21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Easy Depth Sensor Calibration

Hirotake Yamazoe, Hitoshi Habe, Ikuhisa Mitsugami, and Yasushi Yagi
ISIR, Osaka University, Japan
{*yamazoe,habe,mitsugami,yagi*}*@am.sanken.osaka-u.ac.jp*

Abstract

This paper proposes a depth measurement error model of consumer depth cameras such as Microsoft KINECT, and its calibration method. These devices are originally designed for video game interface, thus, the obtained depth map are not enough accurate for 3D measurement. To decrease these depth errors, several models have been proposed, however, these models consider only camera-related parameters. Since the depth sensors are based on projector-camera systems, we should consider projector-related parameters. Therefore, we propose the error model of the consumer depth cameras especially the KINECT, considering both intrinsic parameters of the camera and the projector. To calibrate the error model, we also propose the parameter estimation method by only showing a planar board to the depth sensors. Our error model and its calibration are necessary step for using the KINECT as a 3D measuring device. Experimental results show the validity and effectiveness of the error model and its calibration.

1 Introduction

The consumer depth camera such as the Microsoft KINECT[1] has opened a new vista to computer vision and pattern recognition researchers. Its ability to capture both a depth map and a color image simultaneously and recent depth analysis techniques have stimulated a wide variety of applications.

When applying them to certain applications, technical problems of accuracy in depth measurement occasionally occurs. For example, when a depth camera is used to capture a planar object, noticeable artifacts are observed. This is particularly crucial when registering multiple depth maps captured from different depth cameras.

To tackle this issue, this paper proposes a simple but effective method for calibrating a depth sensor. Our method makes use of a parameterized noise model based on the underlying acquisition process of a depth map. To derive the parameters of the model, we use a planar calibration board similar to the color camera calibration widely used in the community[10]. This model and an optimization process ensure the calibration is effective and efficient.

Because most of the consumer depth sensors were originally designed as an interface for a video game, it is not necessary to capture an accurate depth map. For example, a human pose estimation using a single depth map[8] can provide sufficient results even when the depth map is not particularly accurate. In contrast, the proposed calibration enables us to use the depth sensors as a accurate sensing device. We believe this makes it possible to use the sensors in a wider range of applications,which is the primary contribution of this paper.

In this paper, Section 2 discusses the details of the technical problem we are focusing on by following the proposed error model in Section 3 and its parameter estimation in Section 4. Section 5 shows the experimental results that prove the effectiveness of the proposed method and Section 6 summarizes the paper.

2 Depth Sensor Calibration

Depth measurement has been an active topic among computer vision researchers and various methods have been proposed. Typical examples include a traditional (passive) stereo camera[7] that uses point correspondence between two images, and a ToF (Time of Flight) sensor[6] that measures the time for light to travel from a source to an object and back to a sensor.

Recent consumer depth sensors are also based on triangulation similar to classical stereo cameras. They replace one camera with a projector that casts a structured pattern onto the surface of a target object. This enables us to find corresponding points between the coordinates of a camera and a projector in a robust manner. Although this robustness and its low cost have had a great impact on computer vision, the resultant depth maps do have some errors as shown in the subsequent section. We need to calibrate these errors to exploit the sensors for acquiring accurate depth maps. This is the prime technical issue we tackle in this paper.

To solve the problem, Herrera et al. proposes a depth correction method for the KINECT[3]. However, since their method employs a nonparametric model for the depth correction, they need to estimate a large number of parameters. In addition, their optimization scheme requires an additional high-resolution color camera.

Instead, our method first introduces a parametric model to represent the errors occurring in a resultant depth map. Because this model is comprised of simple models representing respective processes in depth measurement, the model can be compact and easily optimized. We also propose a simple parameter estimation method. In a similar manner to the common color camera calibration method[10], once we show a planar calibration board in front of a depth camera and capture a group of images, our method efficiently optimize a set of parameter.

In this paper, we confine our attention to the calibration of the depth map. Needless to say, combining the proposed method with other techniques such as external calibration and surface refinement enables us to capture accurate and precise geometric information of a target object. We believe this work is an important step toward this goal.

We have to note that the calibration method introduced in subsequent sections is designed for the KINECT because it is the most common depth sensor. However, we can say our calibration method has generality because it is based on the principle common to the other depth sensors.

3 Depth Error Model

3.1 Depth Measurement of KINECT

Before describing our proposed error model, we outline the depth measurement of the KINECT, the depth sensor of which consists of an IR camera and an IR projector. The IR projecter projects special fixed patterns (so called speckle pattern) on the targets, and the IR camera observes the projected pattern. By comparing the observed patterns and the reference patterns captured in advance, the KINECT estimates the depth information of the targets. The reference patterns are the observation results of the IR camera when the IR projector casts the pattern on the reference plane Π_0 [4] (Fig. 1).

Here, we assume a pattern $P(\boldsymbol{x}_i^{(p)})$ is projected in the direction of point $\boldsymbol{x}_i^{(p)}$ onto the reference plane Π_0, and the pattern $P(\boldsymbol{x}_i^{(p)})$ on the Π_0 is projected on the 2D position $\boldsymbol{x}_{0,i}^{(c)}$ at the IR camera. We obtain the following

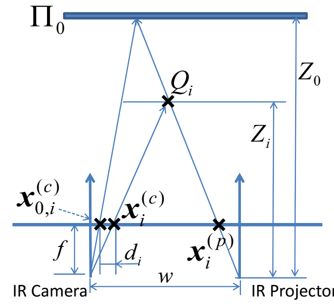

Figure 1. Depth measurement of KINECT

relationship:

$$\boldsymbol{x}_{0,i}^{(c)} = \boldsymbol{x}_i^{(p)} + f \cdot w / Z_0 \tag{1}$$

where w is the baseline distance between the camera and the projector, f is the focal length of the IR camera (and the IR projector), and Z_0 is the distance between the reference plane Π_0 and the KINECT.

Next, we consider target observation measured at point Q_i, and assume a pattern $P(\boldsymbol{x}_i^{(p)})$ is observed at $\boldsymbol{x}_i^{(c)}$ using the IR camera. By referring to the reference patterns, $\boldsymbol{x}_{0,i}^{(c)}$, the observed position of the pattern when the pattern is projected onto the reference plane Π_0 can be obtained. Thus we can calculate the disparity d_i from the reference plane observation at $\boldsymbol{x}_i^{(c)}$ as follows:

$$d_i = \boldsymbol{x}_i^{(c)} - \boldsymbol{x}_{0,i}^{(c)} = \boldsymbol{x}_i^{(c)} - \boldsymbol{x}_i^{(p)} - f \cdot w / Z_0 \tag{2}$$

Then, $\boldsymbol{X}_i$, the 3D positions of the point Q_i can be calculated as

$$\boldsymbol{X}_i = [\frac{x_i^{(c)} \cdot Z_i}{f}, \frac{y_i^{(c)} \cdot Z_i}{f}, \frac{f \cdot w}{f \cdot w / Z_0 + d_i}]^t \tag{3}$$

However, in the KINECT, we cannot directly obtain such disparity values and can only obtain normalized depth values d'_i from 0 to 2047 (where $d_i = m \cdot d'_i + n$) [4, 5]. Therefore, Z_i should be obtained as

$$Z_i = \frac{f \cdot w}{f \cdot w / Z_0 + m \cdot d'_i + n} = \frac{f \cdot w}{\alpha d'_i + \beta} \tag{4}$$

where $\alpha = m, \beta = f \cdot w / Z_0 + n$.

3.2 Depth Error Model

The depth measurement model described above holds only in an ideal case. In practice, when the KINECT observes a planar plane, the depth maps have errors shown in Fig.3 top, and similar errors are also reported in [9]. To compensate these errors, we consider not only camera distortions but also distortions of the projector in our model.

As is well known, a lens distortion model can be expressed as follows[10]:

$$\breve{\boldsymbol{x}}_i^{(c)} = \begin{bmatrix} x_i^{(c)} + x_i^{(c)} \left(k_{c1} \boldsymbol{x}_i^{(c)\prime} \boldsymbol{x}_i^{(c)} + k_{c2} (\boldsymbol{x}_i^{(c)\prime} \boldsymbol{x}_i^{(c)})^2 \right) \\ y_i^{(c)} + y_i^{(c)} \left(k_{c1} \boldsymbol{x}_i^{(c)\prime} \boldsymbol{x}_i^{(c)} + k_{c2} (\boldsymbol{x}_i^{(c)\prime} \boldsymbol{x}_i^{(c)})^2 \right) \end{bmatrix}, \tag{5}$$

where $\boldsymbol{x}_i^{(c)}$ and $\breve{\boldsymbol{x}}_i^{(c)}$ are the ideal and distorted 2D positions, and k_{c1}, k_{c2} are the distortion parameters of the IR camera.

We assume the same distortion model to the projector distortion.

$$\breve{\boldsymbol{x}}_i^{(p)} = \begin{bmatrix} x_i^{(p)} + x_i^{(p)} \left(k_{p1} \boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)} + k_{c2} (\boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)})^2 \right) \\ y_i^{(p)} + y_i^{(p)} \left(k_{p1} \boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)} + k_{c2} (\boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)})^2 \right) \end{bmatrix}, \tag{6}$$

where $\boldsymbol{x}_i^{(p)}$ and $\breve{\boldsymbol{x}}_i^{(p)}$ are the ideal and distorted 2D positions, and k_{p1}, k_{p2} are the distortion parameters of the IR projector.

Here, we focus on a pattern $P(\boldsymbol{x}_i^{(p)})$ that is projected in the direction of point $\boldsymbol{x}_i^{(p)}$ (Fig.2). However, the pattern $P(\boldsymbol{x}_i^{(p)})$ is acutually projected in the direction of point $\breve{\boldsymbol{x}}_i^{(p)}$ by the projector distortions, and is projected onto the point Q_i'. In the camera, the pattern $P(\boldsymbol{x}_i^{(p)})$ is acutually projected onto the position $\breve{\boldsymbol{x}}_i^{(c)}$ because of the camera distortion. Therefore, the disparity d_i at $\breve{\boldsymbol{x}}_i^{(c)}$ can be obtained as $d_i = \breve{\boldsymbol{x}}_i^{(c)} - \boldsymbol{x}_i^{(p)}$.

On the other hand, focusing on Q_i' in Fig.2, the ideal disparity $\hat{d}_i$ corresponding to the point Q_i' should be

$$\hat{d}_i = \boldsymbol{x}_i^{(c)} - \breve{\boldsymbol{x}}_i^{(p)} = d_i - \epsilon_c - \epsilon_p, \tag{7}$$

where $\epsilon_c = \breve{\boldsymbol{x}}_i^{(c)} - \boldsymbol{x}_i^{(c)}$ and $\epsilon_p = \breve{\boldsymbol{x}}_i^{(p)} - \boldsymbol{x}_i^{(p)}$.

Thus, by estimating the distortion parameters of the IR camera and IR projector, we can correct the disparity and depth observation errors of the KINECT. Note that the observable data are only $\breve{\boldsymbol{x}}_i^{(c)}$ and d_i, we can calculate ϵ_c and ϵ_p as follows:

$$\epsilon_c = -\breve{x}_i \left(k_{c1} \breve{\boldsymbol{x}}_i^{(c)\prime} \breve{\boldsymbol{x}}_i^{(c)} + k_{c2} (\breve{\boldsymbol{x}}_i^{(c)\prime} \breve{\boldsymbol{x}}_i^{(c)})^2 \right) \tag{8}$$

$$\epsilon_p = x_i^{(p)} \left(k_{p1} \boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)} + k_{p2} (\boldsymbol{x}_i^{(p)\prime} \boldsymbol{x}_i^{(p)})^2 \right) \tag{9}$$

where $\boldsymbol{x}_i^{(p)} = \breve{\boldsymbol{x}}_i^{(c)} + \alpha d_i' + \beta$, and we employ the approximated undistortion model [2].

Figure3 shows an example of the disparity errors observed by the KINECT and the error simulation using our proposed error model. As can be seen, our model can approximate the disparity errors of the actual KINECT.

In the next section, we will describe the parameter estimation of these error model.

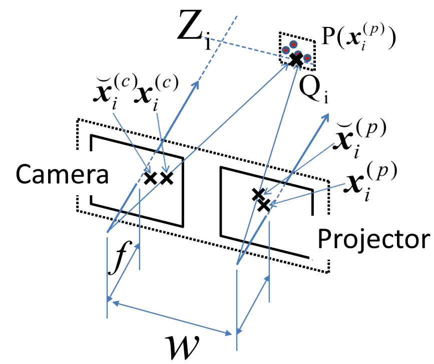

Figure 2. Proposed error model

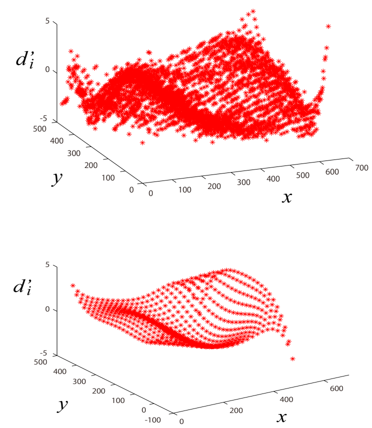

Figure 3. Top: Actual disparity errors, Bottom: Simulated results

4 Parameter Optimization

In the previous section, we described the proposed error model of the KINECT. However, not only these model parameters but also variations of parameters included in Eqs.(3) and (4), such as the focal length, cause depth measurement errors.

Therefore, we propose a method that estimates all the related parameters by only showing a planar board to the depth sensor with different poses and distances. Since it is difficult to optimize all the parameters at once, we employ two-step optimization scheme. First, we estimate the focal length f, and the disparity conversion parameters α and β. Second, the all the parameters including the proposed error model parameters, are estimated.

In the first step. we estimate the optimal values by minimizing the following equation using nonlinear optimization methods,

$$[\hat{\boldsymbol{A}}, \hat{\Pi}] = \arg\min_{A,\Pi} \sum D(\boldsymbol{X}_{(A, \breve{x}_i^{(c)}, d_i')}, \Pi_{(X_i)}) \tag{10}$$

Table 1. Estimated model parameters

	k_{c1}	k_{c2}	k_{p1}	k_{p2}
KINECT1	-0.068	0.137	0.050	-0.075
KINECT2	-0.023	0.048	0.025	-0.039

Table 2. Comparison of average plane fitting errors (mm)

	with correction	w/o correction
KINECT1	46.4	87.9
KINECT2	57.2	83.3

where $\boldsymbol{A} = [f, \alpha, \beta]$, $\boldsymbol{X}_{(A,x,d)}$ is calculated 3D position from $\boldsymbol{A}$, $\boldsymbol{x}$, and d. $D(X_i, \Pi)$ is the distance between the point X_i and the plane Π, and $\Pi_{(X_i)}$ is the plane estimated from points X_i. Here, we employ the values described in [5] as the initial values of the nonlinear optimization.

Using the parameters obtained in the first step as the initial values, we then estimate the optimal values of all parameters by minimizing the following equation,

$$[\hat{\boldsymbol{k}}, \hat{\boldsymbol{A}}, \hat{\Pi}] = \arg \min_{k,A,\Pi} \sum D(\boldsymbol{X}_{i(k,A,\breve{x}_i^{(c)},d'_i)}, \Pi_{(X_i)}). \tag{11}$$

where $\boldsymbol{k} = [k_{c1}, k_{c2}, k_{p1}, k_{p2}]$.

5 Experimental Results

To confirm the validity of the proposed error model and parameter estimation, we performed the following experiments. We first estimated the error model parameters of two KINECTs using our method. For the parameter estimation, we captured eleven observations of the plane with different poses and distances (a white plane board, 300×900 mm, is used).

Next, we compared the plane fitting errors of the observation depth values with and without error correction based on our model.

Table 1 shows the estimated model parameters of the two KINECTs, and Table 2 gives the comparison results of plane fitting errors with and without error correction. As can be seen, the plane fitting errors decreased using the error correction based on the proposed model, which successfully modeled the KINECT depth errors.

On the other hand, although our model seems to be able to represent the actual depth errors of the KINECT as shown in Fig.3, the plane fitting errors with error correction are not small enough. These suggest we need further investigation of our model such as inclusion of distortion centers of the camera and projectors, and so on.

6 Summary

In this paper, we proposed and evaluated a depth error model of the KINECT. In our method, we modeled depth measurement errors based on the distortions of IR camera and IR projector. The optimal model parameters can be estimated by only showing a known-width plane to the depth sensor with different distances and poses.

Experimental results show that the proposed error model can represent the depth mesurement errors of the KINECT. By using our model, about $30 - 50\%$ errors are decreased (about 4 [cm] errors). These results show that the proposed model is essential for use of KINECT as a 3D measuring device.

Future works include further investigation of the error model, and improvement of the optimization for the parameter estimation.

Acknowledgement

This work was partly supported by the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

References

[1] KINECT for Xbox 360, 2010.

[2] J. Heikkila. Geometric camera calibration using circular control points. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, 2000.

[3] D. Herrera C., J. Kannala, and J. Heikkila. Joint depth and color camera calibration with distortion correction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PrePrint.

[4] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12:1437–1454, 2012.

[5] K. Knolige and P. Mihelich. `http://www.ros.org/wiki/kinect_calibration/technical`.

[6] M. Lindner, A. Kolb, and T. Ringbeck. New insights into the calibration of ToF-sensors. *Int'l Conf. Comp. Vis. and Patt. Recog. Workshops*, 1-3(1):1–5, 2008.

[7] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int'l J of Comp. Vis.*, 47(1-3):7–42, 2002.

[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-Time Human Pose Recognition in Parts from Single Depth Images. In *CVPR 2011*, pages 1297–1304, 2011.

[9] J. Smisek, M. Jancosek, and T. Pajdla. 3D with kinect. In *Proc.1st IEEE Workshop on Consumer Depth Cameras for Computer Vision*, pages 1154–1160, 2011.

[10] Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.