2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Recognition of disordered workpieces based on 3D Laser scanner and RS-CNN

Sikui He, Bin Ye, Huijun Li, Yong Gao
School of Information and Control Engineering
China University of Mining and Technology
Xuzhou, China
Email: 1362220974@qq.com; yebin@cumt.edu.cn;
plutoli@163.com; 1075452304@qq.com

Abstract—In industrial production, the disordered grasping operation of the robotic arm is mostly for grasping a single type of workpiece. Effective grasping is not easy when multiple overlapping workpieces are mixed together. The mutual occlusion between workpieces causes the loss of geometric shape information, which makes it difficult to obtain the precise grasping pose of each workpiece. In this paper, a 3D laser scanner is used to acquire the point cloud features of the workpiece. At first, the point clouds are filtered and segmented, and then they are input into the RS-CNN network for recognition and classification. According to the classification results, different models are used to register the point clouds in the scene. At last, the final pose of the workpiece to be grasped is obtained, which realizes the disorderly grasping of various workpieces.

Keywords- workpiece; classification; point cloud; recognition

I. INTRODUCTION

In the manufacturing industry, the task of grasping of disordered workpieces is the picking of parts using industrial robots. Analyze and grab workpiece stacking scenarios such as bearings, metal tubes, and crankshafts placed in material box. The random placement of various workpieces will cause overlapping and occlusion between each other, resulting in the laser scanner can only obtain the local geometric features of the workpiece, it is difficult to realize the classification and recognition of the workpiece. Using neural network to learn the features of the neighboring points of the point cloud can effectively improve the recognition rate of the workpiece.

Depending on the type of input data to the neural network, the usual 3D shape classification methods can be divided into multi-view based, volume-based and point-based methods. MVCNN [1] simply maximizes the multi-view feature as a global feature. But max pooling only retains the maximum value of the features under a specific view, which leads to the loss of a large amount of data information. MHBN [2] coordinates bilinear pooling to integrate local convolutional features, and produces compact global features by this method. Recent advances in multi-view based methods have been made in the research direction of 3D shape recognition and retrieval [3-6]. However, the method based on multi-view projection is more sensitive to viewpoint selection and object occlusion, which causes information loss during 3D to 2D projection. OctNet [7] uses an unbalanced octree to subtly segment the point cloud structure, each leaf node of the octree stores a pooled feature, which reduces the memory requirements and computation of 3D convolutional networks. Qi improved PointNet [8] and proposed the PointNet++ [9], which gradually learns from a larger local area and solves the problems caused by point cloud inhomogeneity and density variation. PointConv [10] regards the convolution kernel as a nonlinear function of local neighborhood point coordinates, which consists of a weight function and a density function, the density function is learned by kernel density estimation.

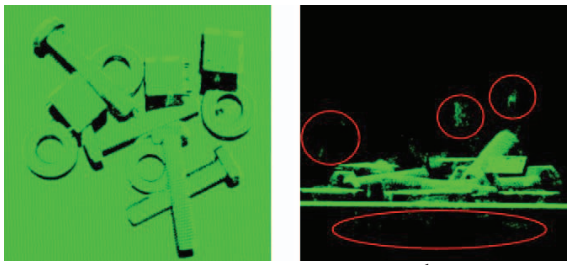
The 6-DoF GraspNet [11] algorithm takes the point cloud as input, uses the designed grasping evaluation model to refine the grasping posture and selects the optimal grasping posture, but it takes a long time. HongZhou [12] proposed a lightweight end-to-end grasp evaluation network PointNetGPD, which can directly evaluate the point cloud in the actuator, but when the point cloud is sparse, overfitting and performance degradation will occur. The Fast-RCNN network [13] realizes the recognition of bearing and door sheet metal parts, combines the two-dimensional image with the point cloud, obtains the point cloud information of the recognized workpiece and performs fast pose estimation.

In this paper, a 3D laser scanner is used to perceive the scene and build an experimental platform for disorderly grabbing. The research contents are as follows: first, obtain the scene point cloud, then preprocessed using statistical filtering, segmentation and clustering. Second, Analogy to the structure of convolutional neural networks in images, learning the features of nearest neighbors in point clouds based on RS-CNN network. Third, train the collected actual point cloud data to realize the recognition of clustering point cloud and get classification labels.

II. DISORDERLY CAPYURE SCENE POINT CLOUD SEGMENTATION

In this paper, a line laser scanning camera is used to study the disordered grasping technology. In order to classify and identify the point cloud of workpiece, the point cloud must first be preprocessed and segmented. Place the material boxes containing different types of workpieces on the side of the conveyor belt, then start the conveyor belt and the laser scanner, waiting for the grasped workpiece to pass through the camera scanning area, the perception information of the scene is obtained, the obtained point cloud of the real scene is shown in Figure 1.

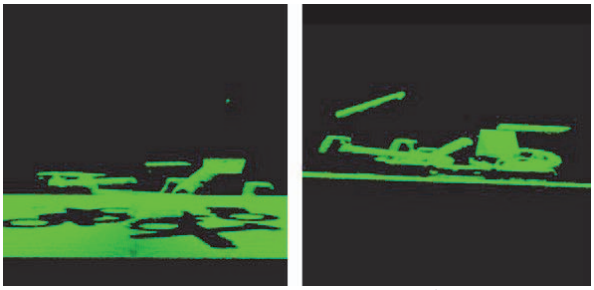
2473-3636/22/$31.00 ©2022 IEEE
DOI 10.1109/DCABES57229.2022.00052



a b

Fig.1 point cloud with noise

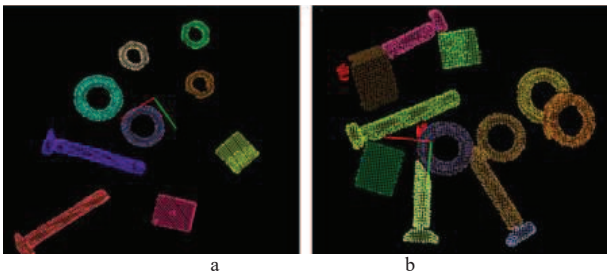
In the figure, there is no obvious noise from the top view, but from the side view, it can be seen that the circled area of the point cloud contains scattered noise points, which are scattered in the scene point cloud disorderly. This is due to the influence of the material properties of the workpiece surface or the external environment such as light and other factors. Using the features of outliers, define a point cloud somewhere less than a certain density as an invalid point cloud. Count the neighborhood of each point in the point cloud and calculate its average distance from all nearby points. The distance of all points in the point cloud should form a Gaussian distribution. Its shape is determined by the mean and standard deviation, so the points with average distance outside the standard range are judged as outliers and removed from the data. The removal effect of noise point cloud is shown in Figure 2.



a b

Fig. 2 Statistical filtering effect

After filtering out the noise points, the point cloud image includes the background and the target. The background plane points cannot play a role in feature extraction for subsequent segmentation recognition and pose estimation, a large number of useless points will increase the calculation time. Using the RANSAC algorithm can effectively Fit and filter out the conveyor belt plane. The characteristics of stacking and occlusion of the workpiece make the acquired point clouds stick, that is, the point clouds of two objects have continuity at the edge. Therefore, a more detailed segmentation of the point cloud is required to distinguish part of the point cloud of each object. Figure 3 shows the effects of the two scenarios obtained after filtering out planes and clustering segmentation.



a b

Fig.3 Euclidean clustering segmentation result

III. Point Cloud Classification Network Based on RS-CNN

RS-CNN [14] learns from the geometric topological relationships and constraints between point clouds and can encode meaningful shape information in point clouds. The learning from part to whole has achieved great success in image CNN. However, Inductive learning is very dependent on the shape perception of irregular subsets in point clouds. so to solve this problem, RS-CNN performs a modeling as shown in Figure 4.

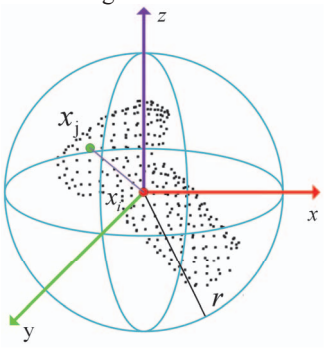


Fig.4 Model of local point subset

The local point set is represented by $\boldsymbol{P_{ush}} \subset \mathbf{R}^3$, which is a sphere formed with the sampling point $\boldsymbol{x_i}$ as the center and the radius is $\mathbf{r}$, the points contained in it are used as the local point set. The neighbors of the point $\boldsymbol{x_i}$ are set as $\boldsymbol{x_j} \in \boldsymbol{N(x_i)}$. After modeling the local shape of the point cloud, it is necessary to inductively represent this local point set $\boldsymbol{f_{P_{sub}}}$ to distinguish the potential shape information it contains. The convolution operation is formulated here as

$$\boldsymbol{f_{P_{sub}}} = \boldsymbol{\sigma}\left(\boldsymbol{A}\left(\left\{\boldsymbol{\tau}\left(\boldsymbol{f_{x_j}}\right), \forall \boldsymbol{x_j}\right\}\right)\right), \boldsymbol{d_{ij}} < \boldsymbol{r}, \forall \boldsymbol{x_j} \in \boldsymbol{N(x_i)} \quad (1)$$

Where $\boldsymbol{x}$ represents the 3D point, $\boldsymbol{f}$ represents the feature vector. $\boldsymbol{d_{ij}}$ represents the Euclidean distance between the sampling point $\boldsymbol{x_i}$ and the neighborhood point $\boldsymbol{x_j}$. A feature transformation of $\boldsymbol{N(x_i)}$ under the function $\boldsymbol{\tau}$, and then using the function $\boldsymbol{A}$ and the nonlinear activation factor $\boldsymbol{\sigma}$ to aggregate the features to obtain $\boldsymbol{f_{P_{sub}}}$. In this formula, the functions $\boldsymbol{A}$ and $\boldsymbol{\tau}$ are the key factors to obtain the eigenvectors. The permutation invariance of the point cloud is satisfied only when the function $\boldsymbol{A}$ acts as a symmetric function and $\boldsymbol{\tau}$ is shared over $\boldsymbol{N(x_i)}$.

In classic CNN, τ Functions can be expressed as

$$\tau(f_{x_j}) = w_j \cdot f_{x_j} \quad (2)$$

Where w_j is learnable weight. When this function is put into the point cloud, w_j does not share weights for each point in $N(x_i)$, nor can it handle irregular $f_{P_{sub}}$, w_j in backpropagation is only related to the independent point x_j, This doesn't give good shape awareness, so this function is modified.

For the sampling point x_i, the distance between it and its neighbor points $x_j \in N(x_i)$ can well represent the local shape and the spatial distribution of the points, so the weight w_j in the traditional CNN is changed to w_{ij}, and the low-level relationship is set, For example, the Euclidean distance is defined as h_{ij}. Define the τ function as

$$\tau(f_{x_j}) = w_{ij} \cdot f_{x_j} = M(h_{ij}) \cdot f_{x_j} \quad (3)$$

M stands for transforming the low-level relational representation between two points into a high-level relational representation in order to encode their spatial distribution, because the multilayer perceptron (MLP) has a better mapping ability, so here we use more Layer Perceptron (MLP) for feature mapping.

By designing the network, it is obtained that the proposed RS-Conv overcomes the permutation invariance, while points are not isolated from each other. Points and their immediate neighbors form meaningful shapes in the geometric space and encode the relationships between points. Interaction relations between point pairs are obtained and weight sharing is achieved, which allows the same learning function to be applied on different rule subsets.

Learning from classical convolutional neural networks, namely RS-CNN, formulas for developing point cloud classification and recognition as

$$F^{l}_{P_{N_l}} = RS - Conv(F^{l-1}_{P_{N_{l-1}}}) \quad (4)$$

In the formula, $F^{l}_{P_{N_l}}$ represents the feature of the l layer of the sampling point set P_{N_l} with N_l points, which is obtained by applying RS-Conv to the features of the $l-1$ layer.

The overall network structure of point cloud classification using RS-CNN is shown in Figure 5. In the figure, N represents the number of points, C represents the number of channels. A fully connected layer is added at the end of the classification network to achieve end-to-end training.

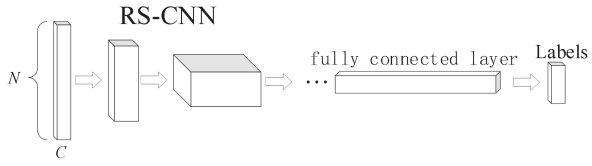


Fig.5 The architectures of RS-CNN applied in the classification

After designing the network structure, the specific implementation details of the convolution operation are as follows: max poling is used as the symmetric function A. ReLU [15] as nonlinear activation function σ, a three-layer multilayer perceptron (MLP) is used to realize the feature mapping with the distance between two points, The vector of two points and the coordinates of two points ($\|x_i - x_j\|$, $x_i - x_j$, x_i, x_j) ten-dimensional feature as input low-level point cloud relationship, using single-layer MLP for channel lift mapping.

The local point cloud subset is sampled using the farthest point sampling method proposed by PointNet, within each neighborhood, a fixed number of point sets are selected for batching and normalization.

IV. Experimental Analysis of Disordered Grasp Classification

A. Data Set Production

Collect the data of different workpieces to make a data set, scan a single object, perform point cloud filtering, plane removal, retain the object point cloud, then obtain a scanning model of a single workpiece. Label information is assigned to each type of artifact point cloud as training labels for the neural network. The collected workpiece point cloud for network training is shown in Figure 6.

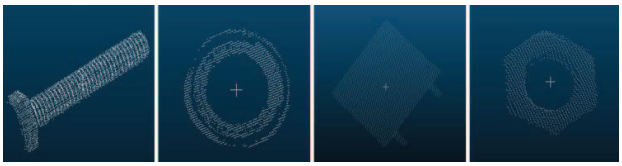


Fig.6 Example of training data

B. Training Process

In order to enhance the effect of training, data enhancement is performed on the training data, the point cloud is rotated and translated. The training process on the grasped artifact dataset is shown in Figure 7, which shows the change in the accuracy of artifact recognition during training.

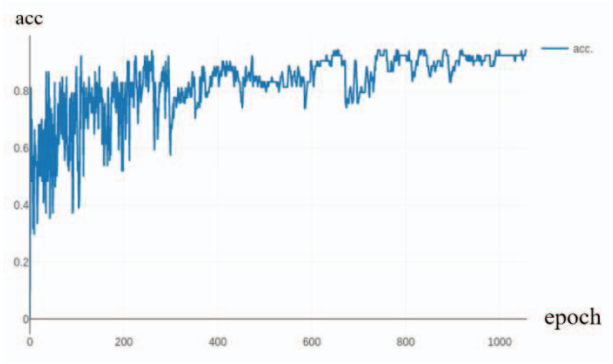


Fig.7 Neural network training process

The process of using the trained model for prediction is: separate the point cloud clusters generated in the point cloud segmentation process, then put the clustered point cloud into the network, regress the label of the point cloud type .Visualize the segmentation results and attach the prediction results of the model, as shown in Figure 8.

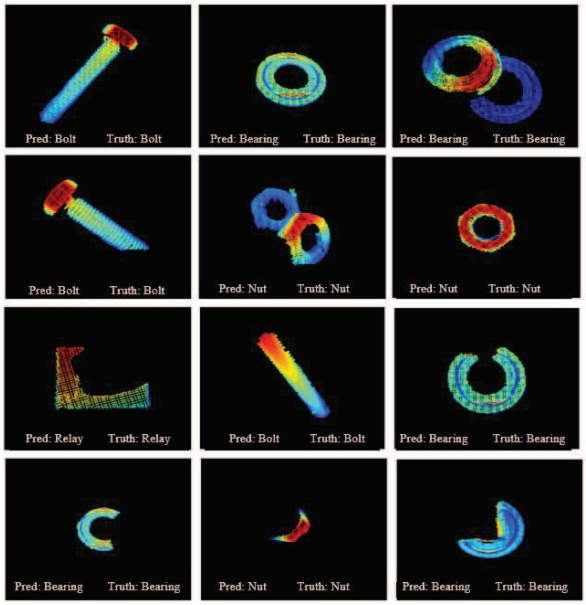


Fig.8 example of correct recognition

The wrong identification samples are shown in Figure 9.

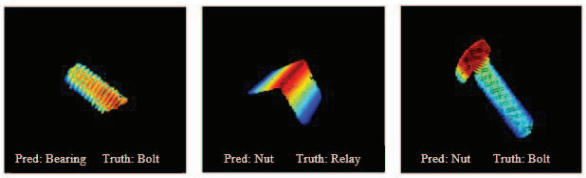


Fig.9 example of mistake recognition

C. Result Analysis

Using the segmentation map of the real point cloud for prediction, the prediction accuracy of the model is: 94.4%. RS-CNN performs well in point cloud classification. For the occlusion problem caused by the stacking scene, as shown in Figure 8, a large number of point clouds caused by the occlusion problem are not scanned completely and the Euclidean clustering is not segmented due to stacking. The neural network is still Its type can be better identified. For the case of wrong classification, one is that the segmented part accounts for a small proportion of the whole, as shown in the first picture of Figure 9, because the segmented point cloud only occupies a small part of the entire bolt model, resulting in a recognition error, The reason for the identification error in the second picture is the identification error caused by the large gap between the segmented point cloud clustering and the relay model in the training set. For recognition errors, in order to improve the recognition rate, the data set can be continuously expanded or the number of training times can be increased to improve the accuracy. On the whole, the recognition of segmented point cloud can meet the requirements of disordered grasping.

V. CONCLUSION

In this paper, the visual processing algorithm based on point cloud is used to identify and classify various types of workpieces in the disorderly grasping task, which provides target information for the robot to grasp. The point cloud classification network proposed in this paper not only improves the visual recognition success rate of workpiece but also enhances the sorting recognition capability and environmental adaptation capability of the disordered grasping system. However, it is cumbersome to perform both point cloud segmentation and feature recognition in sequence. The next improvement direction is to design a semantic segmentation network based on RS-CNN according to the structure of RS-CNN. By using this network, the scene point cloud can be directly segmented, which can reduce the processing complexity.

REFERENCES

[1] H. Su, S. Maji, E. Kalogerakis and E. Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 945-953, doi: 10.1109/ICCV.2015.114.

[2] T. Yu, J. Meng and J. Yuan, "Multi-view Harmonized Bilinear Network for 3D Object Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 186-194, doi: 10.1109/CVPR.2018.00027.

[3] C. Qi, et al., "Volumetric and Multi-view CNNs for Object Classification on 3D Data," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016 pp. 5648-5656.doi: 10.1109/CVPR.2016.609.

[4] Y. Feng, Z. Zhang, X. Zhao, R. Ji and Y. Gao, "GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 264-272, doi: 10.1109/CVPR.2018.00035.

[5] Chu W , Pelillo M , Siddiqi K . Dominant Set Clustering and Pooling for Multi-View 3D Object Recognition[J]. 2017.10.5244/C.31.64.

[6] Ma C , Guo Y , Yang J , et al. Learning Multi-View Representation With LSTM for 3-D Shape Recognition and Retrieval[J]. IEEE Transactions on Multimedia, 2019, 21(5):1169-1182.

[7] G. Riegler, A. O. Ulusoy and A. Geiger, "OctNet: Learning Deep 3D Representations at High Resolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6620-6629, doi: 10.1109/CVPR.2017.701.

[8] R. Q. Charles, H. Su, M. Kaichun and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 77-85, doi: 10.1109/CVPR.2017.16.

[9] Qi C R , Yi L , Su H , et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[J]. arXiv e-prints,2017: arXiv:1706.02413.

[10] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3D point clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 9621–9630, 2019.

[11] A. Mousavian, C. Eppner and D. Fox, "6-DOF GraspNet: Variational Grasp Generation for Object Manipulation," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 2901-2910, doi: 10.1109/ICCV.2019.00299.

[12] HongZhou Liang, XiaoJian Ma, et al. PointNetGPD: Detecting Grasp Configurations from Point Sets[J], arXiv:1809.06267,2018.

[13] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[14] Liu Y, Fan B, Xiang S, et al. Relation-shape convolutional neural network for point cloud analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8895-8904.

[15] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In ICML, pages 807–814, 2010.4 .