Color and Depth Image Correspondence for Kinect v2

Changhee Kim[1], Seokmin Yun[1], Seung-Won Jung[2], and Chee Sun Won[1,*]

[1] Dongguk University-Seoul,
Dept of Electronics and Electrical Engineering,
[2] Dongguk University-Seoul, Dept of Multimedia Engineering,
Dongguk University-Seoul, 30, Pildong-ro 1gil, Jung-gu, Seoul 100-715, Korea
Kimchda210@naver.com, {smyun,swjung83,cswon}@dongguk.edu

Abstract. Kinect v2, a new version of Kinect sensor, provides RGB, IR (Infra-Red) and depth images like its predecessor Kinect v1. However, the depth measurement mechanism and the image resolutions of the Kinect v2 are different from those of Kinect v1, which requires a new transformation matrix for the camera calibration of Kinect v2. In this paper, we correct the radial distortion of the RGB camera and find the transformation matrix for the correspondence between the RGB and depth image of the Kinect v2. Experimental results show that our method yields accurate correspondence between the RGB and depth images.

Keywords: Kinect v2, registration, camera calibration.

1 Introduction

Kinect from Microsoft is a very popular depth sensor. Thanks to its low price, it has been widely adopted for various computer vision applications, including 3D reconstruction, object recognition, and object tracking. To adopt the Kinect for more sophisticated computer vision applications, however, RGB and depth sensors of Kinect need to be precisely calibrated. The calibration for the first version of the Kinect (Kinect v1) has been proposed [1].

Recently, the second version of Kinect (Kinect v2) has been released. The Kinect v2 adopts a different sensing method for the depth measurement and provides higher image resolutions. Specifically, the Kinect v2 uses 'Time of Flight' (TOF) method instead of 'light coding' of the Kinect v1 for the depth measurements. Also, the image resolutions for both RGB and depth of the Kinect 2 are higher than those of Kinect v1. See Table 1 for more comparisons between Kinect v1 and v2 [2]. Because of the differences listed in Table 1, the calibration parameters developed for the Kinect v1 are not applicable for the Kinect v2 sensors. In this paper, we correct the image distortions caused by the lens of the Kinect v2 and provide its calibration matrix for the correspondence between the RGB, depth, and IR (Infra Red) images.

This paper is composed of the following sections. In Section 2, the radial distortions of the Kinect are corrected. The holes in the depth image are filled and the correspondence matrix between the RGB and the depth images is provided. In Section 3 we show the accuracy of our method by comparing with the results by SDK tools.

---

[*] Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Advanced Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 352, DOI: 10.1007/978-3-662-47487-7_17

Table 1. Comparative specifications of Kinect v1 and Kinect v2

	Kinect v1	Kinect v2
Resolution of color image	640x480 (pixel)	1920x1080 (pixel)
Resolution of IR and depth image	320x240 (pixel)	512x424 (pixel)
Field of view of color image	62°x48.6°	84.1°x53.8°
Field of view of IR and depth image	57.5°x43.5°	70.6°x60°
Maximum skeletal tracking	2	6
Method of depth measurement	Light coding	Time of Flight
Working range	0.8m~3.5m	0.5m~8m

2 Kinect v2 Calibration

2.1 Distortion Correction

Camera lens causes image distortions, where Kinect v2 is not an exception. Both IR and color cameras have distortions. Calibration tools such as Camera Calibration Toolbox for Matlab are available to determine the intrinsic parameters of the cameras. A set of checkerboard images from both color and IR cameras are used to identify the corners of the checkerboard pattern in RGB and IR images. Then, by solving the equations from the correspondences of the corner points and by using the non-linear optimization technique to reduce the reprojection errors, the intrinsic camera parameters such as focal length, principal points, and skew can be determined [3]. Then, the radial distortions of the color camera can be corrected. The results of our calibration parameters for the Kinect v2 are listed in Table 2.

Table 2. Camera parameters for Kinect v2

Camera / Parameter	RGB	IR
Focal Length (f_c)	[1053.622 1047.508] ± [4.6884 4.5323]	[376.6518 371.4936] ± [1.8265 1.8015]
Principal points (c_c)	[950.3941 527.3442]	[265.5583 206.6131]
Skew (α_c)	[0.000] ± [0.000] →angle of pixel=90.00 degrees	[0.000 ± [0.000] →angle of pixel=90.00 degrees
Distortion (k_c)	[0.0042 -0.0019 -0.0038 -0.0026 0] ± [0.0038 0.0033 0.0007 0.0008 0]	[-0.0094 -0.0431 0.0004 -0.0003 0] ± [0.0094 0.0144 0.0015 0.0017 0]

Zoomed images before and after the correction of the Kinect v2 distortions are shown in Fig. 1. As one can see in Fig. 1 (a) and (c), the images captured by the Kinect v2 sensor suffer from the radial distortions. These distortions are corrected by the calibration parameters in Table 2 (see Fig. 1 (b) and (d)).

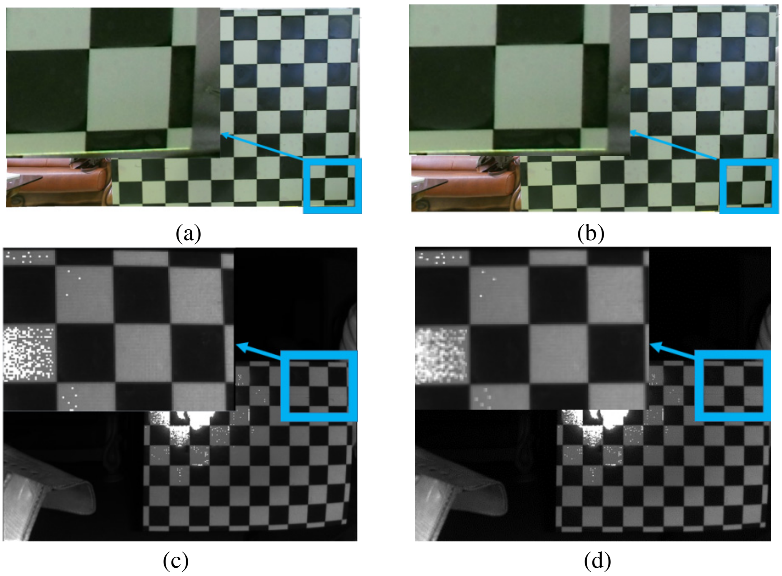


Fig. 1. Images with distortions (a) and (c) and after distortion correction (b) and (d). Above images are taken by color camera and below ones are IR images.

2.2 Correspondence of Color and IR (Depth) Images

The correspondence between the RGB sensor and the IR (depth) sensor can be done by a transformation matrix between them. Since the depth image is generated by the IR sensor, we can use either the depth image or the IR image for the registration. Here, since the depth image cannot show the pattern on the planer checkerboard, our registration is based on the checkerboard images of the RGB and the IR images.

The projective matrix which converts the RGB image coordinate (x,y) into the IR image coordinate(X,Y) is given as follows [4].

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

Matrix equation (1) with eight unknown parameter values can be rewritten as follows

$$X = a_1 x + a_2 y + a_3 - a_7 xX - a_8 yX \tag{2}$$

$$Y = a_4 x + a_5 y + a_6 - a_7 xY - a_8 yY \tag{3}$$

Four corresponding pairs of (x,y) and (X,Y) must be known to solve equations (2) and (3). If we use more than four pairs, the parameters can be found using a least square method like the direct linear transform. We use the checkerboard to find the pairs of corresponding points between the RGB and IR images as shown in Fig. 2. The corresponding points are selected manually. Using these points, we calculate the eight parameters for the transformation matrix. Specifically, we use 515 pairs of points extracted from indoor images of limited distances to calculate equations (2) and (3), and we get the transformation matrix as shown in (4). So, the matrix in (4) is good for the calibration of Kinect v2 images captured in the near distances, say up to 3m.

$$T = \begin{bmatrix} 0.3584 & 0.0089 & 0.0000 \\ 0.0031 & 0.3531 & 0.0001 \\ -101.5934 & 13.6311 & 0.9914 \end{bmatrix} \quad (4)$$

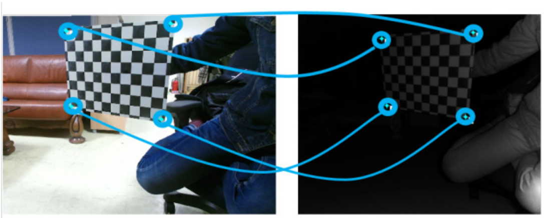


Fig. 2. Corresponding points in RGB and IR images

Fig. 3. Above: original color and IR images. Below: registered color and IR images obtained by the transformation and cropping.

Note that the size of RGB image is bigger than the IR image in Kinect v2. Also, they have different field of views (FOV). Therefore, color images need to be cropped after registration. Fig. 3 shows an example of registration results. Above images are raw images and below ones are the results after the registration and cropping. Color image is cropped at both sides to fit the FOV of the IR image. The IR image is also cropped at top-bottom parts because the top-bottom FOV of color image is bigger than IR image. After the calibration and cropping the size of IR image and color image is changed to 512x360. Original size of color is 1920x1080 and that of IR is 512x424.

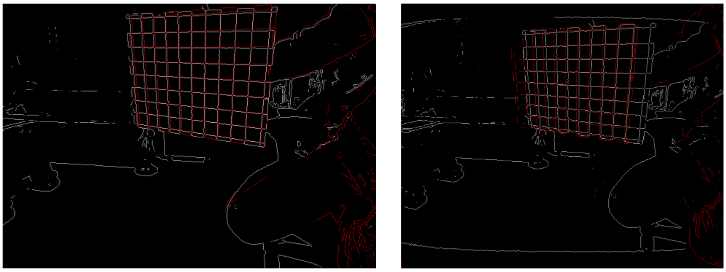
2.3 Hole Filling

Since the depth image is captured by the IR sensor in Kinect v2, the registration between the RGB and the IR image automatically yields the registration among the RGB, IR, and depth images. So, after the registration and cropping, we can generate the pixel-by-pixel correspondence images of RGB, IR, and depth from the Kinect v2 sensor. Here, to make a perfect correspondence, the holes in the depth image are to be filled.

As in Kinect v1, Kinect v2 has depth holes with missing depth measurements. Although the holes of the Kinect v2 along the object boundary are usually thinner than those of the Kinect v1, the holes near the object boundary in the Kinect v2 depth image can be still filled by the method of Kinect v1. In particular, since the RGB image is already aligned with the depth image, we can exploit the edge information in the RGB image to determine the direction of the hole filling in the depth image [5].

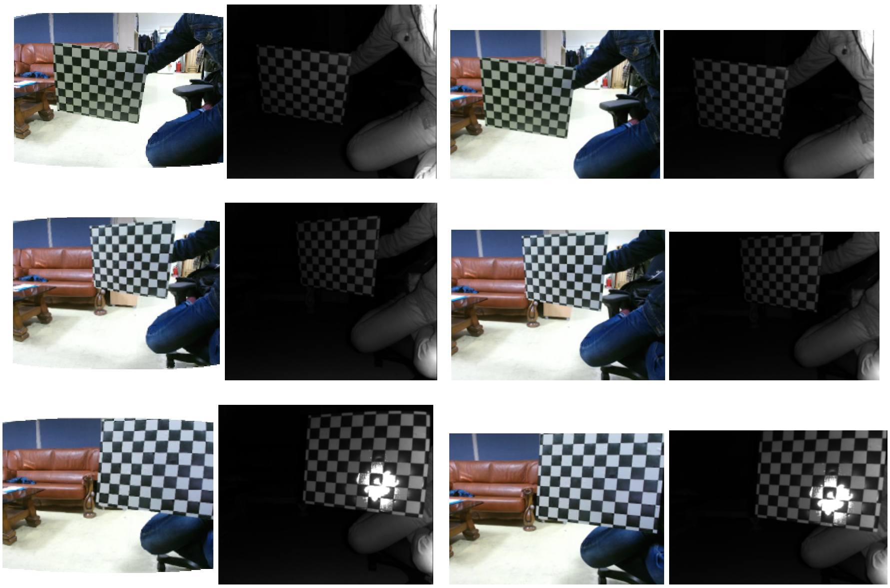
Table 3. Caparison of pixel distance

Image set \ Method	SDK	Ours
A	32.5038	5.7118
B	21.3204	3.0104
C	24.8821	7.8142
Average	26.3254	5.5121



(a) (b)

Fig. 4. Superimposed images (red lines are edges of IR image and white ones are edges of color image): (a) our method, (b) the SDK function



(a) (b)

Fig. 5. Three sets of color and IR images that are used in pixel position error calculation: (a) Result images from the SDK and (b) Result images from our method

3 Results

Fig. 4-(a) shows the superimposed images after the correction of the radial distortion and the registration. Compared to the result of Fig. 4-(b) obtained by the software development kit (SDK) [2] function of "*MapColorFrameToDepthSpace*", the mismatch errors of our method are much smaller than those of the SDK method. Specifically, the checkerboard lines in the left image (Fig. 4-(a)) are well aligned compared to the checkerboard lines in right image (Fig. 4-(b)).

We also calculate the pixel distances of correspondence points of checkerboard in the registered color and IR images to compare the pixel position errors. The results using three sets of color and IR images are shown in Table 3. Also there are images used for calculating pixel position errors in Fig. 5. The average distance of the SDK method is 26.2354 and that of our result is 5.5121. We can notice that our method is about five times more accurate than the SDK method in terms of the pixel mismatches.

4 Conclusion

In this paper, we performed a case study of aligning color, IR, and depth images using Kinect v2 sensor. We obtained the intrinsic parameters for Kinect v2 and calibrated the RGB and IR sensors via the transformation matrix between the two sensors. High accuracy of the proposed registration was confirmed by comparing to the results of the SDK function visually and numerically. As a future work we need a universal calibration matrix for all near and far distances covered by the Kinect v2.

Acknowledgments. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2005024) and by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2015-H0301-14-4007) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Smisek, J., Jancosek, M., Pajdla, T.: 3D with Kinect. Consumer Depth Cameras for Computer Vision, pp. 3–25. Springer, London (2013)
2. Microsoft, Kinect for Windows, `http://www.microsoft.com/en-us/kinectforwindows`
3. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 1330–1334 (2000)
4. Rothwell, C., Forsyth, D.A., Zisserman, A., Mundy, J.L.: Extracting Projective Structure from Single Perspective Views of 3D Point Sets. In: Fourth International Conference on Computer Vision, pp. 573–582. IEEE Press (1993)
5. Le, A.V., Jung, S.W., Won, C.S.: Directional Joint Bilateral Filter for Depth Images. Sensors 14, 11362–11378 (2014)