Metrol. Meas. Syst., Vol. XXII (2015), No. 4, pp. 469–478.

PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229
www.metrology.pg.gda.pl

M&MS

METROLOGICAL ANALYSIS OF MICROSOFT KINECT IN THE CONTEXT OF OBJECT LOCALIZATION

Andrzej Skalski[1)], Bartosz Machura[1)]

AGH University of Science and Technology, Department of Measurement and Electronics, Al. Mickiewicza 30, 30-059 Kraków, Poland (✉ skalski@agh.edu.pl, +48 12 617 2828, machura@agh.edu.pl)

Abstract

This paper presents a comprehensive metrological analysis of the Microsoft Kinect motion sensor performed using a proprietary flat marker. The designed marker was used to estimate its position in the external coordinate system associated with the sensor. The study includes calibration of the RGB and IR cameras, parameter identification and image registration. The metrological analysis is based on the data corrected for sensor optical distortions. From the metrological point of view, localization errors are related to the distance of an object from the sensor. Therefore, the rotation angles were determined and an accuracy assessment of the depth maps was performed. The analysis was carried out for the distances from the marker in the range of 0.8−1.65 m. The maximum average error was equal to 23 mm for the distance of 1.6 m.

Keywords: Kinect motion sensor, localization, camera calibration, measurement precision, robustness.

1. Introduction

In the last few years there has been a significant development of optical sensors, which enabled tracking of 3D objects, their localization and surface reconstruction. This development led to a growing number of applications. As an example, the following can be listed: object and people tracking, motion capture and analysis, character animation, 3D scene reconstruction, gesture-based user interfaces and scientific applications [1–6]. The requirements for spatial and temporal resolution and accuracy depend on the type of a task [1–6].

Most of available tools are based on the use of structured light and infrared cameras with IR filters. A characteristic infrared radiation pattern is emitted by a laser, and a projection pattern is recorded using a corresponding camera. Reconstruction of a 3D scene is performed by analyzing the deformation of the pattern projected on the objects in the camera view field.

Owing to their price and available additional tools (*e.g.* SDK) Microsoft Kinect motion sensors are widely used. The Kinect uses structured light where depth measurement is based on the triangulation principle. The Kinect system consists of an IR camera, an IR projector and a camera working in the visible spectrum (RGB) (Fig. 1). The IR projector emits a pseudo-random pattern. Structured light falls on the scene, thereby forming a pattern projection of objects located in the emitter field. Next, follows the pattern projection acquisition using a camera with an infrared filter is performed, followed by analysis of the resulting image. The images are compared with a reference, which is an image of background taken with the IR camera for a known distance from the device. The reference is stored in the motion sensor memory.

Speckles, *i.e.* small elements of the IR projection, are deformed when they encounter obstacles: they change their shape, size, and location in the image [7]. As a result, it is possible

- 10.1515/mms-2015-0050

Article history: received on Jun. 15, 2015; accepted on Oct. 02, 2015; available online on Dec. 07, 2015; DOI: 10.1515/mms-2015-0050.

to construct a disparity map by assessing correlation between the obtained image and the reference one [8].

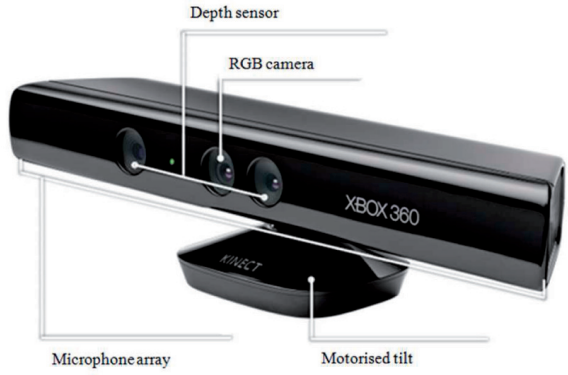

Fig. 1. Components of the Kinect motion sensor.

The article presents the results of research and detailed analysis of the metrological properties of the Microsoft Kinect motion sensor in the context of using it to localize and track objects. Distortions introduced by the optics of the system were taken into account.

Toth *et al*. [7] tested a system with the Kinect motion sensor. Camera calibration and model identification were not carried out in the work; the work was focused on the depth accuracy. In [9], the authors used a spatial marker with 5 spheres to assess the accuracy. The results of the accuracy study depending on the distance were presented for three values of angles between the optical axis and markers (450°, 900° and 1350°). Neither analysis of the impact of optical distortion on the accuracy was carried out in this work, nor information whether the camera calibration took place was given. In [2], authors focused on analysis of the repeatability error and on estimation of the localization error using a robot. Similarly to [9], no camera optics distortion was probably included that affected the accuracy of position estimation. Dutta [10] performed a study in a typical environment, determining the average localization error along the axes x, y, z. However, no specification of the influence of the optics, distances and angles between the optical axis and the subject was given. The Vicon system [11] was adopted as a reference system.

2. Experimental environment

All experiments and measurements were performed in a confined, normally sunlit space. The Kinect cameras were calibrated and images were matched. The process of calibration and matching algorithm are described in further parts of this chapter. In the tests there was used a planar marker designed by the authors, enabling, inter alia, to determine rotation (the vector direction and sense) of the marker in the external coordinate system associated with the Kinect. The marker is shown in Fig. 2.

2.1. Calibration of cameras

In the case of a stereo system it is necessary to determine the internal parameters of each camera, and then define a rigid transformation binding mutual positions of the devices. In the case of the Kinect sensor, this process enables moving the RGB video camera system to an external coordinate system represented by the system associated with the IR camera.

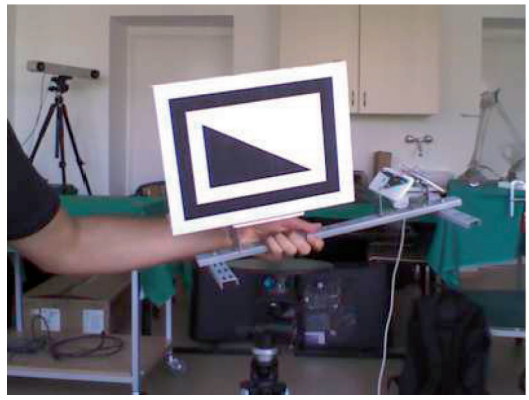

Fig. 2. The designed marker seen by the Kinect RGB camera.

Calibration of a single camera takes into consideration the linear perspective transformation and physical parameters of a matrix of pixels which leads to determining the matrix of internal parameters of the device. The relationship between the pixel and actual coordinates of a point in space can be described using the model (1–3). The resulting internal parameters are defined in the matrix **K**, which has been estimated for each camera [12, 13].

$$\mathbf{Z} = const. \Rightarrow p' \sim \mathbf{K}_s \, \mathbf{K}_f \, \mathbf{\Pi}_o \, \mathbf{P}, \tag{1}$$

$$\mathbf{Z}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & f \end{bmatrix}\begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2}$$

$$\mathbf{K} = \mathbf{K}_s\mathbf{K}_f = \begin{bmatrix} f \cdot s_x & f \cdot s_\theta & o_x \\ 0 & f \cdot s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_x & f_\theta & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where: $\mathbf{K}_s$ – the internal parameters resulting from the parameters of pixels; $\mathbf{K}_f$ – the internal parameters resulting from the perspective transformation; **K** – the matrix of the internal parameters of the camera; f_x, f_y, f_θ – the resulting calibration coefficients of the camera.

In addition, a correction of optical deformation is necessary. Deformations occur mainly in the form of radial distortion, which manifests itself in a so-called pincushion or barrel distortion effect [13]. The tangential distortions, resulting from the fact that the lens of the optical system is not perfectly parallel to the imaging plane, are less common. These deformations can be described by the distortion coefficients which enable correcting the coordinates of a point in the image. The distortion compensation was based on the (4) to (6) [13, 14]:

$$r = \sqrt{(x_u - x_c)^2 + (y_u - y_c)^2}\,, \tag{4}$$

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} x_d\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \\ y_d\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} x_d + \left(2p_1 x_d y_d + p_2\left(r^2 + 2x_d^2\right)\right) \\ y_d + \left(p_1\left(r^2 + 2y_d^2\right) + 2p_2 x_d y_d\right) \end{bmatrix}, \tag{6}$$

- 10.1515/mms-2015-0050

where: x_u, y_u – the coordinates of the point after the distortion removal; x_d, y_d – the coordinates of the point before distortion removal; x_c, y_c – the center of distortion; k_i – the i-th coefficient of radial distortion; p_i – the i-th coefficient of tangential distortion.

The calibration process was carried out based on the analysis of the corresponding characteristic points in the images of the same scene taken from two positions in space. The calibration board was made as an 8 x 6 chessboard grid pattern with a 5 cm edge of a single square. 210 shots taken for various board positions in space: the distance ranged from about 0.8 m to 3.0 m. The chessboards were being placed in various areas of the IR camera view (due to the larger focal length) at different angles, and two shots of a given scene were being taken. Both cameras were analysed for the defects in optical systems, *i.e.*, the tangential and radial distortions. Figs. 3 and 4 show the amount of change of the radial and tangential components for the RGB camera (Fig. 3) and the IR camera (Fig. 4), respectively.

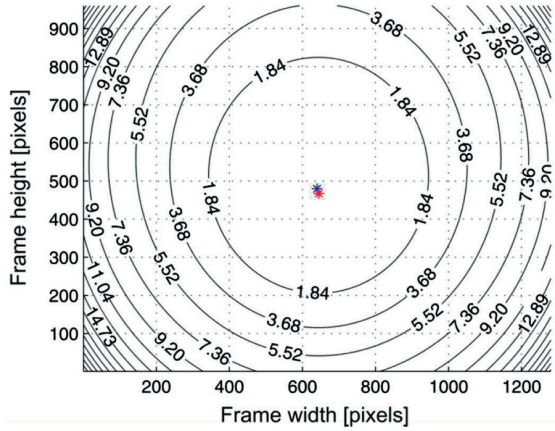

Fig. 3. The map of radial and tangential distortions of the RGB camera optical system: blue and red indicate the centre of the image and the main point, respectively.

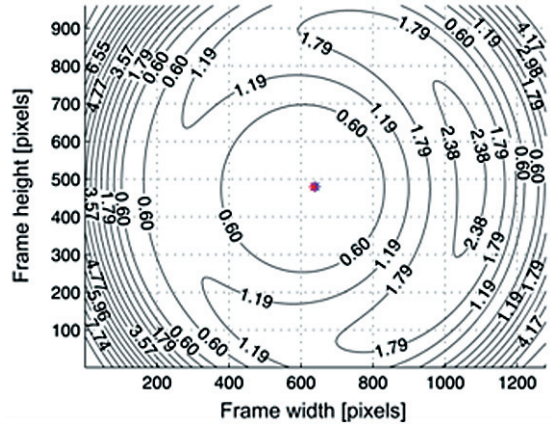

Fig. 4. The map of radial and tangential distortions of the IR camera optical system: blue and red indicate the centre of the image and the main point, respectively.

As can be seen in Figs. 3 and 4, the systems has a low optical distortion: the differences do not exceed approximately 13 pixels on the edges of the RGB camera image and approximately 7 pixels at the corners of the IR camera image on the left side (both top and bottom). The RGB

- 10.1515/mms-2015-0050

image distortions are of a barrel type, in contrast to those of the IR image, where small pincushion distortions occur. The range of the IR image distortions is smaller and their distribution is much more heterogeneous. Despite small optical defects of both cameras, it was decided to compensate distortion and the final analysis was performed on images with distortion removed. Table 1 shows the obtained calibration parameters based on a methodology presented in [15].

Table 1. The identified internal parameters of the Kinect motion-sensor cameras.

PARAMETER	RGB CAMERA	IR CAMERA
f_x	1103.07	1257.93
f_y	1103.43	1259.45
f_θ	0.0	0.0
o_x	644.81	635.74
o_y	467.81	479.01
k_1	0.11	–0.11
k_2	–0.38	0.79
k_3	0.56	–1.14
p_1	–0.0028	–0.00021
p_2	0.000032	–0.0017

2.2. Image registration

The estimated camera parameters have been analyzed and used to register sample images from the RGB camera to the corresponding images taken with the IR camera and generated depth maps.

Because the view fields of the cameras are different and the resulting coordinates x and y must correspond to consecutive natural numbers, interpolation of the values of adjacent pixels in the original image was necessary. This process was based on the bilinear interpolation. Each point from a video camera was transformed according to (7) and (8).

$$\mathbf{P}_W = \mathbf{R} * \mathbf{K}_{RGB}^{-1} * \mathbf{p}_{RGB} + \mathbf{T}, \tag{7}$$

$$\mathbf{p}_{IR} = \mathbf{K}_{IR} * \mathbf{P}_w, \tag{8}$$

where: the subscripts *RGB* and *IR* refer to the parameters of the cameras working in the visible and infrared ranges, respectively, **K** denote the matrices of internal parameters of each camera, **p** – points in the images of each camera, $\mathbf{P}_w$ – the actual coordinates of a point in the external coordinate system.

In this case matching consisted in transforming a two-dimensional image into a different two-dimensional image. The transformation was performed by the (9) to (11).

$$X_w = \frac{\left(x_{RGB} - o_x^{RGB}\right)}{f_x^{RGB}}, \; Y_w = \frac{\left(y_{RGB} - o_y^{RGB}\right)}{f_y^{RGB}}, \tag{9}$$

$$\begin{bmatrix} X_w \\ Y_w \end{bmatrix} = \begin{bmatrix} r_{11}X_w + r_{12}Y_W + r_{13} + t_1 \\ r_{21}X_w + r_{22}Y_W + r_{23} + t_2 \end{bmatrix}, \tag{10}$$

$$x_{IR} = X_W * f_x^{IR} + o_x^{IR}, \; y_{IR} = X_W * f_y^{IR} + o_y^{IR}. \tag{11}$$

- 10.1515/mms-2015-0050

Table 2 shows the obtained external calibration parameters.

Table 2. The identified external parameters of the Kinect motion-sensor cameras.

PARAMETER	VALUE	PARAMETER	VALUE	PARAMETER	VALUE
r_{11}	1.0	r_{21}	–0.0014	r_{31}	0.0017
r_{21}	0.0014	r_{22}	1.0	r_{23}	–0.0057
r_{31}	–0.0016	r_{32}	0.0057	r_{33}	1.0
t_1 [mm]	23.5	t_2 [mm]	–0.3	t_3 [mm]	–3,5
r_x [°]	0.33	r_y [°]	0.09	r_z [°]	0.08

The quality of the IR and RGB camera image matching is shown in Figs. 5 and 6. In Fig. 5 there can be observed the differences in the distribution of the distance between the pairs of characteristic points in both images. When the matching algorithm with correction of distortion was applied, the differences, and thus the errors in the representation of the scene between the cameras were substantially reduced.

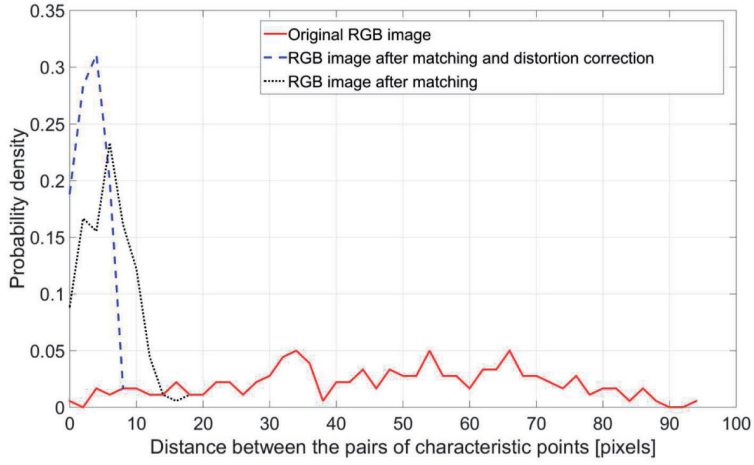

Fig. 5. The probability density function of the distance error between the corresponding points per image from the two cameras.

Fig. 6. Comparison of images from RGB and IR cameras before (left) and after (right) image matching and distortion removal.

Without registration of images and distortion correction the distance errors of about 80–90 pixels occurred.

Figure 6 shows that the highest efficiency of image registration is obtained in the case of image matching with optical distortions removed. When the transformation and matching are applied, the maximum error is almost 10 times lower than in the case without matching, and 2 times lower than in the case of the image with optical distortions. In the case of matching the images with removed distortions and the reference ones, the percentage of perfectly matched points (the difference of 0 pixels) is large (Fig. 5).

3. Accuracy analysis

The distance reference measurements were carried out with the use of a Leica DISTO D8 laser rangefinder (the distance accuracy over the studied range is +/– 1 mm). The test stand was designed in such a way so as to enable a gradual increase of the distance between the marker board and the sensor, starting from the sensor minimum distance (approx. 0.85 m), and ending with its maximum available value (approx. 3.95 m). Due to a variable nature of depth maps [16], five-second series of images (150 frames) were measured for each depth and the points in the middle of the maps were analyzed. On the basis of the recorded values there were determined the resolution, mean value and standard deviation from a given reference distance for a given value. The results are shown in Table 3.

Table 3. Analysis of the depth maps vs. the reference distance. All values in [mm].

DISTANCE FIXED	DISTANCE MEASURED WITH RANGEFINDER	AVERAGE DISTANCE	STANDARD DEVIATION
850	850.2	847.2	3.3
1000	1000.1	998.0	2.1
1500	1500.7	1496.4	4.6
2000	2003.7	2002.5	5.5
2500	2501.1	2523.3	22.3
3000	3001.1	3006.0	4.9
3500	3506.9	3530.2	24.9
3950	3952.5	3972.3	48.0

The estimation of the positioning accuracy as a function of the angle of rotation and the allowed range of rotation was carried out using a flat marker and a precision rotary table. The measurements were performed in four directions of rotation relative to the camera optical axis (left, right, up, down). For each recorded position 5 measurements of the resulting coordinates were recorded and averaged for the final analysis. The obtained mean values are presented in Tables 4 through 7. The determined angles of rotation largely agree with the expected values.

Table 4. Analysis of the marker orientation determination with the use of a rotary table for inclinations to the left relative to the optical axis.

THE ACTUAL ANGLE OF ROTATION ABOUT y AXIS [^{0}]	MEASURED ANGLE OF ROTATION ABOUT AXIS [^{0}]		
	x	*y*	*z*
10	1.03	10.77	0.94
20	0.63	21.69	1.71
30	1.10	32.56	3.02
40	0.51	41.70	3.58
50	0.88	51.52	5.09

Table 5. Analysis of the marker orientation determination with the use of a rotary table for inclinations to the right.

THE ACTUAL ANGLE OF ROTATION ABOUT y AXIS [^{0}]	MEASURED ANGLE OF ROTATION ABOUT AXIS [^{0}]		
	x	*y*	*z*
10	0.26	11.20	0.72
20	0.21	21.99	1.35
30	0.85	32.77	2.17
40	0.28	42.80	3.06
50	0.36	52.14	3.30

Table 6. Analysis of the marker orientation determination with the use of a rotary table for upward inclinations.

THE ACTUAL ANGLE OF ROTATION ABOUT y AXIS [^{0}]	MEASURED ANGLE OF ROTATION ABOUT AXIS [^{0}]		
	x	*y*	*z*
10	0.07	11.53	1.01
20	0.53	22.73	1.32
30	0.69	33.05	1.42
40	1.51	43.62	1.55
50	1.89	52.01	2.88

Table 7. Analysis of the marker orientation determination with the use of a rotary table for downward inclinations.

THE ACTUAL ANGLE OF ROTATION ABOUT y AXIS [^{0}]	MEASURED ANGLE OF ROTATION ABOUT AXIS [^{0}]		
	x	*y*	*z*
10	0.33	9.32	0.81
20	1.08	20.76	1.78
30	1.24	30.74	2.036
40	3.00	40.14	3.90

In order to determine the positioning accuracy a series of measurements were performed at a certain distance with a precision slider whose position was additionally verified with a laser rangefinder. Individual distances were determined as the distances between the origins of the coordinate systems associated with the marker in each measurement. The initial position of the slider carriage was set at a distance of 0.9 m from the motion sensor. The values registered in these periods were subjected to comprehensive analysis, as shown in Figs. 7 and 8.

The most significant error in this test was about 23 mm for the maximum distance from the initial position (750 mm); the marker was then at a distance of approx. 1650 mm from the motion sensor. On the basis of determined values it can be concluded that the localization error increases in proportion to the marker distance from the motion sensor, which is consistent with a linear relationship between the distances generated by the sensor and the actual distances in the analysed range. The value of the R^2 coefficient for a linear fit is 0.99 at $RMSE = 0.6554$ for the distance and localization error. The raw depth measurements provided by the sensor are non-linear. In our experiments we used the Microsoft SDK which delivers linearization [16].

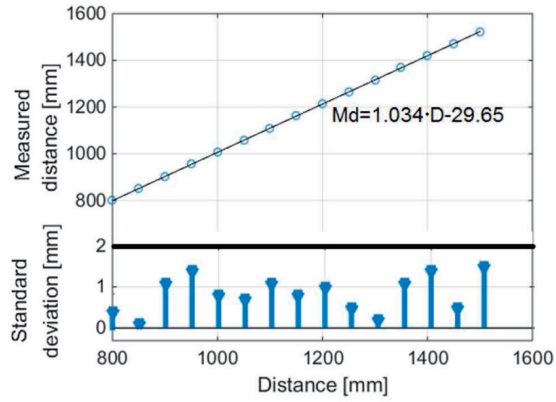

Fig. 7. The accuracy of distance determination based on the marker localization (Md) with the standard deviation for each measurement.

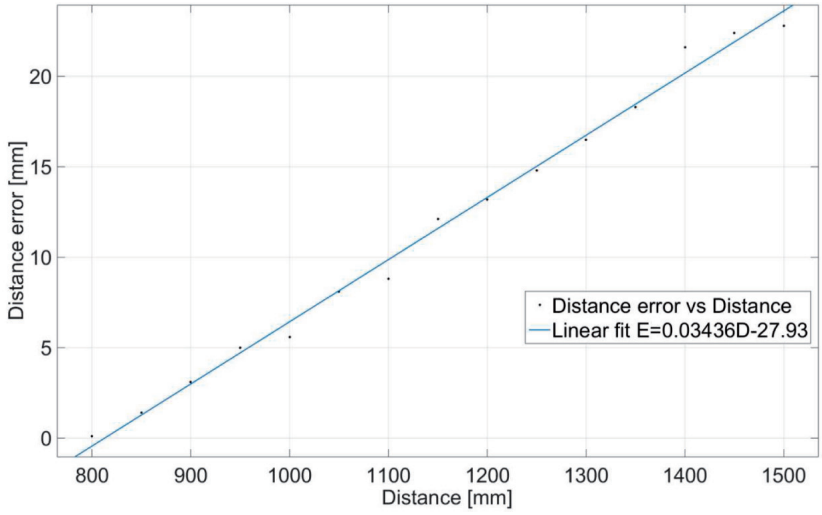

Fig. 8. The average error of distance determination.

4. Conclusions

The article presents a process of calibration and identification of the Microsoft Kinect sensor optics model parameters using a proprietary flat marker. Then the static parameters of the device have been determined using a rotary table and a slider with an electric drive. For analysis with a slider, a linear relationship was observed between the localization error and the location change. The maximum error in this case amounted to approximately 23 mm for a distance of 1.65 m from the sensor. The results obtained are accompanied with low values of standard deviations that do not depend on the marker distance, which confirms good reproducibility of the measurements. Using the rotary table, the marker angular working range was determined, which amounted to approx. 110^0 horizontally and approx. 90^0 vertically. The maximum measured difference from the set reference angle was approximately 3.5^0. The proposed methodology can be used to compare the Kinect sensor with a new Kinect for Xbox One [17].

Acknowledgements

The work was financed from the Dean Grants (statutory activity) and the AGH Rector Grant.

References

[1] Berger, K., *et al.* (2013). A state of the art report on kinect sensor setups in computer vision. Time-of-Flight and Depth Imaging. *Sensors, Algorithms, and Applications*, Springer Berlin Heidelberg, 257–272.

[2] Pedro, L.M., de Paula Caurin, G.A. (2012). Kinect evaluation for human body movement analysis, Biomedical Robotics and Biomechatronics (BioRob). *4th IEEE RAS & EMBS International Conference on*, 24–27, 1856–1861.

[3] Kar, A. (2010). Skeletal tracking using microsoft kinect. *Methodology*, 1, 1–11.

[4] Ross, A.C., *et al*. (2012). Validity of the Microsoft Kinect for assessment of postural control. *Gait & posture*, 36(3), 372–377.

[5] Baek-Lok, O., *et al.* (2014). Validity and reliability of head posture measurement using Microsoft Kinect. *British Journal of Ophthalmology*, 98(11), 1560–1564.

[6] Yu, C., Verkhoglyad, A., Poleshchuk, A., *et al.* (2014). 3D Optical Measuring Systems and Laser Technologies for Scientific and Industrial Applications. *Measurement Science Review*, 13(6), 322–328.

[7] Toth, K.C., *et al*. (2012). Calibrating the MS Kinect Sensor. *ASPRS Annual Conference*, Sacramento, USA.

[8] Khoshelham, K., Elberink, S.O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2), 1437–1454.

[9] Gonzalez-Jorge, H., *et al*. (2013). Metrological evaluation of Microsoft Kinect and Asus Xtion sensors. *Measurement*, 46(6), 1800–1806.

[10] Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics*, 43(4), 645–649.

[11] Vicon system: *http://www.vicon.com/*

[12] Fiala, M., Shu, Ch. (2008). Self-identifying patterns for plane-based camera calibration. *Machine Vision and Applications*, 19(4), 209–216.

[13] Hartley, R., Zisserman, A. (2001). *Multiple View Geometry in Computer Vision*. Prentice Hall.

[14] Stein, G.P. (1997). Lens distortion calibration using point correspondences, Computer Vision and Pattern Recognition. *Proc. of 1997 IEEE Computer Society Conference on*, 602–608.

[15] Heikklä, J., Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1106–1112.

[16] Andersen, M.R., *et al.* (2012). *Kinect depth sensor evaluation for computer vision applications. Department of Engineering. Aarhus University*. Denmark – Technical report ECE-TR-6.

[17] Pascoal, P.B., *et al*. (2015). Retrieval of Objects Captured with Kinect One Camera. *Eurographics Workshop on 3D Object Retrieval*.