High Quality Photometric Reconstruction using a Depth Camera

Sk. Mohammadul Haque, Avishek Chatterjee, Venu Madhav Govindu

Indian Institute of Science

Bangalore - 560012, India

smhaque|avishek|venu@ee.iisc.ernet.in

Abstract

In this paper we present a depth-guided photometric 3D reconstruction method that works solely with a depth camera like the Kinect. Existing methods that fuse depth with normal estimates use an external RGB camera to obtain photometric information and treat the depth camera as a black box that provides a low quality depth estimate. Our contribution to such methods are two fold. Firstly, instead of using an extra RGB camera, we use the infra-red (IR) camera of the depth camera system itself to directly obtain high resolution photometric information. We believe that ours is the first method to use an IR depth camera system in this manner. Secondly, photometric methods applied to complex objects result in numerous holes in the reconstructed surface due to shadows and self-occlusions. To mitigate this problem, we develop a simple and effective multiview reconstruction approach that fuses depth and normal information from multiple viewpoints to build a complete, consistent and accurate 3D surface representation. We demonstrate the efficacy of our method to generate high quality 3D surface reconstructions for some complex 3D figurines.

1. Introduction

The availability of inexpensive depth cameras such as the Kinect has opened up new areas of work and also enabled researchers to revisit some classical problems. While the Kinect was originally intended for estimating human pose it has been used to reconstruct three-dimensional representations of a scene. We may broadly group such approaches into two categories, i.e. 3D model acquisition by a moving depth camera and depth-guided photometric methods. This paper is a contribution towards the later category of photometric approaches for which prior work includes [16, 15, 5].

Following the original contribution of [11], the methods of [16, 15, 5] develop a variety of approaches to combine depth information with estimates of the surface normals as given by photometric constraints. All of these approaches use an RGB image in combination with depth information. While [15] uses the Kinect's RGB image, the methods of [16, 5] use an additional external high-resolution RGB camera to obtain radiometric information of the scene. As will be described below, we use the same infra-red (IR) camera on the Kinect sensor to obtain both depth information as well as images for the photometric constraints. Like other methods we use the raw depth information provided by the Kinect's IR projector-camera pair. However, by switching off the IR projector through software, we are also able to obtain a raw IR radiometric image using the IR camera of the Kinect. This enables us to use the Kinect's IR camera for both depth *and* radiometric sensing, obviating the need for any additional RGB camera. We believe that our work is the first approach to use the Kinect's IR camera as a radiometric sensor by switching off the projector. Apart from not requiring an additional camera, using the IR camera for both depth and radiometric sensing has specific advantages that will be described later.

1.1. Relevant Literature

In our brief survey of relevant literature, we will only describe methods of direct relevance to our approach. Assuming a point source of lighting and a Lambertian reflectance model, the pioneering work of Woodham [14] estimated the normal of a surface point by observing it under lighting from three or more directions. Once the normals are estimated, they can be integrated to obtain a 3D representation of the surface being imaged. While effective, the photometric stereo method of [14] required precisely known calibration of multiple light sources. An alternate method proposed by Horn [8, 7] takes a single image and solves for a surface representation by variational minimisation using the Lambertian photometric constraint on the surface normals. This approach, known as shape-from-shading (SFS) has resulted in a very large body of literature of which [4, 9] are only some representative examples. Surveys of such approaches are available in [17, 3].
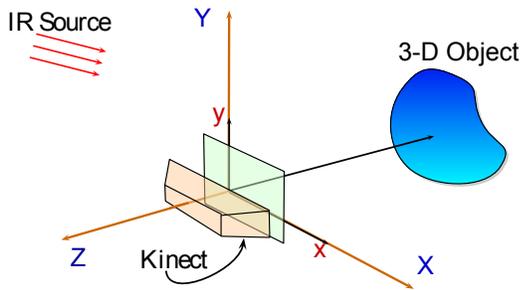
Figure 1. Device setup and axis conventions used in this paper.

Despite the large body of literature, the SFS problem is difficult to solve as the ill-posedness of the photometric constraint results in a local bas-relief ambiguity. To ameliorate this problem, Nehab *et. al.* [11] proposed to combine two independent sources of information, i.e. a depth map (obtained by stereo estimation or range imaging) and a normal map of the surface (estimated from photometric constraints). We follow this approach whereby the low frequency information obtained from a noisy Kinect depth map is fused with the high frequency feature information provided by photometric constraints obtained from the Kinect's IR camera. In [16], the authors modify the method of [11] to incorporate adaptive weighting for surface gradient estimation and also incorporate a Laplacian smoothness penalty. In [15], the authors use clustering of pixels to estimate relative albedos for an SFS method with hole repairing to improve the surface normals obtained from the Kinect. The method of [5] uses a combination of a global lighting model and a local lighting model to optimize normal estimation over a patch of pixels that are fused with a depth map.

2. Photometric Estimation

The first novelty of our approach is that we use the Kinect's IR camera for both depth and normal estimation. As mentioned earlier, our approach appears to be the first such method which uses the same IR camera for depth *and* radiometric sensing. A significant advantage of this approach is that we do not require an additional RGB camera. Consequently, apart from being cheaper, we do not need an additional step to calibrate the Kinect's depth map and the external RGB camera as is required by the methods of [16, 5]. In addition, when viewing complex surfaces, the Kinect's depth map would contain holes due to self-occlusion of the surface. Since the external RGB camera is displaced with respect to the Kinect, the RGB image would also contain additional occlusion regions and

holes. In contrast, in our setup, since we use the Kinect's IR camera, both the depth map and radiometric image are obtained in the same co-ordinate system. This reduces the number of holes in the surface representation.

Using the Kinect's IR camera : The Kinect consists of a depth sensor (IR projector and camera) and an RGB camera. While the depth map and RGB image outputs of the Kinect are 640×480 pixels, both the IR and RGB cameras actually have a true higher resolution of 1280×1024 pixels. Since the raw depth maps provided by Kinect are of lower resolution and tend to be rather noisy, we exploit the higher resolution of the IR camera to extract high resolution details with the aid of an IR source, see Fig. 1. By turning off the Kinect's IR projector and turning on the IR source, we can obtain high resolution IR images of the scene that can be utilised in a photometric sense to obtain high resolution details that cannot be observed in the coarse Kinect depth map.

2.1. Photometric Relations

Assuming a point source at infinity and Lambertian reflectance of the surface, we arrive at the simplest radiometric model for the observed image intensity,

$$I_p = A\alpha_p \max(0, \mathbf{N}_p \cdot \mathbf{V}) \tag{1}$$

where A is the strength of the light source at infinity and I_p and α_p are the observed image intensity and albedo of the p-th pixel. Furthermore, $\mathbf{N}_p$ is the normal at the surface point that is imaged at pixel p and $\mathbf{V}$ is the direction of the lighting source. Throughout we shall assume that the lighting directions are such that $\mathbf{V}$ lies in the half-sphere defined by $\mathbf{N}_p$, i.e. $\mathbf{N}_p.\mathbf{V} > 0$. Assuming uniform albedo we can simplify Eqn. 1 to account for the observed image under different lighting conditions as

$$I_p^k = \mathbf{N}_p \cdot \mathbf{S}^k \tag{2}$$

where $\mathbf{S} = A\alpha\mathbf{V}$ and the superscript k indexes different lighting conditions.

2.2. Estimating Lighting

Since the observed intensities are a product of both lighting and normal directions, we can solve for one given the other. Since a noisy depth map is provided by Kinect, we can utilise it to obtain a preliminary estimate of the surface normals. In our approach, we apply a bilateral filter to smoothen the Kinect's depth map and obtain normal estimates which can be used in Eqn. 2 to linearly solve for the individual lighting $\mathbf{S}^k$. In practice, to account for various errors, we use a robust estimator to solve for $\mathbf{S}^k$,

(c) Our Normal Weighting

(a) Edge Weighting   (b) 3D Model
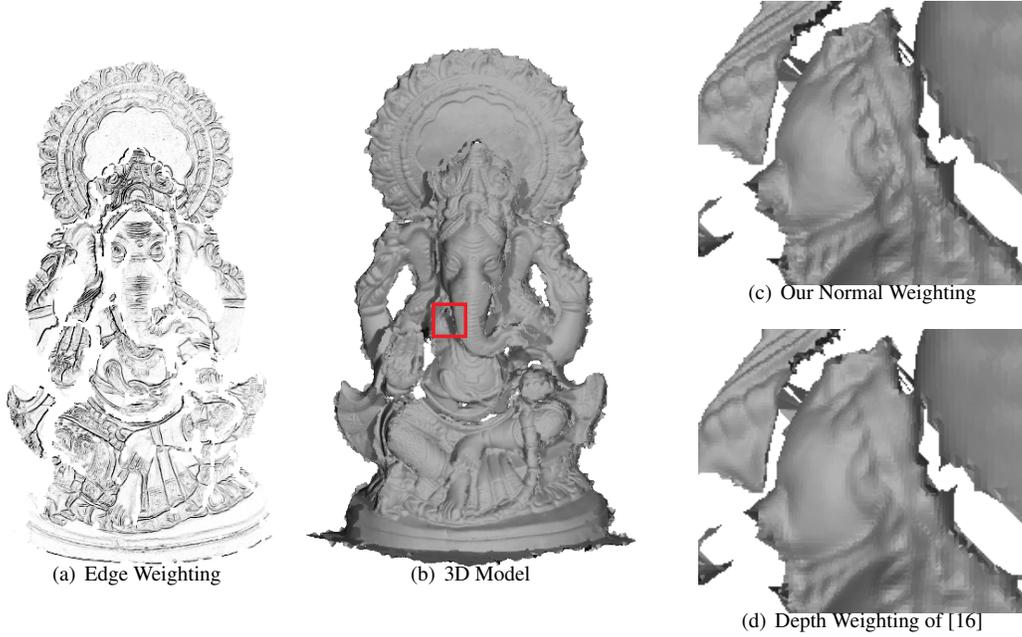
(d) Depth Weighting of [16]

Figure 2. Weighting functions used in our fusion algorithm. (a) shows the edge weighting measure used in Eqn. 7. (c) and (d) show the reconstructions obtained using our normal weighting function of Eqn. 10 and that of [16] respectively. A higher level of detail is preserved in our reconstruction. (b) shows the 3D object where the region corresponding to (c) and (d) is indicated in red. Please view this figure in colour.

$$\min_{\mathbf{S}^k} \sum_i \rho_\sigma \left(I_p^k - \mathbf{N}_p \cdot \mathbf{S}^k \right) \qquad (3)$$

where $\rho(x)$ is the Huber loss function given by

$$\rho_\sigma(x) = \begin{cases} x^2 & if \quad |x| \leq \sigma \\ 2\sigma |x| - \sigma^2 & if \quad |x| > \sigma \end{cases} \qquad (4)$$

In solving Eqn. 3 we iteratively estimate the appropriate scale factor of σ from the data itself. From the residual errors of fitting Eqn. 3, we compute the median absolute deviation (MAD) and set $\sigma = 1.48$ MAD. This normalized median absolute deviation is a robust statistical measure of the variability of a univariate sample [12]. The estimated σ is iteratively updated till convergence.

For the Lambertian assumption used, $\mathbf{S} \in \mathbb{R}^3$ represents the global lighting function. Despite the fact that the Kinect depth map is not very accurate, we can accurately recover $\mathbf{S}$ since it has only 3 unknowns and we have as many observations as 640×480 depth map pixels. Thus, unlike photometric setups that require accurate calibration of the lighting sources, we are able to estimate the lighting direction from the Kinect depth map itself.

2.3. Normal Estimation

Having obtained the light sources $\mathbf{S}^k$, we can re-estimate the normals using photometric constraints. Given 3 or more

lighting conditions, from Eqn. 2 we can linearly and independently solve for the normals of all pixels. Using the Huber loss function allows us to account for shadows (i.e. $I_p^k = 0$) and robustly estimate the normals as

$$\min_{\mathbf{N}_p} \sum_k \rho \left(\mathbf{I}_p^k - \mathbf{S}^k \cdot \mathbf{N}_p \right). \qquad (5)$$

3. Fusion of Depth and Normal Estimates

Once we have the normals estimated for individual pixels, we can fuse the depth and normal estimates to generate a 3D representation of the surface. In this section we describe our fusion method that is adapted from those presented in [11, 16] but also crucially differs in certain key attributes. Specifically, we choose appropriate weighting schemes that carefully account for the relative reliabilities of the depth and normal estimates used. Consider the 2D depth map $Z(x, y)$ which represents the depth of the surface point that is imaged at pixel location (x, y). Therefore, for focal length f we have the location of the 3D point as

$$\mathbf{P}(x, y) = \left[-\frac{x}{f} Z(x, y) \quad -\frac{y}{f} Z(x, y) \quad Z(x, y) \right]^T \qquad (6)$$

The cost function that we optimise in our fusion scheme consists of three terms that capture different attributes that we desire in our final reconstruction.

3

Depth Information : For the observed depth $\mathbf{Z}$ and estimated depth $\hat{\mathbf{Z}}$, the depth component of our cost function E_d is given as

$$E_d\left(\hat{\mathbf{Z}}\right) = \sum_p w_p \|\mu_p\|^2 \left(Z_p - \hat{Z}_p\right)^2 \qquad (7)$$

where for the p-th pixel, $\mu_p = \left[-\frac{x}{f} \quad -\frac{y}{f} \quad 1\right]^T$ and w_p is an adaptive weighting function. The weighting by the norm $\|\mu_p\|$ penalises the net displacement in 3D for a given change in Z. While existing fusion approaches [11, 16] use a uniform weighting for all depth pixels, in our approach we use an adaptive weight that reflects the expected reliability of the depth map. Most depth map estimates, including that of the Kinect, are less reliable at a depth discontinuity and the depth information should correspondingly have a lower influence. Since the normal estimates provided by Eqn. 5 are locally more reliable than the Kinect's depth estimate, we use this normal information to adaptively weight the depth estimate. For every pixel p in the depth map, we compute the normal tensor $\mathcal{N} = \sum_{q \in S(p)} \mathbf{N}_q \mathbf{N}_q{}^T$ where $S(q)$ is a patch centered on p. Following [10], we define an edge saliency measure based on the eigen-values of $\mathcal{N}$. For eigen-values $\lambda_3 \geq \lambda_2 \geq \lambda_1$, our edge saliency is given as $\frac{\lambda_2 - \lambda_1}{\lambda_3}$ which is large for strong edges and small for flat regions. In turn, we define our weighting function as i.e. $\eta = 1 - \frac{\lambda_2 - \lambda_1}{\lambda_3} \in [0, 1]$. Accordingly we adaptively vary the weighting function $w_p = w_0 + (1 - w_0)\eta_p$. We typically set $w_0 = 0.9$. Fig. 2(a) shows a representative weighting function obtained using our edge saliency measure where the depth information is given lower weightage at sharp changes or discontinuities in depth.

Normal Information : For fusing normal information, we need to account for the consistency of the estimated depth with the observed normals given by solving Eqn. 5. From Eqn. 6 we have surface tangents at 3D point $\mathbf{P}(x, y)$ specified as

$$\mathbf{T}_x = \frac{\partial \mathbf{P}}{\partial x} = \left[-\frac{1}{f}\left(Z + x\frac{\partial Z}{\partial x}\right) \quad -\frac{1}{f}y\frac{\partial Z}{\partial x} \quad \frac{\partial Z}{\partial x}\right]^T \quad (8)$$

$$\mathbf{T}_y = \frac{\partial \mathbf{P}}{\partial y} = \left[-\frac{1}{f}x\frac{\partial Z}{\partial y} \quad -\frac{1}{f}\left(Z + y\frac{\partial Z}{\partial y}\right) \quad \frac{\partial Z}{\partial y}\right]^T.$$

It will be observed that the tangents defined in Eqn. 8 are linear in the depth Z. If for a given estimate of depth $\hat{\mathbf{Z}}$, the tangents are to be consistent with the normal information, then ideally the projection of the tangents along the normal direction have to be equal to zero. With this in mind we can define a penalty term based on the normal information as

$$E_n\left(\hat{\mathbf{Z}}\right) = \sum_p (\mathbf{N}_p \cdot \mathbf{T}_x)^2 + (\mathbf{N}_p \cdot \mathbf{T}_y)^2 \qquad (9)$$

(a) RGB Image of a T-shirt

(b) 3D Reconstruction

Figure 3. RGB image of a T-shirt draped on a chair and its 3D reconstruction. In (a), we have cropped out the background in the RGB image for easy visualisation. Our method is able to accurately recover minute changes in depth such as the horizontal stitch at the bottom of the T-shirt.

where the tangent vectors $\mathbf{T}_x$ and $\mathbf{T}_y$ are evaluated on the estimated depth map $\hat{\mathbf{Z}}$ at pixel p. While our normal penalty term is similar in spirit to that of [11] we use a weighting scheme to adaptively improve the estimates for the tangent vectors. In [16], the authors compute the depth derivatives by adaptively weighting the forward and backward difference. Their weights are dependent on the similarity of pixel depths, see Eqn. 10 of [16]. In contrast we choose to weight the forward and backward difference terms of the derivative by a term that depends on the similarity of the estimated normals of the pixels. Thus, while

calculating derivatives for estimating tangents, for two adjacent pixels p and q that are part of a difference function, our weighting term is given as

$$w(p, q) = exp^{-\frac{1}{2\sigma^2}(1 - \mathbf{N}_p^T \mathbf{N}_q)} \qquad (10)$$

It will be observed from Eqn. 10 that our weighting function can adaptively refine the tangent estimates depending on the local curvature of the surface. In Fig. 2(c) and (d) we show a comparison of results obtained using our normal based weighting function of Eqn. 10 and the depth-based weighting function of [16] respectively. As can be observed, our weighting function helps preserve fine features to a greater extent than that of [16].

Laplacian Smoothness : In addition, we also use a smoothness cost E_s given by the discrete version of the conventional Laplacian penalty $\nabla^2(\hat{\mathbf{Z}})$.

Combining the three penalty terms we can represent the solution for fusion of depth and normal estimates as the minimisation of the cost function

$$E\left(\hat{\mathbf{Z}}\right) = E_d\left(\hat{\mathbf{Z}}\right) + \lambda_n E_n\left(\hat{\mathbf{Z}}\right) + \lambda_s E_s\left(\hat{\mathbf{Z}}\right) \qquad (11)$$

We note that Eqn. 11 can be efficiently solved as a linear system of equations for the resultant depth map $\mathbf{Z}$. The reader may also note that since we optimize over the depth map, unlike the methods of [9, 5], we do not need to explicitly enforce integrability of the normal map.

4. Results

In this section we present some results of using our method for 3D surface reconstruction. In all cases we use a single Kinect and a single bright light source that also has an IR component. Since we robustly estimate the lighting $\mathbf{S}$ we can freely place the light source without requiring a calibrated setup. Throughout we first upsample the raw Kinect depth map to match the IR image size of 1280×960 pixels and apply a bilateral smoothing filter to remove noise and high frequency artifacts that are often present in raw Kinect depth maps. Subsequently we further upsample both the IR and depth map images by a factor of 2. Such upsampling provides our reconstruction method with an increased number of vertices. The extra vertices allow us to accurately recover narrow 3D edges that might be as thin as a single pixel in the original IR images. We also note here that in our scenario, to acquire fine scale details, we need to place the Kinect close to the object. This results in the blurring of the Kinect's projected pattern that is used in depth estimation. To mitigate this problem, we reduce the brightness of the projector by software and are able to place the Kinect as close as 40 cm to the object being reconstructed.

In Fig. 3 we show our reconstruction of a T-shirt draped on a chair. This reconstruction is obtained by using 5 different lighting conditions to estimate the pixelwise normals in Eqn. 5 which is used in the fusion step of minimizing Eqn. 11. As can be seen we are able to accurately capture the 3D shape of the T-shirt including a fine scale horizontal stitch at the bottom. We further illustrate the ability of our method in Fig. 4(a) that shows our reconstruction of a complex clay figurine of the Hindu god Ganesh that is about 40 cm in height. As the reconstructions shown in Fig. 4(b) demonstrate, we are able to obtain an accurate 3D reconstruction that preserves many of the fine scale details and depth discontinuities.

5. Multiview Reconstruction

Although the results in Sec. 4 demonstrate that our fusion method can accurately recover 3D information, reconstructing surfaces from a single viewpoint is often inadequate. The 3D objects we have chosen for our experiments are challenging due to their shape complexity. As a result, despite using the Kinect's IR camera for depth and normal estimation, the reconstructions from a single viewpoint contain a large number of holes due to shadow regions, self-occlusions etc. and are inadequate for representing the full object. This is quite evident from the reconstruction in Fig. 4(a) where the 3D details are accurately preserved but many holes are also present. Consequently, we need to reconstruct the object surface from multiple viewpoints and then merge the individual reconstructions to present a unified representation for the object of study. It may be noted here that all the results presented in [11, 16, 15, 5] are single viewpoint reconstructions applied to relatively simpler surfaces compared to the objects used for our results.

The canonical geometric approach to merging individual 3D representations is the Iterative Closest Point (ICP) algorithm [13]. However, we note that the reconstructions from individual viewpoints obtained by optimizing Eqn. 11 are independent of each other. Therefore, while each reconstruction estimates the 3D surface by fusing depth and normal information, they do not enforce consistency across the reconstructions. The result is that the individual reconstructions are no longer rigidly related to each other. The ICP algorithm assumes that the surfaces being registered are rigidly related by a single 3D rotation and translation. In Fig. 5 we show an example of the registration of two reconstructions of a Buddha figurine which are represented in two different colours. While the two surfaces are well aligned in a global sense, the local surface non-rigidity

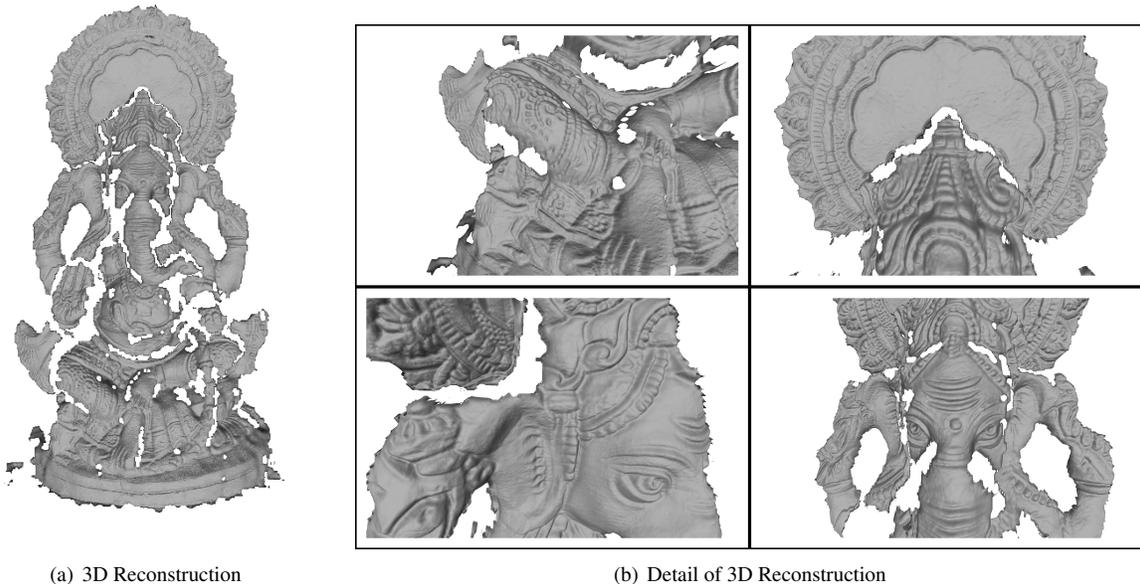(a) 3D Reconstruction   (b) Detail of 3D Reconstruction

Figure 4. (a) shows our reconstruction of a clay figurine of the Hindu god Ganesh. (b) shows some of the details of this reconstruction. Please note that these details are from different viewpoints and all do not corresponding to the reconstruction in (a).

of the two surfaces is evident from the patchiness of the coloured regions. In this figure we have also indicated a cross-section of the registered representations which clearly shows the non-rigid deformations. An important consequence of the non-rigid nature of our reconstructions is that a straightforward geometric registration of individual reconstructions is inadequate. Instead, we require an approach to unify the information across the individual reconstructions to generate a single 3D object representation.

Although the use of multiview photometric stereo is limited in the literature, there do exist some significant methods that utilise all photometric constraints available to solve for a unified object representation [6, 1]. In our case, since we can accurately estimate the surface normals in each individual reconstruction we choose to use them directly instead of indirectly fitting all available photometric constraints. In our experiments, we took multiple views of the objects, each with 5 different lighting conditions. To generate a unified 3D representation for the object, we first carry out individual reconstructions and create meshes from each of the individual viewpoints. These individual reconstructions are subsequently aligned by using a robust version of the point-to-plane distance based Iterative Closest Point (ICP) [13] algorithm. From these registered meshes, we build a single representation using the volumetric merging algorithm (VCG) of [2] that provides an *intermediate* reconstruction. Although

the merged model is a complete representation, since the individual reconstructions are inherently non-rigid and the ICP routine may introduce registration inaccuracies we need to refine this reconstruction to develop a unified and consistent estimate of the 3D surface. This refinement can be achieved by recovering the normal information from the individual reconstructions as the normals are an accurate representation of the local nature of the surface.

For every vertex in the mesh generated by VCG, we shoot rays back into the individual IR camera planes and recover the corresponding normals estimated using Eqn. 5. It will be noted that since we have a unified representation, each vertex will recover normals from only those individual views that are imaging the said vertex. Since the VCG representation is an intermediate one that averages the individual non-rigid representations, averaging these normals will result in the loss of detail. Instead, to be able to preserve detail, we select only one of the available normals using a priority ordering of the individual reconstructions. In our scheme, we prefer normals from the frontal scan over those with a larger viewing angle with respect to the frontal scan. Thus, for each vertex, whenever it is available, we pick a corresponding normal from the frontal scan and select normals from other scans only when necessary. Such a priority ordering ensures that except at viewpoint transition boundaries, neighbouring vertices on the mesh will have normals obtained from the same
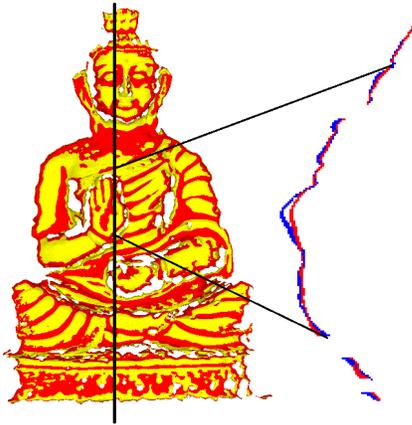
Figure 5. Individual reconstructions are non-rigidly related to each other. The two registered reconstructions are shown in different colours. A zoomed-in cross-section also illustrates the non-rigid deformations present. Please view this figure in colour.

individual reconstruction. Now we have both a depth value and normal associated with each mesh vertex which are fused using the mesh-based method of [11]. This results in a single unified and consistent 3D reconstruction.

We show the results of our full multiview reconstruction method on two complex clay figurines of Ganesh and the Buddha in Fig. 6. As can be seen, our approach provides an accurate and complete photometric reconstruction of complex objects using an IR-based depth camera for both depth and normal estimation.

6. Conclusions

In this paper we have presented an approach to combine depth and normal estimates. Apart from introducing a novel technique that utilises a depth camera system for both depth and radiometric measurements, we estimate lighting direction as well as develop adaptive weighting mechanisms that carefully balance the relative reliability of depth and normal estimates. To estimate the 3D representation of complex objects we develop a multiview reconstruction method that solves for a complete, consistent and accurate representation of complex objects.

References

[1] M. Beljan, J. Ackermann, and M. Goesele. Consensus multi-view photometric stereo. Pattern Recognition, pages 287–296, 2012.

[2] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In SIGGRAPH, SIGGRAPH '96, pages 303–312. ACM, 1996.

[3] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. Computer Vision and Image Understanding, 109(1):22–43, 2008.

[4] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10:439–451, 1988.

[5] Y. Han, J.-Y. Lee, and I. S. Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. In Proceedings of IEEE International Conference on Computer Vision (ICCV), 2013.

[6] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(3):548–554, 2008.

[7] B. K. Horn. Obtaining shape from shading information. MIT press, 1989.

[8] B. K. Horn and M. J. Brooks. The variational approach to shape from shading. Computer Vision, Graphics, and Image Processing, 33(2):174–208, 1986.

[9] M. Johnson and E. Adelson. Shape estimation in natural illumination. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011.

[10] G. Medioni, M. Lee, and C. Tang. A Computational Framework for Segmentation and Grouping. Elsevier Science, 2000.

[11] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. ACM Transactions on Graphics (TOG), 24(3):536–543, 2005.

[12] P. J. Rousseeuw and A. M. Leroy. Robust regression and outlier detection. John Wiley & Sons, Inc., 1987.

[13] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, pages 145–152. IEEE, 2001.

[14] R. J. Woodham. Photometric method for determining surface orientation from multiple images. Optical engineering, 19(1):191139–191139, 1980.

[15] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of rgb-d images. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, pages 1415–1422. IEEE Computer Society, 2013.

[16] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2472–2479. IEEE, 2012.

[17] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21(8):690–706, 1999.

(a) Different views of our reconstruction of the Ganesh figurine

(b) Different views of our reconstruction of the Buddha figurine

Figure 6. Unified multiview photometric reconstructions of (a) Ganesh and (b) Buddha both of which are about 40 cm in height.