Accuracy analysis of 3D object reconstruction using RGB-D sensor

A N Ruchay[1,2], K A Dorofeev[2], A V Kober[1]

[1]Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences, 9 Yanvarya street 29, Orenburg, Russia, 460000
[2]Department of Mathematics, Chelyabinsk State University, Bratiev Kashirinykh street 129, Chelyabinsk, Russia, 454001

Abstract. In this paper, we propose a new method for 3D object reconstruction using an RGB-D sensor. The RGB-D sensor provides RGB images as well as depth images. Since the depth and RGB color images are captured with one sensor of an RGB-D camera placed in different locations, the depth image should be related to the color image. After matching of the images (registration), point-to-point corresponding between two images is found, and they can be combined and represented in the 3D space. In order to obtain a dense 3D map of the 3D object, we design an algorithm for merging information from all used cameras. First, features extracted from color and depth images are used to localize them in a 3D scene. Next, Iterative Closest Point (ICP) algorithm is used to align all frames. As a result, a new frame is added to the dense 3D model. However, the spatial distribution and resolution of depth data affect to the performance of 3D scene reconstruction system based on ICP. The presented computer simulation results show an improvement in accuracy of 3D object reconstruction using real data.

1. Introduction

The 3D reconstruction of objects is a popular task, with applications in the field of medicine, architecture, games, agriculture, 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, 2, 3, 4, 5, 6, 7, 8, 9, 10]. The geometry of an object can be reconstructed from laser range scans, a set of different photos, monocular cameras and stereo cameras; each technology possesses disadvantages and limitations. RGB-D cameras like Microsoft Kinect or ASUS Xtion Pro are sensing systems equipped with an RGB camera, infrared projector, and infrared sensor. They can gather RGB information and the depth map simultaneously.

The ICP (Iterative Closest Point) is one of the most commonly used methods for pairwise alignment, where the rotation and translation to align two point clouds are determined. Many variants on the base ICP algorithm were proposed because of the limitations in the ICP algorithm: different cost functions [11] and others [12]; two-Pass ICP with color constraint to improve the error minimization process [13]; 3D representation using a set of planes to perform the registration [14]; point-to-plane matchings instead of point-to-point matchings in standard

ICP algorithm [15]; a new registration algorithm based on the Matching Signed Distance Fields [16].

In order to fill small holes and to eliminate noise, the median and binomial filters were used [17, 18, 19, 20, 21, 22]. 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 ICP 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 that with help of low precision sensors as Kinect 3D smooth surface of objects can be reconstructed [23].

A new low-cost approach to reconstruct real-time of a 3D object with Kinect sensor uses a SLAM algorithm (Simultaneous Localization and Mapping) [24]. SLAM provides an approximated solution of 3D reconstruction because the accuracy of the system often depends on a heuristic algorithm for obtaining relevant reference points.

In order to improve the accuracy and robustness of the ICP algorithm, a regularization by incorporating the spatial distances of SIFT feature pairs with dynamically adjusted weights to balance errors is proposed in [12]. A new outlier rejection method based on a dynamic thresholding and leverages the structure and sparse feature pairs from the texture of the RGB images were also suggested [12]. A new method to robustly estimates the camera motion in a dynamic environment based on RANSAC algorithm was proposed in the article [25].

We propose to utilize a graph-based SLAM algorithm [24] with loop closure detection using dense color and depth images obtained from the RGB-D camera. We show that our system is able to perform the SLAM for three-dimensional modeling in real-time.

The paper is organized as follows. Section 2 discusses related works. Section 3 describes the proposed system for object 3D reconstruction using an RGB-D sensor. In Section 4 the experimental results are discussed. Finally, Section 5 presents our conclusions.

2. Related work

An approach for realistic surface geometry reconstruction using high-frequency features of color images from Kinect is given in [26]. In order to achieve a better accuracy in the global alignment of scans, it was used a weighted ICP algorithm. An efficient 3D reconstruction approach combining a depth information based 3D model with RGB information to refine the reconstruction results when the camera fails to acquire the correct depth information is presented in [27].

The efficiency problems and real-time processing are discussed in [28]; that is, highly resource consuming of dense point cloud registration algorithm based on an ICP method and slow surfel update procedure for sufficiently large data clouds. For resolving these problems a sparse ICP algorithm a frustum culling procedure exploits the hierarchical map structure. A novel real-time algorithm for simultaneously reconstructing the geometry using a single RGB-D camera is suggested in [29]. Moreover, this method can provide global anchors by including sparse SIFT features. It allows us to potentially achieve better results for loop-closing motions.

A patch-based illumination invariant visual odometry was proposed in [30], which works well in the irregular illumination change. The planar patch selection process is employed and the illumination change model is adopted in each extracted patch in order to consider the partial light variations. Using the robust weighting function and the efficient second-order minimization (ESM) image alignment method, the proposed cost function reflecting the illumination changes is minimized. As a result, the proposed method can accurately estimate the motion of the camera regardless of the partial lighting changes. A method for visualization of occluded objects using multiple Kinect sensors at different locations is proposed in [31].

A novel robust 3D reconstruction system with an RGB-D camera was proposed in [32]. The visual and geometry features and combined SFM technique to make registration more robust

especially in depth map missing cases were used. In order to solve the drift problem, 3D information is used to detect the loop closure and to perform global refinement.

Prior-based Multi-Candidates RANSAC (PMCSAC) algorithm was proposed [32] to make the feature matching more robust and efficient in order to handle the repeated textures/structures. Missing geometry due to depth missing can be effectively completed by combining multi-view stereo and mesh deformation techniques [32].

For solving the problem of a wider range of object deformations merging a sequence of images from a single range sensor into a unified 3D model, without requiring an initial template was proposed in [33]. However, although complex scene topologies can be handled, a regular topology is restricted to be constant throughout the sequence. So, if the coarse-scale reconstruction does not correctly choose the topology, it cannot switch to a fine scale, and the computational cost is also high. A novel approach to the template-driven capture of dense detailed non-rigid deformations from video sequences to carry out simultaneously 2D dense registration and 3D shape inference is presented in [34].

In work [35] a model-based scalable 3D scene reconstruction system based on a CAD model was proposed. The accuracy of KinectFusion algorithm and analysis the noise effect of reconstruction and localization errors based on a CAD model was tested in [36].

An iterative low-cost method for 3D body registration, dealing with unconstrained movements and accuracy is suggested in [37]. A novel method for 3D object reconstruction from RGB-D data that applies sub mapping to 3D bundle adjustment is presented in [38]. A 3D shape descriptor for object recognition with RGB-D sensors is proposed in [39].

A novel volumetric multi-resolution mapping system for RGB-D images was proposed in the paper [40]. Proposed approach generates a textured triangle mesh from a signed distance function that it continuously updates as new RGB-D images arrive. For this, an octree uses as the primary data structure which allows us to represent the scene at multiple scales, and it allows us to grow the reconstruction volume dynamically.

A system for autonomous flight using RGB-D sensors was suggested [41]. It uses a combination of visual odometry techniques and mapping and is able to conduct all sensing and computation required for local position control.

A method to estimate both odometry and scene flow with RGB-D cameras is presented in the paper [42]. The main advantage of the proposed approach is that it provides accurate results with a very low runtime.

A new visual odometry system based on the alignment between consecutive frames by minimization both on the photometric and geometric error was proposed [43], where the original ICP algorithm for frame alignment and visual odometry computation was completely substituted by the proposed method. The use of the inverse depth instead of the depth to parametrize the geometric error is the main contribution of the proposed method.

A new RGB-D based method was proposed to improve the accuracy and robustness of visual odometry [44], where the set of line segments generate from maximum-clique filtered point correspondences.

In the paper [45], a robust RGB-D DVO algorithm BaMVO was proposed for use in a dynamic environment, where consecutive depth images were warped using the preobtained ego-motions to equalize the viewpoints. The background image was estimated using a nonparametric model from the differences between consecutive pairs in the warped depth images. The energy function between successive RGB-D images was represented by the form of the weighted least squares for the calculation of DVO.

A novel 3D geometry enhanced superpixel algorithm for RGB-D data is presented in [46]. The depth map is converted into 3D geometrical information, and superpixels are iteratively clustered according to a distance metric designed from the color information and 3D geometry. By introducing the geometrical information, the proposed superpixel method overcomes the

difficulty in distinguishing adjacent objects with similar colors.

3. The proposed system

This section provides the proposed system with a fusion of information from a moving RGB-D sensor for object 3D reconstruction.

3D reconstruction with our system consists of the following steps:

(i) Registration a point cloud PC_i using RGB and Depth data.

(ii) Detection and matching of global key points in RGB data of PC_i and PC_{i-1} with SURF algorithm.

(iii) Remove outliers with RANSAC.

(iv) Count transformation matrix with ICP using the associate 3D points of the inliers.

(v) Apply hybrid approach that combines ICP odometry with the RGB-D odometry.

(vi) Adding results to a general model (dense 3D map).

The ICP odometry has a problem in tracking drifting or failure for the presence of smooth surfaces [47]. Thus, we have implemented a hybrid approach that combines ICP odometry with the RGB-D odometry, which aims to minimize the photometric error between consecutive RGB-D frames [48]. The hybrid approach combines frame-to-model registration with the increased stability provided by the use of both geometric and photometric cues [49].

The hybrid odometry approach estimates transformation T_i for each frame i by minimizing the following function:

$$R(T_i) = R_{ICP}(T_i) + \lambda R_{RGBD}(T_i),$$

where $R_{ICP}(T_i)$ is based on the point-to-plane error used in ICP algorithm:

$$R_{ICP}(T_i) = \sum_{(p,q)\in K} \|(p - T_i q)^\top n_p\|^2,$$

with n_p is the surface normal at p and K is a set of corresponding point pairs found by projective data association. $R_{RGBD}(T_i)$ is the photometric error between frames $i-1$ and i given as

$$R_{RGBD}(T_i) = \sum_x \|I_i(\pi(T_i^{-1} T_{i-1} \pi^{-1}(x, D_{i-1}))) - I_{i-1}(x)\|^2,$$

where D_k is the depth image from frame k, x passes over all pixels in frame $i-1$, π is a projection operator that projects a 3D point in the camera coordinate frame to the image data, π^{-1} is an operator, that produces a 3D point in the camera coordinate frame that corresponds to a given pixel, and I_k is the intensity of x pixel in color image from frame k. The coefficient λ balances the two counted errors terms. It is calculated empirically and is identical for all reconstructions.

4. Experimental results

In this section, we present experimental results for evaluation of the performance of the proposed system with a fusion of information from an 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.

The object to be mapped is a small box (Kinect V2 adapter box). Fig. (1) shows a real-world object and one frame from RGB-D device (Asus Xtion Pro).

Figure 1. Real world object and one frame from the RGB-D sensor.
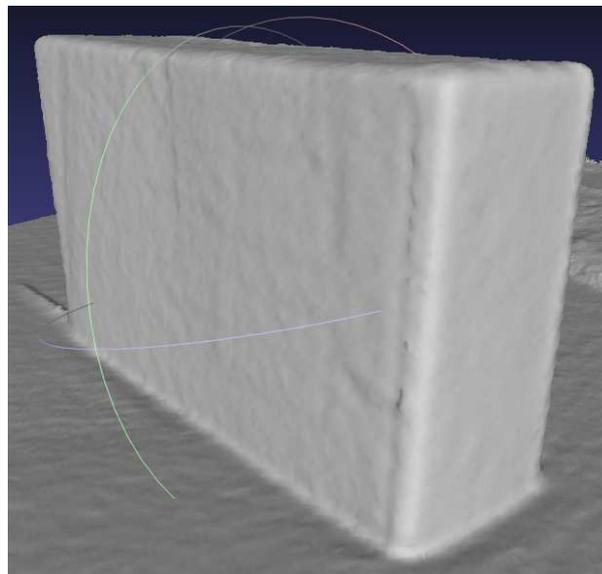
Figure 2. 3D model of the object.

Using the iterative ICP algorithm we obtain the corresponding 3D model of the object as shown in Fig. 2.

Linear distances between 2 points in the point cloud are measurements shown in Fig. 3. The geodesic distance between two points passing through surfaces and the surface area is also shown in Fig. 3.

The average values over five measurements acquired by our system. Corresponding RMSE values calculated for two sensors (Asus Xtion Pro and Kinect V2) are shown in Table 1.

The results show that Kinect V2 yields a more 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
We proposed a system for reconstruction of the 3D model of a real-life object with a small error in measurements. We have also developed a software to perform arbitrary measurements on
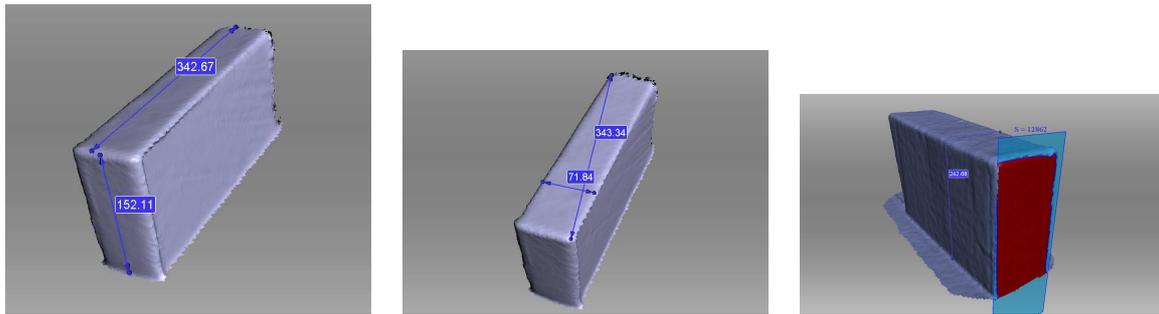
Figure 3. Measurements of the box.

Table 1. Results of measurements

	Real World	AsusXtionPro	RMSE	Kinect V2	RMSE
Length	345	342.67	2.33	344	1.00
Width	76	71.84	4.16	75.6	0.40
Height	158	152.11	5.89	155.87	2.13
Diagonal	352	343.34	8.66	348.78	3.22
Geodetic	228	242.68	14.68	234.85	6.85
Area	12008	12862	854.00	12003	5.00

the 3D model of the object. Our system is sensitive to poor illumination. The obtained results showed that it is possible to use multiple snapshots (sequence of frames) from low precision sensors such as Kinect for accurate reconstruction of the 3D model of objects. The accuracy of 3D reconstructed models is confirmed by the measured distances: linear (distance between two points), geodetic (distance over surfaces), and surface area.

In future, we improve the performance of the proposed algorithms to get more accurate 3D models of objects. We also develop more complicated measurement methods such as a section of an object, calculation of circle perimeter and section area, calculation of arbitrary geodetic distances and curvature of the lines.

6. References

[1] Echeagaray-Patron B A, Miramontes-Jaramillo D and Kober V 2015 *International Conference on Computational Science and Computational Intelligence (CSCI)* 843-844
[2] Echeagaray-Patron B A and Kober V 2015 *Proc. SPIE* **9598** 95980V
[3] Sochenkov I and Vokhmintsev A 2015 *Procedia Engineering* **129** 440-445
[4] Vokhmintsev A, Makovetskii A, Kober V, Sochenkov I and Kuznetsov V 2015 *Proc. SPIE* **9599** 959929
[5] Tihonkih D, Makovetskii A and Kuznetsov V 2016 *Proc. SPIE* **9971** 99712D
[6] Sochenkov I, Sochenkova A, Vokhmintsev A, Makovetskii A and Melnikov A 2016 *Proc. SPIE* **9971** 997124
[7] Picos K, Diaz-Ramirez V, Kober V, Montemayor A and Pantrigo J 2016 *Optical Engineering* **55** 55-55-11
[8] B Adriana and Echeagaray-Patron V K 2016 *Proc. SPIE* **9971** 9971-9976
[9] Echeagaray-Patrón B A, Kober V I, Karnaukhov V N and Kuznetsov V V 2017 *Journal of Communications Technology and Electronics* **62** 648-652
[10] Smelkina N, Kosarev R, Nikonorov A, Bairikov I, Ryabov K, Avdeev A and Kazanskiy N 2017 *Computer Optics* **41(6)** 897-904 DOI: 10.18287/2412-6179-2017-41-6-897-904
[11] Xie J, Hsu Y F, Feris R S and Sun M T 2013 *IEEE International Symposium on Circuits and Systems* 2904-2907
[12] Xie J, Hsu Y F, Feris R S and Sun M T 2015 *Journal of Visual Communication and Image Representation* **32** 194-204
[13] Rhee S M, Lee Y B and Lee H E 2014 *IEEE International Conference on Consumer Electronics* 89-90
[14] Thomas D and Sugimoto A 2013 *IEEE International Conference on Computer Vision* 2800-2807
[15] Chen Y and Medioni G 1991 *Proceedings IEEE International Conference on Robotics and Automation* **3** 2724-2729
[16] Masuda T 2002 *Computer Vision and Image Understanding* **87** 51-65
[17] Aguilar-Gonzalez P M and Kober V 2011 *Optical Engineering* **50** 50-59

[18] Aguilar-Gonzalez P M and Kober V 2012 *Optics Communications* **285** 574-583

[19] Ruchay A and Kober V 2016 *Proc. SPIE* **9971** 99712Y

[20] Ruchay A and Kober V 2017 *Proc. SPIE* **10396** 10396

[21] Ruchay A and Kober V 2017 *Proc. SPIE* **10396** 10396

[22] Ruchay A and Kober V 2018 *Analysis of Images, Social Networks and Texts* (Cham: Springer International Publishing) 280-291

[23] Takimoto R Y, de Sales Guerra Tsuzuki M, Vogelaar R, de Castro Martins T, Sato A K, Iwao Y, Gotoh T and Kagei S 2016 *Mechatronics* **35** 11-22

[24] Aguilar W G, Rodríguez G A, Álvarez L, Sandoval S, Quisaguano F and Limaico A 2017 *Real-Time 3D Modeling with a RGB-D Camera and On-Board Processing* (Cham: Springer International Publ.) 410-419

[25] Dib A and Charpillet F 2015 *International Conference on Advanced Robotics* 1-7

[26] Lee K R and Nguyen T 2016 *Mach. Vision Appl.* **27** 377-385

[27] Pan H, Guan T, Luo Y, Duan L, Tian Y, Yi L, Zhao Y and Yu J 2016 *Neurocomputing* **175** 644-651

[28] Wilkowski A, Kornuta T, Stefanczyk M and Kasprzak W 2016 *Applied Mathematics and Computer Science* **26** 99-122

[29] Guo K, Xu F, Yu T, Liu X, Dai Q and Liu Y 2017 *ACM Trans. Graph.* **36** 32:1-32:13

[30] Kim P, Lim H and Kim H J 2015 *IEEE/RSJ International Conference on Intelligent Robots and Systems* 3688-3694

[31] Nasrin T, Yi F, Das S and Moon I 2014 *Proc. SPIE* **9117** 9117-9115

[32] Wang K, Zhang G and Bao H 2014 *IEEE Transactions on Image Processing* **23** 4893-4906

[33] Dou M, Taylor J, Fuchs H, Fitzgibbon A and Izadi S 2015 *IEEE Conference on Computer Vision and Pattern Recognition* 493-501

[34] Yu R, Russell C, Campbell N D F and Agapito L 2015 *IEEE International Conference on Computer Vision* 918-926

[35] Cheng S C, Su J Y, Chen J M and Hsieh J W 2017 *Model-Based 3D Scene Reconstruction Using a Moving RGB-D Camera* (Cham: Springer International Publishing) 214-225

[36] Jiang S Y, Chang N Y C, Wu C C, Wu C H and Song K T 2014 *IEEE International Conference on Automation Science and Engineering* 1020-1025

[37] Villena-Martinez V, Fuster-Guillo A, Saval-Calvo M and Azorin-Lopez J 2017 *3D Body Registration from RGB-D Data with Unconstrained Movements and Single Sensor* (Cham: Springer International Publishing) 317-329

[38] Maier R, Sturm J and Cremers D 2014 *German Conference on Pattern Recognition (GCPR)* **8753** 54-65

[39] Liu Z, Zhao C, Wu X and Chen W 2017 *Sensors* **17** 451

[40] Steinbrucker F, Sturm J and Cremers D 2014 *IEEE International Conference on Robotics and Automation* 2021-2028

[41] Huang A S, Bachrach A, Henry P, Krainin M, Maturana D, Fox D and Roy N 2017 *Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera* (Cham: Springer International Publishing) 235-252

[42] Jaimez M, Kerl C, Gonzalez-Jimenez J and Cremers D 2017 *IEEE International Conference on Robotics and Automation* 3992-3999

[43] Gutierrez-Gomez D, Mayol-Cuevas W and Guerrero J 2016 *Robot. Auton. Syst.* **75** 571-583

[44] Zhang Y, Hou Z, Yang J and Kong H 2016 *23rd International Conference on Pattern Recognition* 2764-2769

[45] Kim D H and Kim J H 2016 *IEEE Transactions on Robotics* **32** 1565-1573

[46] Yang J, Gan Z, Gui X, Li K and Hou C 2013 *3-D Geometry Enhanced Superpixels for RGB-D Data* (Cham: Springer International Publishing) 35-46

[47] Newcombe R A, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison A J, Kohli P, Shotton J, Hodges S and Fitzgibbon A 2011 *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality* 127-136

[48] Kerl C, Sturm J and Cremers D 2013 *IEEE International Conference on Robotics and Automation* 3748-3754

[49] Choi S, Zhou Q Y, Miller S and Koltun V 2016 *arXiv:1602.02481*

Acknowledgments