https://doi.org/10.2352/ISSN.2470-1173.2018.2.VIPC-263

A Robust and Accurate Calibration Method for Out-of-focus Camera

Xiaowei Hu[1], *Guijin Wang*[1], *Jinnan Wang*[1], *Pengfei Sun*[2], *Jingtao Fan*[2], *Feng Chen*[2], *Yiyuan Xie*[3]
[1] *Department of Electronic Engineering, Tsinghua University, China*
[2] *Department of Automation, Tsinghua University, China*
[3] *School of Electronics and Information Engineering, Southwest University, China*

Abstract

Conventional camera calibration methods regard camera as ideal pinhole model and require well-focused images, which can't be satisfied for long-range photogrammetry or low depth-of-field lens. In this paper, we propose a novel active calibration method for out-of-focus camera using LCD monitor. Firstly, we estimate the defocus map by the temporal coded binary-shift patterns, which makes our method more accurate. Secondly, based on the defocus map, we encode LCD pixel's coordinates into phase-shift patterns with optimal frequency and step properties, and then deblur captured patterns. Finally, deblurred patterns are decoded to generate dense phases map to extract accurate feature points coordinates. Our method significantly improves camera calibrations robustness to lens' defocus, noises, glass refraction compared with state-of-art methods. Experimental results demonstrate that our method is superior to conventional methods whether camera is in- or out-of- focus.

1. Introduction

Camera calibration is one critical issue in photogrammetry[1][2], 3D reconstruct[3][4] and robot visual navigation[5], provides transformation relationship between camera's imaging plane and real world three dimensional(3D) objects. Instead of using carefully fabricated 3D reference target with known dimensions, Zhang[6] had greatly simplified camera calibration procedure by placing a planar chessboard in arbitrary orientations and poses. Since then, accurately extracting of feature points' coordinates of reference plane has been deeply studied. Methods such as planar boards marked with circles or concentric rings[7], iterative refinement of the control points algorithm[8] and "virtual defocus(VD)" with windowed polynomial fitting[9] has been proposed and validated. These methods improved camera calibration accuracy obviously.

However, above methods all require captured images are focused well, which can't been guaranteed in practice. For example, short-range optical measurement requires tiny and precise reference target to focus well. On the other hand, long-range vision system such as autonomous driving vision system focus ranges from less than a meter to a few hundred meters, which requires large enough reference target to cover sufficient field of view(FOV) for detecting feature points in various poses. There is a mutual constraint among size, precision and cost of the manufacture of reference targets.

Recently, several calibration methods have been proposed for out-of-focus camera. For instance, Baba et al. [10] determined camera parameters in a single chessboard image by geometry and width of blurred features. This passive method is not robust and accurate. On the contrast, active methods are more convenient and accurate, which refer to those using display device as reference target. The planar liquid crystal display(LCD) monitor is often adopted. It is manufactured to be in tens of nanometers precision by photolithography with known sizes and high flatness. Ha et al. [11] estimated the blur kernel in each location by unidirectional complementary binary patterns and detected feature points' coordinates from 1D deconvolution. Due to the limitation of feature points' detection, it was sensitive to noises and refraction of the glass panel. Bell et al. [12] took advantage of lens defocus effects and used nearly focused squared binary patterns to carry phases. Their method only worked when lens is under the ideal condition of slight, uniform out-of-focus. Wang et al. [13] used three-step phase-shift circular grating(PCG) arrays as calibration patterns, and extracted the PCG's centers by ellipse fitting of 2π-phase points. Their method's accuracy is not comparable with well-focused calibration for the usage of the ellipse fitting. Above all, a robust and accurate calibration method for out-of-focus camera is in urgent need.

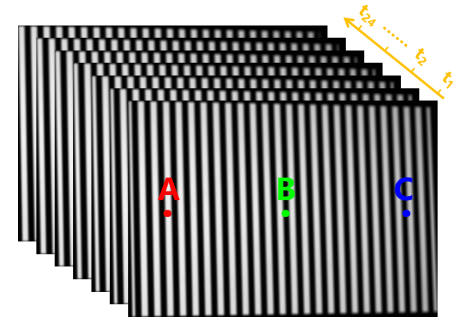
In this paper, we propose a robust and accurate calibration method for out-of-focus camera using LCD monitor. Firstly, we estimate the defocus map by the temporal coded binary-shift patterns, which makes our method more accurate than [12][13]. Secondly, based on the defocus map, we encode LCD pixel's coordinates into phase-shift patterns with optimal frequency and step properties, and deblur captured patterns. Finally, deblurred patterns are decoded to generate dense phase map to extract accurate feature points coordinates. Above steps make camera calibration robust to noises, glass refraction compared with [10][11]. Experimental results demonstrate that our method improves calibration robustness greatly while keeps accuracy the same as the well-focused calibration.

2. Proposed Algorithm

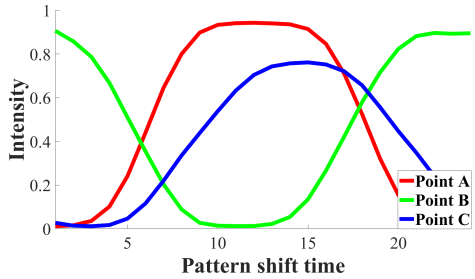
2.1 Calibration Procedure

In this section, we describe the overall calibration procedure as following. We model camera as ideal pinhole imaging model and formulate camera's intrinsic and distortion parameters as described in section 2.2. Then we do following steps for each calibration view:

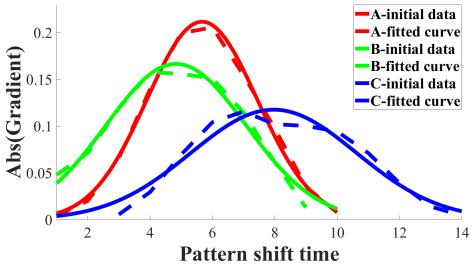
Step 1: We estimate the lens defocus map by the proposed binary-shift patterns firstly. Then we calculate the dense defocus map and its average value $\overline{\sigma}_x$ and $\overline{\sigma}_y$ for following usage. Details are described in section 2.3;



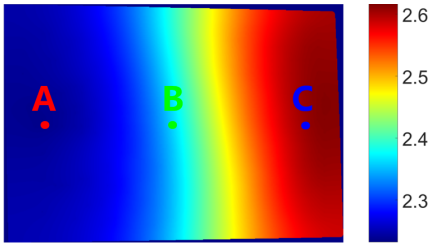
(a) Binary-shift images



(b) Intensity profile



(c) Gaussian fitting



(d) Defocus map

Figure 1. *Example procedure of defocus map estimation. (a)Capture images of binary-shift patterns with period of 24 and 3 sample points A, B and C; (b)Temporal intensity profile of points A(red), B(green) and C(blue); (c)The absolute value of intensity gradient and fitted gaussian function curve of point A, B and C; (d)Estimnated dense defocus map of (a), blur kernel size of points A, B and C equals to 2.21, 2.35 and 2.58 respectively.*

Step 2: We determine the optimal frequency and step parameters of phase-shift patterns based on the defocus map from step1. Then we capture images of encoded phase-shift patterns for step3. Details are described in section 2.4.1 ;
Step 3: We deblur the captured images of phase-shift patterns based on the defocus map estimated from step1 and Wiener filter. Then we decode feature points' coordinates from the deblurred images as described in section 2.4.1 ;
Step 4: We refine feature points' coordinates by finding a local homography $\hat{H}$ based on RANSAC algorithm. The refined feature points' coordinates are calculated by applying $\hat{H}$ again. Details are described in section 2.4.2;
After capturing enough views at various poses and positions, we apply Zhang[4]'s method to calibrate the camera.

2.2 Camera Model

The basic pinhole camera model describes the mapping between the 3D world coordinates and its projection onto the 2D image plane. Let us denote the augmented vector of a 3D point $\boldsymbol{M} = [\boldsymbol{X}\ \boldsymbol{Y}\ \boldsymbol{Z}\ \boldsymbol{1}]$ and its corresponding 2D point $\boldsymbol{m} = [\boldsymbol{x}\ \boldsymbol{y}\ \boldsymbol{1}]$, The mathematical relationship between $\boldsymbol{M}$ and $\boldsymbol{m}$ is given by the following equation:

$$\boldsymbol{sm} = \boldsymbol{K}[\boldsymbol{R}\quad \boldsymbol{t}]\boldsymbol{M}, \boldsymbol{K} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where $\boldsymbol{s}$ is an arbitrary scale factor, camera intrinsic matrix $\boldsymbol{K}$ consists of focal length $[f_u\ f_v]$, principal point$[u_0\ v_0]$ and skewness factor γ of two image axes. The extrinsic camera matrix $[\boldsymbol{R}\quad \boldsymbol{t}]$ is the rotation and translation which relates the world coordinates system to the camera coordinates system.

Camera lens distortion parameters can be modeled as

$$\boldsymbol{D} = [k_1\quad k_2\quad k_3\quad p_1\quad p_2] \tag{2}$$

let $(\boldsymbol{u'}, \boldsymbol{v'})$ be the undistorted coordinates, $\boldsymbol{r} = \sqrt{u^2+v^2}$, then camera radial and tangential lens distortion can be corrected as the following formula:

$$u' = u(1+k_1r^2+k_2r^4+k_3r^6)+2p_1uv+p_2(r^2+2u^2) \tag{3}$$

$$v' = v(1+k_1r^2+k_2r^4+k_3r^6)+2p_1uv+p_2(r^2+2u^2) \tag{4}$$

2.3 Defocus Estimation

2.3.1 Defocus model

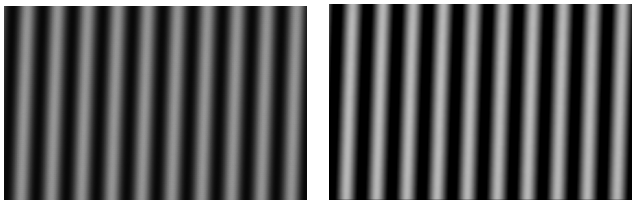
Blurred image can be modeled as a 2D convolution of the initial image with the point spread function(PSF) as

$$\boldsymbol{H} = \boldsymbol{I} \otimes \boldsymbol{G} \tag{5}$$

where $\boldsymbol{I}$ is the initial $m \times n$ well-focused image matrix, $\boldsymbol{G}$ is the point spread function with $r \times r$ blur kernel, and $\boldsymbol{H}$ is the $m \times n$ blurred image. For out-of-focus images, we use a 2D spatially varying, isotropic Gaussian blur kernel

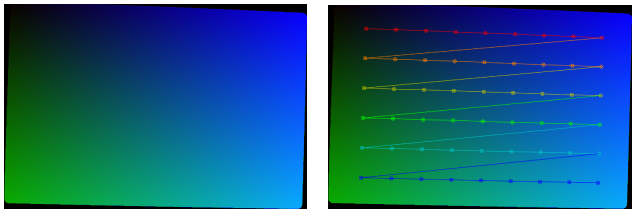
$$G(x_i, y_j) = \frac{1}{2\pi\sigma^2_{x_iy_j}} exp\Big(-\frac{(x_i-x_{i_0})^2+(y_j-y_{j_0})^2}{2\sigma^2_{x_iy_j}}\Big) \tag{6}$$

where $i = 1, \cdots, m, j = 1, \cdots, n$, (x_{i_0}, y_{j_0}) is the focus center, $\sigma_{x_iy_j}$ is the standard deviation of the Gaussian function, which describes the defocus degree of pixel (x_i, y_j). For all pixels (x_i, y_j)



(a) Blurred image (b) Deblurred image

Figure 2. *Example blurred and deblurred phase-shift pattern. (a)One blurred image of phase-shift patterns; (b)Deblurred image of (a) based on defocus map.*



(a) Phase map (b) Feature points

Figure 3. *Example sub-pixel refinement. (a)Decoded dense phase map consists of vertical and horizontal phases; (b)Refined 9×6 feature points.*

displayed on the LCD screen and captured by the camera, we estimate the sparse defocus blur kernel $\sigma_{x_i y_j}$ distribution map by temporal binary-shift patterns, and adopt propagation algorithm to generate the dense defocus kernel distribution map(referred to as defocus map) as section 2.3.2 described.

2.3.2 Defocus map

We estimate the defocus blur kernel size by displaying a series of patterns on a planar liquid crystal display(LCD) device, those patterns are vertical and horizontal binary patterns with 50% duty cycle, wide period and shifted a few LCD pixel per time. This is the work scheme what we called the binary-shift patterns. Fig. 1a is one schematic diagram of blurred images of binary-shift patterns with period equals to 23. The period of the binary-shift patterns should be large enough to avoid being totally blurred caused by the lens defocus. For calibration is not a time-concerned task, the single step of the binary-shift patterns is often set to 1 . To our knowledge, we are the first of adopting binary-shift patterns in the field of out-of-focus camera calibration.

For each view, when camera is well focused, "grid-in-focus" phenomenon may occur, which is caused by finite spatial resolution and LCD dead-zone among adjacent lattice pixels. Therefore, we apply a Gaussian filter with a small standard deviation firstly for each captured image to avoid "grid-in-focus". Then we sample $m_s \times n_s$ points uniformly, model the observed temporal intensity profile as the ideal binary patterns convolved with the Gaussian blur kernel (Fig. 1b). The blur kernel size (σ_{x_i},σ_{y_j}) of pixel (x_i, y_j) is fitted by a 1D Gaussian function on the gradient of temporal intensity profile sampled from the vertical and horizontal binary-shift images separately (Fig. 1c). For simplicity, we use an isotropic Gaussian blur model, and the final σ_{ij} is synthesized as

$$\sigma_{ij} = \frac{\sigma_{x_i}/\eta_{x_i} + \sigma_{y_j}/\eta_{y_j}}{1/\eta_{x_i} + 1/\eta_{y_j}}, \ i = 1, \cdots, m_s, \ j = 1, \cdots, n_s \tag{7}$$

where η_{x_i} and η_{y_i} are the fitting error of each direction. As we have got sparse defocus map, dense defocus map can be generated by propagating the sparse $m_s \times n_s$ sampled points to the entire image(Fig. 1d). Here we solve this problem as [14], then average value $\overline{\sigma}_x$ and $\overline{\sigma}_y$ are calculated based on the dense defocus map for feature detection usage (section 2.4) .

2.4 Feature Detection

2.4.1 Feature point encoding

The main challenge of calibrating out-of-focus camera is how to detect feature points' coordinates accurately. Here we encode feature points' vertical and horizontal coordinates (referred to as phase) by phase-shift algorithms and display encoded patterns on LCD. Phase-shift algorithms are widely used in 3D reconstruct field[15]. For N-step phase-shift fringe patterns with period T, intensity of pixel (x,y) can be described as

$$I_i(x,y) = I_a(x,y) + I_m(cos(\phi(x,y) + 2\pi i/N)) \tag{8}$$

where I_i is the i^{th} step intensity, I_a is the average intensity and I_m is the modulation amplitude, $i = 1, 2, \cdots, N$. Fig. 2a is one blurred phase-shift image, and Fig. 2b is the corresponding deblured image. The phase $\phi(x,y)$ can be solved by

$$\phi(u,v) = tan^{-1}\left[\frac{\sum_{i=1}^{N} I_i sin(2\pi i/N)}{\sum_{i=1}^{N} I_i cos(2\pi i/N)}\right] \tag{9}$$

where $\phi(u,v) \in [-\pi, \pi]$ is called wrapped phase. The absolute phase map without periodicity can be generated by spatial or temporal phase unwrapping method. In this paper, we adopt the absolute phase generated by gray-code patterns as reference to unwrap the wrapped phase $\phi(x,y)$.

In theory, three- or four- step phase-shift patterns with high frequency are robust enough to defocus blur, noise and ambient light. However, due to the reduced brightness contrast caused by lens out-of-focus, the artificially modified gamma curve of LCD and the nonlinear response of camera, decoded feature points' coordinates may produce some unpredictable offset. Therefore, we propose optimizing the period T and the step N of phase-shift patterns specifically to improve performance as following:

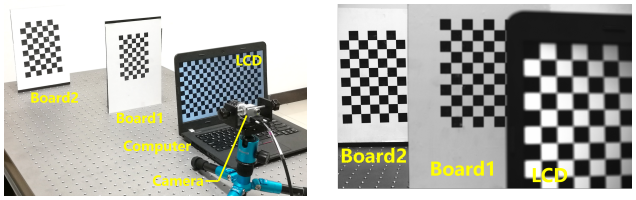
$$\boldsymbol{T_x} \geq s_x \cdot \lambda_x \cdot \overline{\sigma}_x, \ with \ \boldsymbol{N_x} = floor(0.5\boldsymbol{T_x}) \tag{10}$$

$$\boldsymbol{T_y} \geq s_y \cdot \lambda_y \cdot \overline{\sigma}_y, \ with \ \boldsymbol{N_y} = floor(0.5\boldsymbol{T_y}) \tag{11}$$

where s_x and s_y are the scale factor between camera pixel and LCD lattice pixel under certain distance, a value of λ_x and λ_y more than 6 is sufficient empirically, $\overline{\sigma}_x$ and $\overline{\sigma}_y$ are the average value of defocus map calculated by section 2.3, floor(x)=$\lfloor x \rfloor$ means the largest integer value less than or equal to x.

2.4.2 Sub-pixel refinement

After applying decoding algorithm Eq(9), the dense phase map containing feature points' coordinates can be established as



(a) Experimental system (b) Camera view

Figure 4. *Example experimental setup. (a)Experimental scene; (b)Experimental scene under camera view with focusing at Board2 located at a long distance while calibrating at LCD located at a short distance.*

Fig. 2a shown. In estimated phase map, each correspondence which contains integral camera pixel (u_c, v_c) and float LCD pixel (u_l, v_l) can be used to calibrate camera. However, directly applying those correspondence will result in a accuracy reduction, To solve this problem, for only a limited number of points (usually 9×6 or 11×7) are needed, we propose finding a local homography matrix based on RANSAC algorithm[16] instead of using 1D linear or bilinear interpolation to improve robustness and accuracy as following:

$$\hat{H} = \arg\min_{H} \sum_{\forall q \in \mathring{U}} \left\| q - Hp \right\|^2 \tag{12}$$

$$\overline{q} = \hat{H} \cdot \overline{p} \tag{13}$$

here $H \in \Re^{3\times3}$, $p = [u_l, v_l, 1]^T$, $q = [u_c, v_c, 1]^T$, $\mathring{U}$ means the sampled patch. Let $\overline{p}$ be the center of the sampled patch, the refined LCD pixel coordinate $\overline{q}$ in sub-pixel accuracy can be calculated by applying $\hat{H}$ again. Fig. 2b is an example of refined 9×6 feature points array which is in sub-pixel accuracy.

3. Experiments

To validate our proposed method, a camera calibration experimental setup comprising a camera(model: Basler acA1920-155um) with a lens(model: Ricoh FL-CC1614-2M) , a LCD monitor and a computer is shown in Fig. 3a. Camera resolution is 1920×1200 with a pixel size of $5.86\mu m \times 5.86\ \mu m$. LCD monitor resolution is 1366×768 with a pixel size of $234\mu m \times 234\ \mu m$. Lens is a 2/3-inch, 16mm lens with an aperture of F/1.4-F/16.

In order to prove that our proposed method can calibrate out-of-focus camera effectively, we keep camera focusing at a constant distance and realize different defocus effects by changing the distance between LCD monitor and camera while scale up the active areas of the LCD monitor proportionally at the same time. As shown in Fig. 3b, camera focus at reference Board2 located at a long distance, and we calibrate the camera using the LCD monitor located at a short distance with defocus. In this research, we take 4 different defocus effect for camera changing from being well focused to serious defocused. At each distance, we use 9×6 feature points and capture images at 12 different poses. For binary-shift patterns, period is fixed to 36 and single step is fixed to 1. The frequency and step of phase-shift patterns are calculated based on the Eq(10) and Eq(11). For the sake of fairness, we calibrate the camera with this paper proposed method firstly, then we display and capture the chessboard pattern on the monitor with the same pose and position. Table 1 lists the calibration results at four distances.

Table 1. Intrinsic parameters at 4 distances

DS	MD	F_u	F_v	U_0	V_0	DFS(%)
0.15m	Ours	2754.3	2753.1	932.3	588.9	**0.50**
	CB	NW	NW	NW	NW	NW
0.30m	Ours	2754.8	2754.4	934.0	588.7	**0.30**
	CB	2726.3	2724.2	912.0	590.2	5.40
0.45m	Ours	2749.7	2749.3	930.8	591.0	**1.30**
	CB	2735.8	2735.1	931.7	596.1	2.99
0.60m	Ours	2754.0	2753.9	935.3	588.2	0.00
	CB	2752.8	2752.6	941.4	591.9	0.00

In Table 1, DS represents the distance between camera and LCD monitor, here we focus camera at about 0.6m, and calibrate camera at 0.15m, 0.3m, 0.45m, 0.6m respectively. MD represents the method used for calibrating, CB means the method using the chessboard pattern, Ours means the method this paper proposed. NW means not working. We take the result at 0.6m as the real value for each method, and calculate the difference ratio sum(DFS, unit: %) of F_u, F_v, U_0, V_0, the lower DFS is the better. From the table 1, we can find that our method get a far more lower DFS when camera is defocused. The maximal DFS of our method is 1.3% at 0.45m. Our method still works effectively with a 0.5% DFS when calibrating at 0.15m, which means serious defocus. These experimental results clearly demonstrate that our method are more accurate and robust than the chessboard method.

Table 2. Distortion parameters at 4 distances

DS	k_1	k_2	p_1	p_2	MRE(pixels)
0.15m	-0.17	0.25	5.0e-4	2.0e-4	0.041
0.30m	-0.16	0.17	5.5e-4	4.6e-4	0.036
0.45m	-0.15	0.18	6.3e-4	3.9e-4	0.025
0.60m	-0.14	0.17	4.0e-4	5.2e-4	0.021

For camera lens nonlinear distortion, we found taking k_1, k_2, p_1, p_2 is sufficient. Due to the incorrect camera intrinsic parameters estimated by the chessboard method, there is no significance of distortion parameters and mean reprojection errors(MRE, unit: pixels) estimated by the chessboard method. We list the distortion parameters and MRE estimated by our method in Table 2. As listed in Table 2, distortion parameters calibrated at different distance are similar to each other, the maximal MRE is 0.041 pixels, which is relatively small and close to well-focused calibration's 0.021 pixels. These results demonstrate that our method is more robust and accurate compared with the traditional methods.

4. Conclusions

In this paper, we propose a robust and accurate calibration method for out-of-focus camera using LCD monitor. Firstly, we estimate the defocus map by the temporal coded binary-shift patterns, which makes our method more accurate. Secondly, based on the defocus map, we encode LCD pixel's coordinates into phase-shift patterns with optimal frequency and step properties, and deblur captured patterns. Finally, patterns are decoded to generate dense phase map to extract accurate feature points' coordinates. Experimental results demonstrate that our proposed method can calibrate camera robustly and accurately whether camera is in- or out-of- focus. Our proposed method can be used

for short- or long- range vision system in which traditional methods are less effective.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (No. 61327902), Tsinghua University Initiative Scientific Research Program (2016SZ0306), Tsinghua University under Grant 20161080084 and National High tech Research and Development Plan under Grant 2015AA042306.

References

[1] Thomas Luhmann, Stuart Robson, SA Kyle, and IA Harley, *Close range photogrammetry: principles, techniques and applications*, Whittles, 2006.

[2] Chenbo Shi, Guijin Wang, Xuanwu Yin, Xiaokang Pei, Bei He, and Xinggang Lin, "High-accuracy stereo matching based on adaptive ground control points," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1412–1423, 2015.

[3] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern recognition*, vol. 43, no. 8, pp. 2666–2680, 2010.

[4] Guijin Wang, Xuanwu Yin, Xiaokang Pei, and Chenbo Shi, "Depth estimation for speckle projection system using progressive reliable points growing matching," *Applied optics*, vol. 52, no. 3, pp. 516–524, 2013.

[5] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver, "Visual navigation for mobile robots: A survey," *Journal of intelligent and robotic systems*, vol. 53, no. 3, pp. 263, 2008.

[6] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[7] Jun-Sik Kim, Ho-Won Kim, and In So Kweon, "A camera calibration method using concentric circles for vision applications," *ACCV2002, Melbourne, Australia*, 2002.

[8] Ankur Datta, Jun-Sik Kim, and Takeo Kanade, "Accurate camera calibration using iterative refinement of control points," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1201–1208.

[9] Lei Huang, Qican Zhang, and Anand Asundi, "Camera calibration with active phase target: improvement on feature detection and optimization," *Optics letters*, vol. 38, no. 9, pp. 1446–1448, 2013.

[10] Masashi Baba, Masayuki Mukunoki, and Naoki Asada, "A unified camera calibration using geometry and blur of feature points," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. IEEE, 2006, vol. 1, pp. 816–819.

[11] Hyowon Ha, Yunsu Bok, Kyungdon Joo, Jiyoung Jung, and In So Kweon, "Accurate camera calibration robust to defocus using a smartphone," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 828–836.

[12] Tyler Bell, Jing Xu, and Song Zhang, "Method for out-of-focus camera calibration," *Applied optics*, vol. 55, no. 9, pp. 2346–2352, 2016.

[13] Yuwei Wang, Xiangcheng Chen, Jiayuan Tao, Keyi Wang, and Mengchao Ma, "Accurate feature detection for out-of-focus camera calibration," *Applied optics*, vol. 55, no. 28, pp. 7964–7971, 2016.

[14] Shaojie Zhuo and Terence Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011.

[15] Tomislav Pribanić, Saša Mrvoš, and Joaquim Salvi, "Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning," *Image and Vision Computing*, vol. 28, no. 8, pp. 1255–1266, 2010.

[16] Daniel Moreno and Gabriel Taubin, "Simple, accurate, and robust projector-camera calibration," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, 2012, pp. 464–471.

Author Biography

Xiaowei Hu *received his B.S. degree (with honor) from University of Electronic Science and Technology of China, Chengdu, China, in 2015. He is currently pursuing his Ph.D. degree with Tsinghua University, Beijing, China. His research interests include depth sensing ,3D reconstruction, and computational imaging.*

Guijin Wang *received his B.S. and Ph.D. degrees (with honor) from Tsinghua University, China, in 1998 and 2003 respectively, all in electronic engineering. Currently he is an associate professor in the Department of Electronic Engineering, Tsinghua University, China. His research interests focus on wireless multimedia, depth sensing, pose recognition, intelligent human-machine UI.*

Jinnan Wang *received his B.S. in Electronic and Information Engineering at the Changchun University of Science and Technology (2014). He is pursuing his M.E. degree at the Visual Computing Lab in the Department of Electronic Engineer at Tsinghua University. His research interests centers on the defect detection and SLAM.*

Pengfei Sun *received his B.S. degreee in automation from Northeastern Universtiy, Shenyang, China, in 2016. He is currently a Ph.D. student in Tsinghua Universtiy, Beijing, China. His current research interests include computer vision and machine learning.*

Jingtao Fan *received the B.E. and M.E. degrees in computer science and technology and the Ph.D. degree in optical engineering from the Changchun University of Science and Technology, Changchun, China, in 2003, 2007, and 2013 respectively. He currently is a assistant professor with Tsinghua University, Beijing, China. His current research interests include 3D video processing and computer vision.*

Feng Chen *received his B.S. and M.S. degrees in automation from SaintPetersburg Polytechnic University, Saint Petersburg, Russia,in 1994 and 1996, respectively, and the Ph.D. degree from the Automation Department, Tsinghua University, Beijing, China, in 2000. He is currently a Professor with Tsinghua University. His current research interests include computer vision, brain-inspired computing, and inference in graphical models.*

Yiyuan Xie *received the B.S. degree from Southwest Jiaotong University, China, in 2003, and the Ph.D. degree in optical engineering from the Chinese Academy of Sciences in 2009. He is currently a Professor with Southwest University, Chongqing, China. His current research interests include optical networks-on-chip, ultrahigh optical communications, optical data centers, and surface plasmon polaritons.*