Digital model reconstruction through 3D Stereo Depth camera

a faster method exploiting robot poses

Zaki, Ahmed Magdy Ahmed; Carnevale, Marco; Giberti, Hermes; Schlette, Christian

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2022.12.354

Publication date:
2023

Document version:
Final published version

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 217 (2023) 1542–1549

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Industry 4.0 and Smart Manufacturing

Digital model reconstruction through 3D Stereo Depth camera: a faster method exploiting robot poses

Ahmed Magdy Ahmed Zaki[a,*], Marco Carnevale[a], Hermes Giberti[a], Christian Schlette[b]

[a] Università degli Studi di Pavia, Dipartimento di Ingegneria Industriale e dell'Informazione, via A. Ferrata 5, 27100 Pavia, Italy
[b] SDU Robotics, Maersk McKinney Moller Institute, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark

Abstract

In industrial robotic contact-based operations it is necessary to have a detailed geometrical knowledge of the work-piece to automate the generation of the working trajectory. In many cases, the digital model is not available or it differs from its as built actual conditions. Vision sensors, especially 3D vision sensors allows to scan the work-piece and reconstruct its digital copy that is used for the generation of the robotic working trajectory. In this paper we compare two algorithms for the generation of the 3D model of an unknown work-piece based on the use of the RGB-D images taken from different perspectives. The first technique, based on standard image reconstruction methods commonly used for reconstruction of indoor scenes where error in the order of few centimeters is negligible, is exploited in this work in order to be exploited in contact-based robotic operations where the geometrical error has to be limited to few millimeters. It is based on the analysis of the images to estimate the camera pose taking each image. Based on the pose, the algorithm integrates then the information contained in the images in a unique volume representing the scene. The second algorithm directly uses the known poses of the robot, stored while capturing each image, to integrate the information to create the model. The two algorithms are compared to evaluate the accuracy, acquisition time and elaboration time. The analysis is done considering two kinds of objects, the first one with regular shape, the second one corresponding to an actual industrial object.

Keywords: 3D reconstruction, 3D vision, contact-based operations, industrial robots;

1. Introduction

Contact-based robotic applications in manufacturing systems often require a digital model of the work-pieces the robot tool must come into contact with. Although CAD models of the product ”as planned” are often available [1], particularly small and medium-sized enterprises (SMEs) often lack tools and procedures to capture digital models of the products ”as manufactured”. In addition, if product ranges show high variabilities, this gap in the digital coverage

* Ahmed Magdy Ahmed Zaki. Tel.: +39-038-298-5264
E-mail address: ahmedmagdyahme.zaki01@universitadipavia.it

1877-0509 © 2022 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2022.12.354

of robotic manufacturing processes is often even more impacting. For these reasons, the capability of reconstructing a digital model of the unknown objects in unstructured environment is an important issue in the attempt to spread robotic machines in the manufacturing industry. 3D vision techniques based on the use of synchronised color and depth images (RGB-D images) allows to reconstruct the digital model of an unknown objects through several snapped images. In [2] and in [3] two solutions based on the use of fixed 3D vision sensors are used to capture images of a work-piece and and generate the glue deposition trajectory in shoes manufacturing. In the first a 6-axis robot is used while in the second a cartesian robot is used. In [4], the authors proposed a robotic scanning system for adaptive surface mapping that calculates after the acquisition of each image the optimal following scanning pose. In this way a minimum scanning poses are used based on the specific unknown work-piece.

By elaborating the RGB-D images and comparing them through Odometry and pixel features matching algorithms the camera motion can be reconstructed and hence the pose of the camera capturing every image [5]. These calculated poses allow then to integrate the information contained in the different RGB-D images in a unique volume representing the scene [6], or as in the considered case an unknown work-piece.

When the camera is applied on a robot wrist, the camera motion can be easily reconstructed based on the knowledge of the robot poses where the pictures are actually taken. Using robot poses and hence camera poses would significantly speed up the 3D model reconstruction process, making two of the elaboration steps unnecessary.

In [7], the grasping point is calculated elaborating a point-cloud reconstructed exploiting the robot poses to capture multi-view images. The efficiency of the proposed solution is given by the success rate of the system to grasp the objects. In [8], multi-view color and depth images are used to generate the welding trajectory following welding seams of a metal object. In [9] point-cloud captured by a 3D camera is elaborated to generate the welding path in correspondence of v-type grooves in the work-piece. The accuracy of the welding trajectory is done evaluating the overlap of the generated trajectory and the trajectory followed by a professional operator.

For accurate 3D reconstruction for trajectory planning of contact-based operations, usually accurate laser scanners or structured light 3D cameras are used. In [10] solutions for the 3D reconstruction of the work-piece integrating multi-view point-clouds captured by a laser scanner are proposed. The deviation of the reconstructed models with respect to the CAD models are lower than 0.5 millimeters.

We propose two methods to reconstruct the 3D model of unknown work-pieces based on multi-view color and depth images captured by a low-cost 3D camera (Realsense D415). Contact-based operations require an accurate trajectory, for that we propose an evaluation procedure to check the feasibility of using a low-cost 3D camera in such accurate tasks. The evaluation procedure is based on the measurement of the accuracy of the 3D models obtained by applying the proposed two methods. This accuracy evaluation is given by quantifying the degree of similarity of the overall reconstructed 3D models obtained by applying each of the two techniques and the exact CAD model of the work-piece. For the performance evaluation, it is considered also the computational effort and the time necessary to obtain the 3D models that is an important parameter to decrease the process cycle time and possibly increase the production volume.

This paper compares the Odometry-based method based on images feature matching and the robot poses-based method in which the robot poses are directly exploited for images integration and 3D model reconstruction. The algorithms are first tested metal box with known dimensions and regular geometric features. The accuracy of the reconstructed digital models, using the different algorithms, is evaluated through the comparison of each one with the actual box CAD model and the calculation of the errors between the two surfaces using Hausdorff Distance [11] that is a method to evaluate the similarity between two point sets or surfaces. It is shown that the possibility of using the poses of the robot, known in advance, significantly fasten the entire process in the case of a static scene. Secondly the same comparison is carried out on a wooden work-piece actually used in manufacturing of wooden sailboats. The wooden work-piece is selected since it represents more complicated geometrical features in terms of curves, edges and color variation. Also in this case the accuracy of the model reconstructed through robot poses is comparable/more accurate with the odometry-based techniques but the computational time is much lower.

The paper is structured as follows. In section 2, the two techniques used for the 3D reconstruction are explained. In section 3, a technical analysis comparing the performance of the two techniques is shown. In section 4, the two algorithm are applied to reconstruct the 3D model of a complicated object. Finally, the conclusion is drawn in section 5.

2. Proposed solutions

This section describes the developed two solutions for 3D reconstruction of work-pieces in industrial contact-based applications. The two solutions are based on the integration of RGB-D images of the work-pieces. The images are taken from different perspectives in such a way that covers all the faces of the work-piece or the area interested by the robotic contact-based operation such as surface finishing, grinding and sanding.
To capture the dataset of RGB-D images the robotic setup shown in figure 1 is used. The setup consists of a Techman TM5 6-axis collaborative robot. Intel Realsense D415 stereo depth camera is attached to the robot end-effector.

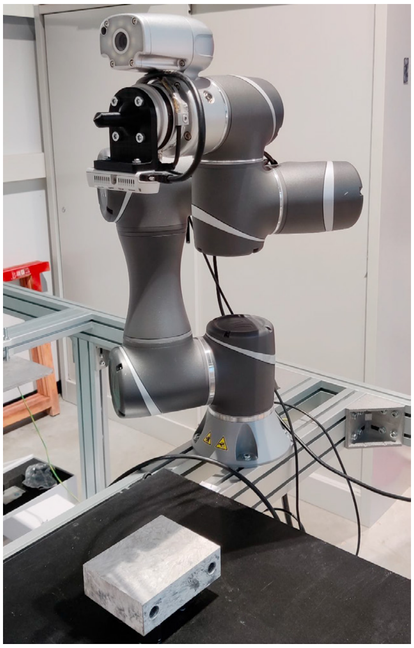


Fig. 1. Setup

The 3D camera uses different sensors to acquire color and depth images of the scene in its field of view. The resolution of the sensors can be configured up to 720x1280 pixels and a frame rate up to 30 frames per second. Based on the resolution chosen, a minimum distance has to be respected between the camera and the observed scene to guarantee the measurement of the depth. The configuration used is the highest resolution of 720x1280 pixels that implies a minimum distance of 450 mm from the work-piece. The two algorithms are developed in Python programming language and using Open source libraries that are Librealsense for images acquisition and elaboration and Open3D for 3D data elaboration [12].
To scan the work-pieces, the robot moves the 3D camera following a predefined trajectory orienting the camera towards the work-piece in all the images. The camera start acquiring from top image and then moves to cover all the four faces of the object. The trajectory chosen is the same for the two developed techniques but the acquisition sequence and timing are different in the two cases.

2.1. Odometry based technique

In Odometry based technique the 3D camera start acquiring the images in a synchronized way with the movement of the robot. During the trajectory following, several images are taken. The Odometry-based algorithm [5, 6] consists of the following steps. Estimation of the camera poses while capturing the different images. This is done using visual Odometry that compares the images two by two to find the translation and the rotation that if applied to one of them make it matches with the other. This process is based on the comparison of pixels visual features in the picked two images and find the optimal pose minimizing the positional error. The poses are then used to calculate a homogeneous transformation matrix for each image that geometrically transforms it to be with respect to a common reference frame. The information of each pair of color and depth images are fetched and then integrated in a unique voxel grid created using Truncated Signed Distance Function (TSDF) [13]. TSDF algorithm is used to represent several point-clouds taken from different point if views in the same volume. If a pixel is visible in several images, TSDF manages the uncertainty of pixel positions calculating a weighted average between the values from the different images giving higher weight to the value in the image in which the pixel is more visible (pixel normal parallel to the camera line of sight taking that image). Finally the constructed voxel grid is elaborated to estimate the smooth surface connecting the voxels of the observed scene.

2.2. Robot based technique

In the robot poses technique, the images are captured only with correspondence to a specific poses that cover specific points of the work-piece. The robot when arrives to the defined poses it stops for 500 ms to allow the 3D camera to capture the RGB-D images. The minimum number of the images is 5 that allows to cover the work-piece from all the four sides and also with an extra top view image. Knowing exactly the pose of the robot end-effector and hence the pose of the camera while capturing a certain image, allows for the reconstruction of the motion and to calculate the homogeneous transformation matrices for each image to refer them to a common reference frame. In this way, the Odometry process for camera pose estimation is not necessary that decrease significantly the time necessary for the elaboration. The calculated homogeneous transformation matrices are then used to integrate the information using TSDF as explained in the other case.

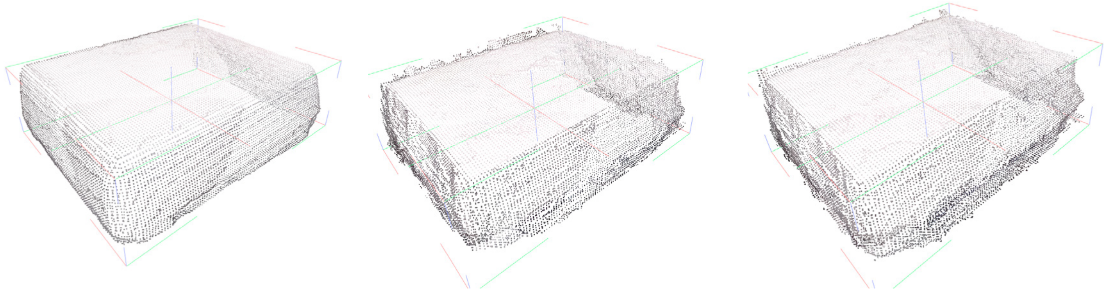
3. Results

In this section, the accuracy of the two algorithms is evaluated comparing the 3D models reconstructed of both of them with respect to the ground truth or the known CAD model. The evaluation is based on three Key Performance Indices (KPI) that are acquisition time, elaboration time and geometrical accuracy with respect to the CAD model.
The object considered for the evaluation process is a metal cube of the dimensions 127 mm * 99 mm * 40 mm that is shown in the figure 2. This kind of objects are mostly to be found in manufacturing systems on which it is necessary to apply a surface finishing procedure.

Fig. 2. Metal box

The first comparison is done between the 3D model reconstructed using RGB-D Odometry-based algorithm. The algorithm integrate 100 pairs of RGB-D images captured of the static work-piece by a moving camera attached to the robot end-effector.
The second comparison is done comparing the 3D reconstructed model of the box using the poses-based algorithm with the 3D CAD model of the work-piece. The pose-based technique is based on the use of minimum number of images and hence minimum elaboration time and needed resources. To be able to cover all the faces of the work-piece the minimum possible number of images is five. That are covering its top face and other four images covering all the

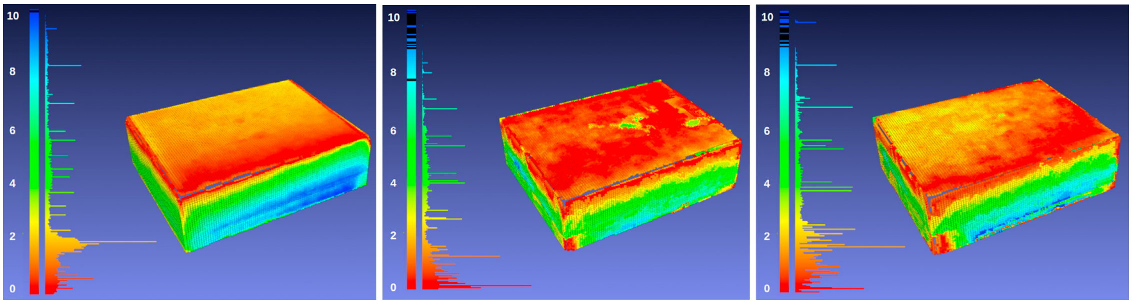
four faces.
We consider also other scanning procedure using eight images of the work-piece. In which, beside the minimum possible number of five images we add extra three images covering the corners of the metal box. In this way all the faces are visible in two images to limit the dependency on possible acquisition noise and reflection of the metal object.
Figure 3 shows the obtained 3D models applying the different 3D reconstruction algorithms. In figure 3a the Odometry-based algorithm result is shown. In figures 3b and 3c shows the results of the poses-based algorithm with five and eight images.



(a) (b) (c)

Fig. 3. Reconstructed 3D models. a)Odometry-based, b)Poses-based 5 images and c)Poses-based 8 images

To quantitatively evaluate the geometrical errors between the models, we use Hausdorff distance to measure the distance between all the points of the point-cloud of the constructed 3D model and the relative point in the point-cloud of the known CAD model of the metal box. To apply Hausdorff Distance algorithm it is necessary to align both of the 3D models and to overlap them. The reconstructed 3D model is always not identical to the 3D CAD model because of 3D camera acquisition accuracy and presence of extra light or object reflections. Applying Iterative Closest point (ICP) automatic registration algorithm for the alignment may fail. To overcome this problem, an alignment procedure based on the alignment of the boundary boxes of the two 3D models is used.
The result of applying Hausdorff Distance algorithm to two point-clouds is represented graphically in a color-map. The color-maps for the three reconstruction cases are shown in figure 4. In the map, the points that have small errors are represented by a red color. With the increase of the error the representation color transforms toward yellow and turns blue for maximum error values. The represented color-map is between 0 mm for red color and 10 mm for blue color. The histogram shown on the left of each image shows the number of points having a certain error level. The error color map can be projected on the point-cloud of the point-cloud of the CAD model. In this way a significant graphical representation is obtained to visualize the areas where the two point-clouds coincide and also where they mismatch. For all the three models most of the points have error lower than 2 mm. The upper face is the face that is having the minimum error and coincides approximately with the real CAD model. Lateral faces present higher errors. In the reconstructed 3D model using the Odometry-based technique, in figure 4a, the right lateral face is overestimated in correspondence to the bigger blue area. When Poses-based reconstruction is used, edges are reconstructed having more realistic values closer to the real value of the real CAD model and this is shown in correspondence with bigger red areas covering the edges and the corners.
Despite the fact that the 3D model reconstructed using Odometry-based technique seems smoother than those reconstructed using Poses-based approach as shown in figure 3. The error analysis using Hausdorff Distance showed that Poses-based approach is more realistic.
The time necessary to use the different techniques is an important evaluation KPI for industrial applications to decrease the cycle time of the 3D reconstruction process. We consider different periods of time for our evaluation. The first is the acquisition time or the time necessary to capture the required data that allow the 3D reconstruction. For the two techniques, the acquisition trajectory followed by the robot is the same. In Odometry-based technique the robot



(a) (b) (c)

Fig. 4. Error representation using Hausdorff Distance. a)Odometry-based, b)Poses-based 5 images and c)Poses-based 8 images

follows all the trajectory and the camera acquires several images of the object during the movement. In Poses-based technique, the same trajectory is followed by the robot with the stop periods in a predefined positions that allow to capture images that cover the work-piece from the interested faces.

Elaboration time is also used for the evaluation. This is the period of time needed to elaborate the images and to integrate all the information included in them to reconstruct the final 3D model. In Odometry-based technique, higher number of images are used and long optimization procedure is used to estimate the camera poses while capturing the images dataset. Hence, its elaboration time increases significantly with the increase of the dataset dimension. Poses-based algorithm integrates directly the information from the images to reconstruct the 3D model since the poses to capture the images are already known.

Table 1 shows a numerical comparison of the three cases considered. For each case, acquisition time, elaboration time and measurements accuracy are shown.

	Acquisition time [s]	Elaboration time [s]	Measurement accuracy [mm]		
			min	mean	max
Odometry-based (100 images)	13.25	864.00	0.0009	3.1588	10.4879
Poses-based (5 images)	10.12	5.70	0.0002	2.4448	10.3795
Poses-based (8 images)	14.15	10.02	0.0001	2.5834	9.1436

Table 1. Results comparison

Acquisition time for all the cases presents similar values. The highest value is in the case of eight images due to the necessity to stop the robot to capture each image for 500 ms to allow the acquisition of a RGB-D stable image. Elaboration time decreases significantly using the poses-based method, around 10 seconds, with respect to the Odometry-based case where the elaboration time is more than 14 minutes.

The table shows also the accuracy evaluation summary using Hausdorff Distance technique. Mean value of geometrical error decreases from 3.15 mm in the case of Odometry-based method to approximately 2.5 mm in case of Poses-based method. Adding more images to cover all the faces in at least two images allows to decrease the maximum error from 10.37 mm to 9.14 mm.

4. Application to an actual component in manufacturing industry

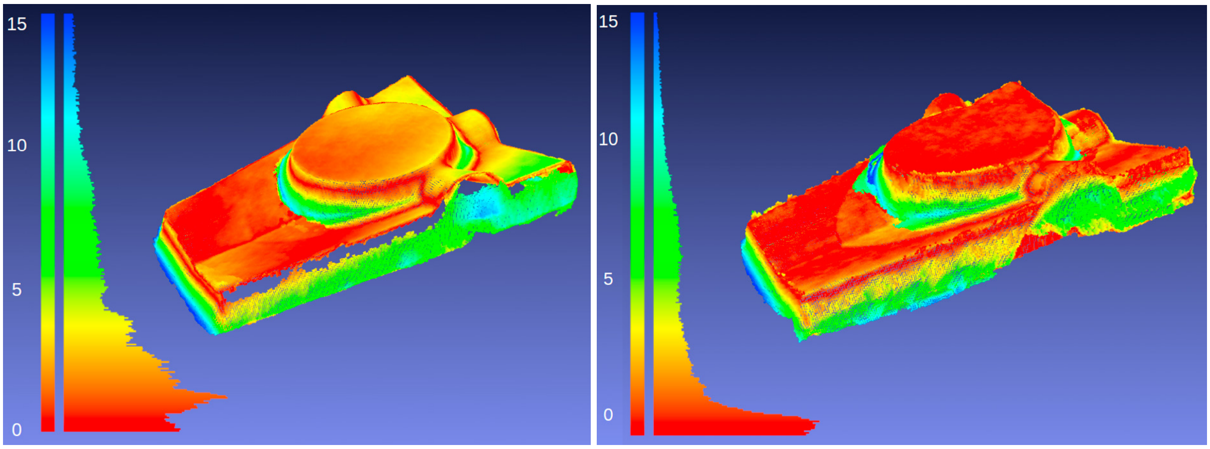
In this section, the two techniques are used to reconstruct the 3D model of a wooden object. The considered object, with respect to the metal box, represents more complicated features in terms of curves, edges and irregular color distribution. The object considered is the one shown in figure 5, used as a mould for sailboat fabrication.

The resulting 3D models are compared to the 3D CAD model of the object to evaluate their accuracy. Hausdorff Distance technique is used to quantify the geometrical errors. The results are shown in figure 6.

The color maps show the error for the Odometry-based technique in figure 6a and for poses-based in figure 6b. The error values represented are between zero and 15 mm. The histogram on the left of each color-map shows the quantity

Fig. 5. Wooden item

of the points at every error level. Using both of the techniques allows to have high number of points with error lower than 5 mm and very few points with error higher than 10 mm.



(a) (b)

Fig. 6. Wooden object analysis. a)Odometry-based reconstruction accuracy, b)Poses-based reconstruction accuracy

Considering the points having an error up to 15 mm, Odometry-based error a mean value of 4.212 mm while for Poses-based algorithm the mean value is 3.890 mm.

Hausdorff analysis shown in figure 6, shows that the two models are similar in the most of the work-piece. Odometry-based technique uses average depth values of many acquisitions to calculate the final value for every voxel. This allows to obtain a smoother flat surface and smoother and overestimated edges. For the areas with lower visibility it tends to overestimate them as shown in correspondence of the blue and yellow areas.

5. Conclusion

In this paper we propose two techniques, Odometry-based and Poses-based, for 3D reconstruction of an unknown work-piece interested by a robotic contact-based operation. The techniques are intended to create a digital copy of work-pieces of which the CAD model is not available or different from the actual fabricated work-piece. The techniques, using a low cost 3d vision sensor and RGB-D images, allows the 3D construction keeping the color of the work-piece that is important for industrial applications that is not possible to obtain using more accurate and very expensive laser scanners. We compare the performance of the two algorithms considering two different kinds of work-pieces. The technical analysis shows that the Poses-based technique decreases significantly the elaboration time with better accuracy with respect to the Odometry-based technique. This because of using a limited number of images with respect to the other technique. Images necessary for Odometry-based technique are between 50 and 100 that require from 8 to 15 minutes for elaboration and integration. While Poses-based technique is able to reconstruct the 3D model in few seconds and requires a number of images equivalent to the number of faces interested by the robotic operation. It is enough to have for every face one or two images where the face is clearly visible. The feasibility to apply the proposed solutions is highly dependent on the size of the considered work-piece. Using a bigger robotic manipulator with a higher reach or Cartesian robot may allow to scan and reconstruct the 3D models of bigger objects. Future works consists of fully automate the process of trajectory generation by developing a search routine to select the area interested by the operation and generate the trajectory to cover it from the reconstructed 3D model. In the proposed solution, the scanning trajectory is fixed in the two proposed methods to cover the robot reachable area and it is independent of the work-piece geometrical features. However, an adaptive scanning procedure to reconstruct the 3D model of the work-piece in a gradual manner can be developed. In this way, if the generated 3D model presents holes or inaccuracies due to shadow or there are undercuts that are not covered in the images an appropriate scanning pose has to be added to the trajectory to cover all the features of the work-piece interested by the contact-based operation.

References

[1] Castelli, Kevin, Ahmed Magdy Ahmed Zaki, Yevheniy Dmytriyev, Marco Carnevale, and Hermes Giberti. 2021. ”A Feasibility Study of a Robotic Approach for the Gluing Process in the Footwear Industry” Robotics 10, no. 1: 6. https://doi.org/10.3390/robotics10010006

[2] A. M. A. Zaki, Y. Dmytriyev, K. Castelli, M. Carnevale and H. Giberti, ”Trajectory Planning for Contact-Based Robotic Applications by Use of a 3D Stereo Depth Camera,” 2021 3rd International Conference on Robotics and Computer Vision (ICRCV), 2021, pp. 68-72, doi: 10.1109/ICRCV52986.2021.9546958.

[3] Stefano Pagano, Riccardo Russo, Sergio Savino, A vision guided robotic system for flexible gluing process in the footwear industry, Robotics and Computer-Integrated Manufacturing, Volume 65, 2020, 101965, ISSN 0736-5845, https://doi.org/10.1016/j.rcim.2020.101965.

[4] Mineo, C., Cerniglia, D., Ricotta, V. et al. Autonomous 3D geometry reconstruction through robot-manipulated optical sensors. Int J Adv Manuf Technol 116, 1895–1911 (2021). https://doi.org/10.1007/s00170-021-07432-5

[5] Park J, Zhou QY, Koltun V. Colored Point Cloud Registration Revisited. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017. p. 143–152

[6] J. Wang et al., ”High quality 3D reconstruction of indoor environments using RGB-D sensors,” 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2017, pp. 1739-1744, doi: 10.1109/ICIEA.2017.8283120.

[7] De Gregorio, D., Tombari, F., Di Stefano, L. (2016). RobotFusion: Grasping with a Robotic Manipulator via Multi-view Reconstruction. In: Hua, G., Jégou, H. (eds) Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science(), vol 9915. Springer, Cham. https://doi.org/10.1007/978-3-319-49409-8_54

[8] Kim, J., Lee, J., Chung, M. et al. Multiple weld seam extraction from RGB-depth images for automatic robotic welding via point cloud registration. Multimed Tools Appl 80, 9703–9719 (2021). https://doi.org/10.1007/s11042-020-10138-7

[9] R. Peng, D. Navarro-Alarcon, V. Wu and W. Yang, ”A Point Cloud-Based Method for Automatic Groove Detection and Trajectory Generation of Robotic Arc Welding Tasks,” 2020 17th International Conference on Ubiquitous Robots (UR), 2020, pp. 380-386, doi: 10.1109/UR49135.2020.9144861.

[10] Madhusudhan Rao R, Deepa Radhakrishna, Usha S, Development of a Robot-mounted 3D Scanner and Multi-view Registration Techniques for Industrial Applications, Procedia Computer Science, Volume 133, 2018, Pages 256-267, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.07.032. (https://www.sciencedirect.com/science/article/pii/S1877050918309773)

[11] N. Aspert, D. Santa-Cruz and T. Ebrahimi, ”MESH: measuring errors between surfaces using the Hausdorff distance,” Proceedings. IEEE International Conference on Multimedia and Expo, 2002, pp. 705-708 vol.1, doi: 10.1109/ICME.2002.1035879.

[12] Zhou, Qian-Yi, Jaesik Park, and Vladlen Koltun. ”Open3D: A modern library for 3D data processing.” arXiv preprint arXiv:1801.09847 (2018).

[13] Werner, D., Al-Hamadi, A., Werner, P. (2014). Truncated Signed Distance Function: Experiments on Voxel Size. In: Campilho, A., Kamel, M. (eds) Image Analysis and Recognition. ICIAR 2014. Lecture Notes in Computer Science(), vol 8815. Springer, Cham. https://doi.org/10.1007/978-3-319-11755-3_40