

A Framework for Ultra High Resolution 3D Imaging

Zheng Lu^{1*}

Yu-Wing Tai²

Moshe Ben-Ezra³

Michael S. Brown¹

¹National University of Singapore. ²Korean Advanced Institute of Science and Technology.
³Microsoft Research Asia.

Abstract

We present an imaging framework to acquire 3D surface scans at ultra high-resolutions (exceeding 600 samples per mm^2). Our approach couples a standard structured-light setup and photometric stereo using a large-format ultra-high-resolution camera. While previous approaches have employed similar hybrid imaging systems to fuse positional data with surface normals, what is unique to our approach is the significant asymmetry in the resolution between the low-resolution geometry and the ultra-high-resolution surface normals. To deal with these resolution differences, we propose a multi-resolution surface reconstruction scheme that propagates the low-resolution geometric constraints through the different frequency bands while gradually fusing in the high-resolution photometric stereo data. In addition, to deal with the ultra-high-resolution images, our surface reconstruction is performed in a patch-wise fashion and additional boundary constraints are used to ensure patch coherence. Based on this multi-resolution reconstruction scheme, our imaging framework can produce 3D scans that show exceptionally detailed 3D surfaces far exceeding existing technologies.

1. Introduction

High resolution 3D imaging is useful in many applications, from engineering analysis, to computer graphics, to the preservation and study of cultural heritage materials. The goal of this work is to design an ultra-high-resolution 3D imaging framework that targets surface sampling at more than 600 samples per mm^2 . To our best knowledge, this is the highest sampling rate demonstrated to date.

The fundamental problem for 3D imaging at these ultra-high-resolutions is that while it is possible to obtain very dense samples of the surface using a high-resolution digital sensor, it is difficult to perform structured-light at these resolutions. Virtually all projector lenses are designed to

*This work was done while Zheng Lu was a visiting student at Microsoft Research Asia.

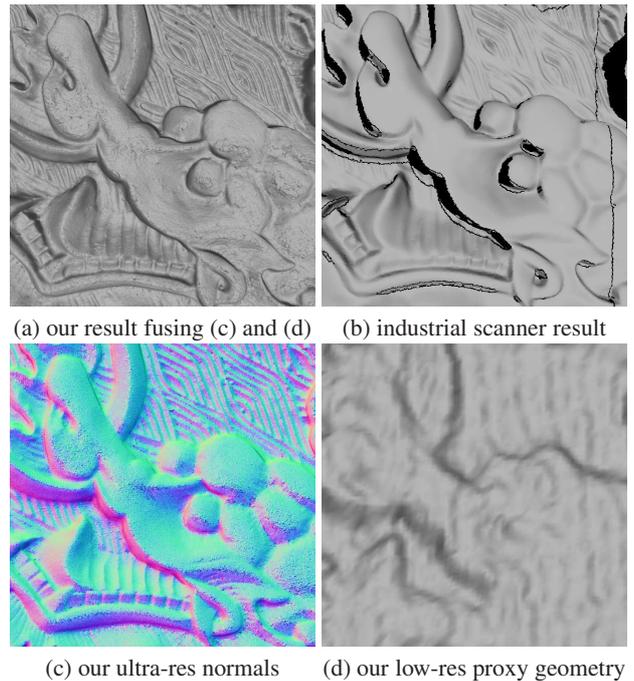


Figure 1. Comparison of our results against a state-of-the-art industrial 3D scanner: (a) our reconstructed surface obtained by combining (c) and (d); (b) surface reconstructed by a Konica Minolta Range 7 (168 samples per mm^2); (c) our input from photometric stereo of (over 600 samples per mm^2); (d) our low-resolution geometry (6.25 samples per mm^2) obtained via structured-light.

magnify the projected screen and therefore cannot produce very dense samples on an object’s surface. While lasers can be focused on an object at much finer resolution, lasers used in 3D scanners rarely can produce an illuminated point less than 10 microns at very close ranges (i.e. a few centimeters away from the object) to only several hundred microns at ranges 50 – 100cm away. Thus, high-resolution laser scanning at these resolutions is limited to small areas that must be stitched together.

To appreciate this difference, Figure 1 shows a 3D surface scanned using ultra-resolution imaging and the same

object scanned by a high-end commercial 3D scanner designed for industrial inspection (Konica Minolta Range 7 [1]). Our 3D surface is scanned at more $3\times$ the reported surface resolution of the commercial scanner. The commercial scanner also had to acquire multiple scans that are stitched together (using Konica’s software) because the object is bigger than its effective area.

To overcome the limitations of structured-light, hybrid imaging approaches that combine positional data from structured-light together with fine surface normals captured from photometric stereo have been proposed (e.g. [5, 13, 19]). We adopt this hybrid imaging approach, however unique to our framework is the significant asymmetry in the resolutions between the two inputs. Existing work dealt with sample differences of $4\times$ resolution between the surface normals and low-resolution geometry ([5]), while our photometric surface detail is scanned at around $100\times$ the resolution of the low-resolution geometry (see Figure 1(a) and (b)). This ultra-high-resolution and significant resolution difference in the hybrid framework creates difficulties not yet encountered by previous approaches.

To address this significant difference in the resolution between the surface normals and low-resolution geometry, we propose a multi-resolution surface reconstruction scheme that fuses the low-resolution geometric with the photometric stereo data at increasing levels of details. To deal with the large data from the ultra-high-resolution input, we adopt a patch-based scheme that uses additional boundary constraints to maintain patch coherence at the boundaries. The results of our approach are 3D images of surface captured at an exceptionally high level of detail.

The remainder of this paper is organized as follows: Section 2 discusses related works; Section 3 describes our system setup to capture ultra high resolution images for normal estimation, and our structured-light system to capture surface geometry; Section 4 presents our main algorithm for surface reconstruction; Section 5 presents our results. A summary of this work is presented in Section 6.

2. Related Work

There is a vast amount of literature on 3D imaging. Readers are directed to [17] and [20] for broad overviews; here only representative examples are mentioned.

3D imaging has been approached using passive triangulation methods such as conventional stereo (e.g. [15]), passive photometric methods such as shape from shading (e.g. [10]), active triangulation methods such as structured-light (e.g. [16]) and active photometric methods such as photometric stereo (e.g. [21]). Hybrid methods that integrate two or more methods include approaches that combine shape from motion and photometric stereo (e.g. [9]), positional (3D points) data and normals (e.g. [18, 3, 7, 12, 11, 5,

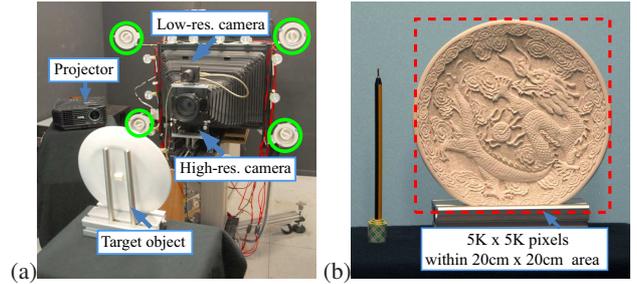


Figure 2. Our 3D imaging setup: (a) Our experimental setup consists of an ultra-high-resolution camera with four lights used photometric stereo. A low-resolution video camera and digital light projector form the structured-light system. (b) shows the effective resolution about one of our objects. Note the scale of the physical object, versus the pixel resolution. This results in a pixel resolution of over 600 samples per mm^2 .

6, 13]), visual hull and normals [8]¹, and recently normals and volume carving [19].

All of the previously mentioned hybrid methods have the potential to be adapted to handle very high-resolution imagery as in our application. We decided on a solution that is closely related to that presented by Nehab et al. [13]. The system presented in [13] used two cameras in a structured-light setup with one of the cameras also used for the photometric stereo. Positional and surface normals were fused using a linear formulation that resulted in a sparse-linear system. In their work, the surface geometry and the photometric data were of the nearly the same resolution. In our work, we have $100\times$ more estimated surface normals than we do 3D points. At our resolutions (e.g. $5K\times 5K$ surface normals), the sparse matrix proposed [13] would have approximately 150 million entries. Solving such a large matrix is not straight forward, even for out-of-core linear solvers such as [14]. As such, we adopt a patch-wise strategy to the fusion process. In addition, to deal with the significant difference in resolutions, we use a multi-resolution pyramid approach to adaptively incorporate the geometric constraint from the low-resolution geometry during the surface integration.

3. System setup

Our hybrid system uses an ultra-high-resolution camera together with four controllable lights. This is combined with a separate structured-light rig composed of a low-res camera and projector. Figure 2(a) shows our setup. The four light sources are calibrated using a mirrored sphere. We placed polarized filters in front of the lights, together with a polarized filter on the ultra-high-resolution camera to reduce light scattering. The two cameras and projectors are calibrated by a physical calibration pattern.

¹Also see <http://carlos-hernandez.org/gallery/>.

The following gives more details to the two main components.

Ultra-high-resolution photometric stereo

High-density photometric stereo images can be captured using a consumer level 35 mm SLR camera equipped with a macro lens. Due to the zoom factor of the lens, this approach will only be able to scan very small surface patches. While it is possible to integrate several surface patches together, we have opted to use a large-format camera. Commercial large-format digital cameras up to 340Megapixel exist on the market [2], however, in this work we use a custom-built 1.6gigapixel camera that uses a translation scanning back with an effective format of 450×300 mm. The details to this camera are outside the scope of this paper and readers are referred to [4] for more information.

To estimate normals, we use the ultra-high-resolution camera to capture four images of the surface with varying illumination from the four light sources. The photometric stereo technique in [21] is used to obtain the surface normals. Figure 2(b) shows the scanning setup for the *dragon plate* used as a running example in this paper. The plate has a diameter of 20cm. These images of the object are at a resolution of $5K \times 5K$ pixels.

Structured-Light geometry Our structured-light system consists of a Benq MP624 projector and a 1024×768 video camera. Standard binary gray-code patterns [16] is used to estimate the low-resolution geometry. In our setup (see Figure 2(b)), we get approximately 6 samples per mm^2 from the structured light scanner, this resolution varies based on the projectors distance to the object but is representative of a typical structured-light system. Figure 1(d) shows a small example of the 3D surface geometry estimated using our method. There are slight pixelization-like artifacts due to inaccuracies in estimating the projected patterns' boundaries, however, since the low-resolution geometry serves only as a soft constraint in the surface reconstruction process our approach is insensitive to these errors. Because our ultra-high-resolution scanning back camera requires roughly a minute to capture an image, we opted to use an auxiliary video camera to perform the structured-light procedure instead of the large-format camera itself.

4. Surface Reconstruction Algorithm

This section describes the surface reconstruction algorithm. The basic algorithm to reconstruct a surface from normals is described first. This is followed by a description on how to include the low-resolution geometry constraint and boundary connectivity constraint into the algorithm. Finally, the steps of the multi-resolution strategy is detailed.

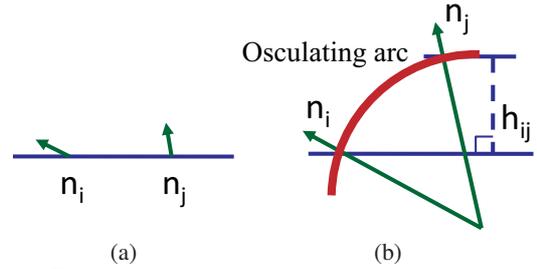


Figure 3. The osculating arc constraint [22] for surface reconstruction. Given the normal configuration $\{n_i, n_j\}$ between neighborhood pixel i and j in (a), we can uniquely defined the relative height h_{ij} in (b) by using an osculating arc to connect n_i and n_j with minimum curvature.

4.1. Surface from normals

Given a dense set of normals the goal is to reconstruct a surface that satisfies the normals' orientation constraints. We use the recent approach presented by Wu et al. [22] for obtaining a surface from normals that constrains the estimated surface using an osculating arc between neighboring normals (see Figure 3). This problem can be cast as a least-square problem that minimizes the following energy function:

$$E(S|\vec{n}) = \sum_i^N \sum_{j \in \mathcal{N}(i)} ((S_i - S_j) - h_{ij})^2 \quad (1)$$

where S is the surface we want to reconstruct, $(S_i - S_j)$ is the first derivative of S in discrete form, h_{ij} is the relative height defined by the osculating arc constraint between neighborhood pixels, $\mathcal{N}(i)$ is the first order neighborhood of a pixel, and N is number of pixels. A qualitative comparison of the osculating arc constraint with other surface from normals algorithms can be found in [22].

Equation (1) can be solved using Gauss-Seidel iteration. At each iteration, the surface height is updated according to the following equations:

$$\begin{aligned} S_i^{t+1} &= S_i^t + \lambda_1 \xi_1, \\ \xi_1 &= \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} (h_{ij} - (S_i^t - S_j^t)) \end{aligned} \quad (2)$$

where $|\mathcal{N}(i)|$ is the number of neighborhood pixels, $\lambda_1 = 0.9$ is the step size and t is iteration index. Note that h_{ij} is the same for all iterations and can be pre-computed.

4.2. Low-resolution geometry constraint

Because photometric stereo inherently captures only local reflection information rather than global structure, many surface from normal reconstruction approaches do not accurately reflect the real surface geometry. As discussed in section 2, one strategy to overcome this is to incorporate positional information in the reconstruction process.

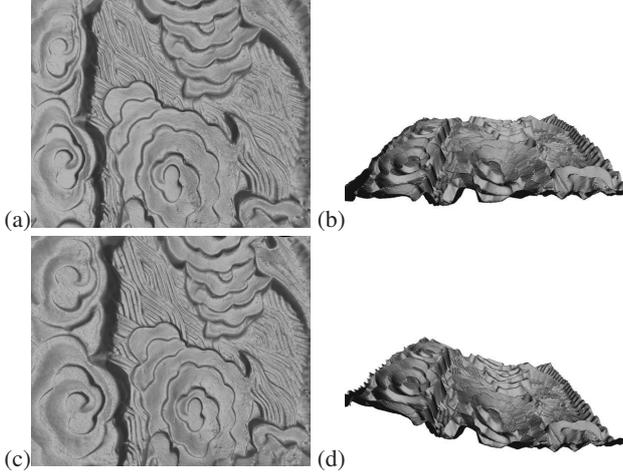


Figure 4. Example of surface reconstruction without including the geometric constraints: (a) reconstructed surface from normals only; (b) a side view of (a); (c) reconstructed surface with low-resolution geometry constraints; (d) a side view of (c).

Our low-resolution geometry constraint is modeled using the following equation:

$$E(S|L) = \sum_i^M (|d(h(S_i)) - L_i| - \Delta)^2 \quad (3)$$

where L is the low-resolution geometry captured by the structured-light setup, M is the number of pixels in low-resolution geometry proxy, $h(\cdot)$ is a Gaussian convolution process with radius equal to two times the downsample rate, $d(\cdot)$ is a downsample operation to match the high-resolution normals to the low-resolution geometry, and $|\cdot|$ is the L1 norm (absolute value) of the errors. The term Δ is a parameter controlling the amount of depth tolerance for surface details to be reconstructed and refined by the normals.

With the additional low-resolution geometry constraint, the iterative update equation in Equation (2) can be updated as follow:

$$S_i^{t+1} = S_i^t + \lambda_1 \xi_1 + \lambda_2 \xi_2, \quad \xi_2 = \begin{cases} h(u(L_i - d(h(S_i)))) & \text{if } |d(h(S_i)) - L_i| > \Delta \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $u(\cdot)$ is an upsample operator. The effect of our low-resolution geometry constraint is shown in Figure 4. The value of Δ can be estimated according to the variance of surface details reconstructed from normals and can be spatially varying.

4.3. Boundary connectivity constraint

As discussed in Section 2, the ultra-high-resolution of the photometric stereo component, makes it challenging to perform integration on the entire surface in one pass. To

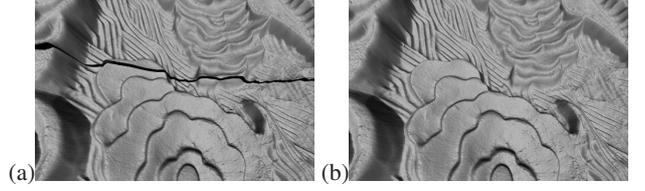


Figure 5. Effect of the boundary connectivity constraints: (a) without the boundary constraint, and (b) with boundary connectivity constraint.

overcome this the surface can be subdivided into more manageable sized patches and each patch reconstructed individually. This leads to a problem that the boundaries of adjacent patches may not be properly aligned after reconstruction. To overcome this, we add a boundary connectivity constraint described by the following equation:

$$E(S|B) = \sum_{i \in \Omega} (S_i - B_i)^2 \quad (5)$$

where Ω is the overlapping area of neighborhood surface patch, B is a surface computed by blending the intermediate reconstructed surface in Ω between neighborhood patches using linear feathering. Adding the boundary constraint into Equation (2), we get:

$$S_i^{t+1} = S_i^t + \lambda_1 \xi_1 + \lambda_2 \xi_2 + \lambda_3 \xi_3, \quad \xi_3 = \begin{cases} B_i - S_i & \text{if } i \in \Omega \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

For this boundary connectivity constraint, the weight of λ_3 during the update iterations needs to be adjusted as the system is iterated. In the initial estimation, λ_3 is equal to zero, and its weight is gradually increased as the number of iterations increases. This allows the surface patch to be reconstructed freely at initial iterations and later refined to meet the boundary of neighborhood patches. In our implementation, λ_3 and B are updated at every 100 iterations. With this boundary connectivity constraint, surface reconstructed can be done in parallel and the problem of resolution is no longer an issue. The effect of this boundary constraint is shown in Figure 5. For the results in this paper, surface patches are taken to be of size 1024×1024 with overlaps of 100 pixel boundary overlap (i.e. 10%).

4.4. Multi-resolution pyramid approach

Due to the very large differences in resolutions between the surface normals and the low-resolution geometry, directly adding the surface normals to the low-resolution geometry will result in noisy reconstruction as shown in Figure 7. To avoid this, our surface reconstruction is done in a multi-resolution pyramid fashion. The main purpose of using the pyramid approach is to correct the low-resolution geometry using normals at the equivalent before we use it as soft constraint at a higher resolution. The multi-resolution

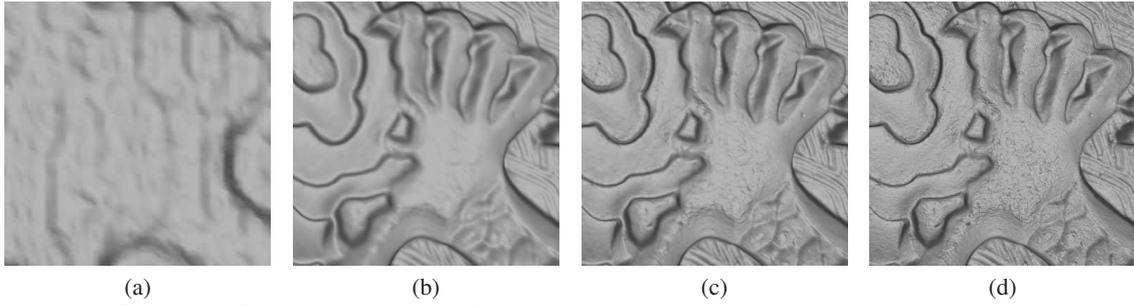


Figure 6. Evolution of our 3D surface up the multi-resolution pyramid: (a) low-resolution geometry; (b) intermediate result at the lowest level of the pyramid; (c) the third level; (d) the last level and final 3D reconstruction.

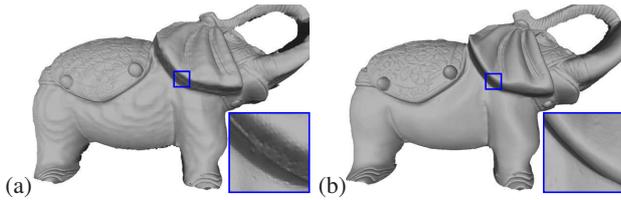


Figure 7. Example of the benefits of the multi-resolution scheme: (a) our surface reconstruction directly using the low-resolution geometry; (b) reconstructed surface using the multi-resolution scheme. The 3D surface in (a) shows noticeable quantization errors due to the low-resolution geometry.

pyramid approach also allows us to resolve small misalignments between the high-resolution normals and low-resolution geometry due to device calibration errors.

We divide the pyramid uniformly into 5 different levels starting at the resolution used to capture the low-resolution geometry (i.e. 1024×768). For each level, instead of down-sampling the estimated high-resolution normals, we down-sampled the ultra-high-resolution input images and estimate the normals from the downsampled images. This down-sampling helps to reduce some of the camera’s sensor noise when estimating the normals. We run our surface reconstruction algorithm described in Equation (6) with the results from previous level as the low resolution constraint. For the lowest resolution, the low-resolution geometry estimated by structured-light is used. Figure 6 shows our intermediate surface reconstruction results (i.e. the evolution) at different levels in the pyramid.

5. Results

This section shows several results captured by our system. Each surface was generated using 800 iterations (per patch) of our surface reconstruction algorithm with the boundary constraint applied once after every 100 steps. Because we are working with resolutions beyond the capabilities of existing devices, providing quantitative comparisons proved very challenging. As a result, we can only show visual results for some of our results. For the *dragon plate* we took the object to an industrial scanning facility using a Konica Minolta Range 7. This serves as our baseline com-

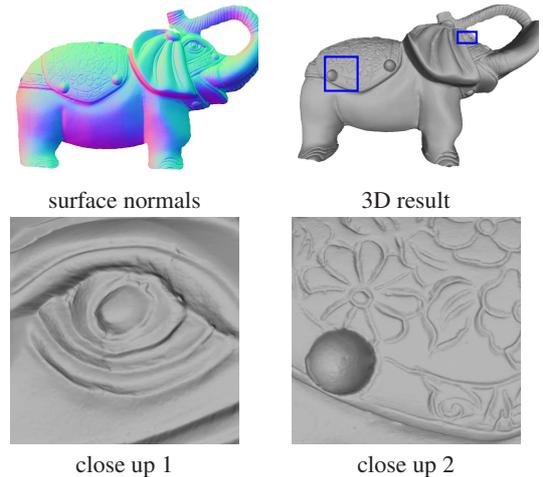


Figure 8. 3D reconstruction of the *elephant* figurine. The zooms show exceptional detail on the surface of the object.

parison against a state-of-the-art industrial laser scanner.

Figure 8 shows a 3D reconstruction of an *elephant* figurine which is approximately 15cm wide. Figure 9 shows example of an *man* figurine of a man roughly 12cm high. The objects required 9 patches and resulted in about 6.5 million reconstruct 3D points, while the man required 12 patches and resulted in about 4.5 million reconstructed 3D points. Both of these results show exceptional surface detail. The *man* figurine is further zoomed to reveal detail that would require a magnifying glass to be seen (properly) with the unaided eye.

Finally, we compare our result a scanned *dragon plate* with that obtained from an industrial standard high end laser scanner (Konica Minolta Range 7) in Figure 10. The finest scanning resolution that can be obtained by the laser scanner is 168 samples per mm^2 , while our sampling rate is 600 samples per mm^2 . The plate required 25 patches and resulted in about 21.5 million reconstructed 3D points. The state-of-the-art scanner reports to have a scanning accuracy of 40 microns. We can see that on the double-zoom of these two surfaces, we reveal detail while the result from the laser scanner is almost completely flat. Note that in order to capture the whole plate by the laser scanner, several scans were performed and stitched together. Our approach, on the other

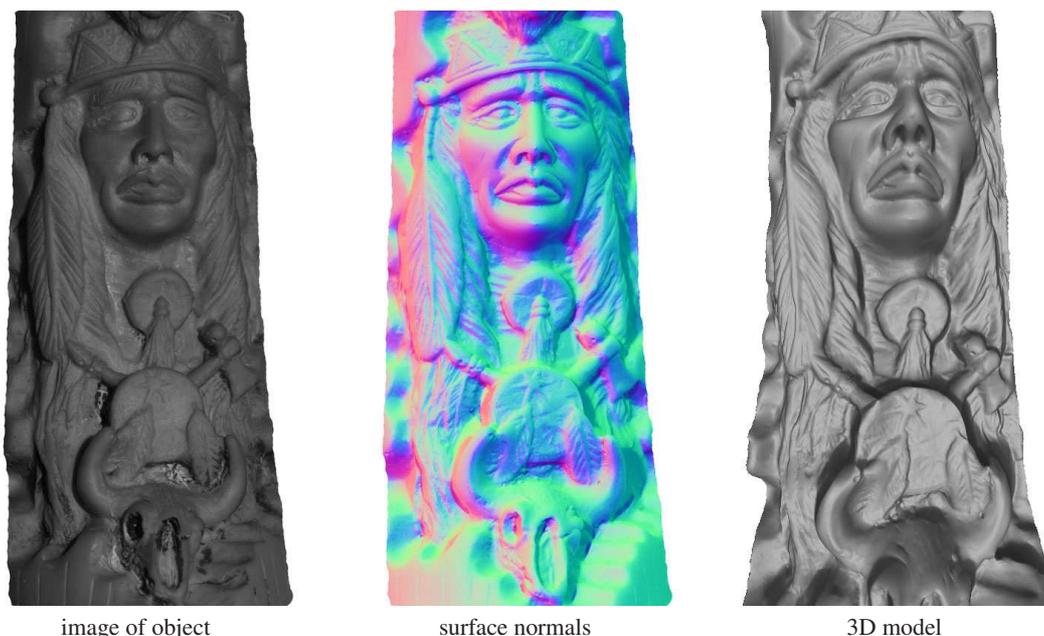
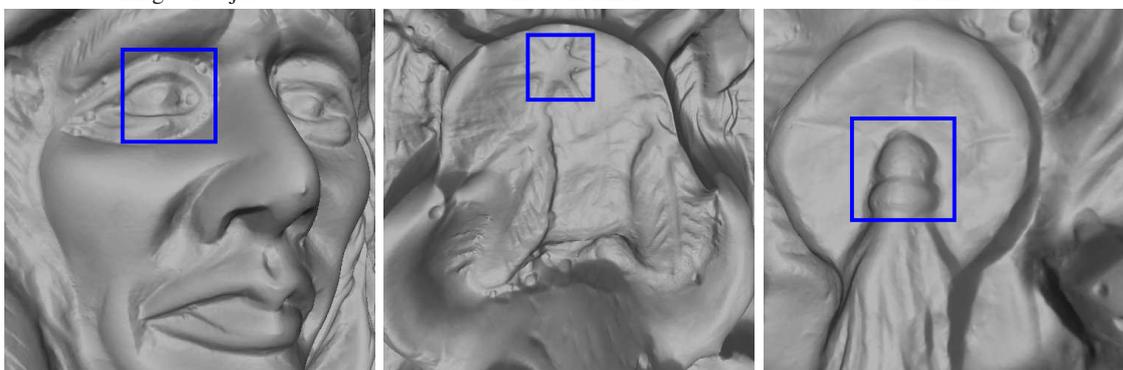


image of object

surface normals

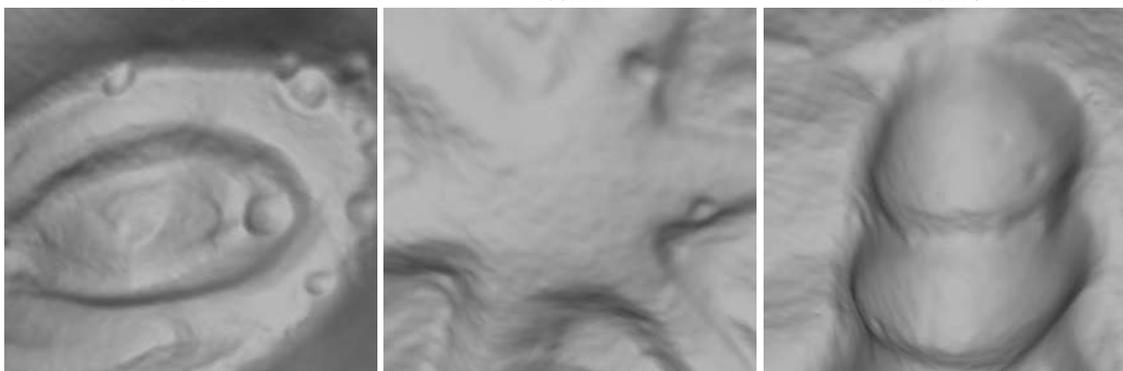
3D model



zoom 1

zoom 2

zoom 3



double zoom 1

double zoom 2

double zoom 3

Figure 9. 3D reconstruction of the *man* figurine. Due to the ultra-resolution of the 3D scan, we can show a zoom and “double zoom” of the 3D surface. This double zoom reveals detail that would require a magnifying glass to see properly.

hand, was able to image the entire 3D object in one pass. In addition, even though we use a patch-wise approach in surface integration, our surface does not contain any blocking or pixelation artifacts. This helps to demonstrate the

effectiveness of our boundary connectivity constraint and multi-resolution pyramid approach. Comparing the surface depth, our estimated surface depths are consistent with surface depths captured by laser scanner.

6. Conclusion

This paper presented a framework to capture ultra-high resolution surfaces using a hybrid system that consists of an ultra-high-resolution photometric stereo system and a structured-light system. Currently, this strategy appears to be one of the only ways to surpass the resolution of structured-light systems. While this imaging framework can be construed as a feat of engineering, there are real concerns that had to be addressed when dealing with such massive amounts of data and with significant resolution asymmetry in the respective subcomponents. To address these issues, a multi-resolution pyramid approach was introduced to reconstruct the high resolution surface progressively and adaptively. We have also discussed how to reconstruct surface in a patchwise fashion and seamlessly stitched the reconstructed surface patches together.

We do can envision that other strategies could be used to produce a similar 3D imaging system. Moreover, as this work is not focused on improving photometric stereo, we inherit all the issues known to affect normal estimations (e.g Lambertian surface assumption, albedo estimation, use of light sources and their calibration, etc). In addition, while combining positional information with estimated normals is a common approach to reduce errors in the surface reconstruction, it may be possible to devise algorithms that can perform surface reconstruction directly from the normals without proxy geometry. To help others address these issues, the dataset used in this paper (input images, estimated normals, and low-res geometry) are publicly available for download and experimentation.

7. Acknowledgement

This work was funded in part by KAIST seed grant (No. G04090064) and NUS RFC (No. R-252-000-423-112). We gratefully acknowledge Judy Brown, Yasuyuki Matsushita, Yi Ma, and the anonymous reviewers for their useful suggestions towards improving this paper.

References

- [1] Konica Minolta Range 7 Laser Scanner Specifications, <http://www.konicaminolta.com>.
- [2] Anagramm and Digital Reproduction, <http://www.linhofstudio.com>.
- [3] S. Banerjee, P. Sastry, and Y. Venkatesh. Surface reconstruction from disparate shading: An integration of shape-from-shading and stereopsis. *IAPR*, 1:141–144, 1992.
- [4] M. Ben-Ezra. High Resolution Large Format Tile-Scan Camera; Design, Calibration, and Extended Depth of Field. In *IEEE International Conference on Computational Photography (ICCP)*, 2010.
- [5] F. Bernardini, H. Rushmeier, I. M. Martin, J. Mittleman, and G. Taubin. Building a digital model of michelangelo’s florentine pieta. *IEEE Computer Graphics and Applications*, 22(1):59–67, Jan.
- [6] C. Y. Chen, R. Klette, and C. F. Chen. Shape from photometric stereo and contours. In *CAIP*, pages 377–384, 2003.
- [7] P. Fua and Y. G. Leclerc. Using 3-dimensional meshes to combine image-based and geometry-based constraints. In *ECCV*, 1994.
- [8] C. Hernández, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE TPAMI*, 30(3):548–554, Mar 2008.
- [9] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3d modeling. In *CVPR*, 2009.
- [10] B. Horn and M. Brooks. *Shape from Shading*. MIT Press, 1989.
- [11] K. Ikeuchi. Determining a depth map using a dual photometric stereo. *International Journal of Robotics Research*, 6(1):15–31, 1987.
- [12] H. Lange. Advances in the cooperation of shape from shading and stereo vision. In *3-D Digital Imaging and Modeling*, 1999.
- [13] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM ToG (SIGGRAPH)*, 24(3):536–543, 2005.
- [14] J. K. Reid and J. A. Scott. An out-of-core sparse cholesky solver. *ACM Transactions on Mathematical Software*, 36(2):1–33, 2009.
- [15] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IEEE TPAMI*, 47:7–42, 2002.
- [16] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, 2003.
- [17] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006.
- [18] D. Terzopoulos. The computation of visible-surface representations. *IEEE TPAMI*, 10(4):417–438, 1988.
- [19] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM ToG (SIGGRAPH-Asia)*, 28(5), Dec. 2009.
- [20] C. Whöler. *3D Computer Vision: Efficient Methods and Applications*. Springer, 2009.

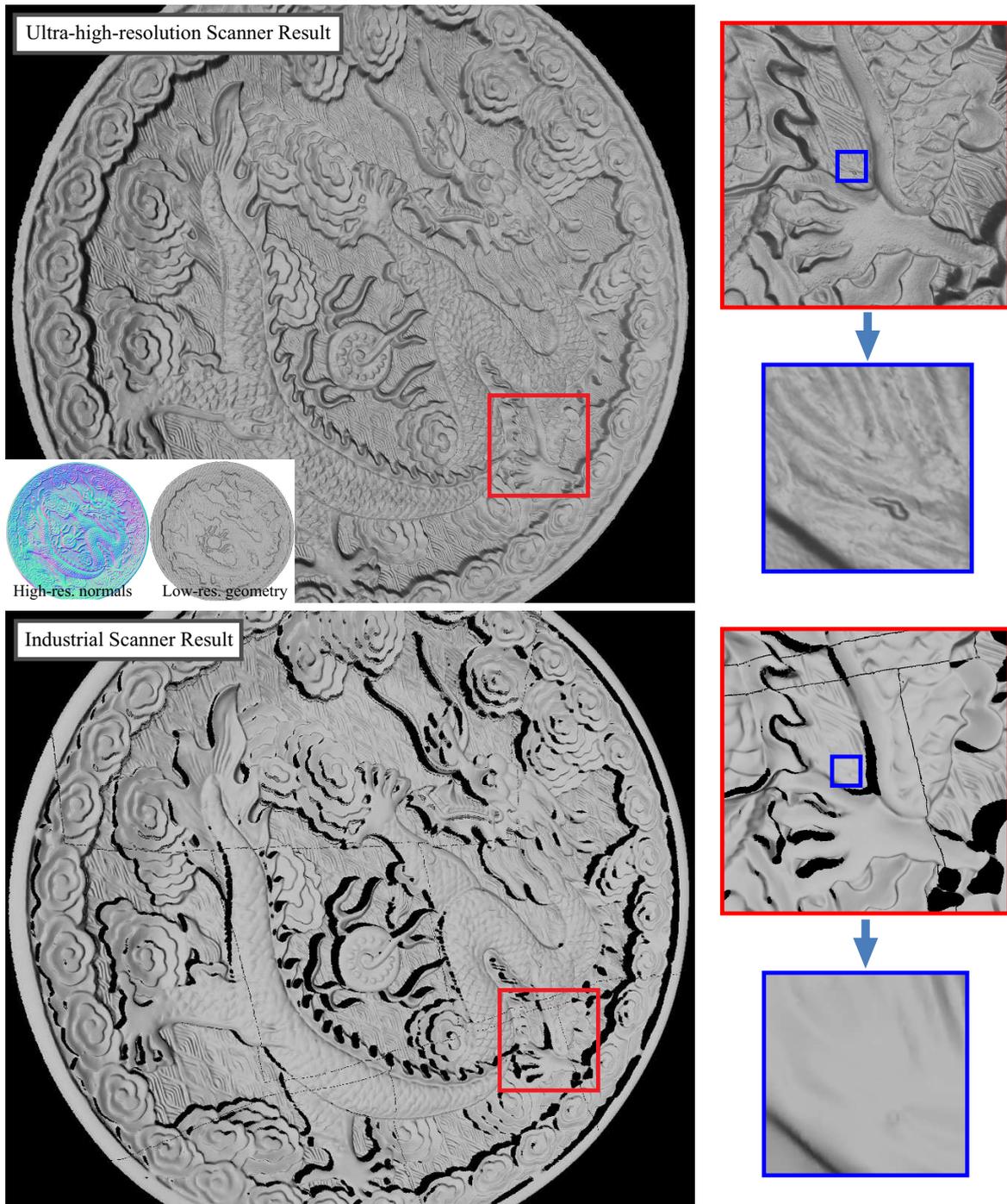


Figure 10. Full-size comparison with an industrial laser scanner. Shown are the full 3D reconstruction from our approach and that from a Konica Minolta Range 7 industrial scanner. Insets for our approach show the surface normals and low-resolution geometry. Zoomed and double zoomed regions show that while the two scans reveal that our results contains considerable more surface detail, while both appear to reflect the correct geometry. Note that the Konica Minolta Range 7 specifications states a scanning accuracy of up to for ± 40 microns.

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

[22] T. Wu, J. Sun, C. Tang, and H. Shum. Interactive normal reconstruction from a single image. *ACM ToG (SIGGRAPH-Asia)*, 27(5), 2008.