A Survey on Training Free 3D Texture-less Object Recognition Techniques

1st Piyush Joshi* † ‡
School of Metallurgy and Materials
University of Birmingham
Birmingham B15 2TT United Kingdom
Email: p.joshi@bham.ac.uk

2nd Alireza Rastegarpanah* †
School of Metallurgy and Materials
University of Birmingham
Birmingham B15 2TT United Kingdom
Email: A.Rastegarpanah@bham.ac.uk

3rd Rustam Stolkin †
School of Metallurgy and Materials
University of Birmingham
Birmingham B15 2TT United Kingdom
Email: r.stolkin@bham.ac.uk

***Abstract*—Local surface feature based 3D object recognition is a rapidly growing research field. In time-critical applications such as robotics, training free recognition techniques are always the first choice as they are free from heavy statistical training. This paper presents an experimental analysis of 3D texture-less object recognition techniques that are free from any training. To our best knowledge, this is the first survey that includes experimental evaluation of top-rated training free recognition techniques on the datasets acquired by an RGBD camera. Based on the experimentation, we briefly present a discussion on potential future research directions.**

***Index Terms*—3D object recognition, Feature descriptors.**

I. Introduction

In the field of computer vision, object recognition is a fundamental research area that has many applications, such as intelligent surveillance, automatic assembly and dismantling, biometric analysis, robotics and medical treatment [1], [2], [3], [4]. For robust recognition, the researchers have focused on the development of 3D recognition techniques. The advantages of 3D object recognition over 2D object recognition made 3D recognition an active research topic [5]. Moreover, low-cost 3D acquisition systems (e.g., Intel Real Sense, Microsoft Kinect etc.) make 3D data more accessible [6], [7], [8].

In literature, 3D object recognition techniques can be divided into two broad categories of global feature based techniques and local feature based techniques [9], [10]. The global feature based techniques consider the object as a whole for the recognition. These techniques compute a set of global features that effectively represent the entire 3D object [11]. They have been widely used for 3D shape retrieval and classification [12]. Given that, these techniques ignore the details of the shape of the object and need a priori segmentation of the object [7]. Therefore, they are not suitable for the recognition of an object in occlusion and cluttered scenes [10]. On the other hand, the local feature based techniques compute features around specific keypoints from the neighborhood. These techniques are capable of handling occlusion and clutter better in comparison to the global feature based techniques [10].

Real-time object recognition is an important and challenging task in the application of robotics. This field has a strong need for computationally efficient techniques to recognize new objects without training. For such time-critical applications, real-time object recognition is an attractive solution because new objects can be easily added and matched online, in contrast to statistical-learning techniques that require many training samples and are often too computationally intensive for real-time performance [13], [14], [15], [16], [17].

Textured objects are often recognized based on their appearance by computing patch descriptor [18]. However, this appearance based recognition will not work for texture-less objects whose appearance is often dominated by their projected contours [19]. Recent advanced technological development in real-time range sensors enables us to obtain high-resolution depth images in real-time. These sensors are very useful, as they are small and light and provide accurate and dense depth measurements for near objects. In spite of popularity, there is still a need of experimentation on data acquired by an RGBD camera [20]. The detailed description of 3D recognition techniques can be accessed from another survey [20]. However, this survey has not included results on data acquired by the RGBD camera.

From the above discussion, this study aims to present an experimental survey on training free 3D texture-less object recognition techniques. Generally, data acquired by the RGBD camera consists of cluttered surfaces and occlusions. And most of the training free, texture-less object recognition techniques in the literature have not experimented on these cluttered RGBD data. Therefore, we analyze the performance of state-of-the-art techniques on such datasets. The rest of the paper is organized as follows. Section II presents related works on training free 3D recognition techniques. Experimental results are discussed in Section III. Future direction is discussed in Section IV. Finally, the paper is concluded in Section V.

II. Training free 3D recognition techniques

In this section, we present a brief survey on training free, local feature based 3D descriptors that can handle texture-less objects. These techniques are discussed as follows. A technique proposed in [21] called Point Feature Histograms (PFH) has encoded the local features of an object. This

* Joshi and Rastegarpanah are identified as joint lead authors of this work.
† The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot, UK.
‡Corresponding Author: Piyush Joshi

TABLE I: Description of considered datasets. #M. and #S. represent the number of models and scenes respectively.

Datasets	#M./#S.	RGBD Camera
Clutter [27]	18/30	Kinect
Challenge [7]	35/176	Kinect

technique has defined a Darboux frame for each pair of points in the object. Further, the technique has computed unique measures using angles between the points' normal and the distance vector between them. Finally, a histogram of 16-bin has been formed using these measures to define a feature descriptor of PFH. The computationally efficient version of PFH is proposed in [22]. The Fast Point Feature Histograms (FPFH) has computed the measures in the same way as PFH but only between the keypoint and its neighbors and this has reduced the time of computation of feature descriptor. The technique proposed in [23] has utilized the radial relationship of the point and its neighborhood to compute the Radius-based Surface Descriptor (RSD). This technique has computed the distance and the difference between normals for every pair of keypoint and a neighbor. The technique named Signature of Histograms of OrienTations (SHOT) is proposed in [24]. It has constructed a Local Reference Frame (LRF) for a keypoint and divided the neighborhood space into 3D spherical volumes. Further, this technique has computed a local histogram using the number of points concerning the angles between the normal at the keypoint and the neighboring points with in the volume. Finally, the technique has fused all local histograms to compute a 3D descriptor. The technique proposed in [25] has utilized the normal of a keypoint as the local reference axis and presented each neighboring point using two variables: signed distance (β) and radial distance (α). This technique has further divided $\alpha - \beta$ space into a 2D array and counted the number of points fell into bins. Finally, the 2D array has bilinearly interpolated to obtain a Spin descriptor. A robust technique has proposed in [26] that represented the LRF by estimating the covariance matrix of all points on the local surface. This technique has computed the descriptor named Rotational Projection Statistics (RoPS) for a keypoint by rotationally projecting the neighboring points onto 2D planes and then low-order central moments and entropy of these projected points have been computed to form the final descriptor.

III. Experimental Evaluation

We have experimented to evaluate the performance of five state-of-the-art 3D descriptors, including Spin Image [25], SHOT [24], RoPS [26], FPFH [22] and RSD [23]. The considered techniques have experimented on the Clutter [27] and Challenge [7] RGBD datasets acquired by an RGBD camera. The description of these datasets is presented in Table I. Clutter and Challenge datasets include RGBD data of typical household objects. The density of cluttered background and occlusions in these datasets is very high. Recognition rate (RR) is commonly used to evaluate the performance of recognition system [20], [28], [26], [29] and [30]. Therefore, we have used RR as our evaluation metric to validate the performance of each technique. RR defines the number of correct matching out of the total instances of a model in the scenes of a dataset. In order to compute RR, a pose is estimated between a model and matched corresponding points provided by a 3D descriptor. This estimated pose is further compared with the pose obtained from the ground truth of the considered dataset to verify the matching (true or false).

TABLE II: Recognition rate on Clutter and Challenge datasets

Techniques	SPIN	SHOT	RSD	RoPS	FPFH
Clutter: RR (%)	7.5	5.8	18.33	10.8	4.1
Challenge: RR (%)	1.70	0.6	5.06	0.4	0.4

A. *Performance on the Clutter and Challenge Datasets*

In real time applications, the heavy cluttered background and occlusions may present in scenes acquired by an RGBD camera. Therefore, there is a need to verify the performance of recognition techniques on very cluttered and occluded data. Clutter and Challenge datasets provide data with heavily cluttered background and high occlusions. For fair experimentation, we have not removed any background or plane surfaces from the scenes. We keep original data as it was in these datasets. Table II shows recognition rate of all techniques on these datasets. From the table, it is clear that the RSD technique has performed better than other techniques, however, it also failed to perform well. This is due to heavy cluttered background and occlusions present in both datasets.

IV. Future Direction

All considered techniques have created a relationship of a keypoint with only its neighbors to form a 3D feature descriptor [20]. That means all techniques have achieved the stability of the descriptor by focusing only on one location i.e. neighbors around the keypoints. The stability achieved from this one place may not be sufficient for the computation of a robust descriptor for a keypoint. Therefore, the development of a robust 3D descriptor using different regions of an object would be the future direction. From our experimentation, we found that the cluttered background is mainly responsible for the poor performance of state-of-the-art techniques. Therefore, automatic filtering of cluttered surfaces will enhance the performance of 3D recognition techniques.

V. Conclusion

We present a survey on training free 3D object recognition techniques that can handle texture-less objects. This paper has demonstrated the experimental results of top-rated techniques on RGBD datasets acquired by an RGBD camera. We have also discussed the reasons for their poor performance and future research directions.

Acknowledgments

This research was conducted as part of the project called "Reuse and Recycling of Lithium-Ion Batteries" (RELIB). This work was supported by the Faraday Institution [grant number FIRG005].

REFERENCES

[1] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," Pattern Recognition, vol. 47, pp. 509–524, 2014.

[2] Y. Guo, J. Wan, M. Lu, and W. Niu, "A parts-based method for articulated target recognition in laser radar data," Optik, vol. 124, no. 17, pp. 2727 – 2733, 2013.

[3] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in Proceedings of European Conference on Computer Vision, 2010, pp. 356–369.

[4] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes," International Journal of Computer Vision, vol. 89, pp. 348–361, 2010.

[5] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 998–1005.

[6] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in Proceedings of IEEE International Conference on Robotics and Automation, 2011, pp. 1–4.

[7] A. Aldoma, F. Tombari, L. Di Stefano, and M. Vincze, "A global hypotheses verification method for 3D object recognition," in Proceedings of European Conference on Computer Vision, 2012, pp. 511–524.

[8] M. Kaiser, X. Xu, B. Kwolek, S. Sural, and G. Rigoll, "Towards using covariance matrix pyramids as salient point descriptors in 3D point clouds," Neurocomputing, vol. 120, pp. 101 – 112, 2013.

[9] E. Akagündüz and I. Ulusoy, "3D object recognition from range images using transform invariant object representation," Electronics Letters, vol. 46, no. 22, pp. 1499–1500, 2010.

[10] N. Bayramoglu and A. A. Alatan, "Shape index SIFT: Range image recognition using local features," in Proceedings of International Conference on Pattern Recognition, 2010, pp. 352–355.

[11] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," Computer Graphics Forum, vol. 27, no. 2, pp. 643–652, 2008.

[12] B. Bustos, D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranić, "Feature-based similarity search in 3D object databases," ACM Computing Surveys, vol. 37, no. 4, p. 345–387, 2005.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 886–893 vol. 1.

[14] R. Fergus, P. Perona, and A. Zisserman, "Weakly supervised scale-invariant learning of models for visual recognition," International Journal of Computer Vision, vol. 71, no. 3, p. 273–303, 2007.

[15] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in Proceedings of IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.

[16] V. Ferrari, F. Jurie, and C. Schmid, "From images to shape models for object detection," International Journal of Computer Vision, vol. 87, no. 3, p. 284–303, 2010.

[17] M. Viola, M. J. Jones, and P. Viola, "Fast multi-view face detection," in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2003.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, p. 91–110, 2004.

[19] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, "Gradient response maps for real-time detection of textureless objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 5, pp. 876–888, 2012.

[20] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2270–2287, 2014.

[21] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, "Persistent point feature histograms for 3D point clouds," in Proceedings of the 10th International Conference on Intelligent Autonomous Systems, 2008.

[22] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in Proceedings of IEEE International Conference on Robotics and Automation, 2009, pp. 3212–3217.

[23] Z. Marton, D. Pangercic, N. Blodow, J. Kleinehellefort, and M. Beetz, "General 3D modelling of novel objects from a single view," in Proceedings of IEEE International Conference on Intelligent Robots and Systems, 2010, pp. 3700–3705.

[24] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in European Conference on Computer Vision, 2010, pp. 356–369.

[25] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 5, pp. 433–449, 1999.

[26] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," International journal of computer vision, vol. 105, no. 1, pp. 63–86, 2013.

[27] J. Glover and S. Popovic, "Bingham procrustean alignment for object detection in clutter," in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 2158–2165.

[28] A. S. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1584–1601, 2006.

[29] B. Taati and M. Greenspan, "Local shape descriptor selection for object recognition in range data," Computer Vision and Image Understanding, vol. 115, no. 5, pp. 681–694, 2011.

[30] P. Bariya, J. Novatnack, G. Schwartz, and K. Nishino, "3d geometric scale variability in range images: Features and descriptors," International Journal of Computer Vision, vol. 99, no. 2, pp. 232–255, 2012.