Calibration of Depth Measurement Model for Kinect-Type 3D Vision Sensors

Branko Karan
Institute of Technical Sciences SASA
Belgrade, Serbia
branko.karan@itn.sanu.ac.rs

ABSTRACT

Accuracy of depth measurement with Microsoft Kinect and similar 3D vision sensors depends on variations in sensor production. Sensor reading may show significant systematic errors that can be compensated in software by using an adequate depth calibration model. This paper presents one such model and a procedure for identification of its parameters. An example calibration is given to illustrate the procedure and the attained improvements.

Keywords

Kinect, 3D sensing, depth camera, RGB-D camera, camera calibration, depth measurement

1. INTRODUCTION

Since its introduction in 2010, Microsoft Kinect sensor gained an enormous popularity in IT community. Prevalent interest has been in its use in computer games and human interface development. Additionally, thanks to its mass production and therefore low market price, there is also a growing interest in using this sensor in other fields, such as robotics (e.g., navigation, fine manipulation) and general 3D sensing.

In emerging fields of applications, accuracy requirements might be generally stronger than with originally envisioned human interfaces. Examples are robotics and similar areas where it is of vital importance to investigate and possibly improve sensor accuracy.

After initial release of Kinect XBOX 360 in 2010, Microsoft has launched an enhanced version named Kinect for Windows in 2012. At the same time, 3D vision sensors with similar design and capabilities have appeared, such as Asus Xtion and PrimeSense Carmine. Thus, it is possible to speak now about Kinect-type sensors that share similar features, both in terms of advantages and drawbacks.

There have been several works dedicated to improvement of accuracy of Kinect-type sensors. Early works, e.g. Burrus [Bur11] and Zhang and Zhang [ZZ11] addressed identification of intrinsic parameters of Kinect cameras. Later, focus was also on refinement of Kinect depth-disparity model, with important works of Khoshelham and Elberink [KE12], Smisek et al. [SJP11], and Herrera C. et al. [HKH12]. These works addressed transformation of Kinect disparity maps into depth maps. However, with new OpenNI [ON13] and Microsoft Kinect SDK [MS13], disparity data are already converted into depth, making the model considered by these authors obsolete.

In this work, an appropriate model for correction of Kinect depth measurements is formulated and a calibration procedure is proposed for identification of model parameters. The procedure is simple in the sense that it relies on commonly adopted camera calibration tools and it does not require specific calibration objects or external measurement devices. The model and the procedure are described in sections 2 and 3. Results of example calibration are given in Section 4, whereas Section 5 summarizes conclusions on the practical aspects of proposed approach.

2. DEPTH MEASUREMENT MODEL

Operation of Kinect depth sensor is grounded on structured light analysis approach. The sensor (Fig. 1) incorporates a laser IR diode for emitting a dotted light pattern and an IR camera for capturing reflected patterns. By using a suitable window size, the sensor compares reflected patterns to reference patterns, obtained for a plane placed at a known distance from the sensor, and uses the position of the best match pattern to infer disparity of reflected bean and further calculate the depth of reflection surface. A supplementary RGB camera is added to provide additional information on color and texture of the surface.

The relationship between depth of reflection surface and the disparity between images of light beans obtained for a reference and measurement (object) surface may be derived in the following manner (the derivation closely follows Khoshelham and Elberink

[KE12]). Looking at Fig. 1, where the reference bean is assumed to pass the path $P-R-C$ and the measurement bean passes the path $P-M-C$, from similarity of triangles $MR''C$ and $M'R'C$ we obtain:

$$\frac{d}{D}=\frac{f}{Z} \tag{1}$$

where Z is the distance of the measurement (object) plane from the sensor, $d=\overline{R'M'}$ denotes the disparity between images of reference R' and measurement M' beans, and f is the focal length of IR camera. From similarity of triangles CPR and

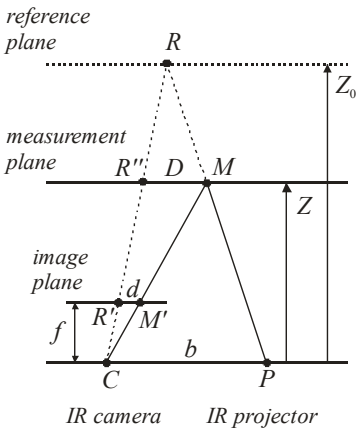

Figure 1. Depth measurement geometry

$R''MR$, another relation is obtained:

$$\frac{D}{b}=\frac{Z_0-Z}{Z_0} \tag{2}$$

in which Z_0 denotes the distance of the reference plane from the sensor. By eliminating D from (1) and (2), the depth measurement model is obtained in the form:

$$\frac{1}{Z}=\frac{1}{Z_0}+\frac{d}{bf} \tag{3}$$

The model (3) is internally used by sensor software to transform detected disparity into depth. However, the value of inferred depth depends on values b, f, Z_0 that are subject to manufacturing variations. Thus, the actual output Z_S from the sensor is really an approximation based on the nominal model:

$$\frac{1}{Z_S}=\frac{1}{Z_{0N}}+\frac{d}{b_N f_N} \tag{4}$$

By combining (3) and (4), the following relationship between Z and Z_S is obtained:

$$\frac{1}{Z}=a_Z\cdot\frac{1}{Z_S}+b_Z \tag{5}$$

where a_Z, b_Z are the values that are characteristics of particular sensor:

$$a_Z=\frac{b_N f_N}{bf};\quad b_Z=\frac{1}{Z_0}-a_Z\frac{1}{Z_{0N}}$$

Ideally, $a_Z=1, b_Z=0$. However, for a specific sensor, depth model parameters a_Z, b_Z may differ from ideal values and this difference may produce significant systematic errors in depth measurement. Therefore, appropriate tuning of depth model parameters may result in improved measurement accuracy.

3. CALIBRATION PROCEDURE

The proposed calibration procedure consists of two steps: the first step comprises standard calibration of sensor's RGB/IR cameras, whereas the calibration of depth model is performed within the second step.

Camera calibration

Camera calibration assumes identification of parameters of functions modeling transformation of coordinates of external objects into coordinates of their images. For a point with homogeneous coordinates $\mathbf{X}_e=\left[X_e,Y_e,Z_e,1\right]^T$ in some external world coordinate frame, transformation of coordinates involves (a) transformation $\mathbf{X}=\mathbf{T}_e\mathbf{X}_e$ into coordinates in camera frame, (b) projection $Z\cdot\mathbf{x}=\left[\mathbf{I}\quad\mathbf{0}\right]\mathbf{X}$ into the point $\mathbf{x}=\left[u,v,1\right]^T$ in normalized image plane, (c) distortion $\mathbf{x}^{(d)}=f^{(d)}(\mathbf{x})$, yielding distorted normalized coordinates $\mathbf{x}^{(d)}$, and finally (d) transformation into pixel coordinates $\mathbf{x}^{(p)}$, using transformation of the form $\mathbf{x}^{(p)}=\mathbf{K}\cdot\mathbf{x}^{(d)}$. Thus, calibration consists in identification of camera matrix $\mathbf{K}$, distortion function $f^{(d)}(\cdot)$ and transformation matrix $\mathbf{T}_e$.

Calibration of Kinect RGB/IR cameras follows a standard procedure of stereo camera calibration whose details is out of scope of this paper. It consists in essence in collecting pairs of images of different views of certain calibration object (e.g., a plane with a checkerboard texture) with known local 3D coordinates of specific feature points (e.g., checkerboard corners). The process proceeds by extracting 2D image coordinates of feature points from individual images and then by performing optimization to find the optimum transformations for all pairs of 3D/2D coordinates of all feature points in all views. The result is obtained in the form of camera matrices $\mathbf{K}_{RGB}$, $\mathbf{K}_{IR}$, distortion functions $f_{RGB}^{(d)}(\cdot)$, $f_{IR}^{(d)}(\cdot)$, and homogeneous transformation matrix ${}^{IR}\mathbf{T}_{RGB}$ for transformation

Figure 2. Calibration images

of homogeneous 3D coordinates from RGB to IR camera frame.

To achieve appropriate light conditions for calibration of IR camera, IR emitter has to be disabled during imaging (in this work, the older XBOX 360 was used, so the projector was simply covered by an ordinary stick tape; the newer model Kinect for Windows allows programmable control over the IR emitter).

Good calibration of both RGB/IR cameras is essential part of the procedure, because cameras are used as measurement devices for the second calibration step.

Depth model calibration

During the second step, a series of images is again collected, with each image containing a view to the calibration checkerboard plane placed in different positions and orientations w.r.t. the sensor. However, instead of pairs of RGB/IR images, this time the IR emitter of the sensor is left enabled and a set of pairs of RGB/depth images is collected.

As in the first step, checkerboard corners are used as calibration feature points with known local coordinates in checkerboard frame, which can be expressed as $\mathbf{X}_C(i,j) = \left[(i-1)\cdot w, (j-1)\cdot h, 0, 1\right]^T$, where w, h is the width/height of checkerboard fields. For each view k, the acquired RGB image is transformed into a grayscale image from which 2D pixel coordinates $\mathbf{x}_{RGB}^{(p)}(i,j,k)$ of all checkerboard corners are extracted and paired to known external coordinates $\mathbf{X}_C(i,j)$ to calculate the checkerboard pose ${}^{RGB}\mathbf{T}_C(k)$. Using the known transformation between camera frames, corner coordinates are expressed in IR camera frame as:

$$\mathbf{X}_{IR}(i,j,k) = {}^{IR}\mathbf{T}_{RGB} \cdot {}^{RGB}\mathbf{T}_C \cdot \mathbf{X}_C(i,j)$$

The *z*-component $Z(i,j,k)$ of such an obtained position is afterward compared to sensor reading. The sensor value is determined by converting $\mathbf{X}_{IR}(i,j,k)$ into pixel coordinates and by searching for the nearest neighbor in sensor depth map m_k:

$$\mathbf{x}_{IR}(i,j,k) = \left\{\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X}_{IR}(i,j,k)\right\}\Big/ Z(i,j,k)$$

$$\mathbf{x}_{IR}^{(p)}(i,j,k) = \mathbf{K}_{IR} \cdot f_{IR}^{(d)}(\mathbf{x}_{IR}(i,j,k))$$

$$Z_S(i,j,k) = m_k(\text{round}(\mathbf{x}_{IR}^{(p)}(i,j,k)))$$

Finally, the obtained set of pairs $Z(i,j,k)$ and $Z_S(i,j,k)$ is employed to fit the parameters a_Z, b_Z of depth calibration model (5).

4. CALIBRATION EXAMPLE

An experiment has been made to calibrate Kinect using a simple 9×8 checkerboard with 30mm square fields. First, a set of close-up views of the checkerboard placed in different orientations was employed to identify camera parameters. Afterward, for the second phase, a set of 27 RGB/depth image pairs as shown in Figure 2 was prepared. The images were divided in 9 groups, each containing three RGB/depth images of the calibration table placed approximately at the same distance from Kinect but in different orientations. To increase precision, all images were made with the highest resolution of Kinect cameras: 1280x960 for the RGB and 640x480 for the IR camera. For each image acquired by the RGB camera, corner coordinates were then extracted for inner 10×9 corners. In this manner, 9 point clusters, each containing 270 points were generated. The extracted coordinates were used to calculate orientation and position of the table and further to find 3D coordinates of all corners. The values of depth obtained from RGB images are finally compared to depth values obtained from Kinect. The differences are shown in Figure 3, where the points corresponding to particular clusters are displayed in different colors.

It is readily seen that both average values of errors and error deviations obtained from Kinect measurements increase with distance. Moreover, the averages show strong regularity that is in accordance to depth measurement model (5).

Using all collected points, parameters of the depth measurement model were computed by least square fitting as $a_Z = 0.9968$, $b_Z = 4.3651\cdot10^{-6}$. The corresponding correction curve is also shown in Figure 3.

Table 1 summarizes statistics on errors obtained before and after correction of depth measurement. It is seen that average values dramatically decreased. Average errors, which were in the range from 1mm for 1m distance (relative error of 0.1%) to 57mm on distance of 3.7m (relative error of 1.5%), practically diminished after calibration, that is, they have been reduced to a level of precision attainable by calibrated

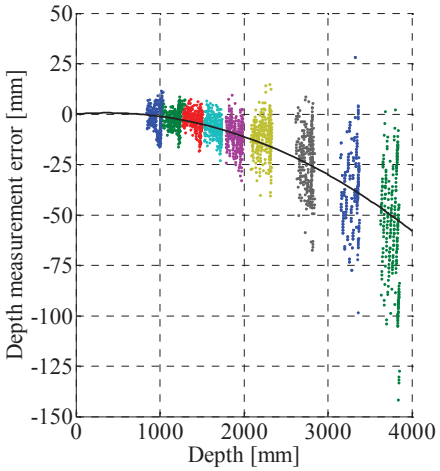

Figure 3. Depth measurement error

RGB camera. On the other side, as expected, confidence intervals (the values in Table 1 are empiric 3σ intervals) underwent only slight changes.

Good matching between experimentally obtained average errors and values predicted by the model offers a possibility to significantly reduce number of calibration measurements. Since the model (5) has only two parameters, it may be expected that similar results could be obtained by using only two clusters of points. Indeed, by selecting clusters 5 and 7, parameters are obtained as $a_Z = 0.9989$, $b_Z = 3.4618 \cdot 10^{-6}$.

	Depth measurement error [mm]		
Depth [mm]	Before calibration	After calibration	Approx. calibration
957	-1.2±18.3	-0.2±18.2	1.0±18.2
1182	-2.4±14.6	-0.1±14.7	1.2±14.7
1411	-3.5±14.2	0.7±13.5	2.0±13.5
1654	-7.0±16.1	-0.3±15.2	0.7±15.3
1907	-11.0±21.1	-1.1±20.5	-0.3±20.6
2228	-11.6±28.1	3.0±28.1	3.3±28.1
2756	-23.2±43.9	1.4±42.7	0.4±42.8
3283	-37.0±57.1	0.0±56.7	-2.7±56.7
3761	-56.9±83.6	-6.2±80.1	-11.0±80.6

Table 1. Improvement in depth measurement

The corresponding statistics calculated for all data points is shown in the last column of Table 1.

5. CONCLUSION

This paper has demonstrated that calibration of Kinect depth sensor can be achieved using the same calibration tools that are commonly used in ordinary camera calibration. Moreover, it has been shown that it is enough to perform calibration imaging at only two distances in the expected range of operation.

It is important to underline that the error reduction demonstrated in this paper is indeed a reduction of differences between reading of Kinect depth sensor and depth values obtained using Kinect RGB camera. Therefore, well calibrated camera is a prerequisite for a quality calibration.

6. ACKNOWLEDGMENTS

This work was supported in part by the Serbian Ministry of Education, Science and Technology Development under grants III-44008 and TR-35003.

All the analysis conducted in this work has been performed in Matlab environment, using Bouguet's Camera Calibration Toolbox [Bou10].

7. REFERENCES

[Bou10] Bouguet, J. Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/-calib_doc/ (last modified 09.07.2010)

[Bur11] Burrus, N. Kinect calibration. http://nicolas.-burrus.name/index.php/Research/Kinect-Calibration (last modified 07.08.2011)

[HKH12] Herrera C., D., Kannala, J., and Heikkilä, J. Joint depth and color camera calibration with distortion correction. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 34, pp. 2058–2064, 2012.

[KE12] Khoshelham, K. and Elberink, S.O. Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors, Vol. 12, pp. 1437–1454, 2012.

[MS13] Microsoft Corporation. Kinect for Windows SDK. http://msdn.microsoft.com/en-us/library/-hh855347.aspx (accessed 27.03.2013)

[ON13] OpenNI Consortium. OpenNI, the standard framework for 3D sensing, http://www.openni.-org/ (accessed 27.03.2013)

[SJP11] Smisek, J. Jancosek, M., and Pajdla, T. 3D with Kinect. Proc. IEEE Int. Conf. Computer Vision Workshops, Barcelona, Spain, pp. 1154–1160, 2011.

[ZZ11] Zhang, C. and Zhang, Z. Calibration between depth and color sensors for commodity depth cameras. 2011 IEEE Int. Conf. Multimedia and Expo (ICME), pp. 1-6, 2011.