2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
June 28-30, 2023. Seattle, Washington, USA

Pose Estimation Based on Point Pair Features with Optimized Voting and Verification Strategies

Gaoming Chen, Ao Gao, Wenhang Liu, Chao Liu, and Zhenhua Xiong

***Abstract*—In many industrial scenarios, such as automotive production lines, sheet stamping parts are widely used. However, due to the weak texture and strong reflections of these typical parts, pose estimation is hard to achieve, resulting in difficulties of grasping automatically. To deal with this problem, we propose a novel point pair feature (PPF) based pose estimation method to facilitate grasping. Firstly, a three-level structure downsampling method is introduced to seek the balance between the number of model points and significant features. Secondly, in order to reduce the interference of placement plane and other objects in the scene, a two-dimensional voting accumulator is constructed with weighted voting. Based on the voting results, the probability map is accordingly established, which guides keypoints sampling and voting again. Finally, edge points and model points are enrolled for pose verification to remove the wrong results. Our method is implemented in physical experiments, and the results show that the proposed method can be effectively applied to pose estimation of sheet stamping parts such as tire lock plates. Moreover, the ablation study demonstrates the criticality of each process.**

***Index Terms*—Pose estimation, Point pair features, Sheet stamping parts.**

I. Introduction

Recently, with the improvement of sensor accuracy, object recognition and pose estimation have developed rapidly, which can be used for bin picking, augmented reality, autonomous driving, etc [1]. For bin picking task, pose estimation is mainly to recognize the target object from the scene and estimate its pose relative to the camera. Then, the pose of the object relative to the manipulator is obtained according to the results of hand-eye calibration, and we can control the manipulator to grasp. Sheet stamping parts are usually composed of simple geometric primitives such as plane, and have a wide range of applications in the automobile production. Due to the characteristics of parts with weak texture, strong reflection and multi-plane, pose estimation and manipulator automatic grasping are facing challenges. Therefore, it is necessary to develop appropriate pose estimation method for sheet stamping parts to realize the automation of automobile production line.

The existing pose estimation methods are mainly divided into template matching based methods, learning based methods and feature based methods [2]. For the unstructured environment, due to parts stacking, the traditional template matching algorithm requires plenty of memory [3]. The deep learning approach performs poorly because of the imperfections and errors in the point cloud captured. In feature based methods, most 3D feature descriptors describe local features, while point pair feature (PPF) describes the model globally, which has rotation invariance and high robustness. Therefore, the point pair feature matching algorithm is used as the basic algorithm to estimate the pose of sheet stamping parts. In PPF based method, there are three main sections that affect the efficiency and accuracy of the algorithm, namely preprocessing, voting and pose verification.

Because models and scenes are usually represented by numerous points, downsampling the model is an important step in the preprocessing section of the offline stage, and the completeness of the model description by the downsampling points directly affects the results. PPF matching algorithm was first proposed by Drost et al. in [4]. Voxel filtering was applied to downsample the model, and each voxel only retained its center point. In [5], the points whose angles between the normals were larger than 30 degrees in the same voxel were reserved, which reduced the loss of valid information. In [6], the visible points from N viewpoints were extracted and aligned using the CAD model of the target, and then reduced model description through voxel filtering. In [7], poisson disk sampling was used to improve the efficiency of matching and prevent point pairs from being too close.

Sheet stamping parts are composed of simple geometric primitives typically, usually with many planar point pair features, thus vulnerable to the interference of the plane in the scene. In [7], the voting weight was calculated according to the angle between the normals of the point pair. The smaller the angle, the smaller the voting weight. In [8], the direction of the point used the tangent direction. The edge points of the scene were only chosen to calculate point pair features, thus partly reducing the influence of the plane. [9] adopted RANSAC approach to extract planes from the scene and removed them larger than the model diameter, but this method is difficult to separate thin parts from the plane. In [10], deep learning method was used to segment the instance, and then sampled and matched in the masks, which resulted in efficient sampling of point pairs. In [11], each point pair was ray tracing in four directions, and these distances obtained were added to the PPF.

An effective pose verification method can delete the mismatched pose, thus making the results more accurate. In [5], the object was rendered according to the estimated pose and filtered based on the matching number of pixels. The final verification was carried out by judging the coverage between the areas with significant depth change and the transformed model. In [12], voxel-based pose verification method was proposed to detect multiple parts efficiently owing to boolean

This work was supported in part by the Major Science and Technology Projects for Self-Innovation of FAW (20210301032GX).

Gaoming Chen, Ao Gao, Wenhang Liu, Chao Liu, and Zhenhua Xiong are with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China (e-mail: cgm1015@sjtu.edu.cn; gagagaga152@sjtu.edu.cn; liuwenhang@sjtu.edu.cn; aalon@sjtu.edu.cn; mexiong@sjtu.edu.cn)

**Corresponding author: Zhenhua Xiong, Chao Liu.*

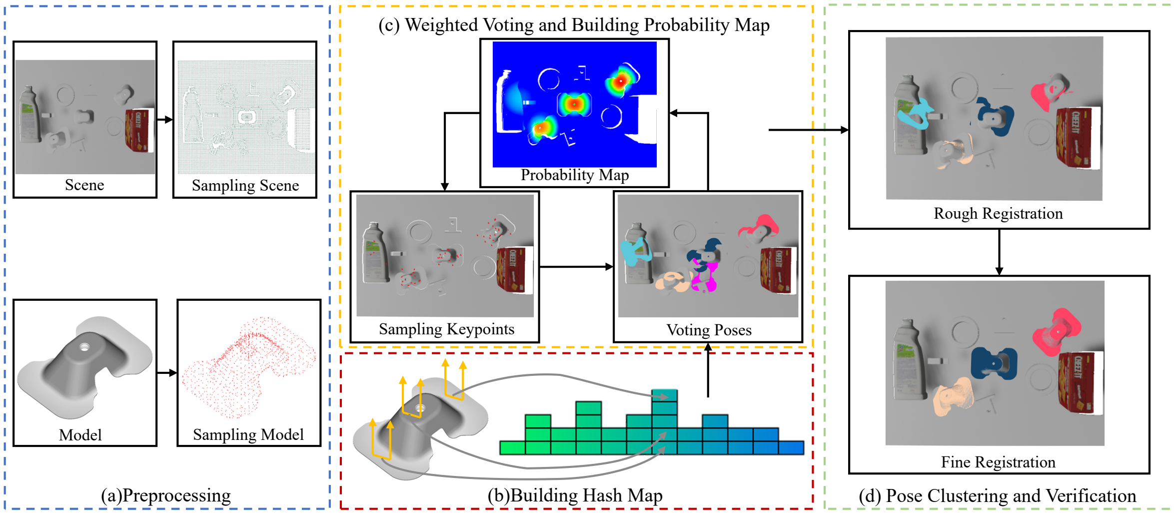


Fig. 1. Illustration of the proposed pipeline. (a) In preprocessing, downsample the model and the scene. (b) Hash map construction based on model point pair features. (c) The probability map is established according to the voting results, which guides the keypoints sampling and voting again. (d) Improved pose verification method is conducted to eliminate invalid results.

operation, which reduced program runtime. In [13], if the distance between the corresponding points of the transformed model and the scene, and the angle between their normals were both less than the threshold, they were considered to be matched, and the number of matched points is calculated for evaluation. In [14], semantic PPF method could remove the ambiguity of the pose hypotheses because the semantic information could be applied to match the right surface.

However, in the above works, the downsampling methods lack the explicit extraction of edge points, which will lead to the loss of details. Also, because the sheet stamping parts have many planar point pair features typically, and the thickness of the part is within the sensor error range, the placement plane and other object surfaces in the scene may obtain higher erroneous votes. Furthermore, due to the highly reflective features of parts, point cloud captured by the structured light camera are incomplete. The general pose verification methods tend to perform poorly in complex scenes where parts stacking.

In this paper, we mainly focus on the pose estimation of sheet stamping parts in the case of scattered stacking, and propose a PPF based method with optimized voting and verification strategies. Firstly, the model is downsampled by our three-level structure downsampling method. Then, through weighted voting and probability map construction, keypoints are obtained, and the voting strategy is executed again to obtain the initial pose. Finally, according to pose verification method considering edge points and model points, the wrong poses can be removed. The results show that the proposed method can be effectively applied to estimate the pose of sheet stamping parts in industrial applications quickly and accurately, which has the ability to complete the bin picking task of 50 parts.

The remainder of this paper is organized as follows. In Section II, the proposed method consisting of preprocessing, weighted voting, probability map building and pose verification are explained in detail. Then, experimental results and discussion are given in Section III. Section IV concludes the paper, and outlines the future work.

II. Methodology

Drost-PPFM [4] proposed by Drost et al. performed model matching on point pair features, and then estimated the pose of the target object in the scene. Given a point pair (p_r, p_i) whose normals are respectively n_r and n_i, the distance vector between two points is $d = p_i - p_r$. The four-dimensional point pair feature descriptor is shown in Eq. (1).

$$F(p_r, p_i) = \left(\|d\|_2, \angle(n_r, d), \angle(n_i, d), \angle(n_r, n_i)\right) \quad (1)$$

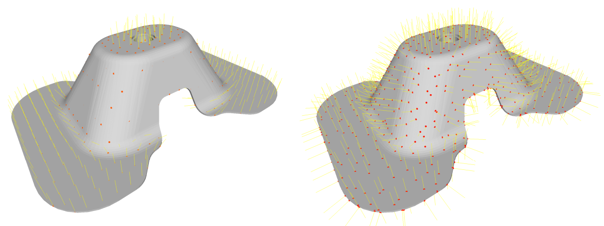
The method we proposed mainly consists of four sections, and its pipeline is shown in Fig. 1. First of all, the model and scene are preprocessed to obtain the hash map described by the model point pairs. Then, the weighted voting is performed, and the probability map is constructed according to the voting results. The keypoints are sampled according to probability and voted again to obtain initial poses. Finally, cluster each initial result, perform iterative closest point (ICP) for each average pose, and get the final results through pose verification.

A. Preprocessing

Since both model point cloud $\mathcal{M}$ and scene point cloud $\mathcal{S}$ have a large amount of data, effectively reducing the number of points can speed up matching. However, the traditional downsampling algorithm based on voxel filtering only retains the center of gravity of each voxel, which will obviously lead to the absence of significant features of the model, such as the cambered surface, thus making the registration worse.

In order to solve this problem, we propose a three-level structure downsampling method. According to the algorithm proposed by Bendels et al. [15], we get the model edge points. For each voxel v_i of model, we not only retain the point $p_{c,i}$ closest to the center of v_i, but also store the points $p_{n,i}$ whose angles between their normals and other storage normals are larger than 30 degrees and the equidistant edge points $p_{e,i}$ of $\mathcal{M}$ in v_i. Consider the voxel index space $\mathbb{V}$ of $\mathcal{M}$, three levels of downsampling point set $\mathcal{M}_d = \{p_{c,i}, p_{n,i}, p_{e,i}, i \in \mathbb{V}\}$ with size N_m. Use the same method to obtain the scene downsampling point set $\mathcal{S}_d$ with size N_s.

Assume that the diameter of the model is D, and set the length of the downsampled voxel to $\delta_V D$, where $\delta_V = 0.04$. The three-level structure downsampling method can increase the number of tire lock plate downsampling points from 632 to 1281, as shown in Fig. 2. It can be seen that our method retains much information about the upper and lower surfaces, the cambered surface and the edge of the part.



(a) Traditional method (b) Our method

Fig. 2. Comparison of downsampling methods. (a) Traditional voxel filtering downsampling. (b) Our three-level structure downsampling method.

According to the camera projection imaging principle, when the angle between the normals of two points p_r and p_i is within $[180° - \varepsilon, 180° + \varepsilon]$, such as one point on the inner surface and the other point on the outer surface, they will hardly appear in the same view. Therefore, for these point pairs, we do not calculate the point pair features, so as to reduce the storage.

B. Weighted Voting and Probability Map

Because sheet stamping parts often have many planar features, the point pair features on the placement planes of the scene have a bad influence on pose estimation, which often leads to wrong matching. When generating model hash map in offline stage, the distance and angle elements in PPF are discretized with a quantization step size of Δd_{dist} and Δd_{angle}. As shown in Fig. 3, when the PPF descriptor extracted from the scene is $F(s_r, s_i) = (r \pm \frac{\Delta d_{dist}}{2}, 90° \pm \frac{\Delta d_{angle}}{2}, 90° \pm \frac{\Delta d_{angle}}{2}, 0° \pm \frac{\Delta d_{angle}}{2})$, it can correspond to multiple point pairs of the model, such as $(p_{r,1}, p_{i,1})$, $(p_{r,2}, p_{i,2})$ and so on, thus generating matching ambiguity.

When building the voting accumulator, we adopt the improved method mentioned in [5]. For any reference point $s_r \in \mathcal{S}_d$, we only build the point pair with the point s_i whose distance from s_r is less than $\frac{D}{2}$. Reducing the number of votes of planar point pair can lighten the interference of plane to a certain extent. At the same time, point pair with large angle of normals provide approximate information gain for pose estimation and should have the same weight. Therefore, we propose a discrete weighted voting method to achieve the above purposes. Due to the noise of the sensor, when the angle $\angle(n_r, n_i)$ between the normals of (s_r, s_i) in the scene is less than 30 degrees, vote 1 for the corresponding model point and the rotation angle in the voting accumulator. Otherwise, it is considered that the point pair is not in the same plane, and

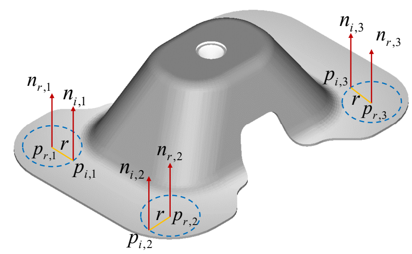


Fig. 3. Planar point pair features in sheet stamping part.

vote 5. The voting rule is described in Eq. (2).

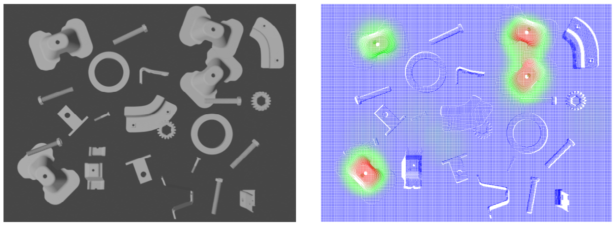
$$V(m_r, \alpha) = \begin{cases} 1, if \angle(n_r, n_i) < 30° \\ 5, if \angle(n_r, n_i) \geq 30° \end{cases} \tag{2}$$

According to the voting results, a large number of estimated pose will be generated, but many of them are inaccurate. Relatively speaking, the higher the number of votes obtained, the higher the reliability, which can be considered to be closer to the ground truth. Inspired by [9], when a scene reference point gets a high number of votes, it means that this point is more likely to be located on the target, which can provide more accurate results. They build the probability map based on the scene reference points, but we consider that the pose center is more effective in the face of large plane interference.

Therefore, we propose the probability map representation for pose center. The whole architecture is given in Algorithm 1. For the kth reference point $s_{r,k}$ in $\mathcal{S}_d$, the model reference point $m_{r,k}$ and pose transformation matrix T_k are obtained according to the highest number of votes $V_{\max,k}$. The nearest neighbor search algorithm is applied to find the nearest point $s_{c,k}$ of the tranformed model center point $m_{c,k}$ in the scene, and a scoring ball $S_{b,k}$ with diameter D is established with $s_{c,k}$ as the center. The updated score at each point s_j in $S_{b,k}$ decreases according to the Euclidean distance from $s_{c,k}$, where the score of $s_{c,k}$ is $V_{\max,k}$. Assuming that the score of scenario point s_j after the kth vote is $S(s_j)_k$, the score after next vote is updated as shown in Eq. (3). After all voting, normalize the score of each point in $\mathcal{S}$, and finally obtain the probability map representation for pose center, as shown in Fig. 4, where the probability from 0 to 1, color from blue to red. It can be seen that there are obviously four regions of interest with high probability in the scene, corresponding to the four parts, which can well sense the distribution of target.

$$S(s_j)_{k+1} = S(s_j)_k + \left\lfloor \frac{\|s_j - s_{c,k}\|_2}{D/2} V_{\max,k} \right\rfloor \tag{3}$$

For the obtained probability map representation for pose center, we conduct keypoint sampling according to the discrete probability distribution. A point s_j is obtained by simple random sampling from the scene. If the probability P_j of the point s_j is greater than the random probability P_{random}, s_j is retained as a keypoint. Repeat the above operations until the size of keypoint set $\mathcal{K}$ is 1/100 of the size of $\mathcal{S}_d$. In this way, $\mathcal{K}$ can not only have a certain global view, but also focus on the local region of interest. The weighted voting is performed again to obtain the initial estimated pose set $\mathcal{T}_1$.



(a) Original image (b) Probability map

Fig. 4. Construction of probability map. (a) Image captured by simulation camera. (b) The probability map representation for pose center.

Algorithm 1 Keypoints Sampling with Probability Map

Input: Model PPF hash map; Model downsampling point set $\mathcal{M}_d$; Scene downsampling point set $\mathcal{S}_d$.
Output: Keypoint set $\mathcal{K}$.

Initializing: $(S(s_1)_0, \ldots, S(s_{N_s})_0) \leftarrow 0$
for $k = 1$ to N_s **do**
  $T_k \leftarrow$ weighted voting with Eq. (2)
  $m_{c,k} \leftarrow$ transformed model center point m_c
  $s_{c,k} \leftarrow$ the nearest point of $m_{c,k}$
  $N_{b,k} \leftarrow$ the size of point in $S_{b,k}$
  for $j = 1$ to $N_{b,k}$ **do**
    $S(s_j)_k$ updated with (3)
  end for
end for
$N_k \leftarrow$ the size of $\mathcal{K}$
while $N_k < 1/100 N_m$ **do**
  $s_j \leftarrow$ random sampling in $\mathcal{S}_d$
  compute random probability P_{random}
  if $P_j > P_{random}$ **then**
    s_j is added to $\mathcal{K}$
  end if
end while

C. Pose Clustering and Verification

Since the keypoints may be in the background, and the sensor noise will interfere with the PPF matching, there are still some wrong poses in $\mathcal{T}_1$. In order to remove some inaccurate results, we select clusters according to the number of votes for each pose. Suppose that after the keypoints voting, the highest number of votes is $Votes_{\max}$, only the poses with more votes than $0.5Votes_{\max}$ are retained, and other results are considered inaccurate and deleted. The Euclidean distance D_T of the translation vector and the angle difference α_T of the quaternion between each pose and the cluster centers are used as the metrics. If D_T is less than the threshold ΔD_T and α_T is less than the threshold $\Delta\alpha_T$, the pose will be merged into the existing cluster, otherwise a new cluster will be created. The translation vector of each cluster takes the mean value of the translation vector of all poses in it, and the rotation is represented by the eigenvector corresponding to the maximum eigenvalue of matrix A in Eq. (4), where n_i represents the number of poses and q_j represents the quaternion corresponding to each pose in the cluster. The clustered pose transformation set $\mathcal{T}_2$ is obtained. For each pose in $\mathcal{T}_2$, ICP is applied to refine the pose to obtain the set $\mathcal{T}_3$.

$$A = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(q_j^T \cdot q_j\right) \tag{4}$$

Since the edge points provide much information in model matching, we propose a pose verification method fusion of edge points and model points. The Canny operator is used for edge detection of the RGB image, and the edge is mapped to $\mathcal{S}$ to obtain the edge points. For $T_i \in \mathcal{T}_3$, transform the downsampling points and equidistant edge points of the model into the scene. KD-tree is applied for nearest neighbor search. If the transformed model downsampling point $m_{i,j}$ has an adjacent point $s_j \in \mathcal{S}$, and the angle between their normals is smaller than 30 degrees, it is considered that they correspond, and the total number of corresponding points Nd_i is obtained. Voxelize the edge points of scene point cloud. If there are edge points in the voxel where the transformed equidistant model edge point are located, it is considered that this edge point is matched successfully, and the number of all matched edge points Ne_i is finally obtained. The final score of each pose is calculated in Eq. (5), where ω_d and ω_e are weight coefficients. In this paper, 1 and 3 are used respectively. Retain the poses with 0.5 times the highest score as the final estimation results.

$$Score_i = \omega_d \cdot Nd_i + \omega_e \cdot Ne_i \tag{5}$$

III. Experiments

In this section, the experiments verify the necessity of each module in our proposed algorithm, and then compare it with other algorithms. Finally, the proposed method is applied to bin picking experiments in real scenes. The sheet stamping part used in the experiment is the tire lock plate, which is applied to fix the tire and the vehicle.

A. Ablation Study

We conduct ablation study to verify the necessity of each component proposed above, including four experiments: using our method as shown in Fig. 5(a), downsampling the model with traditional voxel filtering as shown in Fig. 5(b), no weighted voting and establishment of probability map as shown in Fig. 5(c), and no pose verification as shown in Fig. 5(d). The algorithms applied to the last three experiments are the same as the proposed method except for the above changes.

As shown in Fig. 5(a), our method almost perfectly matches all the target objects in the scene, achieving the highest accuracy. When traditional voxel filtering is used for downsampling, the model is smaller with only 632 points so the program runs faster, but the result has a certain deviation from the ground truth, which is caused by ignoring the details of the part. Compared with our method, the method with traditional voxel filtering has an average greater error of 0.213mm and 0.035rad in translation and rotation respectively, and obtains the second highest accuracy. The latter two methods have obvious mismatches, so only qualitative analysis is carried out. When the weighted voting and probability map construction are not carried out, the object recognition is seriously affected

(a) Ours (b) Without Section II.A

(c) Without Section II.B (d) Without Section II.C

Fig. 5. Ablation study. (a) Our method. (b) With traditional downsampling. (c) Without weighted voting and building probability map. (d) Without pose verification.

by the placement plane, and PPF matching generates many wrong poses. Even if the pose verification is carried out, the effect is still poor, because the wrong poses are matched with other objects in the scene, which ultimately leads to the pose verification failure, as shown in the red marking areas in Fig. 5(c). When the pose verification is not applied, the incorrect poses cannot be removed effectively. There are obvious mismatches in Fig. 5(d), which shows the necessity of optimized pose verification. Therefore, the above three improvements can improve the accuracy of pose estimation.

B. Comparisons

In order to calculate the accuracy of the proposed algorithm, we use three kinds of sheet stamping parts for testing, and the Blender software is applied to construct 20 simulation scenes respectively, in which the parts are stacked in disorder. The RGB image and point cloud are obtained through simulation, and some of the scenes are shown in Fig. 6.

As the final task is oriented to bin picking, we use the three results with the highest score in each scene to compare with the ground truth to calculate the translation and rotation errors. Taking the X direction as an example, the error of the algorithm in this direction is shown in Eq. (6), where $x_{i,j}$ and $xt_{i,j}$ represent the prediction result and ground truth of the jth pose in the X direction of the ith scene respectively.

$$e_x = \frac{1}{180}\sum_{i=1}^{60}\sum_{j=1}^{3}|x_{i,j} - xt_{i,j}| \tag{6}$$

The proposed algorithm is compared with Drost-PPFM [4], Fast-PPFM [12] and PPF-MEAM [8]. All algorithms finally adopt ICP for pose refinement. The experimental results are shown in Table I. The results show that in the three degrees of freedom of translation, our method has the highest accuracy in the prediction of X direction and Z direction, which is at least 2.168 times and 1.324 times higher than other methods, respectively. At the same time, our method has the second highest accuracy in Y direction prediction, whose error is only 0.085mm greater than the most accurate result. In addition, for the three degrees of freedom of rotation, our method has the highest accuracy at $roll$ and $pitch$ angles, 1.449 times and 1.854 times higher than the second highest accuracy, separately. Finally, for the yaw angle, our method achieves the second highest accuracy, which is 1.4×10^{-4}rad greater than the error of the best result. The above results show that the proposed algorithm has higher accuracy in the prediction of multiple degrees of freedom, and the overall pose estimation results are more accurate. It is worth mentioning that the precision accuracy of PPF-MEAM [8] is close to that of the algorithm in this paper, which also shows the effectiveness of edge information in the pose estimation of sheet stamping parts.

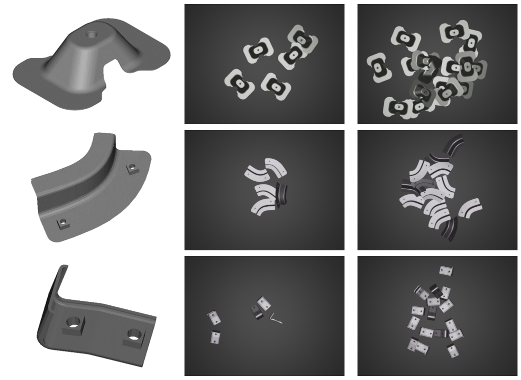


Fig. 6. Three kinds of sheet stamping parts and some simulation scenarios.

TABLE I
POSE ESTIMATION ERROR OF DIFFERENT METHODS

Methods	e_x [1]	e_y [1]	e_z [1]	e_{roll} [2]	e_{pitch} [2]	e_{yaw} [2]
Drost-PPFM	1.581	1.563	1.826	5.207	1.649	0.642
Fast-PPFM	1.352	**0.328**	1.127	1.794	1.482	0.785
PPF-MEAM	1.249	0.615	0.633	1.072	1.107	**0.576**
Ours	**0.576**	0.413	**0.478**	**0.740**	**0.597**	0.590

[1] the unit of variable is mm
[2] the unit of variable is 10^{-2}rad

C. Bin Picking

We apply the algorithm to the real scene for bin picking experiment. The Universal Robot (UR) is equipped with an electromagnet for grasping with structured light camera in a way that eye to hand. According to the electromagnet structure and the shape of the tire lock plate, the left, middle and right planes of each part are taken as candidate grasping areas. Because the algorithm will produce many results, we adopt the height of the part H_i, the matching degree between the part and the scene point cloud M_i, and the number of interference points I_i above the part to comprehensively evaluate after normalization, and finally select the candidate area with the highest score as the target area with Eq. (7). At the same time, if there is collision in this area, choose the pose with the second highest score until it meets the requirement of no collision.

$$PScore_i = \omega_H \cdot H_i + \omega_M \cdot M_i - \omega_I \cdot I_i \tag{7}$$

where ω_H, ω_M and ω_I are weight coefficients, while 1, 1 and 2 are respectively adopted in this paper.

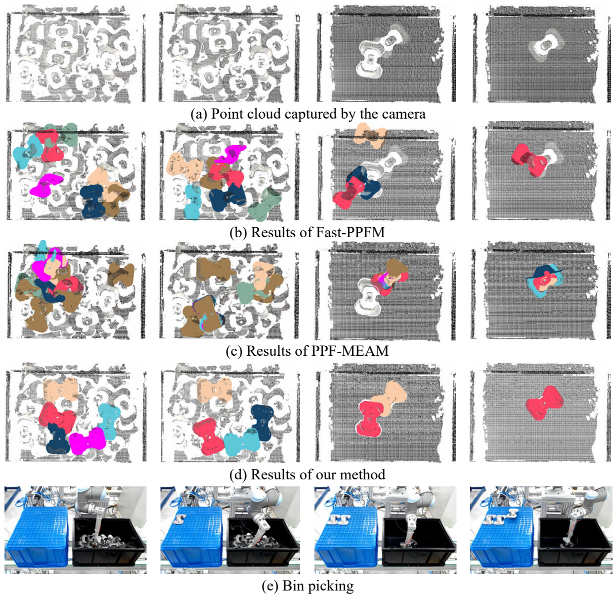


Fig. 7. The experimental process of bin picking. (a) Some point clouds captured by the structured light camera. (b) Pose estimation results of Fast-PPFM. (c) Pose estimation results of PPF-MEAM. (d) Pose estimation results of our method. (e) Bin picking with our method.

Some results during the experiment are shown in Fig. 7. As shown in Fig. 7(a), due to the high specular reflection of the part slope, the point cloud captured has some defects, which brings great challenges to the algorithm. When there are few parts left, it can be seen from Fig. 7(b) that the Fast-PPFM [12] is seriously disturbed by the placement plane and cannot recognize the target from the scene. In contrast, PPF-MEAM [8] can identify some targets when there are few sheet stamping parts, because relatively accurate edge information can be obtained, but it is not effective when there are many parts stacked in disorder owing to the messy edges. In spite of this, even in the face of incomplete point clouds, the algorithm proposed in this paper can still guarantee better pose estimation results, no matter when there are more or fewer parts as shown in Fig. 7(d). Finally, our experiment shows that the proposed algorithm can grasp a whole box of 50 parts stacked in disorder.

D. Discussion

Through ablation study, we verify the effectiveness of the improvements in the proposed method, which can deal with the interference of plane and other objects in the scene. Moreover, as shown in Table I, compared with other three methods, our method has the smallest prediction error at four degrees of freedom. Even the errors of other two degrees of freedom are a little larger than the best result. However, because the keypoints are random sampling according to the probability map, when there are many target objects in the scene, our method cannot guarantee that keypoints are distributed on all targets, which may lead to incomplete pose estimation in the scene.

IV. Conclusion and Future Work

We propose a PPF based pose estimation method with optimized voting and verification strategies for sheet stamping parts recognition and grasping. The three-level structure downsampling method can effectively extract salient features of the model while reducing model points. Weighted voting and probability map construction can reduce plane interference. The pose verification method combining edge and model points can filter wrong poses. Our algorithm has achieved good results in real scenes and can complete bin picking tasks.

The future work is to explore more efficient pose estimation methods. For the parts composed of simple geometric primitives, we can use the graph matching method for target recognition and pose estimation. Also, ensemble learning can be applied to take advantage of multiple algorithms.

References

[1] Z. He, W. Feng, X. Zhao, and Y. Lv, "6d pose estimation of objects: Recent technologies and challenges," *Applied Sciences*, vol. 11, no. 1, p. 228, 2020.

[2] 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 and Vision Computing*, vol. 96, p. 103898, 2020.

[3] C. Ye, K. Li, L. Jia, C. Zhuang, and Z. Xiong, "Fast hierarchical template matching strategy for real-time pose estimation of texture-less objects," in *International Conference on Intelligent Robotics and Applications*. Springer, 2016, pp. 225–236.

[4] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *2010 IEEE computer society conference on computer vision and pattern recognition*. Ieee, 2010, pp. 998–1005.

[5] S. Hinterstoisser, V. Lepetit, N. Rajkumar, and K. Konolige, "Going further with point pair features," in *European conference on computer vision*. Springer, 2016, pp. 834–848.

[6] D. Liu, S. Arai, Y. Xu, F. Tokuda, and K. Kosuge, "6d pose estimation of occlusion-free objects for robotic bin-picking using ppf-meam with 2d images (occlusion-free ppf-meam)," *IEEE Access*, vol. 9, pp. 50 857–50 871, 2021.

[7] N. Wang, J. Lin, X. Zhang, and X. Zheng, "Fast and robust object pose estimation based on point pair feature for bin picking," in *2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*. IEEE, 2021, pp. 528–533.

[8] D. Liu, S. Arai, J. Miao, J. Kinugawa, Z. Wang, and K. Kosuge, "Point pair feature-based pose estimation with multiple edge appearance models (ppf-meam) for robotic bin picking," *Sensors*, vol. 18, no. 8, p. 2719, 2018.

[9] J. Guo, X. Xing, W. Quan, D.-M. Yan, Q. Gu, Y. Liu, and X. Zhang, "Efficient center voting for object detection and 6d pose estimation in 3d point cloud," *IEEE Transactions on Image Processing*, vol. 30, pp. 5072–5084, 2021.

[10] 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," *Robotics and Autonomous Systems*, vol. 141, p. 103775, 2021.

[11] M. Ziegler, M. Rudorfer, X. Kroischke, S. Krone, and J. Krüger, "Point pair feature matching: Evaluating methods to detect simple shapes," in *International Conference on Computer Vision Systems*. Springer, 2019, pp. 445–456.

[12] M. Li and K. Hashimoto, "Fast and robust pose estimation algorithm for bin picking using point pair feature," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1604–1609.

[13] X. Xing, J. Guo, L. Nan, Q. Gu, X. Zhang, and D.-M. Yan, "Efficient mspso sampling for object detection and 6-d pose estimation in 3-d scenes," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 10, pp. 10 281–10 291, 2021.

[14] C. Zhuang, Z. Wang, H. Zhao, and H. Ding, "Semantic part segmentation method based 3d object pose estimation with rgb-d images for bin-picking," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102086, 2021.

[15] G. H. Bendels, R. Schnabel, and R. Klein, "Detecting holes in point set surfaces," 2006.