

# Chapter 1

## 3D with Kinect

Jan Smisek, Michal Jancosek, and Tomas Pajdla

**Abstract** We analyze Kinect as a 3D measuring device, experimentally investigate depth measurement resolution and error properties, and make a quantitative comparison of Kinect accuracy with stereo reconstruction from SLR cameras and a 3D-TOF camera. We propose a Kinect geometrical model and its calibration procedure providing an accurate calibration of Kinect 3D measurement and Kinect cameras. We compare our Kinect calibration procedure with its alternatives available on Internet, and integrate it into an SfM pipeline where 3D measurements from a moving Kinect are transformed into a common coordinate system, by computing relative poses from matches in its color camera.

### 1.1 Introduction

Kinect [4, 14, 22] has become an important 3D sensor. It has received a lot of attention thanks to the rapid human pose recognition system developed on top of 3D measurement [17]. Its low cost, reliability and speed of the measurement promise to make Kinect the primary 3D measuring device in indoor robotics [25], 3D scene reconstruction [7], and object recognition [12].

In this chapter we provide a geometrical analysis of Kinect, design its geometrical model, propose a calibration procedure and demonstrate its performance. We extend here our preliminary results presented in [18].

Approaches to modeling Kinect geometry, which have appeared recently, provide a good basis for understanding the sensor. There exists the following most relevant work: The authors of [2] combined OpenCV camera calibration [24] with Kinect

---

J. Smisek (✉) · M. Jancosek · T. Pajdla  
Center for Machine Perception, Dept. of Cybernetics, FEE, Czech Technical University  
in Prague, Prague, Czech Republic  
e-mail: [smisek.jan@gmail.com](mailto:smisek.jan@gmail.com)

M. Jancosek  
e-mail: [jancom1@cmp.felk.cvut.cz](mailto:jancom1@cmp.felk.cvut.cz)

T. Pajdla  
e-mail: [pajdla@fel.cvut.cz](mailto:pajdla@fel.cvut.cz)



**Fig. 1.1** Kinect consists of infrared (IR) projector, IR camera and RGB camera (illustration from [11])

inverse disparity measurement model [3] to obtain the basic Kinect calibration procedure. The project did not study particular features of Kinect sensors and did not correct for them. An almost identical procedure [11] is implemented in ROS, where an apparent shift between the infrared and depth images is corrected. Another variation of that approach appeared in [8], where OpenCV calibration is replaced by Bouguet’s [1] calibration toolbox. We build on top of previous work and design an accurate calibration procedure based on considering geometrical models as well as on “learning” of an additional correction procedure accounting for remaining non-modeled errors. We use the full camera models and their calibration procedures as implemented in [1], the relationship between Kinect inverse disparity and depth as in [3], correct for depth and infrared image displacement as in [11], and add additional corrections trained on examples of calibration boards. We demonstrate that a calibrated Kinect can be combined with Structure from Motion to get 3D data in a consistent coordinate system allowing to construct the surface of the observed scene by Multiview Stereo. Our comparison shows that Kinect is superior in accuracy to SwissRanger SR-4000 3D-TOF camera and close to a medium resolution SLR Stereo rig. Our results are in accordance with [10] that mentions compatible observations about the Kinect depth quantization.

## 1.2 Kinect as a 3D Measuring Device

Kinect is a composite device consisting of a near-infrared laser pattern projector, an IR camera and a color (RGB) camera, Fig. 1.1. The IR camera and projector are used as a stereo pair to triangulate points in 3D space. The RGB camera can be then used to texture the 3D points or to recognize the image content. As a measuring device Kinect delivers three outputs: IR image, RGB image, and an (inverse) depth image.



Fig. 1.2 A rig with a Kinect and two Nikon D60 SLR cameras

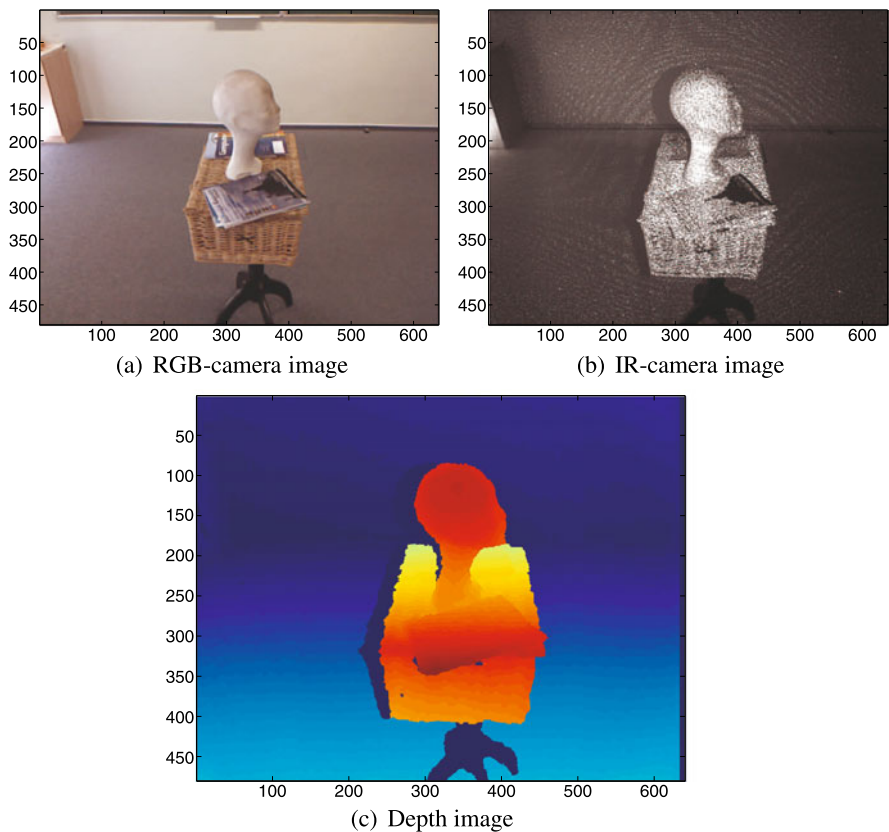


Fig. 1.3 Example of Kinect output images

### 1.2.1 IR Image

The IR camera, Fig. 1.3(b), ( $1280 \times 1024$  pixels for  $57 \times 45$  degrees FOV, 6.1 mm focal length,  $5.2 \mu\text{m}$  pixel size) is used to observe and decode the IR projection pat-

tern to triangulate the 3D scene. If suitably illuminated by a halogen lamp [19, 23] and with the IR projector blocked, Fig. 1.7(c, d), it can be reliably calibrated by [1] using the same checkerboard pattern used for the RGB camera calibration. The camera exhibits non-negligible radial and tangential distortions, see Sect. 1.4.

### 1.2.2 RGB Image

The RGB camera, Fig. 1.3(a), (1280 × 1024 pixels for 63 × 50 degrees FOV, 2.9 mm focal length, 2.8 μm pixel size) delivers medium quality images. It can be calibrated by [1] and used to track relative poses between subsequent images by using an SfM system, e.g. [6, 20].

### 1.2.3 Depth Image

The main raw output of Kinect is an 11-bit image, Fig. 1.3(c), which corresponds to the depth in the scene. Rather than providing the actual depth  $z$ , Kinect returns “inverse depth”  $1/z$ , as shown in Fig. 1.4(a). Taking into account the depth resolution achievable with a Kinect (Sect. 1.2.4), we adopted the model suggested in [11]. The depth image is constructed by triangulation from the IR image and the projector and hence it is “carried” by the IR image, as shown in Eq. 1.5.

The depth image has a vertical stripe of pixels on the right (8 pixels wide) where no depth is calculated, see Fig. 1.3(c). This is probably due the windowing effect of block correlation used in calculating the disparity [11]. We have estimated the size of the correlation window (see Sect. 1.3.1) to be  $9 \times 7$  pixels.

### 1.2.4 Depth Resolution

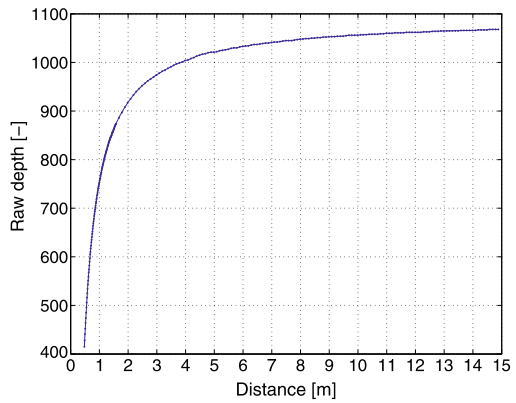
Figure 1.4(b, c) shows the resolution of the measured depth as a function of the true depth. The depth resolution was measured by moving Kinect away (0.5 m–15 m) from a planar target in sufficiently fine steps to record all the values returned in a view field of approximately  $5^\circ$  around the image center.

The size of the quantization step  $q$  [mm], which is the distance between two consecutive recorded values, was found to be the following function of the depth  $z$  [m]:

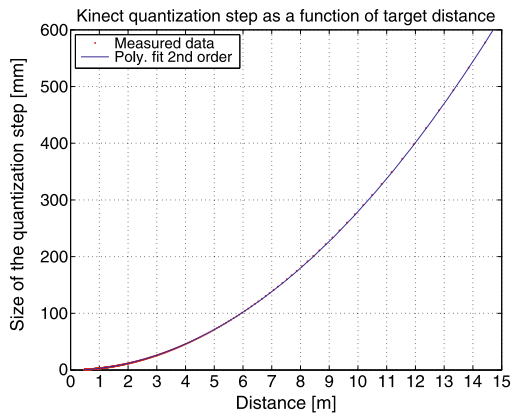
$$q(z) = 2.73 z^2 + 0.74 z - 0.58. \quad (1.1)$$

This is in accordance with the expected quadratic depth resolution for triangulation-based devices. The values of  $q$  at the beginning, resp. at the end, of the operational range were  $q(0.50 \text{ m}) = 0.65 \text{ mm}$ , resp.  $q(15.7 \text{ m}) = 685 \text{ mm}$ . These findings are in accordance with [10].

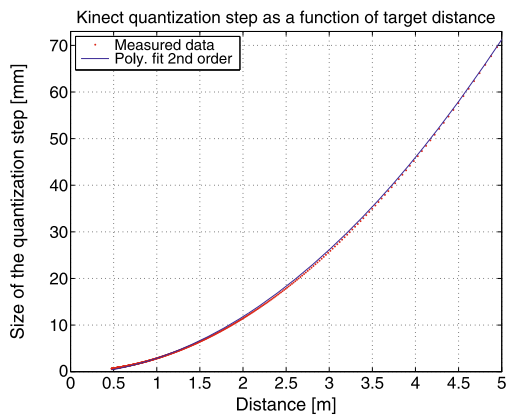
**Fig. 1.4** The estimated size of the Kinect quantization step  $q$  as a function of target distance for 0–5 m



(a) Kinect inverse depth as a function of the real depth.



(b) Kinect depth quantization step  $q$  (0-15 m).



(c) Kinect depth quantization step (0-5 m detail).

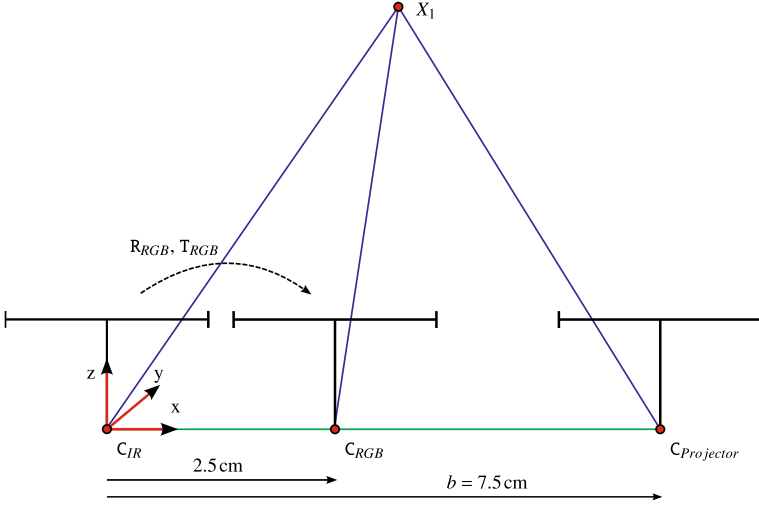


Fig. 1.5 Geometrical model of Kinect

### 1.3 Kinect Geometrical Model

We model Kinect as a multi-view system consisting of RGB, IR and Depth cameras. A Geometrical model of RGB and IR cameras, which project a 3D point  $X$  into an image point  $[u, v]^T$ , is given by [1]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \quad (1.2)$$

$$\begin{bmatrix} s \\ t \\ 1 \end{bmatrix} = \underbrace{(1 + k_1 r^2 + k_2 r^4 + k_5 r^6)}_{\text{radial distortion}} \begin{bmatrix} p \\ q \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} 2k_3 p q + k_4 (r^2 + 2p^2) \\ 2k_4 p q + k_3 (r^2 + 2q^2) \\ 1 \end{bmatrix}}_{\text{tangential distortion}} \quad (1.3)$$

$$r^2 = p^2 + q^2, \quad \begin{bmatrix} p z \\ q z \\ z \end{bmatrix} = R(X - C) \quad (1.4)$$

with distortion parameters  $k = [k_1, k_2, \dots, k_5]$ , camera calibration matrix  $K$ , rotation  $R$  and camera center  $C$  [5].

The Depth camera of Kinect is associated to the geometry of the IR camera. It returns the inverse depth  $d$  along the  $z$ -axis, as visible in Fig. 1.5, for every pixel

**Table 1.1** IR to Depth-camera pixel position shift

Circle	1	2	3	4	5	6	7	Mean
$u_0$	4.1	4.3	4.0	4.0	3.9	4.2	4.1	4.1
$v_0$	3.0	2.9	3.5	3.1	3.0	3.3	2.8	3.1

$[u, v]^\top$  of the IR cameras as

$$\begin{bmatrix} x \\ y \\ d \end{bmatrix} = \begin{bmatrix} u - u_0 \\ v - v_0 \\ \frac{1}{c_1} \frac{1}{z} - \frac{c_0}{c_1} \end{bmatrix}, \quad (1.5)$$

where  $u, v$  are given by Eq. 1.3, true depth  $z$  by Eq. 1.4,  $[u_0, v_0]^\top$  by Table 1.1,  $\mathbf{X}$  stands for 3D coordinates of a 3D point, and  $c_1, c_0$  are parameters of the model. We associate the Kinect coordinate system with the IR camera and hence get  $\mathbf{R}_{\text{IR}} = \mathbf{I}$  and  $\mathbf{C}_{\text{IR}} = 0$ . A 3D point  $\mathbf{X}_{\text{IR}}$  is constructed from the measurement  $[x, y, d]$  in the depth image by

$$\mathbf{X}_{\text{IR}} = \frac{1}{c_1 d + c_0} \text{dis}^{-1} \left( \mathbf{K}_{\text{IR}}^{-1} \begin{bmatrix} x + u_0 \\ y + v_0 \\ 1 \end{bmatrix}, \mathbf{k}_{\text{IR}} \right) \quad (1.6)$$

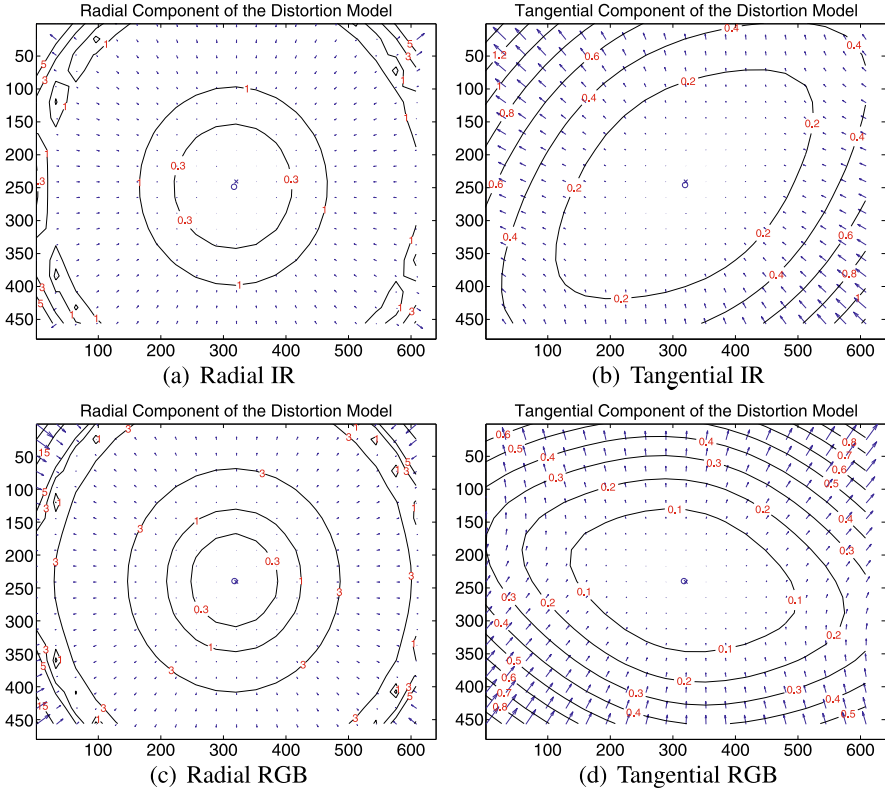
and projected to the RGB images as

$$\mathbf{u}_{\text{RGB}} = \mathbf{K}_{\text{RGB}} \text{dis}(\mathbf{R}_{\text{RGB}}(\mathbf{X}_{\text{IR}} - \mathbf{C}_{\text{RGB}}), \mathbf{k}_{\text{RGB}}) \quad (1.7)$$

where  $\text{dis}$  is the distortion function given by Eq. 1.3,  $\mathbf{k}_{\text{IR}}, \mathbf{k}_{\text{RGB}}$  are the respective distortion parameters of the IR and RGB cameras,  $\mathbf{K}_{\text{IR}}$  is the IR camera calibration matrix and  $\mathbf{K}_{\text{RGB}}, \mathbf{R}_{\text{RGB}}, \mathbf{C}_{\text{RGB}}$  are the calibration matrix, the rotation matrix and the center of the RGB camera, respectively.

### 1.3.1 Shift Between IR Image and Depth Image

IR and Depth images were found to be shifted. To determine the shift  $[u_0, v_0]^\top$ , circular targets spanning the field of view were captured from different distances in the IR and Depth images, Fig. 1.8(a). Edges of the targets were computed in the IR and Depth images using the Sobel edge detector. In order to mitigate the effect of the unstable Depth image edges, reconstruction circles were fit to the measured data, Fig. 1.8(b). The pixel distances between centers of the fitted circles are shown in Table 1.1. The shift was estimated as the mean value of the distances over all the experiments. We conclude that there is a shift of about 4 pixels in the  $u$  direction and of 3 pixels in the  $v$  direction.

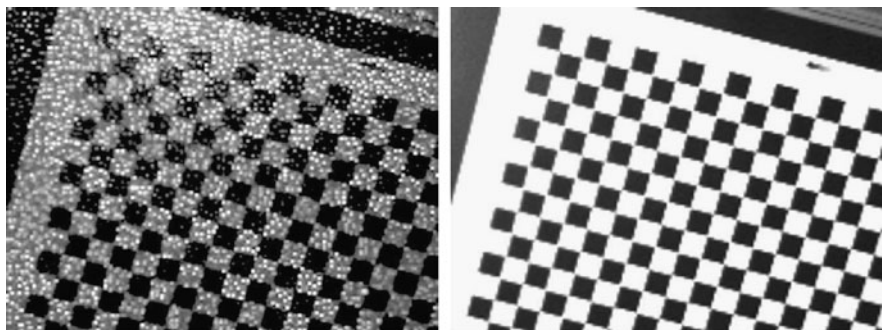


**Fig. 1.6** Estimated distortions of the Kinect cameras. The *red numbers* denote the sizes and the *arrows* denote the directions of pixel displacements induced by the lens distortion. The *cross* indicates the image center, the *circle* marks the location of the principal point

### 1.3.2 Identification of the IR Projector Geometrical Center

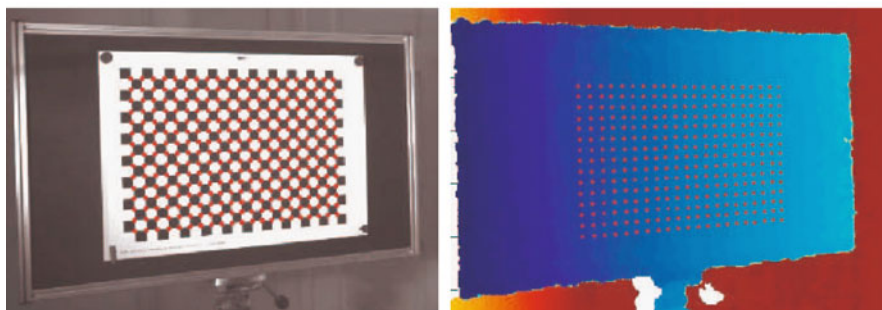
We have first acquired seven IR and Depth images of a plane positioned at different distances. The projected pattern contains nine brighter and easily identifiable speckle dots, Fig. 1.9(a). These points were formed by  $r = 1, \dots, 9$  rays  $l_r$  transmitted from the IR projector. Each point was reconstructed in the 3D space and grouped by its ray of origin  $X_{IR_{i,r}}$ . The IR projector center  $C_P$  is located in the common intersection of the nine rays. We formulated a nonlinear optimization problem to find the projector center  $C_P$  by minimizing the perpendicular distances of the reconstructed points  $X_{IR_{i,r}}$  from a bundle of rays passing through  $C_P$ . Figure 1.9(b) shows the resulting ray bundle next to the IR cameras frame. Figure 1.9(c) shows the residual distances from the points  $X_{IR_{i,r}}$  to their corresponding rays of the optimal ray bundle. All residual distances are smaller than 2 mm. The estimated projector center has coordinates  $C_P = [74.6, 1.1, 1.3]^T$  [mm] in the IR camera reference frame.





(a) IR image of a calibration checkerboard

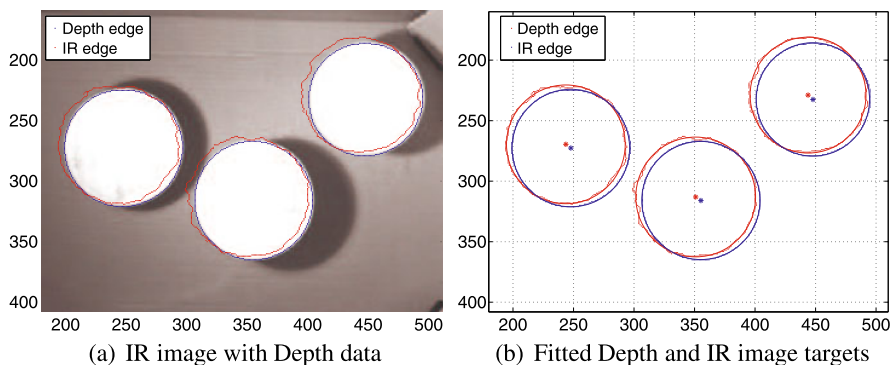
(b) IR image of the calibration checkerboard illuminated by a halogen lamp with the IR projection blocked.



(c) Calibration points extracted in the RGB image.

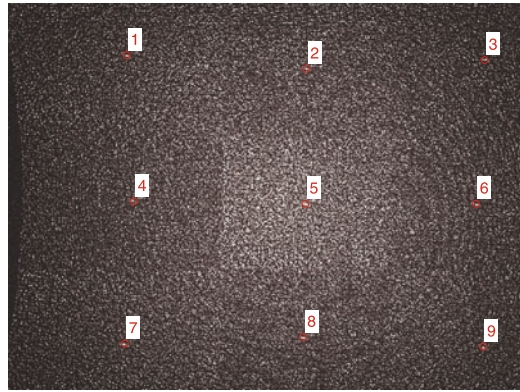
(d) Calibration points in the Depth image.

**Fig. 1.7** The calibration board in the IR, RGB and Depth images

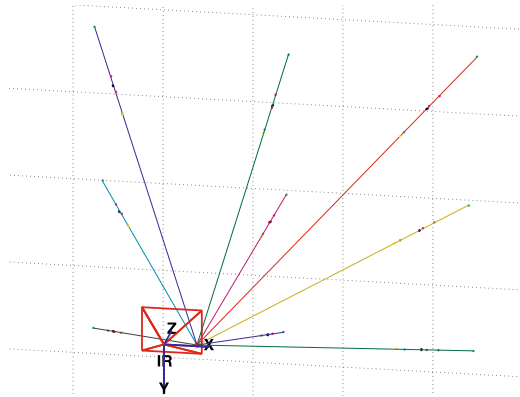


**Fig. 1.8** Illustration of the IR to Depth image shift

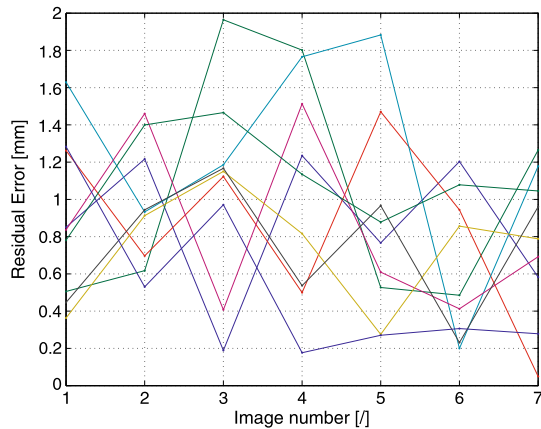
**Fig. 1.9** Identification of the geometrical model



(a) IR image with 9 extracted brighter dots

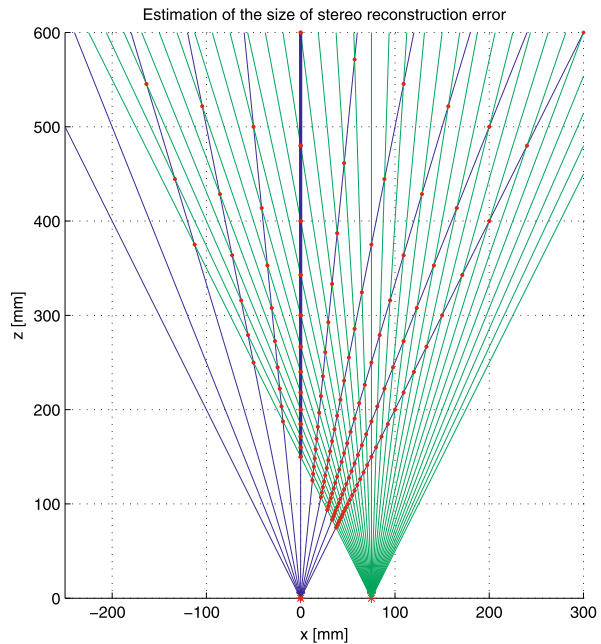


(b) Estimated IR projector geometrical position



(c) Residual error between reconstructed points and rays passing through them to estimated projector position

**Fig. 1.10** Kinect IR camera (*blue*) and projector (*green*) view fields and ray distribution in the  $x$ - $y$  plane estimated in Sect. 1.3.3. For clarity, we plot only every 64th camera ray, i.e. there are 11 rays for the IR camera, and every 150th projector ray, i.e. there are 32 projector rays. *Red dots* illustrate the sampling of the space by points that can be reconstructed. The *bold blue line* marks the center ray of the IR camera where the distance resolution shown in Fig. 1.11 was estimated. Note that the closest point, which is actually measured by the real device, is at a depth of about 40 cm



### 1.3.3 Identification of Effective Depth Resolutions of the IR Camera and Projector Stereo Pair

In this section the view fields of the Kinect IR camera and of the Kinect projector and their effective resolution, which determines the distribution of the resolution in 3D measurement, will be investigated.

The size of the IR image and of the depth image is known to be  $640 \times 480$  pixels with  $10.4 \mu\text{m}$  pixel size, spanning approx.  $60^\circ \times 45^\circ$  view angle. This gives an angular resolution of  $0.0938^\circ/\text{pixel}$  in the IR camera.

Counting the speckle dots on the projected pattern yields about 800 dots along the central horizontal line across the projector field of view. Projector FOV and IR camera FOV are approximately the same. Hence we get 800 dots per  $60^\circ$  and an angular resolution of  $0.0750^\circ/\text{ray}$  for the projector rays. The green curve in Fig. 1.11 shows the simulated depth quantization along the central IR camera ray (the blue line in Fig. 1.10) for the camera and projector resolution described above. It clearly does not correspond to the red curve measured on a real Kinect.

To get our simulation closer to reality, we assume that ray detection is done with higher accuracy by interpolating rays from the projected patterns. The blue curve in Fig. 1.11 corresponds to detecting rays with  $1/8$  pixel accuracy, as was hypothesized in [11]. Hence we get the effective resolution of  $5120 = 640 \times 8$  rays per  $60^\circ$ , i.e.  $0.00938^\circ/\text{ray}$ , in the projector. This corresponds to our measurement on a real Kinect.

**Fig. 1.11** Comparison of stereo reconstruction uncertainty measured with Kinect and simulated using identified parameters of the stereo system

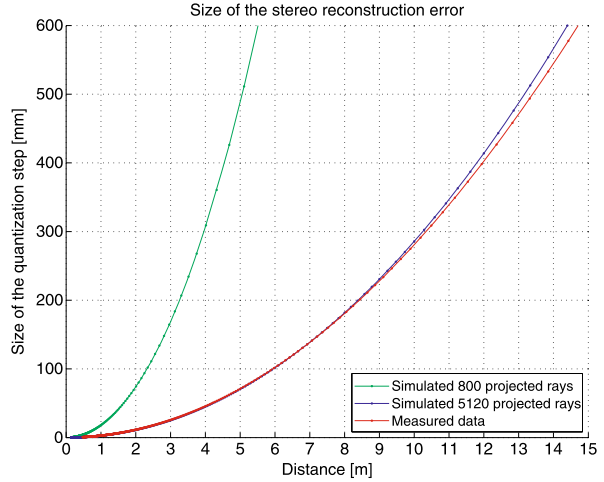


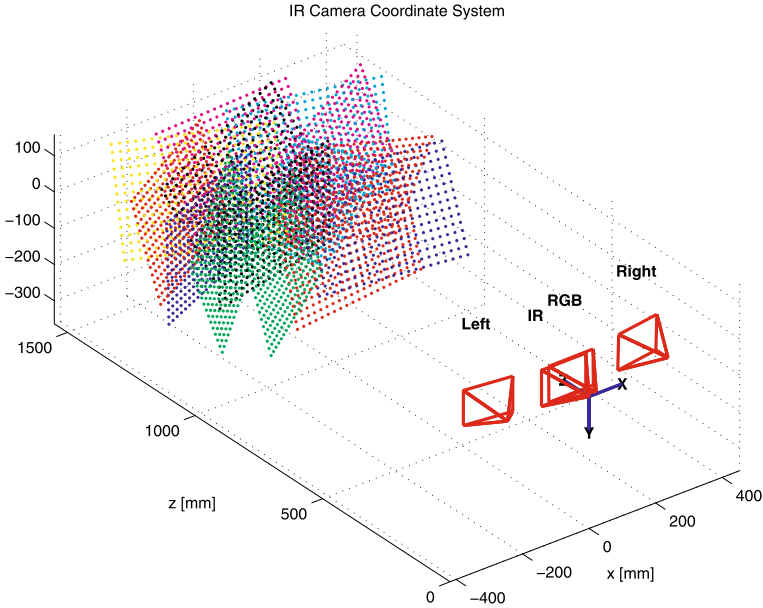
Figure 1.10 illustrates view fields and ray arrangements for Kinect IR camera (blue) and projector (green). The bold blue line marks the center of the IR camera view where the distance resolution was evaluated. For clarity, we show only every 64th camera ray and every 150th projector ray and their intersections as red dots.

## 1.4 Kinect Calibration

We calibrate, as proposed in [1], Kinect cameras together by showing the same calibration target to the IR and RGB cameras, Fig. 1.7(c). This allows to calibrate both cameras w.r.t. the same 3D points and hence the poses of the cameras w.r.t. the points can be chained to give their relative pose, Fig. 1.12. Taking the Cartesian coordinate system of the IR camera as the global Kinect coordinate system makes the camera relative pose equal to  $\mathbb{R}_{\text{RGB}}, \mathbb{C}_{\text{RGB}}$ .

Tables 1.2 and 1.3 show the internal parameters and Fig. 1.6 shows the effect of distortions in the cameras. We included the tangential distortion since it non-negligibly increased the overall accuracy of 3D measurements. Figure 1.7(a) shows the IR image of the calibration board under the normal Kinect operation when it is illuminated by its IR projector. A better image is obtained by blocking the IR projector and illuminating the target by a halogen lamp Fig. 1.7(b).

Parameters  $c_0, c_1$  of the Depth camera are calibrated as follows: We get  $n$  measurements  $X_{D_i} = [x_i, y_i, d_i]^\top$ ,  $i = 1, \dots, n$ , of all the calibration points from the depth images, Fig. 1.7(d). The Cartesian coordinates  $X_{\text{IR}_i}$  of the same calibration points were measured in the IR Cartesian system by intersecting the rays projecting the points into IR images with the best plane fits to the reconstructed calibration points. Parameters  $c_0, c_1$  were optimized to best fit  $X_{D_i}$  to  $X_{\text{IR}_i}$  using Eq. 1.6.



**Fig. 1.12** Position and orientation of Kinect IR and RGB cameras and the SLR stereo pair (*Left, Right*) altogether with 3D calibration points reconstructed on planar calibration targets

**Table 1.2** Intrinsic parameters of the Kinect IR camera

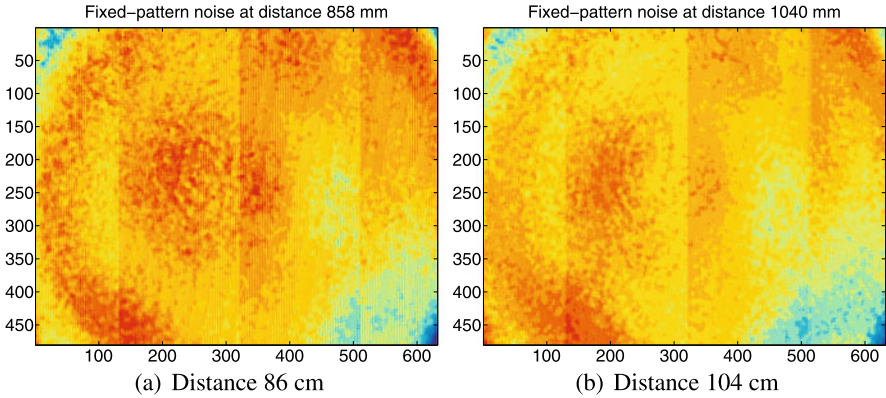
Focal length		Principal point		Distortion coefficients			
f [px]	f [mm]	$x_0$ [px]	$y_0$ [px]	$k_{c1}$	$k_{c2}$	$k_{c3}$	$k_{c4}$
585.6	6.1	316	247.6	-0.1296	0.45	-0.0005	-0.002

**Table 1.3** Intrinsic parameters of the Kinect RGB camera

Focal length		Principal point		Distortion coefficients			
f [px]	f [mm]	$x_0$ [px]	$y_0$ [px]	$k_{c1}$	$k_{c2}$	$k_{c3}$	$k_{c4}$
524	2.9	316.7	238.5	0.2402	-0.6861	-0.0015	0.0003

### 1.4.1 Learning Complex Residual Errors

It has been observed that a Kinect calibrated with the above procedure still exhibited small but relatively complex residual errors for the close range measurements. Figure 1.13 shows residuals after fitting the plane to the calibrated Kinect measurement of a plane spanning the field of view. The target has been captured from 18 different distances ranging from 0.7 to 1.3 meters. Highly correlated residuals were accounted.



**Fig. 1.13** Residuals of the plane fitting showing the fixed-pattern noise on depth images from different distances

**Table 1.4** Evaluation of the  $z$ -correction. The standard deviation of the residuals of the plane fit to the measurement of a planar target has been reduced

Data-set	Standard deviation [mm]	
	Original $\sigma$	Corrected $\sigma$
Even images	2.18	1.54
Odd images	1.98	1.34

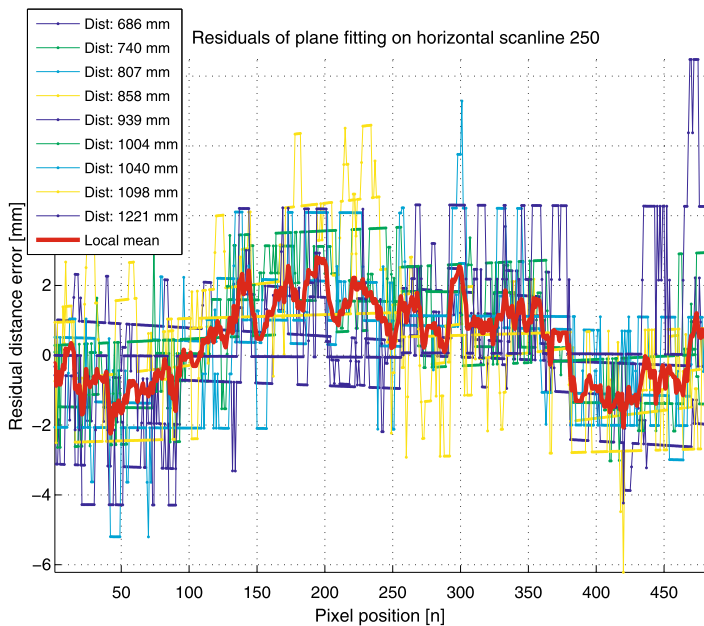
Residuals along the 250th horizontal Depth image row are shown in Fig. 1.14(a). Note that the residual values do not depend on the actual distance to the target plane (in this limited range). The values are consistently positive in the center and negative at the periphery. To compensate for this residual error, we form a  $z$ -correction image of  $z$  values constructed as the pixel-wise mean of all residual images. The  $z$ -correction image is subtracted from the  $z$  coordinate of  $X_{IR}$  computed by Eq. 1.6.

To evaluate this correction method, the  $z$ -correction image was constructed from residuals of even images and then applied to odd (the first row of Table 1.4) and to even (the second row of Table 1.4) depth images. The standard deviation of residuals decreased.

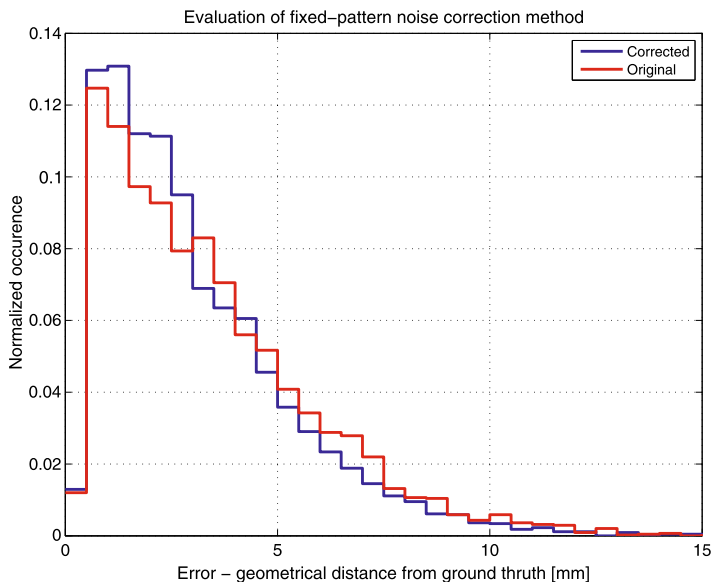
After applying the  $z$ -correction to Kinect measurements from the experiment described in Sect. 1.5.2, the mean of the residual errors decreased by approximately 0.25 mm, Fig. 1.14(b). The residuals were evaluated on 4410 points spanning the field of view.

## 1.5 Validation

In this section, different publicly available Kinect depth models are tested and compared to our method on a 3D calibration object. Furthermore, we provide a com-



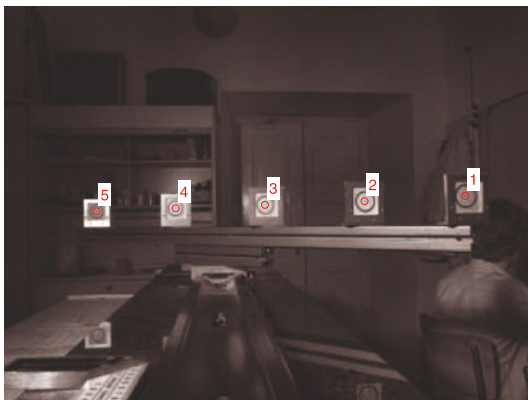
(a) Residuals of plane fitting on the 250<sup>th</sup> horizontal row in the center of the Depth image. The local mean is shown as a solid red line.



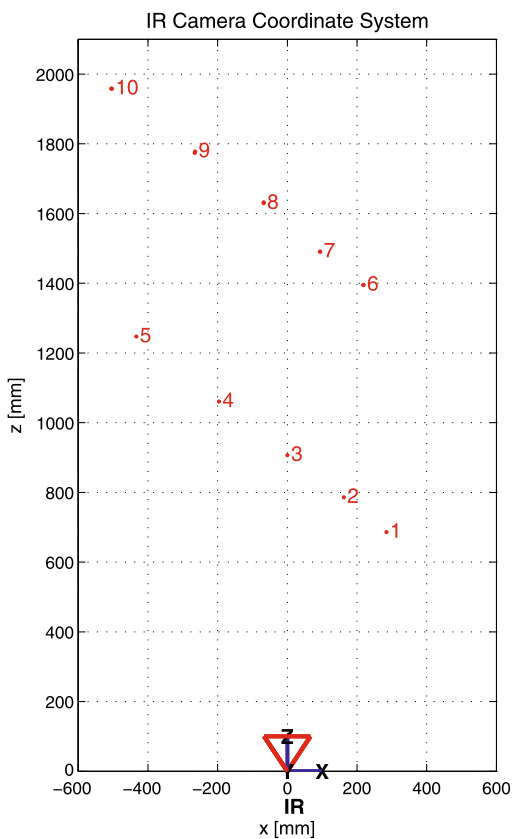
(b) Normalized histogram of the residual errors of calibrated Kinect with (blue) and without (red) complex error correction.

Fig. 1.14 Correcting complex residual errors

**Fig. 1.15** Kinect accuracy evaluation on a 3D reference object with five flat targets mounted on a rigid bench



(a) IR image



(b) 3D reconstruction

parison of the accuracy of Kinect measurements against stereo triangulation and 3D measurements based on Time-of-Flight. Finally, we demonstrate the functionality of our Kinect calibration procedure by integrating it into an SfM pipeline.



**Table 1.5** Accuracy evaluation of different reconstruction methods on a reference 3D object. Kinect 1 is the device for which we made complete calibration as described in this chapter. Kinect 2 was evaluated with the calibration from Kinect 1

Device	Method	Distance difference $d$ [mm]										$\mu(d)$	$\sigma(d)$
Kinect 1	Our	1.17	3.52	-1.98	-2.56	1.73	-7.09	-1.31	5.21	<b>-0.16</b>	<b>3.89</b>		
	ROS	2.69	5.69	1.22	1.91	4.21	-3.33	3.44	11.33	3.39	4.17		
	Burrus	10.64	15.29	12.03	13.41	9.84	3.28	10.49	18.28	11.66	4.41		
	Tangent	3.15	6.19	1.95	3.35	5.30	-1.45	6.02	15.00	4.94	4.78		
	OpenNI	-3.08	9.37	-0.69	-12.12	2.55	-5.85	0.59	11.99	0.34	7.82		
	Kinect SDK	N.A. <sup>a</sup>	N.A.	-2.62	-6.98	7.10	-12.22	5.98	1.99	-1.12	7.58		
Kinect 2	Our	2.09	9.90	-6.49	-11.82	2.81	2.58	0.47	-7.31	-0.97	7.02		
	ROS	3.16	11.32	-4.32	-8.82	4.38	4.73	3.37	-3.38	1.30	6.38		
	Burrus	10.97	20.56	6.33	2.38	9.68	11.04	9.85	2.86	9.21	5.75		
	Tangent	3.47	11.77	-3.58	-7.10	5.58	6.55	6.19	0.96	2.98	<b>6.05</b>		
	OpenNI	-1.14	12.22	-9.12	-11.45	5.24	6.09	-7.40	2.68	<b>-0.36</b>	8.37		
	Kinect SDK	N.A.	N.A.	-2.36	-7.83	7.00	-0.47	-2.97	12.22	0.93	7.34		

<sup>a</sup>Kinect SDK currently limits the measurement range to 0.8–4 m

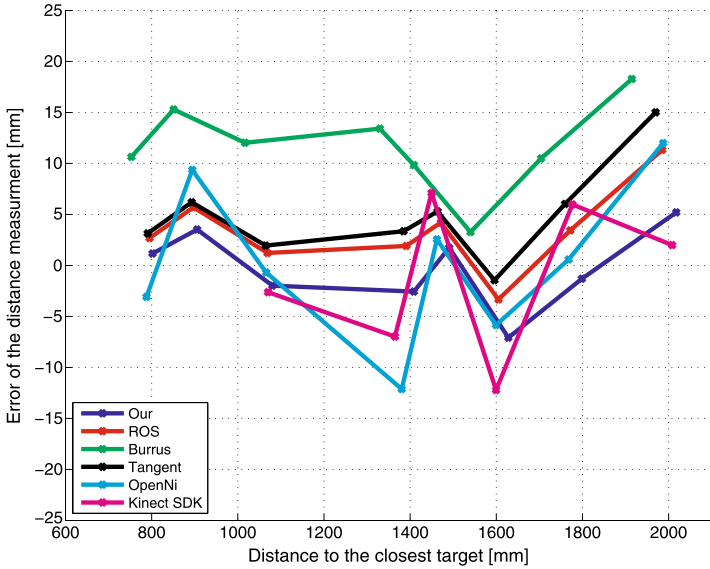
### 1.5.1 Kinect Depth Models Evaluation on a 3D Calibration Object

We evaluate the accuracy of the calibration by measuring a reference 3D object. The 3D object consisted of five flat targets that were rigidly mounted together along a straight line on a rigid bench, Fig. 1.15(a). As ground truth, the distances between centers of the targets were carefully measured by a measure tape with accuracy better than 1 mm.

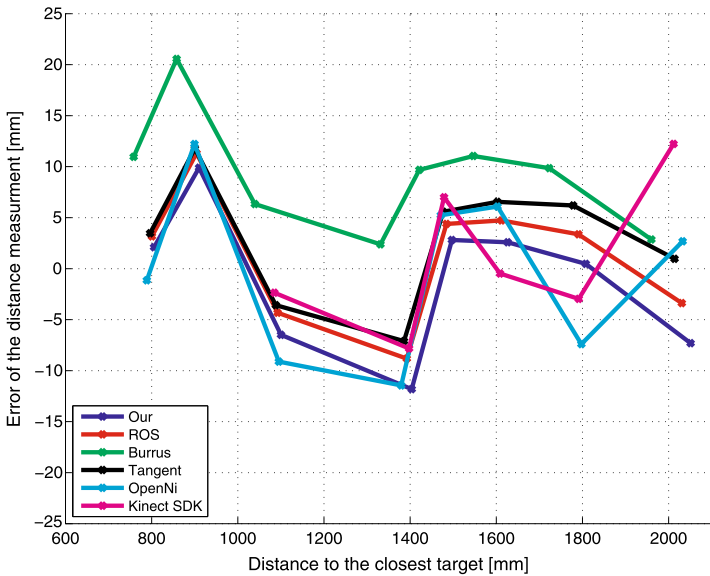
The object was then captured using Kinect from two different distances to get measurements in the range between 0.7 m to 2 m, Fig. 1.15(b). After extracting the central points of the targets in the IR image, Fig. 1.15(a), several different reconstruction methods were used to get their 3-dimensional positions, Fig. 1.15(b).

Our Kinect calibration model, which was described in Sect. 1.4, was compared to the ROS calibration [11], Burrus calibration [2], Magnenat calibration [21], OpenNi calibration [16] and Microsoft Kinect SDK calibration [15].

Distances between the reconstructed target points were compared to the ground truth measurements in Table 1.5 and in Fig. 1.16. The experiment was performed on two Kinect devices. Kinect 1 is the device for which the complete calibration, as described in this chapter, was made. Kinect 2 was evaluated with the calibration from Kinect 1, to determine whether it is possible to transfer calibration parameters of one device to another. We see that our method is the best for Kinect 1 and among the best three for Kinect 2.



(a) Kinect 1



(b) Kinect 2

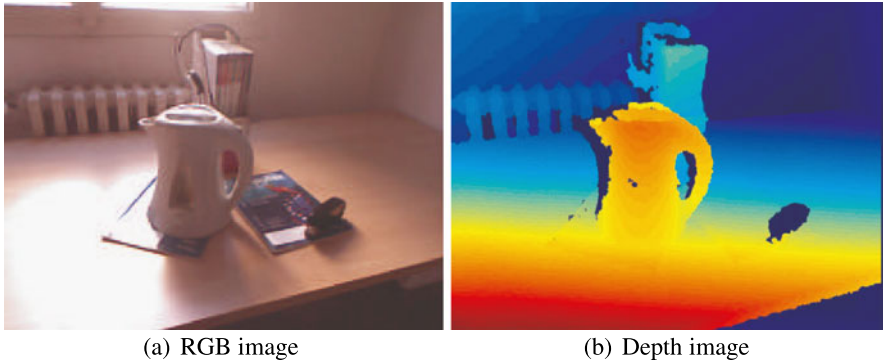
Fig. 1.16 Accuracy evaluation of different reconstruction methods on a 3D calibration object

### 1.5.2 Comparison of Kinect, SLR Stereo and 3D TOF

We have compared the accuracy of Kinect, SLR Stereo and 3D TOF cameras on the measurements of planar targets: Kinect and SLR Stereo (image size  $2304 \times 1536$

**Table 1.6** Comparison of SLR Stereo triangulation, Kinect and SR-4000 3D TOF depth sensing

Method	Geometrical error $e$ [mm]		
	$\mu(e)$	$\sigma(e)$	$\max(e)$
SLR Stereo	1.57	1.15	7.38
Kinect	2.39	1.67	8.64
SR-4000	27.62	18.20	133.85

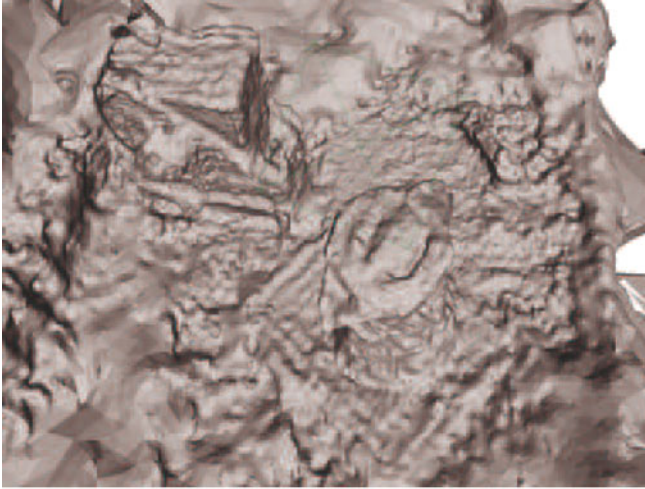


**Fig. 1.17** Example of images from Kinect RGB cameras and the corresponding depth that were used for scene reconstruction

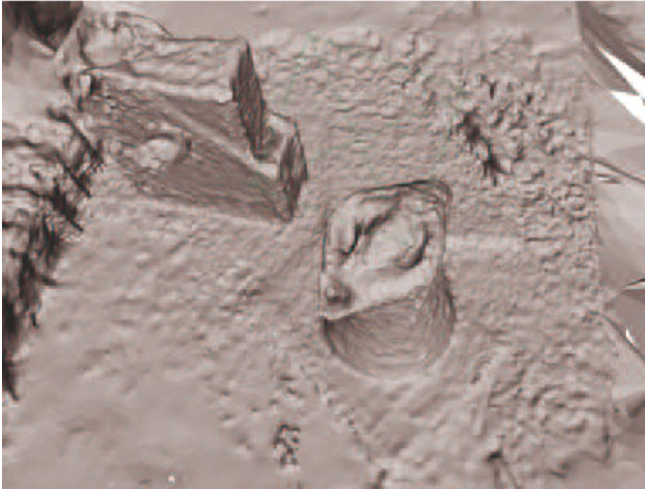
pixels) were rigidly mounted (Fig. 1.2) and calibrated (Fig. 1.12) together. SLR Stereo was performed by reconstructing calibration points extracted by [1] and triangulated by the linear least squares triangulation [5]. They measured the same planar targets in 315 control calibration points on each of the 14 targets. SR-4000 3D TOF [13] measured different planar targets but in a comparable range of distances 0.9–1.4 meters from the sensor in 88 control calibration points on each of the 11 calibration targets. The error  $e$ , Table 1.6, corresponds to the Euclidean distance between the points returned by the sensors and points reconstructed in the process of calibration of the cameras of the sensors. SLR Stereo is the most accurate, Kinect follows and SR-4000 is the least accurate.

### 1.5.3 Combining Kinect and Structure from Motion

Figure 1.17 shows a pair of 1/2-resolution ( $640 \times 480$ ) Kinect RGB and depth images (where the original depth image was reprojected using Eq. 1.7 to correspond with the RGB image pixels). A sequence of 50 RGB-Depth image pairs has been acquired and the relative poses of the RGB cameras have been computed by a SfM pipeline [6, 20]. Figure 1.18(a) shows a surface reconstructed from 3D points obtained by mere Multiview stereo [9] using only Kinect RGB images. Utilizing



(a) Only visual data were used to reconstruct the scene by [10].

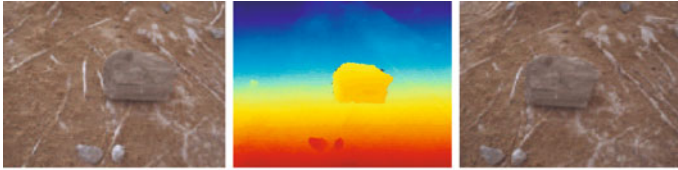


(b) Improved reconstruction using Kinect depth data registered by SfM applied to Kinect RGB images.

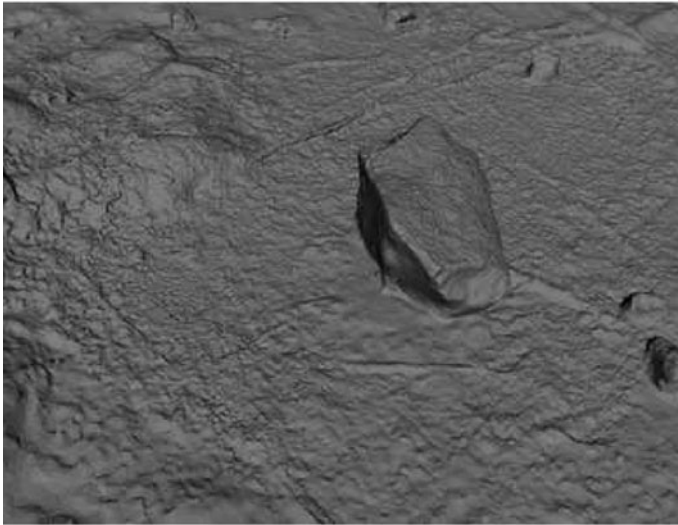
**Fig. 1.18** Scene reconstruction from Kinect RGB camera. The figure shows a comparison of reconstruction quality when the scene is reconstructed only using *Multiview stereo* and the case when the 3D data from Kinect are also available

retrieved relative poses, depth data were registered together and used in the same method to provide improved reconstruction, Fig. 1.18(b).

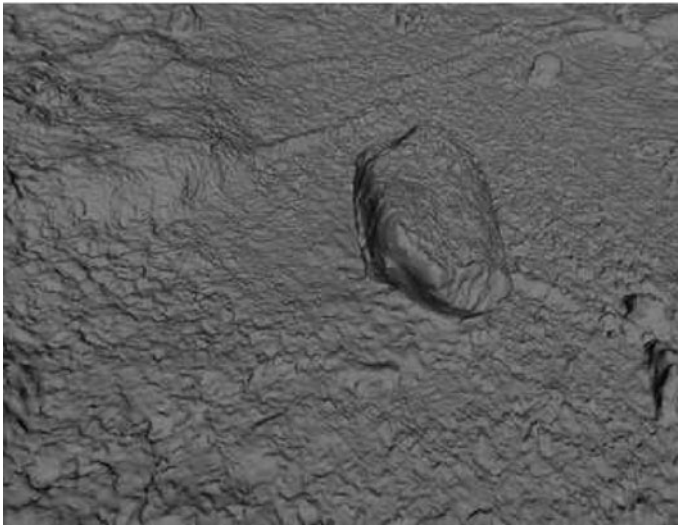
Figure 1.19 compares a 3D surface reconstruction from point cloud computed by plane sweeping [9] with 70 Kinect 3D data processed by surface reconstruction of [9] ( $2304 \times 1536$  pixels). Kinect 3D data were registered into a common coordinate system via SfM [6, 20] applied to Kinect image data. We see that when multiple



(a) Left SLR image. (b) Kinect true depth. (c) Right SLR image.



(d) Multiview reconstruction [10].



(e) Reconstruction from registered Kinect 3D data.

**Fig. 1.19** Comparison of Kinect with Multiview reconstruction [9]

measurements are used, the Kinect result is quite comparable to more accurate Multiview stereo reconstruction.

## 1.6 Conclusion

We have provided an analysis of Kinect 3D measurement capabilities and its calibration procedure allowing to combine Kinect with SfM and Multiview Stereo, which opens a new area of applications for Kinect. It was interesting to observe that in the quality of the multi-view reconstruction, Kinect over-performed SwissRanger SR-4000 and was close to 3.5 M pixel SLR Stereo.

**Acknowledgements** This research was supported by TA02011275—ATOM—Automatic Three-dimensional Terrain Monitoring and FP7-SPACE-241523 ProViScout grants.

## References

1. Bouguet, J.Y.: Camera calibration toolbox. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/) (2010)
2. Burrus, N.: Kinect calibration. <http://nicolas.burrus.name/index.php/Research/KinectCalibration> (2010)
3. Dryanovski, I., Morris, W., Magnenat, S.: kinect\_node. [http://www.ros.org/wiki/kinect\\_node](http://www.ros.org/wiki/kinect_node) (2010)
4. Freedman, B., Shpunt, A., Machline, M., Arieli, Y.: Depth mapping using projected patterns. US Patent (2010)
5. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2003)
6. Havlena, M., Torii, A., Pajdla, T.: Efficient structure from motion by graph optimization. doi:10.1007/978-3-642-15552-9\_8
7. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: RGB-D mapping: using Kinect-style depth cameras for dense 3d modeling of indoor environments. *Int. J. Robot. Res.* (2012). doi:10.1177/0278364911434148
8. Herrera, D.C., Kannala, J., Heikkila, J.: Accurate and practical calibration of a depth and color camera pair. [http://www.ee.oulu.fi/~dherrera/kinect/2011-depth\\_calibration.pdf](http://www.ee.oulu.fi/~dherrera/kinect/2011-depth_calibration.pdf) (2011)
9. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
10. Khoshelham, K.: Accuracy analysis of Kinect depth data. In: ISPRS Workshop Laser Scanning, vol. XXXVIII (2011)
11. Konolige, K., Mihelich, P.: Technical description of Kinect calibration. [http://www.ros.org/wiki/kinect\\_calibration/technical](http://www.ros.org/wiki/kinect_calibration/technical) (2011)
12. Lai, K., Bo, L., Ren, X., Fox, D.: Sparse distance learning for object recognition combining RGB and depth information. In: IEEE International Conference on Robotics and Automation (2011)
13. MESA Imaging: SwissRanger SR-4000. <http://www.mesa-imaging.ch/> (2011)
14. Microsoft: Kinect for X-BOX 360. <http://www.xbox.com/en-US/kinect> (2010)
15. Microsoft: Kinect for Windows. <http://www.kinectforwindows.org/> (2012)
16. Openni: <http://openni.org/> (2011)

17. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from a single depth image. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
18. Smisek, J., Jancosek, M., Pajdla, T.: 3D with Kinect. In: International Conference on Computer Vision—Workshop on Consumer Depth Cameras for Computer Vision (2011)
19. Smisek, J., Pajdla, T.: 3D camera calibration. M.Sc. thesis, Czech Technical University in Prague (2011)
20. Snavely, N., Seitz, S., Szeliski, R.: Modeling the world from internet photo collections. *Int. J. Comput. Vis.* (2007)
21. Magnenat, S.: Stéphane Magnenat’s distance model. [http://groups.google.com/group/openkinect/browse\\_thread/thread/31351846fd33c78/e98a94ac605b9f21](http://groups.google.com/group/openkinect/browse_thread/thread/31351846fd33c78/e98a94ac605b9f21) (2011)
22. Wikipedia: Kinect. <http://en.wikipedia.org/wiki/Kinect>
23. Willow Garage: ROS—Kinect calibration: code complete. <http://www.ros.org/news/2010/12/kinect-calibration-code-complete.html> (2010)
24. Willow Garage: Camera calibration and 3D reconstruction. [http://opencv.willowgarage.com/documentation/cpp/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://opencv.willowgarage.com/documentation/cpp/camera_calibration_and_3d_reconstruction.html) (2011)
25. Willow Garage: Turtlebot. <http://www.willowgarage.com/turtlebot> (2011)