A Convex Framework for High Resolution 3D Reconstruction

Min Li, Changyu Diao, Song Lv and Dongming Lu
College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China

Keywords: Structured-light, Photometric Stereo, Convex Optimization, 3D Reconstruction.

Abstract: We present a convex framework to acquire high resolution surfaces. It is typical to couple a structure-light setup and a photometric method to reconstruct a high resolution 3D surface. Previous methods often get stuck in a local minima for the appearance of occasional outliers. To address this issue, we develop a convex variational model by incorporating a total variation (TV) regularization term with a data term to generate the surface. Through relaxing the model to an equivalent high dimensional variational problem, we obtain a global minimizer of the proposed problem. Results on both synthetic and real-world data show an excellent performance by utilizing our convex variational model.

1 INTRODUCTION

The highly detailed reconstruction of 3D shape has been one of the classic topics in computer vision, from computer graphics, to reverse rendering, and to the digital preservation of cultural heritage materials. It is a challenging task especially when we have to reconstruct millions of 3D points. In our paper, we intend to reconstruct high resolution detailed surfaces via a convex framework, which make us avoid getting stuck in local minima and obtain a high quality surface.

The fundamental difficulty of highly detailed surface reconstruction is that it is impossible to acquire dense and accurate samples of a surface via only one method. While methods of laser scanners or structured light often obtains accurate surfaces, it is difficult to perform highly detailed reconstructions due to limitations of hardware. In contrast, surface reconstruction from gradient fields is capable of obtaining detailed surfaces, however, there are still some problems with them. Typically, it is possible to capture high-resolution images via methods such as photometric stereo or shape from shading, while the resulting gradient filed of the above methods is usually non-integrable due to gradient manipulations, presence of noises or outliers in gradient estimation. Consequently, it is difficult to reconstruct an accurate surface only through photometric stereo or shape from shading on account of the lack of global information.

To overcome limitations of structured light and photometric stereo, hybrid methods have been proposed for a decade. Similar to (Lu et al., 2010), We adopt an approach which combines structured light and normal information estimated by photometric stereo. Rather than employing the classical ℓ-2 methods such as the least square method, we present a new convex framework to overcome the problem of ℓ-2 methods and maintain high-quality surface details, through which we avoid the influence of outliers and improve the results substantially. Existing work dealt with sample differences of only $4\times$ resolution between the detailed surface and the low-resolution geometry (Nehab et al., 2005), while our resolution differences are much more than that. A recent work (Lu et al., 2010) presented a framework to deal with the ultra-high-resolution 3D reconstruction, however, the algorithm is sensitive to outliers and the estimate is skewed by outliers for the reason that they have employed the least square method in a multi-resolution surface reconstruction scheme.

In contrast with ℓ-2 methods such as the least quare method, global optimization strategies such as convex optimization overcome the problem of local minima by a global optimization. Classic computer vision problems are usually defined on a discrete domain, keeping them away from convex properties. Nonetheless, a recent work (Pock et al., 2008) demonstrate that mulit-label problems such as stereo matching and image restoration in computer vision areas can utilize convex optimization by relaxing the original problem from a discrete domain to a continuous one.

Inspired by the idea of matching, we consider the

Li M., Diao C., Lv S. and Lu D..
A Convex Framework for High Resolution 3D Reconstruction.
DOI: 10.5220/0005306503170324
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 317-324
ISBN: 978-989-758-091-8

problem of fusing geometry information and normal resolution as a multi-label problem. By incorporating the TV regularization term, we develop a convex variational framework that takes advantage of both global geometry information and local normal information. Previous the least square method has regarded the low resolution geometry as an initialization of the surface reconstruction from gradient method. Instead, while we observe that we are able to acquire gradients from both high resolution normal information and low resolution geometry information, we propose a convex framework that match gradient information between the high resolution normal and the low resolution geometry, which turns into a variational problem. In order to solve the variational problem, we employ the level-sets method through which we lift the original problem to a higher dimensional space, leading to that our algorithm is more robust and effective than the existing ones.

The remainder of this paper is organized as follows: Section 2 discusses related works; Section 3 presents our main surface construction problem; Section 4 presents our results. We give our conclusion in Section 5.

2 RELATED WORK

A variety of approaches (Herbort and Wöhler, 2011; Scharstein and Szeliski, 2002; Seitz et al., 2006; Salvi et al., 2010) have been developed for 3D reconstruction. We usually classify the 3D reconstruction methods as passive methods such as conventional stereo (Bernardini et al., 2002) and shape from shading (Horn and Brooks, 1989), and active methods such as structured light (Salvi et al., 2010) and photometric stereo (Woodham, 1980). Moreover, hybrid methods, focusing on combining position information with normal information are presented to acquire dense 3D points, such as ((Nehab et al., 2005; Lu et al., 2010; Banerjee et al., 1992; Bernardini et al., 2002; Lange, 1999; Aliaga and Xu, 2010; Wu et al., 2011; Birkbeck et al., 2006)) or visual hull with normals (Hernández et al., 2008), and so on.

Among the above approaches, a popular one was presented by (Nehab et al., 2005) which fused position information and normal information into a linear formulation. Although they have obtained better results compared with the single reconstruction approach, the position data and the normal information have approximately the same resolution in their work. As a result, it is impossible to apply their approach into the high resolution reconstruction problem directly. To overcome the asymmetry between the high resolution normal information and low resolution position information, (Lu et al., 2010) proposed a multi-resolution pyramid framework and used the least square method in every layer of the pyramid. We already discussed that the least square method is sensitive to outliers, leading to a local minima problem. However, errors are also magnified during the propagation from one layer to another in Lu's method.

Opposite to the proposed methods, global optimization such as graph cuts (Sinha and Pollefeys, 2005; Hornung and Kobbelt, 2006; Ladikos et al., 2008; Higo et al., 2009; Yu et al., 2006; Vogiatzis et al., 2007) or convex optimization (Kolev et al., 2010; Pock et al., 2008) overcomes the limitation of local minima problem. Here, only representative examples are mentioned. We reconsider the hybrid method as a multi-label problem, which in general cannot be globally minimized. However, (Ishikawa, 2003) showed that one can compute the exact solution of the multi-label problem if the pairwise interactions are convex in terms of a linearly ordered label set. Based on this, researchers shifted the discrete multi-label problem to its continuous counterpart, the variational approach (Yuan et al., 2010; Pock et al., 2008). It is well known that if the energy functional is convex and the minimization is carried out on a convex set, the globally optimal solution can be computed. Thus (Pock et al., 2008) shifted the original variational model to a higher dimensional space and developed a convex formulation.

There is great potential for all the above mentioned hybrid methods to implement global optimization by adding a regularization term. Using the setup which is similar to (Lu et al., 2010), we intend to present a convex framework and adopt the sub-pixel continuous formulation proposed by (Pock et al., 2008), which makes use of continuous optimization techniques.

3 SURFACE RECONSTRUCTION WITH A CONVEX FRAMEWORK

3.1 Conventional Surface Reconstruction from a Gradient Field

Estimating Surface Normals. According to conventional photometric stereo (Woodham, 1980), given a lambertian surface, we are able to estimate surface

Figure 1: Our setup. Our experimental setup consists of a digital camera, four lights and a digital light projector. We use the camera for both photometric stereo and structured-light scan.

normals from the following equation:

$$I = \rho L \cdot n \tag{1}$$

where n is the normal we want to compute, I is the image intensity, ρ is the surface albedo which is a constant here, and L is the light source direction which can be calibrated by a mirrored sphere.

Surface from Gradient Field. Consider the equation of an object as $Z = f(x,y)$ and $\Omega \in \mathbb{R}$ is the image domain. $\mathbf{x} = (x,y)^T \in \Omega$ is the pixel coordinate. Let (p,q) denote the observed gradient field over this surface. Then we can easily get $n = (\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y}, 1) = (p,q,-1)$. We estimate p and q by photometric stereo. A common approach to obtain surface Z is to minimize the LS function by (Horn, 1990; Agrawal et al., 2006). Then we get:

$$\mathcal{D}(p,q,Z) = \int_{\Omega} ((\frac{\partial Z}{\partial x} - p)^2 + (\frac{\partial Z}{\partial y} - q)^2)\,\mathrm{d}\mathbf{x} \tag{2}$$

The associated Euler-Lagrange differential equation of (2) is the poisson equation $\nabla^2 Z = div(p,q)$, where div is the divergence operator. This is the well-known poisson solver method. However, we can discretize the problem as a least problem and solve it using Gauss-Seidel iteration (Lu et al., 2010).

3.2 Low Geometry Constraints with a Convex Framework

In practice, the gradient field obtained from photometric stereo is rarely integrable due to the inherent noise in the estimation process, manipulation of gradient fields or ambiguities in the solution. Consequently, many surface from gradient algorithms do not reconstruct the surface geometry accurately. As discussed before, we can overcome this by incorporating positional information in the reconstruction process. Specifically, we estimate the orientation consistency between the gradient field acquired from photometric stereo and the real orientation.

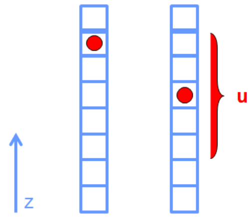


Figure 2: Representation of u. Red dots represent locations of each point on the computed by the surface-from-gradient algorithm (left red dot) and structured-light (right one).

Typically, we are able to acquire accurate surface orientation via photometric stereo as well as accurate position information through triangular methods such as stereo (Bernardini et al., 2002) and structured light (Salvi et al., 2010). As has mentioned in section 3.1, we consider the surface as $Z = f(x,y)$ while (p,q) denote the observed gradient filed over the surface.

Inspired by other computer vision problems such as stereo estimation and image segmentation, which are usually treated as labeling problems, we convert our fusing problem into a labeling problem, which includes The regularization term $\mathcal{R}(.)$ and the data term $\mathcal{D}(.)$. Our goal is to minimize the following energy functional:

$$\min_u \{\mathcal{R}(u) + \mathcal{D}(u)\} \tag{3}$$

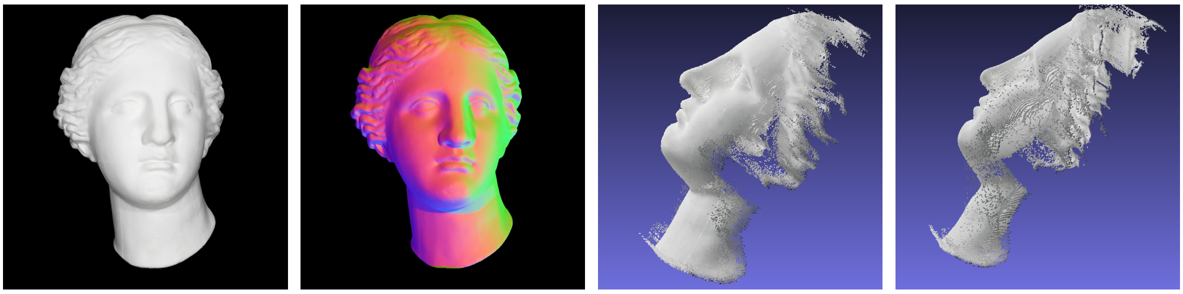
Specifically, for our problem, we utilize the offset u in the z axis position, with which the depth of each pixel on the image domain shifts up and down in order to match the related gradients (Fig.2). We utilize the TV regularization term as the regularization term in order to obtain smooth results while we update (2) as the Data term $\mathcal{D}(p,q,u)$. Then our low resolution constraint is modeled as the following variational estimation:

$$\min_u E(u) = \int_{\Omega} |\nabla u(\mathbf{x})|\,\mathrm{d}\mathbf{x} + \lambda \int_{\Omega} \rho(\mathbf{x},u)\mathrm{d}\mathbf{x} \tag{4}$$

where the right term is $D(u)$ to measure the orientation consistency and defined as:

$$\rho(\mathbf{x},u) = \left|\frac{\partial(Z+u)}{\partial x} - d(p)\right| + \left|\frac{\partial(Z+u)}{\partial y} - d(q)\right| \tag{5}$$

In (4) and (5), due to the prior of low resolution geometry $Z(x,y)$, we take the depth to shift along z



(a) image of object (b) surface normals (c) 3D reconstruction (Our) (d) 3D reconstruction (LS)

Figure 3: 3D reconstruction of the venus figure. The reconstructions show our method is more robust to outliers than the LS method.

coordinate up and down to match the orientation obtained from photometric stereo. $u(\mathbf{x})$ denote the offsets of the depth, which is somewhat similar to the disparity in stereo. $d(\cdot)$ is the downsample operator to match the low geometry and the high resolution and $|\cdot|$ is the ℓ-1 norm.

Generally, the variational model (4) is not convex due to the non-convex of the data term $\mathcal{D}(u)$. However, we can develop a convex formulation via lift variational model (4) to a higher dimensional space by representing u in terms of its level sets, which allows us to compute the exact solution of the original non-convex problem.

We utilize the functional lifting method of (Pock et al., 2008). For simplify, We employ the expressions of the paper in (Pock et al., 2008), where $\mathbf{l}_{u>\gamma}$ is the indicator for the γ-*super-levels* of u and $\phi(\mathbf{x},\gamma) = \mathbf{l}_{u>\gamma}(\mathbf{x})$ denotes the binary function to resemble the graph of u. The variational model (4) is equivalent to the following high dimensional variational problem:

$$\min_{\phi\in D}\{|\nabla\phi(\mathbf{x},\gamma)| + \rho(\mathbf{x},\gamma)|\partial_\gamma\phi(\mathbf{x},\gamma)|d\Sigma\} \quad (6)$$

In (6), D is the relaxed feasible set of ϕ from binary interval $\{0,1\}$ to $[0,1]$ as:

$$D = \{\phi : \Sigma \to [0,1] | \phi(\mathbf{x},\gamma_{min}) = 1, \phi(\mathbf{x},\gamma_{max}) = 0\} \quad (7)$$

Consequently, (6) is convex in ϕ and minimization is carried over D, which is convex, the overall problem is convex. (Pock et al., 2008)

Then we solve the associated Euler-Lagrange function of (4), to find its global minimizer, which we compute as:

$$-div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \partial_\gamma\left(\rho\frac{\partial_\gamma\phi}{|\partial_\gamma\phi|}\right) = 0, s.t. \quad \phi \in D \quad (8)$$

To avoid that denominators of (6) become zeros, we replace them with the robust function:

$$f(|s|) = \sqrt{s^2+\xi^2} \quad (9)$$

where ξ is a small constant.

3.3 Discretization

Considering a three-dimensional regular cartesian grid in our numerical implementation, we get:

$$D_s = \Big\{ (i\cdot\Delta x, j\cdot\Delta y, k\cdot\Delta\gamma) \Big| \; 0 \le i < M, 0 \le j < N, 0 \le k < O \Big\} \quad (10)$$

where $M \times N$ denotes the grids of image domain and O is the range of depth value. We utilize standard forward differences to approximate the gradient operator, which is:

$$(\nabla_3\phi)_{i,j,k} = \left(\frac{\phi_{i+1,j,k}-\phi_{i,j,k}}{\Delta x}, \frac{\phi_{i,j+1,k}-\phi_{i,j,k}}{\Delta y}, \frac{\phi_{i,j,k+1}-\phi_{i,j,k}}{\Delta\gamma}\right)^T \quad (11)$$

and the divergence operator is:

$$(div_3(\nabla_3\phi))_{i,j,k} = \frac{(\nabla_3\phi)^1_{i,j,k+1}-(\nabla_3\phi)^1_{i,j,k}}{\Delta x} + \frac{(\nabla_3\phi)^2_{i,j,k+1}-(\nabla_3\phi)^2_{i,j,k}}{\Delta y} + \frac{(\nabla_3\phi)^3_{i,j,k+1}-(\nabla_3\phi)^3_{i,j,k}}{\Delta\gamma} \quad (12)$$

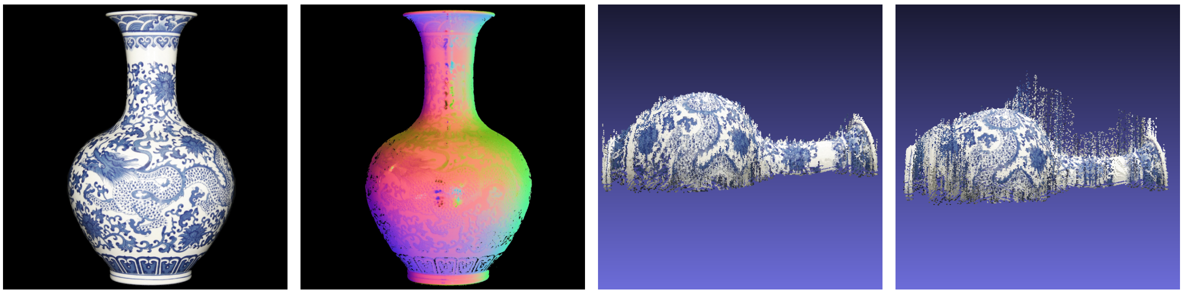
and we use center differences to approximate the partial derivative as the following:

$$\partial_\gamma\phi = \frac{\phi_{i,j,k+1/2}-\phi_{i,j,k-1/2}}{\Delta\gamma} \quad (13)$$

where Δx, Δy denote the width of spatial discretization and $\Delta\gamma$ denotes the height of the depth discretization.

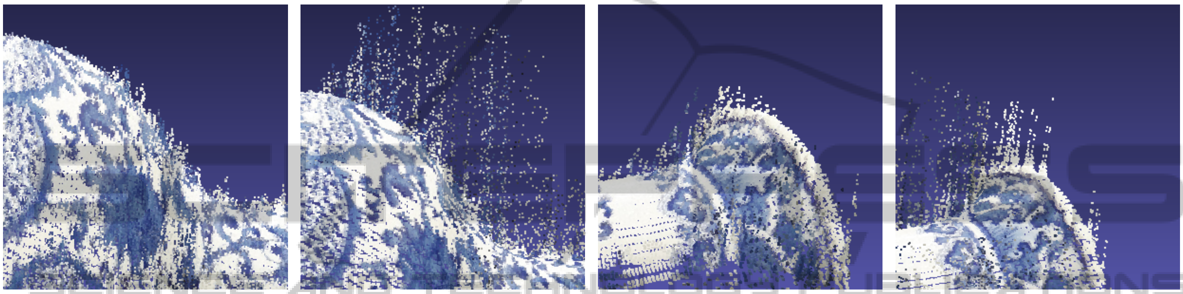
4 RESULTS

We estimate the high-resolution surface normals through photometric stereo images captured by four



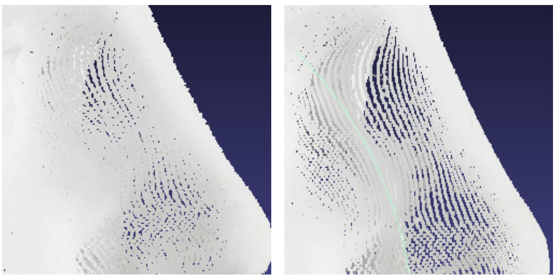
(a) image of object (b) surface normals (c) 3D model (Our) (d) 3D model (LS)

Figure 5: 3D reconstruction of the vase figure.



(a) zoom1 (Our) (b) zoom1 (LS) (c) Zoom2 (Our) (d) Zoom2 (LS)

Figure 6: Details of vase.



(a) Zoom (Our) (b) Zoom (LS)

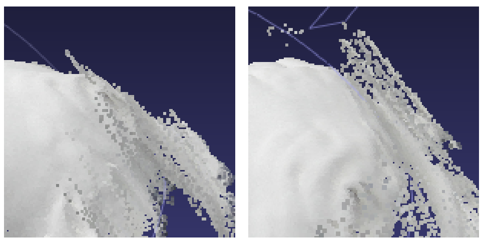
Figure 4: "Nose" details of venus. The zoom details show our method acquires better results than the LS method.

LED light sources. We estimate the low-resolution geometry using a typical structured-light system, which includes a Benq MP624 projector and a M5D camera (Fig.1). This section shows results of both synthetic data-set and data captured by our system including a 5D camera and a Benq projector (1024×768).

In order to address the difference between the surface normals and the geometry, we also employ the multi-resolution pyramid approach. However, we handle the initial depth manually in a certain degree to ensure the position of every point stands not so far away from its accurate position. We find that results of least square (LS) method are nosier than our method since the LS method is much more sensitive to outliers than ours. After the simple deposing step, we confine the offset u of z axis value to a particular range instead of shifting the unknown values from zero. Besides, we discretize depth value with a subpixel interval so as to acquire the final reconstructed surface as accurate as possible.

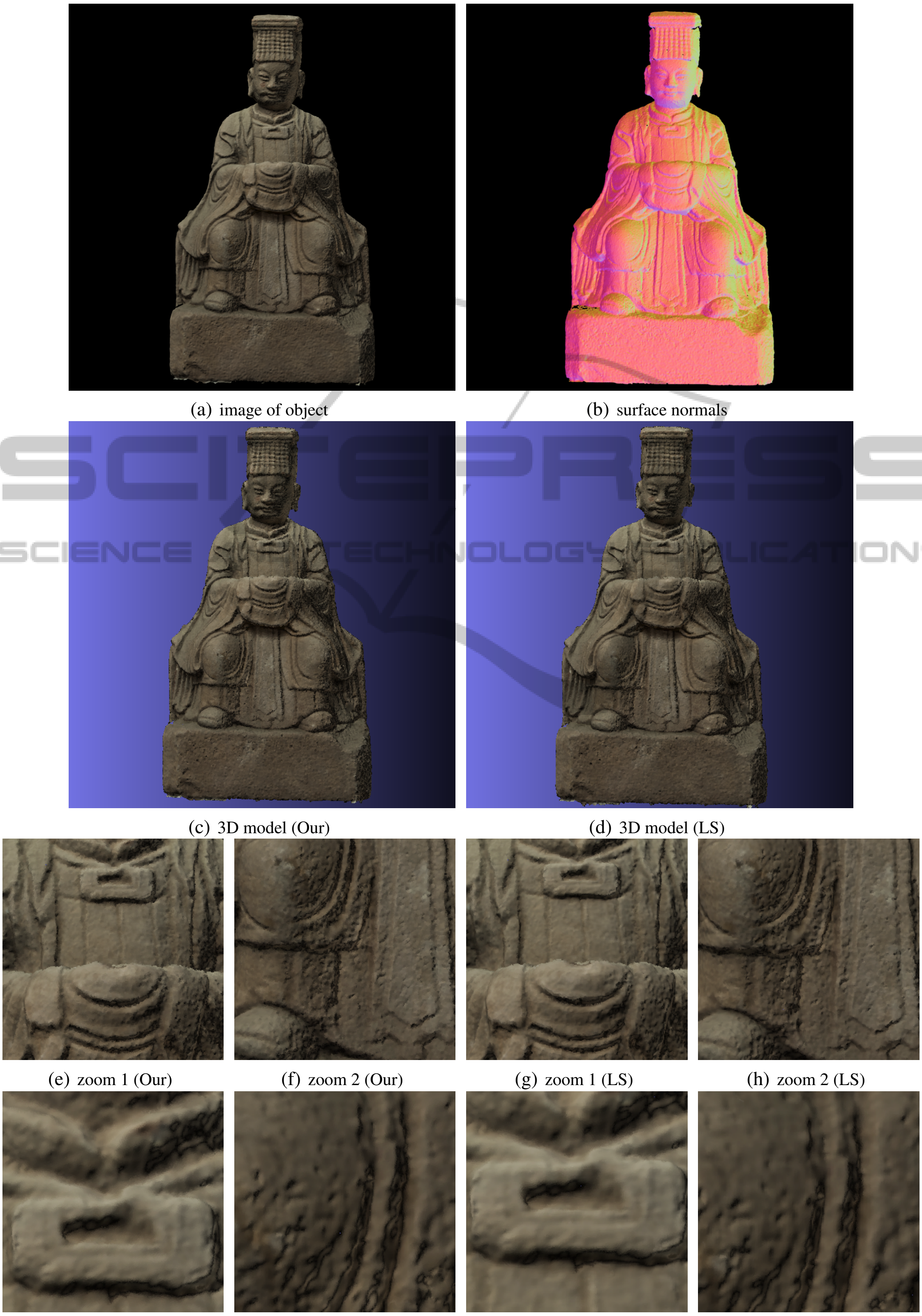
Fig.3 shows a 3D reconstruction of the "venus" statue which is approximately 30cm wide and Fig.5 shows a china of about 35cm high, both captured by our system. The initial model contains tens of thousands points and the final results is more than twenty times than that. All of these results shows that our method is much less sensitive to the outliers than the LS method. Meanwhile we also retain the local surface details excellently. In Fig.4 and Fig.6, it reveals that we obtain more delicate details especially in the sharp positrons.

Finally, in order to demonstrate that our method is also able to achieve ultra-high resolution, we make



(a) Our method (b) LS method

Figure 7: The top of the head of venus.



(a) image of object (b) surface normals

(c) 3D model (Our) (d) 3D model (LS)

(e) zoom 1 (Our) (f) zoom 2 (Our) (g) zoom 1 (LS) (h) zoom 2 (LS)

(i) double zoom 1 (Our) (j) double zoom 2 (Our) (k) double zoom 1 (LS) (l) double zoom 2 (LS)

Figure 8: 3D reconstruction of the buddha figure. And we show “zoom” and “double-zoom” surfaces.

use 3DMAX to render data of a 14cm-high buddha. We compare the results in Fig.8. The original 3D model we use to render is sampled about 150 samples per mm^2 and we promote the sampling rate of our synthetic data to 510 samples per mm^2. The object is reconstructed about 5.7 millions of 3D points and we double zoom the buddha to show that both our method and LS method can reveal the delicate details. However, our results are much insensitive to the outliers, especially on some boundary areas, which is shown in Fig.7 for example.

5 CONCLUSIONS

We demonstrate that we are able to reconstruct high-quality surfaces through global optimized framework and to represent good surface details at the same time. Although certain hybrid systems have been presented to address the problem of fusing normal information and geometry data, only few of them attempted to surpass the resolution of structured-light systems to the ultra-high resolution level. However, the results of the ultra-high resolution construction are inevitable to influence by outliers due to the inherent nature of LS method. To overcome these issues, we introduce a convex framework to ensure that the high-quality surface is reconstructed progressively and the error is reasonably under control. Consequently, we are able to implement ultra high resolution 3D reconstruction while retaining subtle details close to local methods such as the LS method.

Even our method do handle quite a vast of materials, the normal estimation process still needs to be improved for the extremely specular surfaces. In future we will improve the normal information by utilizing an alternative BRDF model and incorporate color information to make the estimation more reliable. In addition, we also consider to develop a more efficient numerical algorithm to minimize our variational model, so as to avoid the "zero-denominator" problem of the fixed point algorithm and to accelerate the convergence rate.

ACKNOWLEDGEMENTS

This work is supported by Issue of National Science and Technology Support Plan (No. 2012BAH03F02, No. 2013BAK08B08), Important Project Subject to Tender of The National Social Science Fund (No. 12&ZD32), and Science and Technology Project on Heritage Conservation (Detection and Simulation Technology for Heritage Preservation Environment in Archaeological Sites), Zhejiang Province.

REFERENCES

Agrawal, A., Raskar, R., and Chellappa, R. (2006). What is the range of surface reconstructions from a gradient field? In *ECCV*, pages 578–591. Springer.

Aliaga, D. G. and Xu, Y. (2010). A self-calibrating method for photogeometric acquisition of 3d objects. *IEEE Transactions. Pattern Analysis and Machine Intelligence*, 32(4):747–754.

Banerjee, S., Sastry, P., and Venkatesh, Y. (1992). Surface reconstruction from disparate shading: An integration of shape-from-shading and stereopsis. *IAPR*, 1:141–144.

Bernardini, F., Rushmeier, H., Martin, I. M., Mittleman, J., and Taubin, G. (2002). Building a digital model of michelangelo's florentine pieta. *IEEE Transactions. Computer Graphics and Applications*, 22(1):59–67.

Birkbeck, N., Cobzas, D., Sturm, P., and Jagersand, M. (2006). Variational shape and reflectance estimation under changing light and viewpoints. In *ECCV*, pages 536–549. Springer.

Herbort, S. and Wöhler, C. (2011). An introduction to image-based 3d surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3):1–17.

Hernández, C., Vogiatzis, G., and Cipolla, R. (2008). Multiview photometric stereo. *IEEE Transactions. Pattern Analysis and Machine Intelligence*, 30(3):548–554.

Higo, T., Matsushita, Y., Joshi, N., and Ikeuchi, K. (2009). A hand-held photometric stereo camera for 3-d modeling. In *ICCV*, pages 1234–1241. IEEE.

Horn, B. K. (1990). Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75.

Horn, B. K. and Brooks, M. J. (1989). *Shape from shading*. MIT press.

Hornung, A. and Kobbelt, L. (2006). Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In *CVPR*, volume 1, pages 503–510. IEEE.

Ishikawa, H. (2003). Exact optimization for markov random fields with convex priors. *IEEE Transactions. Pattern Analysis and Machine Intelligence*, 25(10):1333–1336.

Kolev, K., Pock, T., and Cremers, D. (2010). Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *ECCV*, pages 538–551. Springer.

Ladikos, A., Benhimane, S., and Navab, N. (2008). Multiview reconstruction using narrow-band graph-cuts and surface normal optimization. In *BMVC*, pages 1–10.

Lange, H. (1999). Advances in the cooperation of shape from shading and stereo vision. In *Second International Conference on 3-D Digital Imaging and Modeling*, pages 46–58. IEEE.

Lu, Z., Tai, Y.-W., Ben-Ezra, M., and Brown, M. S. (2010). A framework for ultra high resolution 3d imaging. In *CVPR*, pages 1205–1212. IEEE.

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

Pock, T., Schoenemann, T., Graber, G., Bischof, H., and Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *ECCV*, pages 792–805. Springer.

Salvi, J., Fernandez, S., Pribanic, T., and Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, volume 1, pages 519–528. IEEE.

Sinha, S. N. and Pollefeys, M. (2005). Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *ICCV*, volume 1, pages 349–356. IEEE.

Vogiatzis, G., Hernández, C., Torr, P. H., and Cipolla, R. (2007). Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions. Pattern Analysis and Machine Intelligence*, 29(12):2241–2246.

Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139.

Wu, C., Liu, Y., Dai, Q., and Wilburn, B. (2011). Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *IEEE Transactions. Visualization and Computer Graphics*, 17(8):1082–1095.

Yu, T., Ahuja, N., and Chen, W.-C. (2006). Sdg cut: 3d reconstruction of non-lambertian objects using graph cuts on surface distance grid. In *CVPR*, volume 2, pages 2269–2276. IEEE.

Yuan, J., Bae, E., and Tai, X.-C. (2010). A study on continuous max-flow and min-cut approaches. In *CVPR*, pages 2217–2224. IEEE.