

Article

# Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications

Kouros Khoshelham \* and Sander Oude Elberink

Faculty of Geo-Information Science and Earth Observation, University of Twente, P.O. Box 217, Enschede 7514 AE, The Netherlands; E-Mail: oudeelberink@itc.nl

\* Author to whom correspondence should be addressed; E-Mail: khoshelham@itc.nl;  
Tel.: +31-53-487-4477; Fax: +31-53-487-4335.

Received: 14 December 2011; in revised form: 6 January 2012 / Accepted: 31 January 2012 /  
Published: 1 February 2012

---

**Abstract:** Consumer-grade range cameras such as the Kinect sensor have the potential to be used in mapping applications where accuracy requirements are less strict. To realize this potential insight into the geometric quality of the data acquired by the sensor is essential. In this paper we discuss the calibration of the Kinect sensor, and provide an analysis of the accuracy and resolution of its depth data. Based on a mathematical model of depth measurement from disparity a theoretical error analysis is presented, which provides an insight into the factors influencing the accuracy of the data. Experimental results show that the random error of depth measurement increases with increasing distance to the sensor, and ranges from a few millimeters up to about 4 cm at the maximum range of the sensor. The quality of the data is also found to be influenced by the low resolution of the depth measurements.

**Keywords:** range camera; calibration; sensor; RGB-D; point cloud; triangulation; imaging; error budget; laser scanning

---

## 1. Introduction

Low-cost range sensors are an attractive alternative to expensive laser scanners in application areas such as indoor mapping, surveillance, robotics and forensics. A recent development in consumer-grade range sensing technology is Microsoft's Kinect sensor [1]. Kinect was primarily designed for natural interaction in a computer game environment [2]. However, the characteristics of the data captured by

Kinect have attracted the attention of researchers from other fields [3–11] including mapping and 3D modeling [12–15]. A demonstration of the potential of Kinect for 3D modeling of indoor environments can be seen in the work of Henry *et al.* [16].

The Kinect sensor captures depth and color images simultaneously at a frame rate of up to 30 fps. The integration of depth and color data results in a colored point cloud that contains about 300,000 points in every frame. By registering the consecutive depth images one can obtain an increased point density, but also create a complete point cloud of an indoor environment possibly in real time. To realize the full potential of the sensor for mapping applications an analysis of the systematic and random errors of the data is necessary. The correction of systematic errors is a prerequisite for the alignment of the depth and color data, and relies on the identification of the mathematical model of depth measurement and the calibration parameters involved. The characterization of random errors is important and useful in further processing of the depth data, for example in weighting the point pairs or planes in the registration algorithm [17,18].

Since Kinect is a recent development—it was released in November 2010—little information about the geometric quality of its data is available. Geometric investigation and calibration of similar range sensors, such as the SwissRanger [19] and PMD [20], has been the topic of several previous works [21–26]. However, these sensors are based on the time-of-flight measurement principle, and are fundamentally different from the Kinect which is a triangulation sensor.

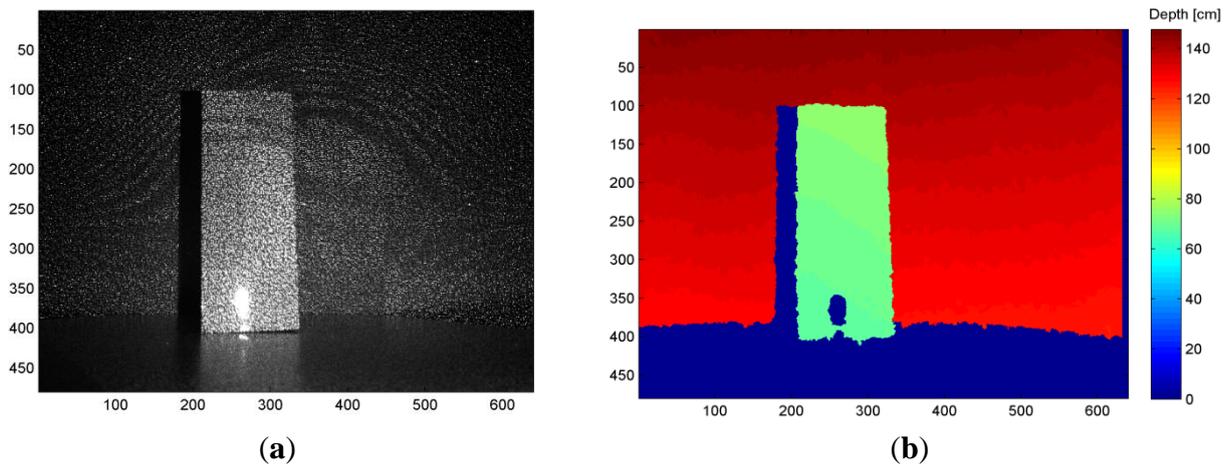
In this paper our primary focus is on the depth data. The objective of the paper is to provide an insight into the geometric quality of the Kinect depth data through calibration and an analysis of the accuracy and density of the points. We present a mathematical model for obtaining 3D object coordinates from the raw image measurements, and discuss the calibration parameters involved in the model. Further, a theoretical random error model is derived and verified by an experiment.

The paper proceeds with a description of the depth measurement principle, the mathematical model and the calibration parameters in Section 2. In Section 3, the error sources are discussed, and a theoretical error model is presented. In Section 4, the models are verified through a number of experiments and the results are discussed. The paper concludes with some remarks in Section 5.

## 2. Depth Measurement by Triangulation

The Kinect sensor consists of an infrared laser emitter, an infrared camera and an RGB camera. The inventors describe the measurement of depth as a triangulation process [27]. The laser source emits a single beam which is split into multiple beams by a diffraction grating to create a constant pattern of speckles projected onto the scene. This pattern is captured by the infrared camera and is correlated against a reference pattern. The reference pattern is obtained by capturing a plane at a known distance from the sensor, and is stored in the memory of the sensor. When a speckle is projected on an object whose distance to the sensor is smaller or larger than that of the reference plane the position of the speckle in the infrared image will be shifted in the direction of the baseline between the laser projector and the perspective center of the infrared camera. These shifts are measured for all speckles by a simple image correlation procedure, which yields a disparity image. For each pixel the distance to the sensor can then be retrieved from the corresponding disparity, as described in the next section. Figure 1 illustrates the depth measurement from the speckle pattern.

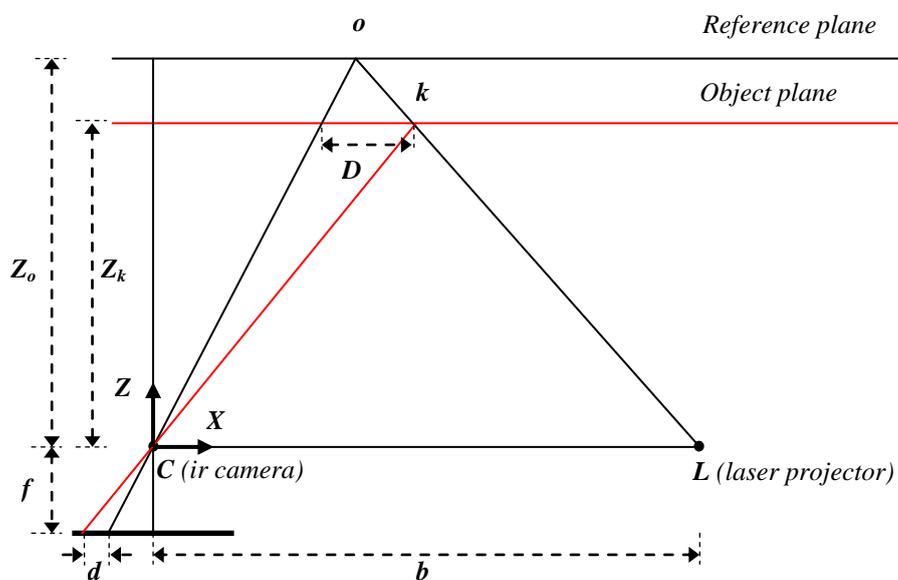
**Figure 1.** (a) Infrared image of the pattern of speckles projected on a sample scene. (b) The resulting depth image.



### 2.1. Mathematical Model

Figure 2 illustrates the relation between the distance of an object point  $k$  to the sensor relative to a reference plane and the measured disparity  $d$ . To express the 3D coordinates of the object points we consider a depth coordinate system with its origin at the perspective center of the infrared camera. The  $Z$  axis is orthogonal to the image plane towards the object, the  $X$  axis perpendicular to the  $Z$  axis in the direction of the baseline  $b$  between the infrared camera center and the laser projector, and the  $Y$  axis orthogonal to  $X$  and  $Z$  making a right handed coordinate system.

**Figure 2.** Relation between relative depth and measured disparity.



Assume that an object is on the reference plane at a distance  $Z_o$  to the sensor, and a speckle on the object is captured on the image plane of the infrared camera. If the object is shifted closer to (or further away from) the sensor the location of the speckle on the image plane will be displaced in the  $X$

direction. This is measured in image space as disparity  $d$  corresponding to a point  $k$  in the object space. From the similarity of triangles we have:

$$\frac{D}{b} = \frac{Z_o - Z_k}{Z_o} \quad (1)$$

and:

$$\frac{d}{f} = \frac{D}{Z_k} \quad (2)$$

where  $Z_k$  denotes the distance (depth) of the point  $k$  in object space,  $b$  is the base length,  $f$  is the focal length of the infrared camera,  $D$  is the displacement of the point  $k$  in object space, and  $d$  is the observed disparity in image space. Substituting  $D$  from Equation (2) into Equation (1) and expressing  $Z_k$  in terms of the other variables yields:

$$Z_k = \frac{Z_o}{1 + \frac{Z_o}{fb} d} \quad (3)$$

Equation (3) is the basic mathematical model for the derivation of depth from the observed disparity provided that the constant parameters  $Z_o$ ,  $f$ , and  $b$  can be determined by calibration. The  $Z$  coordinate of a point together with  $f$  defines the imaging scale for that point. The planimetric object coordinates of each point can then be calculated from its image coordinates and the scale:

$$\begin{aligned} X_k &= -\frac{Z_k}{f}(x_k - x_o + \delta x) \\ Y_k &= -\frac{Z_k}{f}(y_k - y_o + \delta y) \end{aligned} \quad (4)$$

where  $x_k$  and  $y_k$  are the image coordinates of the point,  $x_o$  and  $y_o$  are the coordinates of the principal point, and  $\delta x$  and  $\delta y$  are corrections for lens distortion, for which several models with different coefficients exist; see for instance [28]. Note that here we assume that the image coordinate system is parallel with the base line and thus with the depth coordinate system.

## 2.2. Calibration

As mentioned above, the calibration parameters involved in the mathematical model for the calculation of 3D coordinates from the raw image measurements include:

- focal length ( $f$ );
- principal point offsets ( $x_o$ ,  $y_o$ );
- lens distortion coefficients (in  $\delta x$ ,  $\delta y$ );
- base length ( $b$ );
- distance of the reference pattern ( $Z_o$ ).

In addition, we may consider a misalignment angle between the  $x$ -axis of the image coordinate system and the base line. However, this does not affect the calculation of the object coordinates if we define the depth coordinate system to be parallel with the image coordinate system instead of the base line. We may, therefore, ignore this misalignment angle.

From the calibration parameters listed above the first three can be determined by a standard calibration of the infrared camera. In practice, however, the calibration parameters of the infrared camera do not directly correspond to the disparity images, because the size of the disparity images computed by the internal processor of Kinect is  $640 \times 480$  pixels, which is smaller than the actual size of the infrared sensor ( $1,280 \times 1,024$  pixels) [29]. Due to the bandwidth limitation of the USB connection, the images of the infrared sensor are also streamed in a reduced size of  $640 \times 480$  pixels corresponding to the disparity images (that is the images are resized and cropped).

Therefore, a convenient approach to the calibration is to estimate the calibration parameters from the reduced infrared images instead of the actual sensor, provided that a pixel-to-pixel correspondence exists between the reduced infrared images and the disparity images. By examining the images we observed a shift of 4 pixels in the  $x$  direction between the disparity and infrared images (supposedly implying the application of a 9-pixel wide correlation window for calculating disparities [30]). Once this shift is corrected for, the calibration parameters estimated from the reduced infrared images can be applied to the measurements in the disparity images.

The determination of the base length and the reference distance is more complicated for the following reason. In practice, the raw disparity measurements are normalized and quantized between 0 and 2,047, and streamed as 11 bit integers. Therefore, in Equation (3)  $d$  should be replaced with  $md' + n$  with  $d'$  the normalized disparity and  $m, n$  the parameters of a (supposedly) linear normalization (in fact denormalization). Including these in Equation (3) and inverting it yields:

$$Z_k^{-1} = \left(\frac{m}{fb}\right)d' + \left(Z_o^{-1} + \frac{n}{fb}\right) \quad (5)$$

Equation (5) expresses a linear relation between the inverse depth of a point and its corresponding normalized disparity. By observing the normalized disparity for a number of object points (or planes) at known distances to the sensor the coefficients of this linear relation can be estimated in a least-squares fashion. However, the inclusion of the normalization parameters does not allow determining  $b$  and  $Z_o$  separately.

The calibration parameters mentioned above completely define the relation between the image measurements ( $x, y, d'$ ) and object coordinates ( $X, Y, Z$ ) of each point. Once estimated, they enable the generation of a point cloud from each disparity image. Note that these parameters do not describe the internal geometry of the infrared camera as they are estimated from the resized and cropped images.

### 2.3. Adding Color to the Point Cloud

The integration of the depth and color data requires the orientation of the RGB camera relative to the depth coordinate system. Since we defined the depth coordinate system at the perspective center of the infrared camera we can estimate the orientation parameters by a stereo calibration of the two cameras. The parameters to be estimated include three rotations between the camera coordinate system of the RGB camera and that of the infrared camera, and the 3D position of the perspective center of the RGB camera in the coordinate system of the infrared camera. In addition, the interior orientation parameters of the RGB camera, *i.e.*, the focal length, principal point offsets and the lens distortion parameters must be estimated.

In practice, the images of the RGB camera are also streamed in a reduced size; therefore, it is more convenient to perform a stereo calibration of the reduced images instead of the physical cameras. The resulting parameters describe the relation between the 3D coordinates of each point and its corresponding pixel-coordinates in the reduced RGB image. Once these parameters are estimated, we can add color to the point cloud by projecting each 3D point to the RGB image and interpolating the color.

### 3. Depth Accuracy and Resolution

Accuracy and point density are two important measures for evaluating the quality of a point cloud. In the following sections factors influencing the accuracy and density of Kinect data are discussed, and a theoretical random error model is presented.

#### 3.1. Error Sources

Error and imperfection in the Kinect data may originate from three main sources:

- The sensor;
- Measurement setup;
- Properties of the object surface.

The sensor errors, for a properly functioning device, mainly refer to inadequate calibration and inaccurate measurement of disparities. Inadequate calibration and/or error in the estimation of the calibration parameters lead to systematic error in the object coordinates of individual points. Such systematic errors can be eliminated by a proper calibration as described in the previous section. Inaccurate measurement of disparities within the correlation algorithm and particularly the quantization of the disparities also influence the accuracy of individual points.

Errors caused by the measurement setup are mainly related to the lighting condition and the imaging geometry. The lighting condition influences the correlation and measurement of disparities. In strong light the laser speckles appear in low contrast in the infrared image, which can lead to outliers or gap in the resulting point cloud. The imaging geometry includes the distance to the object and the orientation of the object surface relative to the sensor. The operating range of the sensor is between 0.5 m to 5.0 m according to the specifications, and, as we will see in the following section, the random error of depth measurement increases with increasing distance to the sensor. Also, depending on the imaging geometry, parts of the scene may be occluded or shadowed. In Figure 1, the right side of the box is occluded as it cannot be seen by the infrared camera though it may have been illuminated by the laser pattern. The left side of the box is shadowed because it is not illuminated by the laser but is captured in the infrared image. Both the occluded areas and shadows appear as gaps in the point cloud.

The properties of the object surface also impact the measurement of points. As it can be seen in Figure 1, smooth and shiny surfaces that appear overexposed in the infrared image (the lower part of the box) impede the measurement of disparities, and result in a gap in the point cloud.

### 3.2. Theoretical Random Error Model

Assuming that in Equation (5) the calibration parameters are determined accurately and that  $d'$  is a random variable with a normal distribution we can propagate the variance of the disparity measurement to obtain the variance of the depth measurement as follows:

$$\sigma_Z^2 = \left( \frac{\partial Z}{\partial d} \right)^2 \sigma_{d'}^2 \quad (6)$$

After simplification this yields the following expression for the standard deviation of depth:

$$\sigma_Z = \left( \frac{m}{fb} \right) Z^2 \sigma_{d'} \quad (7)$$

with  $\sigma_{d'}$  and  $\sigma_Z$  respectively the standard deviation of the measured normalized disparity and the standard deviation of the calculated depth. Equation 7 basically expresses that the random error of depth measurement is proportional to the square distance from the sensor to the object. Since depth is involved in the calculation of the planimetric coordinates, see Equation (4), we expect the error in  $X$  and  $Y$  to be also a second order function of depth. By propagating the errors in Equation (4), and assuming that the random error of image coordinates  $x, y$  can be ignored, we obtain the random error of  $X$  and  $Y$ :

$$\begin{aligned} \sigma_x &= \left( \frac{mx}{f^2 b} \right) Z^2 \sigma_{d'} \\ \sigma_y &= \left( \frac{my}{f^2 b} \right) Z^2 \sigma_{d'} \end{aligned} \quad (8)$$

### 3.3. Depth Resolution and Point Density

The resolution of the infrared camera, or more precisely the pixel size of the disparity image, determines the point spacing of the depth data on the  $XY$  plane (perpendicular to camera axis). Since each depth image contains a constant  $640 \times 480$  pixels, the point density will decrease with increasing distance of the object surface from the sensor. Considering the point density as the number of points per unit area, while the number of points remains constant the area is proportional to the square distance from the sensor. Thus, the point density on the  $XY$  plane is inversely proportional to squared distance from the sensor.

The depth resolution refers to the minimum depth difference that can be measured, and is determined by the number of bits per pixel used to store the disparity measurements. The Kinect disparity measurements are stored as 11-bit integers, where one bit is reserved to mark the pixels for which no disparity is measured, so-called no-data. Thus, a disparity image contains 1,024 levels of disparity. Since depth is inversely proportional to disparity, the resolution of depth is also inversely related to the levels of disparity. Let  $Z(d')$  denote depth as a function of normalized disparity  $d'$ , then depth resolution is simply the depth difference corresponding to two successive levels of disparity; *i.e.*,  $\Delta_Z(d') = Z(d') - Z(d' - 1)$ , and as we learned the differential yields:

$$\Delta_Z = \left( \frac{m}{fb} \right) Z^2 \quad (9)$$

Thus, the depth resolution is also a quadratic function of depth, and decreases with increasing distance from the sensor.

#### 4. Experiments and Results

Experiments were carried out to first determine the calibration parameters of the sensor and then investigate the systematic and random errors in the depth data. The following sections describe the tests and discuss the results.

##### 4.1. Calibration Results

A standard camera calibration was performed using the reduced images of both the infrared camera and the RGB camera to estimate the calibration parameters in the Photomodeler<sup>®</sup> software. A total of eight images of a target pattern were taken by both cameras from different angles. To avoid the disturbance of the laser speckles in the infrared images the aperture of the laser emitter was covered by a piece of opaque tape. To model the lens distortion we used the well-known model of Brown [31] with three radial distortion parameters ( $K1$ ,  $K2$ ,  $K3$ ) and two decentering parameters ( $P1$ ,  $P2$ ). The calibration was first performed with all lens distortion parameters as unknowns. Then, those parameters whose standard deviation was large compared to the estimated parameter value were removed from the estimation model, and the remaining parameters were estimated again. As a result, parameter  $K3$  was excluded from the parameter sets of both cameras, and  $P2$  was excluded from the parameter set of the RGB camera. Table 1 summarizes the results of the calibration procedure. The overall calibration accuracy as the RMS of point marking residuals in image space was 0.14 pixels for the IR images and 0.09 pixels for the RGB images. In the parameters of the RGB images notice the very large principal point offset ( $y_o$ ) of 0.327 mm corresponding to 35 pixels; see also Figure 3(b). This value is close to 32 pixels, which is the offset we would expect if a reduced image was obtained by resizing a full resolution image to one-half and cropping it at the top ( $1024/2 - 480 = 32$ ). The infrared images however do not have large principal point offsets, meaning that they were cropped at the center. The reason for this apparently inconsistent cropping and the large  $y_o$  in the RGB images is not known to the authors.

**Table 1.** Calibration parameters estimated for the infrared and RGB images.

Calibration parameter		IR images	RGB images
Focal length	$f$	$5.453 \pm 0.012$ [mm]	$4.884 \pm 0.006$ [mm]
Principal point offset	$x_o$	$-0.063 \pm 0.003$ [mm]	$0.032 \pm 0.002$ [mm]
	$y_o$	$-0.039 \pm 0.008$ [mm]	$-0.327 \pm 0.005$ [mm]
Frame dimension	$w$	$5.996 \pm 0.001$ [mm]	$6.012 \pm 0.002$ [mm]
	$h$	4.5 [mm]	4.5 [mm]
Pixel size	$p_x$	9.3 [ $\mu$ m]	9.3 [ $\mu$ m]
	$p_y$	9.3 [ $\mu$ m]	9.3 [ $\mu$ m]
Radial lens distortion	$K1$	$2.42e-3 \pm 1.2e-4$	$-5.75e-3 \pm 6.4e-5$
	$K2$	$-1.70e-4 \pm 1.2e-5$	$4.42e-4 \pm 5.8e-6$
	$K3$	0	0
Decentering lens distortion	$P1$	$-3.30e-4 \pm 3.7e-5$	$-1.07e-4 \pm 2.8e-5$
	$P2$	$5.25e-4 \pm 7.5e-5$	0

**Figure 3.** Lens distortions of (a) IR camera and (b) RGB camera. The principal points are marked by x and the image centers by +.

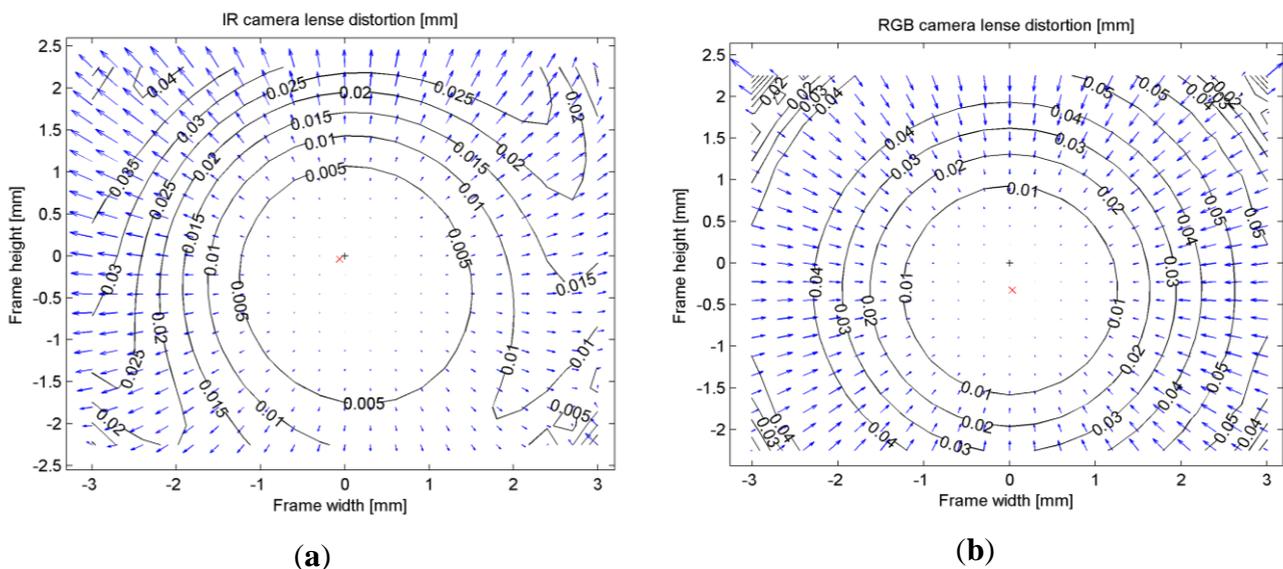
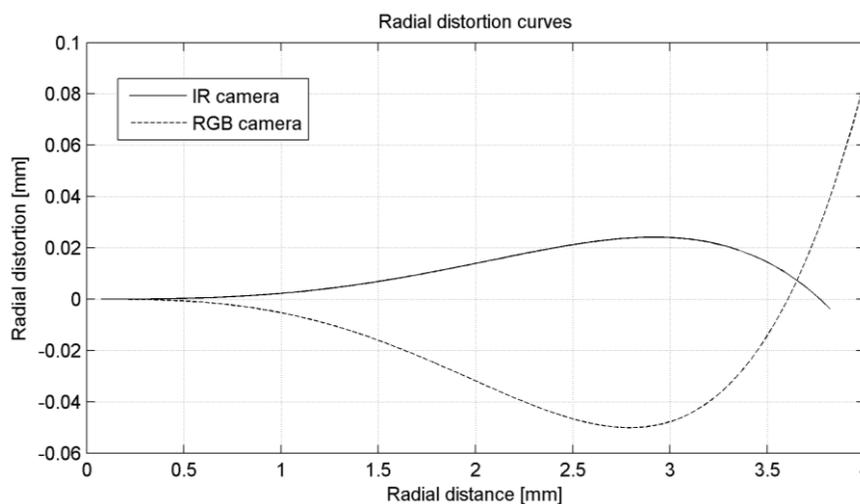


Figure 3 shows the combined effect of radial and decentering distortions for both the IR and the RGB images. Notice the larger effect of decentering distortions in the IR image as compared to the RGB image. The magnitude of radial distortions however is larger in the RGB image, particularly in the upper left corner where radial distortions reach 9 pixels (0.08 mm). This can be verified by examining the radial distortion curves in Figure 4, which show that radial distortions of the RGB camera are generally larger than those of the IR camera. In practice, radial distortions in the RGB images lead to misalignments between the color and depth data in the point cloud. A distortion of 0.08 mm in the image space corresponds to a misalignment of 8 cm at the maximum range of the sensor (5 m).

**Figure 4.** Radial distortion curves for the IR and RGB images.



For the stereo calibration images of a checkerboard pattern were taken simultaneously by the two cameras, and the relative orientation parameters were estimated in a bundle adjustment. Table 2 lists the resulting parameters. As it can be seen the rotations are quite small, and the relative position

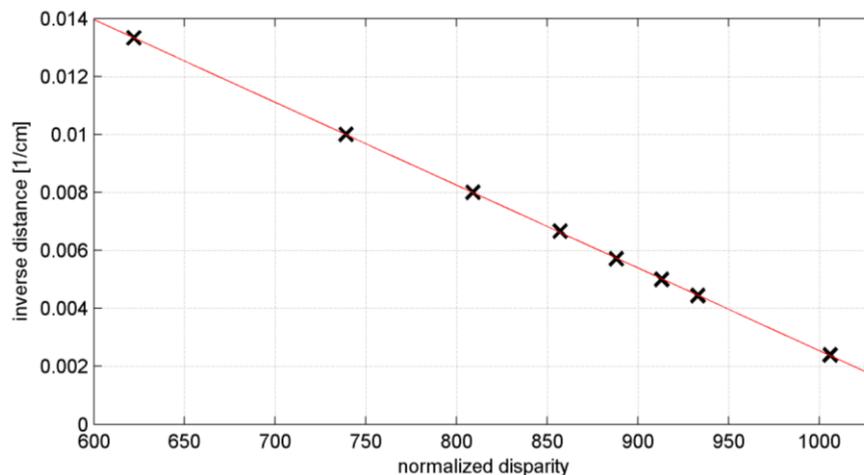
parameters indicate that the center of the RGB camera is approximately on the base line between the IR camera and the laser emitter.

**Table 2.** Exterior orientation parameters of the RGB camera relative to the IR camera.

Rotation parameters [degree]			Position parameters [mm]		
$r_x$	$r_y$	$r_z$	$t_x$	$t_y$	$t_z$
0.56	0.07	0.05	-25.60	0.34	2.91

To determine the parameters involved in the disparity-depth relation (Equation 5) depth values were measured at eight different distances to the sensor using a measuring tape. The inverse of the measured distances were then plotted against the corresponding normalized disparities observed by the sensor, see Figure 5. As it can be seen the relation is linear as we expected from the mathematical model in Equation (5). The parameters of the disparity-depth relation were obtained by a simple least-squares line regression. These were found to be  $-2.85e-5$  as the slope and 0.03 as the intercept of the line. Using these parameters we can now calculate depth values from the observed normalized disparities.

**Figure 5.** Linear relation of normalized disparity with inverse depth.



#### 4.2. Comparison of Kinect Point Cloud with the Point Cloud of a High-End Laser Scanner

To investigate the systematic errors in Kinect data a comparison was made with a point cloud obtained by a high-end laser scanner. The Kinect point cloud was obtained from the disparity image using Equations (4) and (5) and the calibration parameters from the previous step. The laser scanner point cloud was obtained of the same scene by a calibrated FARO LS 880 laser scanner. The nominal range accuracy of the laser scanner is 0.7 mm for a highly reflective target at a distance of 10 m to the scanner [32,33]. The average point spacing of the laser scanner point cloud on a surface perpendicular to the range direction (and also the optical axis of the infrared camera of Kinect) was 5 mm. It was therefore assumed that the laser scanner point cloud is sufficiently accurate and dense to serve as reference for the accuracy evaluation of the Kinect point cloud. In the absence of any systematic errors the mean of discrepancies between the two point clouds is expected to be close to zero.

To enable this analysis, first, an accurate registration of the two point clouds is necessary. The registration accuracy is important because any registration error may be misinterpreted as error in the

Kinect point cloud. To achieve the best accuracy two registration methods were tested. The first method consisted of a manual rough alignment followed by a fine registration using the iterative closest point (ICP) algorithm [34]. To make ICP more efficient a variant suggested by Pulli [35] was followed in which 200 randomly selected correspondences (closest points) with a rejection rate of 40% were used. In the second method the two roughly-aligned point clouds were segmented into planar surfaces and 20 corresponding segments were manually selected. Then, a robust plane fitting using RANSAC [36,37] was applied to obtain plane parameters and the inlying points. The registration was then performed by minimizing the distances from the points in one point cloud to their corresponding planes in the other point cloud [38].

In both registrations the estimated transformation parameters consisted of a 3D rotation and a 3D translation. To reveal a possible scale difference between the point clouds a third registration was performed using the plane-based method augmented with a scale parameter.

Table 3 summarizes the registration residuals pertaining to the three methods. It is clear that the methods perform similarly, all yielding very comparable residuals. Furthermore, the scale parameter obtained from the third registration was found to be 1.01. The largest effect of such a scale on the furthest point of the point cloud is 5 cm, which is not larger than the random error and depth resolution of the data as will be shown later. Thus, we can conclude that there is no scale difference between the Kinect point cloud and the laser scanner point cloud.

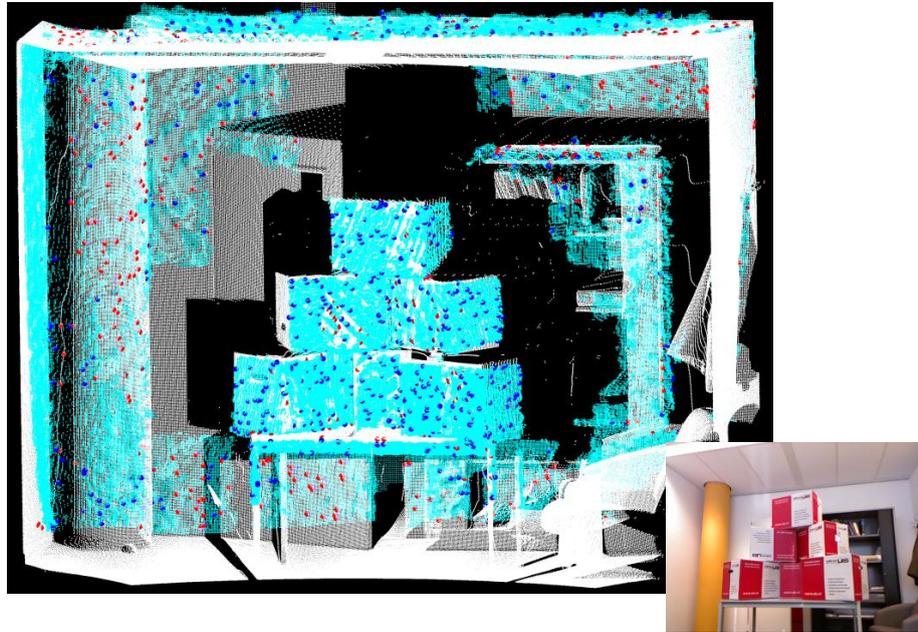
**Table 3.** Registration results of the three methods.

	Transformation parameters			Residuals				
	$s$	$r_x, r_y, r_z$ [deg]	$t_x, t_y, t_z$ [cm]	Min [cm]	Mean [cm]	Med [cm]	Std [cm]	Max [cm]
point-point distances (icp)	-	-88.16, 0.03, 0.07	1.20, -0.81, 3.56	0.1	1.2	0.9	0.9	4.4
point-plane distances without scale	-	-91.52, 0.14, -0.23	0.07, -0.32, 0.82	0.0	1.1	0.8	0.9	7.1
point-plane distances with scale	1.01	-90.64, -0.02, 0.04	-0.27, -3.53, -5.60	0.0	1.1	0.9	0.9	7.0

For the comparison between the two point clouds the result of the ICP registration method was used. A total of 1,000 points were randomly selected from the Kinect point cloud and for each point the nearest neighbor was found in the laser scanner point cloud. These closest point pairs were the basis for evaluating the accuracy of the Kinect point cloud. However, it was considered that the point pairs might contain incorrect correspondences, because the two sensors had slightly different viewing angles, and therefore, areas that could not be seen by one sensor might be captured by the other and vice versa. Figure 6 shows the two point clouds and the closest point pairs.

Figure 7 shows the histograms of discrepancies between the point pairs in X, Y and Z. Table 4 lists the statistics related to these discrepancies. The mean and median discrepancies are close to zero, which is an indication that there are no systematic shifts between the two point clouds. For comparison, the last three rows of Table 4 show the discrepancies between the laser scanner point cloud and an uncalibrated Kinect point cloud, measured on the same number of sampled point pairs. The discrepancies are clearly larger when the uncalibrated point cloud is used, indicating the effect of calibration.

**Figure 6.** Comparison of Kinect point cloud (cyan) with the point cloud obtained by FARO LS880 laser scanner (white). The larger points are samples randomly selected from the Kinect data (blue) and their closest point in the laser scanner data (red). The thumbnail on the lower right is a color image of the setup.



**Figure 7.** Histograms of discrepancies between the closest point pairs in X, Y and Z direction.

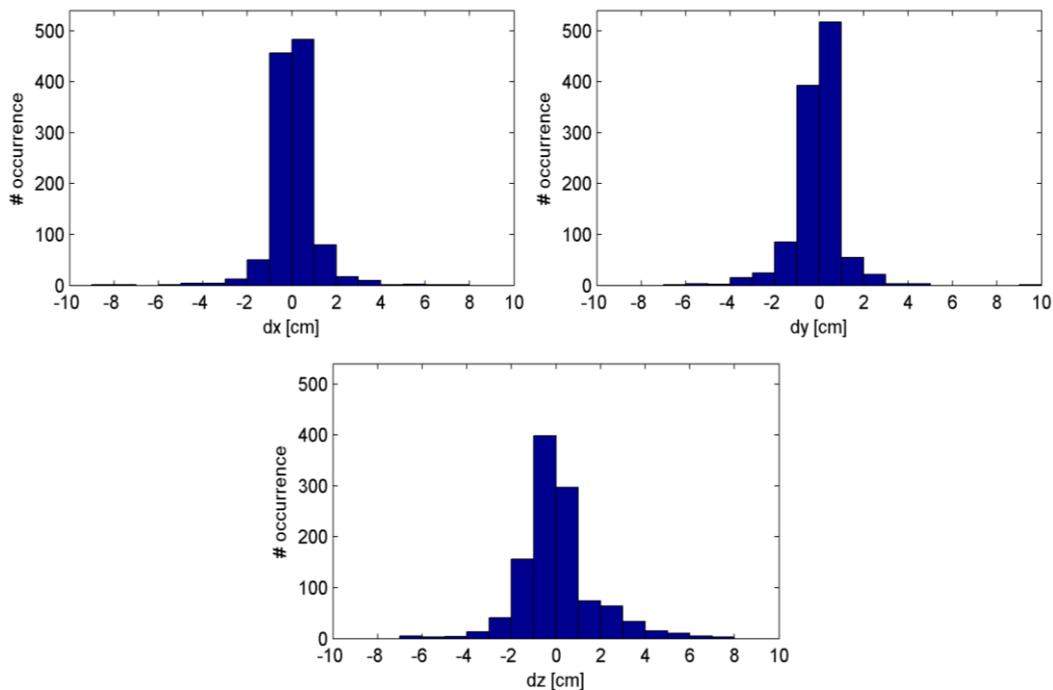
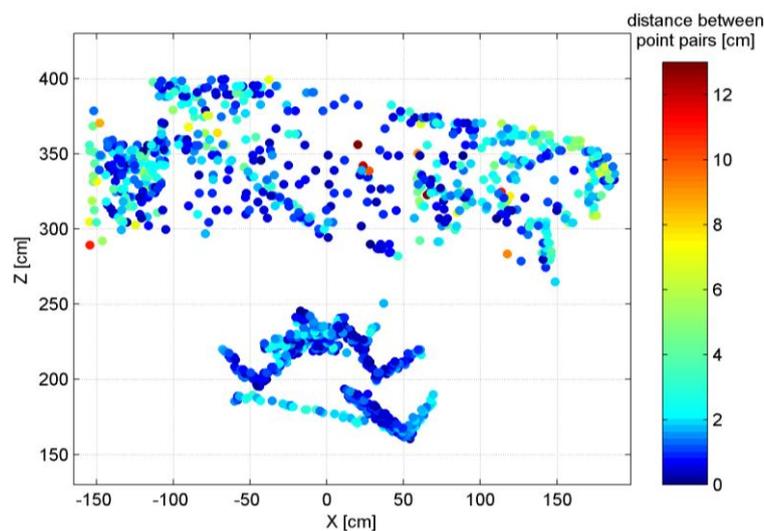


Figure 8 shows the distribution of the point pair distances in the X-Z plane. In general, points that are located further away from the sensor, particularly those at the sides of the point cloud, show larger discrepancies. This is what we expected based on the theoretical random error model. Overall, the comparison of the two point clouds shows that about 84% of the point pairs are less than 3 cm apart.

**Table 4.** Statistics of discrepancies between the closest point pairs. The last three rows show, for comparison, the same statistics obtained for an uncalibrated Kinect point cloud.

	Mean [cm]	Median [cm]	Standard deviation [cm]	Interquartile range [cm]	Percentage in [−0.5 cm, 0.5 cm]	Percentage in [−1.0 cm, 1.0 cm]	Percentage in [−2.0 cm, 2.0 cm]
dx	0.1	0.0	1.0	0.6	63.4	83.4	95.0
dy	0.0	0.0	1.1	0.6	63.4	80.7	93.2
dz	0.1	−0.1	1.8	1.3	38.9	61.6	82.1
dx *	−0.5	−0.2	1.4	1.0	55.0	74.3	90.9
dy *	−0.6	−0.1	1.5	1.1	56.8	72.7	82.9
dz *	−0.1	−0.4	1.8	1.8	25.1	51.6	81.2

**Figure 8.** Distribution of point pair distances in the X-Z plane.

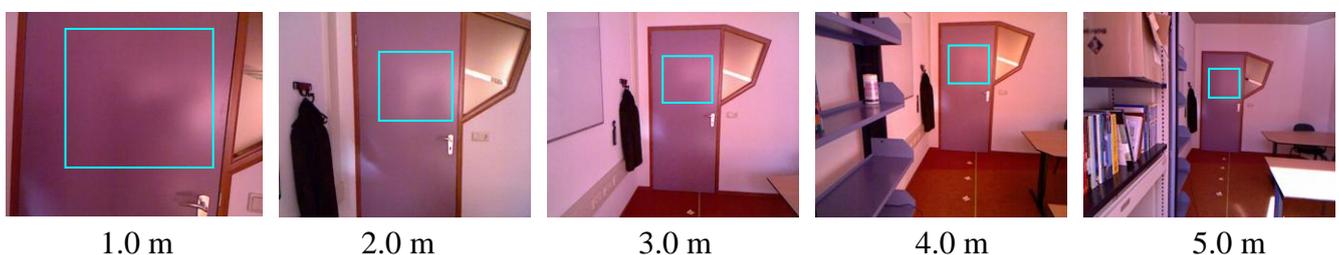


### 4.3. Plane Fitting Test

To verify the relation between the random error and the distance to the sensor a plane fitting test was carried out. The planar surface of a door was measured at various distances from 0.5 m to 5.0 m (the operation range of the sensor) with 0.5 m intervals.

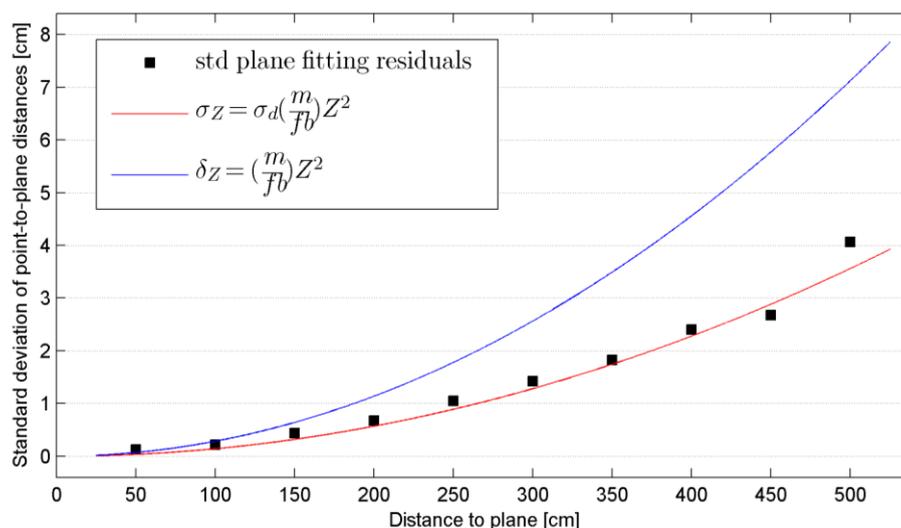
In each resulting point cloud a same part of the door was selected and a plane was fitted to the selected points. The RANSAC plane fitting method was used to avoid the influence of outliers. Figure 9 shows the measurement setup.

**Figure 9.** The planar surface of a door measured at different distances to the sensor. The boxes show the plane fitting area.



Since in all measurements the selected planar surface was approximately perpendicular to the optical axis of the sensor the residuals of the plane fitting procedure can be seen as a representation of the depth random error. To evaluate this random error at different distances an equal number of samples (4,500 samples) were randomly selected from each plane, and the standard deviation of the residuals was calculated over the selected samples. Figure 10 shows the calculated standard deviations plotted against the distance from the plane to the sensor (the black squares). It can be seen that the errors increase quadratically from a few millimeters at 0.5 m distance to about 4 cm at the maximum range of the sensor. Although the plane fitting residuals can be seen as an indication of random error of depth measurement, they are also influenced by the depth resolution at each plane. Having determined the calibration parameters we can now evaluate Equations (7) and (9) to obtain the theoretical random error and resolution of individual depth measurements at different distances from the sensor. In Figure 10, the red curve shows the theoretical random error obtained from Equation (7) with  $|m/fb| = 2.85e-5$  from the depth calibration result (Figure 5) and assuming a disparity measurement error ( $\sigma_d$ ) of  $\frac{1}{2}$  pixel. The blue curve is a plot of depth resolution obtained by evaluating Equation (9). The disparity error of  $\frac{1}{2}$  pixel seems a fair estimate since the theoretical random error curve is consistent with the observed standard deviations, considering that the low depth resolution has a minor effect on the estimate of the standard deviation of plane fitting residuals when a large number of samples are used.

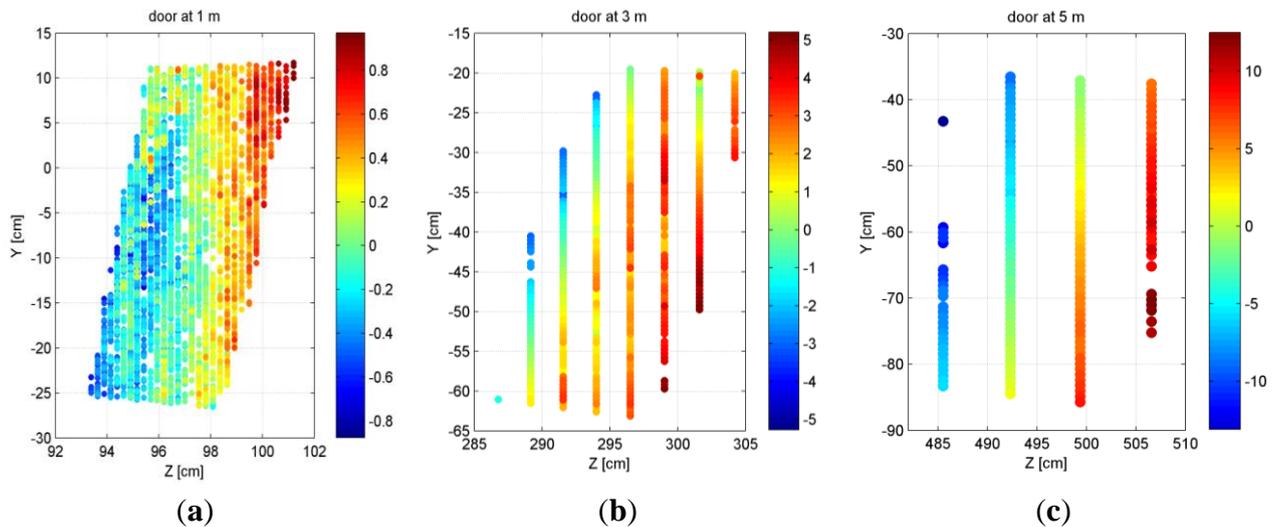
**Figure 10.** Standard deviation of plane fitting residuals at different distances of the plane to the sensor. The curves show the theoretical random error (**red**) and depth resolution (**blue**).



Although depth resolution does not have a large influence on the standard deviation of plane fitting residuals, its effect on the level of individual points should not be understated. This effect is more pronounced at larger distances from the sensor such that at the maximum range of 5 meters the point spacing in the depth direction is 7 centimeters. The combined effect of the random error and low depth resolution at large distances results in a planar surface (perpendicular to the sensor) appearing as several layers of points in the data when seen from side-view. Figure 11 shows the point clouds of the door plane at three distances projected on the Y-Z plane (Z being the depth direction; see Section 2.1 for the definition of the coordinate system). While at 1 m the point spacing in the depth direction is

quite small (about 2 mm), at 3 m and 5 m the points are clearly distributed in several layers at intervals corresponding to the depth resolution, which is about 2.5 cm for the plane at 3 m distance and close to 7 cm at 5 m.

**Figure 11.** Point cloud of a planar surface at 1 meter (a), 3 meter (b) and 5 meter (c) distance from the sensor projected on the Y-Z plane. Colors represent distance to the best-fit plane in centimeters.



## 5. Conclusions

The paper presented a theoretical and experimental analysis of the geometric quality of depth data acquired by the Kinect sensor. The geometric quality measures represent the depth accuracy and resolution for individual points. Indoor mapping applications are often based on the extraction of objects instead of an irregular set of points. In order to describe the quality of extracted objects, some basic error propagation would be needed. While fitting geometric object models to the data can reduce the influence of random errors and low depth resolution, the effect of systematic errors can only be eliminated through a proper calibration procedure.

From the results of calibration and error analysis the following main conclusions can be drawn:

- To eliminate distortions in the point cloud and misalignments between the colour and depth data an accurate stereo calibration of the IR camera and the RGB camera is necessary;
- The random error of depth measurements increases quadratically with increasing distance from the sensor and reaches 4 cm at the maximum range of 5 meters;
- The depth resolution also decreases quadratically with increasing distance from the sensor. The point spacing in the depth direction (along the optical axis of the sensor) is as large as 7 cm at the maximum range of 5 meters.

In general, for mapping applications the data should be acquired within 1–3 m distance to the sensor. At larger distances, the quality of the data is degraded by the noise and low resolution of the depth measurements.

## Acknowledgements

The authors would like to thank the Open Kinect community and in particular Nicolas Burrus of the Robotics Lab for providing an open-source interface to stream the Kinect data.

## References

1. Microsoft. Kinect. Available online: <http://www.xbox.com/en-us/kinect/> (accessed on 14 December 2011).
2. PrimeSense. Available online: <http://www.primesense.com/> (accessed on 14 December 2011).
3. Amit, B.; Dagan, E.; Gershon, K.; Alon, L.; Yinon, O.; Yaron, Y. Enhanced interactive gaming by blending full-body tracking and gesture animation. In *Proceedings of the ACM SIGGRAPH ASIA 2010 Sketches*, Seoul, Korea, 15–18 December 2010.
4. Andrew, D.W. Using a depth camera as a touch sensor. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, Saarbrücken, Germany, 7–10 November 2010.
5. Chang, Y.J.; Chen, S.F.; Huang, J.D. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Res. Dev. Disabil.* **2011**, *32*, 2566–2570.
6. Gottfried, J.M.; Fehr, J.; Garbe, C.S. Computing range flow from multi-modal Kinect data. In *Proceedings of the 7th International Symposium on Visual Computing, ISVC 2011*, Las Vegas, NV, USA, 26–28 September 2011; Volume 6938, pp. 758–767.
7. Stowers, J.; Hayes, M.; Bainbridge-Smith, A. Altitude control of a quadrotor helicopter using depth map from Microsoft Kinect sensor. In *Proceedings of the IEEE International Conference on Mechatronics, ICM 2011*, Istanbul, Turkey, 13–15 April 2011; pp. 358–362.
8. Walker, B.; Caroline, P.; William, D.S. Using depth information to improve face detection. In *Proceedings of the 6th International Conference on Human-Robot Interaction*, Lausanne, Switzerland, 6–9 March 2011.
9. Benavidez, P.; Jamshidi, M. Mobile robot navigation and target tracking system. In *Proceedings of the 6th International Conference on System of Systems Engineering: SoSE in Cloud Computing, Smart Grid, and Cyber Security, SoSE 2011*, Albuquerque, NM, USA, 2011; pp. 299–304.
10. Raptis, M.; Kirovski, D.; Hoppe, H. Real-time classification of dance gestures from skeleton animation. In *Proceedings of the 10th Annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2011*, Vancouver, BC, Canada, 5–7 August 2011; pp. 147–156.
11. Riche, N.; Mancas, M.; Gosselin, B.; Dutoit, T. 3D Saliency for abnormal motion selection: The role of the depth map. In *Proceedings of the 8th International Conference on Computer Vision Systems, ICVS 2011*, Sophia Antipolis, France, 20–22 September 2011; Volume 6962, pp. 143–152.
12. Herbst, E.; Ren, X.; Fox, D. Toward object discovery and modeling via 3-D scene comparison. In *Proceedings of IEEE International Conference on Robotics and Automation*, Shanghai, China, 9–13 May 2011.
13. Menna, F.; Remondino, F.; Battisti, R.; Nocerino, E. Geometric investigation of a gaming active device. In *Videometrics, Range Imaging, and Applications XI*; Remondino, F., Shortis, M.R., Eds.; SPIE: Munich, Germany, 2011; Volume 8085, p. 80850G.
14. Zollhöfer, M.; Martinek, M.; Greiner, G.; Stamminger, M.; Süßmuth, J. Automatic reconstruction of personalized avatars from 3D face scans. *Comput. Animat. Virtual Worlds* **2011**, *22*, 195–202.

15. Izadi, S.; Kim, D.; Hilliges, O.; Molyneaux, D.; Newcombe, R.; Kohli, P.; Shotton, J.; Hodges, S.; Freeman, D.; Davison, A.; *et al.* KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of ACM Symposium on User Interface Software and Technology*, Santa Barbara, CA, USA, 16–19 October 2011.
16. Henry, P.; Krainin, M.; Herbst, E.; Ren, X.; Fox, D. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Proceedings of International Symposium on Experimental Robotics (ISER)*, Delhi, India, 18–21 December 2010.
17. Khoshelham, K. Automated localization of a laser scanner in indoor environments using planar objects. In *Proceedings of International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zürich, Switzerland, 15–17 September 2010.
18. Rusinkiewicz, S.; Levoy, M. *Efficient Variants of the ICP Algorithm*; IEEE Computer Soc.: Los Alamitos, CA, USA, 2001; pp. 145–152.
19. MESA Imaging. Available online: <http://www.mesa-imaging.ch/> (accessed on 14 December 2011).
20. PMD [vision] CamCube 3.0. Available online: <http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmdvisionr-camcube-30/> (accessed on 14 December 2011).
21. Breuer, P.; Eckes, C.; Müller, S. Hand gesture recognition with a novel IR time-of-flight range camera—A pilot study. In *Lecture Notes in Computer Science*; Gagalowicz, A., Philips, W., Eds.; Springer: Berlin, Germany, 2007; Volume 4418, pp. 247–260.
22. Kahlmann, T.; Ingensand, H. Calibration and development for increased accuracy of 3D range imaging cameras. *J. Appl. Geod.* **2008**, *2*, 1–11.
23. Kahlmann, T.; Remondino, F.; Ingensand, H. Calibration for increased accuracy of the range imaging camera SwissRanger. In *Proceedings of ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, Dresden, Germany, 25–27 September 2006; pp. 136–141.
24. Lichti, D.D. Self-calibration of a 3D range camera. In *Proceedings of International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, Beijing, China, 3–11 July 2008; pp. 927–932.
25. Lindner, M.; Schiller, I.; Kolb, A.; Koch, R. Time-of-flight sensor calibration for accurate range sensing. *Comput. Vis. Image Underst.* **2010**, *114*, 1318–1328.
26. Shahbazi, M.; Homayouni, S.; Saadatseresht, M.; Sattari, M. Range camera self-calibration based on integrated bundle adjustment via joint setup with a 2D digital camera. *Sensors* **2011**, *11*, 8721–8740.
27. Freedman, B.; Shpunt, A.; Machline, M.; Arieli, Y. Depth Mapping Using Projected Patterns. U.S. Patent 2010/0118123, 13 May 2010.
28. Fraser, C.S. Digital camera self-calibration. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 149–159.
29. Chipworks. Teardown of the Microsoft Kinect—Focused on motion capture. Available online: <http://www.chipworks.com/en/technical-competitive-analysis/resources/recent-teardowns/2010/12/teardown-of-the-microsoft-kinect-focused-on-motion-capture/> (accessed on 14 December 2011).
30. Konolige, K.; Mihelich, P. Kinect operation. Available online: [http://www.ros.org/wiki/kinect\\_calibration/technical](http://www.ros.org/wiki/kinect_calibration/technical) (accessed on 14 December 2011).
31. Brown, D.C. Close-range camera calibration. *Photogramm. Eng.* **1971**, *37*, 855–866.
32. Faro. Laser Scanner LS 880 Techsheet. Available online: [http://faro.com/FaroIP/Files/File/Techsheets%20Download/UK\\_LASER\\_SCANNER\\_LS.pdf](http://faro.com/FaroIP/Files/File/Techsheets%20Download/UK_LASER_SCANNER_LS.pdf) (accessed on 14 December 2011).

33. Khoshelham, K.; Altundag, D.; Ngan-Tillard, D.; Menenti, M. Influence of range measurement noise on roughness characterization of rock surfaces using terrestrial laser scanning. *Int. J. Rock Mech. Mining Sci.* **2011**, *48*, 1215–1223.
34. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256.
35. Pulli, K. Multiview Registration for large data sets. In *Proceedings of Second International Conference on 3D Digital Imaging and Modeling*, Ottawa, ON, Canada, 4–8 December 1999.
36. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395.
37. Khoshelham, K.; Li, Z.L.; King, B. A split-and-merge technique for automated reconstruction of roof planes. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 855–862.
38. van der Sande, C.; Soudarissanane, S.; Khoshelham, K. Assessment of relative accuracy of AHN-2 laser scanning data using planar features. *Sensors* **2010**, *10*, 8198–8214.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).