

FUSION OF INFORMATION FROM MULTIPLE KINECT SENSORS FOR 3D OBJECT RECONSTRUCTION

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

¹ *Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences, Orenburg, Russia,*

² *Department of Mathematics, Chelyabinsk State University, Chelyabinsk, Russia*

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 the depth data. Next, the precision of coordinate mapping between sensors data for registration of depth and color images is evaluated. We test a proposed system for 3D object reconstruction with four Kinect V2 sensors and present reconstruction accuracy results. Experiments and computer simulation are carried out using Matlab and Kinect V2.

Key words: multiple sensors, Kinect, 3D object reconstruction, fusion.

Citation: Ruchay AN, Dorofeev KA, Kolpakov VI. Fusion of information from multiple Kinect sensors for 3D object reconstruction. *Computer Optics* 2018; 42(5): 898-903. DOI: 10.18287/2412-6179-2018-42-5-898-903.

Acknowledgments: The Russian Science Foundation (project #17-76-20045) financially supported the work.

Introduction

The 3D reconstruction of objects is a popular task, with applications in the field of medicine, architecture, games, and film industry. The 3D reconstruction has many applications in object recognition, object retrieval, scene understanding, object tracking, autonomous navigation, human-computer interaction, telepresence, telesurgery, reverse engineering, virtual maintenance and visualization [1–9].

Accurate 3D reconstruction of objects is an important ingredient for many robotic, assisted living, surveillance, and industrial applications [10]. Consumer depth cameras, such as Microsoft's 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. The object at several instants can be captured using a single sensor such as in [11, 12] where the object rotates, and the camera is still or as in [13, 14] where the sensor moves around a still object. Another approach uses multiple sensors that capture the object simultaneously as in [15–17]. However, fusing information from multiple depth cameras for object detection has not been investigated in detail. In this paper, we investigate 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 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.

1. Related work

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

In order to fill small holes and to eliminate noise, the median and binomial filters were used [18–23]. Moreover, the use of the color information in the point correspondence process avoids false positives matches and, therefore, leads to a more reliable registration. Note that by adjusting iterative closest point (ICP) algorithm and reconstruction parameters it is possible to improve the registration and appearance of details that were invisible with just one scan due to the sensor-limited precision. Finally, it was shown [24] that 3D smooth surface of objects can be reconstructed using low precision sensors such as Kinect.

In [15] 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 [25].

The interference problem of multiple Kinect cameras dramatically degrades the depth quality. In paper [26] 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 [27]. This work is based on the fusion of infor-

mation from Kinect and a thermal sensor (thermopile) mounted on top of a mobile platform.

In [28] authors proposed a human action recognition system using multiple Kinect sensors based on multi-view skeleton integration. In [13] 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 movement's rehabilitation therapy system for hemiplegic patients was proposed in [29]. 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 [30]. Multiple Kinect Fusion allows head motion tracking in the case when partial and complete occlusions of the face occur. To increase the accuracy of human joint position estimation for rehabilitation and physiotherapy information-fusion from multiple Kinects was used [31]. 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 [16]. This paper describes 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 issues 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 [32], 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 planar regions from the environment.

A method to calibrate multiple Kinect sensors was developed in [33]. 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 were investigated in [34]. 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 simultaneous calibration of relative poses of a Kinect and three external cameras by op-

timizing a cost function and adding corresponding weights to the external cameras in different locations is proposed in [35]. 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 [36]. 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 [17]. 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 [17].

RGB-D datasets for different applications including object reconstruction and 3D-simultaneous localization and mapping were proposed in [10]. We choose the TUM Benchmark dataset for evaluating visual odometry and visual SLAM (Simultaneous Localization and Mapping) [37].

2. The proposed system

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

We consider the following steps of the proposed system: 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.

Camera space refers to the 3D coordinate system used by Kinect, where the coordinate system is defined as follows [17]: the origin is located at the center of the infrared sensor on the Kinect; the positive X direction goes to the sensor's left; the positive Y direction goes up; the positive Z goes out in the direction the sensor is facing; and the units are in meters.

The data acquisition procedure follows the next steps [17]:

1. On each sensor, we use the color camera to capture images synchronously in order to detect the colored markers. For this, we use small color blobs that satisfy

- certain constraints: on the 1D calibration pattern, the color points should lie on a line with the fixed length; and for the 2D pattern, the color points must have fixed distances given by the 2D pattern. The red, green, blue, and yellow markers are defined as $a, b, c,$ and d respectively. All the cameras must find the three or four markers to count as a valid frame.
- Map the coordinates of $a, b, c,$ and d by the coordinate mapping from color space to camera space obtaining $A_i, B_i, C_i,$ and D_i where $A_i, B_i, C_i, D_i \in R^3$ and $i \in \{1, 2, 3, 4\}$ is the Kinect identification.
 - Map the complete depth frame (424×512 depth values) to camera space points producing a point cloud PC_i for each Kinect sensor where $PC_i \in R^{217088 \times 3}$ using the depth camera and the coordinate mapping.
 - Transfer the 3D coordinates of the points (A_i, B_i, C_i, D_i) and the point clouds PC_i .

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 algorithm and markers as the calibration object [34, 33, 17]. We define the depth camera of the first Kinect as the reference.

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 [17], 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. 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.

The pose is represented by a 4×4 rigid transformation containing the rotation R and translation t that align each of the cameras with the reference. To obtain these transformations we used the camera space points A_i, B_i, C_i, D_i from each Kinect sensor obtained in the data acquisition step. Setting the first Kinect ($i=1$) as the reference, the problem of obtaining the pose boils down to obtaining the best rotations R_i and translations t_i that align the points from the Kinect sensors $M_i = [A_i B_i C_i D_i], i \in \{2, 3, 4\}$ to the points in the reference Kinect (M_1).

The calibration of the extrinsic matrix procedure consists of the following steps:

- Solve for R_i and t_i :

$$M_1 = R_i M_i + t_i,$$

where R_i and t_i are rotations and translations applied to each set of points $M_i, i \in \{2, 3, 4\}$ to align them with the reference M_1 .

- Find the centroids of the 3D points M_i :

$$centroid_i = (A_i + B_i + C_i + D_i) / 4.$$

- Move the points to the origin and find the optimal rotation R_i :

$$H_i = \sum_{j=1}^3 (M_i^j - centroid_i)(M_i^j - centroid_i)^T,$$

$$[U_i, S_i, V_i] = SVD(H_i), \quad R_i = V_i U_i^T,$$

where H_i is the covariance matrix of the i -th Kinect and SVD denotes the singular value decomposition.

- Find the translation t_i as

$$t_i = -R_i centroid_i + centroid_1.$$

- Apply a refining step using Iterative Closest Point (ICP) on each aligned point cloud with the reference, to minimize the difference between them. The aligned point clouds will be denoted as $PC_i, i \in \{2, 3, 4\}$.

The matching points between the reference Kinect and the rest (obtained in the point cloud matching step) are the 3D points in the world reference and will be denoted by P_{wi} for $i \in \{2, 3, 4\}$, the 2D projections of these points on the image plane are denoted by $u_i = (u, v)$ and are known from the acquisition step.

In homogeneous coordinates, the mapping between points $P_w = (x, y, z)$ and their 2D projections $u = (u, v)$ in the image plane is given by

$$\lambda u = (K \ 0) \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} P_w,$$

where K is the intrinsic parameters matrix or camera parameters, and $[R, t]_{w \rightarrow c}$ the extrinsic parameters, R is a 3×3 rotation matrix that defines the camera orientation and t is a translation vector that describe the position of the camera in the world.

Our goal is to compute the intrinsic parameters K which contains the focal length (α, β), a skew factor (γ), and the principal point (u_0, v_0) for fixed extrinsic parameters $[R, t]$ obtained with the pre-calibration step. We estimate the intrinsic parameters as proposed in [33]

$$[\alpha, \beta, \mu, u_0, v_0]^T = argmin_h \left[\sum_{i=1}^J (A_i h - u_i)^T (A_i h - u_i) \right],$$

where J is a number of 3D point matches $(x_i, y_i, z_i), i \in 1, \dots, J$ and the matrix A_i^T is

$$\begin{pmatrix} r_{11}x_i + r_{12}y_i + r_{13}z_i + t_x & 0 \\ r_{31}x_i + r_{32}y_i + r_{33}z_i + t_z & 0 \\ r_{21}x_i + r_{22}y_i + r_{23}z_i + t_x & 0 \\ r_{31}x_i + r_{32}y_i + r_{33}z_i + t_z & 0 \\ 1 & 0 \\ 0 & r_{21}x_i + r_{22}y_i + r_{23}z_i + t_y \\ 0 & r_{31}x_i + r_{32}y_i + r_{33}z_i + t_z \\ 0 & 1 \end{pmatrix},$$

where r_{jk} and t_x, t_y, t_z are the rotation and translation elements of the known pose transformation between the reference camera and the camera which we want to estimate the intrinsic parameters.

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 [17]. 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 an 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 [37]. The object was placed in the field of view of the four cameras, wherein depth map and RGB frames were acquired by each Kinect V2 sensor. Fig. 1 shows RGB images and depth maps of a chair taken by four Kinect sensors using real data.

The Kinect accuracy is not very good and degrades with distance [38]. However, our calibration method with computed distortion parameters yields a better accuracy than the Kinect’s 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.

Also, we have experimental results for evaluation of the performance of the proposed system with fusion of information from RGB-D sensor for object 3D reconstruction. The metric of evaluation is the root mean square error (RMSE) of measurements.

$$RMSE = \sqrt{(E(ED - RD)^2)},$$

where ED is the estimated measurement by a device and RD is the real known measurement of the object.

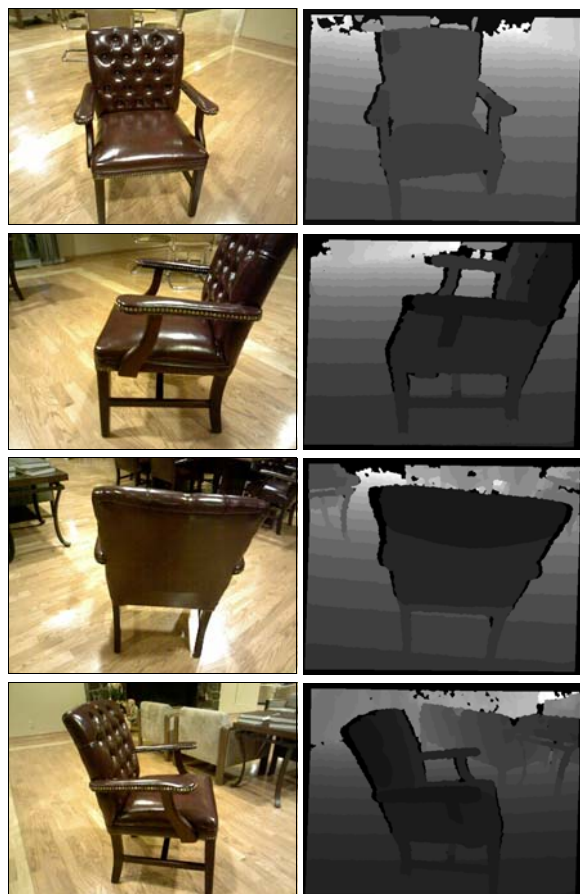


Fig. 1. The RGB images and depth maps of a chair are scanned by four Kinect sensors

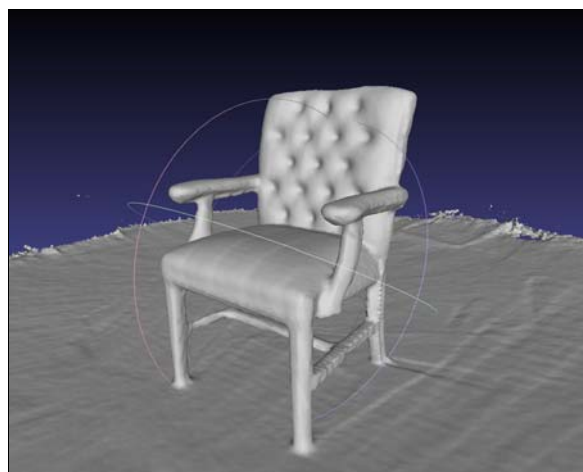


Fig. 2. 3D Reconstruction of objects using our proposed method

By taking five measurements, the average values were calculated. Corresponding RMSE values calculated for Kinect V2 are shown in Table 2.

Table 1. Results of measurements

	Real World	Kinect V2	RMSE
Length	734	732	2
Width	867	872	5
Height	895	889	6
Diagonal	639	647	8
Geodetic	678	682	4
Area	67960	68010	50

The results show that Kinect V2 yields a necessary accurate 3D model of the object. The obtained accuracy allows us to make all measurements on the 3D model as on a real object.

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 of 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.

References

- Echeagaray-Patron BA, Miramontes-Jaramillo D, Kober V. Conformal parameterization and curvature analysis for 3D facial recognition. 2015 International Conference on Computational Science and Computational Intelligence (CSCI) 2015; 843-844. DOI: 10.1109/CSCI.2015.133.
- Echeagaray-Patron BA, Kober V. 3D face recognition based on matching of facial surfaces. Proc SPIE 2015; 9598: 95980V. DOI: 10.1117/12.2186695.
- Smelkina NA, Kosarev RN, Nikonorov AV, Bairikov IM, Ryabov KN, Avdeev AV, Kazanskiy NL. Reconstruction of anatomical structures using statistical shape modeling [In Russian]. Computer Optics 2017; 41(6): 897-904. DOI: 10.18287/2412-6179-2017-41-6-897-904.
- Vokhmintsev A, Makovetskii A, Kober V, Sochenkov I, Kuznetsov V. A fusion algorithm for building three-dimensional maps. Proc SPIE 2015; 9599: 959929. DOI: 10.1117/12.2187929.
- Kotov AP, Fursov VA, Goshin YeV. Technology for fast 3D-scene reconstruction from stereo images [In Russian]. Computer Optics 2015; 39(4): 600-605. DOI: 10.18287/0134-2452-2015-39-4-600-605.
- Sochenkov I, Sochenkova A, Vokhmintsev A, Makovetskii A, Melnikov A. Effective indexing for face recognition. Proc SPIE 2016; 9971: 997124. DOI: 10.1117/12.2238096.
- Picos K, Diaz-Ramirez V, Kober V, Montemayor A, Pantrigo J. Accurate three-dimensional pose recognition from monocular images using template matched filtering. Opt Eng 2016; 55(6): 063102. DOI: 10.1117/1.OE.55.6.063102.
- Echeagaray-Patrón BA, Kober V. Face recognition based on matching of local features on 3D dynamic range sequences. Proc SPIE 2016; 9971: 997131. DOI: 10.1117/12.2236355.
- Echeagaray-Patrón BA, Kober VI, Karnaukhov VN, Kuznetsov VV. A method of face recognition using 3D facial surfaces. J Commun Technol Electron 2017; 62(6): 648-652. DOI: 10.1134/S1064226917060067.
- Cai Z, Han J, Liu L, Shao L. RGB-D datasets using microsoft kinect or similar sensors: a survey. Multimed Tools Appl 2017; 76(3): 4313-4355. DOI: 10.1007/s11042-016-3374-6.
- Dou M, Taylor J, Fuchs H, Fitzgibbon A, Izadi S. 3D scanning deformable objects with a single RGBD sensor. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015; 493-501. DOI: 10.1109/CVPR.2015.7298647.
- Guo K, Xu F, Yu T, Liu X, Dai Q, Liu Y. Real-time geometry, albedo, and motion reconstruction using a single RGB-D camera. ACM Trans Graph 2017; 36(4): 44a. DOI: 10.1145/3072959.3083722.
- Namitha N, Vaitheeswaran SM, Jayasree VK, Bharat MK. Point cloud mapping measurements using kinect RGB-D sensor and kinect fusion for visual odometry. Procedia Computer Science 2016; 89: 209-212. DOI: 10.1016/j.procs.2016.06.044.
- Jun C, Kang J, Yeon S, Choi H, Chung T-Y, Doh NL. Towards a realistic indoor world reconstruction: Preliminary results for an object-oriented 3D RGB-D mapping. Intelligent Automation & Soft Computing 2017; 23(2): 207-218. DOI: 10.1080/10798587.2016.1186890.
- Susanto W, Rohrbach M, Schiele B. 3D object detection with multiple kinects. ECCV'12 Proceedings of the 12th international conference on Computer Vision 2012; 2: 93-102. DOI: 10.1007/978-3-642-33868-7_10.
- Kowalski M, Naruniec J, Daniluk M. Livescan3D: A fast and inexpensive 3D data acquisition system for multiple Kinect v2 Sensors. 2015 International Conference on 3D Vision (3DV) 2015: 318-325. DOI: 10.1109/3DV.2015.43.
- Córdova-Esparza D-M, Terven JR, Jiménez-Hernández H, Herrera-Navarro A-M. A multiple camera calibration and point cloud fusion tool for Kinect V2. Science of Computer Programming 2017; 143: 1-8. DOI: 10.1016/j.scico.2016.11.004.
- Aguilar-Gonzalez PM, Kober V. Design of correlation filters for pattern recognition with disjoint reference image. Opt Eng 2011; 50(11): 117201. DOI: 10.1117/1.3643723.
- Aguilar-Gonzalez PM, Kober V. Design of correlation filters for pattern recognition using a noisy reference. Opt Commun 2012; 285(5): 574-583. DOI: 10.1016/j.optcom.2011.11.012.
- Ruchay A, Kober V. Clustered impulse noise removal from color images with spatially connected rank filtering. Proc SPIE 2016; 9971: 99712Y. DOI: 10.1117/12.2236785.
- Ruchay A, Kober V. Removal of impulse noise clusters from color images with local order statistics. Proc SPIE 2017; 10396: 1039626. DOI: 10.1117/12.2272718.
- Ruchay A, Kober V. Impulsive noise removal from color video with morphological filtering. Proc SPIE 2017; 10396: 1039627. DOI: 10.1117/12.2272719.
- Ruchay A, Kober V. Impulsive noise removal from color images with morphological filtering. In Book: van der Aalst W, Ignatov DI, Khachay M, Kuznetsov SO, Lempitsky V, Lomazova IA, Loukachevitch N, Napoli A, Panchenko A, Pardalos PM, Savchenko AV, Wasserman S, eds. Analysis of Images, Social Networks and Texts (AIST 2017). Cham: Springer; 2018: 280-291. DOI: 10.1007/978-3-319-73013-4_26.
- Takimoto RY, Tsuzuki M de SG, Vogelaar R, Martins T de C, Sato AK, Iwao Y, Gotoh T, Kagei S. 3D reconstruction and multiple point cloud registration using a low precision RGB-D sensor. Mechatronics 2016; 35: 11-22. DOI: 10.1016/j.mechatronics.2015.10.014.
- Nasrin T, Yi F, Das S, Moon I. Partially occluded object reconstruction using multiple Kinect sensors. Proc SPIE 2014; 9117: 91171G. DOI: 10.1117/12.2053938.
- Xiang S, Yu L, Liu Q, Xiong Z. A gradient-based approach for interference cancelation in systems with multiple Kinect cameras. 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013) 2013; 13-16. DOI: 10.1109/ISCAS.2013.6571770.
- Susperregi L, Arruti A, Jauregi E, Sierra B, Martinez-Otzeta JM, Lazkano E, Ansuategui A. Fusing multiple image transformations and a thermal sensor with kinect to improve person detection ability. Engineering Applications of Artificial

- Intelligence 2013; 26(8): 1980-1991. DOI: 10.1016/j.engappai.2013.04.013.
28. Kwon B, Kim D, Kim J, Lee I, Kim J, Oh H, Kim H, Lee S. Implementation of human action recognition system using multiple Kinect sensors. In Book: Ho YS, Sang J, Ro Y, Kim J, Wu F, eds. *Advances in Multimedia Information Processing – PCM 2015*. Cham: Springer; 2015; I: 334-343. DOI: 10.1007/978-3-319-24075-6_32.
29. Du H, Zhao Y, Han J, Wang Z, Song G. Data fusion of multiple kinect sensors for a rehabilitation system. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2016: 4869-4872. DOI: 10.1109/EMBC.2016.7591818.
30. Noonan PJ, Ma J, Cole D, Howard J, Hallett WA, Glocker B, Gunn R. Simultaneous multiple Kinect v2 for extended field of view motion tracking. 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) 2015: 1-4. DOI: 10.1109/NSSMIC.2015.7582070.
31. Pathirana PN, Li S, Trinh HM, Seneviratne A. Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation. *IEEE Transactions on Industrial Electronics* 2016; 63(3): 1822-1833. DOI: 10.1109/TIE.2015.2497662.
32. Nakazawa M, Mitsugami I, Habe H, Yamazoe H, Yagi Y. Calibration of multiple kinects with little overlap regions. *IEEJ Transactions on Electrical and Electronic Engineering* 2015; 10(S1): S108-S115. DOI: 10.1002/tee.22171.
33. Córdova-Esparza D-M, Terven JR, Jiménez-Hernández H, Vázquez-Cervantes A, Herrera-Navarro A-M, Ramírez-Pedraza A. Multiple Kinect V2 calibration. *Automatika* 2016; 57(3): 810-821. DOI: 10.7305/automatika.2017.02.1758.
34. Tsui KP, Wong KH, Wang C, Kam HC, Yau HT, Yu YK. Calibration of multiple Kinect depth sensors for full surface model reconstruction. *Proc SPIE* 2016; 10011: 100111H. DOI: 10.1117/12.2241159.
35. Liao Y, Sun Y, Li G, Kong J, Jiang G, Jiang D, Cai H, Ju Z, Yu H, Liu H. Simultaneous calibration: A joint optimization approach for multiple kinect and external cameras. *Sensors* 2017; 17(7): 1491. DOI: 10.3390/s17071491.
36. Li H, Liu H, Cao N, Peng Y, Xie S, Luo J, Sun Y. Real-time RGB-D image stitching using multiple Kinects for improved field of view. *Int J Adv Robot Syst* 2017; 14(2): 1-8. DOI: 10.1177/1729881417695560.
37. Choi S, Zhou Q-Y, Miller S, Koltun V. A large dataset of object scans. arXiv:1602.02481. 2016.
38. Khoshelham K, Elberink SO. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* 2012; 12(2): 1437-1454. DOI: 10.3390/s120201437.

Authors' information

Alexey N. Ruchay (b. 1986) graduated from the Chelyabinsk State University in 2008, PhD. Currently he works as the leading researcher at Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences, and associate professor at Chelyabinsk State University. Research interests include machine vision, processing of signal and image, and biometrics. E-mail: ran@csu.ru.

Konstantin A. Dorofeev (b.1989) graduated from the Chelyabinsk State University in 2011. Engineer-researcher at Chelyabinsk State University. Research interests: machine vision, and processing of signal and image. E-mail: kostuan1989@mail.ru.

Vladimir I. Kolpakov (b. 1987) graduated from the Orenburg State University in 2010, PhD. Currently working as a research fellow researcher at Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences. Scientific interventions include breeding, genetics, and evaluation of cattle. E-mail: vkolpakov056@yandex.ru.

*Code of State Categories Scientific and Technical Information (in Russian – GRNTI): 28.17.19.
Received September 21, 2018.*