Model-Based 3-D Object Recognition Methods

Ipsita Parida
20918570
iparida@uwaterloo.ca

Overview

In response to the desire to build systems with visual capabilities, as well as the practical need to assist a wide range of real-world problems like robot navigation, automated visual recognition, and so on, 3-D object recognition has become an active area of study. Sensors are used to gather precise surrounding information and the next process involves object detection, feature extraction, and object recognition. The standard 2-D image object identification cannot satisfy the practical application when the complexity of the object to be identified rises, but 3-D object recognition can characterize the shape and structure and improve the recognition rate. Among various methodologies, this tutorial aims to study model-based object recognition. Voxel-based methods and point-set based methods are referred to as model-based methods[1][2]. Model-based approach extracts a 3-D geometric feature description from the sensed data and compares it with the model description. In this tutorial, two such methods will be studied in detail.

Methodologies

Object recognition in 2-D and 3-D space has long been a topic of study in computer vision. Object recognition in the two-dimensional image domain has been investigated for decades. The most widely used method for identifying an object in an image is to calculate SIFT (Scale Invariant Feature Transform) characteristics[3] from the picture and compare them to those of a model object. Meanwhile in 3-D scenario, depending on the input types, 3-D object recognition systems are categorized into five groups based on their input modes[4]: geometric or model-based method, appearance or view-based method, feature matching-based method, depth image-based method, and intelligent algorithm-based approach. Geometric or model-based method requires a prior information about the attributes of a object such as shape and structure. It usually involves complex computations and geometrical modeling. Due to recent developments in 3-D depth cameras- Light detection and ranging (LiDAR) Cameras and RGB-D cameras, object recognition in the 3-D data domain is being studied widely.

In autonomous systems, fast and accurate detection of objects has been a huge challenge[5] as it involves dealing with large number of obstacles with different features. LiDAR sensors are used to sense accurate and appropriate environmental information. Large-scale 3-D point clouds (data points in the 3-D plane) get collected fast and with few mistakes using LiDAR sensors[6]. Before running an object recognition operation, it's important to get the features of the object right. For use cases as such, the ground points occupy a large portion of the scene in the point cloud. They also form a horizontal plane for the objects near the earth's surface which can be misjudged as connected. Hence it is important to filter out these ground point clouds to expedite the computing and recognition process. The object point cloud is then converted to a voxel model to get the geometries and attributes distribution features (Figure 1). In each voxel, multiple features such as point count, density, centroid, variance, covariance, eigenvector, eigenvalues, surface curvature, and divergence degree, can be calculated in parallel to form the feature vectors which act as object features datasets. The voxel counts for different objects can be different based on their sizes. A scale filter is used to transform feature vectors from uncertain count voxels to a normalized object feature matrix (3x3x3). For object identification, this data can be loaded into an initialized multilayer Neural Network model[7]. The feature extraction process can be sped up with the help of a graphics processing unit (GPU), which can aid in real-time object detection. Unmanned ground vehicles (UGVs) can benefit from this methodology[8] as the local course can be computed fast with accuracy.

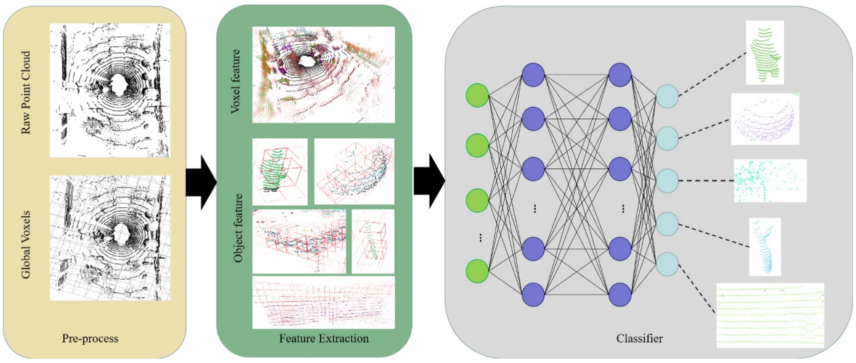

Figure 1: Flowchart of a object recognition system with multiple extracted features[8]

Another approach to 3-D object recognition, shown in Figure 2, computed a 3-D map, a point-cloud data into N planar patches by using a plane extraction technique built on the Random Sample Consensus (RANSAC) approach [9]. Each patch has some intrinsic properties (features) and Inter-Plane-Relationship (IPRs) (a geometric arrangement with another planer patch). The features, represented by the feature vector will be used to determine the object type for a patch. The features are divided into three categories: (i) Basic Features (BF): intrinsic attributes of a planer patch, Orientation, Area, and Height (OAH) (ii) Low-Level Features (LLF): defined all IPRs between two planar patches. 9 IPRs are computed- plane-distance, plane-angle, parallel-distance, projection-overlap-rate, projection-distance, is-parallel, is-perpendicular, is-coplanar, and is-adjacent (iii) High Level Features (HLF): a set of IPRs that is present in the object model. Each patch is assigned an HLF vector. This is constructed from the BFs *[O, A, H]* of each patch with some extra parameters by analyzing the LLF matrices of the planar patches. Next, the Gaussian-Mixture-Model-based Plane Classifier (GMM-PC) [10] does the plane classification on the HLF vector assignment. However, for n-type of objects, 'n' GMMs are assigned, one for each type. Each GMM is trained with a particular object in different scenes. This training helps in determining the configuration of a GMM. This is then used to classify each patch into a certain object type. A clustering procedure then clusters the classified plane into corresponding objects. Qian et al. [11] used this method to develop a robotic navigation aid. The object-level information could substantially aid a visually impaired person's ability to navigate in an indoor environment more successfully. The recognized objects can be viewed as reference points for navigation or obstacle while navigation. This can be extended to autonomous navigation and object manipulation.

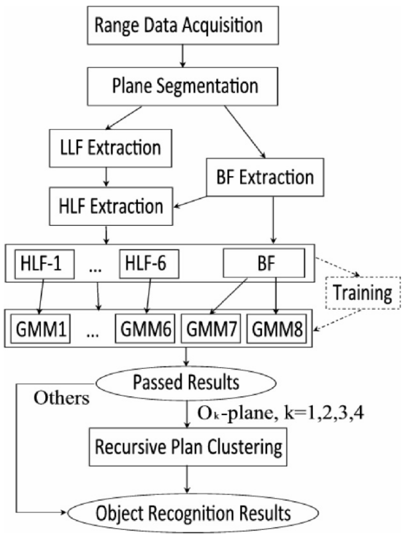

Figure 2: Diagram of a object recognition method with plane segmentation[9]

Conclusion

In the realm of computer vision, vision-based 3-D object recognition has always been a hotspot for research. Object recognition for uniform point cloud distribution or scenes with fewer items is now the most extensively utilized recognition method. The 3-D point cloud scene data is susceptible to noise, and the density distribution can be uneven in some cases. An important study path will be how to reduce point cloud noise, lessen the influence of uneven density distribution, and apply advance recognition technology to 3-D point cloud data.

References

[1] Z. Yang and L. Wang, "Learning relationships for multi-view 3d object recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7505–7514.

[2] S. Qi, X. Ning, G. Yang, L. Zhang, P. Long, W. Cai, and W. Li, "Review of multi-view 3d object recognition methods based on deep learning," *Displays*, vol. 69, p. 102053, 2021.

[3] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. Ieee, 1999, pp. 1150–1157.

[4] T. Dong, X. Qi, Q. Zhang, W. Li, and L. Xiong, "Overview on vision-based 3d object recognition methods," in *International Conference on Image and Graphics*. Springer, 2019, pp. 243–254.

[5] A. K. Jain and C. Dorai, "3d object recognition: Representation and matching," *Statistics and Computing*, vol. 10, no. 2, pp. 167–182, 2000.

[6] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.

[7] W. Song, S. Zou, Y. Tian, S. Fong, and K. Cho, "Classifying 3d objects in lidar point clouds with a back-propagation neural network," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, pp. 1–12, 2018.

[8] Y. Tian, W. Song, S. Sun, S. Fong, and S. Zou, "3d object recognition method with multiple feature extraction from lidar point clouds," *The Journal of Supercomputing*, vol. 75, no. 8, pp. 4430–4442, 2019.

[9] X. Qian and C. Ye, "Ncc-ransac: A fast plane extraction method for navigating a smart cane for the visually impaired," in *2013 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2013, pp. 261–267.

[10] T. M. Nguyen and Q. J. Wu, "Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 1, pp. 193–202, 2011.

[11] X. Qian and C. Ye, "3d object recognition by geometric context and gaussian-mixture-model-based plane classification," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 3910–3915.