

Pose Estimation with Mismatching Region Detection in Robot Bin Picking

Zhe Wang, Lei Jia, Lei Zhang, and Chungang Zhuang^(✉)

State Key Laboratory of Mechanical System and Vibration,
School of Mechanical Engineering, Shanghai Jiao Tong University,
Shanghai 200240, China
cgzhuang@sjtu.edu.cn

Abstract. 3D object detection and pose estimation based on 3D sensor have been widely studied for its applications in robotics. In this paper, we propose a new clustering strategy in Point Pair Feature (PPF) based 3D object detection and pose estimation framework to further improve the pose hypothesis result. Our main contribution is using Density Based Spatial Clustering of Applications with Noise (DBSCAN) and Principle Component Analysis (PCA) in PPF method. It was recently shown that point pair feature combined with a voting framework was able to obtain a fast and robust pose estimation result in heavily cluttered scenes with occlusions. However, this method may fail in the mismatching region caused by false features or features with insufficient information. Our experimental results show that the proposed method can detect mismatching region and false pose hypotheses in PPF method, which improves the performance in robot bin picking application.

Keywords: Pose estimation · Mismatching region detection · Point pair feature (PPF) · Robot bin picking

1 Introduction

In the last two decades, industrial robots have been widely used in various fields of manufacture to enable efficient automated production processes. Vision-guided industrial robots are one of key parts in the state-of-the-art manufacturing processes. One of the main tasks for industrial robot is picking object and placing it to the specific location. This task has a good solution in 2D placement case while in 3D scattered stacking case, as illustrated in Fig. 1, remains a challenge. Indeed, 3D object detection and pose estimation are significant for robot bin picking in 3D case. The goal is to locate some pickable object instances and determine their 6D poses, i.e. rotation matrix and translation matrix.

Various algorithms have been proposed for 3D object detection and pose estimation based on 2D images, range images or 3D point clouds. However, few of them are suited for the real-world industrial robot bin picking application for 3D scattered stacking case. There are some reasons: (a) the industrial objects are

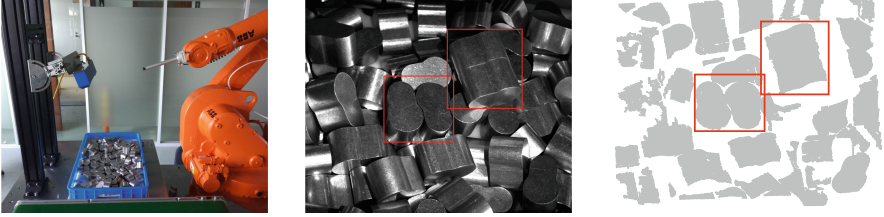


Fig. 1. Left: robot bin picking system. Middle: objects are stacked randomly in a bin. Right: scene point clouds from 3D sensor. The mismatching regions are marked with the red bounding box. (Color figure online)

generally texture-less objects with simple shape. Thus, traditional 2D/3D key point based object detection and pose estimation are not appropriate. (b) 3D scattered stacking objects form a heavy cluttered environment with occlusions. (c) 3D sensor may not obtain complete scene information because of reflection, transmission and absorption. (d) the industrial bin picking requires eliminating the wrong pose hypothesis. Therefore, no satisfactory general solution to this task exists yet.

In literatures, while several methods have been proposed to tackle with this problem based on different input data, we focus here on the approach from [1], which relies on point cloud data. This algorithm has been improved and extended by many authors [2–7]. Also, it has been used in various robot bin picking applications [5, 8]. Recent studies [4] have shown its potential to deal with sensor noise, background clutters and occlusions, which is the major disadvantage of another popular pose estimation algorithm, LINEMOD [9, 10]. Moreover, smarter sampling strategy [2–4] and GPU acceleration [6] enable PPF method to obtain a fast and robust pose estimation result in heavily cluttered scenes with occlusions. However, this method may fail in the mismatching region caused by false features or features with insufficient information. As shown in Fig. 1, the mismatching regions are marked with the red bounding boxes where the wrong pose may be evaluated a high score. Thus, robot may try to pick these high score wrong detected objects, which causes a decline in system robustness. Actually, even human may not able to identify the correct poses from these regions in point clouds. To tackle with this problem, we propose a new clustering strategy in Point Pair Feature (PPF) based 3D object detection and pose estimation framework to further improve the pose hypothesis result. Our main contribution is using Density Based Spatial Clustering of Applications with Noise (DBSCAN) [11, 12] and Principle Component Analysis (PCA) in PPF method. Through further studying on the clustering step, we give a probable pose distribution perspective for mismatching region detection.

The remainder of this paper is organized as follows. Section 2 gives an extensive overview of various cutting-edge 3D object detection and pose estimation methods as well as their applications for robot bin picking. Our method for

improved PPF method and mismatching region detection is demonstrated in Sect. 3. Experimental results are presented and discussed in Sect. 4 and the conclusion follows in Sect. 5.

2 Related Work

An extensive overview of the state-of-the-art approaches in 3D object detection and pose estimation is provided. This section also introduces their application in robot bin picking.

2.1 3D Object Detection and Pose Estimation

Many algorithms have been proposed for 3D object detection and pose estimation. We only focus on the recent work here and split them in several categories.

Point Wise Correspondence Based Methods. Iterative Closest Point (ICP) [13] is a classic point wise correspondence based method for point cloud registration, which can also be used for pose estimation. However, it requires a good initialization to avoid being trapped in a local minimum during optimization. Go-ICP [14] provides a global optimal registration by using global branch-and-bound (BnB) optimization in ICP. Since ICP enables a high accuracy point cloud registration, many algorithms combine ICP into a coarse-to-fine pose estimation framework.

Template Based Methods. Template based methods, which are widely used in 2D object detection, are extended in 3D case in recent years [9, 10]. The templates store the different viewpoints of object appearances. An object is detected when a template matches the image and gives its 3D pose. LINEMOD, presented in [9, 10], renders 3D object model to generate a large number of templates covering full view hemisphere. It combines 2D image gradients and 3D surface normals to build multiple modalities. Though such template based approaches can work accurately and quickly in practice, they still show robust problem in cluttered environment with occlusions.

Local Feature Based Methods. Local feature based methods, including point pair feature (PPF) [1], point feature histograms (PFH) [15] and 3D key point that are extended from traditional 2D key point like 3D-SURF [16], are robust to occlusions. A review can be found in [17]. These features are usually combined with other techniques like voting for object detection. However, they are usually computationally expensive and have difficulty in scenes with heavy clutters. In this paper, we focus one of the promising methods, PPF based method [1]. Several improvements and extensions of PPF based method have been proposed in recent years. [2–4] improve PPF method [1] in almost every stage by smarter subsampling strategy, smarter voting strategy and smarter pose evaluation strategy. [4] shows its potential to tackle with scene noise, cluttered environment with occlusions and loss of information caused by subsampling compared with original PPF method [1] and other state-of-the-art algorithms [3, 9]. [5] discusses

different point pair features for voting. [6] uses color point pair feature and gives its GPU implementation. RANSAC integrated PPF method is proposed in [7].

Learning Based Methods. With the rapid development of learning techniques, several learning based methods are proposed for this task. [18,19] use Iterative Hough Forest on local patches to detect and estimate 3D poses. [20–22] extend traditional convolutional neural network to 3D case for object detection and pose estimation. These recent studies show the potential to use the state-of-the-art learning techniques on this task.

2.2 3D Object Pose Estimation in Robot Bin Picking

Since most industrial objects are texture-less, LINEMOD [9,10] and PPF method [1,5,8] are popular in robot bin picking application. The first Amazon Picking Challenge [23,24] presents some state-of-the-art robot bin picking frameworks combined with various methods and tricks. SegICP, presented in [25], shows the potential to combine CNN-based object detection with ICP on this task.

3 Proposed Method

We describe here our contributions to improve the PPF method [1] more robust for robot bin picking. This paper focuses on PPF methods in pose clustering stage. Compared with traditional PPF method and its improvements, we give a probable pose distribution perspective to rethink about the PPF method and use it to detect mismatching region in the scene. This section introduces our improved PPF method first and then describes our clustering strategy in detail.

3.1 Improved PPF Method

We revisit PPF based object detection and pose estimation pipeline and describe our improved PPF method here. The improved PPF method can be divided into five stages.

Preprocessing the Model and Scene Data. Generally, the model data is CAD format and scene data is a point cloud without normals. In order to use PPF method, remeshing techniques are used to transform model and scene data to uniform dense point clouds with normals. We also use subsampling to accelerate the PPF method.

Model Training. Point pair feature is used to create a hash table for voting framework in this step. Every point pair on the model surface calculates the feature $\mathbf{F}_m(\mathbf{m}_i, \mathbf{m}_j)$ and the corresponding angle α_{ij}^{model} . The distance and angle in PPF are discretized in step d_{dist} and $d_{angle} = 2\pi/n_{angle}$. The indices of hash table is hashed PPF and the stored element is $(\mathbf{m}_i, \mathbf{m}_j, \alpha_{ij}^{model})$.

Voting. Voting framework is similar to the Generalized Hough Transform. Every local coordinates (\mathbf{m}_r, α) is presented in a two-dimensional accumulator array. α is discretized using d_{angle} . Here we subsample scene points as reference points. For each reference point s_r , the point pair feature $\mathbf{F}_s(s_r, s_j)$ and the corresponding angle α_{rj}^{scene} is calculated. Then corresponding $\mathbf{F}_m(\mathbf{m}_r, \mathbf{m}_j)$ is found and $\alpha = \alpha_{rj}^{model} - \alpha_{rj}^{scene}$. The accumulator position (\mathbf{m}_r, α) vote one. Compared to the original version, we compute vote value as

$$vote = \begin{cases} 1 - \lambda|n_1 \cdot n_2|, & 1 - \lambda|n_1 \cdot n_2| \geq v_{min} \\ v_{min}, & 1 - \lambda|n_1 \cdot n_2| < v_{min} \end{cases} \quad (1)$$

where $|n_1 \cdot n_2|$ denotes the absolute value of the dot product of two normals, λ is a weighting parameter in default set to 1 and v_{min} is the minimum boundary of the vote in default set to 0.1. The weighted vote will weaken the planar vote at some level. In addition, if the distance between two points is larger than the size of the model, it's unnecessary to process them because only point pairs belong to the same object shall be vote. After all points s_i are processed, the peak in the accumulator array are the corresponding local coordinate to s_r . Then, we obtain a set of possible object poses, which we called probable pose distribution in this paper.

Pose Clustering. In the original version of PPF method, the retrieved poses are sorted in descending order of the number of votes to make sure group most likely pose hypotheses first. Then it uses euclidean distance to cluster the proximal poses and the score of the cluster is the sum of the scores of the contained poses. Finally, it removes the isolated poses with low score and averages the clustered pose hypotheses to get the final pose hypotheses. In this paper, we use new clustering strategy combined with DBSCAN and PCA to further improve the pose hypotheses result. The following part will describe it in detail.

Post Processing the Pose Hypotheses. After the high score pose hypotheses are obtained, post processing is suggested to obtain higher accuracy of the detected poses. The pose hypothesis scores will be reevaluated to reject the wrong pose hypotheses and ICP algorithm will be used to get accuracy poses. One of the popular evaluation strategies is computing how many points on the model are detected on the scene point cloud. However, false pose hypotheses can get high evaluation scores in mismatching region, which will give wrong pose result to robot. Hence, we evaluate pose hypotheses in clustering stage first to detect the pose hypotheses in mismatching region.

3.2 Mismatching Region

Given a scene point cloud and a model point cloud, we define the mismatching region as region where wrong pose hypothesis can be evaluated a high score. For example, traditional evaluation strategies in the post processing stage, such as re-rendering the depth image or computing area of the model surface detected

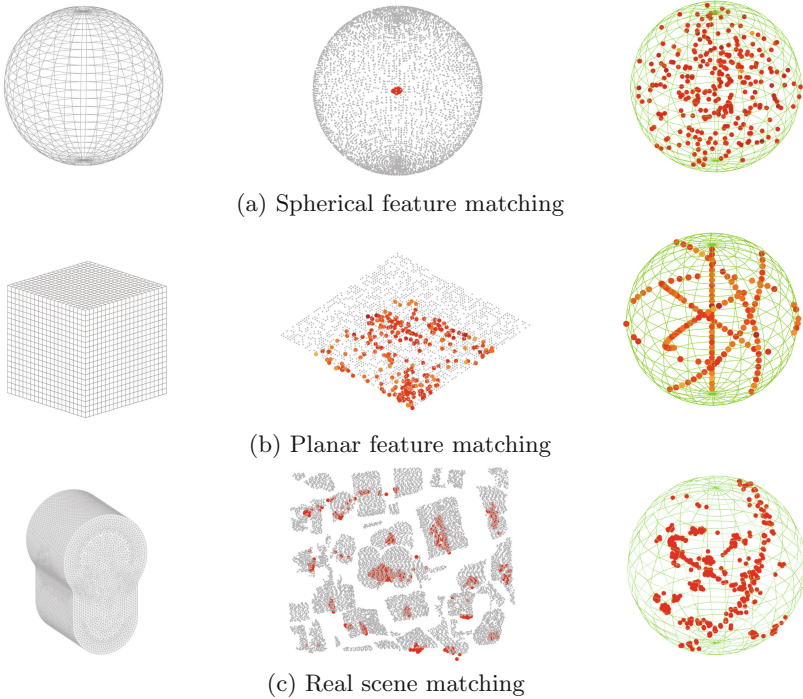


Fig. 2. The probable pose distribution visualization on the mismatching region. Left: model point cloud. Middle: scene point cloud with pose positions visualization. Right: pose orientations visualization in normalized axis-angle space (Color figure online)

in the scene, will apparently evaluate a wrong pose hypothesis a high score in the mismatching region shown in Fig. 1. The mismatching region is caused by false features or features with insufficient information. This false features may appear when the scene features are similar to the model features, multiple model features couple or scene features couple with model features.

To vividly describe the cause of the mismatching region, we use PPF method in the scene with mismatching region and visualize the probable pose distribution, which is a set of high vote pose hypotheses after voting stage. The result is illustrated in Fig. 2. In the spherical feature matching, the pose positions gather at one point while the orientations randomly distributed in the normalized axis-angle space because the overlapping surface on the model can slide on the scene with small overlapping area change. Also, for the cube with six planar features, the pose positions can slide in a wide range on the planar scene while the orientations are constraint to be six spatial curves. And for real scene matching with the object in our robot bin picking system, it's obvious to see that the mismatching region contains a large pose position sliding area. Therefore, we know the mismatching region is the region where the overlapping surface on the model can slide on the scene with small overlapping area change. And we can

detect the mismatching region by detecting the pose hypotheses sliding area in the probable pose distribution. In fact, the mismatching region highly depend on the shape of the detected objects and the way of stacking the detected objects. If the shape feature is simple, such as cube and cylinder, and the sensor can only scan the partial shape, the features cannot denote a specific pose because the pose hypotheses lack enough reliable constraints, which show up as false features or features with insufficient information. For example, in our robot bin picking system, the object, shown in Fig. 2, will cause the mismatching region problem.

To detect the pose in the mismatching region, we propose a new clustering strategy on the probable pose distribution. Since we only focus on the robot bin picking application, we believe that pose position is more significant than orientation, which may vary with symmetric structure but does not affect the bin picking. We use DBSCAN to cluster the probable pose by position distribution and use PCA to find the size of the clustered pose set. Then we use a threshold to detect the mismatching region by pose size.

3.3 DBSCAN: Density-Based Clustering

With the appropriate setting on the sampling step, here we list several assumptions for PPF method: (a) Every reference point on the object will return a high vote true pose hypothesis. (b) High visibility object sampled many reference points will gather plenty of high vote pose hypotheses on its true pose, as illustrated in the Fig. 2. (c) Scene feature dissimilar to the object will return a low vote wrong pose hypothesis. (d) High vote pose hypotheses, i.e. probable pose distribution, contain all visible object poses in the scene. (e) Mismatching region, where overlapping can slide, is a dense pose hypotheses cluster of arbitrary shape. Original PPF method uses greedy strategy in the pose clustering stage. It uses highest score pose hypotheses as a cluster center and cluster its nearest pose hypotheses. By repeating this loop, we get a descending order clustered pose hypotheses, which will be used to post processing. The cluster shape is similar to K-means, which limits the use of probable pose distribution.

Hence, we handle probable pose distribution using Density Based Spatial Clustering of Applications with Noise (DBSCAN) [11], which can discover clusters of arbitrary shape and also handle outliers effectively. DBSCAN algorithm defines cluster as a densely connected points by *eps* and *minpts*. *eps* is the maximum distance neighborhood for given point and *minpts* is the minimum number of points required in the *eps-neighborhood* of a point to form a cluster. This paper implements the DBSCAN algorithm in [11, 12] for clustering probable pose position. As we state before, only the position of the pose is used for clustering. The basic pipeline of DBSCAN is shown in Algorithm 1.

Figure 3 shows the result of clustered pose in the position space. Scene data is shown in Fig. 1. The high density clusters are detected and marked with different color while the isolated or low density points are detected as noise. Apparently, a clustered pose hypothesis with small size present a specific true pose.

Algorithm 1. DBSCAN($P, eps, minpts$)

```

Input: probable pose set  $P$ ,  $eps$ ,  $minpts$ 
Output: pose clusters  $C$ , pose cluster id  $I$ 
1: mark all poses in  $P$  as unvisited,  $C \leftarrow \emptyset$ ,  $I \leftarrow \emptyset$ 
2: for each unvisited pose  $p$  in  $P$  do
3:   mark  $p$  as visited
4:   Neighbors  $\leftarrow$  FindNeighbors( $p, eps$ )
5:   if  $|Neighbors| \leq minpts$  then
6:     mark  $p$  as noise,  $I(p) \leftarrow$  noise
7:   else
8:      $c \leftarrow$  next cluster
9:     add  $p$  to cluster  $c$ ,  $I(p) \leftarrow c$ 
10:    for each pose  $p'$  in Neighbors do
11:      if  $p'$  is unvisited then
12:        mark  $p'$  as visited
13:        pNeighbors  $\leftarrow$  FindNeighbors( $p', eps$ )
14:        if  $|pNeighbors| \geq minpts$  then
15:          Neighbors  $\leftarrow$  Neighbors joined with pNeighbors
16:        end if
17:        if  $p'$  is not yet member of any cluster then
18:          add  $p'$  to cluster  $c$ ,  $I(p') \leftarrow c$ 
19:        end if
20:      end if
21:    end for
22:    add  $c$  to  $C$ 
23:  end if
24: end for

```

3.4 Mismatching Region Detection

After we obtain clustered pose hypotheses by DBSCAN algorithm, we use Principle Component Analysis (PCA) to extract cluster features. PCA is a popular machine learning and statistics technique to extract the main directions of the dataset. In our approach, we first compute the center and main directions of the cluster, and then project the points in the cluster on the main directions to get the cluster size, which presents the overlapping sliding area. The sizes of all clusters are illustrated with 3D oriented bounding box in Fig. 3. Also, the mismatching region is detected by the sum of box size and marked with red color.

To obtain final pose hypotheses result, we use original greedy strategy in further clustering. The highest vote pose hypothesis in the cluster will be used to cluster its neighborhood and the final vote will be the sum of all neighbor votes. This cluster result will present the final pose hypothesis of the cluster. Then, we apply this method to every cluster as well as the mismatching region. Each cluster outputs only one good pose hypothesis with vote and the pose hypothesis from mismatching region will be marked as unreliable pose hypothesis. Finally, our approach only return high vote reliable pose hypotheses for the post processing stage.

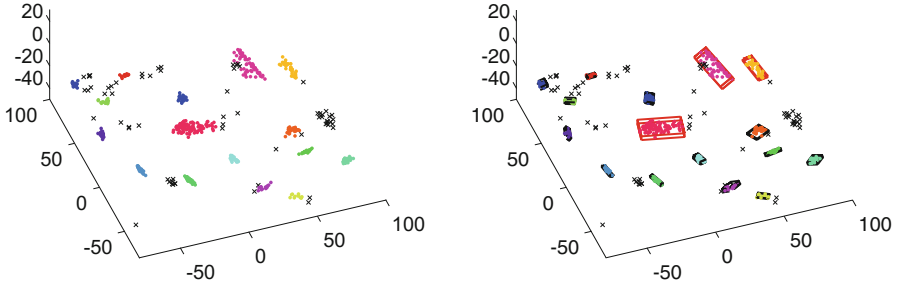


Fig. 3. Left: visualization of DBSCAN result. Right: visualization of PCA result. The oriented bounding boxes are drawn and the mismatching region are shown in red. The noise poses are marked as cross. (Color figure online)

4 Experimental Results

We evaluate our approach on the real robot bin picking application with mismatching region detection. Given a scene point cloud data, our algorithm not only detects the objects and estimates their poses, but also marks the poses in the mismatching region, as shown in Fig. 4. Multiple high score pose hypotheses are detected and wrong pose hypotheses with high evaluation score are marked with red color. Compared with the state-of-the-art PPF method, our approach improves the pose hypotheses result and enables robust 3D object detection and pose estimation in robot bin picking application. As we state in the Sect. 3, the mismatching region problem is caused by the shape of objects, the way of stacking objects and the capability of the sensor. If the mismatching region doesn't exist, our method works like the original version because DBSCAN works like original clustering. According to our analysis in Sect. 3, this method shall work in other object detection and pose estimation task.

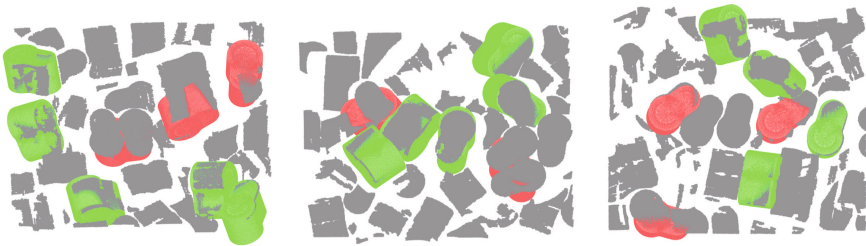


Fig. 4. 3D object pose estimation with mismatching region detection on the real scene data for robot bin picking. The objects in mismatching region are marked with red color. (Color figure online)

The algorithm takes several parameters in DBSCAN and filter stage, which are related to sampling step and reference point selection strategy in PPF method. Also, since the clustering algorithm complexity stay same and the size of probable pose hypotheses is generally small, our approach dose not increase the much runtime compared with original version. However, this problem remains a challenge if the overlapping can not slide in some cases.

5 Conclusion

In this paper, we revisit the state-of-the-art PPF method and propose an improved PPF method with mismatching region detection for 3D object detection and pose estimation. We demonstrate the mismatching problem when using traditional PPF method for robot bin picking. To improve the pose hypothesis result, a new clustering strategy combined with DBSCAN and PCA is proposed. Also, we define the mismatching region and probable pose distribution to obtain more information in PPF method. Experiments on the real scene point clouds show that our approach provides the capability of mismatching region detection compared with original version.

We believe that PPF method still remain promising potential if we combine it with 2D image for 3D object detection and pose estimation problem. And the probable pose distribution has the potential to offer more constraints for other pose estimation techniques.

Acknowledgements. This work is partially supported by the National Natural Science Foundation of China (51375309). The support is gratefully acknowledged.

References

1. Drost, B., Ulrich, M., Navab, N., Ilic, S.: Model globally, match locally: efficient and robust 3D object recognition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 998–1005. IEEE (2010)
2. Pavel, Z.: Detection and localization of texture-less objects in RGB-D images (2015)
3. Birdal, T., Ilic, S.: Point pair features based object detection and pose estimation revisited. In: 2015 International Conference on 3D Vision (3DV), pp. 527–535. IEEE (2015)
4. Hinterstoisser, S., Lepetit, V., Rajkumar, N., Konolige, K.: Going further with point pair features. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 834–848. Springer, Cham (2016). doi:[10.1007/978-3-319-46487-9_51](https://doi.org/10.1007/978-3-319-46487-9_51)
5. Choi, C., Taguchi, Y., Tuzel, O., Liu, M.-Y., Ramalingam, S.: Voting-based pose estimation for robotic assembly using a 3D sensor. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 1724–1731. IEEE (2012)
6. Choi, C., Christensen, H.I.: RGB-D object pose estimation in unstructured environments. *Robot. Auton. Syst.* **75**, 595–613 (2016)

7. Papazov, C., Burschka, D.: An efficient RANSAC for 3D object recognition in noisy and occluded scenes. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010. LNCS, vol. 6492, pp. 135–148. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-19315-6_11](https://doi.org/10.1007/978-3-642-19315-6_11)
8. Abbeloos, W., Goedemé, T.: Point pair feature based object detection for random bin picking. In: 2016 13th Conference on Computer and Robot Vision (CRV), pp. 432–439. IEEE (2016)
9. Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., Lepetit, V.: Gradient response maps for real-time detection of textureless objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 876–888 (2012)
10. Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N.: Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 548–562. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-37331-2_42](https://doi.org/10.1007/978-3-642-37331-2_42)
11. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD 1996, pp. 226–231 (1996)
12. Kumar, K.M., Reddy, A.R.M.: A fast DBSCAN clustering algorithm by accelerating neighbor searching using groups method. *Pattern Recogn.* **58**, 39–48 (2016)
13. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992)
14. Yang, J., Li, H., Jia, Y.: Go-ICP: Solving 3D registration efficiently and globally optimally. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1457–1464 (2013)
15. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3D registration. In: IEEE International Conference on Robotics and Automation, ICRA 2009, pp. 3212–3217. IEEE (2009)
16. Knopp, J., Prasad, M., Willems, G., Timofte, R., Van Gool, L.: Hough transform and 3D SURF for robust three dimensional classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 589–602. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15567-3_43](https://doi.org/10.1007/978-3-642-15567-3_43)
17. Guo, Y., Bennamoun, M., Sohel, F., Min, L., Wan, J.: 3D object recognition in cluttered scenes with local surface features: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(11), 2270–2287 (2014)
18. Kouskouridas, R., Tejani, A., Doumanoglou, A., Tang, D., Kim, T.-K.: Latent-class hough forests for 6 DoF object pose estimation. arXiv preprint [arXiv:1602.01464](https://arxiv.org/abs/1602.01464) (2016)
19. Sahin, C., Kouskouridas, R., Kim, T.-K.: A learning-based variable size part extraction architecture for 6D object pose recovery in depth. arXiv preprint [arXiv:1701.02166](https://arxiv.org/abs/1701.02166) (2017)
20. Maturana, D., Scherer, S.: VoxNet: a 3D convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 922–928. IEEE (2015)
21. Song, S., Xiao, J.: Deep sliding shapes for amodal 3D object detection in RGB-D images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 808–816 (2016)
22. Qi, C.R., Su, H., Nießner, M., Dai, A., Yan, M., Guibas, L.J.: Volumetric and multi-view CNNs for object classification on 3D data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5648–5656 (2016)

23. Correll, N., Bekris, K.E., Berenson, D., Brock, O., Causo, A., Hauser, K., Okada, K., Rodriguez, A., Romano, J.M., Wurman, P.R.: Analysis and observations from the first Amazon picking challenge. *IEEE Trans. Autom. Sci. Eng.* (2016)
24. Zeng, A., Yu, K.-T., Song, S., Suo, D., Walker Jr., E., Rodriguez, A., Xiao, J.: Multi-view self-supervised deep learning for 6D pose estimation in the amazon picking challenge. arXiv preprint [arXiv:1609.09475](https://arxiv.org/abs/1609.09475) (2016)
25. Wong, J.M., Kee, V., Le, T., Wagner, S., Mariottini, G.-L., Schneider, A., Hamilton, L., Chipalkatty, R., Hebert, M., Segicp, D.J., et al.: Integrated deep semantic segmentation and pose estimation. arXiv preprint [arXiv:1703.01661](https://arxiv.org/abs/1703.01661) (2017)