

# Exploiting Shading Cues in Kinect IR Images for Geometry Refinement

Gyeongmin Choe Jaesik Park\* Yu-Wing Tai In So Kweon

Korea Advanced Institute of Science and Technology, Republic of Korea

[gmchoe, jspark]@rcv.kaist.ac.kr, yuwing@kaist.ac.kr, iskweon77@kaist.ac.kr

## Abstract

In this paper, we propose a method to refine geometry of 3D meshes from the Kinect fusion by exploiting shading cues captured from the infrared (IR) camera of Kinect. A major benefit of using the Kinect IR camera instead of a RGB camera is that the IR images captured by Kinect are narrow band images which filtered out most undesired ambient light that makes our system robust to natural indoor illumination. We define a near light IR shading model which describes the captured intensity as a function of surface normals, albedo, lighting direction, and distance between a light source and surface points. To resolve ambiguity in our model between normals and distance, we utilize an initial 3D mesh from the Kinect fusion and multi-view information to reliably estimate surface details that were not reconstructed by the Kinect fusion. Our approach directly operates on a 3D mesh model for geometry refinement. The effectiveness of our approach is demonstrated through several challenging real-world examples.

## 1. Introduction

Over the past few years, the Microsoft Kinect has become a popular input device in depth acquisition for human pose recognition [18], and 3D reconstruction [8]. The Kinect utilizes active range sensing by projecting a structured light pattern, *i.e.* the speckle pattern, on a scene in infrared (IR) spectrum. Through analyzing the displacement of the speckle pattern, depth map of a scene can be estimated. The success of Kinect relies heavily on the usage of IR pattern and the narrow band IR camera which filtered out most undesired ambient light that makes the depth acquisition robust to natural indoor illumination. Although the narrow band IR camera is one of the key component to the success of Kinect, after depth acquisition, these IR images are discarded and not used in any post-processing applications. In this paper, we show that the narrow band IR camera of Kinect is not only useful in capturing the speckle pattern for depth estimation, but also useful to capture shading cues of

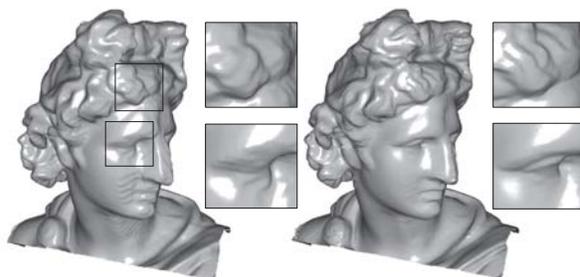


Figure 1. Comparison of a real data - Apollo. Left: 3D model from the Kinect fusion. Right: Our refined 3D model using the IR shading cues. The 3D mesh is rendered with the Phong-shaded model.

a scene which allows higher quality reconstruction than the popular Kinect fusion [8] that uses only the estimated depth map for 3D reconstruction.

We model the light from the IR projector of Kinect as a near point light source where its illumination falloffs according to the inverse square law. With the Lambertian BRDF assumption about the scene materials in the narrow band IR spectrum, we define a near light IR shading model which describes the captured intensity as a function of surface normals, albedo, lighting direction, and distance between a light source and surface points. The proposed model has ambiguity in normals and distance estimation using a single shading image. Therefore, we utilize an initial 3D mesh from the Kinect fusion and shading images from different view point. Our approach operates directly on a 3D mesh and optimizes the geometry refinement process subject to the shading constraint of our model. The result is a high quality mesh model which captures surface details that were not reconstructed by the Kinect fusion as shown in Figure 1. Thanks to the usage of the Kinect narrow band IR camera, our approach is also robust to indoor illumination which works well in both dark room or natural lighting environment. In addition, our approach does not require additional camera nor complicated setup which makes it useful in practical scenario as an add-on to enhance reconstruction results from the Kinect fusion.

\*The first and the second authors provided equal contributions to this work.

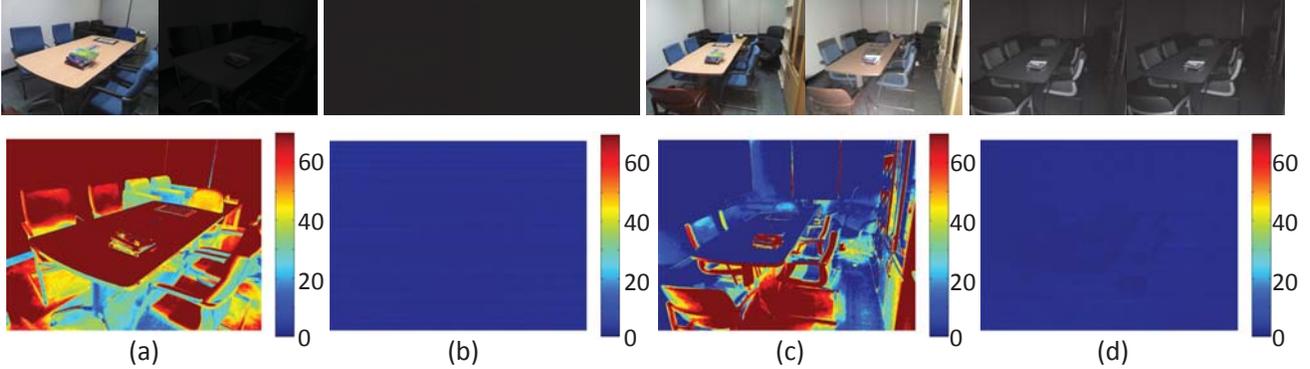


Figure 2. Invariability of Kinect IR images under different lighting conditions. (a) RGB images under ambient light and dark room. (b) The corresponding Kinect IR images of (a). (c) RGB images under ambient light and dark room with an additional wide spectrum light source. (d) The corresponding Kinect IR images of (c). The difference images are shown below each of the image pairs. Enormous differences are observed in the RGB image pairs while the IR image pairs are almost identical.

## 2. Related Works

Representative works that utilize shading information for geometry refinement are reviewed in this section. These techniques can be classified into 2D depth map refinement methods and 3D mesh refinement methods.

In depth map refinement, Nehab *et al.* [12] first proposed to use photometric stereo [6] to estimate surface normals to refine depth maps captured by a 3D laser scanner to achieve high quality geometry refinement. Work in [11] utilizes a giga-pixel camera to estimate ultra high resolution surface normals from photometric stereo to refine a low resolution depth map captured by using a structured light. Work in [1] uses shading information to improve depth map from ToF camera, but they assumed the only light source is from the ToF camera which limited their application in a controlled environment. Thanks to the recent development of Kinect, depth map of a scene can be easily acquired at low cost. But the Kinect depth map usually contains holes and noise which makes it less useful when a high quality depth map is required. Utilizing the additional RGB image from Kinect, methods in [22, 2, 14, 17] define a smoothness cost according to image structures in RGB image for depth map refinement, but their approaches do not use any shading information. Hence, they also lose fine depth details during the smoothing process. In [25, 13], they also use normals from photometric stereo to refine depth map with additional consideration to depth discontinuities [25] and first-order derivative of surface normals [13]. Recent works in [23, 4] propose to use shape-from-shading [7] from an RGB image instead of using photometric stereo to estimate surface details for depth map refinement.

In 3D mesh refinement methods, they typically start with a rough 3D mesh model estimated by using stereo matching [16], or structure from motion [10]. Similar to the 2D depth map refinements, Hernandez *et al.* [5] demonstrate to use normals from photometric stereo to refine mesh model

from multi-view stereo. Lensch *et al.* [9] introduce a generalized method for modeling non-Lambertian surfaces by wavelet-based BRDFs and use it for mesh refinement. Vlasic *et al.* [20] first integrate per-view normal maps into partial meshes, then deforms them using thin-plate offsets to improve the alignment while preserving geometric details. Wu *et al.* [21] use the multi-view stereo to solve the shape-from-shading ambiguity. They demonstrated high-quality 3D geometry under the arbitrary illumination but assume the captured objects contain only a single albedo. Park *et al.* [15] refine 3D mesh in parameterized space and demonstrate state-of-the-art quality in geometry refinement results using normals from photometric stereo.

Comparing our work to the previous works, especially for the 3D mesh refinement methods, most of them utilize photometric stereo to estimate normal details. Although high quality surface details can be estimated by photometric stereo, as demonstrated in the experimental setting in [5, 20, 15, 24], they require a complicated setting to calibrate lighting direction and to control environment illumination. In contrast, our work utilizes the Kinect narrow band IR camera which makes our approach robust to natural indoor illumination. In addition, we define a near light shading model which fits perfectly to our problem setting to utilize a near point light source instead of directional light source for normal estimation. Since our work directly operates on a mesh model, our approach is also efficient and effective in mesh model refinements.

## 3. Kinect IR images and data capturing

In this section, we first analyze the Kinect IR images. After that, we define our near light IR shading model based on our observation about Kinect IR images. Our processes for IR radiometric calibration and data capturing are also included in this section.

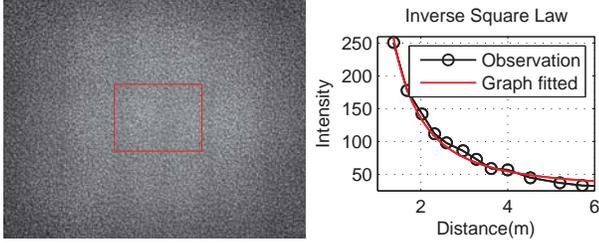


Figure 3. Validation of the inverse square law. Left: Region of Interested (ROI) in IR image. Various images at different depth are captured. The median intensity within each ROI is plotted. Right: Intensity falloffs with increasing distance. The falloff rate follows the inverse square law.

### 3.1. Kinect IR Images

We verify the invariability of Kinect IR images under different lighting conditions. In Figure 2, we blocked the Kinect IR projector and then capture IR images under ambient light and dark room environment. The RGB image pairs in (a) show enormous intensity differences under the two different lighting conditions, but the IR image pairs in (b) are almost identical. Next, we put a wide spectrum light source and then capture the RGB and IR images again under the same ambient light and dark room environment. Again, enormous intensity differences are shown in RGB image pairs in (c) and the IR image pairs in (d) have almost no difference. The differences in IR images in (d) is only caused by the additional wide spectrum light source. This example shows that common indoor lighting does not cover the IR spectrum captured by the Kinect IR camera. Unless a wide spectrum light source is presented in a scene, the Kinect IR images is unaffected by indoor ambient lighting.

Our second analysis verifies the inverse square law property of the Kinect IR projector. We capture IR images at various distance from a white wall. Figure 3 shows a captured IR image<sup>1</sup> and the plotting of median intensity with the red box area. The decay of observed intensity follows the inverse square law.

### 3.2. Near light IR shading Model

We define our near light IR shading model as follows:

$$I(\mathbf{u}) = \frac{c\rho}{d^2}(\mathbf{n} \cdot \mathbf{l}) + I_{Ambient}, \quad (1)$$

where  $I(\mathbf{u})$  is observed intensity at point  $\mathbf{u}$ ,  $\rho$  is albedo of surface,  $c$  is global brightness,  $\mathbf{n}$  is surface normal,  $\mathbf{l}$  is lighting direction, and  $d$  is the distance between the surface point and center of light source. Here, we assume the captured materials in IR spectrum follow the Lambertian BRDF model.

Following our analysis in the previous subsection, we set  $I_{Ambient} = 0$  since most indoor ambient lights are invisible

<sup>1</sup>The IR image is radiometrically calibrated.

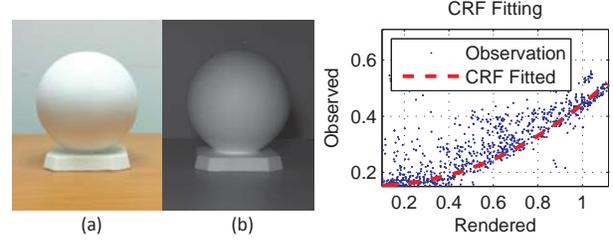


Figure 4. Estimation of camera response function (CRF) of Kinect IR camera. Left: (a) Color and (b) IR image of the spherical object for calibration. Right: Curve fitting for CRF estimation. The x-axis shows the rendered intensities from base mesh and the y-axis shows the measured intensities in (b).

to the Kinect IR camera and we denote  $\mathbf{l}$  as the only point light source in a scene. The inverse square term  $d$  is added to account for the light falloff property of Kinect IR projector. Note that different pairs of  $d$  and  $n$  can produce identical intensity assuming known albedo and lighting direction. Utilizing initial mesh from Kinect, and multiple view point information, we will show that this shading model is an effective constraint for geometry refinement without solving  $d$  and  $n$  for each view individually.

### 3.3. Radiometric Calibration of IR camera

We note the responses of the Kinect IR camera is not strictly linear to the luminance of incoming light. Therefore, we need to radiometrically calibrate the Kinect IR camera. In previous works for radiometric calibration [3], multiple different exposure images can be easily captured for calibration. However, the Kinect IR camera can only capture single exposure image. In addition, there is no calibration pattern for IR camera calibration. Here, we propose a radiometric calibration method which makes use of multiple photometric observations of a known geometry to estimate the camera response function (CRF) of the Kinect IR camera.

We use a white Lambertian sphere as shown in Figure 4(Left) for our calibration. The white sphere has a known geometry and a complete observation of surface normals in every directions. We use the Kinect fusion to obtain a base mesh of the sphere, and then capture the IR shading images of the sphere. Since the geometry, the distance, the lighting direction and the albedo are known for this calibration object, we can synthetically render a predicted observation using Equation (1). By comparing the measured intensities,  $I_{obs}$ , with the predicted intensities,  $I_{ren}$ , we can estimate the CRF,  $f$ , by fitting a curve that minimize the least square errors,  $\|I_{obs} - f(I_{ren})\|^2$ , as illustrated in Figure 4(Right). Here, we assume  $f$  is a gamma function where  $I_{obs} = (I_{ren})^\gamma$ . In our estimation, we find that the gamma value is approximately equal to 0.8.

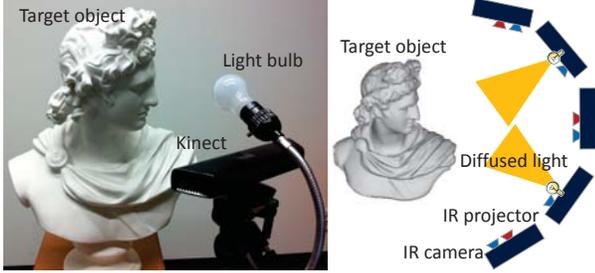


Figure 5. Our data capturing system. We use the Kinect fusion to obtain an initial base mesh. During the Kinect Fusion process, at certain camera positions, IR camera is blocked and a diffuse light is turned on for capturing shading images.

### 3.4. Data Capturing

Figure 5 shows our data capturing system. Our data capturing process is composed of continuous Kinect fusion and discrete IR shading image acquisition. An additional wide spectrum point light source is used because we cannot switch the speckle pattern to uniform IR light from Kinect IR projector using the Kinect SDK. During the Kinect fusion, we block the Kinect IR projector and turn on the wide spectrum light source to capture shading images at multiple locations. Note that, this process can be simplified and the Kinect IR projector can be used if the pattern from the Kinect IR projector is programmable. Since indoor ambient light does not affect the captured IR image, our data acquisition is performed under natural indoor lighting.

After data capturing, we have a base mesh from the Kinect fusion, and IR shading images from multiple discrete view points. Note that the locations where we captured shading images belongs to the subset of camera poses during Kinect fusion. The camera poses are estimated using the Kinect SDK by registering Kinect depth map with the current reconstructed surface. The relative location of the additional wide spectrum point light source and the Kinect IR camera is fixed and pre-calibrated. Therefore, lighting direction,  $\mathbf{l}$  in Equation (1), is known after data capturing. The Kinect IR images contain considerable amount of noise. Therefore, we average more than 30 input images from the same view point to obtain a denoised shading image.

## 4. Geometry Refinement

In this section, we describe our geometry refinement process via optimization. We begin this section by estimating surface albedo of geometry. Next, we describe our mesh refinement process by finding an optimal displacement of mesh vertices that satisfied our near light IR shading model.

We denote  $\mathbf{x}_i$  as a vertex on the base mesh  $M$  from the Kinect fusion,  $\mathbf{x}_j \in N(\mathbf{x}_i)$  as the neighboring vertices that directly connect to  $\mathbf{x}_i$ ,  $\mathbf{K}$  and  $\mathbf{P}_m$  are the intrinsic and extrinsic projection matrices of camera poses from the  $m$ -th

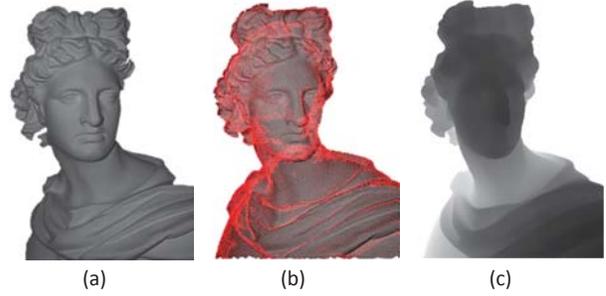


Figure 6. (a) One of our input shading images. (b) Projected mesh vertex (red dots) on (a). (c) Depth map derived from a projected mesh model. Note that the derived depth map from the Kinect fusion is far more accurate than the RAW depth map from Kinect. In our geometry refinement process, we use this depth map instead of the Kinect RAW depth map for mesh optimization.

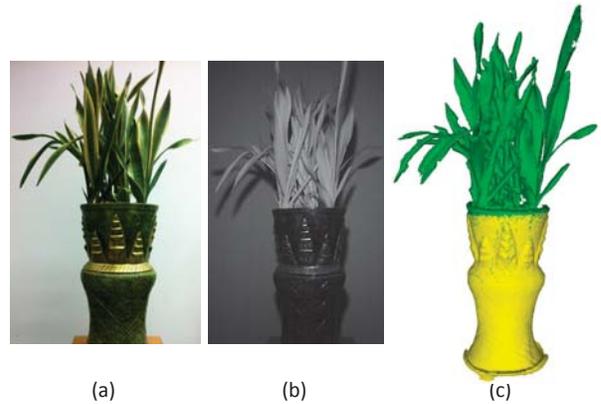


Figure 7. Albedo grouping. (a) Color image (b) IR shading image. (c) Color labels of grouped albedo.

view, and  $\mathbf{u}_{i,m} = \mathbf{K}\mathbf{P}_m\mathbf{x}_i$  is the image coordinate of vertex  $\mathbf{x}_i$  projected on the  $m$ -th view. We also define  $V_{i,m}$  which represent the visibility of  $\mathbf{x}_i$  on the  $m$ -th view. Figure 6 shows an example of vertices projection on one of the input shading images.

### 4.1. Albedo Estimation

**Global Albedo** If we assume the surface albedo consists of a single value, we can estimate the surface albedo of vertices globally using the inversion of Equation (1). Given the measured intensity,  $I$ , initial normals  $\mathbf{n}$  and initial depth map  $d$  from projected mesh model and the known lighting direction,  $\mathbf{l}$ , we can obtain:

$$c\rho = \frac{1}{Z} \sum_{i=1}^N \sum_{\substack{m=1, \\ \mathbf{u}_{i,m} \in V_i}}^M \frac{d_{i,m}^2}{\mathbf{n}_{i,m} \cdot \mathbf{l}_{i,m}} I_m(\mathbf{u}_{i,m}) \quad (2)$$

where  $N$  is total number of vertices,  $M$  is total number of shading images, and  $Z$  is a normalization factor. To avoid the effect of cast shadow and specular saturation, we drop the measurements where intensity values are either too

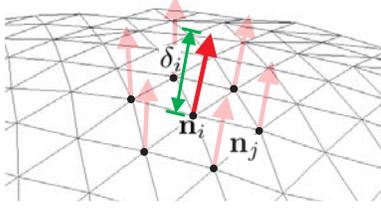


Figure 8. Relationship of variables in mesh optimization in Section 4.2.

small or too large.

**Local Albedo** When the captured object has multiple albedos, we assume surface albedo are locally invariant. Therefore, we can estimate the local albedo by extending Equation (2) as:

$$c\rho_i = \frac{1}{Z_i} \sum_{\mathbf{x}_j \in N(\mathbf{x}_i)} \sum_{\substack{m=1, \\ \mathbf{u}_{i,m} \in V_i}}^M \frac{d_{j,m}^2}{\mathbf{n}_{j,m} \cdot \mathbf{l}_{j,m}} I_m(\mathbf{u}_{j,m}) \quad (3)$$

After the local albedo estimation, we group the mesh vertices using K-means clustering based on the vertex position and local albedo  $(\mathbf{x}_i, c\rho_i)$ . We then re-estimate the albedo within each group globally. This process gives us a more reliable albedo estimation since local albedo estimation does not handle outliers in shading image that violate our shading model. Figure 7 shows an example of our albedo grouping. Note that the grouping is according to shading image in IR image instead of colors in RGB image.

## 4.2. Mesh Optimization

In geometry refinement, we refine the initial mesh model by estimating an optimal displacement of vertex along its normal direction subject to the shading constraint from the Kinect IR images. Our cost function for mesh refinement is defined as follows, which is composed of a data term  $E_p(\delta)$  and a smoothness term  $E_s(\delta)$ :

$$\arg \min_{\delta} E_p(\delta) + E_s(\delta), \quad (4)$$

$$E_p(\delta) = \sum_{i=1}^p \sum_{k \in V_i} w_{i,k} \left( i_{i,k} - c\rho \frac{\mathbf{n}_{i,k}(\delta_{i,k}) \cdot \mathbf{l}_{i,k}}{d_{i,k}^2} \right)^2, \quad (5)$$

$$E_s(\delta) = \sum_{i=1}^p \sum_{j \in N_i} \lambda_1 (\delta_i - \delta_j)^2 + \sum_{i=1}^p \lambda_2 (\delta_i)^2, \quad (6)$$

where  $\delta$  denotes the displacement of vertices that we want to optimize, and  $\mathbf{n}$  is normal direction of a vertex. Figure 8 illustrates the relationship of variables in Equation (4).

The data term  $E_p(\delta)$  in Equation (5) is designed with the near light IR shading model which is covered in Section 3.2. At the beginning of our work, IR camera center is initially estimated in the world coordinate  $X$ . Since we utilize the

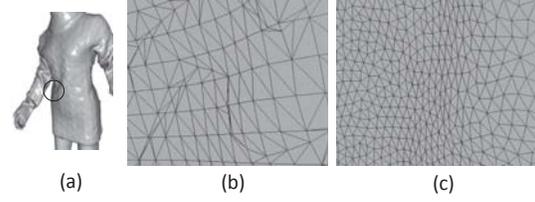


Figure 9. Mesh comparison of before and after remeshing. (a) Region of interest(ROI) of mesh (b) Initial mesh from the Kinect fusion. (c) Our mesh after remeshing. Since the mesh is more clear and dense than that of (b), we can optimize the displacement of vertices to recover fine details effectively.

calibrated IR camera and the light source, the light direction  $\mathbf{l}_{i,k}$  at the each light positions can be estimated using IR camera poses which we obtained from the Kinect fusion. The distance  $d$  between the light source and the vertex position uses depth from the vertex projection as illustrated in Figure 6.  $w_{i,k}$  is a confidence weight expressed by  $\mathbf{n}_{i,k} \cdot \mathbf{l}_{i,k}$  which means that the observation of vertex normal near to light direction is more confidently used for the optimizing. Since the estimated  $d$  is in  $mm$  unit metric depth and has large value compared to the other terms, the optimization is sensitive to the depth  $d$ . Therefore, we begin our optimizing process with the depth-multiplied shading image  $I_c = I * D$  where  $i_i \in I, d_i \in D$  and fix  $d$  as constant at every iteration so that  $d$  is set to be independent of  $\delta$  in this process.

The smoothness term  $E_s(\delta)$  in Equation (6) is composed of two terms. The first term ensures the change of displacement is locally smooth between neighboring vertices,  $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{N}_i$ , and the second term in Equation (6) is to ensure the estimated displacement  $\delta_i$  would not be too large since the initial mesh from the Kinect fusion is already quite accurate. The  $\lambda_1$  and  $\lambda_2$  are determined using the vertex visibility  $V$  and mesh scale.

Comparing our method with the conventional method [5], our method has an advantage to optimize only a single variable  $\delta$  for each vertex, which simplifies the optimizing process and makes our process more stable while method in [5] needs to optimize 3 variables, *i.e.*  $x, y, z$  displacements, for each vertex. By adjusting  $\delta_i$  of the each vertex  $\mathbf{x}_i$ , position of vertex  $\mathbf{x}_i$  are iteratively updated toward minimizing our optimization cost in Equation (4):

$$\mathbf{x}_i^t = \mathbf{x}_i^{t-1} + \delta_{i,t} \mathbf{n}_i^{t-1}. \quad (7)$$

We optimize Equation (4) by utilizing a sparse non-linear least square optimization tool<sup>2</sup>. At the time  $t$ ,  $\delta$  is determined by minimizing the cost in Equation (4) at the previous step  $t-1$ . By doing so, we update  $\delta$  iteratively to obtain a high quality mesh with many fine geometry details.

Since the amount of details we can recover are limited by the mesh complexity. We applied remeshing [19] to

<sup>2</sup>SparseLM: Sparse Levenberg-Marquardt nonlinear least squares <http://users.ics.forth.gr/~lourakis/sparseLM/>

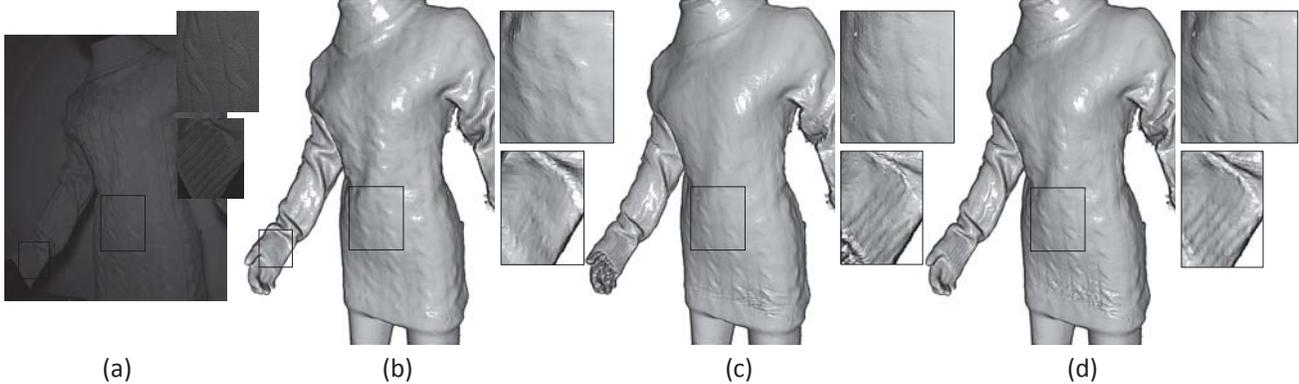


Figure 10. Result comparison of real data - Sweater. (a) IR shading image (b) Initial mesh model (c) Ours: Enhanced mesh result using 4 shading images. (d) Ours: Enhanced mesh result using 17 shading images.

improve initial mesh quality obtained from Kinect fusion as shown in Figure 9. The number of vertices are set to be 200K which does not affect the initial geometry while allowing us to recover fine geometry details that were not reconstructed by the Kinect fusion.

## 5. Experimental Result

We use the Kinect for windows and Kinect SDK 1.7 in our experiments. The implementation of the Kinect fusion is provided in the Kinect SDK. For each real world object, we capture around 20 IR shading images with resolution  $640 \times 480$  for our geometry refinement. Qualitative comparisons between the initial and the refined meshes for several challenging real world dataset are provided in Figure 10 and Figure 11. The example real world objects are: Apollo, Cicero, Sweater, Towel and Vase model. These examples contain fine geometry details which were never been captured in Kinect RAW depth maps and in the mesh model reconstructed by the Kinect fusion. After our geometry refinement, the fine details are presented in our refined mesh model. Mesh models are rendered with the Phong-shaded model.

**Sweater** Sweater is made of wool and it has repetitive twisted pattern on it. It is  $0.8m$  high and  $0.4m$  wide. The measured depth variation of the twisted pattern is  $1mm$ . Figure 10 shows IR shading image, initial mesh and our results. In Figure 10 (b), the geometry from the Kinect fusion does not fully express the twisted pattern on the sweater. On the other hand, our result (c) and (d) recovers detailed pattern. By comparing (c) and (d), we observed that how the number of shading image used affect the result in our framework. Those two are the enhanced mesh result using 4 shading images and 17 shading images respectively. As we can see in enlarged box, the twisted pattern and the sleeve show more precise details in the result of using 17 images. We observed that more shading observations makes our mesh refinement more stable and result in better geometry.

**Apollo** We captured the statue of Apollo which is made of plaster and has fine details on its geometry especially face and hair region. The size of Apollo is  $0.75m \times 0.65m$ . In Figure 1 and Figure 11, the initial mesh from the Kinect fusion and enhanced mesh from our method are compared. The IR shading image shows that Apollo has double eyelid on its eye but it is not expressed in the mesh from the Kinect fusion. Apollo has also the fine detail of hair but is not appeared in the input mesh. This occurs from the limited voxel resolution of the Kinect fusion. Our refinement on the initial mesh shows enhanced double eyelid and hair geometry. We used 24 IR shading images for the result.

**Cicero** Statue of Cicero (size of  $0.7m \times 0.45m$ ) is also used to verify our algorithm. The back-head of the Cicero exhibits very fine level of detail and it is not shown in initial mesh at all. In our result, the fine hair is recovered and it becomes to have more realistic geometry. 22 IR shading images are used here.

**Towel** We also verified that our method works well on small object with the subtle details. A towel, size of  $0.2m \times 0.2m$ , were used for our experiment. As shown in Figure 11, result of towel, initial mesh loses its fine check-pattern and it shows flat surface geometry. However, our method effectively recovers the check-pattern in detail and the surface of our result mesh becomes rather similar to the geometry of real object.

**Vase** We also tested our algorithm with multi-albedo object. The target object is plant with a pot, denoted here as a vase and it has size of  $1.2m \times 0.3m$ . We grouped the albedo as described in Section 4.1. As shown in Figure 7, plant leaves and a pot have different observation in terms of surface albedo in the IR image. We observed that the plant leaves have smooth geometry and there was less room for refine geometric details. On the other hand, the pot has complex geometry. We apply our method on the initial mesh from the Kinect fusion. In this case, albedo grouping in Sec. 4.1 are applied prior to mesh optimization. The cross stripes on the pot is recovered using our method.

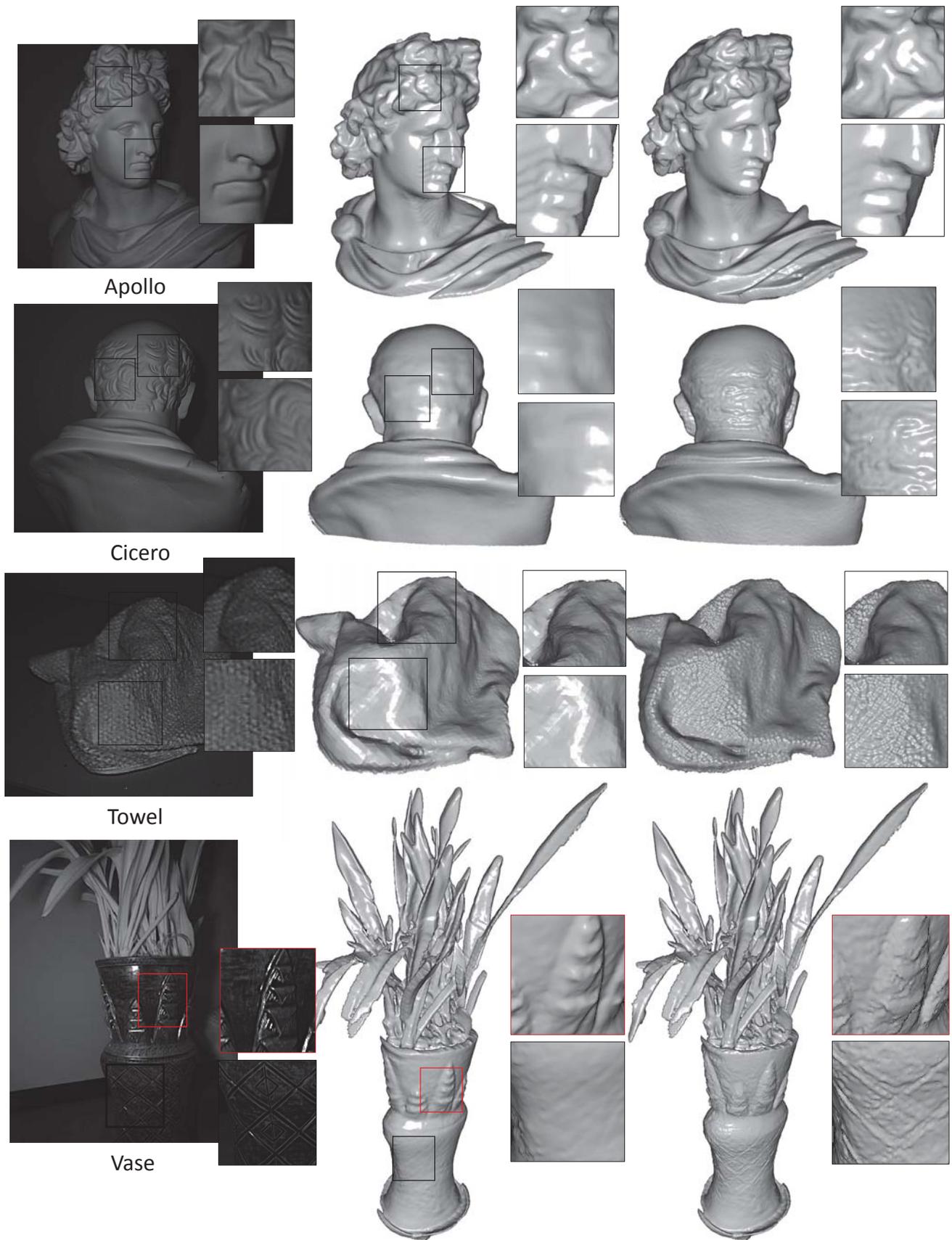


Figure 11. Enhanced mesh result of real world objects - Apollo, Cicero, Towel and Vase. From the left, each column represents IR shading images, initial mesh from the Kinect fusion and our mesh result respectively.

However, the region which is marked with the red box is not refined well. In this region, specular exists and it does not follow the Lambertian shading model in Equation (1).

## 6. Conclusion

In this paper, we have presented a framework to utilize shading information from the Kinect IR image for geometry refinement. As far as we are aware, this is the first work that seriously studies the shading information inherent in the Kinect IR image and utilizes it for geometry refinement. As demonstrated in our study, the captured spectrum of Kinect IR images does not have any overlapping with visible spectrum which makes our acquisition unaffected by indoor illumination condition. Since there is almost no ambient light in IR spectrum, the captured intensity can be accurately modeled by our near light IR shading model assuming the captured materials follow the Lambertian BRDF. We have also described a method to radiometrically calibrate the Kinect IR image using a diffuse sphere, a method to estimate geometry albedo and albedo grouping, and a new mesh optimization method to refine geometry by estimating a displacement vector along vertex normal direction. Our experimental results show that our framework is effective and demonstrates high quality mesh model via our geometry refinements. As a limitation of our work, we assume Lambertian BRDF which makes our results error-prone to specular highlight. Due to the usage of the Kinect fusion, we also assume the reconstructed object is static. In future, we will study how to extend our work to handle non-Lambertian BRDF objects, and mesh refinement for dynamic object reconstructions.

## 7. Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2010-0028680). Yu-Wing Tai was supported by the Microsoft Research Asia.

## References

- [1] M. Bohme, M. Haker, T. Martinetz, and E. Barth. Shading constraint improves accuracy of time-of-flight measurements. *CVIU*, 114(12):1329–1335, 2010.
- [2] J. Dolson, J. Baek, C. Plagemann, and S. Thrun. Upsampling range data in dynamic environments. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [3] M. Grossberg and S. Nayar. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(10), 2004.
- [4] Y. Han, J.-Y. Lee, and I. Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. In *Proc. of Int'l Conf. on Computer Vision (ICCV)*, 2013.
- [5] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):548–554, 2008.
- [6] B. Horn, R. Woodham, and W. Silver. Determining shape and reflectance using multiple images. In *MIT AI Memo*, 1978.
- [7] B. K. P. Horn and M. J. Brooks. *Shape from shading*. MIT Press, Cambridge, MA, USA, 1989.
- [8] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *ACM Symposium on User Interface Software and Technology*, 2011.
- [9] H. P. A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H. Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. on Graph.*, 22:234–257, 2003.
- [10] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 193:133–135, 1981.
- [11] Z. Lu, Y.-W. Tai, M. Ben-Ezra, and M. S. Brown. A framework for ultra high resolution 3d imaging. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [12] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, 24(3):536–543, 2005.
- [13] T. Okatani and K. Deguchi. Optimal integration of photometric and geometric surface measurements using inaccurate reflectance/illumination knowledge. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [14] J. Park, H. Kim, Y.-W. Tai, M. Brown, and I. Kweon. High quality depth map upsampling for 3d-tof cameras. In *Proc. of Int'l Conf. on Computer Vision (ICCV)*, 2011.
- [15] J. Park, S. Sinha, Y. Matsushita, Y.-W. Tai, and I. Kweon. Multiview photometric stereo using planar mesh parameterization. In *Proc. of Int'l Conf. on Computer Vision (ICCV)*, 2013.
- [16] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2000.
- [17] J. Shen and S. ching S. Cheung. Layer depth denoising and completion for structured-light rgb-d cameras. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [18] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from a single depth image. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [19] V. Surazhsky and C. Gotsman. Explicit surface remeshing. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*.
- [20] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popovic, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. on Graph.*, 28(5), 2009.
- [21] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 969–976, 2011.
- [22] Q. Yang, R. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [23] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of rgb-d images. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [24] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. Chan. Outdoor photometric stereo. In *IEEE Conference on Computational Photography (ICCP)*, 2013.
- [25] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2012.