Accurate 3D reconstruction using a Multi-phase ToF Camera

Ding Liang, Yebin Liu, Wei Yan, Chenggang Yan and Qionghai Dai

Automation Department, Tsinghua University, Beijing, China

ABSTRACT

The depth quality of a time-of-flight (ToF) camera is influenced by many systematic and non-systematic errors[1] . In this paper we present a simple method to correct and reduce these errors and propose a multi-phase approach to improve the depth acquisition accuracy. Compared with traditional calibration methods, we take the position of light source into account, and calibrate the light source together with the camera to reduce depth distortion. To ameliorate the sensor errors caused in the manufacturing process, a Look-up Table (LUT) is used to correct pixel-related errors. Besides, we capture images with multiple phases and apply FFT to get the true depth. By the proposed approach, we are able to reconstruct an accurate 3D model with RMSE of the measured depth belowing 1.2mm.

Keywords: Time of flight camera, TOF, multi-phases, camera calibration, high-accuracy 3D-reconstruction

1. INTRODUCTION

Time-of-Flight(ToF) camera is a widely used depth sensor, which is popular for its high framerate, small size and high maneuverability. However ToF camera is not suitable for getting high quality depth image because of its random noises and systematic errors. To increase the signal-to-noise ration (SNR), many methods have been proposed, such as low amplitude filtering[2] , accuracy thresholding based on pixel variance[3] and photometric stereo.[4] Though noises are truly decreased, depth image is smoothed at the same time. Our method is based on the analysis of the origins of systematic and non-systematic errors[1] . Traditional camera calibration methods based on intensity image are applied in[5–7] . 25 infra-red LEDs are utilized in[8] to assist high accuracy calibration. All of these methods consider only the geometric calibration of the camera sensor itself. We try to model the object together with the camera sensor and the light source to reduce the errors induced by inappropriate approximation as much as possible. Besides, we explore how the number of phases of the modulation light source relates with the depth sensing performance and propose a method to improve the depth resolution by using multiple modulation phases under an as high as possible modulation frequency. The source of systematic and non-systematic errors is introduced in section 2. Then we will present our method in detail in section 3. Section 4 gives results of our method with different parameters, following with conclusion in section 5.

2. SOURCE OF ERRORS

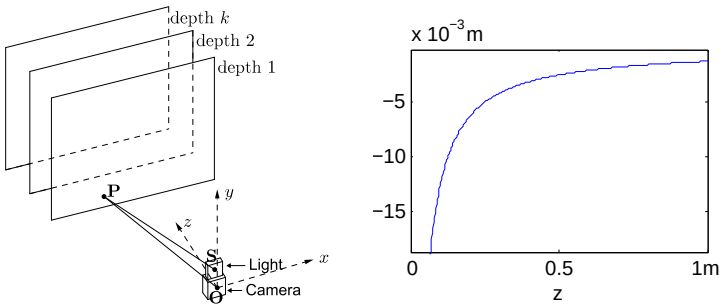
This section analysis some of the errors and noises in a time-of-flight depth sensing system. We consider the geometry problem of the light source and the sensor, the fly time response error due to the manufacturing process, as well as the random noises.

2.1 Geometry of the Light Source and the Sensor

One of the biggest differences between ToF cameras and traditional RGB color cameras is that ToF cameras use their own light sources. In other words, the light source is part of a ToF system. As such, the position of light source has to be considered when designing and using this system. Available methods usually treat the center of light source as the center of the camera sensor, which is reluctantly acceptable when these two centers are near enough and accurate reconstruction is not required. However, if high quality depth image is desired, the position of the camera and the light source should be calibrated before hand.

Moreover, different pixels have different light travel time for the same depth value plane. This is extremely essential when the field of view is large, see Fig. 2a. This problem can be solved by a calibration method to get the intrinsic parameters. Traditional methods to calibrate the ToF camera are mostly like calibration of color

Optoelectronic Imaging and Multimedia Technology III, edited by Qionghai Dai, Tsutomu Shimura,
Proc. of SPIE Vol. 9273, 92733J · © 2014 SPIE · CCC code: 0277-786X/14/$18
doi: 10.1117/12.2071837



(a) (b)

Figure 1: Figure 1a shows our set of camera and light. We use planes at different depths to assist calibrating the position of the light source. Light flies along $\boldsymbol{S} \to \boldsymbol{P} \to \boldsymbol{O}$, from the light source to the camera sensor, reflected by object which is a plane here. Figure 1b illustrates the systematic error cost by the light source. $z = tc$ stands for the light travel time. Here we let $s_x = 0, s_y = 0.05, s_z = 0, \alpha = 0, \beta = 0$.

cameras using a check-board. We propose in this paper a calibration method that jointly calibrate the light source and the camera.

The fly time captured by the ToF camera can be formulated as

$$t = \left(\|\boldsymbol{S} - \boldsymbol{P}\|_2 + \|\boldsymbol{P}\|_2\right)/c \tag{1}$$

where $\boldsymbol{S}, \boldsymbol{P}$ is the position of the light source and the object point, c is the constant light speed. The equation $t = 2\|\boldsymbol{P}\|_2/c$, which is used in all applications, holds only when $\boldsymbol{S}$ is extremly near the camera (The camera center is set to be the origin). Here we assume $\boldsymbol{S} = (s_x, s_y, s_z)^T$, $\boldsymbol{P} = z\,(\alpha, \beta, 1)^T$, where α, β are parameters obtained by calibration. Given the pixel position (i, j), α and β can be caculated by the camera intrinsic parameters using the following camera projection model:

$$\begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \Rightarrow \begin{cases} \alpha = \frac{X}{Z} = \frac{x - c_x}{f_x} \\ \beta = \frac{Y}{Z} = \frac{y - c_y}{f_y}. \end{cases} \tag{2}$$

and

$$\|\boldsymbol{P}_{true}\|_2 - \|\boldsymbol{P}_{false}\|_2 = \left(s_x^2 + s_y^2 + s_z^2 - (tc)^2\right) \frac{\sqrt{\alpha^2 + \beta^2 + 1}}{2\left(\alpha s_x + \beta s_y + s_z - tc\sqrt{\alpha^2 + \beta^2 + 1}\right)} - \frac{tc}{2}$$

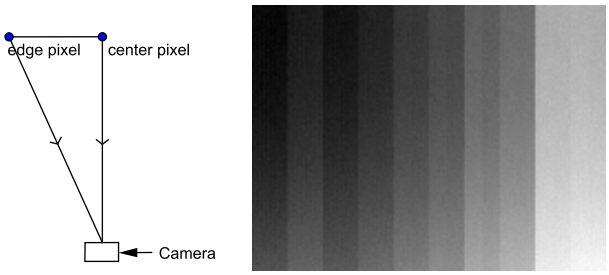
The plot of this kind of error is illustrated in Fig. 1b. Our proposed method for geometric calibration of the light source and the sensor is given in section 3.1.

2.2 Manufacturing Errors

Another type of significant systematic errors come from the manufacturing error of the ToF sensor, mainly due to different material properties in CMOS-gates. Theoretically, all the pixels should response the same phase difference for a scene point with the same depth. However, the manufacturing error causes different phase responses (phase differences correspond to depth value) between pixels. For the TOF sensor we used, such error causes 10 discrete depth partitions as shown in Fig. 2b. We therefore divide the image into 10 regions according to the measured partitions and assume that pixels in the same region share the same properties, though in fact they changes gradually from top to bottom.

2.3 Random Noises

Random noises are non-systematic errors, which are hard to be removed through sensor calibration. In this paper, we propose a multi-phase method working under a high modulation frequency to decrease these errors, see section 3.3 for detail.



(a) (b)

Figure 2: Edge pixels and center pixels of the same depth cause different travel distances (Fig. 2a). Figure 2b shows the fly time response errors causing 10 discrete depth partitions.

3. SOLUTIONS

This section presents three key techniques aiming to solve the three corresponding challenges described in the section 2.

3.1 Light Source and Sensor Plane Calibration

Here we propose a novel method to compensate for the errors mentioned in section 2.1 by calibrating both the ToF sensor and the light source. Directly measuring the position of the light source is inaccurate, since the light souce is composed of multiple laser diodes, which locate some distance apart from each other. We therefore regard the multiple light sources as a virtual light source, and propose to use a diffuse plane to assist estimating the light source position.

We set our plane at k different depths roughly in parallel with the sensor plane, then

$$\begin{cases} D_k & = \|\boldsymbol{S} - \boldsymbol{P}_k\|_2 + \|\boldsymbol{P}_k\|_2 \\ \boldsymbol{n}^T \cdot \boldsymbol{P}_k & = d_k \end{cases} \tag{3}$$

where $\boldsymbol{n}$ is the normal of the plane to be estimated together with $\boldsymbol{S}$, and D_k is the distance image we measured at kth depth d_k. Even though we try to set the plane normal perpendicular to the sensor plane, the exact normals $\boldsymbol{n}$ of the diffuse planes are still unknown and need to be estimated. Since we use a slider to move the plane at different regular locations, it is reasonable to assume that the normal $\boldsymbol{n}$ keeps constant for all depth d_k.

Besides light source calibration, the ToF sensor itself should also be geometrically calibrated. Using an available traditional camera calibration method*, we are able to map every pixels at coordinates (i, j) in the image to their world coordinates given the depth values z by

$$\boldsymbol{P}(i, j) = z\boldsymbol{\eta}_{ij} \tag{4}$$

where $\boldsymbol{\eta}_{ij} = (\alpha_{ij}, \beta_{ij}, 1)^T$ is the calculated camera parameters. Using all the pixels at i, j on different depth d_k, we get hundreds of thousands points $\boldsymbol{P}$ and are able to estimate $\boldsymbol{n}, \boldsymbol{S}$ by a non-linear regression method.[9] Once we know the exact $\boldsymbol{S}$, the global corrdinates (x_{ij}, y_{ij}, z_{ij}) of any 3D point in the scene can be computed by

$$\begin{cases} z_{ij} & = \dfrac{D_{ij}^2 - \|\boldsymbol{S}\|_2^2}{2\left(D_{ij}\left\|\boldsymbol{\eta}_{ij}\right\|_2 - \boldsymbol{\eta}_{ij}^T \cdot \boldsymbol{S}\right)} \\ x_{ij} & = z_{ij}\alpha_{ij} \\ y_{ij} & = z_{ij}\beta_{ij} \end{cases} \tag{5}$$

where D_{ij} is the distance value on the captured distance image D of a real scene. Equation 5 solves the problem of light source calibration and the rectification of non-uniform travel time on the sensor plane.

*http://www.vision.caltech.edu/bouguetj/calib_doc/

3.2 Rectifying Manufacturing Errors

The manufacturing error described in section 2.2 can be compensated by a Look-up Table (LUT), as proposed by Kahlmann et al.[8] Given a depth value D_{ij} of a pixel (i, j), the LUT returns an offset d_{ij} for it and the accurate depth value can be obtained by adding this offset to D_{ij}. Different from previous approaches, we measure the LUT by setting the light source facing direct to the TOF camera. Under such a setting, the light directly comes into the camera without reflection, so that the LUT is not influenced by any other systematic errors. It should also be noted that the LUT depends on frequencies and should be captured under the same frequency.

3.3 Multiple-Phase TOF Depth Sensing

ToF camera uses a modulated light source to light the scene, and get the fly time of light by integrating the received signal and the reference signal at 4 different phases. This method is called four-bucket sampling.[1] Mathematically, the measured image signal $H(\phi)$ of a paricular phase ϕ can be formulated as

$$H(\phi) = \int_0^{NT} S_e(\tau - \phi) S_r(\tau) \mathrm{d}\tau, \tag{6}$$

where S_e, S_r stands for the reference and the recieved signal, ϕ is the controllable phase, and NT stands for N periods. Most of the ToF cameras use a sine wave to modulate light, i.e., $S_e(t) = A_1 \sin(\omega t)$ and $S_r(t) = A_2 \sin(\omega(t + \Delta t)) + B$, then considering the random noise in the imaging process, the measured image signal can be written as

$$H(\phi) = A \cos(2\pi f \Delta t + \phi) + \delta \tag{7}$$

where δ is the random noise and f is the frequency of the light. By changing ϕ 4 times at equal intervals, The noise can be roughly removed and D and Δt can be caculated by

$$D = c\Delta t = \frac{c}{2\pi f} \arctan\left(\frac{H(270°) - H(90°)}{H(0°) - H(180°)}\right) \tag{8}$$

Practically, due to electronics limitations and artifacts, the modulation signal is usually periodic but not sinusoidal. For example, square wave, pulse wave and coded wave[10] could be existed. Besides, waveforms may vary with the changing of frequencies.

To address this issue, we adopt the Fourier analysis on the time-of-flight signals[11] to expand the measured signal into Fourier series,

$$H(\phi) = \sum_{n=1}^{\infty} A_n \cos(2\pi n f \Delta t + n\phi). \tag{9}$$

In this way, Δt can be calculated by fetching the phase in the first term of the Fourier transform on $H(\phi)$ as[11]

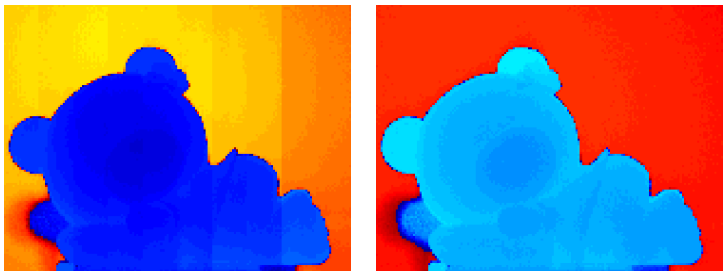
$$\Delta t = \frac{\mathcal{P}(\mathcal{F}_N(H(\phi)))}{2\pi f} \tag{10}$$

where N is the number of ϕ used for Fourier transform $\mathcal{F}$, $\mathcal{P}$ retrieval the first phase term. Compared with equation 8, multiple phases remove the random noises in a more controllable way, with improved results demonstrated in section 4.

4. RESULTS

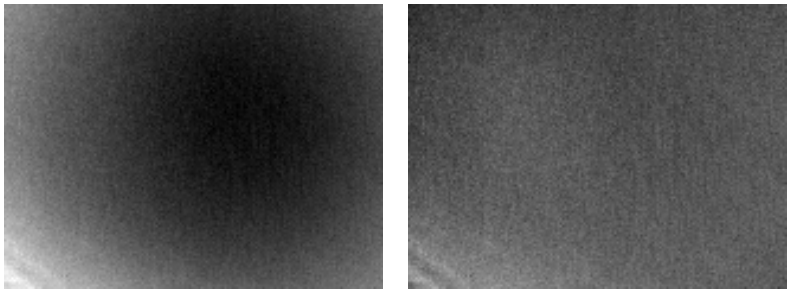
We utilize a modified PMD-nano camera used in Heide et al.[12] and Lin et al.[11] The modified camera is equipped with 12 laser diodes working at wavelength of 640nm. These lasers can produce modulated square waves with frequency range in 0∼180Mhz.

Fig. 3 shows the importance of rectification of the manufacturing errors using the LUT. Before correction, the captured scene depth is segmented into pieces. Only with this rectification, the right depth map can be obtained. Fig. 4 compares the captured depth signal of a plane with (right) and without (left) our proposed



(a) (b)

Figure 3: Depth image before (3a) and after (3b) pixel-related manufacturing error rectification.



(a) (b)

Figure 4: These two depth images present the difference between depth maps of a plane with and without light source and sensor plance calibration (section 3.1). It could be seen in (4a) that the minimum depth is not exactly the center of the image because of the position of light source.

light source and camera plane calibration. Based on the estimated position of light source, we are able to get a right depth map without distortion.

The influence of multiple phases is illustrated in Fig.5. Under working frequency $100MHz$, the quality of depth measurement substantially improves with the increasing of the number of phases used. However, the quality saturated when the number of phase reaches 32. In this case, we are able to get a depth map of a plane with $RMSE < 1.2mm$, which is far smaller than existing methods. Fig.6 illustrates how the number of the phases and the selection of the frequency effect the depth quality. we capture a plane under different number of phases and under different frequencies. Though higher frequencies lead to higher depth accuracy, as shown in Fig. 6a. when $f < 120MHz$, amplitudes of the signal shrunk with the increasing of modulation frequency,

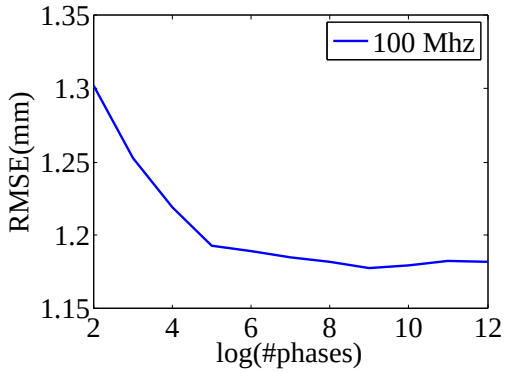
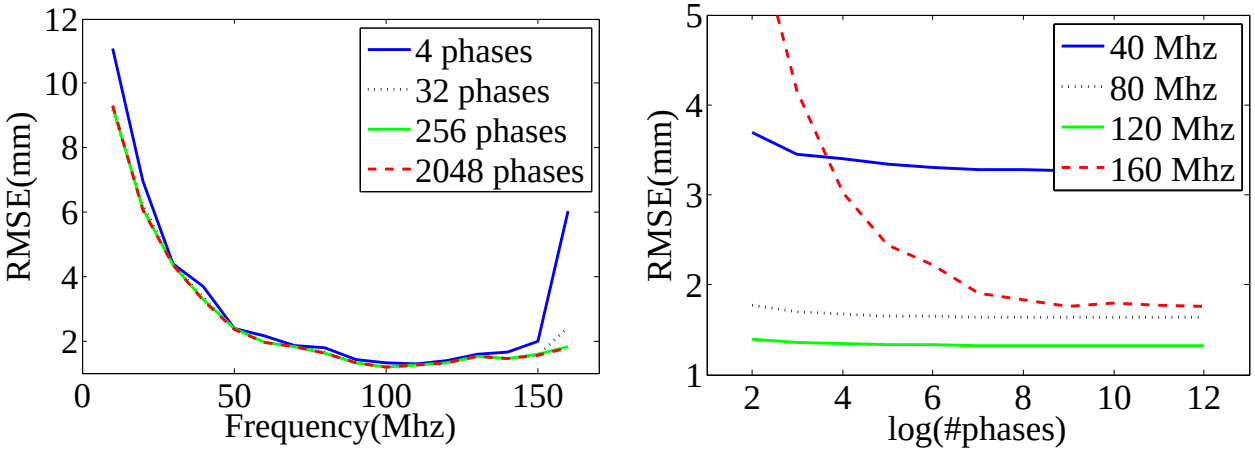


Figure 5: RMSE under different numbers of phases. The more phases, the better. $2^5 = 32$ phases are enough to get nearly the best performence.



(a) impact of frequencies

(b) impact of number of phases

Figure 6: RMSE drops dramatically with frequency raises in the normal working range $[0, 120MHz]$, but bounces up when frequency is extremly high. In contrast, the improvement of multiple phases is not that obvious. These results are calculated of a plane at 80cm. Similar results can be obtained for other depths.

i.e. $A_n \propto \frac{1}{f}$, and thereby, RMSE increases when the working frequency is too high ($f > 120MHz$). One way to conquer this is to use more powerful light source. Moreover, it can be seen from Fig. 6b. that in the case of low signal SNR under the extremely high working frequency (e.g., $160MHz$), the multi-phase approach substantially increases the depth accuracy. Fig.7 shows the 3D mesh reconstruction of a toy bear.

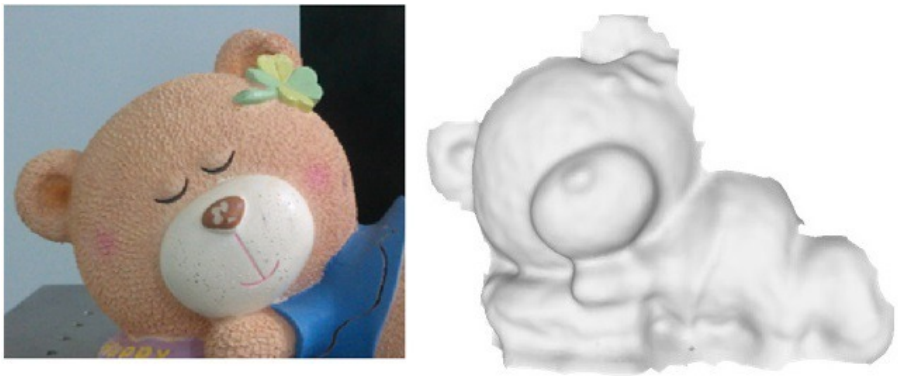


Figure 7: A 3d model of a toy. The nose can be clearly distinguished from its big mouth.

5. CONCLUSION

This paper proposes a novel method for ToF camera calibration, taking the position of light source, non-uniform sensor plane response, manufacturing errors, and random noises into account. With the proposed method to remove all these systematic and non-systematic errors in a TOF system, an accurate 3D point cloud can be obtained. It should be noted that the problems of multi-path[13,14] can not be solved through this method and shall be conquered in the future.

6. ACKNOWLEDGMENTS

This work was supported by the 863 Program (No.2013AA01A604) and The National key foundation for exploring scientific instrument No. 2013YQ140517.

REFERENCES

1. Foix, S., Alenya, G., and Torras, C., "Lock-in time-of-flight (tof) cameras: a survey," *Sensors Journal, IEEE* **11**(9), 1917–1926 (2011).
2. Wiedemann, M., Sauer, M., Driewer, F., and Schilling, K., "Analysis and characterization of the pmd camera for application in mobile robotics," in [*Proceedings of the 17th IFAC World Congress*], 6–11 (2008).
3. Falie, D. and Buzuloiu, V., "Noise characteristics of 3d time-of-flight cameras," in [*Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on*], **1**, 1–4, IEEE (2007).
4. Kim, S. K., Kang, B., Heo, J., Jung, S.-W., and Choi, O., "Photometric stereo-based single time-of-flight camera," *Optics letters* **39**(1), 166–169 (2014).
5. Fuchs, S. and May, S., "Calibration and registration for precise surface reconstruction with time-of-flight cameras," *International Journal of Intelligent Systems Technologies and Applications* **5**(3), 274–284 (2008).
6. Lindner, M. and Kolb, A., "Lateral and depth calibration of pmd-distance sensors," in [*Advances in Visual Computing*], 524–533, Springer (2006).
7. Lindner, M., Kolb, A., and Hartmann, K., "Data-fusion of pmd-based distance-information and high-resolution rgb-images," in [*Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on*], **1**, 1–4, IEEE (2007).
8. Kahlmann, T., Remondino, F., and Ingensand, H., "Calibration for increased accuracy of the range imaging camera swissrangertm," *Image Engineering and Vision Metrology (IEVM)* **36**(3), 136–141 (2006).
9. Seber, G. and Wild, C., "Nonlinear regression," (2003).
10. Kadambi, A., Whyte, R., Bhandari, A., Streeter, L., Barsi, C., Dorrington, A., and Raskar, R., "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles," *ACM Transactions on Graphics (TOG)* **32**(6), 167 (2013).
11. Lin, J., Liu, Y., Matthias, H., and Dai, Q., "Fourier analysis on transient imaging by multifrequency time-of-flight camera," in [*Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*], (2014).
12. Heide, F., Hullin, M. B., Gregson, J., and Heidrich, W., "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.* **32**(4), 45 (2013).
13. Fuchs, S., Suppa, M., and Hellwich, O., "Compensation for multipath in tof camera measurements supported by photometric calibration and environment integration," in [*Computer Vision Systems*], *Lecture Notes in Computer Science* **7963**, 31–41, Springer Berlin Heidelberg (2013).
14. Fuchs, S., "Multipath interference compensation in time-of-flight camera images," in [*Pattern Recognition (ICPR), 2010 20th International Conference on*], 3583–3586, IEEE (2010).