ORIGINAL ARTICLE

CrossMark

A self-calibrated photo-geometric depth camera

Liang Xie[1] · Yuhua Xu[1] · Xiaohu Zhang[2] · Wei Bao[3] · Chenpeng Tong[4] · Boxin Shi[5]

Abstract

Compared with geometric stereo vision based on triangulation principle, photometric stereo method has advantages in recovering per-pixel surface details. In this paper, we present a practical 3D imaging system by combining the near-light photometric stereo and the speckle-based stereo matching method. The system is compact in structure and suitable for multi-albedo targets. The parameters (including position and intensity) of the light sources can be self-calibrated. To realize the auto-calibration, we first use the distant lighting model to estimate the initial surface albedo map, and then with the estimated albedo map and the normal vector field fixed, the parameters of the near lighting model are optimized. Next, with the optimized lighting model, we use the near-light photometric stereo method to re-compute the surface normal and fuse it with the coarse depth map from stereo vision to achieve high-quality depth map. Experimental results show that our system can realize high-quality reconstruction in general indoor environments.

Keywords Photometric stereo · Stereo matching · Lighting calibration · Random speckle

1 Introduction

3D surface imaging technique has widely applications such as 3D modeling, reverse engineering, 3D printing, human body measurement, movies and animation, and humanC-machine interaction [1]. Phase-shifting-based systems with the digital light processing (DLP) projector have high measurement accuracy, but the cost is relatively high [2,3]. In recent years, the consumer depth sensors (e.g., Kinect [4]) greatly reduce the cost of 3D data acquisition. However, these low-cost sensors only have limited accuracy. In order to improve reconstruction quality, one kind of method is fusing the aligned depth maps with a volumetric, truncated signed distance function (TSDF) representation, such as KinectFu-

sion [5]. Another popular approach is to combine the normal information obtained by the photometric stereo (PS) method with the depth map obtained by these depth sensors [6–8].

Conventional photometric stereo (PS) estimates surface normal from an image sequence taken from a fixed viewpoint under varying directional (distant) lightings. Instead of directly measuring position/depth, PS estimates surface orientations by measuring the shading variations of surface under different illuminations. A 3D shape up to a scale can be obtained via integration over the obtained surface normal field. PS is an excellent solution for obtaining surface normal but not directly for depth, largely because integrating normals is prone to introducing low-frequency biases [9]. Another reason that PS is not as popular as other 3D modeling approach is because it works best for laboratory setup which needs controlled environment [9].

Combining advantages from both depth and normal sensors has been studied in recent years [6–8,10–12]. Such fusion approach achieves high-quality 3D reconstruction by integrating the coarse base geometry estimated from the depth sensor with high-resolution surface normal details from PS. However, in these methods, the point light source-based systems have compact structures, but usually require specific calibration devices (e.g., mirror spheres) to precisely calibrate the parameters of the light sources [13–15]. With the directional lighting assumption, self-calibration of the light

✉ Yuhua Xu
xyh_nudt@163.com

1 College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, Hunan, China

2 School of Aeronautics and Astronautics, Sun Yat-sen University, Guangzhou 510275, Guangdong, China

3 School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, Anhui, China

4 DeepInTech, Shenzhen 518055, Guangdong, China

5 Institute of Digital Media, School of EECS, Peking University, Beijing 100871, China

sources can be realized [7]. To obtain the directional lighting, the light sources have to be placed far away from the targets to be measured, which makes the systems uncompact. Besides, some methods assume that the target surface has a uniform albedo [6], so they cannot be applied to multi-albedo targets which are very common in practical applications.

Motivated by the shortcomings of existing systems, our goal is to design a practical PS system that meets the following conditions:

(1) The near point light source model is adopted for making a compact system.

(2) The parameters of the point light sources can be calibrated automatically.

(3) The system can be adapted to the multi-albedo objects.

(4) The PS method and the stereo matching method are combined to reconstruct surfaces with rich details and high accuracy.

According to the guideline above, we design a 3D imaging system combining the PS method and the traditional stereo vision, called the photo-geometric depth camera. The system consists of two CMOS cameras, a near-infrared speckle projector, four near-infrared LED, and a synchronous circuit, as shown in Figs. 1 and 5. The experimental hardware cost is about $400.

The main contributions of this paper are:

(1) A high-precision 3D imaging system is introduced, which combines the photometric stereo and the binocular stereo. The system is low cost and compact and can be adapted to multi-albedo targets in the general indoor environment.

(2) A point light source auto-calibration algorithm is proposed. The traditional point light source calibration method usually requires specific calibration objects (such as the mirror spheres). In our method, we firstly use the distant lighting model [7] to estimate the initial surface albedo map. Then, with the estimated albedo map and the normal vector field fixed, the parameters of the near lighting model are optimized. Next, with the optimized lighting model, we use the near-light photometric stereo (NLPS) method to re-compute the albedo map and use the method [9] to compute higher-quality depth map. We repeat the above two steps iteratively until convergence or the iteration times reaching a predefined maximum number.

The remainder of this paper is organized as follows: The details of the proposed system are presented in Sect. 2, and experimental results are provided in Sect. 3. Finally, Sect. 4 concludes the paper.
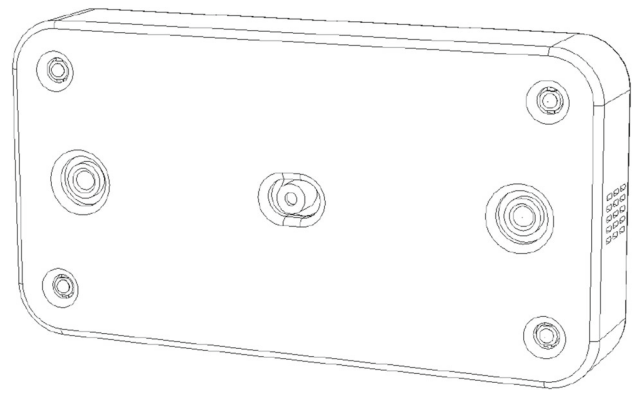

Fig. 1 The proposed photo-geometric depth camera

2 Photo-geometric depth camera

2.1 Hardware

Our photo-geometric depth camera consists of two CMOS cameras, a Kinect-type near-infrared (NIR) speckle projector, four NIR LEDs with a wavelength of 830 nm, and a microcontroller-based circuit, as shown in Fig. 1. The cameras have a frame rate of 60 fps and a resolution of 1280×960. The cameras are connected to the PC via Gigabit Ethernet interfaces. The narrowband filters are mounted on the 4-mm-focal-length lenses to filter out ambient light. The two cameras and the four LEDs are mounted on a rigid structure to keep the stable relative positional relationship. The cameras, the LEDs, and the projector are synchronized by the trigger signal of the microcontroller-based circuit, which is controlled by the PC via USB 2.0 interface. The NIR projector can emit a large number of random speckles to enhance the surface texture.

As illustrated in Fig. 2, in a reconstruction cycle, the projector is lit firstly and the cameras are triggered simultaneously to capture a pair of images of the speckles. The stereo image pair is used to generate the initial depth map. Then, the four LEDs are lit one by one, and the cameras are triggered to capture images under the illumination of each LED. In these images, the images of the left camera are used for photometric computation.

2.2 Near-light photometric stereo

We use the near point light source assumption in our image formation model [11]. In the jth image, the light vector $\mathbf{l}_{ij}$ from the surface point $\mathbf{x}_i$ to the light source $\mathbf{s}_j$ is written as

$$\mathbf{l}_{ij} = \mathbf{s}_j - \mathbf{x}_i \tag{1}$$

With the near light source assumption, intensity observation o_i is computed with accounting the inverse square law

Fig. 2 The timing diagram of a reconstruction cycle
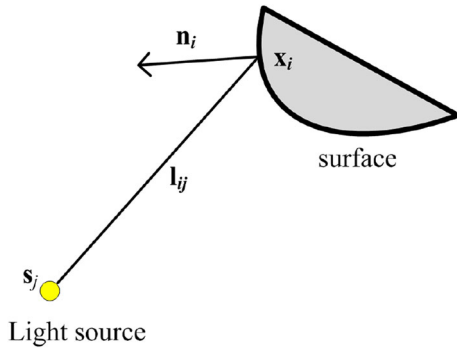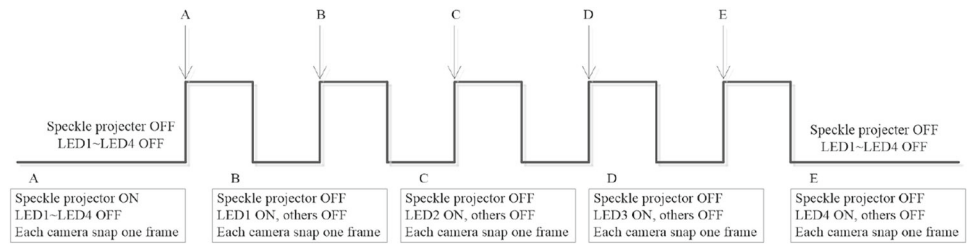

Fig. 3 Near-light photometric stereo model

as

$$o_i = E_j \rho_i \frac{\mathbf{l}_{ij} \cdot \mathbf{n}_i}{\|\mathbf{l}_{ij}\|^3} = E_j \rho_i \frac{(\mathbf{s}_j - \mathbf{x}_i) \cdot \mathbf{n}_i}{\|\mathbf{s}_j - \mathbf{x}_i\|^3} \tag{2}$$

where E_j is the light source intensity at a unit distance, ρ_i is surface albedo, and $\mathbf{n}_i$ is the surface normal vector (Fig. 3).

Once we know the light source parameters, we can estimate the normal vector $\mathbf{n}_i$ and the albedo ρ_i according to Eq. (2) from at least three observations. In order to balance the efficiency and the quality, we use four point light sources in our setup.

2.3 Reconstruction pipeline

We use the stereo camera calibration method in [16] to calibrate our stereo cameras. The light source will be self-calibrated using the method to be discussed in Sect. 2.5.

Given the calibrated cameras and lightings, the 3D reconstruction pipeline is illustrated in Fig. 4. Firstly, the stereo matching method is applied to the speckle image pair to generate the initial depth map. Then, the photometric computation is applied to the four images under the illumination of the four LEDs, respectively, to generate the surface normal vector field. Finally, the obtained initial depth map and the normal vector field are integrated to generate the higher-quality depth map.

2.4 Initial depth generation

The semi-global matching (SGM) method [17] performs an energy minimization using dynamic programming on multiple 1D paths. The energy function consists of three terms: a data term for photo-consistency, a smoothness term for slanted surfaces that change the disparity slightly (parameter P_1), and a smoothness term for depth discontinuities (parameter P_2). Due to that SGM has a good balance in efficiency and accuracy, we use this method to estimate the initial depth map.

2.5 Point light source self-calibration

This section proposed a new calibration method for point light source including the geometric parameters and the light intensity. With the self-calibration method, our system does not rely on the fixed calibration such as mirror spheres [13–15], which makes the system more flexible and practical.

To estimate E_j and $\mathbf{s}_j$, we use the system described in Sect. 1 to capture five image pairs of the target according to the timing diagram shown in Fig. 2. Our light source calibration firstly makes a distant lighting assumption and estimate a rough albedo map. Then, an iterative manner optimization is applied to estimate the parameters of the near light sources. The calibration algorithm is summarized as follows.

Algorithm 1: Point light source calibration

(1) Initialization

Rough depth map generation Each stereo pair is rectified to obtain a row-aligned epipolar geometry. The stereo matching method described in Sect. 2.4 is applied to the speckle image pair to generate the initial depth map $\mathbf{D}_0$ of the target. A bilateral filter [18] is applied to the raw depth map to obtain a discontinuity preserved depth map with reduced noise $\mathbf{D}_0'$.

$$\mathbf{D}_0'(\mathbf{u}) = \frac{1}{W_p} \sum_{q \in nb(\mathbf{u})} N_{\sigma_s}(\|\mathbf{u} - \mathbf{q}\|) N_{\sigma_s}(\|\mathbf{D}_0(\mathbf{u}) - \mathbf{D}_0(\mathbf{q})\|) \mathbf{D}_0(\mathbf{q}) \tag{3}$$

where $N_\sigma = exp(-t^2 \sigma^{-2})$, $nb(\mathbf{u})$ denotes the neighborhood of the pixel $\mathbf{u}$ and W_p is a normalizing constant.

Initial position estimation According to the mounting position of the LEDs relative to the reference camera (we

Fig. 4 The pipeline of the
reconstruction
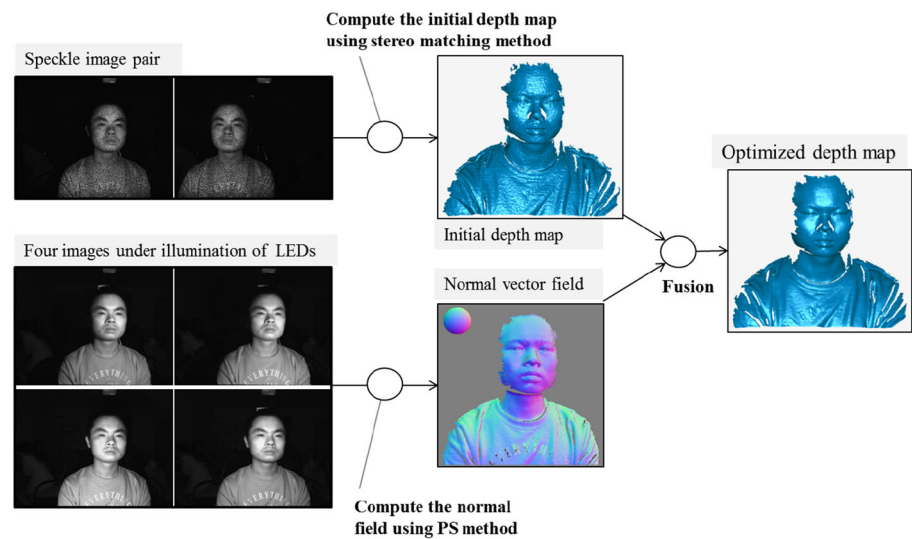
use the left camera as the reference camera), we can estimate
an initial value $\mathbf{s}_{j,0}$ for $\mathbf{s}_j$.

Initial albedo estimation We follow the automatic calibra-
tion method [7] with distant lighting assumption to estimate
the initial albedo map ρ. We first robustly estimate a rank-3
approximation of the observed brightness matrix using an
iterative re-weighting method and then factorize this rank
reduced brightness matrix into the corresponding lighting,
albedo and surface normal components.

Initial intensity estimation With the depth map $\mathbf{D}_0'$ and the
camera parameters, the point cloud $\mathbf{x}_i$ of the target surface
can be generated. Furthermore, the surface normal vectors $\mathbf{n}_i$
can be estimated with the point cloud [19]. Up to now, using
Eq. (2), we can estimate the initial value of E_j with the linear
least square method.

$$E_j = \frac{\sum_{i=1}^{N} o_i \|\mathbf{l}_{ij}\|^3}{\sum_{i=1}^{N} \rho_i \mathbf{l}_{ij} \cdot \mathbf{n}_i} \tag{4}$$

where N is the number of the surface points.

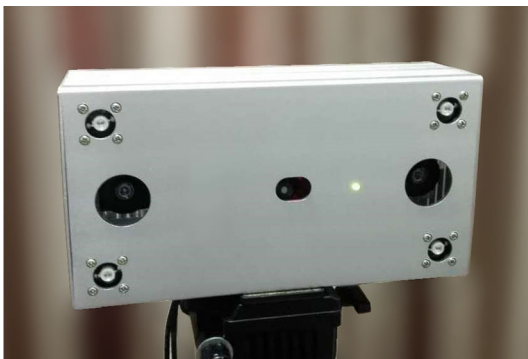(2) **Position and intensity refinement**

Fig. 5 Prototype of the proposed photo-geometric depth camera

Equation (2) is a typical nonlinear least squares problem.
With the estimated initial values, we use the Levenberg–
Marquardt (LM) algorithm [20] to optimize E_j and $\mathbf{s}_j$ with
the albedo map and the normal field fixed. The cost function
is defined as:

$$cost(\mathbf{E}, \mathbf{S}) = \sqrt{\frac{1}{4N} \sum_{j=1}^{4} \sum_{i=1}^{N} \left(o_i - E_j \rho_i \frac{(\mathbf{s}_j - \mathbf{x}_i) \cdot \mathbf{n}_i}{\|\mathbf{s}_j - \mathbf{x}_i\|^3} \right)^2} \tag{5}$$

where $\mathbf{E} = \{E_j\}$, $\mathbf{S} = \{\mathbf{s}_j\}$, $j = 1, 2, 3, 4$.

(3) **Updating albedo map, surface points, and normals**

With the optimized E_j and $\mathbf{s}_j$, we use the near lighting
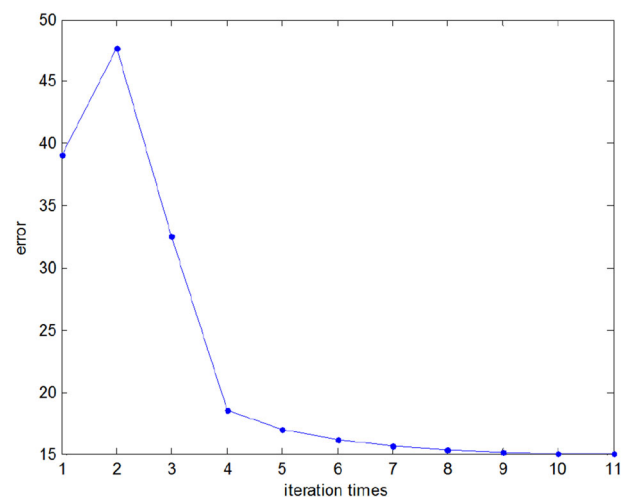model [(Eq. (2)] to re-compute the albedo map and the nor-

Fig. 6 Convergence curve of the iterative optimization of the light
source parameters for the dataset shown in Fig. 4. The error is defined
in Eq. (5)

Fig. 7 Comparisons of the estimated albedo maps and the normal vector fields. **a** The gray images of the three targets and the results of the estimated albedo maps. **b** Results of the estimated normal vector fields. In **a**, the first column shows the gray images, the second column shows the results of [7], and the last column shows our results. In **b**, the left shows the results of [7], and the right shows ours
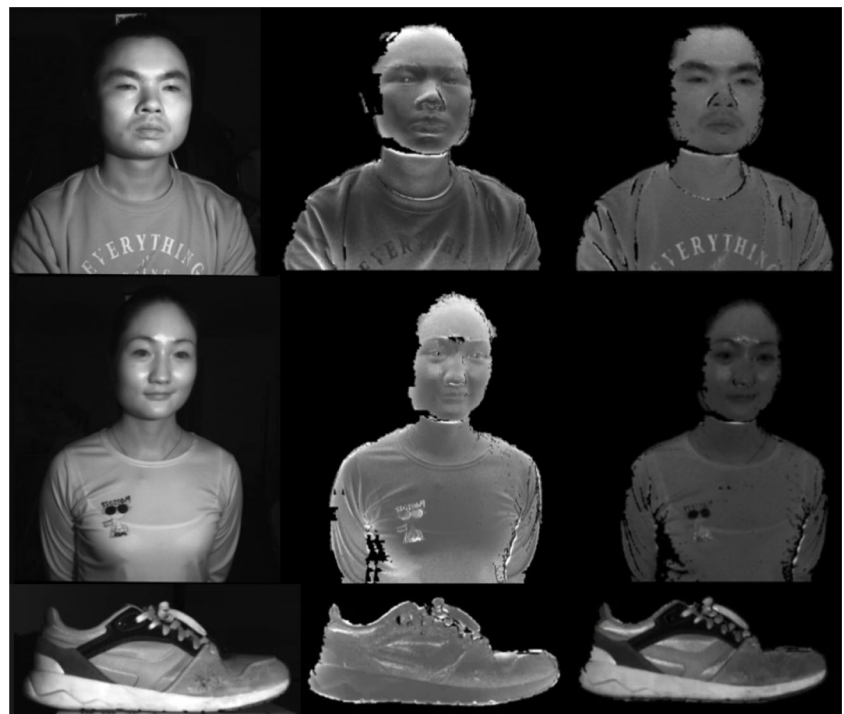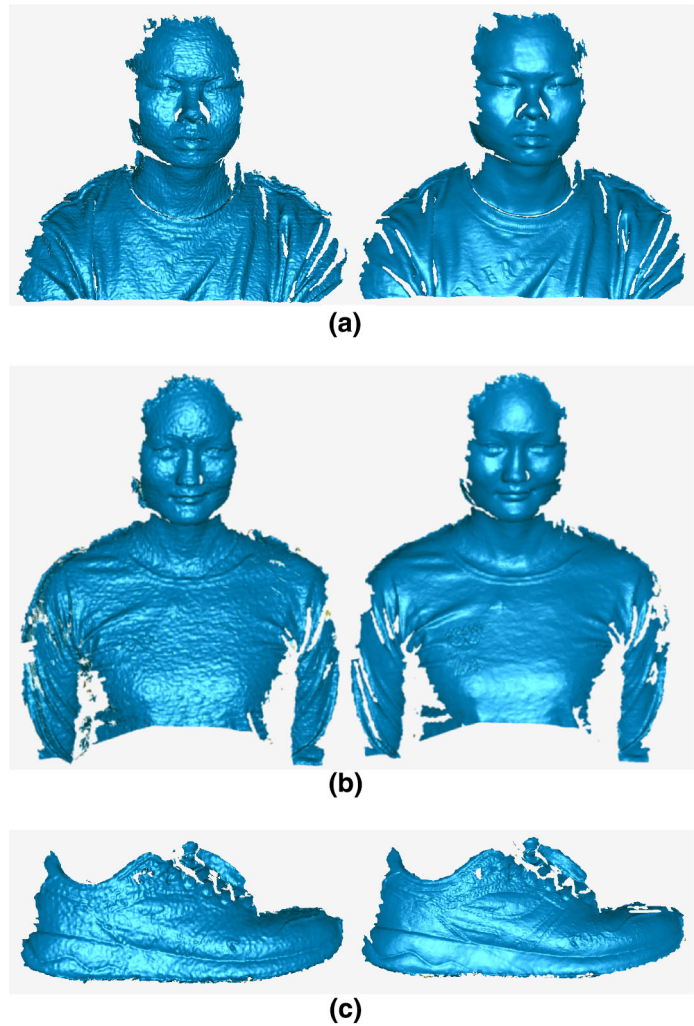

(a)

(b)

Fig. 8 Comparisons of
reconstruction results. **a** Results
of the male. **b** Results of the
female. **c** Results of the shoe.
The left column is the results of
the stereo matching, and the
right column is the results after
fusion

(a)

(b)

(c)

mal field. By combining the rough depth map and the normal field with the method in [9], the higher-quality depth map can be acquired. Note that the normal field used for the following optimization is obtained from the optimized depth map, rather than the PS method.

(4) **Iterative optimization**

Jump to Step (2) until convergence or the iteration times reaching the predefined maximum number.

2.6 Depth normal fusion

To estimate the optimal depth by combining the normal vector field by the PS method and the rough depth map by the stereo matching, we can form a linear system of equations as [9] to refine the quality of the reconstructed surface:

$$\begin{bmatrix} \lambda \mathbf{I} \\ \nabla^2 \end{bmatrix} [\mathbf{D}^*] = \begin{bmatrix} \lambda \mathbf{D} \\ \partial \mathbf{N}^* \end{bmatrix} \tag{6}$$

where $\mathbf{D}^*$ is the refined depth map, ∇^2 is a Laplacian operator, $\mathbf{I}$ is an identity matrix, and λ is a weighting parameter controlling the contribution of depth constraint. $\partial \mathbf{N}^*$ is the stacks of $-\frac{\partial}{\partial x} \frac{n_x}{n_y} - \frac{\partial}{\partial y} \frac{n_y}{n_z}$ for each normal $\mathbf{n} \in \mathbf{N}^*$. While it forms a large linear system of equations, because the left matrix is sparse, it can be efficiently solved using existing sparse linear solvers (e.g., CHOLMOD [21]).

3 Experimental results

Figure 5 shows the prototype of the proposed photogeometric depth camera. The baseline length of the stereo system is 176.92 mm. The dimension of the depth camera is 260 mm × 76 mm × 150 mm.

3.1 Qualitative evaluation

We firstly use three targets including a male, a female, and a shoe to evaluate our depth cameras. The gray images of the three targets are shown in Fig. 7a.

Figure 6 shows the convergence curve of the iterative optimization process for the target in Fig. 4. The Y-axis is the root-mean-square error defined in Eq. (4). After 10 iterations, the error converges.

We compare our method with the distant lighting model [7]. Figure 7a shows the estimated albedo maps, and Fig. 7b shows the estimated normal vector fields. From Fig. 7a, we can know that the albedo of the eyebrows of the two persons is relatively low and the albedo of the words on clothes of the male is relatively high. Our results correctly reflect these facts. However, the method in [7] cannot show these. The albedo bias of [7] is also severe for the shoe. Furthermore, the estimated surface normal vectors in face regions of the method in [7] are severely biased. These results show that the auto-calibration method in Sect. 2.5 improves the quality of the estimated albedos and normals greatly.

Figure 8 shows the reconstruction results of three targets. Figure 8a shows the results of the male shown in Figs. 7a and 8b shows the reconstruction result of a female, and Fig. 8c shows the results of a shoe. The left is the result using the speckle images, and the right is the result after combining the initial depth map by the stereo matching and the normal vector field by the PS method. These results show that the reconstruction quality can be improved remarkably by combining the near-light PS method, in which the parameters of the point light sources are calibrated automatically using our calibration method described in Sect. 2.5.

We also compare our depth camera with Kinect, a popular depth camera. Figure 9a shows the reconstruction results of a 30-cm tall David sculpture, and Fig. 9b shows the results of our depth camera. The volume voxels resolution of Kinect-Fusion is set to 512^3. We can see that the result of our depth camera has higher quality in reconstruction details.

3.2 Quantitative evaluation

Figure 10 shows the quantitative evaluation results, where the result of a commercial phase-shifting system with nominal accuracy of 0.025 mm is treated as the reference model. The David sculpture is scanned in 14 different views using the system. The obtained point clouds are stitched together, and then, the triangular mesh model is calculated utilizing Geomagic Qualify [22]. Figure 10a shows the comparison with KinectFusion in geometric accuracy. The root-mean-square error (RMSE) of Kinect is 1.58 mm, and the RMSE of our depth is 0.36 mm. Figure 10b, c shows the comparison results of the estimated normals and albedos with [7]. To evaluate the estimated normals quantitatively, the nor-
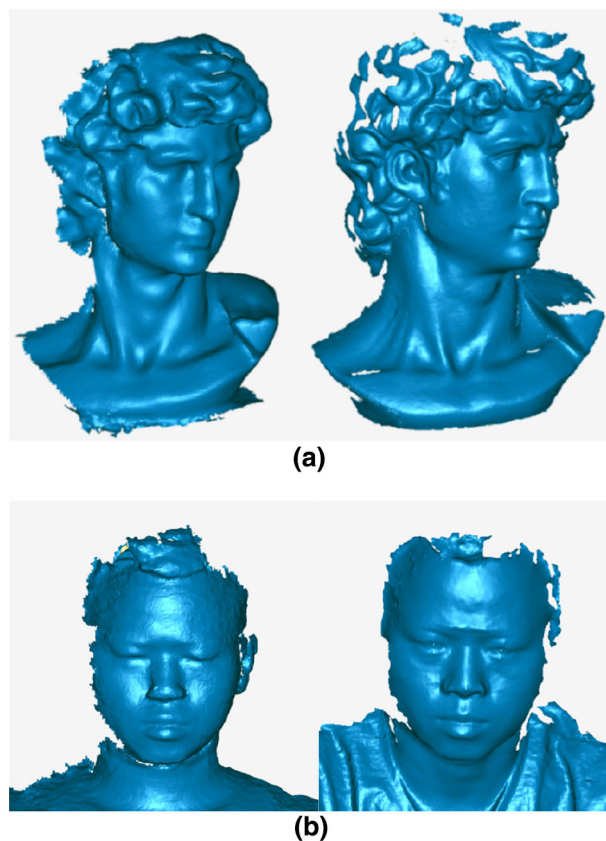

(a)

(b)

Fig. 9 Comparison with Kinect. **a** Results of a David sculpture. **b** Result of a human face. The left shows the results of Kinect, and the right shows the results of our depth camera

mals of the reference model are taken as the ground truth. We first align the reconstructed point cloud with the reference model and find the closest point in reference model for each point in the reconstructed point cloud as the corresponding point and then compare their normals. The mean angle error of the normals using our method is 8.4° comparing with 16.3° from [7]. For albedos, because the surface of the David sculpture is uniform, we assume the ground-truth albedos are one everywhere. For the estimated albedos of the two methods, they have no unified scale. So we first align them to the ground-truth albedos by estimating a optimal scale before error evaluation. The RMSE of our method is 0.232, and the RMSE of [7] is 0.547.

Furthermore, we calibrate the 3D locations of the four LEDs, respectively, using a mirror sphere as in [14]. The calibration method requires the mirror sphere to be placed at least two different locations. To get more accurate results, we capture five images of the sphere at five different locations. The mean distance deviation of the four 3D points between the manual calibration and the automatic calibration method is 3.42 cm. In Fig. 11, the left is the normal field computed from the smoothed rough depth, where the low-frequency components are accurate but lack of high-

Fig. 10 Quantitative comparison. **a** The left is the reference model obtained by the high accuracy phase-shifting system, the middle is the error map of KinectFusion, and the right is the error map of our depth camera. The gray regions in error maps indicate the invalid points. **b** The first is the "ground-truth" normal map calculated from the reference model, the second and the third are the normal maps obtained by our method and [7], respectively, and the fourth and the fifth are the corresponding error maps. **c** The first two images show the albedo maps obtained by our method and [7], respectively, and the last two are the corresponding error maps
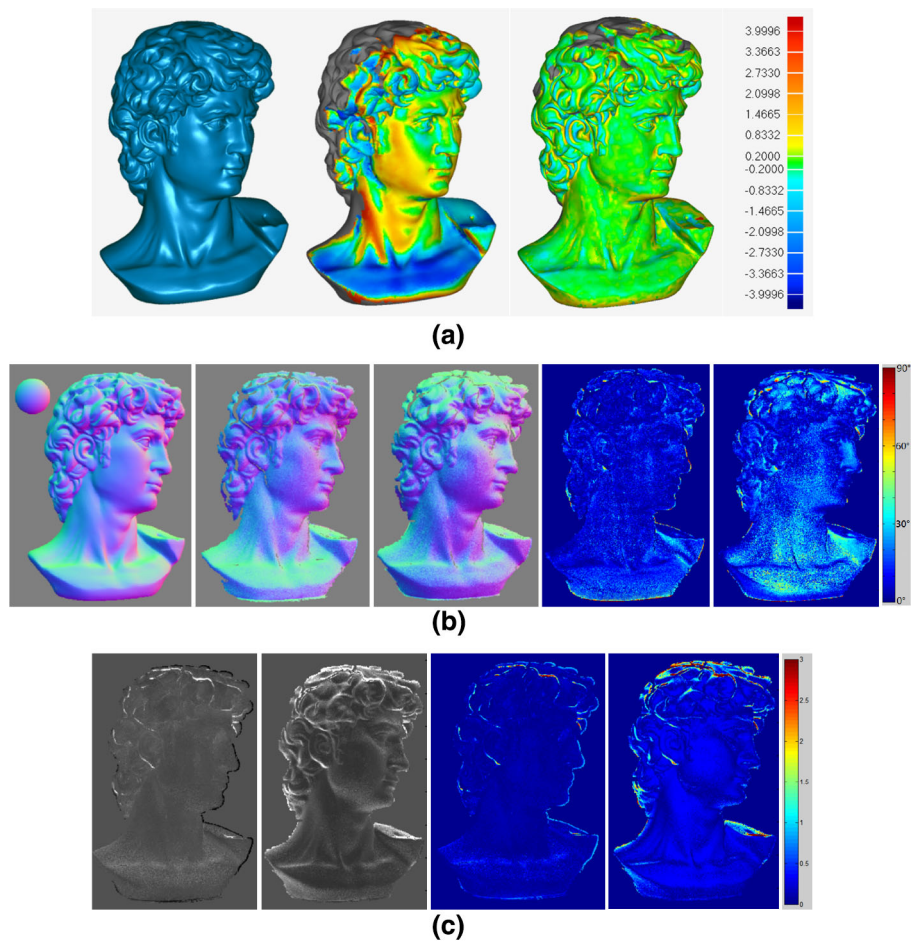

Fig. 11 Comparisons between the manual method [14] and the proposed method. The left is the normal field computed from the smoothed rough depth map, where the low-frequency components are accurate but lack of high-frequency components. The middle and the right are the normal fields computed by the PS method using the lighting parameters from the manual and automatic calibration methods, respectively

frequency components. The middle and the right are the normal fields computed by the PS method using the lighting parameters from the manual and automatic calibration methods, respectively. Visually, the quality of the automatic method is only a little worse than the manual method. The PS method recovers the high-frequency components, but there is a deviation in low-frequency components. So in the fusion process [9], the low-frequency components of the normal field form the depth are used to correct the normals from the PS method.

4 Summary

In this paper, we design a photo-geometric depth camera by combining the near point light source photometric stereo and the speckle-based stereo matching method. The depth camera is compact in structure and suitable for multi-albedo targets. The parameters (including position and intensity) of the light sources can be self-calibrated. To realize the auto-calibration of the point light sources, we firstly use the distant lighting model [7] to estimate the initial surface albedo map.

Then, with the estimated albedo map and the normal vector field fixed, the parameters of the near lighting model are optimized. Next, with the optimized lighting model, we use the NLPS method to re-compute the albedo map and use the method [9] to compute higher-quality depth map. Repeat the above two steps iteratively until convergence or the iteration times reaching the predefined maximum number. Experiments have demonstrated that the depth camera we designed can reconstruct the multi-albedo targets with high fidelity in general indoor environment.

In the current implementation, the images captured by the cameras are transmitted to the computer before processing. In future work, we will design an embedded system based on FPGA in our depth camera to process the images so that the depth maps can be generated by the depth camera directly.

References

1. Blais, F.: Review of 20 years of range sensor development. J. Electron. Imaging **13**(1), 231–243 (2004)
2. Zhang, S., Van Der Weide, D., Oliver, J.: Superfast phase-shifting method for 3-D shape measurement. Opt. Express **18**(9), 9684–9689 (2010)
3. Jiang, C., Bell, T., Zhang, S.: High dynamic range real-time 3D shape measurement. Opt. Express **24**(7), 7337–7346 (2016)
4. Kinect. http://www.xbox.com
5. Newcombe, R.A., Izadi, S., Hilliges, O., Fitzgibbon, A.: KinectFusion: real-time dense surface mapping and tracking. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality. IEEE, pp. 127–136 (2011)
6. Haque, M., Chatterjee, A., Govindu, V. Madhav.: High quality photometric reconstruction using a depth camera. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2275–2282 (2014)
7. Chatterjee, A., Govindu, V.M.: Photometric refinement of depth maps for multi-albedo objects. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 933–941 (2015)
8. Han, Y., Lee, J.Y., Kweon, I.S.: High quality shape from a single RGB-D image under uncalibrated natural illumination. In: Proceedings of IEEE International Conference on Computer Vision. IEEE, pp. 1617–1624 (2013)
9. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3D geometry. ACM Trans. Graph. **24**(3), 536–543 (2005)
10. Quau, Y., Mecca, R., Durou, J.D.: Unbiased photometric stereo for colored surfaces: a variational approach. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 4359–4368 (2016)
11. Higo, T., Matsushita, Y., Joshi, N., Ikeuchi, K.: A hand-held photometric stereo camera for 3-d modeling. In: Proceedings of IEEE International Conference on Computer Vision. IEEE, pp. 1234–1241 (2009)
12. Wang, C., Wang, L., Matsushita, Y.: Binocular photometric stereo acquisition and reconstruction for 3d talking head applications. In: Proceedings of International Speech Communication Association, pp. 2748–2752 (2013)
13. Shi, B., Inose, K., Matsushita, Y., Tan, P., Yeung, S., Ikeuchi, K.: Photometric stereo using internet images. In: Proceedings of International Conference on 3D Vision (3DV) (2014)
14. Powell, M.W., Sarkar, S., Goldgof, D.: A simple strategy for calibrating the geometry of light sources. IEEE Trans. Pattern Anal. Mach. Intell. **23**(9), 1022–1027 (2001)
15. Zhou, W., Kambhamettu, C.: Estimation of illuminant direction and intensity of multiple light sources. In: Proceedings of European Conference on Computer Vision, pp. 206–220 (2002)
16. Ackermann, J., Fuhrmann, S., Goesele, M.: Geometric point light source calibration. In: Vision Modeling and Visualization, pp. 161–168 (2013)
17. Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/
18. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2), 328–341 (2008)
19. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of IEEE International Conference on Computer Vision. IEEE, pp. 839–846 (1998)
20. Mitra, N.J., Nguyen, A.: Estimating surface normals in noisy point cloud data. In: Proceedings of ACM Symposium on Computational Geometry. ACM, pp. 322–328 (2003)
21. Madsen, K., Nielsen, H.B., Tingleff, O.: Methods for non-linear least squares problems, 2nd edn. Informatics and Mathematical Modeling, Technical University of Denmark, Lyngby, Denmark, 24C29 (2004)
22. CHOLMOD. http://www.suitesparse.com
23. Geomagic Qualify. http://www.geomagic.com

Liang Xie is currently working toward the Ph.D. degree at College of Aerospace Science and Engineering, National University of Defense Technology (NUDT), China. He received the B.S. degree and M.S. degree at NUDT in 2012 and 2014, respectively. His research interests include 3D reconstruction and human body motion capture.
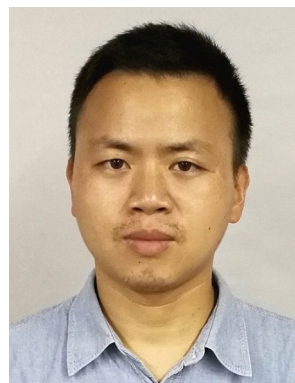
Yuhua Xu is currently a postdoctor at College of Aerospace Science and Engineering, National University of Defense Technology (NUDT). He received the B.S. degree and Ph.D. degree at Hefei University of Technology (HFUT), China, in 2006 and 2011, respectively. In 2014, he was supported by the National Natural Science Foundation of China. His research interest covers 3D reconstruction and computer vision.

Xiaohu Zhang is currently a professor at School of Aeronautics and Astronautics, Sun Yat-sen University. He received the Ph.D. degree at College of Aerospace Science and Engineering, National University of Defense Technology in 2006. His research interest covers videometrics and computer vision.
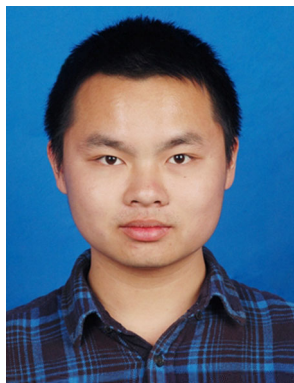
Chenpeng Tong is an engineer at DeepInTech, Shenzhen, China. He received the B.S. degree at Shijiazhuang Tiedao University in 2015. His research interest covers 3D reconstruction and computer vision.

Wei Bao is an Associate Professor at School of Electrical Engineering and Automation, HFUT. He received the B.S. degree, M.S. degree, and Ph.D. degree at HUFT in 2003, 2006, and 2009, respectively. His research interest covers control theory and computer vision.

Boxin Shi is currently an Assistant Professor (1000 Youth Talents Professorship) at Peking University, where he leads the Camera Intelligence Group. Before joining PKU, he did postdoctoral research at MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University, from 2013 to 2016 and worked as a Researcher at the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. He received the B.E. degree from Beijing University of Posts and Telecommunications in 2007, M.E. degree from Peking University in 2010, and Ph.D. degree from the University of Tokyo in 2013. He won the Best Paper Runner-up award at International Conference on Computational Photography 2015.