Ho Chuen Kam, Kin Hong Wong and Baiwu Zhang, "Dual Back-to-Back Kinects for 3-D Reconstruction", ISVC'16 12th International Symposium on Visual Computing December 12-14, 2016, Las Vegas, Nevada, USA .

Dual Back-to-Back Kinects for 3-D Reconstruction

Ho Chuen Kam[1(✉)], Kin Hong Wong[1], and Baiwu Zhang[2]

[1] Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong
hckam@cse.cuhk.edu.hk
[2] University of Toronto, Toronto, ON M5S, Canada

Abstract. In this paper, we investigated the use of two Kinects for capturing the 3-D model of a large scene. Traditionally the method of utilising one Kinect is used to slide across the area, and a full 3-D model is obtained. However, this approach requires the scene with a significant number of prominent features and careful handling of the device. To tackle the problem we mounted two back-to-back Kinects on top of a robot for scanning the environment. This setup requires the knowledge of the relative pose between the two Kinects. As they do not have a shared view, calibration using the traditional method is not possible. To solve this problem, we place a dual-face checkerboard (the front and back patterns are the same) on top of the back-to-back Kinects, and a planar mirror is employed to enable either Kinect to view the same checkerboard. Such an arrangement will create a shared calibration object between the two sensors. In such an approach, a mirror-based pose estimation algorithm is applied to solve the problem of Kinect camera calibration. Finally, we can merge all local object models captured by the Kinects together to form a combined model with a larger viewing area. Experiments using real measurements of capturing an indoor scene were conducted to show the feasibility of our work.

1 Introduction

In recent years visual reality is becoming popular, and many applications are developed for industrial and domestic use. Virtual tour in museums and tourist attractions is one of the potential applications. This requires the capturing of the environments and turning them into various 3-D models. With the range cameras, images and depth information are easily aggregated to construct virtual scenes.

A variety of 3-D range cameras are already available in the market, such as Microsoft Kinect, etc. It is economical so that it is extensively used in 3-D vision research. In 3-D reconstruction, normally one Kinect is employed to scan the entire environment. KinectFusion [1,2] are examples of the renowned algorithms for capturing the virtual scene. However, this kind of one-Kinect method suffers from some undesirable effects. As the algorithm is largely based on feature matching among frames, it will not work on the following cases:

1. The Kinect twitches, i.e., translates and rotates too fast to a great extent.
2. The object surface is too plain and lack of features (e.g. plain wall).

Under these circumstances the one-Kinect algorithm fails to converge, yielding undesirable results.

In this paper, a simple and efficient way is proposed to tackle the 3-D reconstruction problem. First, two back-to-back Kinects are mounted on top of a robot, and each of them captures a point cloud. Finally, they are merged to form the whole scene. To accomplish the task, we have to find the relative locations of the cameras. However, as they share no common views, traditional algorithms for camera calibration [3] do not work. We propose putting a dual-face checkerboard (the front and back patterns are the same) on top of the two Kinects, and a planar mirror is employed to recover the images of checkerboard for the cameras. By doing so, we created a calibration object for the RGB cameras of the two Kinects without shared view.

This paper is organised as follows. The related work will be discussed at Sect. 2 and theories used are explained in Sect. 3. Synthetic and real experiments are conducted, and the results are shown in Sect. 4. Section 5 concludes our work.

2 Related Work

2.1 KinectFusion

KinectFusion [1] by Newcombe et al. is one of the notable and first of the 3-D volumetric reconstruction techniques. It totally relies on the object features for registration and calculates the correspondences by the estimation algorithm Iterative Closet Point (ICP) [4]. By using one Kinect and sliding it across a scene, a unique 3-D model is generated by the following steps:

1. Surface Extraction of 3-D object.
2. Alignment of sensor.
3. Volumetric integration to fill 3-D model.
4. Raycasting.

Although the algorithm can achieve an extraordinary accuracy, it suffers from various limitations which are unstable in some scenarios. The major working principle relies on feature matching step using ICP. Therefore it fails when the cameras move too fast, or the object it captures contains few features.

2.2 Pose Estimation Without a Direct View

First, the problem is coined by Sturm and Bonfort [5] in 2006. They are the first to suggest the use of a planar mirror for pose estimation. However using a mirror, the calibration object can become a virtual image of a camera viewing through the mirror. Besides, they pointed out that the motion between two

consecutive virtual views is on the intersection line of two planes. This motion can be described as *fixed-axis rotation*.

Later, Kumar et al. [6] formulated a linear method to calibrate cameras using a planar mirror. This method requires five virtual views to be incorporated in the linear system, but this does not require the constraints of fixed-axis rotation.

Hesch et al. [7] put forward the use of the maximum-likelihood estimator to compute the relative poses of a camera using a planar mirror. They tried to minimise the solving system from five virtual views to three points viewed in three planes when compared to the method of Kumar et al. [6].

Later in 2010 Rodrigues et al. [8] further extended the linear method to a better closed-form solution. The mirror planes positions can be simply and unambiguously solved by a system of linear equations. The method enabled a minimum of three virtual views to converge.

In 2012 Takahashi et al. [9] introduced a new algorithm of using Perspective-3-Point (P3P) to return solutions from three virtual images. The solutions can then be computed by an orthogonality constraint, which was proven to be a significant improvement on accuracy and robustness.

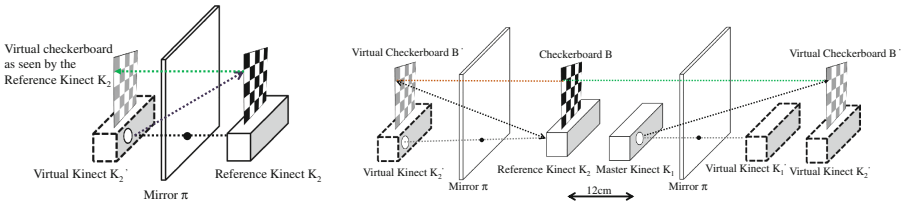
In our paper, the algorithm originated from Rodrigues et al. [8] is used in the localisation of two non-overlapping Kinects.

3 Theory

3.1 Overview of Our Proposed System

To demonstrate our idea, we have built the system for capturing the scene and reconstruction. The subsequent subsections cover the details.

Two Kinects, "Master Kinect K_1" and "Reference Kinect K_2", are placed in the scene as shown in Fig. 1. They are positioned in a back-to-back manner such that their views are non-overlapping. A checkerboard is placed on the top of Kinect K_2. Each Kinect can capture the point clouds (P_1 and P_2) of its field of view respectively. To merge the point clouds together to form the complete scene, point cloud P_2 should be translated to the coordinate system of K_1. The pose of K_2 must be known to K_1 in order to perform this task. In addition, the



(a) Reference Kinect K_2 and the virtual checkerboard B'

(b) Master Kinect K_1 and the virtual checkerboard B'

Fig. 1. Overview of our proposed system

position of checkerboard is not as same as that of Kinect K_2 cameras. Therefore, the pose of the checkerboard with respect to the K_2 camera must also be known. In summary, the proposed algorithm finds out the following to recover the whole 3-D environment:

– The relative pose between K_1 and K_2.
– The relative pose between the checkerboard B and the camera of K_2.

To achieve the tasks, we used a planar mirror to recover the calibration pattern images and the estimation is conducted by the linear methods. In the following sections, we will describe (1) the geometry of the symmetric reflection and (2) the linear method for recovering poses of the back-to-back Kinects.

3.2 Geometry of Mirror Reflection

The formation of the mirror image and the camera geometry are shown in this section. Theories from Rodrigues et al. [8] are summarized and presented here.

First, we define the 3-D point projection. From bringing back the points from the world coordinate to a camera coordinate, the transformation matrix T is applied to the points.

$$T = \begin{pmatrix} R & t \\ \mathbf{0} & 1 \end{pmatrix} \tag{1}$$

T is a 4×4 transformation matrix containing a 3×3 rotation matrix R and a 3×1 translation matrix t.

Moreover, two parameters are needed to define the mirror π. They are the normals in unit vector $\boldsymbol{n}$ and the orthogonal distance d from the mirror to the origin. An arbitrary point x is on the plane π if and only if:

$$\boldsymbol{n}^T x = d \tag{2}$$

Now we define the point projection properties. Assume P is the point in the world that cannot be seen by the camera directly, and $\widehat{P}$ is the reflected point of P. The projection on the plane can be represented by:

$$p \sim K \begin{pmatrix} I & 0 \end{pmatrix} T \begin{pmatrix} \widehat{P} \\ 1 \end{pmatrix} \tag{3}$$

From the Fig. 2, we can establish the relationship between the 3-D point P and its reflection $\widehat{P}$.

$$\widehat{P} = P + 2(d - \boldsymbol{n}^T P)\boldsymbol{n} \tag{4}$$

By simplifying it to matrix form:

$$\begin{pmatrix} \widehat{P} \\ 1 \end{pmatrix} = S \begin{pmatrix} P \\ 1 \end{pmatrix} \tag{5}$$

where S is an symmetry transformation caused by mirror π.

$$S = \begin{pmatrix} 1 - 2\boldsymbol{n}\boldsymbol{n}^T & 2d\boldsymbol{n} \\ 0 & 1 \end{pmatrix} \tag{6}$$

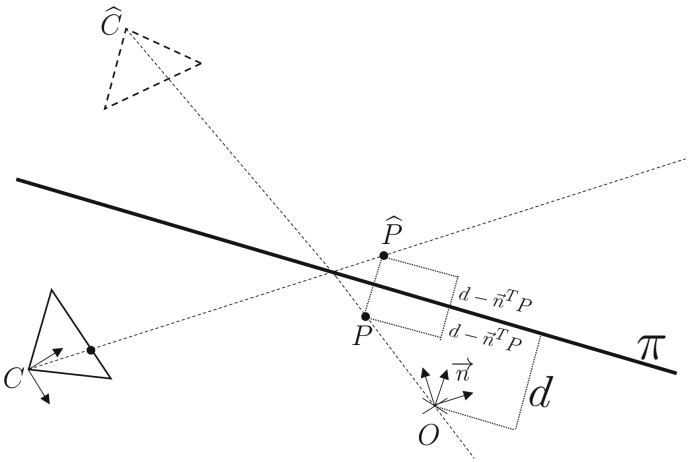


Fig. 2. Overview of mirror geometry. P is a 3-D point and $\widehat{P}$ is its reflection.

Now we define the geometry of the reflected camera model - virtual camera. By combining Eqs. 3 and 5,

$$p \sim K \begin{pmatrix} I & 0 \end{pmatrix} TS \begin{pmatrix} P \\ 1 \end{pmatrix} \tag{7}$$

From Eq. 7, TS transforms the points in world coordinates to the respective mirrored camera frame $\widehat{C}$. There is a remark from Kumar et al. [6] that the handiness changes caused by any symmetry transformation. More Importantly, the transformation from the real space camera C to the virtual space camera $\widehat{C}$ is defined by S' symmetry matrix,

$$S' = TST^{-1} \tag{8}$$

The transformation S' is involutionary, i.e. S' can also be applied to transformation from the virtual space to the real space.

3.3 Problem Formulation

After the reflection geometry is defined, the calibration problem can then be formulated.

Figure 3 shows the geometry of virtual cameras and mirrors. In the followings we take the virtual camera $\widehat{C}_1$ as the reference frame.

$T_{i=1..N}$ are the rigid transformations among virtual cameras. Note that Gluckman and Nayar [10] revealed that those transformation are always planar.

Let $\widehat{P}_i$ and P_r be the same 3-D point expressed with respect to $\widehat{C}_i$ and C_r respectively. From Eqs. 5 and 8,

$$\begin{pmatrix} P_r \\ 1 \end{pmatrix} = T_i S_i T_i^{-1} \begin{pmatrix} \widehat{P}_i \\ 1 \end{pmatrix} \tag{9}$$

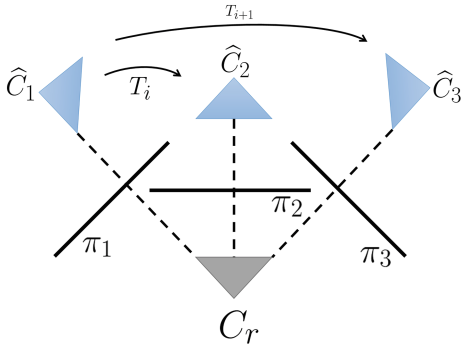


Fig. 3. Mirrors π_i and virtual cameras $\widehat{C}_i$

After all, the following sets of linear constraints can then be established:

$$t_i = 2(d_1 - 2d_i cos(\frac{\theta_i}{2}))\boldsymbol{n}_1 + 2d_i\boldsymbol{n}_i \tag{10}$$

$$t_i^T\boldsymbol{n}_1 - 2d_1 + 2cos(\frac{\theta_i}{2})d_i = 0 \tag{11}$$

$$[t_i]_x\boldsymbol{n}_1 - 2sin(\frac{\theta_i}{2})\boldsymbol{w}_i d_i = 0 \tag{12}$$

With more than 3 virtual views, we can form the following matrix:

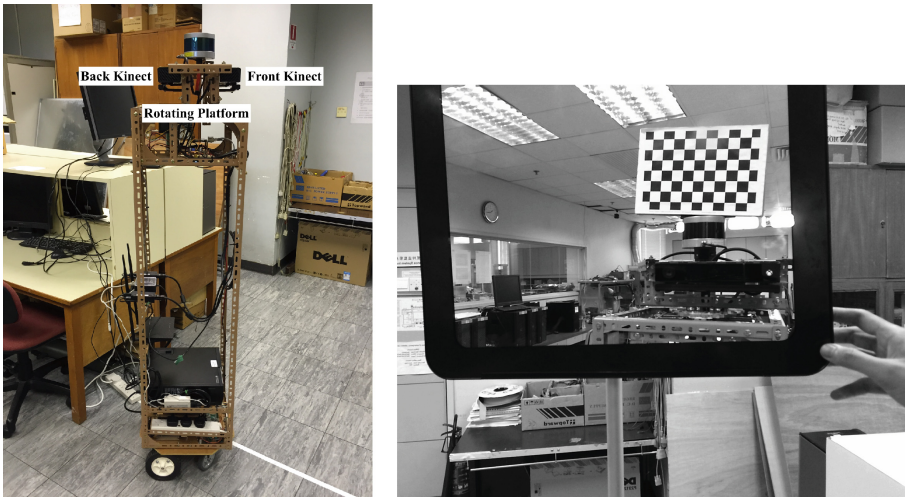
$$\begin{pmatrix} B_1 & b_1 & 0 & \cdots & 0 \\ B_2 & 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{N-1} & 0 & 0 & \cdots & b_{N-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{n}_1 \\ d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_N \end{pmatrix} = 0, \tag{13}$$

$$\text{where } B_i = \begin{pmatrix} t_i^T & -2 \\ [t_i]_x & 0 \end{pmatrix}, b_i = \begin{pmatrix} 2cos(\frac{\theta_i}{2}) \\ -2sin(\frac{\theta_i}{2})\boldsymbol{w}_i \end{pmatrix}$$

By applying SVD to the system, the least square solution can be obtained and hence the positions of mirror planes can then be calculated. With this information, we can further determine the symmetry matrix S and locates the real camera C_r.

4 Experiments

To illustrate our idea our group had built a 160-cm-tall robot with two Kinects on the rotation platform. Figure 4a shows the robot. They are placed in a back-to-back manner, are separated by a distance of 25 cm. The back-to-back Kinect



(a) Our robot (b) Calibrating the Kinect by a mirror

Fig. 4. Using our scene capturing robot

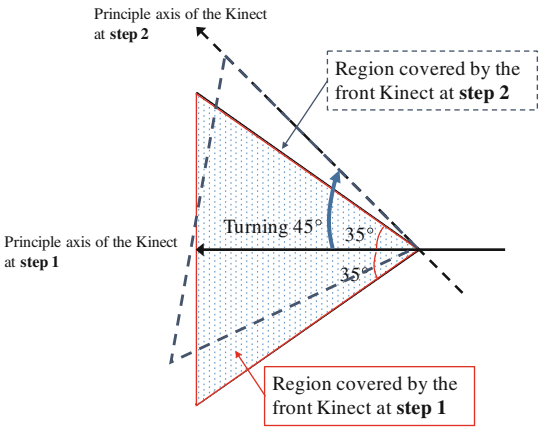
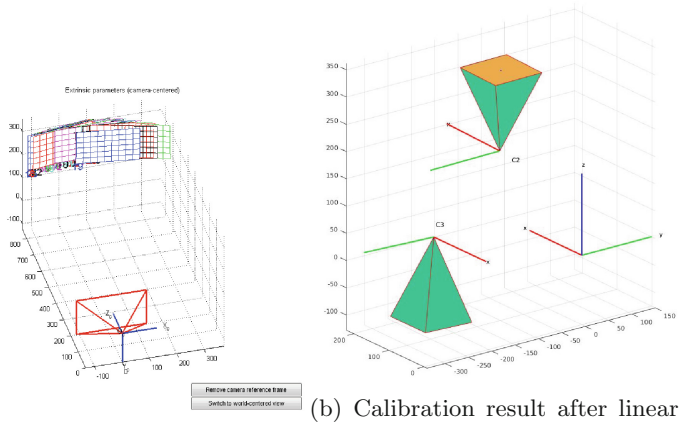


Fig. 5. The region captured by the front Kinect at **step 1** and **step 2**

pair is placed on a rotating platform controlled by a computer. At step 1, we first capture the front and back region by the Kinect pair. Since the Kinect can only cover a region of 70°, so we need to rotate the Kinect pair to cover a larger region. Thus we take another 3-D view at step 2 by turning the Kinect pair 45° horizontally. We repeat till step 4. So the front view ($4\times45°+35°\times2=250°$) will be covered by the front Kinect. Since the back Kinect also capture the back scene with the same scope, the full 360° scene can be covered. The regions covered by the front Kinect at step 1 and step 2 are shown in Fig. 5. The final process is



(a) Direct calibration (b) Calibration result after linear method[8]

Fig. 6. Calibration results

to merge the point clouds captured at all the steps according to their relative positions (shown in Fig. 6b).

4.1 Calibration Using Mirror

In order to verify the algorithm, we performed experiments by manoeuvring mirrors in different key positions and capturing the images. After we had aggregated all the results, we tested and analysed them by Bouguet's Matlab camera calibration toolbox [11].

Figure 6a shows the result of direct calibration. As the calibration algorithm was not aware of the mirrors, the checkerboard patterns appeared to be far behind of them. The process was repeated with the second Kinect at the back. After we had collected all the samples, the linear method by Rodriduges et al. [8] was used to estimate the camera positions.

After the Kinects are calibrated, we applied it to our scanning robot. The rotation platform will turn so that the two Kinects can capture the whole 360-degree scene. Here shows the result of aggregation and merging of point cloud in Fig. 7. Our proposed solution successfully reconstruct an indoor environment, without the need of displacing the Kinect around the scene such as required by KinectFusion [1].

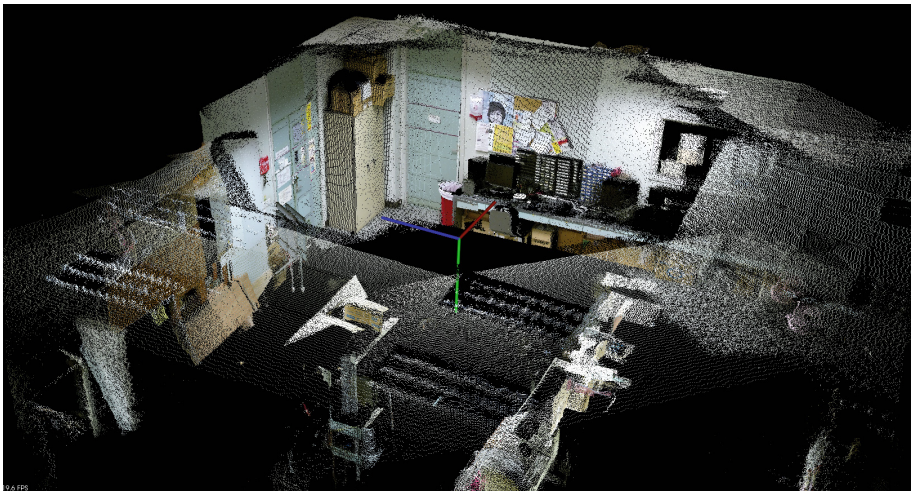


Fig. 7. Point cloud merging results

5 Conclusion

In this paper, we proposed a multiple-Kinect approach to solving the problem of 3-D scene reconstruction. Two back-to-back Kinects were mounted on top of a robot and performed scanning. The aggregated result is stable and accurate as compared to one-Kinect methods such as KinectFusion [1]. In addition, the system is easy to deploy and requires low-cost hardware only. In the future, we are confident that the complete system can be built for various virtual reality applications such as building virtual models for virtual tourism or virtual museums.

Acknowledgement. This work is supported by a direct grant (Project Code: 4055045) from the Faculty of Engineering of the Chinese University of Hong Kong.

References

1. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion: real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 127–136. IEEE (2011)
2. Kim, S., Kim, J.: Occupancy mapping and surface reconstruction using local Gaussian processes with kinect sensors. IEEE Trans. Cybern. **43**, 1335–1346 (2013)
3. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000)
4. Whelan, T., Johannsson, H., Kaess, M., Leonard, J.J., McDonald, J.: Robust real-time visual odometry for dense RGB-D mapping. In: 2013 IEEE International Conference on Robotics and Automation (ICRA), pp. 5724–5731. IEEE (2013)

5. Sturm, P., Bonfort, T.: How to compute the pose of an object without a direct view? In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3852, pp. 21–31. Springer, Heidelberg (2006). doi:10.1007/11612704_3
6. Kumar, R.K., Ilie, A., Frahm, J.M., Pollefeys, M.: Simple calibration of non-overlapping cameras with a mirror. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–7. IEEE (2008)
7. Hesch, J.A., Mourikis, A.I., Roumeliotis, S.I.: Mirror-based extrinsic camera calibration. In: Chirikjian, G.S., Choset, H., Morales, M., Murphey, T. (eds.) Algorithmic Foundation of Robotics VIII, pp. 285–299. Springer, Heidelberg (2009)
8. Rodrigues, R., Barreto, J.P., Nunes, U.: Camera pose estimation using images of planar mirror reflections. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 382–395. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_28
9. Takahashi, K., Nobuhara, S., Matsuyama, T.: A new mirror-based extrinsic camera calibration using an orthogonality constraint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1051–1058. IEEE (2012)
10. Gluckman, J., Nayar, S.K.: Catadioptric stereo using planar mirrors. Int. J. Comput. Vis. **44**, 65–79 (2001)
11. Bouguet, J.Y.: Camera calibration toolbox for Matlab (2004)