Computer Vision: A Review on 3D Object Recognition

Yeray Mezquita[1](✉), Alfonso González-Briones[1], Patricia Wolf[2], and Javier Prieto[1]

[1] BISITE Research Group, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo 2, 37007 Salamanca, Spain
yeraymm@usal.es, alfonsogb@usal.es, javierp@usal.es

[2] Department of Marketing & Management, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
pawo@sam.sdu.dk

Abstract. Three-dimensional (3D) object recognition is one of the fundamental tasks in computer vision applications, paving the way for all other image understanding operations. Although the potential of 3D object recognition is enormous, the information required for the spatial processing of the information means that the practical applications that end up being developed are very limited by the computational cost of the algorithms and frameworks used. This manuscript seeks to collect information on the most current review works in the literature. Thanks to this, researchers and developers who start working in the field of 3D object identification can find a compilation of the most important points to understand the current context in this field.

Keywords: Review · 3D object recognition · Computer vision

1 Introduction

Three-dimensional (3D) object recognition is one of the fundamental tasks in computer vision applications, paving the way for all other image understanding operations [9,13,19,25,26,36,41]. The demand for 3D object recognition is increasing due to the widespread of its applications in areas such as artificial intelligence robots, automated driving, medical image analysis, and virtual/augmented reality among other areas [3,18,20,21,23,24,32,35].

Although the potential of 3D object recognition is enormous, the information required for the spatial processing of the information means that the practical applications that end up being developed are very limited by the computational cost of the algorithms and frameworks used [40]. This is why over time more and more ways of working with three-dimensional image data have been developed to optimize the applications that require their use [29]. Besides, the need for better data bases that help in the creation of those frameworks and algorithms

J. M. Machado et al. (Eds.): DCAI 2022, LNNS 585, pp. 117–125, 2023.
https://doi.org/10.1007/978-3-031-23210-7_11

has driven the apparition of works like the one presented in [32], where the databases created are the main focus.

This manuscript seeks to collect information on the most current review works in the literature. Thanks to this, researchers and developers who start working in the field of 3D object identification can find a compilation of the most important points to understand the current context in this field: the algorithms and databases used to develop and evaluate them. Since this work is part of the "Technological Consortium TO develop sustainability of underwater Cultural heritage" project (TECTONIC), which makes use of 3D images of the seabed, we will also talk about related studies more specifically in this area.

The paper continues with an explanation of the 3D computer vision in Sect. 2. This is followed by the study model in the Sect. 3 followed by a discussion of the work found in the Sect. 4. Finally, the conclusions are detailed in the Sect. 5.

2 Background on 3D Object Identification

In this section it is going to be described a general background of the 3D computer vision field. 3D computer vision can be broadly defined as technologies that enable three-dimensional measurement or inspection of objects or surfaces in 3D. One of the advantages of 3D computer vision over 2D computer vision is that, in the case of occlusion, the views from different viewpoints can complement each other's detail features of the object and achieve excellent recognition performance. The data-sets used in 3D computer vision are created in different ways:

- Laser profiling: Laser profiling is one of the most popular 3D imaging techniques. The object being measured is moved through a laser beam as a camera positioned at a known angle records the changing profile of the laser as the object moves through it. This configuration is particularly popular on factory production floors or packaging lines as it relies on the movement of the object relative to the laser, meaning it is well suited to products on conveyor belts.
- Stereo imaging: Another popular 3D imaging technique is stereo imaging, in which two cameras are used to record 2D images of an object that can then be triangulated and converted into a 3D image. Like laser profiling, these techniques also allow for object movement during measurement and registration. The use of a random static illumination pattern can also give arbitrary texture to flat surfaces and objects that do not have natural edges, which many stereo reconstruction algorithms require.
- Fringe projection: In fringe projection, a fringe pattern is projected over the entire surface to be measured. The image is recorded by a camera positioned perpendicular to the object being measured. The point cloud created is capable of giving a height resolution up to two orders of magnitude higher than a laser profiling method is capable of providing. Fringe projection is also more scalable with a measurement area ranging from one millimeter to over one meter.

– Time of Flight: The time of flight method measures the time it takes for a pulse of light to reach the object being measured and then return. The time required to measure each point in the image will vary depending on the size and depth of the object and therefore each point will provide this information as they are measured.

There are different computer vision tasks, which can be classified as follows:

– Image classification. In this task, the image is assigned a label.
– Object localization. Once an object is located, a rectangle is drawn around it.
– Object detection. In this task the two previous tasks are combined, an object is located and a label is added to it.
– Instance segmentation. In this task, it is differentiated individuals of the same category.
– Semantic segmentation. Distinguishes high-level categories of significance, usually objects.
– Object recognition. Is a term that is used to referring all of the previous tasks together [30].

Model based methods can be divided in: i) voxel-based methods, where the objects can be represented as a 3D mesh, or ii) point-set-based methods, where a set of unordered points is used for prediction tasks [7,16,39,40].

3 Research Model

In this section, we will explain the development of the study model carried out in this work. To begin with, we want to answer the question: What has been published so far in the field of object identification in 3D images? Therefore, we have made use of the following macros: ("3D") AND ("object recognition") AND ("review").

Using the macros defined to search for academic articles in the ScienDirect database, we found 4764 results. Of which 3 have been used for this work:

– Review of multi-view 3D object recognition methods based on deep learning [29]. This manuscript presents an updated review and classification on deep learning methods for multi-view 3D object recognition. It also test these methods on mainstream datasets to provide some results and insights on the methods studied.
– Object recognition datasets and challenges: A review [32]. In this paper it is provided a detailed analysis of 160 datasets that has been widely used in the object recognition areas. It is also presented an overview of the object recognition benchmarks and the metrics adopted for evaluation purposes in the computer vision community.
– Review on deep learning techniques for marine object recognition: Architectures and algorithms [37]. The survey presented in this work is mainly focused in the marine object recognition methods based on deep learning for both,

surface and underwater targets. It is described typical deep network frameworks in three parts: image preprocessing, feature extraction, and recognition and model optimization.

4 Discussion

In this paper, we have surveyed some of the reviews in the literature on object recognition on 3D images. This work is necessary to find methods that have worked in this field and that can be used for the TECTONIC project, which will need to make use of these methods to label and identify objects in 3D images of the seafloor.

Based on what's said in [29], the use of deep learning techniques in multi-view 3D object recognition has become one of the most researched topics. It is because the deep learning techniques researched [2,6,17,22] can directly use the pretrained and successful advanced classification network as the backbone network, while thanks to the views obtained from multiple viewpoints it is possible to complement each one of the features of any object. There are still some challenges existing in this research topic, that's why many methods are being proposed in the literature to tackle them [8,15,28].

There are some works that propose the use of a mature classification network, such as VGG-M [4], VGG19 [33], GoogLeNet [34], AlexNet [14], ResNet-18/50 [10], and DenseNet [12]. The classification network is pretrained by the large-scale 2D datasets, to extract features at the view level, used then as the backbone network [38]. To provide accurate object classification, view-level features extracted need the use of feature fusion techniques. In the original work, the method used was a simple max-pooling of the features, but it ignored the relationship between the features. Other works of the literature proposed methods such as Recurrent Neural Networks (RNN) [1,27], Long Short-Term Memory (LSTM) [11], dynamic routing [31], Graph Convolutional Neural Network (GCN) [38]. In real-world applications, active camera viewpoints selection methods do the task of object recognition by optimizing the number of view inputs, and solving the occlusion problem, while reducing the cost of mobile robot [5].

From the study done in [29] it can be concluded that the existing 3D object recognition methods, see Table 1, have demonstrated their advantages although some aspects still need to be improved. E.g. passive selection methods are computationally intensive, limiting their performance in multi-view 3D object recognition. Another example is that the active views selection method requires all views to be inputted, from which arise practical problems such as occlusion. Because of these problems, advances in computer vision will require the automatic selection of the best viewing angle of the object [41]. A view to be selected as the best one should be the one with the most abundant image information and the highest distinguishability. This requirement will help the network to achieve the best recognition performance with as few views as possible while reducing the cost of the mobile robot and getting rid of the occlusion problem.

Table 1. Table summary of the techniques listed.

Type of method	Techniques
Convolutional neural networks	VGG-M [4] VGG19 [33] GoogLeNet [34] AlexNet [14] ResNet-18/50 [10] DenseNet [12]
Recurrent neural networks	Deep recurrent attention model (DRAM) [1] Visual attention model [27] Long short-term memory (LSTM) [11]
Capsules	Dynamic routing [31]
Graphs in CNN	View-graph convolutional neural network (GCN) [38]
Multi-view deep neural network	Veram [5]

In the manuscript [32] it's been shown the importance of the size and quality of datasets as the use of deep-learning techniques, which heavily rely on training data, spreads on the object recognition area. Also, the datasets are needed to provide a fair benchmarking mean for competitions, proving to be instrumental to the advancements in the field, by providing quantifiable benchmarks for the developed models. Also, it is being found the important of developing new and more challenging datasets, as the algorithms mature and the existing datasets become saturated. As a conclusion to the review, the authors state that researchers need to find the appropriate training and testing mediums for their desired applications.

In [37] authors exclusively focus on the deep-learning-based marine object recognition. According to this manuscript, marine object recognition propose various subproblems, such as the resolution (low, moderate and high) problem of images, samples starvation in image and video data, complex marine environmental factors, different degrees of model architectures, and optimizations based on the supervised and unsupervised learning models. Based on the literature review provided, and to provide guidelines for the researchers, they list the issues and challenges found as: i) the need to improve the public marine datasets, ii) necessity of pre-training, iii) the need of a unified framework, iiii) the fusion of multi-source features, v) fusion of multi-deep model, vi) sub-class recognition, and vii) the general model structure. Furthermore, it is proven that deep-learning methods, regardless of supervised or unsupervised, can be used in object recognition underwater and the surface.

5 Conclusion

In this article, we have conducted a study on the current context of computer vision for object recognition in three-dimensional environments. From the study, we found that deep learning techniques are the most widely used and investigated so far. They have proven to be effective and practical. Furthermore, we can conclude that the study, design, and development of new databases help enormously in this area since deep learning techniques are highly dependent on the data used and the current databases are starting to be saturated. Finally, we have studied the context in which computer vision is found in the maritime field. It can be said that object detection, both underwater and on the surface, is following the same development steps as in other fields, with deep learning techniques being the most widespread at present, which makes the development of new public databases for this field even more necessary.

The present study is far from perfect, it could be improved, not only by analyzing more works, but also to point out the advantages and disadvantages of each of the works in the literature. Besides, a comparison between the techniques, in different databases, applied to the maritime field could be a good way to differentiate the work while improving its scientific contribution.

Acknowledgements. The research of Yeray Mezquita is supported by the predoctoral fellowship from the University of Salamanca and co-funded by Banco Santander. This research was also partially supported by the project "Technological Consortium TO develop sustainability of underwater Cultural heritage (TECTONIC)", financed by the European Union (Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 873132). Authors declare no conflicts of interest.

References

1. Ba, J., Mnih, V., Kavukcuoglu, K.: Multiple object recognition with visual attention. arXiv preprint arXiv:1412.7755 (2014)
2. Casado-Vara, R., Novais, P., Gil, A.B., Prieto, J., Corchado, J.M.: Distributed continuous-time fault estimation control for multiple devices in IoT networks. IEEE Access **7**, 11972–11984 (2019)
3. Castellanos-Garzón, J.A., Mezquita Martín, Y., Jaimes, S.J.L., López, S.M.: A data mining approach applied to wireless sensor networks in greenhouses. In: International Symposium on Distributed Computing and Artificial Intelligence, pp. 431–436. Springer, Berlin (2018)
4. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. arXiv preprint arXiv:1405.3531 (2014)
5. Chen, S., Zheng, L., Zhang, Y., Sun, Z., Xu, K.: Veram: view-enhanced recurrent attention model for 3d shape classification. IEEE Trans. Vis. Comput. Graph. **25**(12), 3244–3257 (2018)
6. Fatima, N.: Enhancing performance of a deep neural network by comparing optimizers experimentally. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. **9**(2), 79–90 (2020). ISSN: 2255-2863

7. Feng, Y., Xiao, J., Zhuang, Y., Yang, X., Zhang, J.J., Song, R.: Exploiting temporal stability and low-rank structure for motion capture data refinement. Inf. Sci. **277**, 777–793 (2014)
8. González-Briones, A., Castellanos-Garzón, J.A., Mezquita Martín, Y., Prieto, J., Corchado, J.M.: A framework for knowledge discovery from wireless sensor networks in rural environments: a crop irrigation systems case study. Wirel. Commun. Mobile Comput. **2018** (2018)
9. Gupta, S., Meena, J., Gupta, O., et al.: Neural network based epileptic EEG detection and classification (2020)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
13. Hussain, A., Hussain, T., Ali, I., Khan, M.R., et al.: Impact of sparse and dense deployment of nodes under different propagation models in manets (2020)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (2012)
15. Li, T., Corchado, J.M., Sun, S.: Partial consensus and conservative fusion of gaussian mixtures for distributed PHD fusion. IEEE Trans. Aerosp. Electron. Syst. **55**(5), 2150–2163 (2018)
16. Liu, A.A., Zhou, H., Nie, W., Liu, Z., Liu, W., Xie, H., Mao, Z., Li, X., Song, D.: Hierarchical multi-view context modelling for 3d object classification and retrieval. Inf. Sci. **547**, 984–995 (2021)
17. López, A.B.: Deep learning in biometrics: a survey. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. **8**(4), 19–32 (2019)
18. Martín, Y.M., Parra, J., Pérez, E., Prieto, J., Corchado, J.M.: Blockchain-based systems in land registry, a survey of their use and economic implications. CISIS **2020**, 13–22 (2020)
19. Mezquita, Y., Alonso, R.S., Casado-Vara, R., Prieto, J., Corchado, J.M.: A review of k-nn algorithm based on classical and quantum machine learning. In: International Symposium on Distributed Computing and Artificial Intelligence, pp. 189–198. Springer, Berlin (2020)
20. Mezquita, Y., Casado-Vara, R., González Briones, A., Prieto, J., Corchado, J.M.: Blockchain-based architecture for the control of logistics activities: pharmaceutical utilities case study. Logic J. IGPL **29**(6), 974–985 (2021)
21. Mezquita, Y., Gil-González, A.B., Prieto, J., Corchado, J.M.: Cryptocurrencies and price prediction: a survey. In: International Congress on Blockchain and Applications, pp. 339–346. Springer, Berlin (2021)
22. Mezquita, Y., Gil-González, A.B., Martín del Rey, A., Prieto, J., Corchado, J.M.: Towards a blockchain-based peer-to-peer energy marketplace. Energies **15**(9), 3046 (2022)
23. Mezquita, Y., González-Briones, A., Casado-Vara, R., Chamoso, P., Prieto, J., Corchado, J.M.: Blockchain-based architecture: a mas proposal for efficient agri-food supply chains. In: International Symposium on Ambient Intelligence, pp. 89–96. Springer, Berlin (2019)

24. Mezquita, Y., González-Briones, A., Casado-Vara, R., Wolf, P., de la Prieta, F., Gil-González, A.B.: Review of privacy preservation with blockchain technology in the context of smart cities. In: Sustainable Smart Cities and Territories International Conference, pp. 68–77. Springer, Berlin (2021)
25. Mezquita, Y., Parra, J., Perez, E., Prieto, J., Corchado, J.M.: Blockchain-based systems in land registry, a survey of their use and economic implications. In: Computational Intelligence in Security for Information Systems Conference, pp. 13–22. Springer, Berlin (2019)
26. Mezquita, Y., Valdeolmillos, D., González-Briones, A., Prieto, J., Corchado, J.M.: Legal aspects and emerging risks in the use of smart contracts based on blockchain. In: International Conference on Knowledge Management in Organizations, pp. 525–535. Springer, Berlin (2019)
27. Mnih, V., Heess, N., Graves, A., et al.: Recurrent models of visual attention. In: Advances in Neural Information Processing Systems, vol. 27 (2014)
28. Pimpalkar, A.P., Raj, R.J.R.: Influence of pre-processing strategies on the performance of ml classifiers exploiting TF-IDF and bow features. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. **9**(2), 49 (2020)
29. Qi, S., Ning, X., Yang, G., Zhang, L., Long, P., Cai, W., Li, W.: Review of multi-view 3d object recognition methods based on deep learning. Displays **69**, 102053 (2021)
30. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
31. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
32. Salari, A., Djavadifar, A., Liu, X.R., Najjaran, H.: Object recognition datasets and challenges: a review. Neurocomputing (2022)
33. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
34. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
35. Valdeolmillos, D., Mezquita, Y., González-Briones, A., Prieto, J., Corchado, J.M.: Blockchain technology: a review of the current challenges of cryptocurrency. In: International Congress on Blockchain and Applications, pp. 153–160. Springer, Berlin (2019)
36. Vergara, D., Extremera, J., Rubio, M.P., Dávila, L.P.: The proliferation of virtual laboratories in educational fields. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. **9**(1), 85 (2020)
37. Wang, N., Wang, Y., Er, M.J.: Review on deep learning techniques for marine object recognition: architectures and algorithms. Control Eng. Pract. **118**, 104458 (2022)
38. Wei, X., Yu, R., Sun, J.: View-GCN: view-based graph convolutional network for 3d shape analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1850–1859 (2020)
39. Yang, Z., Wang, L.: Learning relationships for multi-view 3d object recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7505–7514 (2019)

40. Zhou, H.Y., Liu, A.A., Nie, W.Z., Nie, J.: Multi-view saliency guided deep neural network for 3-d object retrieval and classification. IEEE Trans. Multimedia **22**(6), 1496–1506 (2019)
41. Zubair, S., Al Sabri, M.A.: Hybrid measurement of the similarity value based on a genetic algorithm to improve prediction in a collaborative filtering recommendation system. ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. **10**(2), 165–182 (2021)