3D with Kinect

Jan Smisek, Michal Jancosek and Tomas Pajdla
CMP, Dept. of Cybernetics, FEE, Czech Technical University in Prague
{smiseja1,jancom1,pajdla}@fel.cvut.cz

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 Kinect geometrical model and its calibration procedure providing an accurate calibration of Kinect 3D measurement and Kinect cameras. We demonstrate the functionality of Kinect calibration by integrating 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 color camera.

1. Introduction

Kinect[1] [5, 14, 18] is becoming an important 3D sensor. It is receiving a lot of attention thanks to rapid human pose recognition system developed on top of 3D measurement [15]. The low cost, reliability and speed of the measurement promises to make Kinect the primary 3D measuring devices in indoor robotics [21], 3D scene reconstruction [1], and object recognition [12].

In this paper we provide a geometrical analysis of Kinect, design its geometrical model, propose a calibration procedure, and demonstrate its performance.

Approaches to modeling Kinect geometry, which appeared recently, provide a good basis for understanding the sensor. There are the following most relevant works. Work [3] combined OpenCV camera calibration [20] with Kinect inverse disparity measurement model [4] to obtain the basic Kinect calibration procedure. He did not study particular features of Kinect sensors and did not correct for them. 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 [2] calibration toolbox. We build on top of the 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 [2], the relationship between Kinect inverse disparity and depth as in [4], correct for depth and infrared image displacement as in [11], and add additional correction trained on examples of calibration boards. We demonstrated that calibrated Kinect can be combined with Structure from Motion to get 3D data in a consistent coordinate system allowing to construct 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] where compatible observations about Kinect depth quantization were mentioned.

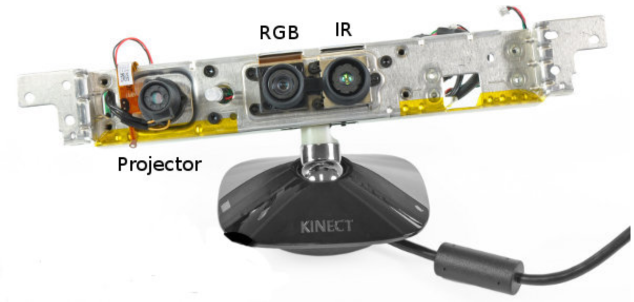


Figure 1. Kinect consists of Infra-red (IR) projector, IR camera and RGB camera (illustration from [11]).

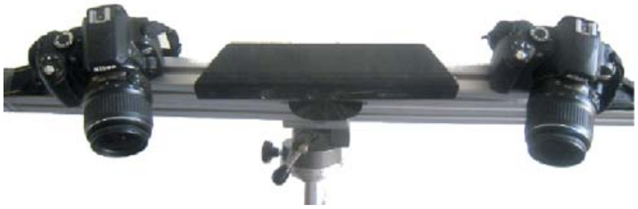


Figure 2. Kinect and two Nikon D60 SLR cameras rig.

[1] This research was supported by grants ME CR MSM 6840770038, FP7-SPACE-241523 ProViScout, and by EU SpaceMaster programme.

2. Kinect as a 3D measuring device

Kinect is a composite device consisting of an IR projector of a pattern and IR camera, which are used to triangulate points in space. It works as a depth camera, and a color (RGB) camera, which can be used to recognize image content and texture 3D points, Fig 1. As a measuring device, Kinect delivers three outputs: IR image, RGB image, and (inverse) Depth image.

2.1. IR image

IR (1280×1024 pixels for 57×45 degrees FOV, 6.1 mm focal length, $5.2\,\mu m$ pixel size) camera is used to observe and decode the IR projection pattern to triangulate 3D scene. If suitably illuminated by a halogen lamp [16, 19] and with the IR projector blocked, Fig. 7(c, d), it can be reliably calibrated by [2] using the same checkerboard pattern used for the RGB camera. The camera exhibits non-negligible radial and tangential distortions, Tab. 2.

2.2. RGB image

RGB (1280×1024 pixels for 63×50 degrees FOV, 2.9 mm focal length, $2.8\,\mu m$ pixel size) camera delivers medium quality images. It can be calibrated by [2] and used to track the camera motion by SfM systems, e.g. [17, 7].

2.3. Depth image

The main raw output of Kinect is an image that corresponds to the depth in the scene. Rather than providing the actual depth z, Kinect returns "inverse depth" d, Fig. 3(a). Taking into account the depth resolution achievable with Kinect (section. 2.4), we adopted the model Eqn. 5 suggested in [3]. The Depth image is constructed by triangulation from the IR image and the projector and hence it is "carried" by the IR image, Eqn. 5.

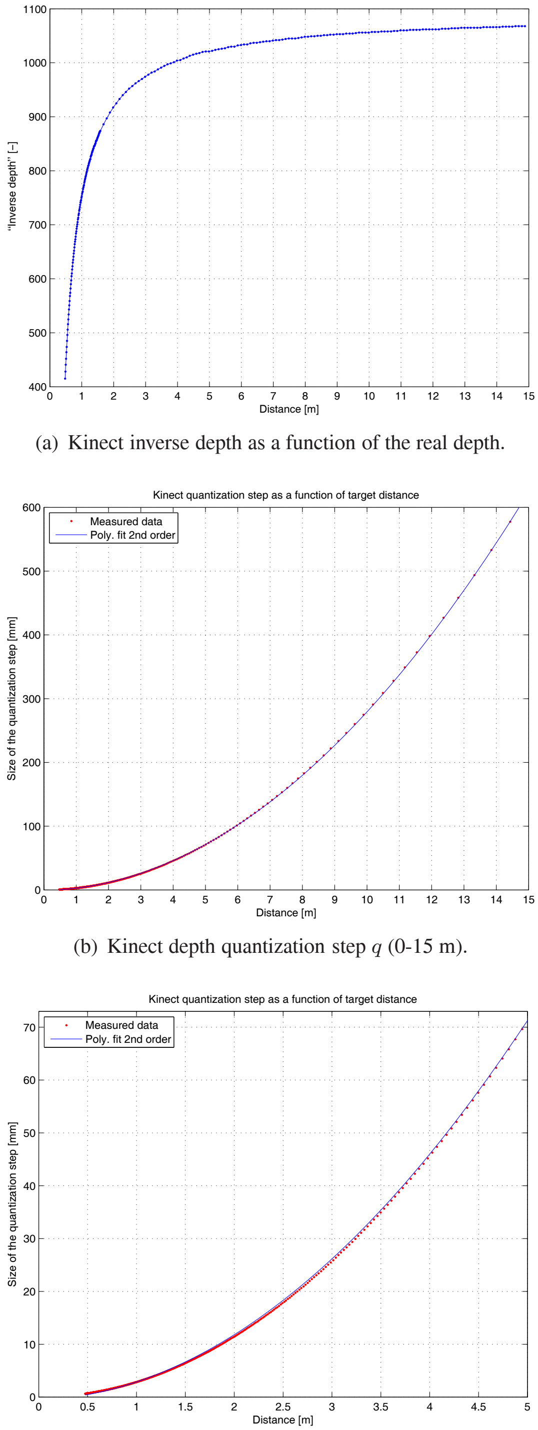
2.4. Depth resolution

Fig. 3(b,c) show the resolution in depth as a function of the distance. The depth resolution was measured by moving Kinect away (0.5 m–15 m) from a planar target sufficiently finely to record all values returned in approximately 5° view field around the image center.

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

$$q(z) = 2.73\,z^2 + 0.74\,z - 0.58 \text{ [mm]}. \tag{1}$$

with z in meters. The values of q at the beginning, resp. at the end, of the operational range were $q(0.50 \text{ m}) = 0.65$ mm, resp. $q(15.7 \text{ m}) = 685$ mm.



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

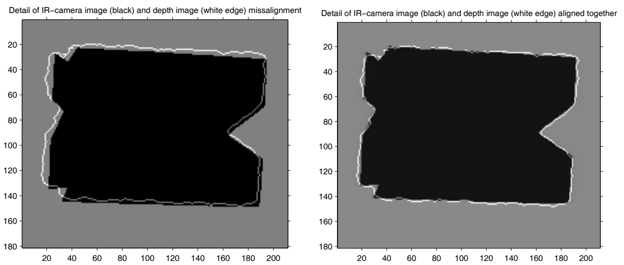
(b) Kinect depth quantization step q (0-15 m).

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

Figure 3. Estimated size of the Kinect quantization step q as a function of target distance for $0 - 5$ m.

2.5. Shift between IR image and Depth image

IR and Depth images were found to be shifted. To determine the shift $[u_0, v_0]^\top$, several different targets were



(a) Before alignment (b) After alignment

Figure 4. Illustration of IR to Depth image shift and its correction. The IR image of a target is shown in black. The Depth image of the target is represented by its white edge.

Image	1	2	3	4	Mean
u_0	2.8	2.9	3.0	3.4	3.0
v_0	3.0	2.7	2.8	3.1	2.9

Table 1. IR to Depth-camera pixel position shift.

captured in the IR and Depth images, Fig. 4(a). The contrast target was segmented out from the background and the shifts was determined by bringing the segmented shapes in the best alignment, Fig. 4(b).

The results of several experiments with targets of different shapes are shown in Tab. 1. The shift was estimated as the mean value over all experiments. Our result suggests using a correlation window of size 7 × 7 pixels in the depth calculation process. This is close to 9 × 9 window size estimated in [11].

3. Kinect geometrical model

We model Kinect as a multi-view system consisting of RGB, IR and Depth cameras. Geometrical model of RGB and IR cameras, which project a 3D point $\mathtt{X}$ into image point $[u, v]^\top$, is given by [2]

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathtt{K} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} s \\ t \\ 1 \end{bmatrix} = \underbrace{(1 + k_1 r^2 + k_2 r^4 + k_5 r^6) \begin{bmatrix} p \\ q \\ 0 \end{bmatrix}}_{\text{radial distortion}} + \underbrace{\begin{bmatrix} 2\,k_3\,p\,q + k_4\,(r^2 + 2\,p^2) \\ 2\,k_4\,p\,q + k_3\,(r^2 + 2\,q^2) \\ 1 \end{bmatrix}}_{\text{tangential distortion}} \tag{3}$$

$$r^2 = p^2 + q^2, \quad \begin{bmatrix} p\,z \\ q\,z \\ z \end{bmatrix} = \mathtt{R}\,(\mathtt{X} - \mathtt{C}) \tag{4}$$

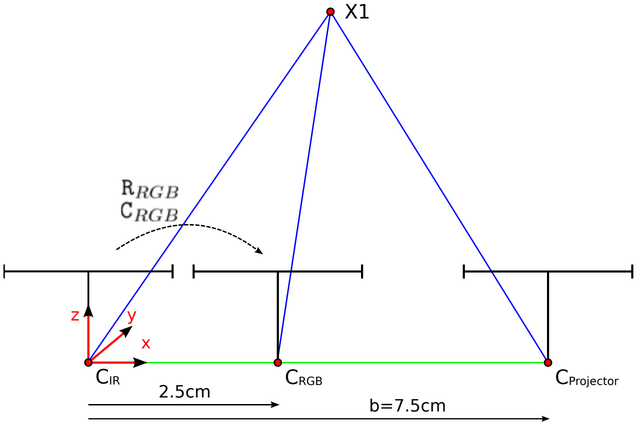


Figure 5. Geometrical model of Kinect.

with distortion parameters $\mathtt{k} = [k_1, k_2, \ldots, k_5]$, camera calibration matrix $\mathtt{K}$, rotation $\mathtt{R}$ and camera center $\mathtt{C}$ [6].

The Depth camera of Kinect is associated to the geometry of the IR camera. It returns the inverse depth d along the z-axis, Fig. 5, for every pixel $[u, v]^\top$ of the IR cameras as

$$\begin{bmatrix} x \\ y \\ d \end{bmatrix} = \begin{bmatrix} u - u_0 \\ v - v_0 \\ \frac{1}{c_1}\,\frac{1}{z} - \frac{c_0}{c_1} \end{bmatrix} \tag{5}$$

where u, v are given by Eq. 3, true depth z by Eq. 4, $[u_0, v_0]^\top$ by Tab. 1, $\mathtt{X}$ 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 $\mathtt{R}_{IR} = \mathtt{I}$ and $\mathtt{C}_{IR} = \mathtt{0}$. A 3D point $\mathtt{X}_{IR}$ is constructed from the measurement $[x, y, d]$ in the depth image by

$$\mathtt{X}_{IR} = \frac{1}{c_1 d + c_0}\, dis^{-1}\left(\mathtt{K}_{IR}^{-1} \begin{bmatrix} x + u_0 \\ y + v_0 \\ 1 \end{bmatrix}, \mathtt{k}_{IR}\right) \tag{6}$$

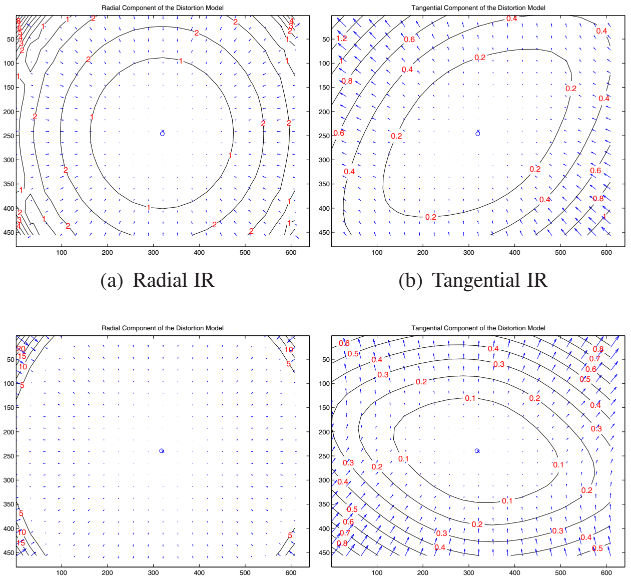
and projected to the RGB images as

$$\mathtt{u}_{RGB} = \mathtt{K}_{RGB}\, dis\,(\mathtt{R}_{RGB}(\mathtt{X}_{IR} - \mathtt{C}_{RGB}), \mathtt{k}_{RGB}) \tag{7}$$

where dis is the distortion function given by Eqn. 3, $\mathtt{k}_{IR}$, $\mathtt{k}_{RGB}$ are respective distortion parameters of the IR and RGB cameras, $\mathtt{K}_{IR}$ is the IR camera calibration matrix and $\mathtt{K}_{RGB}, \mathtt{R}_{RGB}, \mathtt{C}_{RGB}$ are the calibration matrix, the rotation matrix and the center, of the RGB camera, respectively.

4. Kinect calibration

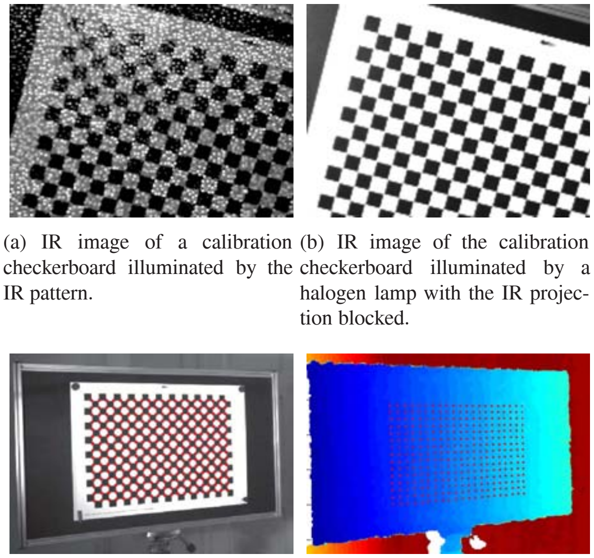
We calibrate [2] Kinect cameras together by showing the same calibration target to the IR and RGB cameras, Fig. 7(c). In this way, both cameras are calibrated w.r.t. the same 3D points and hence the poses of the cameras w.r.t. points can be chained to give their relative pose, Fig. 8. Taking the cartesian coordinate system of the IR camera as the



(a) Radial IR (b) Tangential IR

(c) Radial RGB (d) Tangential RGB

Figure 6. Estimated distortions effects of both Kinect cameras. The red numbers denote the size and arrows the direction of the pixel displacement induced by the lens distortion. The cross indicates the image center the circle the location of the principal point.



(a) IR image of a calibration checkerboard illuminated by the IR pattern.

(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.

Figure 7. The calibration board in the IR, RGB and Depth images.

Focal length		Principal point	
f [px]	f [mm]	x_0 [px]	y_0 [px]
585.6	6.1	316	247.6
Distortion coefficients			
k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}
-0.1296	0.45	-0.0005	-0.002

Table 2. Intrinsic parameters of Kinect IR camera.

Focal length		Principal point	
f [px]	f [mm]	x_0 [px]	y_0 [px]
524	2.9	316.7	238.5.
Distortion coefficients			
k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}
0.2402	-0.6861	-0.0015	0.0003

Table 3. Intrinsic parameters of Kinect RGB camera.

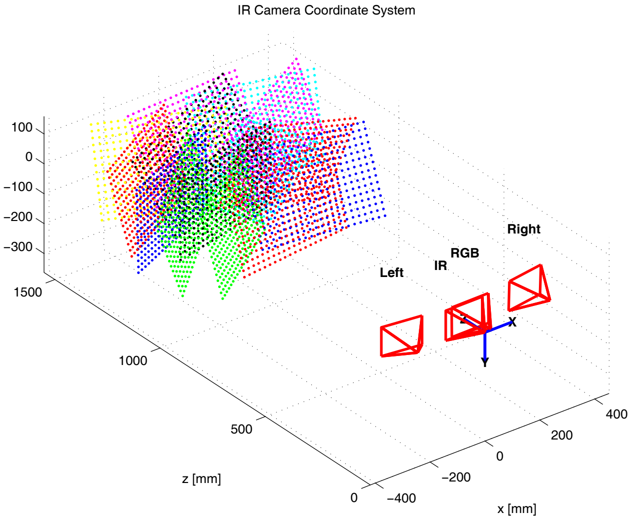
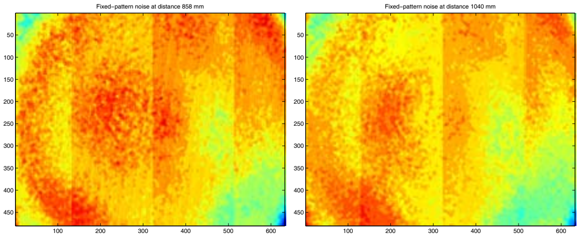


Figure 8. Position and orientation of Kinect IR and RGB cameras and the SLR stereo pair (Left, Right) altogether with 3D calibration points reconstructed on planar calibration targets.

global Kinect coordinate system makes the camera relative pose equal to $\mathtt{R}_{RGB}, \mathtt{C}_{RGB}$.

Tab. 2, 3 show internal parameters and Fig. 6 shows effect of distortions in the cameras. We included the tangential distortion since it non-negligibly increased the overall accuracy of 3D measurement. Fig. 7(a) shows the IR image of the calibration board under the normal Kinect operation when it is illuminated by its IR projector. Much better image is obtained by blocking the IR projector and illuminating the target by a halogen lamp Fig. 7(b).

Parameters c_0, c_1 of the Depth camera are calibrated as follows. We get n measurements $\mathtt{X}_{D_i} = [x_i, y_i, d_i]^\top$, $i = 1, \dots, n$, of all the calibration points from Depth images, Fig. 7(d). The cartesian coordinates $\mathtt{X}_{IR_i}$ 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 $\mathtt{X}_{D_i}$ to $\mathtt{X}_{IR_i}$ using Eqn. 6.



(a) Distance 86 cm (b) Distance 104 cm

Figure 9. Residuals of plane fitting showing the fixed-pattern noise on depth images from different distances.

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

Table 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.

4.1. Learning complex residual errors

It has been observed that Kinect calibrated with the above procedure still exhibited small but relatively complex residual errors for close range measurements. Fig. 9 shows residuals after plane fitting to the calibrated Kinect measurement of a planar target spanning the field of view. The target has been captured from 18 different distances ranging from 0.7 to 1.3 meters and highly correlated residuals were accounted.

Residuals along 250 horizontal Depth image rows are shown in Fig. 10(a). Residuals 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 $\mathtt{X}_{IR}$ computed by Eqn. 6.

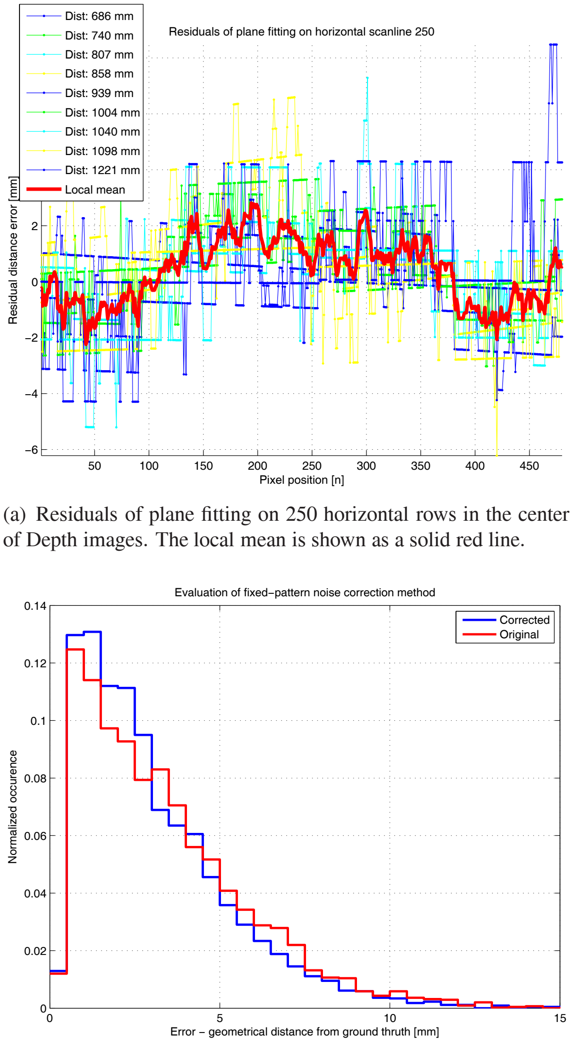
To evaluate this correction method, the z-correction image was constructed from residuals of even images and then applied to odd (the first row of Tab. 4) and to even (the second row of Tab. 4) depth images. The standard deviation of residuals decreased.

After applying the z-correction to Kinect measurements from the experiment described in Section 5.1, the mean of the residual errors decreased by approx. 0.25 mm, Fig. 10(b). The residuals were evaluated on 4410 points spanning the field of view.

5. Validation

5.1. Comparison of Kinect, SLR Stereo and 3D TOF

We compare the accuracy of the measurement of planar targets by Kinect, SLR Stereo and 3D TOF cameras. Kinect and SLR Stereo (image size 2304×1536 pixels) were rigidly mounted (Fig. 2) and calibrated (Fig. 8) together. SLR Stereo was done by reconstructing calibration points extracted by [2] and triangulated by the linear least squares triangulation [6]. 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, Tab. 5, corresponds to the Euclidean distance between the points returned by the sensors and points reconstructed in the process of calibration of cameras of the sensors. SLR Stereo is the most accurate, Kinect follows and SR-4000 is the least accurate.



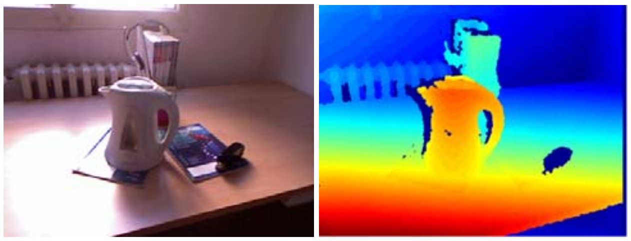
(a) Residuals of plane fitting on 250 horizontal rows in the center of Depth images. The local mean is shown as a solid red line.

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

Figure 10. Correcting complex residual errors.

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

Table 5. Comparison of SLR Stereo triangulation, Kinect and SR-4000 3D TOF depth sensing.



(a) RGB image

(b) Depth image

Figure 11. Example of images from Kinect RGB cameras and the corresponding depth that were used for scene reconstruction.

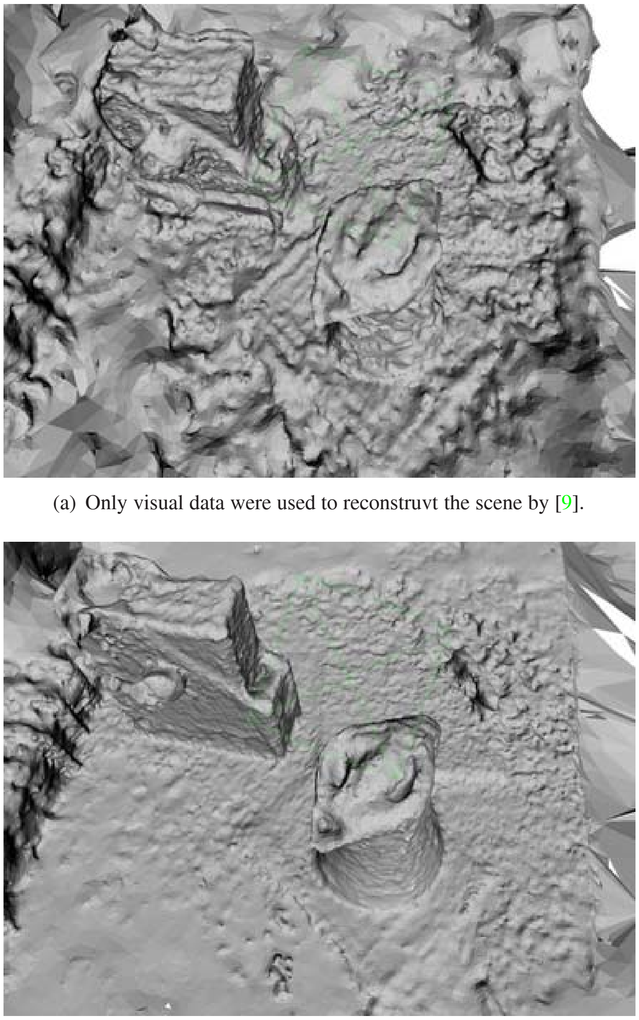
5.2. Combining Kinect and SfM

Fig. 11 shows a pair of $1/2$-resolution (640×480) Kinect RGB image and the depth along the rays through the RGB image pixels computed by Eqn. 7. A sequence of 50 RGB-Depth image pairs has been acquired and the the poses of the RGB cameras have been computed by an SfM pipeline [17, 7]. Using the Kinect calibration, true depth images have been computed and registered together. Fig. 12(a) shows a surface reconstructed from 3D points obtained by mere Multiview stereo [9] on Kinect RGB images. Fig. 12(b) shows the improvement obtained by using the same method after adding registered Kinect 3D data.

Fig. 13 compares a 3D surface reconstructions from point cloud computed by plane sweeping [9] with 70 Kinect 3D data processed by surface reconstruciton of [9] (2304×1536 pixels). Kinect 3D data were registered into a common coordinate system via SfM [17, 7] applied to Kinect image data. We can see that when multiple measurements are used, the result with Kinect is quite comparable to more accurate Multi-view atereo reconstruction.

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 overperformed SwissRanger SR-4000 and was close to 3.5 M pixel SLR Stereo.



(a) Only visual data were used to reconstruvt the scene by [9].

(b) Improved reconstruction using Kinect depth data registered by SFM applied to Kinect colour images.

Figure 12. Scene Reconstruction from Kinect RGB Camera. The figure shows a comparison of reconstruction quality when the scene is reconstructed only using *multi view stereo* and the case when the 3D data from Kinect is also available.

References

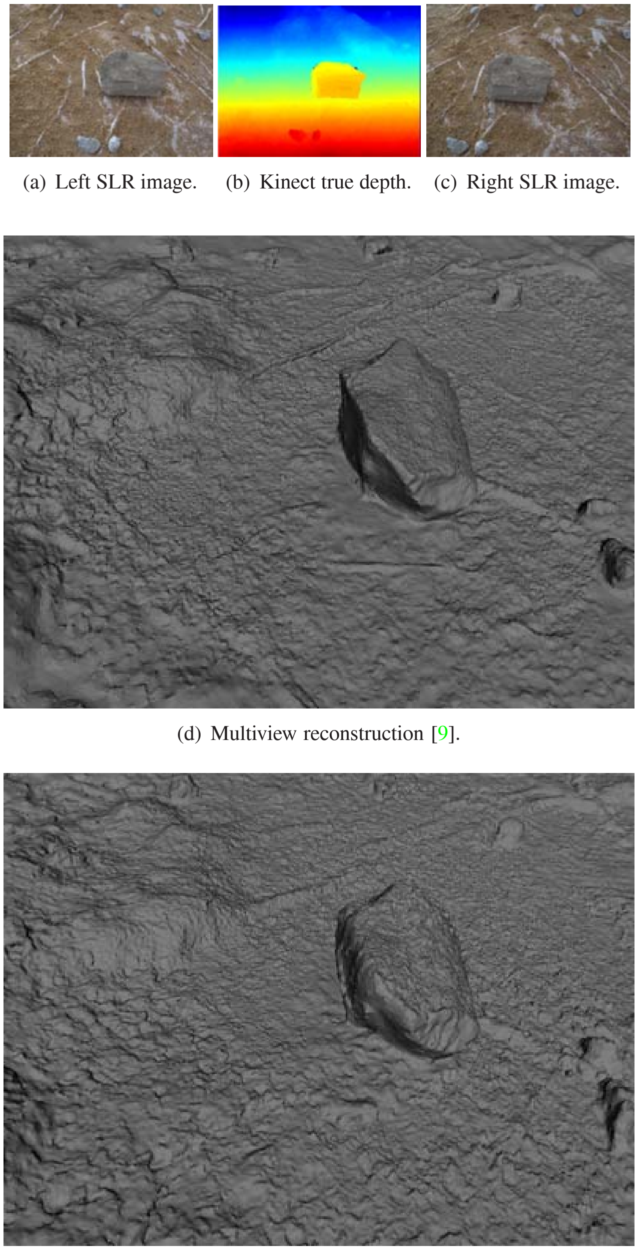
[1] D. Avrahami et al. RGB-D: Techniques and usages for Kinect style depth cameras. http://ils.intel-research.net/projects/rgbd, 2011. 1

[2] J. Y. Bouguet. Camera calibration toolbox. http://www.vision.caltech.edu/bouguetj/calib_doc/, 2010. 1, 2, 3, 5

[3] N. Burrus. Kinect calibration. http://nicolas.burrus.name/index.php/Research/KinectCalibration. 1, 2

[4] I. Dryanovski, W. Morris, and S. Magnenat. kinect_node. http://www.ros.org/wiki/kinect_node, 2010. 1

[5] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli. Depth mapping using projected patterns, 2010. 1



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

(d) Multiview reconstruction [9].

(e) Reconstruction from registered Kinect 3D data.

Figure 13. Comparison of Kinect with Multi-view reconstruction [9].

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2nd edition, 2003. 3, 5

[7] M. Havlena, A. Torii, and T. Pajdla. Efficient structure from motion by graph optimization. In *ECCV*, 2010. 2, 6

[8] D. C. Herrera, J. Kannala, and J. Heikkila. Accurate and practical calibration of a depth and color camera pair. http://www.ee.oulu.fi/~dherrera/kinect/2011-depth_calibration.pdf, 2011. 1

[9] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. 2011. 6, 7

[10] K. Khoshelham. Accuracy analysis of kinect depth data. In *ISPRS Workshop Laser Scanning*, volume XXXVIII, 2011. 1

[11] K. Konolige and P. Mihelich. Technical description of Kinect calibration. http://www.ros.org/wiki/kinect_calibration/technical, 2011. 1, 3

[12] K. Lai, L. Bo, X. Ren, and D. Fox. Sparse distance learning for object recognition combining RGB and depth information. In *IEEE International Conference on Robotics and Automation*, 2011. 1

[13] MESA Imaging. SwissRanger SR-4000. http://www.mesa-imaging.ch/, 2011. 5

[14] Microsoft. Kinect for X-BOX 360. http://www.xbox.com/en-US/kinect, 2010. 1

[15] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from a single depth image. In *CVPR*, 2011. 1

[16] J. Smisek and T. Pajdla. *3D Camera Calibration*. MSc. thesis, Czech Techical University in Prague, 2011. 2

[17] N. Snavely, S. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 2007. 2, 6

[18] Wikipedia. Kinect. http://en.wikipedia.org/wiki/Kinect. 1

[19] Willow Garage. ROS - Kinect calibration: Code complete. http://www.ros.org/news/2010/12/kinect-calibration-code-complete.html, 2010. 2

[20] Willow Garage. Camera calibration and 3D reconstruction. http://opencv.willowgarage.com/documentation/cpp/camera_calibration_and_3d_reconstruction.html, June 2011. 1

[21] Willow Garage. Turtlebot. http://www.willowgarage.com/turtlebot, 2011. 1