An Integrated Procedure for Calibrating and Distortion Correction of the Structure Sensor and Stereo-Vision Depth Sensors

Dariusz Rzeszotarski$^{(\boxtimes)}$ and Pawel Strumillo

Institute of Electronics, Łódź University of Technology, 90-924 Łódź, Poland
{dariusz.rzeszotarski,pawel.strumillo}@p.lodz.pl

Abstract. The paper presents a calibration procedure of a 3D scene reconstruction system consisting of an active depth sensor (Structure Sensor) and a stereo camera with a wide view angle lenses. The wide angle lenses with large radial distortions used in the stereoscopic part of the system require application of a fish-eye model for correcting geometric distortions while for the infrared camera of the Structure Sensor a traditional pinhole model is sufficient. Calibration of the system comprises also a procedure for correcting depth distortions introduced by the Structure Sensor device. A simple yet efficient method for calibrating the cameras using functions provided by OpenCV library is proposed. The system is a part of a device helping visually impaired people to navigate in the environment.

1 Introduction

Nowadays 3D computer vision finds wider and wider applications. In case of mobile and autonomous systems the 3D data about the environment geometry is usually provided either by passive sensors - stereo cameras or active depth sensors like the Structure Sensor (SS) or Kinect devices [2,12]. Both approaches have their advantages and shortcomings.

Active depth sensors are based mainly on infrared structured-light or Time-of-Flight (ToF) technologies [2,4,12]. An example of a ToF camera is Kinect 2 device while the first versions of the Kinect and the Structure Sensor use infrared structured light pattern projector and a low range infrared CMOS camera. The main constraint of active infrared depth sensors is that their applications are limited to indoors spaces or at least to lighting conditions in which CMOS camera is not blinded and is able to recognize infrared pattern reflected by an object being observed. In case of the SS device the range of depth reconstruction is 400–4000 mm. The 3D depth map is delivered by the SS device with 30 fps rate with 640×480 spatial resolution and with high depth accuracy of 0.12–1% depending on the distance to an object (the higher the distance the lower the accuracy).

On the other hand stereo-vision systems are more customizable. There are many stereo-vision camera manufactures providing ready to use cameras with

different base length and distance range. There are also stereo-vision kits available with two synchronized cameras that can be mounted with the customizable base length and optics. However, stereo-vision systems contrary to active depth sensor perform poorly in low lighting conditions. Therefore the two depth sensor types are complementary regarding lighting conditions.

A 3D reconstruction module presented in this paper is a part of the device helping visually impaired people to navigate [1]. It should reconstruct the 3D scene in front of the user in different lighting conditions in indoor and outdoor environments. Therefore the two passive and active types of sensors are being used.

To take advantage of complementary features of the two sensors and to reconstruct 3D structure of a scene regardless of lighting conditions a fusion of the depth maps from both types of the sensors is applied. To that end calibration of the system is necessary. Calibration of the system includes evaluation of optical parameters of the sensors and relative positions of reference frames associated with each sensor and also evaluation and correction of distortions of depth data provided by the SS device.

2   Literature Review

Calibration of a 3D reconstruction system refers to estimation of the optical parameters of the cameras and their relative positions [11]. Calibration of an active depth sensor also includes establishing a relation between the depth data provided by the sensor and their true values. Recent literature is abundant with articles describing various methods of calibration between the depth sensor and single RGB camera. Smisek [7] et al. calibrate the Kinect device in three steps: (1) they estimate cameras intrinsic parameters (focal length in pixels, principal point coordinates, geometrical distortion coefficients) and the relative position of RGB and IR cameras, (2) they introduce the depth camera model describing a relation between the disparity value provided by the Kinect and the depth value, (3) depth model coefficients are computed by optimizing their values to best fit the model using the calibration points and their projections onto the best plane.

Herrera et al. [6] propose a calibration method of the Kinect device from RGB images and depth images pairs. The main disadvantage of the method, however, is the requirement of manual marking of four corners of the calibration object in depth images. Zhang and Zhang [13] also propose the calibration method using only RGB and the depth images. They introduce linear depth model relating the depth values with the corrected ones. All the parameters are finally refined by non-linear optimization.

Darwish [3] calibrates the SS device along with RGB camera mounted in the iPad and introduces analytical spatial distortion model of disparity from the depth sensor assuming radial and tangential nature of the distortions.

The distortion models used in the reviewed literature use disparity values. As Karan [9,10] pointed out current active depth sensor devices use OpenNI API

which provides only depth values, and taking this into account, he introduces the first nonlinear correction model in depth domain although omitting spatial distortions.

The system introduced in this article consists of stereo camera. The field of view (FoV) of this camera differs significantly from FoV of the SS depth sensor. Our method enables to build fused depth or disparity maps in real time. The calibration method described in this article addresses the following issues: (1) choosing an appropriate camera model describing optical properties of wide-angle lenses; (2) evaluation of relative position of sensors defined by different camera models (pinhole and fish-eye); (3) correction of spatial depth distortions of the SS device instead of just spatial disparity distortions (as the OpenNI library provides for the depth data).

3   Description of the System

The system consists of stereo camera LI-OV580-STEREO from Leopard Imagining and the Structure Sensor from Occipital Fig. 1a. The relative position of the cameras is shown in Fig. 1b. The figure shows top view of the coordinate frames associated with each camera. L, R denote left and right images of the stereo camera, IR - is the CMOS camera of the SS device that is sensitive to infrared radiation and thus is able to visualise the pattern being projected onto the scene by the infrared structure light projector (IR_P).
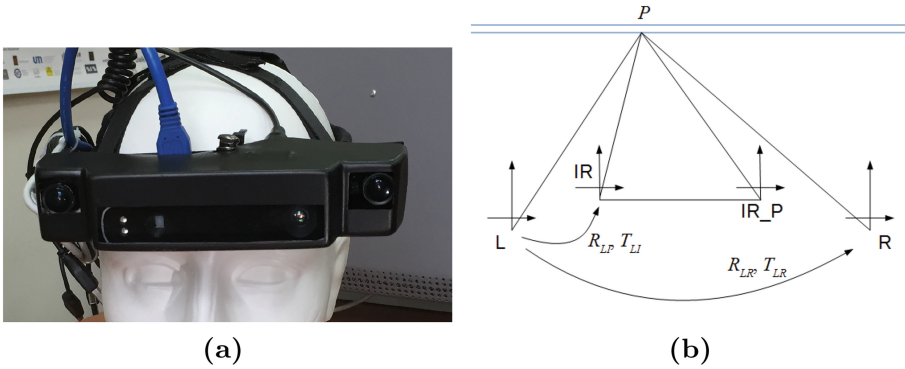

(a)   (b)

Fig. 1. The 3D reconstruction system: **a** The headgear with stereo and SS depth sensors. **b** Relative positions of the sensors and the parameters describing their relative positions.

3D coordinates of point P in the world reference frame are: $\boldsymbol{X} = [X, Y, Z]^T$. Point P is visualised and its image coordinates are $\boldsymbol{x_L} = [x_L, y_L]^T$, $\boldsymbol{x_R} = [x_R, y_R]^T$ in the left and right camera respectively.

The input images of the stereo camera are rectified i.e. their horizontal lines are aligned, i.e. the projections of point P defined in a 3D scene onto the rectified images have the same vertical components of their image coordinates $\boldsymbol{x_L} = [x_L, y_L]^T$, $\boldsymbol{x_R} = [x_R, y_R]^T$, $y_L = y_R = y$. Knowing the rectified image coordinates $\boldsymbol{x_L}$, $\boldsymbol{x_R}$ of a given point P its 3D coordinates in the left camera reference frame can be calculated using the following formula:

$$Z_L = f_x \frac{||T_{LR}||}{d}, \quad X_L = Z_L \frac{x_L - c_x}{f_x}, \quad Y_L = Z_L \frac{y_L - c_y}{f_y}, \tag{1}$$

where $d = x_L - x_R$ is called disparity, f_x, f_y, c_x, c_y are intrinsic parameters of the left and right rectified images which are the same for both (L and R) cameras; f_x, f_y are horizontal and vertical focal lengths expressed in pixels, c_x, c_y are image coordinates of the principal points (principal point is defined as intersection of Z axis with image plane), $||T_{LR}||$ is length of the translation vector between the left and right camera.

The same point P can be located with the use of the SS device: a specific infrared pattern is projected by IR_P onto point P and its image is visible in infrared camera IR. The IR image containing the infrared pattern is compared to the image of a plane placed at a given distance and containing the same infrared pattern. For each pixel p_I with coordinates $\boldsymbol{x_I} = [\boldsymbol{x_I}, \boldsymbol{y_I}]^T$ represented by a surrounding area with a distinguishable infrared pattern a corresponding pixel $\boldsymbol{x_I'}$ with the same pattern is found in the saved image. The difference in horizontal components of coordinates of the matched pixels is a disparity value inversely proportional to the distance likewise as in a stereo vision system as given in Eq. (1). The matching algorithm is implemented in the SS device and the output is a depth map, i.e. a 2-D array with depth values of respective pixels.

R_{LR}, T_{LR} denote extrinsic parameters (the rotation matrix and the translation vector) relating $\boldsymbol{X_L}$ coordinates in the reference frame of the left camera to $\boldsymbol{X_R}$ coordinates in the reference frame of the right camera:

$$\boldsymbol{X_L} = R_{LR}\boldsymbol{X_R} + T_{LR} \tag{2}$$

R_{LI}, T_{LI} denote extrinsic parameters (the rotation matrix and the translation vector) relating $\boldsymbol{X_L}$ coordinates in the reference frame of the left camera to $\boldsymbol{X_I}$ coordinates in the reference frame of the infrared camera (IR) of the SS device:

$$\boldsymbol{X_L} = R_{LI}\boldsymbol{X_I} + T_{LI} \tag{3}$$

The relation between 3D coordinates $\boldsymbol{X_{L||}}$ in the rectified left camera reference frame and 3D coordinates $\boldsymbol{X_I}$ in the infrared camera reference frame is as follows:

$$\boldsymbol{X_{L||}} = R_L R_{LI}\boldsymbol{X_I} + R_L T_{LI}, \tag{4}$$

where R_L is a rotation matrix relating 3D coordinates after and before rectification: $\boldsymbol{X_{L||}} = R_L \boldsymbol{X_L}$.

It is assumed that the depth map refers to a 2-D array with Z components of 3D coordinates. The point cloud map term refers to 2-D array with X, Y and Z

3D coordinates. The depth map produced by the SS device is a 2-D array that associates each pixel $p(x_I, y_I)$ from the corresponding infrared image with a Z_I component of $\boldsymbol{X_I}$ coordinate of a point from the object being observed. The X_I and Y_I components of a given element from the depth map can be retrieved using the formula:

$$X_I = \frac{Z_I}{f_{Ix}}\left(x_I - c_{Ix}\right), \quad Y_I = \frac{Z_I}{f_{Iy}}\left(y_I - c_{Iy}\right), \tag{5}$$

where IR camera intrinsic parameters f_{Ix}, f_{Iy}, c_{Ix}, c_{Iy} can be retrieved from the SS device using the OpenNI SDK or from the calibration procedure described in the next section. Using Eq. (1) the depth or a 3D point cloud map can be also computed from the disparity map based on the rectified images from the stereo camera.

4   Calibration

Calibration of the system requires evaluation of the following parameters: (1) intrinsic parameters of each single camera: the left and right camera of the stereo pair and an infrared CMOS camera from the SS device. (2) extrinsic parameters describing a relative position of the left and right camera of the stereo vision system and relative position of the left and infrared camera from the SS device; the extrinsic parameters are defined by Eqs. (2) and (3).

Calibration results of the three cameras (i.e. two cameras from the stereo vision system and one infrared camera of the SS device) with the use of a standard pinhole model with radial and tangential distortions are strongly vulnerable to positioning of the calibration board.

This is because we have used calibration procedures with fish-eye camera model for the left and right camera. The model was introduced in [8] and is available in fish-eye module of the OpenCV library.

The OpenCV library provides procedures for evaluating intrinsic parameters of a camera described with pinhole or fish-eye camera models. It also provides procedures for evaluation of relative position of two cameras described either by the pinhole or fish-eye camera models. The problem is that the left-infrared pair consists of cameras depicted by two different models: pinhole and fish-eye respectively. Therefore first intrinsic parameters of the left and right cameras and their relative position (R_{LR}, T_{LR}) was computed with the use of OpenCV calibration procedures for fish-eye model. Then using pairs of infrared images and rectified left images of the calibration board, intrinsic parameters of the infrared camera and its relative position (from Eq. (4)) against rectified left camera were computed by applying OpenCV calibration procedures for pinhole camera model.

The quality of the applied rectification procedure is evaluated as an average difference between $y-$components of corners found in the left and right rectified image in each position of the calibration board:

$$\Delta_{||y} = \sum_{j=1}^{M}\sum_{i=1}^{N}\frac{|y_{Lji} - y_{Rji}|}{MN} \tag{6}$$

where j is the index of the calibration board position, M is the number of calibration board positions, i is the index of corners pair, N is the number of corners in the calibration board.

The quality of evaluation of the extrinsic parameters relating positions of the infrared camera from the SS device to the left RGB camera is measured by average difference between image coordinates $\boldsymbol{x}'_{Ii}$ of corners found in the infrared image and image coordinates $\boldsymbol{x}_{Ii}$ of corners reconstructed from the stereo camera and reprojected onto the infrared image:

$$\Delta_{LRI} = \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{||\boldsymbol{x}'_{Iji} - \boldsymbol{x}_{Iji}||}{MN} \tag{7}$$

where j is index of calibration board position, M is the number of the calibration board positions, i is the index of corners, N is the number of corners in the calibration board.

Calibration Results

Table 1 shows average reprojection error in pixels for each calibration procedure used for estimation of extrinsic parameters $(R_{LR}, T_{LR}, R_{LI}, T_{LI})$. The values of the stereo camera reprojection errors Δ_{LR} are smaller for the fish-eye camera model than for the pinhole camera model.

Table 1. The average re-projection error in pixels of the calibration procedure: Δ_{LR}, Δ_{LI} - left-right, left-infrared relative position. Δ_{LRI} is the average reprojection error between stereo and IR camera. $\Delta_{||y}$ is the average difference in y-components of corners image coordinates in the rectified left-right pair.

Camera model	$\Delta_{LR}[pix]$	$\Delta_{LI}[pix]$	$\Delta_{LRI}[pix]$	$\Delta_{||y}[pix]$
pinhole	0.780	0.790	0.631	0.436
fish-eye	0.520	0.678	0.611	0.190

The reprojection error Δ_{LRI} calculated with use of Eq. (7) can be applied to evaluate the quality of R_{LI}, T_{LI} used for re-projecting of depth maps from the SS device onto the left rectified RGB image. Comparing the values of Δ_{LRI} from Table 1 shows that the reprojection error does not depend on choosing the model of the left RGB camera. This result is not surprising since the model of the infrared camera remains the same. On the other hand Table 1 shows that the quality of rectification evaluated as average difference between y-components of corners found in the left and right image is smaller if the fish-eye camera model is chosen for stereo cameras instead of the pinhole camera model.

5   Evaluation and Correction of Depth Distortions

In order to evaluate the quality of depth data provided by the SS device the following setup is proposed. As Smisek in [7] noticed in case of the Kinect device

a misalignment can occur between the infrared image and the depth image. In case of the Kinect the misalignment was noticed both along the horizontal and vertical axis of the image.

In order to evaluate the misalignment between the infrared and depth images from the SS device a calibration board was used. First, corners from the calibration board were found in an left RGB image. Next, based on the image coordinates of the found corners and their known 3D position in the calibration board reference frame the calibration board position in the left RGB camera reference frame was evaluated. Knowing the dimensions of the calibration board additional reference points were added at the edges of the board (Fig. 2a). Since the system was calibrated the reference points were projected onto the depth image. As the Fig. 2b shows in case of the SS device only the horizontal misalignment can be noticed.
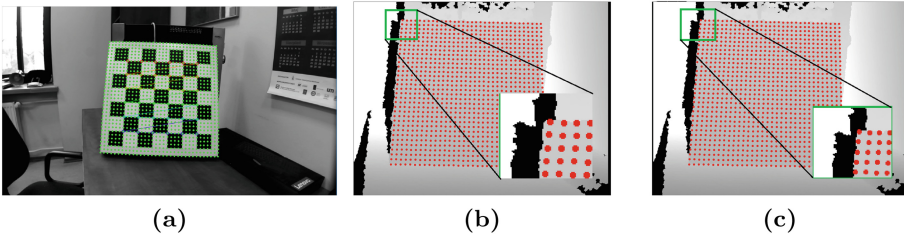

(a)   **(b)**   **(c)**

Fig. 2. (**a**) Reference points for misalignment evaluation. (**b**) Misaligned depth images. (**c**) Aligned depth images

In order to find the value of the misalignment the depth image was shifted by 0.1 pixel till the reprojected reference points were aligned with the edges of the calibration board. Figure 2c shows that the depth image shifted horizontally by 3 pixels can be regarded as aligned.

As noticed in [3,5,7] depth sensors based on a structure light introduce spatial nonlinear depth distortions. To evaluate the spatial depth distortions introduced by the SS device a sequence of depth images with plane object (plane wall) at distances in the range of 1000 to 3000 mm was taken. For each depth image from the sequence (using Eq. 5) a set of 3D coordinates $\boldsymbol{X_D} = [X_D, Y_D, Z_D]$ evenly distributed with 5 pixel step along x and y axis was chosen. As the set of 3D coordinates $\boldsymbol{X_D}$ belongs to the wall plane object it was fitted to a plane equation. For each point $\boldsymbol{X_D}$ from the set a plane fit residual was calculated as:

$$\Delta Z = Z_D - Z_p, \tag{8}$$

where Z_p is the depth component of projection of $\boldsymbol{X_D}$ 3D coordinate onto the fitted plane.

Figures 3a, b show 3D images of spatial distribution of plane fit residuals of depth image (red dots) from the SS device.
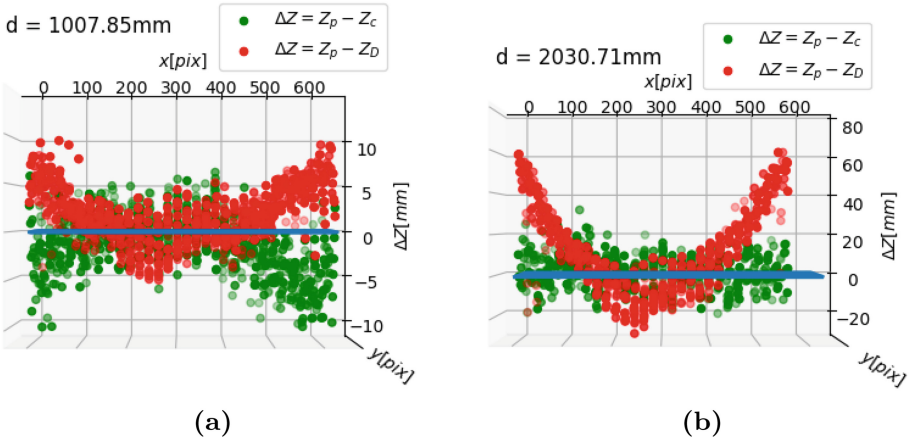
Fig. 3. 3D images of spatial distribution of the plane fit residuals: of distorted depth maps (red dots), of undistorted depth maps (green dots), along with the view of zero level at different distances **a** 1007.85 mm, **b** 2030.71 mm (Color figure online)
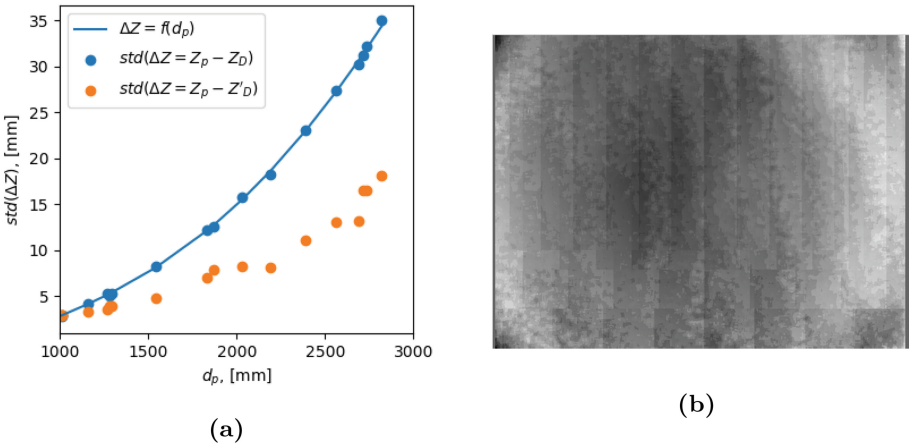

Fig. 4. a Standard deviation of difference between depth values and their projections onto the fitted plane. **b** The evaluated depth distortion pattern

Figure 4a show values of standard deviation of plane fit residuals $std(\Delta Z = Z_p - Z_D)$ for different plane distances d_p. By comparing the figures one can note that the shape of the spatial depth distortions does not change with the distance but the magnitude of the distortions increases as the distance of the plane increases. In order to correct the depth images captured by the SS device the following spatial depth distortion model is proposed. The model is similar to the one proposed by Herrera [5] but it is defined in the depth domain instead of disparity domain and the function depicting the change of distortions magnitude

is approximated by a polynomial. The proposed model can be described by the following formula:

$$Z'_D(x,y) = Z_D(x,y) + f(Z_D(x,y))D(x,y) \tag{9}$$

where $Z'_D(x,y)$ is the corrected depth value at x, y coordinates, $f(Z_D)$ is the function describing the change of the magnitude of spatial depth distortions, $D(x,y)$ is an array storing spatial depth distortions pattern.

To evaluate $f(Z_D)$ for each depth frame from the sequence a standard deviations of plane fit residuals were calculated. The function was described as polynomial approximation of the standard deviation distribution over plane distances. Figure 4a shows the distribution of the standard deviation of plane fit residuals $std(\Delta Z = Z_p - Z_D)$ and its approximation $\Delta Z = f(d_p)$:

$$f(Z) = W_0 + W_1 Z + W_2 Z^2 \tag{10}$$

In order to evaluate depth distortion pattern $D(x,y)$ from the depth sequence four frames with distinctive plane distances were chosen. Each element of depth distortions pattern $D(x,y)$ at coordinates x, y was evaluated using a linear least square approximation:

$$\sum_{i=1}^{N} |\Delta Z_i(x,y) - D(x,y)f(Z_{Di}(x,y))|^2 \tag{11}$$

where i is the index of depth frame, $N = 4$ is the number of depth frames used for approximation, $\Delta Z_i(x,y) = Z_p(x,y) - Z_D(x,y)$

Figure 4b shows the evaluated depth distortions pattern image. Figures 3a, b show 3D images of spatial distribution of plane fit residuals of undistorted (green dots) and distorted (red dots) depth image recorded by the SS device at different distances from plane wall object. Figure 4a shows that the values of standard deviation of plane fit residuals $std(\Delta Z = Z_p - Z'_D)$ for the corrected depth values Z'_D are almost twice as smaller than the plane fit residuals $std(\Delta Z = Z_p - Z_D)$ obtained for raw distorted depth values Z_D.

To evaluate the accuracy of depth values provided by the SS device a sequence of RGB and depth images with calibration board at distances in the range of 500 to 2000 mm was taken. For each j-th RGB-depth frame pair the position of the calibration board was evaluated from the detected chessboard corners in the left RGB image using OpenCV function solvePnP. The j-th position is depicted by rotation matrix R_{Lj} and translation vector T_{Lj}. Knowing the position of the calibration board and the relative position of the left and IR cameras the known 3D coordinates of chessboard corners $\boldsymbol{X_i}$ were transformed into the reference frame of the depth image: $\boldsymbol{X_{Ii}} = [X_{Ii}, Y_{Ii}, Z_{Ii}]^T$.

In order to retrieve for each i-th reference point a corresponding depth value Z_{Di} from the depth image, their image coordinates x_{Ii}, y_{Ii} were calculated by projecting the 3D coordinates $\boldsymbol{X_{Ii}}$ onto the depth image. After projecting the reference points onto the depth image I_D a set of depth values Z_{Di} belonging to the calibration board plane can be calculated. Using Eq. 5 a complete set of 3D

coordinates $\boldsymbol{X_{Di}} = [X_{Di}, Y_{Di}, Z_{Di}]^T$ of the reference points provided by the SS device can be obtained.

To evaluate the accuracy of the depth values provided by the SS device for each j-th depth-RGB frame depth components of two distance vectors were compared: (1) depth component Z_{Bj} of the vector connecting point $[0, 0, 0]$ in the depth reference frame and its projection onto the plane defined by the plane equation obtained from known 3D coordinates $\boldsymbol{X_{Ii}}$ of chessboard corners, (2) depth component Z_{Dj} of the vector connecting the point $[0, 0, 0]$ in the depth reference frame and its projection onto the plane defined by the plane equation obtained from 3D coordinates $\boldsymbol{X_{Di}}$ provided by the SS device.
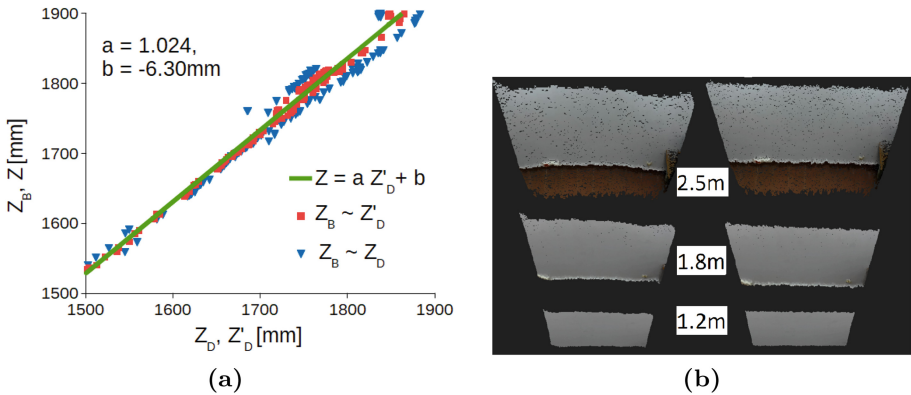

Fig. 5. **a** 3D view of a wall at different distances before and after undistorting depth maps from the SS device. **b** Depth components of plane distance obtained using calibration board position vs. depth components of plane distance obtained using depth values from the SS device along with linear approximation of the relation. (Color figure online)

Figure 5a shows depth components Z_B of the plane distance vector obtained from the calibration board position vs. plane distances Z_D, Z'_D obtained from the SS device distorted depth values (blue triangles) and corrected (red squares) ones. Figure 5a also shows the linear approximation (green line) of the relation between the expected depth components Z and corrected depth values Z'_D. As expected the distribution of expected depth components Z_B vs. Z'_D corrected depth components is less divergent from the linear approximation than the distribution Z_B vs. Z_D depth components before correction.

Figure 5b shows 3D visualizations of point clouds of wall object at different distances before correction (left column) of the depth components and after correction (right column) of the depth components. Figure 6 shows 3D visualizations of point clouds of staircase before (left image) and after (right image) the depth components correction.
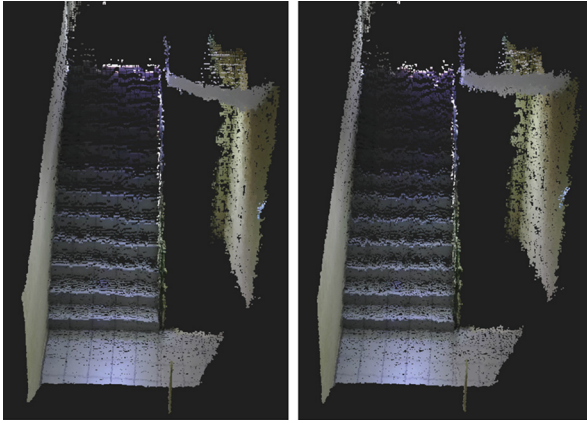
Fig. 6. 3D view of the staircase before and after undistorting depth maps from the SS device

6   Conclusions

Calibration procedure of a 3D reconstruction system consisting of stereo camera and an active depth sensor was presented. The main contribution of the article is addressing issues related to combining depth sensors with different camera models (pinhole and fish-eye) and depth distortions of the Structure Sensor device. Applying the fish-eye model makes the calibration results less vulnerable to calibration board positioning and yields smaller average difference of y components of image coordinates of corresponding points after rectification. The model of spatial depth distortions of the SS device is introduced as an alternative to popular in literature spatial disparity distortions model tested mostly for Kinect active depth sensors. Correction of depth maps from the SS device with the proposed model gives twice as smaller plane fit residuals as before correction.

References

1. Caraiman, S., et al.: Computer vision for the visually impaired: the sound of vision system. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1480–1489 (2017)
2. Dal Mutto, C., Zanuttigh, P., Cortelazzo, G.M.: Time-of-Flight Cameras and Microsoft Kinect™. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-3807-6
3. Darwish, W., Tang, S., Li, W., Chen, W.: A new calibration method for commercial RGB-D sensors. Sensors **17**(6), 1204 (2017). Calibration
4. Hansard, M., Lee, S., Choi, O., Horaud, R.P.: Time-of-Flight Cameras: Principles, Methods and Applications. Springer, London (2013). https://doi.org/10.1007/978-1-4471-4658-2
5. Herrera, D., Kannala, J., Heikkilä, J.: Joint depth and color camera calibration with distortion correction. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 2058–2064 (2012)

6. Herrera, C.D., Kannala, J., Heikkilä, J.: Accurate and practical calibration of a depth and color camera pair. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6855, pp. 437–445. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23678-5_52
7. Smisek, J., Jancosek, M., Pajdla, T.: 3D with kinect. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1154–1160 (2011)
8. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Trans. Pattern Anal. Mach. Intell. **28**(8), 1335–1340 (2006)
9. Karan, B.: Calibration of depth measurement model for kinect-type 3D vision sensors (2013)
10. Karan, B.: Calibration of kinect-type RGB-D sensors for robotic applications. FME Trans. **43**(1), 47–54 (2015)
11. Rzeszotarski, D., Pełczyński, P.: Software application for calibration of stereoscopic camera setups. Metrol. Meas. Syst. **19**(4), 805–816 (2012)
12. Sell, J., O'Connor, P.: The xbox one system on a chip and kinect sensor. IEEE Micro **34**(2), 44–53 (2014)
13. Zhang, C., Zhang, Z.: Calibration between depth and color sensors for commodity depth cameras. In: Shao, L., Han, J., Kohli, P., Zhang, Z. (eds.) Computer Vision and Machine Learning with RGB-D Sensors. ACVPR, pp. 47–64. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08651-4_3