previous method [17] and the proposed method. (top row) Venus model. (bottom row) Bunny model. (left column) mean error. (middle column) the standard deviation. (right column) vertices distribution.

The standard deviations in error distance of the two methods are also plotted with respect to evolution in Figure (middle). We can see that the standard deviation in error distance of the model reconstructed by the proposed method drops faster and converges to a smaller value than Vogiatzis' method.

In Figure (right), we compare the distribution of vertices of the previous method and that of the proposed method for the Venus model. We obtained the models after the two methods fully converged. It can be seen that around 90% of vertices in the model produced by our proposed method fall within a distance of 0.02 to the ground truth, while only 55% of the vertices produced by the previous method fall within the same distance. This indicates the proposed method reaches a higher precision than the previous method.

Figure 5 shows the comparison of reconstructions of the Venus model without any regularization. We can see that the proposed method captures the sharp and deep concavity between the legs. While in the previous method, the triangles were driven to incorrect positions by the wrong photometric normal over concavities. The ill-shaped triangles prevent the mesh from converging onto the real surface.

We also test the proposed method using a real world porcelain figurine sequence, which is the same sequence used in [17]. The sequence contains 36 images, with a resolution of 922×1158 under 3 different illuminations, each illumination is kept fixed over 12 images. Figure 6 shows the previous method produced an over smoothed surface under the chin while the proposed method captures the sharp shape due to the constraint of the
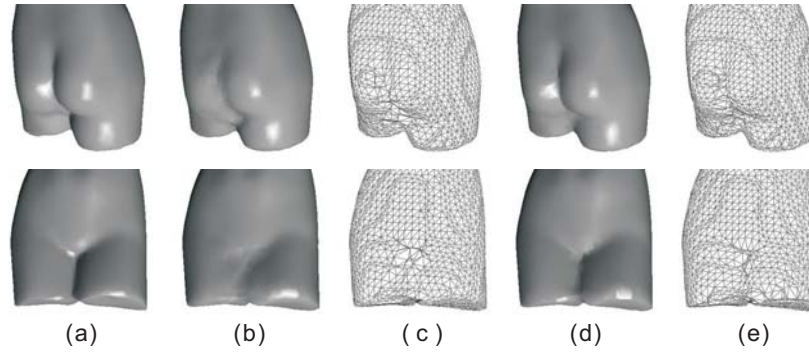
Figure 5: Comparison of the vertices between the previous method and the proposed method without regularization. (a) Ground truth. (b) Previous method. (c) Mesh structure of previous method suffers from serious self-crossing problem due to the incorrect photometric normals over the concavities and fails to recover the concavities without regularization. (d) Proposed method. (e) Mesh structure of the proposed method shows a better control over the mesh structure and recover the concavities successfully.

photometric depth. Figure 7 shows another example of reconstruction obtained using the proposed method.
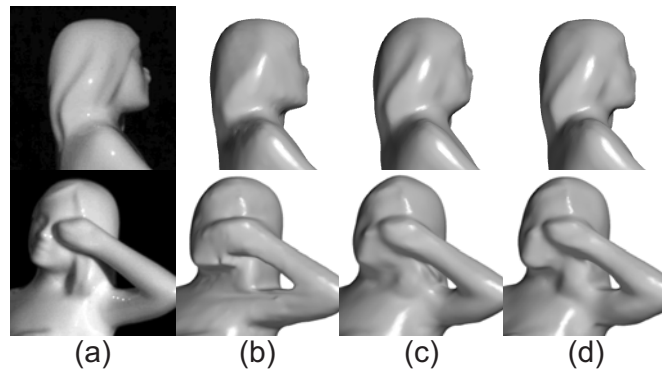

Figure 6: Close-up comparison between the previous method and the proposed method. (a) Real images. (b) Visual hull. (c) Previous method. (d) Proposed method.

6   Conclusion

In this paper, we present a novel method for recovering scene illumination and reconstructing a complete detailed 3D model for textureless objects. The proposed method utilizes the true surface points estimated by a dual space approach and directly recovers
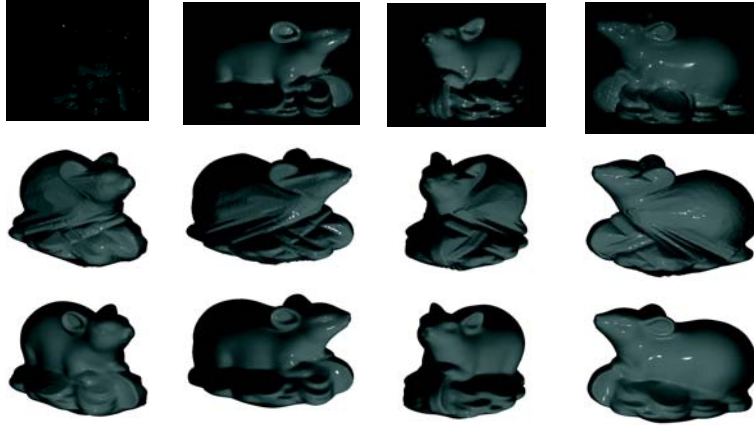
Figure 7: Reconstruction results using the proposed method. From the first row to the last, original image, visual hull, and reconstructed model.

the illumination with a closed-form solution. With the estimated illumination, a surface is recovered with concavities and other details by iteratively optimizing photometric information and vertex positions of the surface. Unlike existing approaches, besides photometric normal, we take the photometric depth into account during the optimization process, and this enables the approach to converge more effectively and be able to handle deep and sharp concavities. Experimental results demonstrate that the proposed method can achieve better result both quantitatively and qualitatively compared with other existing approaches.

Acknowledgement

References

[1] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *Int. Journal of Computer Vision*, 72(3):239–257, 2007.

[2] R. Cipolla, K. E. Åström, and P. J. Giblin. Motion from the frontier of curved surfaces. In *Proc. Int. Conf. on Computer Vision*, pages 269–275, 1995.

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

[4] J. S. Franco and E. Boyer. Exact polyhedral visual hulls. In *Proc. British Machine Vision Conference*, pages 329–338, 2003.

[5] M. Habbecke and L. Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[6] C. Hernndez, G. Vogiatzis, and R. Cipolla. Probabilistic visibility for multi-view stereo. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

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

[8] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo beyond lambert. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 171–178, 2003.

[9] N. Joshi and D.J. Kriegman. Shape from varying illumination and viewpoint. In *Proc. Int. Conf. on Computer Vision*, pages 1–7, 2007.

[10] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume I, pages 103–110, 2001.

[11] B. Kim and P. Burger. Depth and shape form shading using the photometric stereo method. *Computer Vision and Image Understanding*, 54(3):416–427, November 1991.

[12] R. Koch, M. Pollefeys, and L. van Gool. Multi viewpoint stereo from uncalibrated video sequences. In *Proc. European Conf. on Computer Vision*, volume I, pages 55–71, 1998.

[13] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(2):150–162, February 1994.

[14] C. Liang and K.-Y. K. Wong. Complex 3d shape recovery using a dual-space approach. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume II, pages 878 – 884, 2005.

[15] S. M. Seitz, B. Curless, J. Diebel, and D. S. R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume I, pages 519–526, 2006.

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

[17] G. Vogiatzis, C. Hernandez, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume II, pages 1847–1854, 2006.