3D Object Detection Methods

Subjects: Robotics

Contributor: Ahmed El-Dawy, Amr El-Zawawi, Mohamed El-Habrouk

Effective environmental perception is critical for autonomous driving; thus, the perception system requires collecting 3D information of the surrounding objects, such as their dimensions, locations, and orientation in space.

Keywords: 3D object detection ; autonomous driving ; robotics

1. Introduction

The development of driving assistance systems holds the possibility of reducing accidents, reducing environmental emissions, and easing the stress associated with driving [1][2][3][4]. Several levels of automation have been proposed, based on their technology capacities and human interaction [5][6]. The most widely known levels can be broken down into six groups [7]. Beginning with Level 0 (Driver-Only Level), the complete control of the vehicle, including steering, braking, accelerating, and decelerating, is completely under the control of the driver [8]. As the level increases from Level 1 to Level 3, the user interaction is reduced and the level of automation is increased [8]. In contrast to earlier levels of autonomous driving, Level 4 (High Driving Automation) and Level 5 (Full Driving Automation) attain fully autonomous driving, where the vehicle can be operated without the need for any driving experience or even a driving licence [9]. The difference between Level 4 and Level 5 is that autonomous vehicles categorized in Level 5 can drive entirely automatically in all driving domains and require no human input or interaction. Thus, Level 5 prototypes remove the steering wheel and the pedals; hence, the role of the driver is diminished to that of a mere passenger [10]. Consequently, there is a constraint on any vehicle equipped with a driving assistance system in order to evolve into a practical reality. This vehicle must be equipped with a perception system that enables high levels of awareness and intelligence necessary to deal with stressful real-world situations, make wise choices, and always act in a manner that is secure and accountable [5][11][12].

The perception system of autonomous cars performs a variety of tasks. The ultimate purpose of these tasks is to fully understand the vehicle's immediate surroundings with low latency so that the vehicle can make decisions or interact with the real world. Object detection, tracking, and driving area analysis are examples of such activities. 3D object properties detection is a significant area of study in the autonomous vehicles perception system [13]. Current 3D object detection techniques can be primarily classified into Lidar-based or vision-based techniques [14]. Lidar-based 3D object detection techniques are accurate and efficient but expensive, which limits their use in industry [15]. On the other hand, vision-based techniques [16] can also be categorized into two distinct groups: monocular and binocular vision. Vision-based perception systems are widely used due to their low cost and rich features (critical semantic elements within the image that carry valuable information and are used to understand and process the visual data). The most significant downside of monocular vision is that it cannot determine depth directly from image information, which may cause errors in 3D pose estimation in monocular object detection. The cost of binocular vision is higher than that of monocular vision, although it can provide more accurate depth information than monocular vision. Moreover, binocular vision yields a narrower visual field range, which cannot meet the requirements of certain operating conditions [17].

For the case of monocular vision systems, the camera projects a 3D point (defined in the 3D world's coordinate frame of the object) into the 2D image coordinate frame. This is a forward mathematical operation, which removes the projected object's depth information. Thus, the inverse projection of the point from the 2D image coordinate frame back into the 3D world's coordinate frame of the object is not a trivial mathematical task [18].

2. 3D Object Detection Methods

The various methods employed for 3D object detection can be classified into two distinct categories: conventional techniques or deep learning techniques. This section provides a comprehensive summary of the present 3D object detection methodologies as summarized in **Figure 1**. The 3D-data-based techniques in both categories achieve superior detection results over 2D-data-based techniques. However, they come with additional cost related to the sensor setup, as

mentioned earlier. On the other hand, 2D-data-based techniques are cost efficient, which comes with sacrificing the depth estimation accuracy [15][17].

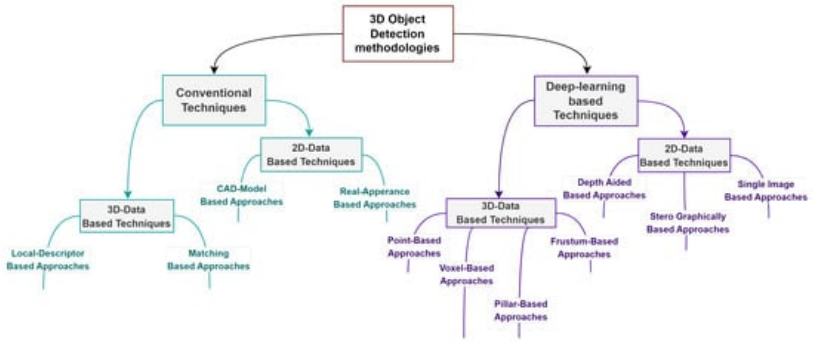

Figure 1. 3D object detection methodologies.

2.1. Conventional Techniques

Conventional techniques depend heavily on a prior knowledge of the object appearance to retrieve the essential suite of object features using hand-crafted feature descriptors, followed by the formation of a database by appropriately attributing features with the 3D models [19][20][21]. In order to provide quick detection results, the model database is organized and searched using effective indexing algorithms such as hierarchical RANSAC search, and voting is performed by using ranked criteria [22]. Other approaches depend on different hashing techniques like geometric hashing [19] and elastic hash table [23]. The conventional techniques can be categorized according to the input information into 3D-data-based techniques and 2D-data-based techniques [24].

2.1.1. 3D-Data-Based Techniques

With the advancement of sensor technology, more devices capable of capturing 3D environmental data, such as depth cameras and 3D scanners [25][26], are developed. When compared to 2D-data techniques, 3D-data techniques retain the object's genuine physical characteristics, which make them a superior choice to quantify the 6D pose [27]. The two primary categories of conventional 3D-data-based approaches are as follows.

Local-Descriptor-Based Approaches: These depend on the local descriptor and utilize an offline global descriptor applied to the model. This global descriptor needs to be rotation and translation invariant. Then, the local descriptor is subsequently generated online and matched with the global descriptor. Super Key 4-Points Congruent Sets (SK-4PCS) [28] can be paired with invariant local properties of 3D shapes to optimize the quantity of data handled. The iterative Closest Point (ICP) [29] approach is a traditional method that can compute the pose transformation between two sets of point clouds. Another approach [30] establishes a correspondence between the online local descriptor and the saved model database by employing the enhanced Oriented Fast and Rotated BRIEF (ORB) [31] feature and the RBRIEF descriptor [32]. A different approach centered around semi-global descriptors [33] can be used to assess the 6D pose of large-scale occluded objects.

Matching-Based Approaches: The objective of matching-based approaches is to find the template in the database that is almost identical to the input sample as well as retrieving its 6D pose. An innovative Multi-Task Template Matching (MTTM) framework was proposed in [34] to increase matching reliability. It locates the closest template of a target object while estimating the masks of segmentation and the object's pose transformation. In order to enhance the data storage footprint and the lookup searching time, fused Balanced Pose Tree (BPT) and PCOF-MOD (multimodal PCOF) under optimum storage space restructuring was proposed in [35] in order to yield memory-efficient 6D pose estimation. A study on the control of micro-electro-mechanical system (MEMS) microassembly was conducted in [36] to allow for an accurate control to enable minimizing 3D residuals.

2.1.2. 2D-Data-Based Techniques

To measure the 6D pose of objects, a variety of features that are represented in the 2D image can be used. These features may include texture features, geometric features, color features, and more [24][37][38]. Speeded Up Robust Features (SURF) [39] and Scale-Invariant Feature Transform (SIFT) features [40] are considered the primary features that can be utilized for object pose estimation. Color features [41], as well as geometric features [42], can be utilized to enhance the pose estimation performance. The conventional 2D-data-based techniques can be further broken down according to the used matching template into CAD-model-based approaches and real-appearance-based approaches.

CAD-Model-Based Approaches: CAD-modelbbased approaches depend mainly on the rendered templates of CAD models. These approaches are suitable for industrial applications, because illumination and blurring have no effect on the rendering process [24]. A Perspective Cumulated Orientation Feature (PCOF) based on orientation histograms was proposed in [43] to estimate a robust object's pose. The Fine Pose parts-based Model (FPM) [44] was implemented to localize objects in a 2D image using the given CAD models. Moreover, the edge correspondences can be used to estimate the pose [45]. The multi-view geometry can be adapted in order to extract the object's geometric features [46], while the epipolar geometry can be used to generate the transformation matrix [47]. Some other approaches suggest visuals-based tracking like [48], which performs tracking approach based on CAD model for micro-assembly and like [49] which employs 6D posture estimation for end-effector tracking in a scanning electron microscope to enable higher-quality automated processes and accurate measurements.

Real-Appearance-Based Approaches: Despite the accuracy that can be achieved by employing 3D CAD models, reliable 3D CAD models are not always obtained. Under these conditions, real object images are used as the template. The 6D pose of an object can be measured depending on multi-cooperative logos [50]. The Histogram Of Gradients (HOG) [51] is an effective technique for improving the pose estimation performance. Another approach [52] suggested template matching in order to provide stability against small image variations. As obtaining sufficient real image templates is laborious and time-consuming, and producing images via CAD model is becoming simpler, approaches that depend on CAD models are much more popular rather than approaches that depend on real appearance [24].

The main downside of the conventional techniques is that they are very sensitive to noise and are not efficient in terms of computational burden and storage resources [53]. The performance of the 2D-data-based techniques heavily depends on the total number of templates, which impacts the pose estimation accuracy, as these approaches depend on the object appearance experience. The greater the number of templates, the more accurate the posture assessment [54][55]. Consequently, a large number of templates necessitates a significant amount of storage and search time [56]. Also, it is not practical to obtain 360° features of the model [57][58]. Because of how the 3D-data-based approaches function, the major limitation for these approaches is that they fail to function adequately whenever the object has a high level of reflected brightness [24]. Another problem is that the efficiency of these approaches is somewhat limited, as point clouds and depth images involve an enormous quantity of data, leading to a large computational burden [24].

2.2. Deep Learning-Based Techniques

Convolution Neural Networks (CNNs) are commonly used in 3D object detection methodologies based on deep learning strategies to retrieve a hierarchical suite of abstracted features from each object in order to record the object's essential information [59]. Unlike conventional 3D object detection techniques, which depend on hand-crafted feature descriptors, deep learning-based techniques utilize learnable weights that can be automatically tuned during the training phase [60]. Thus, these techniques provide more robustness against environmental variations. The deep learning techniques can also be categorized into 3D- and 2D-data-based techniques, as shown in **Figure 1**.

2.2.1. 3D-Data-Based Techniques

In recent years, LiDAR-based 3D detection [61][62][63][64] has advanced rapidly. LiDAR sensors collect accurate 3D measurement data gathered from the surroundings in the pattern of 3D points (x, y, z), at which x, y, z are each point's absolute 3D coordinates. LiDAR point cloud representations, by definition, necessitate an architecture that allows convolution procedures to be performed efficiently. Thus, deep learning-based techniques, which depend on LiDAR 3D detection, can be divided into Voxel-based approaches, Point-based approaches, Pillar-based approaches, and Frustum-based approaches [65][66].

Voxel-Based Approaches: These partition point clouds into similarly sized 3D voxels. Following that, for every single voxel, feature extraction can be used to acquire features from a group of points. The aforementioned approach minimizes the overall size of the point cloud, conserving storage space [66]. Voxel-based approaches, such as VoxelNet [67] and SECOND [68], augment the 2D image characterization into 3D space by splitting the 3D space into voxels. Other deep learning models like Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection [69], Afdet [70], Center-based 3D object detection and tracking [71], and Psanet [72] utilize the Voxel-based approach to perform 3D object detection.

Point-Based Approaches: These were introduced to deal with raw unstructured point clouds. Point-based approaches, such as PointNet [73] and PointNet++ [74] issue raw point clouds as input and retrieve point-wise characteristics for 3D object detection using formations such as multi-layer perceptrons. Other models like PointRCNN [64], PointRGCN [75],

RoarNet [76], LaserNet [77], and PointPaiting [78] are examples of deep learning neural networks that adopt a Point-Based approach.

Pillar-Based Approaches: These structure the LiDAR 3D point cloud into vertical columns termed as pillars. Utilizing pillar-based approaches enables the tuning of the 3D point cloud arranging process in the x–y plane, removing the z coordinate, as demonstrated in PointPillars [79].

Frustum-Based Approaches: These partition the LiDAR 3D point clouds into frustums. The 3D object detector models that crop point cloud regions recognized by an RGB image object detector include SIFRNet [80], Frustum ConvNet [81], and Frustum PointNet [82].

2.2.2. 2D-Data-Based Techniques

RGB images are the main input for deep learning-based techniques that rely on 2D data. Despite the outstanding performance of 2D object detection networks, generating 3D bounding boxes merely from the 2D image plane is a substantially more complicated challenge owing to the lack of absolute depth information [83]. Thus, the 2D-data-based techniques for 3D object detection can be categorized according to the adopted method to generate the depth information into Stereographically Based Approaches, Depth-Aided Approaches, and Single-Image-Based Approaches.

Stereographically Based Approaches: Strategies in [84][85][86][87][88] process the pair of stereo images using a Siamese network and create a 3D cost volume to determine the matching cost for stereo matching using neural networks. The preliminary work 3DOP [89] creates 3D propositions by tinkering with a wide range of extracted features such as stereo reconstruction, and object size previous convictions. MVS-Machine [90] considers differentiable projection and reprojection for improved 3D volume construction handling from multi-view images.

Depth-Aided Approaches: Due to the missing depth information in single monocular image input, several researchers tried to make use of the progress in depth estimation neural networks. Previous research works [91][92][93] convert images into pseudo-LiDAR perceptions by harnessing off-the-shelf depth map predictors and calibration parameters. Then, they deploy established LiDAR-based 3D detection methods to output 3D bounding boxes, leading to lesser effectiveness. D4LCN [94] and DDMP-3D [95] emphasize a fusion-based strategy between image and estimated depth using cleverly engineered deep CNNs. However, most of the aforementioned methods that utilize off-the-shelf depth estimators directly pay substantial computational costs and achieve only limited improvement due to inaccuracy in the estimated depth map [96].

Single-Image-Based Approaches: Recently, several research works [13][97][98][99][100] employed only a monocular RGB image as input to 3D object detection. PGD-FCOS3D [101] establishes geometric correlation graphs between detected objects then utilizes the constructed graphs to improve the accuracy of the depth estimation. Some research works like RTM3D [102], SMOKE [103], and KM3D-Net [104] anticipate key points of the 3D bounding box as an adjacent procedure for establishing spatial information of the observed object. Decoupled-3D [105] presents an innovative framework for the decomposition of the detection problem into two tasks: a structured polygon prediction task and a depth recovery task. QD-3DT [106] presents a system for tracking moving objects over time and estimating their full 3D bounding box from an ongoing series of 2D monocular images. The association stage of the tracked objects depends on quasi-dense similarity learning to identify objects of interest in different positions and views based on visual features. MonoCInIS [107] presents a method that utilizes instance segmentation in order to estimate the object's pose. The proposed method is camera-independent in order to account for variable camera perspectives. MonoPair [108] investigates spatial pair-wise interactions among objects to enhance detection capability. Many recent research works depend on a prior 2D object detection stage. Deep3Dbox [109] suggests an innovative way to predict orientation and dimensions. M3D-RPN [110] considers a depth-aware convolution to anticipate 3D objects and produces 3D object properties with 2D detection requirements. GS3D [111] extends Deep3Dbox with a feature extraction module for visible surfaces. Due to the total loss of depth information and the necessity for a vast search space, it is not trivial to estimate the object's spatial position immediately [17]. As a result, PoseCNN [112] recognizes an object's position in a 2D image while simultaneously predicting its depth to determine its 3D position. Because rotation space is nonlinear, it is challenging to estimate 3D rotation directly with PoseCNN [17].

Although deep learning methods that rely on a single RGB image provide some appropriate projection-based constraints, they fail to achieve promising results due to an absence of depth spatial data [91][113].

References

1. Crayton, T.J.; Meier, B.M. Autonomous vehicles: Developing a public health research agenda to frame the future of transportation policy. J. Transp. Health 2017, 6, 245–252.
2. Yurtsever, E.; Lambert, J.; Carballo, A.; Takeda, K. A Survey of Autonomous Driving: Common Practices and Emerging Technologies. IEEE Access 2020, 8, 58443–58469.
3. Shladover, S.E. Review of the state of development of advanced vehicle control systems (AVCS). Veh. Syst. Dyn. 1995, 24, 551–595.
4. Vander Werf, J.; Shladover, S.E.; Miller, M.A.; Kourjanskaia, N. Effects of adaptive cruise control systems on highway traffic flow capacity. Transp. Res. Rec. 2002, 1800, 78–84.
5. Calvert, S.; Schakel, W.; Van Lint, J. Will automated vehicles negatively impact traffic flow? J. Adv. Transp. 2017, 2017, 3082781.
6. Gasser, T.M.; Westhoff, D. BASt-study: Definitions of automation and legal issues in Germany. In Proceedings of the 2012 Road Vehicle Automation Workshop, Irvine, CA, USA, 25 July 2012.
7. International, S. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE Int. 2018, 4970, 1–5.
8. Varotto, S.F.; Hoogendoorn, R.G.; van Arem, B.; Hoogendoorn, S.P. Empirical longitudinal driving behavior in authority transitions between adaptive cruise control and manual driving. Transp. Res. Rec. 2015, 2489, 105–114.
9. Nassi, D.; Ben-Netanel, R.; Elovici, Y.; Nassi, B. MobilBye: Attacking ADAS with camera spoofing. arXiv 2019, arXiv:1906.09765.
10. Vivek, K.; Sheta, M.A.; Gumtapure, V. A comparative study of Stanley, LQR and MPC controllers for path tracking application (ADAS/AD). In Proceedings of the 2019 IEEE International Conference on Intelligent Systems and Green Technology (ICISGT), Visakhapatnam, India, 29–30 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 67–674.
11. Gupta, A.; Anpalagan, A.; Guan, L.; Khwaja, A.S. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. Array 2021, 10, 100057.
12. Sharma, D. Evaluation and Analysis of Perception Systems for Autonomous Driving. 2020. Available online: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1536525&dswid=-9079 (accessed on 5 November 2023).
13. Liu, Y.; Yixuan, Y.; Liu, M. Ground-aware monocular 3D object detection for autonomous driving. IEEE Robot. Autom. Lett. 2021, 6, 919–926.
14. Li, Z.; Du, Y.; Zhu, M.; Zhou, S.; Zhang, L. A survey of 3D object detection algorithms for intelligent vehicles development. Artif. Life Robot. 2022, 27, 115–122.
15. Wu, Y.; Wang, Y.; Zhang, S.; Ogai, H. Deep 3D object detection networks using LiDAR data: A review. IEEE Sens. J. 2020, 21, 1152–1171.
16. Qian, R.; Lai, X.; Li, X. 3D object detection for autonomous driving: A survey. Pattern Recognit. 2022, 130, 108796.
17. Wu, J.; Yin, D.; Chen, J.; Wu, Y.; Si, H.; Lin, K. A survey on monocular 3D object detection algorithms based on deep learning. J. Phys. Conf. Ser. 2020, 1518, 012049.
18. Gu, F.; Zhao, H.; Ma, Y.; Bu, P. Camera calibration based on the back projection process. Meas. Sci. Technol. 2015, 26, 125004.
19. Lamdan, Y.; Schwartz, J.T.; Wolfson, H.J. Affine invariant model-based object recognition. IEEE Trans. Robot. Autom. 1990, 6, 578–589.
20. Rigoutsos, I.; Hummel, R. Implementation of geometric hashing on the connection machine. In Proceedings of the Workshop on Directions in Automated CAD-Based Vision, Maui, HI, USA, 2–3 June 1991; IEEE Computer Society: Piscataway, NJ, USA, 1991; pp. 76–77.
21. Rigoutsos, I. Massively Parallel Bayesian Object Recognition; New York University: New York, NY, USA, 1992.
22. Biegelbauer, G.; Vincze, M.; Wohlkinger, W. Model-based 3D object detection: Efficient approach using superquadrics. Mach. Vis. Appl. 2010, 21, 497–516.
23. Bebis, G.; Georgiopoulos, M.; da Vitoria Lobo, N. Learning geometric hashing functions for model-based object recognition. In Proceedings of the IEEE International Conference on Computer Vision, Cambridge, MA, USA, 20–23 June 1995; IEEE: Piscataway, NJ, USA, 1995; pp. 543–548.
24. He, Z.; Feng, W.; Zhao, X.; Lv, Y. 6D pose estimation of objects: Recent technologies and challenges. Appl. Sci. 2020, 11, 228.

25. Wang, K.; Xie, J.; Zhang, G.; Liu, L.; Yang, J. Sequential 3D human pose and shape estimation from point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 7275–7284.
26. Li, X.; Wang, H.; Yi, L.; Guibas, L.J.; Abbott, A.L.; Song, S. Category-level articulated object pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 April 2020; pp. 3706–3715.
27. Zhang, Z.; Hu, L.; Deng, X.; Xia, S. Weakly supervised adversarial learning for 3D human pose estimation from point clouds. IEEE Trans. Vis. Comput. Graph. 2020, 26, 1851–1859.
28. Guo, Z.; Chai, Z.; Liu, C.; Xiong, Z. A fast global method combined with local features for 6d object pose estimation. In Proceedings of the 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Hong Kong, China, 8–12 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
29. Chen, Y.; Medioni, G. Object modelling by registration of multiple range images. Image Vis. Comput. 1992, 10, 145–155.
30. Yu, H.; Fu, Q.; Yang, Z.; Tan, L.; Sun, W.; Sun, M. Robust robot pose estimation for challenging scenes with an RGB-D camera. IEEE Sens. J. 2018, 19, 2217–2229.
31. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2564–2571.
32. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. Brief: Binary robust independent elementary features. In Proceedings of the Computer Vision—ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Proceedings, Part IV 11. Springer: Berlin/Heidelberg, Germany, 2010; pp. 778–792.
33. Nospes, D.; Safronov, K.; Gillet, S.; Brillowski, K.; Zimmermann, U.E. Recognition and 6D pose estimation of large-scale objects using 3D semi-global descriptors. In Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 27–31 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
34. Park, K.; Patten, T.; Prankl, J.; Vincze, M. Multi-task template matching for object detection, segmentation and pose estimation using depth images. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 7207–7213.
35. Konishi, Y.; Hattori, K.; Hashimoto, M. Real-time 6D object pose estimation on CPU. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3451–3458.
36. Tamadazte, B.; Marchand, E.; Dembélé, S.; Le Fort-Piat, N. CAD model-based tracking and 3D visual-based control for MEMS microassembly. Int. J. Robot. Res. 2010, 29, 1416–1434.
37. Brachmann, E.; Michel, F.; Krull, A.; Yang, M.Y.; Gumhold, S.; Rother, C. Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3364–3372.
38. Marullo, G.; Tanzi, L.; Piazzolla, P.; Vezzetti, E. 6D object position estimation from 2D images: A literature review. Multimed. Tools Appl. 2022, 82, 24605–24643.
39. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 2008, 110, 346–359.
40. Lowe, D.G. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 2004, 60, 91–110.
41. Miyake, E.; Takubo, T.; Ueno, A. 3D Pose Estimation for the Object with Knowing Color Symbol by Using Correspondence Grouping Algorithm. In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 960–965.
42. Zhang, X.; Jiang, Z.; Zhang, H.; Wei, Q. Vision-Based Pose Estimation for Textureless Space Objects by Contour Points Matching. IEEE Trans. Aerosp. Electron. Syst. 2018, 54, 2342–2355.
43. Konishi, Y.; Hanzawa, Y.; Kawade, M.; Hashimoto, M. Fast 6D pose estimation from a monocular image using hierarchical pose trees. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 398–413.
44. Lim, J.J.; Khosla, A.; Torralba, A. Fpm: Fine pose parts-based model with 3D cad models. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part VI 13. Springer: Berlin/Heidelberg, Germany, 2014; pp. 478–493.

45. Muñoz, E.; Konishi, Y.; Murino, V.; Del Bue, A. Fast 6D pose estimation for texture-less objects from a single RGB image. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5623–5630.
46. Peng, J.; Xu, W.; Liang, B.; Wu, A.G. Virtual stereovision pose measurement of noncooperative space targets for a dual-arm space robot. IEEE Trans. Instrum. Meas. 2019, 69, 76–88.
47. Chaumette, F.; Hutchinson, S. Visual servo control. II. Advanced approaches . IEEE Robot. Autom. Mag. 2007, 14, 109–118.
48. Wnuk, M.; Pott, A.; Xu, W.; Lechler, A.; Verl, A. Concept for a simulation-based approach towards automated handling of deformable objects—A bin picking scenario. In Proceedings of the 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Auckland, New Zealand, 21–23 November 2017; pp. 1–6.
49. Kratochvil, B.E.; Dong, L.; Nelson, B.J. Real-time rigid-body visual tracking in a scanning electron microscope. Int. J. Robot. Res. 2009, 28, 498–511.
50. Guo, J.; Wu, P.; Wang, W. A precision pose measurement technique based on multi-cooperative logo. J. Phys. Conf. Ser. 2020, 1607, 012047.
51. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 886–893.
52. Hinterstoisser, S.; Cagniart, C.; Ilic, S.; Sturm, P.; Navab, N.; Fua, P.; Lepetit, V. Gradient response maps for real-time detection of textureless objects. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 876–888.
53. Solina, F.; Bajcsy, R. Recovery of parametric models from range images: The case for superquadrics with global deformations. IEEE Trans. Pattern Anal. Mach. Intell. 1990, 12, 131–147.
54. Roomi, M.; Beham, D. A Review Of Face Recognition Methods. Int. J. Pattern Recognit. Artif. Intell. 2013, 27, 1356005.
55. Vishwakarma, V.P.; Pandey, S.; Gupta, M. An illumination invariant accurate face recognition with down scaling of DCT coefficients. J. Comput. Inf. Technol. 2010, 18, 53–67.
56. Muñoz, E.; Konishi, Y.; Beltran, C.; Murino, V.; Del Bue, A. Fast 6D pose from a single RGB image using Cascaded Forests Templates. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 4062–4069.
57. Salganicoff, M.; Ungar, L.H.; Bajcsy, R. Active learning for vision-based robot grasping. Mach. Learn. 1996, 23, 251–278.
58. Chevalier, L.; Jaillet, F.; Baskurt, A. Segmentation and Superquadric Modeling of 3D Objects. February 2003. Available online: http://wscg.zcu.cz/wscg2003/Papers_2003/D71.pdf (accessed on 5 November 2023).
59. Vilar, C.; Krug, S.; O'Nils, M. Realworld 3D object recognition using a 3D extension of the hog descriptor and a depth camera. Sensors 2021, 21, 910.
60. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Proceedings of the Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC); Springer: Berlin/Heidelberg, Germany, 2019; Volume 11, pp. 128–144.
61. Li, J.; Sun, Y.; Luo, S.; Zhu, Z.; Dai, H.; Krylov, A.S.; Ding, Y.; Shao, L. P2V-RCNN: Point to voxel feature learning for 3D object detection from point clouds. IEEE Access 2021, 9, 98249–98260.
62. Li, J.; Luo, S.; Zhu, Z.; Dai, H.; Krylov, A.S.; Ding, Y.; Shao, L. 3D IoU-Net: IoU guided 3D object detector for point clouds. arXiv 2020, arXiv:2004.04962.
63. Shi, S.; Guo, C.; Jiang, L.; Wang, Z.; Shi, J.; Wang, X.; Li, H. Pv-rcnn: Point-voxel feature set abstraction for 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10529–10538.
64. Shi, S.; Wang, X.; Li, H. Pointrcnn: 3D object proposal generation and detection from point cloud. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 770–779.
65. Mao, J.; Shi, S.; Wang, X.; Li, H. 3D object detection for autonomous driving: A review and new outlooks. arXiv 2022, arXiv:2206.09474.
66. Fernandes, D.; Silva, A.; Névoa, R.; Simões, C.; Gonzalez, D.; Guevara, M.; Novais, P.; Monteiro, J.; Melo-Pinto, P. Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy. Inf. Fusion 2021, 68, 161–191.

67. Zhou, Y.; Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3D object detection. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4490–4499.
68. Yan, Y.; Mao, Y.; Li, B. Second: Sparsely embedded convolutional detection. Sensors 2018, 18, 3337.
69. Zhu, B.; Jiang, Z.; Zhou, X.; Li, Z.; Yu, G. Class-balanced grouping and sampling for point cloud 3D object detection. arXiv 2019, arXiv:1908.09492.
70. Ge, R.; Ding, Z.; Hu, Y.; Wang, Y.; Chen, S.; Huang, L.; Li, Y. Afdet: Anchor free one stage 3D object detection. arXiv 2020, arXiv:2006.12671.
71. Yin, T.; Zhou, X.; Krahenbuhl, P. Center-based 3D object detection and tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 19–25 June 2021; pp. 11784–11793.
72. Li, F.; Jin, W.; Fan, C.; Zou, L.; Chen, Q.; Li, X.; Jiang, H.; Liu, Y. PSANet: Pyramid splitting and aggregation network for 3D object detection in point cloud. Sensors 2020, 21, 136.
73. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
74. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Adv. Neural Inf. Process. Syst. 2017, 30, 1, 5, 7.
75. Zarzar, J.; Giancola, S.; Ghanem, B. PointRGCN: Graph convolution networks for 3D vehicles detection refinement. arXiv 2019, arXiv:1911.12236.
76. Shin, K.; Kwon, Y.P.; Tomizuka, M. Roarnet: A robust 3D object detection based on region approximation refinement. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Dearborn, MI, USA, 9–12 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2510–2515.
77. Meyer, G.P.; Laddha, A.; Kee, E.; Vallespi-Gonzalez, C.; Wellington, C.K. Lasernet: An efficient probabilistic 3D object detector for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12677–12686.
78. Vora, S.; Lang, A.H.; Helou, B.; Beijbom, O. Pointpainting: Sequential fusion for 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4604–4612.
79. Lang, A.H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; Beijbom, O. Pointpillars: Fast encoders for object detection from point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12697–12705.
80. Zhao, X.; Liu, Z.; Hu, R.; Huang, K. 3D object detection using scale invariant and feature reweighting networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Hawaii, HI, USA, 27 January–1 February 2019; Volume 33, pp. 9267–9274.
81. Wang, Z.; Jia, K. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3D object detection. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1742–1749.
82. Qi, C.R.; Liu, W.; Wu, C.; Su, H.; Guibas, L.J. Frustum pointnets for 3D object detection from rgb-d data. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 918–927.
83. Rahman, M.M.; Tan, Y.; Xue, J.; Lu, K. Notice of violation of IEEE publication principles: Recent advances in 3D object detection in the era of deep neural networks: A survey. IEEE Trans. Image Process. 2019, 29, 2947–2962.
84. Chang, J.R.; Chen, Y.S. Pyramid stereo matching network. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5410–5418.
85. Zhang, F.; Prisacariu, V.; Yang, R.; Torr, P.H. Ga-net: Guided aggregation net for end-to-end stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 185–194.
86. Wang, Y.; Lai, Z.; Huang, G.; Wang, B.H.; Van Der Maaten, L.; Campbell, M.; Weinberger, K.Q. Anytime stereo image depth estimation on mobile devices. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5893–5900.
87. Guo, X.; Yang, K.; Yang, W.; Wang, X.; Li, H. Group-wise correlation stereo network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3273–3282.

88. Kendall, A.; Martirosyan, H.; Dasgupta, S.; Henry, P.; Kennedy, R.; Bachrach, A.; Bry, A. End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 66–75.
89. Chen, X.; Kundu, K.; Zhu, Y.; Berneshawi, A.G.; Ma, H.; Fidler, S.; Urtasun, R. 3D object proposals for accurate object class detection. Adv. Neural Inf. Process. Syst. 2015, 28, 1.
90. Kar, A.; Häne, C.; Malik, J. Learning a multi-view stereo machine. Adv. Neural Inf. Process. Syst. 2017, 30, 2.
91. Ma, X.; Wang, Z.; Li, H.; Zhang, P.; Ouyang, W.; Fan, X. Accurate monocular 3D object detection via color-embedded 3D reconstruction for autonomous driving. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6851–6860.
92. Weng, X.; Kitani, K. Monocular 3D object detection with pseudo-lidar point cloud. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3, 6, 11.
93. Wang, Y.; Chao, W.L.; Garg, D.; Hariharan, B.; Campbell, M.; Weinberger, K.Q. Pseudo-lidar from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8445–8453.
94. Ding, M.; Huo, Y.; Yi, H.; Wang, Z.; Shi, J.; Lu, Z.; Luo, P. Learning depth-guided convolutions for monocular 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 1000–1001.
95. Wang, L.; Du, L.; Ye, X.; Fu, Y.; Guo, G.; Xue, X.; Feng, J.; Zhang, L. Depth-conditioned dynamic message propagation for monocular 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 454–463.
96. Huang, K.C.; Wu, T.H.; Su, H.T.; Hsu, W.H. Monodtr: Monocular 3D object detection with depth-aware transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4012–4021.
97. Simonelli, A.; Bulo, S.R.; Porzi, L.; Ricci, E.; Kontschieder, P. Towards generalization across depth for monocular 3D object detection. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XXII 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 767–782.
98. Simonelli, A.; Bulo, S.R.; Porzi, L.; López-Antequera, M.; Kontschieder, P. Disentangling monocular 3D object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1991–1999.
99. Ma, X.; Zhang, Y.; Xu, D.; Zhou, D.; Yi, S.; Li, H.; Ouyang, W. Delving into localization errors for monocular 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4721–4730.
100. Zhang, Y.; Lu, J.; Zhou, J. Objects are different: Flexible monocular 3D object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3289–3298.
101. Wang, T.; Zhu, X.; Pang, J.; Lin, D. Probabilistic and Geometric Depth: Detecting Objects in Perspective. In Proceedings of the Conference on Robot Learning (CoRL), London, UK, 8 November 2021.
102. Li, P.; Zhao, H.; Liu, P.; Cao, F. Rtm3d: Real-time monocular 3D detection from object keypoints for autonomous driving. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part III 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 644–660.
103. Liu, Z.; Wu, Z.; Tóth, R. Smoke: Single-stage monocular 3D object detection via keypoint estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 996–997.
104. Li, P.; Zhao, H. Monocular 3D detection with geometric constraint embedding and semi-supervised training. IEEE Robot. Autom. Lett. 2021, 6, 5565–5572.
105. Cai, Y.; Li, B.; Jiao, Z.; Li, H.; Zeng, X.; Wang, X. Monocular 3D object detection with decoupled structured polygon estimation and height-guided depth estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10478–10485.
106. Hu, H.N.; Yang, Y.H.; Fischer, T.; Darrell, T.; Yu, F.; Sun, M. Monocular quasi-dense 3D object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 2022, 45, 1992–2008.
107. Heylen, J.; De Wolf, M.; Dawagne, B.; Proesmans, M.; Van Gool, L.; Abbeloos, W.; Abdelkawy, H.; Reino, D.O. Monocinis: Camera independent monocular 3D object detection using instance segmentation. In Proceedings of the

IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2021; pp. 923–934.

108. Chen, Y.; Tai, L.; Sun, K.; Li, M. Monopair: Monocular 3D object detection using pairwise spatial relationships. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12093–12102.
109. Mousavian, A.; Anguelov, D.; Flynn, J.; Kosecka, J. 3D bounding box estimation using deep learning and geometry. In Proceedings of the IEEE Conferencef on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7074–7082.
110. Brazil, G.; Liu, X. M3d-rpn: Monocular 3D region proposal network for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9287–9296.
111. Li, B.; Ouyang, W.; Sheng, L.; Zeng, X.; Wang, X. Gs3d: An efficient 3D object detection framework for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1019–1028.
112. Xiang, Y.; Schmidt, T.; Narayanan, V.; Fox, D. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. arXiv 2017, arXiv:1711.00199.
113. Lu, Y.; Ma, X.; Yang, L.; Zhang, T.; Liu, Y.; Chu, Q.; Yan, J.; Ouyang, W. Geometry uncertainty projection network for monocular 3D object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Event, 11–17 October 2021; pp. 3111–3121.