www.horacongress.com

International Congress on Human-Computer Interaction, Optimization and Robotic Applications

July 5-7, 2019, Urgup, Nevşehir, Turkey

Conference Proceedings
www.set-science.com
8: 192-198, 2019

A Mobile Robot Application for 3D Object Classification via Point Cloud Data in Indoor Environment

Fatma Caran[1*], Güleycan Tuğrul[1], Ümit Göven[1], Kaya Turgut[1] and Burak Kaleci[1]

[1]*Electrical Engineering Department/Eskisehir Osmangazi University, Turkey*
[*]*(caranfatma@gmail.com)*

***Abstract* –**3D object classification has been widely studied in robotic community in recent years. According to types of features that are utilized to describe an object, 3D object classification studies can be separated into two main categories (local and global).The local feature-based studies are robust against partial occlusion and clutter. However, they generally require more computational time and memory consumption when it is compared with global feature-based studies. On the other hand, global-feature based studies are appropriate for 3D shape classification because they observe the entire geometry of objects. In this study, a global feature-based approach, Viewpoint Feature Histogram (VFH), is used to classify table, chair and bookshelf objects. Support Vector Machine (SVM) is applied to classify features. To analyze the classification accuracy, we modelled the ESOGU Electrical Engineering Laboratory building in GAZEBO and a P3-AT mobile robot with RGB-D camera was used to construct Gazebo dataset that includes table, chair, and bookshelf objects. Similarly, a dataset for real-environment is constructed with the same objects in ESOGU Electrical Engineering Laboratory building. The test results show that the implemented method is able to classify each object with a classification rate above 83% and 60% in Gazebo dataset and real-environment dataset, respectively.

Keywords –3D object classification, VFH, point cloud, mobile robot, indoor environments.

I. Introduction

3D object classification has been widely studied in the field of robotics in recent years and it has been used in a wide range of applications, from finding objects for industrial robotic arms to generating semantic maps via mobile robots. Classification of 3D objects in these application areas is important to increase autonomy of robots. In this way, robots can perform the tasks that are expected to them more efficiently. In previous studies, object classification was generally applied to 2D images. Since image data do not contain information about the depth and geometry of the object, these studies can be employed in limited circumstances. In this study, objects such as chair, bookshelf and table that are frequently used in indoor environments are classified using point cloud data captured from a RGB-D sensor.

3D object classification approaches are divided into two main categories, namely local and global, according to types of features that are utilized to describe an object. Local feature-based approaches are robust against the clutter and partial occlusions that are often encountered in real life environment. However, the local feature-based studies generally require more computational time and memory consumption when it is compared with global feature-based studies since they rely on the geometric information of the local neighborhood of each key-point. On the other hand, global-feature based studies are appropriate for 3D shape classification and reproduction of 3D models because they observe the entire geometry of objects [1]. In this study, after a comprehensive analysis about the global-feature based approaches, Viewpoint Feature Histogram (VFH) was selected to extract features. The reason behind that selection was the objects (chair, bookshelf and table) to be classified have different geometric primitives. In addition, global feature-based approaches have low computational complexity compared to local feature-based approaches. In this way, it is intended to decrease the classification time as low as possible. Since VFH has O(n) time complexity, it is appropriate for that purpose. SVM was employed to classify the extracted features.

The main purpose of the study is to measure the efficiency of the implemented method on the simulated and real datasets when it is trained with synthetic dataset. To achieve this, the Computer Aided Design (CAD) models of ModelNet40 for chair, bookshelf and table objects were used as training dataset. Then, ESOGU Electrical Engineering Laboratory building was modeled in GAZEBO simulation environment and a P3-AT mobile robot with RGB-D camera was used to construct Gazebo dataset that includes chair, bookshelf and table objects. Similarly, a dataset for real-environment was constructed with the same objects in ESOGU Electrical Engineering Laboratory building. These Gazebo and real-environment datasets were used in the experimental work. The test results for Gazebo and real-environment datasets are analyzed in terms of classification rate. The test results show that the implemented method is able to classify each object

DOI:10.36287/setsci.4.5.038

with a classification rate above 83% and 60% in Gazebo dataset and real-environment dataset, respectively.

II. Related Works

3D object classification is a well-studied problem in computer vision. The first attempts for object classification utilized 2D descriptors on image data and some of them produced successful results in limited circumstances. After the introduction of 3D sensors such as LIDAR's, 3D lasers and RGB-D cameras, 3D point cloud data has been gained popularity in computer vision and robotic society. As a result, the studies that are aimed to develop 3D descriptors have been increased in recent year. Han et al. [2] reviewed the 3D point cloud descriptors. They categorized the descriptors into three major classes: local-based, global-based and hybrid-based approaches. Local-based descriptors exploit the neighbourhood relationship around the point and features are calculated separately for each point in the point cloud. Therefore, the features produced via local-based descriptors carry the information about the local geometry around the point instead of the whole object. On the other hand, global-based descriptors extract the features by using the entire point cloud model defining the object. For this reason, they provide information about object geometry. Hybrid-based descriptors attempt to combine the advantages of local-based and global-based descriptors while they aim to reduce the disadvantages of these descriptors. Although it is possible to classify objects with local-based, global-based or hybrid-based descriptors, global descriptors are preferred commonly for reasons such as calculation time and memory [3]. In this subsection, local-based, global-based or hybrid-based 3D point cloud descriptors will briefly discussed to explain why the VFH selected as a descriptor.

Point Feature Histogram (PFH) is a local-based descriptor and collects information about the local geometry of a point. It analyses the neighbourhood points within the region around a point. Then, it estimates the angles of pan (α), yaw (φ), and tilt (θ) between the point and each neighbourhood point. To do that, PFH utilizes difference between the directions of normal of each point pair. It then forms a histogram that combines the angles for each pair of points [4].

Rusu et al. [5] introduced a new local-based descriptor, which is called Fast Point Feature Histogram (FPFH), based on the PFH descriptor. FPFH is a fast variant of PFH. This descriptor was tested with noisy outdoor datasets. It is faster than PFH, but when compared with PFH, fine details of objects are observed to disappear in outdoor environments with FPFH.

Rusu et al. [6] extended FPFH to overcome insufficient classification results when the position of object is altered. Viewpoint Feature Histogram (VFH) is a global-based descriptor and it combines FPFH with viewpoint perspective.

Aldoma et al. stated that VFH descriptor could not produce efficient classification success due to its shortcomings when it is trained with synthetic data. They proposed the Clustered VFH (CVFH) descriptor to identify and to estimate the position of the objects for real data captured from the RGB-D camera. To achieve this, they exploited VFH descriptor's recognition and calculation efficiency. They also proposed a roll histogram for determining the position of objects because descriptors such as VFH and CVFH are independent of transformation [7].

Fan et al. [8] extended the CVFH descriptor, which does not perform well in similar geometric structures, using color and texture information and defined it as Color-CVFH. The proposed descriptor performs better than CVFH when it is used objects with similar colors and shapes. In order to reduce the computational complexity, SVM classifier was preferred over feature matching.

Wohlkingerve and Vincze created a synthetic dataset, which is called 3D-Net, with 200 different objects obtained from CAD models in office and kitchen environments. Then, they classified the data obtained in real life according to the closest neighbourhood relationship with partially completed point cloud models obtained from CAD models. In addition, they developed a new global-based descriptor, Ensemble of Shape Functions (ESF), by combining distance, angle, and area shape functions. They compared the ESF with the Shape Distribution on Voxel Surfaces (SDVS), VFH, CVFH, and Signature of Histogram of Orientation (SHOT) descriptors by using 3D-Net dataset. The test results showed that ESF produces promising classification success [9].

Alhamzi et al. proposed a hybrid-based descriptor by combining local-based FPFH and global-based VFH descriptors in their study. While VFH was preferred for object recognition, FPFH was used to estimate the position of the object. The authors stated that classification rate of the objects with VFH was more efficient than other descriptors such as ESF and CVFH [1].

Kasaei et al. presented a descriptor called Global Orthographic Object Descriptor (GOOD). There is a trade-off between the classification rate of this descriptor used for object recognition and object manipulation and computational complexity [3].

Alexandre examined the local-based and global-based descriptors in the Point Cloud Library (PCL) [10] on a dataset [11]. The author recommended that the descriptor should be selected according to desired recognition rate, time requirements and the size of the processed data and the task such as object classification or position estimation.

According to the previous studies mentioned above, it is clear that the descriptor to be selected should be appropriate to the problem and geometric structure of the objects. Since the geometric structure of the objects (chair, bookshelf and table) to be classified in this study was different each other, it was decided to use a global-based descriptor. Among the global-based descriptors, VFH was preferred as a descriptor because it works faster than other descriptors and provides more efficient classification rates in the studies performed.

III. Implemented Method

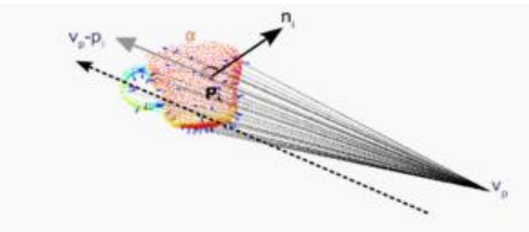
A. Viewpoint Feature Histogram (VFH)

VFH is a global feature extraction method and is based on a local-based descriptor, FPFH. FPFH is insufficient in some conditions because it cannot alter according to the position of the object. Therefore, VFH was developed based on both FPFH and benefiting from the perspective. VFH consists of two components. The first component is the viewpoint direction. For the calculation of this component, first the centroid of the point cloud is determined. Then, the unit vector is calculated between the centroid and the position of the sensor. The angles between the unit vector and the normal of the points in the point cloud are calculated and added on the histogram. 128 features are obtained from this first component. The second component is the expanded FPFH component and is calculated just like the FPFH. Therefore, in order to understand the second component of VFH, firstly PFH and then FPFH should be analyzed.

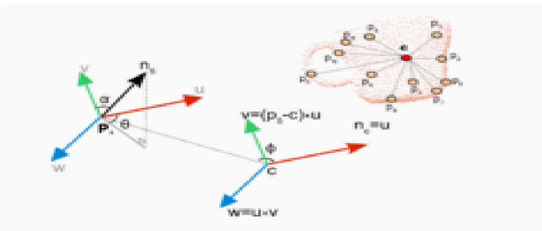
PFH is a local-based descriptor and calculates features for each point in the point cloud data. It uses the neighbourhood points within the region around a point. For each point pair, the point and its neighbourhood, PFH analyses the difference between the directions of normal and gathers information about the geometry surrounding the point. As a result, it estimates the angles of pan (α), yaw (φ), and tilt (θ) between the point and each neighbourhood point and forms a histogram that combines the angles for each pair of points [4]. Since the histogram collects angles of pan (α), yaw (φ), and tilt (θ) between each pair, the time complexity of the PFH is high. Equation 1 depicts the calculation of the pan, yaw, and tilt angles for a pair. In equation 1, n_i and n_j represent surface normals, p_i and p_j describe the selected point pairs where $i \neq j$, d represents the distance between the point pair, and u, v, w represent the Darboux frame coordinate system for the selected point p_i [6].

$$\begin{gathered} \alpha = v.n_i \\ \varphi = u.\frac{(p_i - p_j)}{d} \\ \theta = \tan^{-1}(w.n_j, u.n_j) \end{gathered} \tag{1}$$

FPFH is another local-based descriptor based on PFH. The reference circumference and angles are calculated in the same way as in PFH, but unlike PFH, it takes into account the direct connections between the current point and its neighbourhoods. Although it is faster than PFH, FPFH is slower when it is compared with other 3D descriptors. The reason of that is features are extracted for all points. The second component of VFH is calculated just like as FPFH. However, this calculation does not perform separately for each point. The perspective vector calculated in the first part is used as a normal and the angles are calculated only for the centroid of the point cloud. The second component yields 135 features, 45's in each of three angles (pan (α), yaw (σ), and tilt (θ)).



(a) Viewpoint Component



(b) Extended FPFH component of VFH

Fig. 1 Components of VFH [6]

PCL [9] calculates an additional histogram for the Shape Distribution Component (SDC). This histogram includes 45 features and the resultant descriptor contains 308 features. The SDC measures distances and encodes information about the distribution of points around the centroid. In this way, SDC allows distinguishing objects with similar characteristics according to their size and normal distribution [2]. By combining viewpoint components, FPFH components and the histograms resulting from the SDC, the VFH descriptor is constructed. The components of VFH are shown in Figure 1.

B. *Support Vector Machine (SVM)*

In previous studies, VFH global-based descriptor was widely used together with The Support Vector Machine (SVM) classifier. SVM is a classifier that separates classes via a hyperplane. In particular, it analyses and compares to four different multi-class strategies [12]. In other words, SVM determines optimal hyperplane according to the labelled training data. Then, the optimal hyperplane is utilized to categorize the new examples of test data. In two-dimensional space, this hyperplane is a line that divides a plane into two parts, each class on both sides. Figure 2 shows the linear division of the two classes. The support vector machine automatically determines the centers, weights and threshold that minimize the upper limit of the expected test error [13].

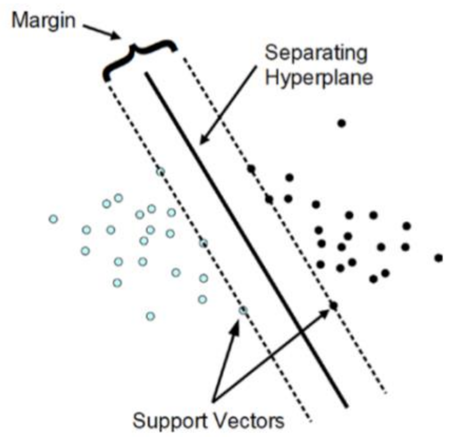


Fig. 2 Linear separation of two separate classes by SVM [14]

IV. Experimental Results

A. *ModelNet40 Dataset*

In the ModelNet40 dataset, there are 3D models belonging to 41 different objects created with Computer Aided Design (CAD) [15]. The CAD data in the ModelNet40 library is presented as a manifold triangular mesh model. In the models, surface model is constructed as possible as minimum numbers of point. Moreover, points are denser in on the edge regions.

Chair, bookshelf and table models that are intended to be classified in this study were obtained from ModelNet40 dataset. 989 chair, 672 bookshelf and 492 table models were prepared for training phase. Since the point cloud data obtained from the Gazebo and real-environment will not be a closed model like ModelNet40 dataset, it is pre-processed to simulate real-environment. Blender simulation environment [16] was used to generate the training data. Blender is a software that enables the creation of 3D objects. Blender can be employed for 3D modelling, animations and simulations. In this study, Blender environment is used to acquire point cloud data from surface of chair, bookshelf and table CAD objects with simulated RGB-D sensor. In this way, it is intended to simulate the data obtained from the real environment and to increase the diversity in the train data.

In the first stage, ModelNet40 model of an object and the RGB-D sensor were placed in the Blender simulation environment. An example of a chair and RGB-D sensor in the Blender simulation environment is shown in Figure 3. Then, point clouds were generated from16 different angles (0°, 30°, 45°, 60°, 90°, 120°, 135 °, 150°, 180°, 210°, 225°, 240°, 270°, 300°, 315°, and 330°) and 3 different distances (1, 2 and 3.5 meters). The procedure was repeated for all objects to obtain the training dataset. After this procedure is completed, the training dataset contains 16000 chair, 11200 bookshelf and 8000 table point cloud data. Point cloud samples of a chair object that are captured at 0°, 90°, 180°, and 270° angles are given in Figure 4.

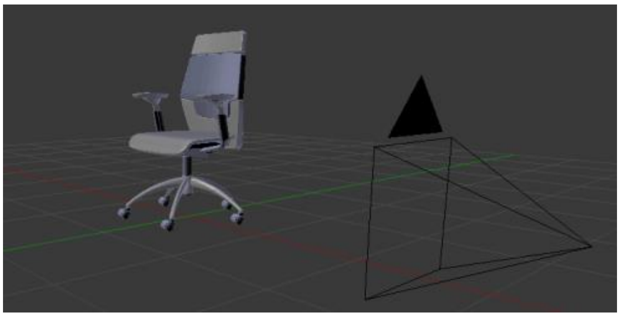
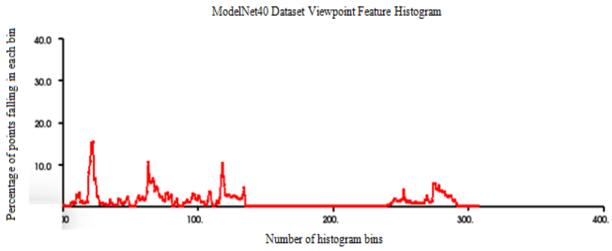


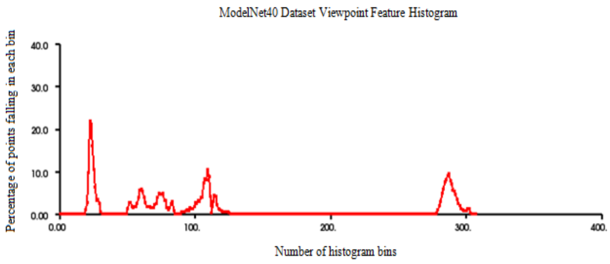
Fig. 3Sample image of a chair and a RGB-D sensor in Blender simulation environment

Fig. 4 Point cloud data of a chair taken from 0 °, 90 °, 180 °, 270 ° angles in Blender simulation environment.

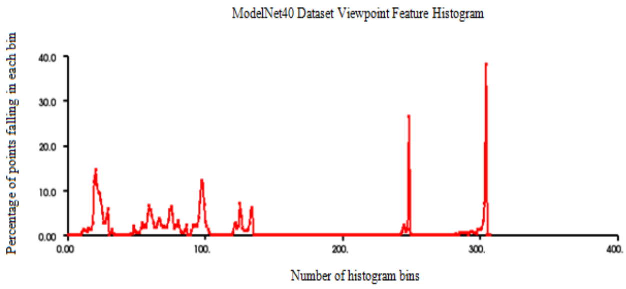
VFH is a descriptor that requires the normal of the points in the point cloud data. To calculate that normal value of a point, neighbourhood points that are enclosed in a certain radius are used. Then, point clouds containing the normal information obtained for chair, bookshelf and table objects were given as input to VFH descriptor and features were determined for each object. Figure 5 shows an example for extracted feature histograms of chair, bookshelf and table objects. In the figure, the horizontal axis and vertical axis show the number of bin and the percentage of points per bin, respectively. When the feature histograms of the chair, bookshelf and table objects were examined, it was observed that extracted features of each object were different from each other. As a result, chair, bookshelf and table objects can be distinguished from each other because different features are generated for each object. After determining the features for chair, bookshelf and table objects via VFH, training was conducted with SVM classifier.



(a) Chair



(b) Bookshelf



(c) Table

Fig. 5 Viewpoint Feature Histogram for chair, bookshelf and table objects obtained from ModelNet40 dataset

B. Gazebo Dataset

The main purpose of this study is to measure the efficiency of the implemented method on the simulated and real-environment datasets when it is trained with synthetic dataset. To achieve that, ESOGU Electrical Engineering Laboratory Building was modelled in GAZEBO simulation environment and a Pioneer P3-AT mobile robot with a RGB-D sensor is launched the environment. Figure 6 shows the modelled environment.

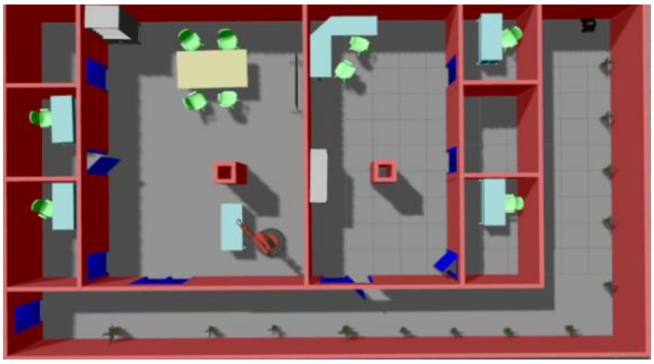
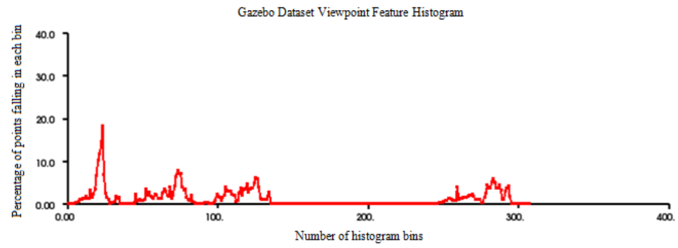


Fig.6 ESOGU Electrical Engineering Laboratory building Gazebo simulation environment

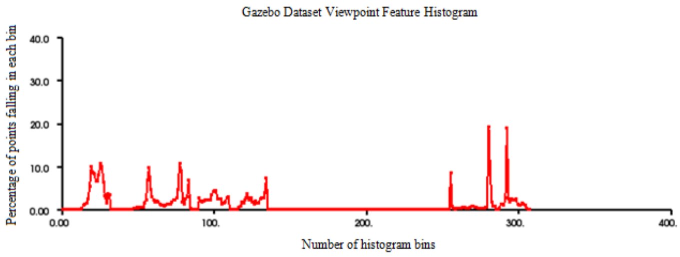
In this environment, the robot was moved and the point cloud data captured by the RGB-D sensor were recorded. Thus, a dataset was created. This dataset contains 38 chair, 42 bookshelf and 30 table point cloud data from different angles and positions. Figure 7 shows RGB images (left) and point cloud data (right) captured from the Gazebo simulation environment.

Fig. 7 RGB images (left) and point cloud data examples (right) for chair, bookshelf and table objects captured from Gazebo simulation environment

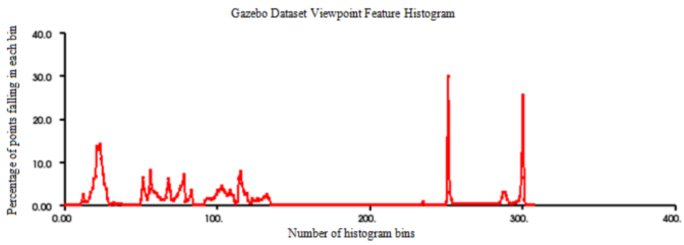
The features of the Gazebo dataset were extracted with VFH. An example feature histogram for chair, bookshelf and table objects is given in Figure 8. Then, all data obtained from the Gazebo simulation environment were classified via SVM. The training phase of the SVM was performed with the dataset prepared using the Blender. The classification rate of the each object for Gazebo dataset is given in Table 1.



(a) Chair



(b) Bookshelf



(c) Table

Fig. 8 Viewpoint Feature Histogram for chair, bookshelf and table objects obtained from Gazebo dataset

Table 1. Gazebo dataset test results

Objects	Classification Rate
Chair	97%
Bookshelf	88%
Table	83%

When the feature histograms of the training dataset (Figure 5) and Gazebo dataset (Figure 8) together with the results in Table 1 are analyzed, it is clear that the feature histograms of chair are almost the same in Figure 5 and Figure 8. For that reason, chair object has the highest classification rate in Table 1. However, in Figure 8, feature histograms for bookshelf and table objects have some similarities. Moreover, when the histograms for bookshelf and table objects in Figure 5 and Figure 8 are examined, there are some differences between them. Therefore, some bookshelf objects are confused with table objects and similarly some table objects are classified as bookshelf objects. As a result, the classification rate of the bookshelf and table objects remains at 88% and 83%, respectively.

C. Real-Environment Dataset

After promising results had been obtained from the Gazebo dataset, it was decided to test the method with a real-environment dataset. Then, a dataset for real-environment was constructed with the same objects in ESOGU Electrical Engineering Laboratory building. This dataset contains 20 data from different angles and positions for each object. Pre-processing steps were applied the dataset before the extraction of features. First, point cloud data were filtered for the outlier points to get only the related objects, which are intended to classify. For filtering process, PassThrough filter was used which is defined in PCL. Then, smoothing was performed to reduce the noise on the filtered point clouds. Figure 9 shows the RGB images (left) and resultant point cloud data (right) captured from the real environment.

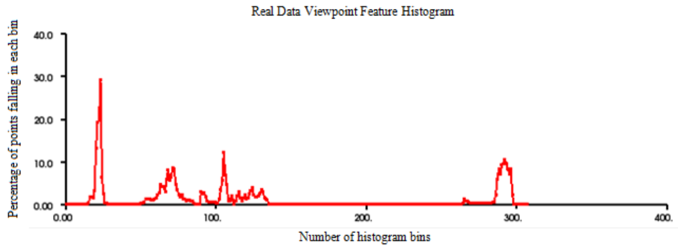
Fig. 9 RGB images (left) and point cloud data examples (right) for chair, bookshelf and table objects captured from real environment

The features of the real-environment dataset were extracted with VFH. An example feature histogram for chair, bookshelf and table objects is given in Figure 10. Then, all data obtained from the real-environment were classified via SVM. The training phase of the SVM was performed with the dataset prepared using the Blender. The classification rate of the each object for real-environment dataset is given in Table 2.

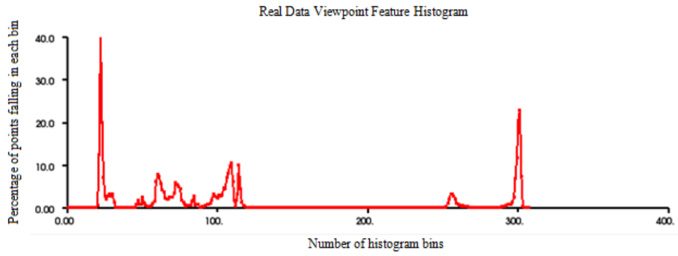
When the results in Table 2 are analyzed together with the feature histograms of training dataset (Figure 5) and feature histograms of real-environment dataset (Figure 10), it is seen that, the classification rate of chair object is higher than the bookshelf and table objects. Since the feature histogram of chair object in Figure 10 mostly preserves the characteristic of the histogram in Figure 5, the classification rate of chair object is an expected result. On the other hand, the feature histograms of bookshelf and especially table objects in Figure 5 and Figure 10 have significant differences. Moreover, when the feature histograms in Figure 10 are analyzed, the characteristic of the bookshelf and table object get lost and they carry similarities with chair object. For these reasons, some bookshelf and table are classified as chair object and the classification rate of the bookshelf and table objects remains at 68% and 60%, respectively.

Table 2. Real-environment dataset test results

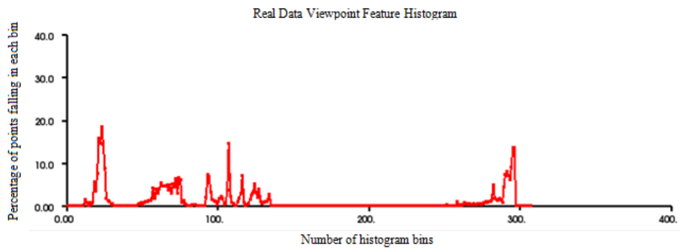
Objects	Classification Rate
Chair	%96
Bookshelf	%68
Table	%60



(a) Chair



(b) Bookshelf



(c) Table

Fig. 10 Viewpoint Feature Histogram for chair, bookshelf and table objects obtained from real-environment dataset

V. CONCLUSION

In this study, it is aimed to classify the chair, bookshelf and table objects, which are frequently used in the indoor environment, via point cloud data. In the implemented algorithm, VFH was employed to extract the features of the objects. The obtained features were classified using SVM.The main purpose of the study is to measure the efficiency of the implemented method on the simulated and real datasets when it is trained with synthetic dataset. To achieve this, the ModelNet40 were used as training dataset. Then, the experimental works were performed by using Gazebo and real-environment datasets. The test results showed that the implemented method produce promising classification rates for each object. Our future plan is to investigate the other global-based descriptors such CVFH and ESF in the same datasets.

ACKNOWLEDGMENT

This paper is supported by the Tubitak 2209-B (2241-A) Industrial Oriented License Graduate Thesis Support Program.

REFERENCES

[1] K. Alhamzi, M. Elmogy and S. Barakat, "3D Object Recognition Based on Local and Global Features Using Point Cloud Library", *International Journal of Advancements in Computing Technology*, vol. 7, no. 3, pp. 43-54, 2015.

[2] X.F. Han, J.X. Jin, J. Xie, M.J. Wang and W. Jiang, "A comprehensive review of 3D point cloud descriptors", *ArXiv*, 2018.

[3] S. H. Kasaei , A. M. Tomé , L. S. Lopes and M. Oliveira, "A Global Orthographic Object Descriptor for 3D Object Recognition and Manipulation", *Pattern Recognition Letters*, vol. 83, no. 3, pp. 312-320, 2016.

[4] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, "Learning Informative Point Classes for the Acquisition of Object Model Maps," *in Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 643-650, 2008.

[5] R.B. Rusu, N. Blodow and M.Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration", *IEEE Int. Conf Robot Automation (ICRA)*, pp. 3212 - 3217, 2009.

[6] R. B. Rusu, G. Bradski, R. Thibauxand and J. Hsu, "Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram",*2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2155-2162, 2010.

[7] A. Aldoma, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu and G. Bradski, "CAD-Model Recognition and 6DOF Pose Estimation Using 3D Cues", *IEEE International Conference on Computer Vision Workshops(ICCV Workshops)*, pp.585-592, 2011.

[8] Z. Fan, Z. Li and W. Li, "Object Detection and Sorting by Using a Global Texture-Shape 3D Feature Descriptor", *eprint arXiv:1802.01116*,2018arXiv180201116F, 2018.

[9] W. Wohlkingerve and M. Vincze, "Ensemble of shape functions for 3D object classification",*2011 IEEE International Conference on Robotics and Biomimetics*, Karon Beach, Phuket, pp. 2987-2992, 2011.

[10] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," *in IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1-4, 2011.

[11] L. A. Alexandre, "3D descriptors for object and category recognition: A comparative evaluation", *In Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

[12] F. Melgani and L. Bruzzone "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines" *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, vol.42, no. 8, 2004.

[13] B. Sch¨olkopf, K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio and V.Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers" *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 45, no. 11, 1997.

[14] K. Ramasubramanian, A. Singh, "Machine Learning Using R", *Apress*, no:1,2017

[15] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao "3D ShapeNets: A Deep Representation for Volumetric Shapes", *Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition(CVPR2015)* Available:http://modelnet.cs.princeton.edu/

[16] Google,Internet: https://www.blender.org/