



Heriot-Watt University
Research Gateway

6D Pose Estimation for Precision Assembly

Citation for published version:

Skeik, O, Erden, MS & Kong, X 2023, 6D Pose Estimation for Precision Assembly. in *5th IEEE International Conference on Image Processing, Applications and Systems 2022.*, 10052989, IEEE, 5th IEEE International Conference on Image Processing, Applications and Systems 2022, Genova, Italy, 5/12/22.
<https://doi.org/10.1109/ipas55744.2022.10052989>

Digital Object Identifier (DOI):

[10.1109/ipas55744.2022.10052989](https://doi.org/10.1109/ipas55744.2022.10052989)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

5th IEEE International Conference on Image Processing, Applications and Systems 2022

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

6D pose estimation for precision assembly*

Ola Skeik¹, Mustafa Suphi Erden and Xianwen Kong

Abstract— The assembly of 3D products with complex geometry and material, such as in a concentrator photovoltaics solar panel unit, are typically conducted manually. This results in low efficiency, precision and throughput. This study is motivated by an actual industrial need and targeted towards automation of the currently manual assembly process. By replacing the manual assembly with robotic assembly systems, the efficiency and throughput could be improved. Prior to assembly, it is essential to estimate the pose of the objects to be assembled with high precision. The choice of the machine vision is important and plays a critical role in the overall accuracy of such a complex task. Therefore, this work focuses on the 6D pose estimation for precision assembly utilizing a 3D vision sensor. The sensor we use is a 3D structured light scanner which can generate high quality point cloud data in addition to 2D images. A 6D pose estimation method is developed for an actual industrial solar-cell object, which is one of the four objects of an assembly unit of concentrator photovoltaics solar panel. The proposed approach is a hybrid approach where a mask R-CNN network is trained on our custom dataset and the trained model is utilized such that the predicted 2D bounding boxes are used for point cloud segmentation. Then, the iterative closest point algorithm is used to estimate the object's pose by matching the CAD model to the segmented object in point cloud.

I. INTRODUCTION

Next generation 3D products with complex geometry and material are becoming more and more popular. Assembly of these products is challenging as precision assembly is required. Currently, the assembly of industrial parts with such complex geometry is conducted manually. However, manual assembly usually results low efficiency and precision [1] and also results low throughput. To improve efficiency and increase throughput high precision robotic assembly systems need to be developed. To this end, a prerequisite to perform a robotic assembly task is to estimate the pose of the objects to be assembled where pose estimation is defined as estimating the object's location and orientation in the real world (with reference to the camera) [2]. 6D pose estimation is a challenging task on its own but when it is a prerequisite for precision assembly, the problem becomes more challenging as high accuracy is critical. Consequently, the choice of the vision system in the robotic assembly system is crucial and plays a vital role in achieving such complex and challenging task. With recent advances in 3D technology, 3D vision systems have gained much attention and have been recently utilized in the 6D pose estimation problem.

In this paper we focus on the problem of 6D pose estimation for precision assembly of complex pieces of an industrial

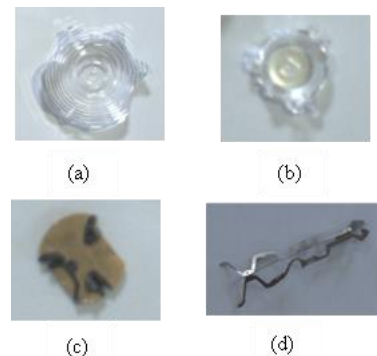


Fig. 1. Pieces of the industrial product to be assembled: (a) primary lens, (b) secondary lens, (c) solar-cell, (d) tripod-leg

product using 3D vision. The pieces of the industrial product are shown in Fig.1. The 3D vision sensor we use is a 3D structured light scanner which can export high quality point cloud data in addition to 2D images.

A. Background

6D pose estimation has been an active field of study and has recently gained the attention of different research communities such the robotics community [3]–[8]. Comprehensive surveys relating to this topic can be found in [9]–[11]. 6D pose estimation is an important task in robotics as it forms the basis for other robotic tasks such as grasping, bin-picking, assembly, etc. A formal definition of the 6D pose estimation problem is given in [2], [10], [12], [13] which we briefly state here. Given two sources of information as an input, a training dataset $T = \{T_1, T_2, \dots, T_n\}$ and an input image I , a 6D pose estimator predicts the 6D pose of an occurrence of object with a confidence score. Fig. 2 shows a block diagram of the 6D pose estimation problem.

Different methods have been proposed in the literature for the 6D pose estimation problem which vary depending on the source of input data whether it is RGB, RGB-D, or point cloud data such as, for example, [7], [14]–[16]. It is well known that using both color and depth information is superior over using only color information to retrieve the 6D pose. Moreover, the 6D pose estimation methods have been categorized as traditional methods, deep learning based methods, and tradition methods combined with deep learning methods (hybrid approach) [11]. As pointed out in [11], one of the main difference between traditional and deep learning methods is in feature extraction which is more powerful in the latter since convolution neural networks (CNNs) are used to

* This work has been supported by EPSRC grant EP/T024844/1

The authors are with the School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AL (ad9423@coventry.ac.uk¹, m.s.erden@hw.ac.uk, X.Kong@hw.ac.uk).

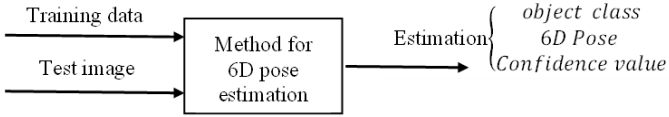


Fig. 2. Block diagram of the 6D pose estimation problem

extract features rather than hand-craft feature extraction as in the former. Moreover, it has been concluded in [11] that deep learning combined with traditional methods can produce better results.

In this work, we use a hybrid approach to estimate the 6D pose of the objects. The method leverages both 2D images and 3D point cloud data to estimate the objects poses. First, we leverage from the power of CNNs and train an instance segmentation model (Mask R-CNN) to locate the objects in the 2D image from which the objects in point cloud are segmented. Then, we utilize a classical pose estimation method, iterative closest point (ICP) algorithm, to estimate the pose of the objects by aligning the CAD model of the object with the segmented object in the scene point cloud.

B. Related work

In this section, a review on related literature is provided. The review is classified as *hybrid approaches*, where approaches on the 6D pose estimation methods using a hybrid approach are reviewed, and *robotic assembly*, which provides a review on literature addressing the 6D pose estimation problem specifically for robotic assembly tasks.

Hybrid approaches: Hybrid approaches to pose estimation typically consist of two stages where in the first stage a deep CNN network is used for object segmentation, and in the second stage different authors propose and utilize different pose estimation techniques to achieve and improve the accuracy of 6D pose estimation. For example, in [17], a hybrid approach was proposed to estimate the 6D pose of rigid objects. A two-stage pipeline was used where instance segmentation via deep learning was performed first, followed by a point-pair voting method to recover the object’s pose. The authors in [3] follow a similar approach to that in [17] and apply their proposed results in a robot pick and place scenario. In [18], a hybrid approach was proposed to address a 6D pose estimation problem. In particular, a high-precision pose estimation method of 3C parts for robotic grasping and assembly was proposed. The algorithm proposed consists of object segmentation followed by a PCA-ICP algorithm for pose estimation. In [19], a hybrid 6D object pose estimation method was proposed where object detection for scene point cloud segmentation was performed first, followed by a template-based method for the 6D pose estimation.

Robotic assembly: A 6D pose estimation method for the robotic assembly of shell parts was proposed in [20] which includes two phases, an initial pose estimation phase which utilizes the principle component analysis (PCA) algorithm and a local pose estimation phase which includes voxel sampling and a weighted point-to-plane ICP algorithm. In [18], a high-precision pose estimation method of 3C parts for robotic grasping and assembly is proposed.

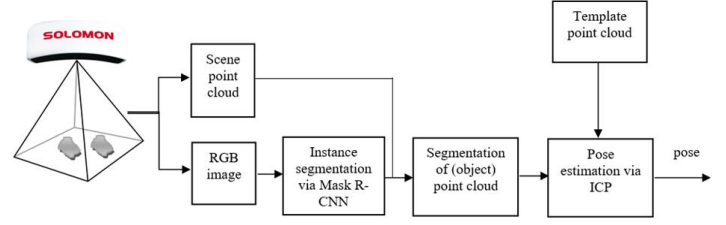


Fig. 3. Pipeline for the 6D pose estimation method

The algorithm proposed consists of object segmentation followed by a PCA-ICP algorithm for pose estimation. A deep learning based approach and a domain randomization approach were proposed in [6] for the 6D pose estimation problem for assembly scenarios using RGB-D images. The ICP algorithm was then used as a refinement step to improve the pose estimation accuracy. The pose estimation problem for high precision robotic assembly task using simulated depth images was addressed in [8]. The proposed algorithm for 6D pose estimation is an end-to-end deep convolution neural network based approach.

II. METHODS

Fig. 3 shows the whole pipeline of the proposed method in this work. The 6D pose estimation method uses both 2D images and 3D point cloud data. A structured light 3D scanner (details of the scanner are discussed in the next section) is used to acquire the 2D images and 3D point cloud data. In the first step, object detection is performed using an instance segmentation network. The target objects are then segmented in the point cloud. In the second step, the 6D pose of the target object is estimated via iterative closet point algorithm.

A. Instance segmentation and point cloud segmentation

In order to estimate the pose of the target object, the first step is to locate the object in the 2D image. In this work, we use instance segmentation via deep learning methods to locate the object in the 2D image where instance segmentation is defined as detecting and locating instance of objects of a certain class in an input image down to the pixel-level [21]. There are a variety of popular networks that can perform object segmentation however, in this work we chose Mask R-CNN [22] since it has achieved state-of-the-art performance on instance-level segmentation [23]. For a Mask R-CNN model, the input is a 2D image and the output is a predicted object mask, a 2D bounding box, an object label class and a confidence score. After performing instance segmentation, the 2D bounding box coordinates are used to segment the target object in the point cloud.

B. Pose estimation

The next step after segmenting the objects in the point cloud is to estimate their 6D poses. The pose estimation algorithm we chose in this work is the Iterative Closest Point algorithm which was first introduced in [24]. The ICP method is the most widely used method for point cloud registration [18], [20]. It is a local point cloud registration method. The ICP algorithm typically finds the rotation

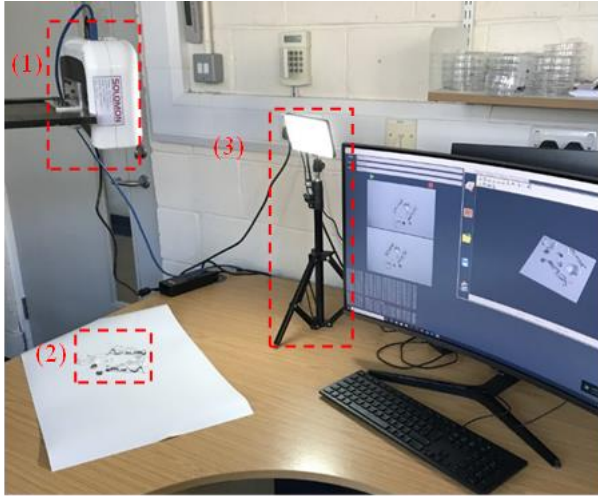


Fig. 4. Experimental setup; (1) 3D structured light scanner, (2) objects, (3) LED panel light

matrix R and translation matrix T that minimizes the square error between the scene model and the CAD model as described in [25] and is re-visited here as follow.

Given a scene model p and a CAD model q , find a transformation T that P can register in q such that the error defined by

$$\sum \|Rp + T - q\|^2 \quad (1)$$

is minimized. Here, we use the ICP algorithm to match the CAD model of the object to the segmented object in the point cloud.

C. Hardware devices

A 3D structured light scanner (SLM 3DSCN-0230) from SOLOMON Technology Corporation is used. This device can accurately generate both 2D images and point cloud data. It has a camera resolution of 1920×1200 and a field of view of $205 \times 128 \sim 914 \times 571$ mm corresponding to the working distance of $450 \sim 2000$ mm. In our setup we mount our scanner at a distance of 860 mm from the table as shown in Fig. 4. The objects are placed on white photography backdrop. A small LED panel light with white, yellow, red, and blue color filters is also used to vary the lighting conditions. The experiments are performed on a PC with NVIDIA GeForce GTX 1660 Ti with 6GB memory and an 11th Gen Intel Core i7 11700F (2.5GHz) processor. Fig. 4 shows the complete setup.

D. Objects to be assembled

The objects to be assembled are provided by Chitendai Ltd which form an assembly unit of concentrator photovoltaics solar panel. Four objects are provided which are: primary lens, secondary lens, solar-cell, and a tripod leg. Fig. 1 shows each of the objects. From Fig. 1 it can be seen that the characteristics of the objects we are dealing with make our pose estimation problem very challenging. The objects we are considering are reflective, refractive/translucent (semi-

transparent objects), symmetric and very small objects in the range of 18mm to 100 mm. It is well known in the literature that such characteristics make the pose estimation problem very challenging. In particular, semi-transparent objects exhibit refractive and reflective nature. As in [9], [15], [26]–[29], the reflective and refractive nature of the objects makes the pose estimation problem very difficult because such objects do not have their own color making their appearance vary according to the background and illumination conditions [26]. In this work we focus on the pose estimation of the solar cell only, as it may be considered the least complex among the four class objects and the remaining objects will be considered in future work taking into account the characteristics of such objects.

III. EXPERIMENTAL RESULTS

A. Instance segmentation and object segmentation in point cloud

Under different lighting conditions, 35 images and corresponding point cloud containing the solar-cell object were collected. The minimum and maximum numbers of instances of the solar-cell object in an image and its corresponding point cloud is 1 and 6 instances respectively. Computer Vision Annotation Tool (CVAT) was used to annotate the objects in the 2D image. Fig. 5 shows a sample of the collected 2D images with their corresponding ground truth annotations. As previously mentioned, we finetune a Mask R-CNN on our dataset. We use PyTorch [30] to train the network. We have one class object (solar-cell) other than the background. The dataset was split as 25 images for training and 9 images for testing. The network was trained with stochastic gradient descent (SGD) optimizer with learning rate of 0.005, momentum of 0.9, weight decay of 0.0005 and the number of epochs was chosen to be 60. The trained model is then utilized. Fig. 6 shows examples of the predicted output (object class, bounding box coordinates, masks and confidence score).

The bounding box coordinates are then used to segment the object from the scene point cloud. The bounding box coordinates are mapped from the 2D intrinsic coordinates to the 2D world coordinate which are then used to find the points of the object of interest in the scene point cloud. Fig. 7-9 show an example of the segmentation process.

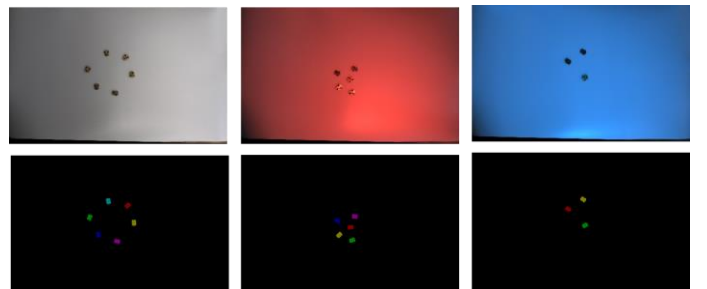


Fig. 5. Sample of collected 2D images and corresponding ground truth annotations

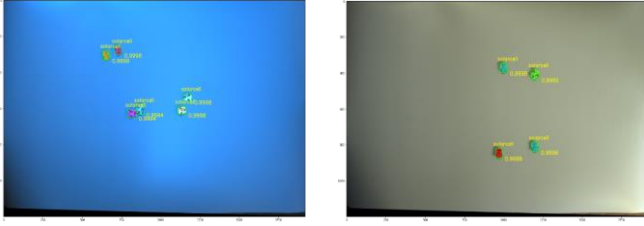


Fig. 6. Examples of the predicted output (object class, bounding box coordinates, masks, and confidence score)

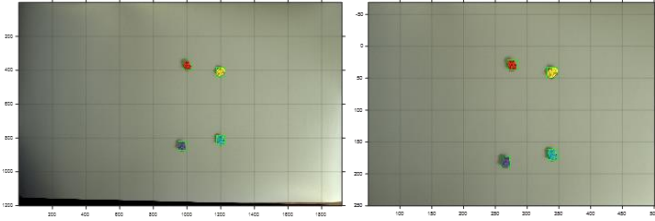


Fig. 7. Mapping from intrinsic to world coordinates, left image: 2D intrinsic coordinates, right image: 2D world coordinate

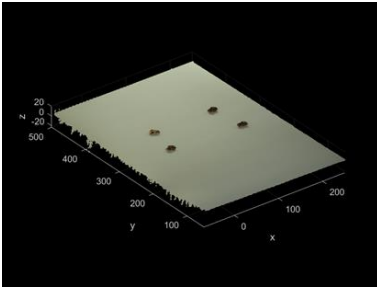


Fig.8. 3D point cloud



Fig. 9. Objects segmented in the point cloud

B. Pose estimation

As mentioned in section II-B, the pose estimation algorithm to be implemented is the ICP algorithm which requires a CAD model of the object. The algorithm should align the CAD model of the object with the segmented scene point cloud. In this work, a CAD model (in the form of a .step/.iges file) of each of the objects and a CAD model of the assembly are available from Chitendai Ltd which were all converted to .dwg files. The CAD models are then converted to point cloud

data as described hereafter. First these CAD files are converted into .stl files, which is a triangular representation of a 3-dimensional surface geometry, using AutoCAD program. Using open3d in python, the triangle mesh is read from each file and the surface normal is computed. Fig. 10 shows the triangle mesh with surface normals of each of the objects and the assembly unit. The triangle mesh is then uniformly sampled to generate the point cloud. The number of points that are sampled from the triangle surface is chosen as 38046. Fig. 11 shows the point cloud which has been uniformly sampled from the triangle mesh.

Note that it is important that the CAD object model and the scene object point cloud have the same scale. If they do not have the same scale then as described in [31] it is required to scale the CAD model to the right scale so that it is equivalent to the object size. This can be done as described in [31] by

$$P_N(V) = \mu P(V), \quad \mu = \frac{S}{M} \quad (2)$$

where $P(V)$ represents the CAD model, $V = v_1, v_2, \dots, v_n$, and n is the number of points, S is the scale of the object model, M is the scale of the CAD model, and $P_N(V)$ represents the new CAD model equivalent to the object size.

The outcome of the pose estimation algorithm should be a rigid transformation which is represented by a 4×4 homogenous transformation matrix $P = [R, t; 0, 1]$ where R is a 3×3 rotation matrix and t is a 3×1 translation vector such that by applying the transformation matrix P , the model

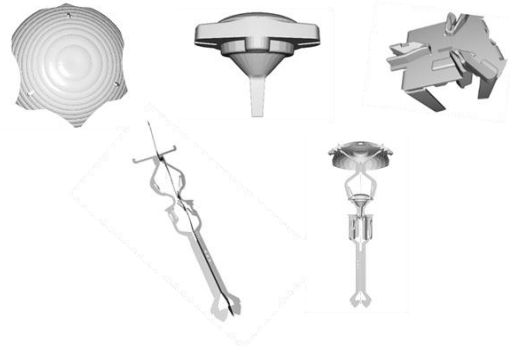


Fig. 10. CAD models of the objects and the assembly unit

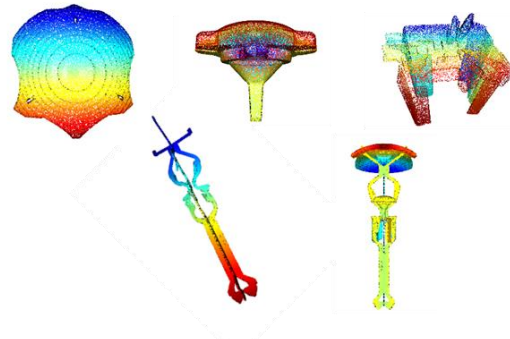


Fig. 11. Point cloud of objects and assembly unit which have been uniformly sampled from the mesh

of the object (3D points in the model coordinate system) is transformed and matched to the segmented object in the scene point cloud.

In this experiment, the point cloud CAD model is rotated by angles 45, 60, 90, 150 and 330 degrees about the z axis. The point-to-point error metric is used and the maximum iteration number of the ICP algorithm is set to 100 if no solution converges earlier. The ICP algorithm is performed with one segmented object point cloud. The root mean square error (rmse), which represents the Euclidean distance between the inlier aligned points, is computed where the registration with minimum error is selected. Fig. 12 shows the result of the ICP algorithm with the minimum rmse value which is 0.97.

IV. DISCUSSION AND FUTURE WORK

In this paper, we addressed the 6D pose estimation problem for precision assembly via a hybrid approach. A structured light 3D scanner was used to acquire accurate and high-quality 3D point cloud data in addition to 2D images of a solar-cell object which is one of four objects of an assembly unit to be assembled. A Mask R-CNN network was trained on our custom dataset and the model was then utilized to predict the 2D bounding boxes, the object mask, the object class label and a confidence score. The bounding box coordinates were used to segment the objects in point cloud and then the ICP algorithm was used to estimate the objects pose by matching the CAD model to the segmented object in point cloud.

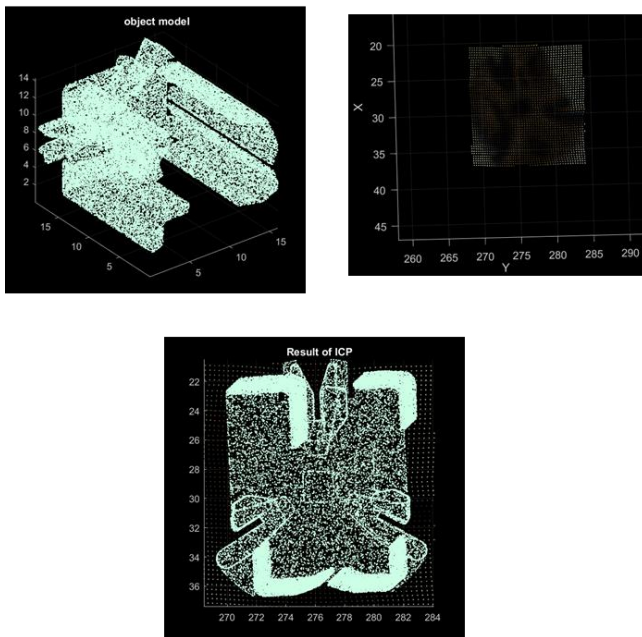


Fig. 12. Point cloud Result of ICP algorithm (bottom figure) where the object model (top left figure) is aligned to the segmented object (top right figure)

Although we benefited from the high-quality point cloud data obtained by the structured light 3D scanner used in this work, the accuracy of the proposed pose estimation method does not yet meet the accuracy required for the assembly unit of the concentrator photovoltaics solar panel under consideration. Therefore, some improvements need to be made to our pose estimation method in order to achieve the required accuracy. Below are some suggestions for improvement: (1) In this work, we used the bounding boxes coordinates for point cloud segmentation. However, it would be better and more accurate to map the masks of the objects obtained from the Mask R-CNN on to the point cloud for point cloud segmentation, (2) In this work, the original ICP algorithm was used as the pose estimation algorithm. It is known that the original ICP algorithm is affected by the initial positions and that it works well when the initial positions are close enough otherwise it may be possible that the estimation results are trapped in to a local minimum resulting incorrect convergence [20], [31]. Many variants and improvements of the original ICP algorithm have been proposed in the literature over the years [32], [33]. Moreover, for achieving better accuracy, global registration algorithms, which can estimate the rigid transformation parameters of two-point clouds without the need for an initial pose, have been used prior to local registration methods to provide them with an initial pose estimate [18], [20], [34]. Accordingly, a global registration method such as PCA could be first implemented followed by an improved ICP algorithm to improve our pose estimation algorithm. Furthermore, different global registration methods other than PCA could be implemented followed by an improved ICP algorithm and a comparison between these methods could be made to see which method provides the most accurate results.

REFERENCES

- [1] Y. Ma, X. Liu, J. Zhang, D. Xu, D. Zhang, and W. Wu, "Robotic grasping and alignment for small size components assembly based on visual servoing," *Int. J. Adv. Manuf. Technol.*, vol. 106, no. 11–12, pp. 4827–4843, 2020.
- [2] M. Rudorfer, "Towards Robust Object Detection and Pose Estimation as a Service for Manufacturing Industries (PhD thesis)," Technische Universität Berlin, 2021.
- [3] T.-T. Le, T.-S. Le, Y.-R. Chen, J. Vidal, and C.-Y. Lin, "6D pose estimation with combined deep learning and 3D vision techniques for a fast and accurate object grasping," *Rob. Auton. Syst.*, vol. 141, 2021.
- [4] G. Liang, F. Chen, Y. Liang, Y. Feng, C. Wang, and X. Wu, "A Manufacturing-Oriented Intelligent Vision System Based on Deep Neural Network for Object Recognition and 6D Pose Estimation," *Front. Neurobot.*, vol. 14, pp. 1–15, 2021.
- [5] S. H. Zabihifar, "Unreal mask : one-shot multi-object class-based pose estimation for robotic manipulation using keypoints with a synthetic dataset," *Neural Comput. Appl.*, vol. 6, 2021.
- [6] S. Stevsic, S. Christen, and O. Hilliges, "Learning to assemble: Estimating 6d poses for robotic object-object manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1159–1166, 2020.
- [7] N. Hajari, G. L. Bustillo, H. Sharma, and I. Cheng, "Marker-less 3d object recognition and 6d pose estimation for homogeneous textureless objects: An rgb-d approach," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–22, 2020.
- [8] Y. Litvak, A. Biess, and A. Bar-Hillel, "Learning Pose Estimation for High-Precision Robotic Assembly Using Simulated Depth Images," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 3521–3527.

- [9] Z. He, W. Feng, X. Zhao, and Y. Lv, "6D pose estimation of objects: Recent technologies and challenges," *Appl. Sci.*, vol. 11, no. 1, pp. 1–18, 2020.
- [10] C. Sahin, G. Garcia-Hernando, J. Sock, and T. K. Kim, "A review on object pose recovery: From 3D bounding box detectors to full 6D pose estimators," *Image Vis. Comput.*, vol. 96, 2020.
- [11] J. Chen, L. Zhang, Y. Liu, and C. Xu, "Survey on 6D Pose Estimation of Rigid Object," in *Proceedings of the 39th Chinese Control Conference*, 2020, pp. 7440–7445.
- [12] T. Hodan, J. Matas, and S. Obdrzalek, "On Evaluation of 6D Object Pose Estimation," in *Computer Vision -- ECCV 2016 Workshops*, 2016, pp. 606–619.
- [13] T. Hodaň et al., "BOP Challenge 2020 on 6D Object Localization," in *ECCV Workshops*, 2020.
- [14] S. Peng, Y. Liu, Q. Huang, H. Bao, and X. Zhou, "PVNet: Pixel-wise voting network for 6DoF pose estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*
- [15] C. Xu, J. Chen, M. Yao, J. Zhou, L. Zhang, and Y. Liu, "6DoF pose estimation of transparent object from a single RGB-D image," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–19, 2020.
- [16] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 77–85.
- [17] R. König and B. Drost, "A Hybrid Approach for 6DoF Pose Estimation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12536 LNCS, pp. 700–706, 2020.
- [18] N. Zhang, Y. Xie, X. Yang, H. Hu, and Y. Lou, "High-precision pose estimation method of the 3C parts by combining 2D and 3D vision for robotic grasping in assembly applications," *2021 IEEE Int. Conf. Real-Time Comput. Robot. RCAR 2021*, pp. 548–553, 2021.
- [19] Y. Zhang, C. Zhang, M. Rosenberger, and G. Notni, "6D object pose estimation algorithm using preprocessing of segmentation and keypoint extraction," *I2MTC 2020 - Int. Instrum. Meas. Technol. Conf. Proc.*, pp. 1–6, 2020.
- [20] H. Hu, W. Gu, X. Yang, N. Zhang, and Y. Lou, "Fast 6D object pose estimation of shell parts for robotic assembly," *Int. J. Adv. Manuf. Technol.*, vol. 118, no. 5–6, pp. 1383–1396, 2022.
- [21] G. Du, K. Wang, S. Lian, and K. Zhao, *Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review*, vol. 54, no. 3. Springer Netherlands, 2021.
- [22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- [23] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Syst. Appl.*, vol. 172, no. October 2020, p. 114602, 2021.
- [24] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, 1992.
- [25] H.-Y. Lin, C.-C. Chang, and S.-C. Liang, "3D Pose estimation using genetic-based iterative closest point algorithm," *Int. J. Innov. Comput. Inf. Control*, vol. 14, no. 2, pp. 537–547, 2018.
- [26] S. Sajjan et al., "Clear Grasp: 3D Shape Estimation of Transparent Objects for Manipulation," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 3634–3642, 2020.
- [27] Y. Qian, M. Gong, and Y. H. Yang, "3D Reconstruction of Transparent Objects with Position-Normal Consistency," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4369–4377, 2016.
- [28] C. J. Phillips, K. G. Derpanis, and K. Daniilidis, "A novel stereoscopic cue for figure-ground segregation of semi-transparent objects," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, pp. 1100–1107, 2011.
- [29] J. Yang, Y. Gao, D. Li, and S. L. Waslander, "ROBI: A Multi-View Dataset for Reflective Objects in Robotic Bin-Picking," 2021.
- [30] "TorchVision Object Detection Finetuning Tutorial." [Online]. Available: https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html. [Accessed: 11-Jan-2022].
- [31] H. Y. Lin, S. C. Liang, and Y. K. Chen, "Robotic grasping with multi-view image acquisition and model-based pose estimation," *IEEE Sens. J.*, vol. 21, no. 10, pp. 11870–11878, 2021.
- [32] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proceedings of International Conference on 3-D Digital Imaging and Modeling, 3DIM*, 2001, pp. 145–152.
- [33] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets: Open-source library and experimental protocol," *Auton. Robots*, vol. 34, no. 3, pp. 133–148, 2013.
- [34] D. Dong, X. Yang, H. Hu, and Y. Lou, "Pose estimation of components in 3C products based on point cloud registration," in *IEEE International Conference on Robotics and Biomimetics, ROBIO 2019*, 2019, pp. 340–345.