

106. The 3D object recognition with environmental adaptability based on VFH descriptor and region growing segmentation

Zhuang Peng¹, Jinbao Chen², Dong Han³, Meng Chen⁴

¹AVIC Chengdu Aircraft Industrial (Group) Co., Ltd., Chengdu, China

^{2,3}Nanjing University of Aeronautics and Astronautics, Nanjing, China

⁴Shanghai Institute of Aerospace Systems Engineering, Shanghai, China

¹Corresponding author

E-mail: ¹zhuanggeph@nuaa.edu.cn, ²chenjbao@nuaa.edu.cn, ³han_dongnuaa@126.com,

⁴workmailcm@126.com

Received 12 May 2015; received in revised form 4 October 2016; accepted 17 October 2016

DOI <https://doi.org/10.21595/jme.2016.16050>

Abstract. 3D object recognition is a basic research in the machine vision field. Microsoft KINECT V2 is utilized to collect external environmental information. The point cloud file is obtained after processing the collected information. In order to filter the point cloud and obtain point cloud model of a single object in the environment after region growing segmentation, the point cloud is applied to point cloud library. Then, the VFH descriptor of the point cloud model is calculated. After inputting point cloud model of the trained target, the point cloud model with the minimum CHI square distance between the VFH descriptor of the target and VFH descriptor of the point cloud model can be found. The 3D object corresponding to the found model is the identified object. For the 3D object recognition in an unfamiliar environment, the algorithm of 3D object recognition with environmental adaptability is proposed. After the 3D object recognition system built, the physical verification is conducted about the proposed algorithm. Giving the target model, the system successfully identifies the 3D object in the unfamiliar environment, that demonstrates the correctness of the algorithm.

Keywords: VFH descriptor, region growing segmentation, CHI square distance, environmental adaptability, 3D object recognition.

1. Introduction

3D object recognition is so versatile, such as space exploration, industrial production and household services [1-4]. So, it is a priority that need to be solved. 3D object recognition is an artificial intelligence technology which mimic human to recognize and identify objects in the environment through the eyes. However, according to the fact that the human vision occupies at least 60 percent of the human brain resources, 3D object recognition is considered as AI-complete problems in academic circles [5].

3D object recognition is generally divided into three steps. Firstly, it trains the target object and calculates the feature vector of the target. Then, it segments various objects in the environment and calculates the feature vector of each object. Finally, it gets the feature vector closest to the target feature vector. And its corresponding object is the target. Y. Guo et al. used rotational projection statistics for 3D local surface description and object recognition [6]. The type of data they used is MESH type which is more difficult to obtain. Steder et al. proposed NARF descriptors for object recognition [7]. In order to make the descriptor invariant to rotation, they extracted a unique orientation from the descriptor and shifted the descriptor according to this orientation. G. Zhou et al. proposed a method which uses the salient points as the key point to recognize the object [8]. But this method is limited for identifying an object with a smooth surface.

In order to identify the object in the unfamiliar environment, this paper proposes a method which adjusts the parameter dynamically so that the region growing segmentation can adapt to the environment. VFH descriptor is used to describe the object feature, and the target object is judged by CHI square distance of VFH descriptors. Finally, the research builds the 3D object recognition

system. It successfully identifies the irregular objects, that verifies the correctness of the method.

2. Point cloud acquisition

This paper utilizes KINECT V2 to integrate depth information and color information into a three-dimensional point cloud with color components. Firstly, a KINECT class is initialized. Then it gets color image frame and depth image frame resources and obtains the coordinates mapper. After creating the viewer classes of those resources, the data of the latest frame are acquired. Integration of color and depth resources needs to use already acquired coordinate mapper, because of the difference between coordinate system origin of two resources. The coordinates of current depth point in the color image frame are obtained in the coordinate mapper. Then, RGB values of the current depth point are obtained. And it gets XYZ values of the current depth point in the camera space coordinate system. The RGB values and the XYZ values of the current depth point are assigned to a Point XYZRGB-type point of PCL. After the above operations are done for each the point in the current depth frame, the model of the convenient-interactive colorful point cloud is obtained (Fig. 1(a)). In the following, this paper will use the point cloud model without color information (Fig. 1(b)).

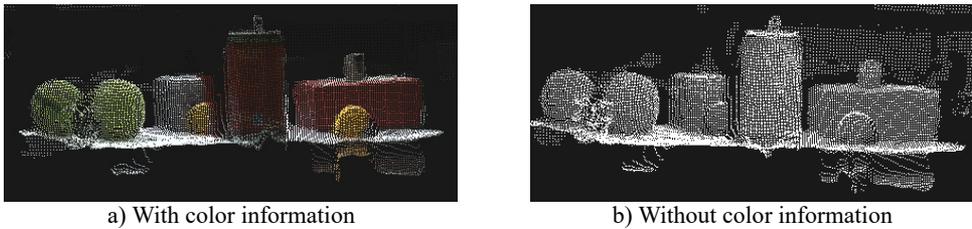


Fig. 1. Point cloud model

3. Region growing segmentation

Due to the instability of KINECT V2 sensor and the environmental factors, there are some noise in the model of raw point cloud (Fig. 1(b)). It may lead to calculation error of partial features of point cloud model, such as curvature and normal, thus resulting in the point cloud segmentation fail. In order to remove those points not meeting the criteria, statistical analysis for near point of each point is done [9]. For each point, the mean distance from it to all its neighbors is calculated. By assuming that the resulted distribution is Normal Distribution with a mean and a standard deviation, those points whose mean distance is outside an interval (Eq. (1)) can be considered as outliers. The interval has following formula:

$$I = \mu \pm \sigma \cdot k, \quad (1)$$

where I represents the interval, μ is the mean and σ is the standard deviation, k is standard deviation coefficient.

The number of point's neighbors is 50 and k is 0.5 in this paper. After outliers trimmed from the raw point cloud model, the new point cloud model becomes more real (Fig. 2).



Fig. 2. Filtered point cloud model

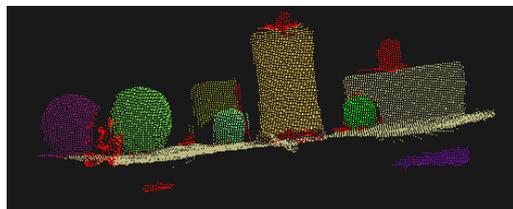


Fig. 3. Point cloud model after region growing segmentation

Then, this paper implements region growing segmentation on filtered point cloud model [10]. First of all, it computes the normal and curvature value of each point in point cloud model. It sorts the points by their curvature value. Region growth begins from the point having the smallest curvature value, because growth from the flattest area allows to reduce the total number of segments. The main process of region growth is as follows:

- 1) It picks up the point with minimum curvature value.
- 2) The picked point is added to the set called seeds.
- 3) For every seed point, the algorithm finds its neighbors.
- 4) Calculating the angle between the normal of seed point and the normal of each neighbor point, if the angle is less than threshold value, then current point is added to the current region.
- 5) For every neighbor, if the curvature is less than threshold value, this point is added to the seeds. And the current seed point is removed from the seeds.

Repeating the main process, if the seeds becomes empty, the growth of the picked point with minimum curvature value is completed. Removing those points of growth region from the model, it sorts the points by their curvature value again. Then it repeats the above-mentioned method. This paper sets the number of neighbors to 30, the angle threshold of normal to 0.05π and curvature threshold to 40. Then it gets those regions after region growing segmentation (Fig. 3). There different colors regions represent different objects, and non-touching regions of the same color also represent different objects.

4. Compute VFH descriptor

VFH descriptor is a novel representation for object features [11]. VFH descriptor consists of an extended FPFH of center point and the angle between the central viewpoint direction translated to each normal. Center point is the point with the minimum mean distance from itself to all points in the point cloud. The extended FPFH includes the relative pan, tilt, yaw angles and the distances between each point and the center point (Fig. 4).

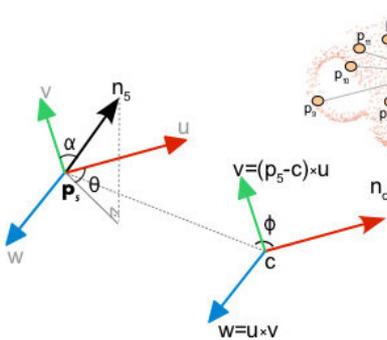


Fig. 4. The extended FPFH

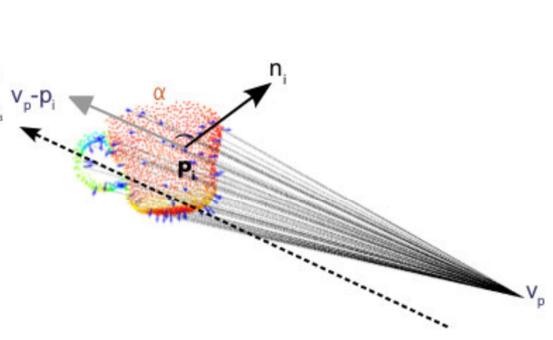


Fig. 5. Viewpoint direction angle

Where c is center point and n_c is normal of c and p_5 is a point in point cloud model and n_5 is normal of p_5 . Defining Cartesian coordinate system, where u is unit vector with the same direction of n_c and v is unit vector with the same direction of $(p_5 - c) \times u$ and w is unit vector with the same direction of $u \times v$. So, there are the following formulas:

$$\alpha = v \cdot n_5, \quad (2)$$

$$\phi = u \cdot \frac{p_5 - c}{d}, \quad (3)$$

$$\theta = \arctan(w \cdot n_5, u \cdot n_5), \quad (4)$$

$$d = \|p_5 - c\|_2. \quad (5)$$

α , ϕ and θ between center point and other points are computed, which is called the Simplified Point Feature Histogram (SPFH). Then:

$$FPFH(c) = SPFH(c) + \frac{1}{k} \sum_{i=1}^k \frac{1}{d_k} \cdot SPFH(p_k), \quad (6)$$

where k is the number of points in the sphere and the sphere's center is c and the sphere's radius is the distance from c to the farthest point and d_k is the distance from c to p_k .

Then, the angle between the central viewpoint direction translated (black dotted line in Fig. 5, not gray line) to each normal is computed by SPFH method (Fig. 5).

The VFH uses 45 binning subdivisions for each of the four extended FPFH values and 128 binning subdivisions for the viewpoint component. There are the VFH descriptors (Fig. 6) of the yellow cylinder region and the biggest blue sphere region in the Fig. 3.

Comparing the two VFH descriptors in Fig. 6, it can demonstrate that VFH descriptors of different shape and pose objects are different. So, this paper will take advantage of this feature for 3D object recognition.

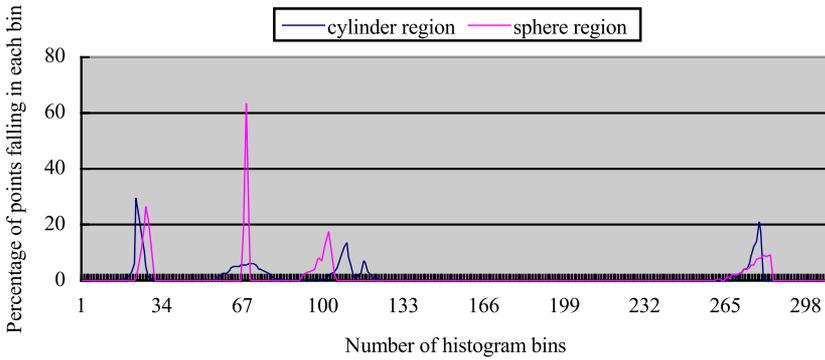


Fig. 6. VFH descriptors

5. 3D object recognition and the algorithm with environmental adaptability

3D object recognition process is as follows:

- 1) Establish point cloud model of the target object and calculate its VFH descriptor.
- 2) Establish point cloud model of the environment.
- 3) Filter point cloud model of the environment and complete region growing segmentation.
- 4) Calculate VFH descriptor of each region.
- 5) Calculate CHI square distance between target's VFH descriptor and each region's.
- 6) Find out the region with minimum CHI square distance that is the target object.

In fact, the algorithm has some faultiness that there are two objects not segmented correctly in Fig. 3. It must improve the algorithm for 3D object recognition in the unfamiliar environment. So, the algorithm of 3D object recognition with environmental adaptability is proposed. Aiming the minimum CHI square distance, the algorithm adjusts parameters to make 3D object recognition system adapt to the unfamiliar environment. According to the characteristics of the algorithm, the parameters of VFH descriptor calculation should be prespecified. After testing, the curvature threshold value is set to 40 that has a small impact on the region growing segmentation. So, the angle threshold of normal could be only adjusted dynamically.

It calculates CHI square distance between VFH descriptors of target and each region after each segmentation, where angle threshold of normal decreases from 30 degrees to 1 degree (Step: 1 degree). So, the region with minimum CHI square distance in 30 times segmentation can be found and its corresponding threshold is the parameter best adapted to the current environment. And the

region is identified as the target object.

6. Experimental verification

There is the point cloud model of the unfamiliar environment in Fig. 7(a) and the target object model in Fig. 7(c) which has been established previously. The best angle threshold of normal is 16 degrees in this unfamiliar environment after algorithm is executed. There are regions after region growing segmentation in Fig. 7(b) and the recognition result in Fig. 7(d).

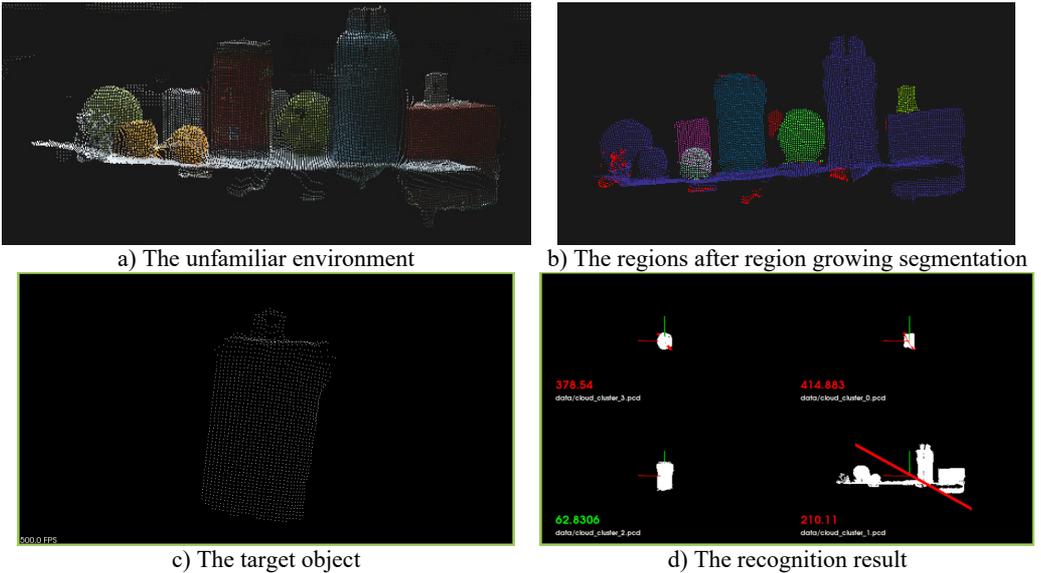


Fig. 7. Point cloud model

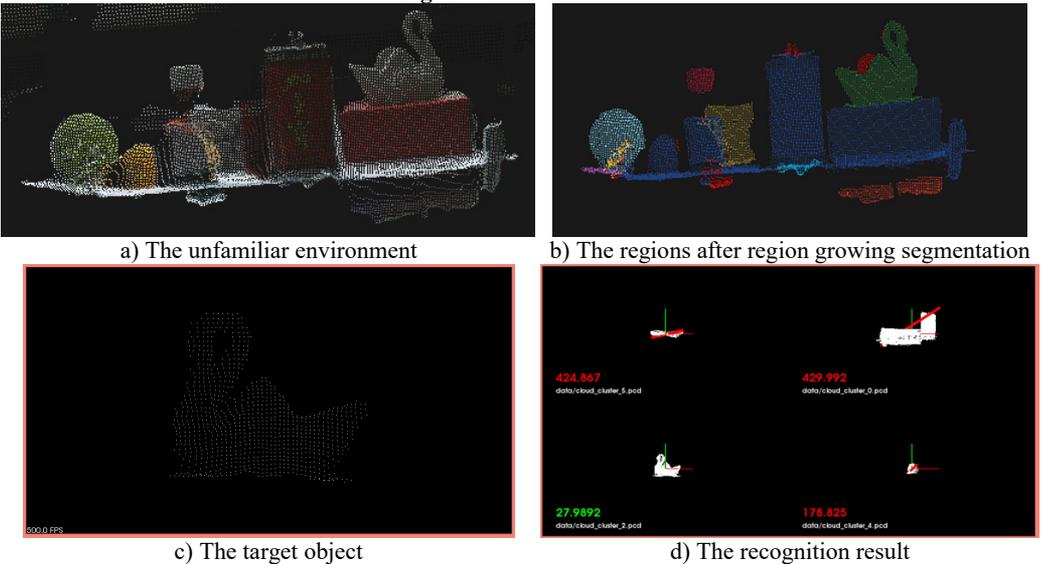


Fig. 8. Point cloud model

The identified object is the region of the lower left corner in Fig. 7(d) which is corresponding to the red cylinder object in the environment. Where 62.8306 is CHI square distance between the VFH descriptors of the target and actual object. The region of the lower right corner is not divided and includes 5 objects. If it includes the target, the angle threshold of normal will be adjusted

adaptively. Obviously, the relationship between the angle threshold of normal and the CHI square distance is nonlinear. Once the target has been segmented, it will be divided out in the next division, or divided into some smaller regions. That is why angle threshold of normal decreases from 30 degrees to 1 degree (Step: 1 degree).

Next, this paper verifies an irregular object recognition. There is the point cloud model of the unfamiliar environment in Fig. 7(a) and the target object model in Fig. 7(c) which has been established previously. The best angle threshold of normal is 14 degrees in this unfamiliar environment after algorithm is executed. There are regions after the region growing segmentation in Fig. 7(b) and the recognition result in Fig. 7(d).

7. Conclusions

Firstly, this paper builds point cloud model and segments the model by region growing segmentation. Then, it calculates the VFH descriptor of each region. The algorithm of 3D object recognition with environmental adaptability is proposed. The algorithm can make 3D object recognition system adapt to the unfamiliar environment and segment each object correctly. By the minimum CHI square distance of the VFH descriptors standards, it determines the angle threshold of normal. Finally, the system completes the 3D object recognition accurately, including regular object and irregular object.

However, it finds that the direction of the viewpoint has a significant impact on the object recognition. So, establishing the point cloud model of the multi-viewpoint direction of the target object is extremely important. In the future, further studies will be launched.

References

- [1] **Dong S., Williams B.** Learning and recognition of hybrid manipulation motions in variable environments using probabilistic flow tubes. *International Journal of Social Robotics*, Vol. 4, Issue 4, 2012, p. 357-368.
- [2] **Tombari F., Salti S., Di Stefano L.** Unique signatures of histograms for local surface description. *Computer Vision*, Springer Berlin Heidelberg, Vol. 6313, 2010, p. 356-369.
- [3] **Mian A., Bennamoun M., Owens R.** On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, Vol. 89, Issues 2-3, 2010, p. 348-361.
- [4] **Sukno F. M., Waddington J. L., Whelan P. F.** Comparing 3D descriptors for local search of craniofacial landmarks. *Advances in Visual Computing*, Springer Berlin Heidelberg, Vol. 7432, 2012, p. 92-103.
- [5] **Liu Guangcan.** Machine Learning Based Object Recognition. Dissertation, Shanghai Jiao Tong University, 2013, (in Chinese).
- [6] **Guo Y., Sohel F., Bennamoun M., et al.** Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision*, Vol. 105, Issue 1, 2013, p. 63-86.
- [7] **Steder B., Rusu R. B., Konolige K., et al.** NARF: 3D range image features for object recognition. *Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 40, 2010.
- [8] **Zou G., Hua J., Dong M., et al.** Surface matching with salient keypoints in geodesic scale space. *Computer Animation and Virtual Worlds*, Vol. 19, Issues 3-4, 2008, p. 399-410.
- [9] **Rusu R. B., Marton Z. C., Blodow N., et al.** Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems*, Vol. 56, Issue 11, 2008, p. 927-941.
- [10] **Rabbani T., van den Heuvel F., Vosselmann G.** Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, Issue 5, 2006, p. 248-253.
- [11] **Rusu R. B., Bradski G., Thibaux R., et al.** Fast 3d recognition and pose using the viewpoint feature histogram. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, p. 2155-2162.