Lighting-up geometry: accurate 3D modelling of museum artifacts with a torch and a camera

G. Vogiatzis[1] and C. Hernández[1] and R. Cipolla[2]

[1]Toshiba Research Europe, Cambridge Research Laboratory, St. George House, 1 Guildhall Streeet, Cambridge CB2 3NH, UK
[2]Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

Abstract

This paper addresses the problem of obtaining complete, very detailed reconstructions of shiny objects such as glazed ceramics. 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 initialise a multi-view photometric stereo scheme to obtain a closed surface reconstruction. The algorithm has been implemented as a practical model acquisition system. Here, we present a number of complete reconstructions of challenging real objects.

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 specialised, 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 [Lev02]. 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 [HS04], 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.

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 generalised photometric stereo scheme.

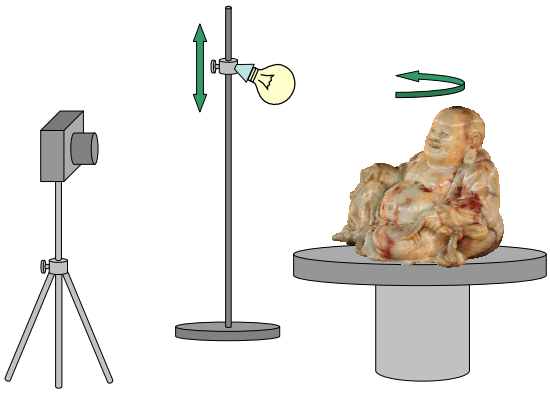

Figure 1: ***Acquisition Setup.*** *The object is rotated on a turntable in front of a camera and a point light-source. A sequence of images are captured while the light-source changes position between consecutive frames. No knowledge of the camera or light-source positions is assumed.*

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 toler-

ate 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.

Shape recovery from images is a well established task with two families of techniques offering the most accurate results, multi-view stereo (e.g. [KZ02, FK98, HS04]) and photometric stereo [Woo80]. 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. 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. 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 piece of sculpture. Due to this limitation existing photometric stereo techniques have so far only been able to extract depth-maps (e.g. [THS04]) with the notable recent exceptions of [ZCHS03, LHYK05], where the authors present techniques for recovering 2.5D reconstructions from multiple viewpoints. The full reconstruction of multi-sided objects is however still not possible by these methods. In previous work [VHC06], we have presented an algorithm that enables the reconstruction of un-textured single-albedo objects. In this work, we extend that algorithm to the general case of multiple albedo objects.

2. Algorithm

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 the technique presented in [MWC01], 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. 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. Both camera pose and illumination are estimated from the object's silhouettes.
- The full 3D geometry of a complex, shiny, 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.

2.1. 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, each surface point projects to a point of intensity given by:

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

where $\mathbf{l}$ is a 3D vector directed towards the light-source and scaled by the light-source intensity, $\mathbf{n}$ is the surface unit normal at the object location and λ is the albedo at that location. Equation (1) provides a single constraint on the three coordinates of the product $\lambda\mathbf{l}$. Then, given three points with an unknown but *equal* albedo λ, their normals, and the corresponding three image intensities, we can construct three such equations that can uniquely determine $\lambda\mathbf{l}$. For multiple images, these same three points can provide the light directions and intensities in each image up to a global unknown scale factor λ. The problem is then how to obtain three such points.

Our approach is to use the 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* [Lau94], 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* [CG99]. 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 equal albedo 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 [FB81] to separate inliers from outliers and estimate illumination direction and intensity. The scheme can be summarised as follows:

1. Pick three points on the visual hull and from their image intensities and normals estimate an illumination hypothesis.
2. All points of the visual hull vote for the illumination hypothesis *if* their predicted image intensities are within a threshold of the observed image intensities.
3. Repeat 1 and 2 a set number of times always keeping the illumination hypothesis with the largest number of votes.

2.2. 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 use an 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 $\mathbf{x_1} \dots \mathbf{x_M}$, 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 [JCYS04, VHC06] where the authors introduce a decoupling between the mesh normals, which depend on $\mathbf{x_1} \dots \mathbf{x_M}$, and the direction vectors used in the Lambertian model equation which become a set of independent variables $\mathbf{v_1} \dots \mathbf{v_F}$ which we call *photometric normals*. The minimisation cost is then composed of two terms, where the first term E_m brings the mesh normals close to the photometric normals through the following equation:

$$E_m\left(\mathbf{x}_{\mathbf{1},\dots,\mathbf{M}}; \mathbf{v}_{\mathbf{1},\dots,\mathbf{F}}\right) = \sum_{f=1}^{F} \|\mathbf{n_f} - \mathbf{v_f}\|^2 A_f, \qquad (2)$$

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

$$E_v\left(\mathbf{v}_{\mathbf{1},\dots,\mathbf{F}}, \lambda_{1,\dots,F}; \mathbf{x}_{\mathbf{1},\dots,\mathbf{M}}\right) = \sum_{f=1}^{F} \sum_{k \in \mathcal{V}_f} \left(\mathbf{l_k}^T \lambda_f \mathbf{v_f} - i_{f,k}\right)^2. \qquad (3)$$

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 and albedos.

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. 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 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. 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.

The algorithm was tested on three challenging shiny objects shown in figures 2 and 3. Thirty-six 3456×2304 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. The visual hull was then processed by the robust light estimation scheme of Section 2.1 to recover the distance light-source directions and intensities in each image. The photometric stereo scheme of section 2.2 was then applied. The results in figure 2 show reconstructions of very fine relief porcelain vases. The reconstructed relief (especially for the vase on the right) is less than a millimetre while their height is approximately 15-20 cm. Figure 3 shows a detailed reconstruction of a Buddha figurine made of polished soapstone. 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 multi-view stereo algorithm [HS04] was executed. The results are shown in the second row of Figure 3. It is evident that, while the low frequency component of the geometry of the figurine is correctly recovered, the high frequency detail 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.

4. Conclusion

We have demonstrated that the 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 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.

Figure 2: ***Reconstructing porcelain vases.*** *Top: sample of input images. Bottom: proposed method. The resulting surface captures all the fine details present in the images, even in the presence of strong highlights.*

Figure 3: ***Reconstructing coloured marble.*** *Top: Two input images. Middle: model obtained by multi-view stereo method from [HS04]. Bottom: proposed method. The resulting surface is filtered from noise while new high frequency geometry is revealed (note the reconstructed surface cracks).*

This paper has presented a novel reconstruction technique using silhouettes and the cue of photometric stereo 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.

References

[CG99] CIPOLLA R., GIBLIN P.: *Visual Motion of curves and surfaces.* Cambridge University Press, 1999.

[FB81] FISCHLER M., BOLLES R.: Random sample consensus: A paradigm for model-fitting with applications to image analysis and automated cartography. *CACM 24*, 6 (1981), 381–395.

[FK98] FAUGERAS O., KERIVEN R.: Variational principles, surface evolution, pdes, level set methods and the stereo problem. *IEEE Transactions on Image Processing 7*, 3 (1998), 335–344.

[HS04] HERNÁNDEZ C., SCHMITT F.: Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding 96*, 3 (December 2004), 367–392.

[JCYS04] JIN H., CREMERS D., YEZZI A., SOATTO S.: Shedding light in stereoscopic segmentation. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.* ((accepted) 2004).

[KZ02] KOLMOGOROV V., ZABIH R.: Multi-camera scene reconstruction via graph-cuts. In *Proc. European Conf. on Computer Vision* (2002), vol. 3, pp. 82–96.

[Lau94] LAURENTINI A.: The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 2 (1994).

[Lev02] LEVOY M.: Why is 3d scanning hard? Invited talk at 3D Processing, Visualization, Transmission, Padua, Italy, 2002.

[LHYK05] LIM J., HO J., YANG M., KRIEGMAN D.: Passive photometric stereo from motion. In *Proc. IEEE International Conference on Computer Vision* (2005).

[MWC01] MENDONÇA P., WONG K., CIPOLLA R.: Epipolar geometry from profiles under circular motion. *IEEE Trans. Pattern Anal. Mach. Intell. 23*, 6 (2001), 604–616.

[THS04] TREUILLE A., HERTZMANN A., SEITZ S.: Example-based stereo with general brdfs. In *Proc. European Conf. on Computer Vision* (may 2004).

[VHC06] VOGIATZIS G., HERNÁNDEZ C., CIPOLLA R.: Reconstruction in the round using photometric normals and silhouettes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2006).

[Woo80] WOODHAM R.: Photometric method for determining surface orientation from multiple images. *Optical Engineering 19*, 1 (1980), 139–144.

[ZCHS03] ZHANG L., CURLESS B., HERTZMANN A., SEITZ S.: Shape and motion under varying illumniation: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proc. 9th Int. Conf. on Computer Vision* (2003).