

Fusion of information from multiple Kinect sensors for 3D object reconstruction

A.N. Ruchay^{1,2}, K.A. Dorofeev², V.I. Kolpakov¹

¹All-Russian Research Institute of Beef Cattle Breeding, 9 Yanvarya street 29, Orenburg, Russia, 460000

²Department of Mathematics, Chelyabinsk State University, Bratiev Kashirinykh street 129, Chelyabinsk, Russia, 454001

Abstract. In this paper, we estimate the accuracy of 3D object reconstruction using multiple Kinect sensors. First, we discuss the calibration of multiple Kinect sensors, and provide an analysis of the accuracy and resolution of its the depth data. Next, the precision of coordinate mapping between sensors data for registration of depth and color images is evaluated. We test a system with four Kinect V2 sensors and present reconstruction accuracy results. Experiments and computer simulation are carried out using Matlab and Kinect V2.

Keywords: multiple sensors, Kinect, 3D object reconstruction, fusion.

1. Introduction

Accurate 3D reconstruction of objects is an important ingredient for many robotic, assisted living, surveillance, and industrial applications [1]. Consumer depth cameras, such as Microsofts Kinect, provide depth information based on an active structured light sensor combined with color images. The additional depth information helps to estimate the position of 3D object in 3D environment; however, the use of a single (color and depth) camera is still limited especially in highly cluttered scenes, with object occlusion and objects which are difficult to distinguish from a single view. Therefore it is necessary to capture an object of interest from several viewpoints, then the data from all viewpoints are subsequently combined to obtain the surface of the entire object. The problem of 3D reconstruction can be solved in several ways. One of them to capture the object at several instants using a single sensor such as in [2, 3] where the object rotates and the camera is still or as in [4, 5] where the sensor moves around a still object. Another approach uses multiple sensors that capture the object simultaneously as in [6, 7, 8]. However, fusing information from multiple depth cameras for object detection has not been investigated in detail. In this paper we investigate the architecture of this system, the calibration procedure, the problem of interference and the accurate 3D reconstruction of objects.

The paper is organized as follows. Section 2 discusses related approaches, scenarios, and datasets. Section 3 describes the proposed architecture of system with fusion of information from multiple Kinect sensors for object 3D reconstruction. In Section 4 experimental results are discussed. Finally, Section 5 presents our conclusions.

2. Related work

This section contains information about feature representations for depth, multi-camera approaches, and available datasets obtained with (consumer) depth cameras.

In [6] authors generate point clouds from the depth information of multiple registered cameras and use the VFH descriptor to describe them. For color images they employ the DPM and combine both approaches with a simple voting approach across multiple cameras.

A new method for the occluded object visualization using two Kinect sensors at different locations was proposed in [9].

The interference problem of multiple Kinect cameras dramatically degrades the depth quality. In paper [10] an algorithm for interference cancelation in systems with multiple Kinect camera was proposed. This algorithm takes advantage of statistic property of depth map, propagates reliable gradient from interference-free region to interfered region, and derives depth values with complete gradient map under the least errors criterion.

A novel approach to combine data from multiple low-cost sensors to detect people in a mobile robot was proposed in [11]. This work is based on the fusion of information from Kinect and a thermal sensor (thermopile) mounted on top of a mobile platform.

In [12] authors proposed a human action recognition system using multiple Kinect sensors based on multi-view skeleton integration. In [4] the Kinect Fusion was designed for the 3D reconstruction of a scene in real-time using the Kinect sensor. The system was applied for Visual Navigation of a Robotic Vehicle when no external reference like GPS is available.

A mirror movements rehabilitation therapy system for hemiplegic patients was proposed in [13]. This system is based on two Kinects to eliminate problems like limb blocking, data loss in one Kinect by Brusa seven-parameter model and RLS method for coordinate transformation. Two networked Kinect sensors are used for real-time rigid body head motion tracking for brain PET/CT in [14]. Multiple Kinect Fusion allows head motion tracking in the case when partial and complete occlusions of the face occurs. To increase the accuracy of human joint position estimation for rehabilitation and physiotherapy information-fusion from multiple Kinects was used [15]. It was shown that the most significant improvement is achieved with two Kinects and the subsequent increase of the number of receivers is not significant.

A system for live 3D reconstruction using multiple Kinect sensors is presented in [7]. This paper describes and discusses a general design of the system architecture, the method of estimating the camera poses in 3D space, the problem of interference between multiple Kinect V2 devices. The following issuers of 3D reconstruction using multiple Kinect are improved: automated markerless calibration, improved noise removal, tessellation of the output point cloud and texturing.

One of the most important problems is a calibration method for multiple Kinects. In paper [16], an accurate and efficient calibration method for multiple Kinects was proposed by overlapping joint regions among Kinects and extra RGB cameras, containing a sufficient number of corresponding points between color images to estimate the camera parameters. The parameters are obtained by minimizing both the errors of corresponding points between color images and the ones of range data of planer regions from the environment.

A method to calibrate multiple Kinect sensors was developed in [17]. The method requires at least three acquisitions of a 3D object from each camera or a single acquisition of a 2D object, and a point cloud from each Kinect obtained with the built-in coordinate mapping capabilities of Kinect. The proposed method consists of the following steps: image acquisition, pre-calibration, point cloud matching, intrinsic parameters initialization, and final calibration.

Different methods to calibrate a 3-D scanning system consisting of multiple Kinect sensors was investigated in [18]. A sphere, a checker board and a cube as the calibration object were proposed. A cubic object for the calibration task is the most suitable for this application.

A novel method for simultaneously calibration of relative poses of a Kinect and three external cameras by optimizing a cost function and adding corresponding weights to the external cameras in different locations is proposed in [19]. A joint calibration of multiple devices is efficiently constructed.

A real-time 3D reconstruction method to extend the limited field of view of the depth map of Kinect sensors by using registration data of color images to construct depth and color panorama is proposed in [20]. An efficient anisotropic diffusion method is proposed to recover invalid depth data in the depth map from the Kinect sensors.

A calibration procedure with Kinects coordinate mapping to extract registered color, depth, and camera space data during the acquisition step was proposed in [8]. Using three acquisitions from each camera, the calibration procedure is capable to obtain the intrinsic and extrinsic parameters for each camera. A method for point cloud fusion after calibrating the cameras is suggested in [8].

RGB-D datasets for different applications including object reconstruction and 3D-simultaneous localization and mapping were proposed in [1]. We choose the TUM Benchmark dataset for evaluating visual odometry and visual SLAM (Simultaneous Localization and Mapping) [21].

3. Architecture of system with information fusion from multiple Kinect sensors for object 3D reconstruction

This section provides the architecture of the proposed system with fusion of information from multiple Kinect sensors for object 3D reconstruction.

We consider the following steps: data acquisition, calibration, point cloud fusion.

Step 1: data acquisition.

The acquisition step obtains two types of data from each Kinect: a point cloud PC_i from each Kinect sensor; the 2D projections of the point cloud on the depth and color cameras. We represent depth information as point clouds, i.e. as a set of points in the 3D space with 3D coordinates. This allows easy aggregation of the depth information available from multiple registered cameras views. We retain RGB color and depth images as well as 3D point clouds for each camera and registered multi-view point clouds computed from the depth images. Note that the depth data provided by Kinect is noisy and incomplete. Noisy data possess to variations between 2 to 4 different discrete depth levels. So, we smooth the depth data by averaging over 9 depth frames. In order to recover incomplete regions, median filtering is utilized. The interference problem of multiple Kinect cameras was solved even for large angles between Kinects.

Step 2: calibration.

Our pose estimation procedure consists of two steps: pre-calibration and calibration.

The pre-calibration procedure provides an initial rough estimate of the camera poses. We calibrate the extrinsic matrix between different Kinect cameras using ICP and a cubic structure as the calibration object [18]. We choose one camera as reference to which other cameras are calibrated.

The second procedure performs a full camera calibration, i.e., it computes intrinsic and extrinsic parameters, by finding numerous 3D point matches between pairs of adjacent cameras. We select a less computationally expensive solution; that is, the R-Nearest Neighbor [8], which is described below.

First, apply the transformations obtained during the pre-calibration step to each of the point clouds $PC_i, i \in \{2, 3, 4\}$ to align them with the reference point cloud from the first sensor PC_1 . Once the point clouds are aligned, search for 3D points matches between the reference point cloud and the rest using the nearest neighbor approach inside a radius $R = 2$ millimeters. A point p is the R-near neighbor of a point q if the distance between p and q is at most R . The algorithm either returns the nearest R-near neighbor or concludes that no such point exists. Note that the point clouds must have overlapped points to find matches between pairs of cameras.

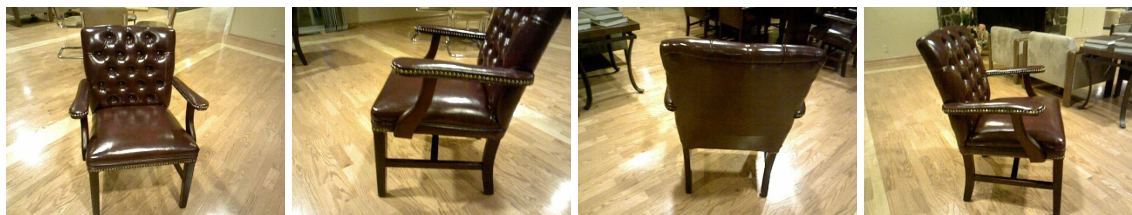


Figure 1. The RGB images of a chair is scanned by four Kinect sensors.

Step 3: point cloud fusion.

Fusing all the point clouds with color into a single one can be done using the calibration data from each camera [8]. After acquiring a depth and color frame from each Kinect sensor, we undistort the depth image and obtain the $[x, y, z]$ coordinates of each pixel in the 3D world space. The $[x, y, z]$ points are mapped onto the color frame using the intrinsic and extrinsic parameters of the color camera to obtain the corresponding color of each 3D point. Finally, to merge the colored 3D data we use the extrinsic parameters of each camera, i.e., the pose between each camera and the reference are utilized to transform all the point clouds into a single reference frame.

4. Experimental results

In this section, we present the results to evaluate the performance of our proposed method of calibration and fusion of multiple Kinect sensors for object 3D reconstruction.

In our experiments four Kinect V2 sensors connected to four computers which featured a Intel core i7 processor with four cores and 16 GB of memory were used. To evaluate the performance of our calibration method, we carried out point cloud fusion and 3D reconstruction of a chair from database [21]. The object was placed in the field of view of the four cameras and depth and color frames by each Kinect V2 sensor was acquired. Fig. 1 shows RGB images of a chair taken by four Kinect sensors using real data.

The Kinect accuracy is not very good and degrades with distance [22]. However, our calibration method with computed distortion parameters yields a better accuracy than the Kinects built-in mapping. To evaluate our calibration results qualitatively, we mapped the $[x, y, z]$ points onto the color frame using the intrinsic and extrinsic parameters of the color camera. In this way the corresponding color of each 3D point is obtained. Finally, by merging the colored 3D data from the four Kinect sensors we got a 3D fused point cloud which then used for reconstruction of a meshed object with MeshLab. Fig. 2 shows the 3D reconstruction of a chair. The 3D model is fine and accurate.

The experiment has shown that the proposed method of calibration and fusion of multiple Kinect sensors is able to provide accurate 3D object reconstruction. All frames from multiple Kinect sensors are fused correctly.

5. Conclusion

In this paper, we proposed the system of fusion of multiple Kinect sensors for object 3D reconstruction. The procedure contains the following steps: data acquisition, calibration multiple Kinect sensors, point cloud fusion from multiple Kinect sensors. The implementation is done in MATLAB using the Kinect V2. We evaluated the performance of our proposed system for object 3D reconstruction using real data. The experiment has shown that the proposed method of calibration and fusion of multiple Kinect sensors for object 3D reconstruction is accurate.

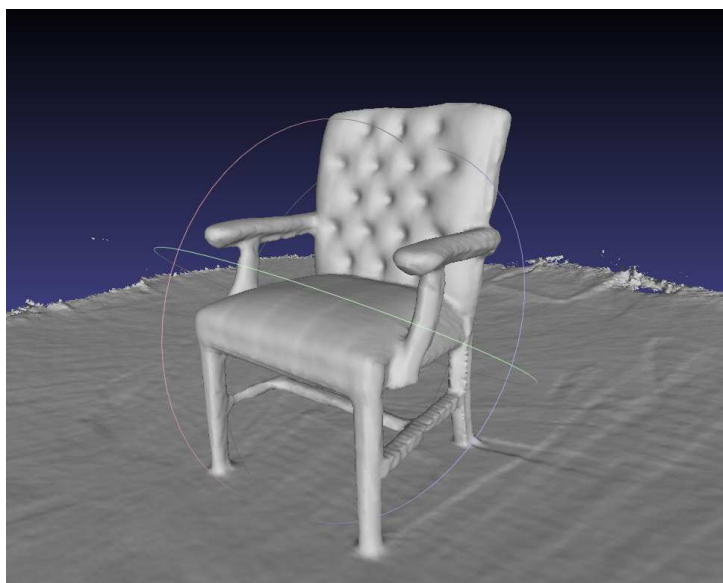


Figure 2. 3D Reconstruction of objects using our proposed method.

6. Acknowledgments

This work was supported by the Russian Science Foundation, grant no. 17-76-20045.

7. References

- [1] Rgb-d datasets using microsoft kinect or similar sensors: a survey / Ziyun Cai, Jungong Han, Li Liu, Ling Shao // *Multimedia Tools and Applications*. — 2017. — Vol. 76(3). — P. 4313–4355.
- [2] 3d scanning deformable objects with a single rgb-d sensor / M. Dou, J. Taylor, H. Fuchs [et al.] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — [S. l. : s. n.], 2015. — P. 493–501.
- [3] Real-time geometry, albedo, and motion reconstruction using a single rgb-d camera / Kaiwen Guo, Feng Xu, Tao Yu [et al.] // *ACM Trans. Graph.* — 2017. — Vol. 36(3). — P. 32:1–32:13.
- [4] Point cloud mapping measurements using kinect rgb-d sensor and kinect fusion for visual odometry / N. Namitha, S.M. Vaitheeswaran, V.K. Jayasree, M.K. Bharat // *Procedia Computer Science*. — 2016. — Vol. 89. — P. 209–212.
- [5] Towards a realistic indoor world reconstruction: Preliminary results for an object-oriented 3d rgb-d mapping / ChangHyun Jun, Jaehyeon Kang, Suyong Yeon [et al.] // *Intelligent Automation & Soft Computing*. — 2017. — Vol. 23(2). — P. 207–218.
- [6] Susanto, W. 3D Object Detection with Multiple Kinects / Wandi Susanto, Marcus Rohrbach, Bernt Schiele // *Computer Vision – ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part II* / Ed. by Andrea Fusiello, Vittorio Murino, Rita Cucchiara. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2012. — P. 93–102.
- [7] Kowalski, M. Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors / Marek Kowalski, Jacek Naruniec, Michal Daniluk // *International Conference on 3D Vision (3DV)*. — Los Alamitos, CA, USA : IEEE Computer Society, 2015. — P. 318–325.
- [8] A multiple camera calibration and point cloud fusion tool for kinect {V2} / Diana-Margarita Cordova-Esparza, Juan R. Terven, Hugo Jimenez-Hernandez, Ana-Marcela Herrera-Navarro // *Science of Computer Programming*. — 2017. — Vol. 143. — P. 1–8.
- [9] Partially occluded object reconstruction using multiple kinect sensors [Text] / Tabassum Nasrin, Faliu Yi, Samarjit Das, Inkyu Moon // *Proc.SPIE*. — Vol. 9117. — [S. l. : s. n.], 2014. — P. 9117–9125.
- [10] A gradient-based approach for interference cancelation in systems with multiple kinect cameras / Sen Xiang, Li Yu, Qiong Liu, Zixiang Xiong // *ISCAS*. — [S. l.] : IEEE, 2013. — P. 13–16.
- [11] Fusing multiple image transformations and a thermal sensor with kinect to improve person detection ability / L. Susperregi, A. Arruti, E. Jauregi [et al.] // *Engineering Applications of Artificial Intelligence*. — 2013. — Vol. 26(8). — P. 1980–1991.

- [12] Implementation of Human Action Recognition System Using Multiple Kinect Sensors / Beom Kwon, Doyoung Kim, Junghwan Kim [et al.] // *Advances in Multimedia Information Processing – PCM 2015: 16th Pacific-Rim Conference on Multimedia*, Gwangju, South Korea, September 16-18, 2015, Proceedings, Part I. — Cham : Springer International Publishing, 2015. — P. 334—343.
- [13] Data fusion of multiple kinect sensors for a rehabilitation system / Huibin Du, Yiwen Zhao, Jianda Han [et al.] // *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016*, Orlando, FL, USA, August 16-20, 2016. — [S. l. : s. n.], 2016. — P. 4869—4872.
- [14] Simultaneous multiple kinect v2 for extended field of view motion tracking / P. J. Noonan, J. Ma, D. Cole [et al.] // *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. — [S. l. : s. n.], 2015. — P. 1—4.
- [15] Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation / P. N. Pathirana, S. Li, H. M. Trinh, A. Seneviratne // *IEEE Transactions on Industrial Electronics*. — 2016. — Vol. 63(3). — P. 1822–1833.
- [16] Calibration of multiple kinects with little overlap regions / Mitsuru Nakazawa, Ikuhisa Mitsugami, Hitoshi Habe [et al.] // *IEEE Transactions on Electrical and Electronic Engineering*. — 2015. — Vol. 10. — P. S108—S115.
- [17] Multiple kinect v2 calibration / Diana-Margarita Cordova-Esparza, Juan Terven, Hugo Jimenez-Hernandez [et al.] // *Automatika—Journal for Control, Measurement, Electronics, Computing and Communications*. — 2017. — 02. — Vol. 60.
- [18] Calibration of multiple kinect depth sensors for full surface model reconstruction / Kwan Pang Tsui, Kin Hong Wong, Changling Wang [et al.] // *Proc.SPIE*. — Vol. 10011. — [S. l. : s. n.], 2016. — P. 10011–10016.
- [19] Simultaneous calibration: A joint optimization approach for multiple kinect and external cameras / Yajie Liao, Ying Sun, Gongfa Li [et al.] // *Sensors*. — 2017. — Vol. 17(7). — P. 1491.
- [20] Real-time rgb-d image stitching using multiple kinects for improved field of view / Hengyu Li, Hang Liu, Ning Cao [et al.] // *International Journal of Advanced Robotic Systems*. — 2017. — Vol. 14(2). — P. 1729881417695560.
- [21] A large dataset of object scans / Sungjoon Choi, Qian-Yi Zhou, Stephen Miller, Vladlen Koltun // *CoRR*. — 2016. — Vol. 1602.02481.
- [22] Khoshelham, K. Accuracy and resolution of kinect depth data for indoor mapping applications / Kouros Khoshelham, Sander Oude Elberink // *Sensors*. — 2012. — Vol. 12(2). — P. 1437–1454.