

A SEMANTIC 3D POINT CLOUD SEGMENTATION APPROACH BASED ON OPTIMAL VIEW SELECTION FOR 2D IMAGE FEATURE EXTRACTION

*A. Adam¹, L. Grammatikopoulos², G.Karras¹, E. Protopapadakis¹, K. Karantzas¹

¹ National Technical University of Athens, NTUA, Department of Rural and Surveying Engineering, 57001 Zografou, Greece-
(katerin, gkarras, eftprot, karank,)@central.ntua.gr

² University of West Attica, UNIWA, Department of Geomatic Engineering and Geoinformatics, 12243 Aigaleo, Greece -
(lazaros)@uniwa.gr

Commission II

KEY WORDS: Point cloud processing, Semantic segmentation, Structure-from-Motion data, City models

ABSTRACT:

3D semantic segmentation is the joint task of partitioning a point cloud into semantically consistent 3D regions and assigning them to a semantic class/label. While the traditional approaches for 3D semantic segmentation typically rely only on structural information of the objects (i.e. object geometry and shape), the last years many techniques combining both visual and geometric features have emerged, taking advantage of the progress in SfM/MVS algorithms that reconstruct point clouds from multiple overlapping images. Our work describes a hybrid methodology for 3D semantic segmentation, relying both on 2D and 3D space and aiming at exploring whether image selection is critical as regards the accuracy of 3D semantic segmentation of point clouds. Experimental results are demonstrated on a free online dataset depicting city blocks around Paris. The experimental procedure not only validates that hybrid features (geometric and visual) can achieve a more accurate semantic segmentation, but also demonstrates the importance of the most appropriate view for the 2D feature extraction.

1. INTRODUCTION

3D segmentation is the joint task of partitioning a point cloud into consistent 3D regions. On the other hand, 3D semantic segmentation is the joint task of partitioning a point cloud into semantically consistent regions and classifying them in their corresponding semantic concepts, deploying a machine learning procedure. While there have been approaches applying 3D semantic segmentation to generic concepts such as table, mug, lamp, rocket, etc. (Qi et al., 2017), the main emphasis has been placed on detecting objects in indoor and outdoor scenes, such as furniture (Badami et al., 2017), cars, buildings and vegetation (Huang and You, 2016, Tchappmi et al., 2017, Lawin et al., 2017), and facade parsing (Martinovic et al., 2015).

Typical approaches for facade parsing, and 3D segmentation in general, rely solely on the structural information of the objects (i.e. the geometry and shape of the objects), without taking into account visual information (i.e. the color and texture of an object). However, geometrical and shape features are not able to distinguish between objects belonging to the same surface (e.g. walls and door). On the other hand, 2D segmentation approaches that rely on visual characteristics may be able to separate this kind of data. Based on the 2D color and texture features, there have been works that attempt facade parsing in 2D space (Hantang Liu, 2017). However, 3D and 2D data contain complementary information that, if effectively combined, can provide more accurate segmentation results.

The joint consideration of geometrical and visual (spectral) characteristics has been an area of intense research efforts lately. This has been also motivated by the popularity of sensors (e.g. Unmanned Aerial Vehicle (UAV)-based image capturing, depth sen-

sors, etc.) that capture also the visual information of the scene along with corresponding algorithms for reconstructing dense 3D point clouds from multiple overlapping images (e.g. Structure from Motion (SfM) and Multiview Stereo (MVS) algorithms). Many works have shown the benefit of combining geometrical and visual information from both spaces regardless of the data capturing technique, e.g. 3D point clouds from multiple overlapping images (through SfM algorithms) (Gadde et al., 2017, Martinovic et al., 2015, Riemenschneider et al., 2014), RBGD data from depth sensors (Song and Xiao, 2014, Valentin et al., 2013, Kahler and Reid, 2013, Munoz et al., 2012).

Particularly for facade parsing, 3D models reconstructed using SfM algorithms and overlapping images is one of the most popular types of data. Combining 3D and 2D information, given the nature of this type of data, a popular approach is to create feature vectors for each point in the 3D point cloud combined with a set of 2D features from an optimal view of the object in the 2D space. Thus focus has been placed on identifying this optimal 2D view from the multiple overlapping images. In this direction, we propose a hybrid point cloud segmentation approach that jointly considers the structural information from 3D representation and the visual information, extracted from the most suitable image. More specifically, having the overlapping images and the resulting 3D model at its disposal, the proposed segmentation approach extracts the geometric features from the point cloud, while formulating a rule for selecting the best image for 2D feature extraction. A hybrid vector is formed and inserted into classifiers for training. Our algorithm is evaluated using RueMongue2014, a free available dataset.

The paper is organized as follows. Section 2 reviews the related work, presenting point cloud segmentation solely relying on geometry, as well as novel, hybrid procedures, combining both 3D

*Corresponding author

and 2D data. Moreover, 2.3 focuses on the methodologies combining 3D and 2D features and aiming at semantic segmentation, sourcing from SFM/MVS data. Section 3 describes in detail our proposed methodology. Section 4 presents the evaluation and experimental results. Finally, Section 5 discusses and concludes the impact of this work and describes our intentions for future research.

2. RELATED WORK

This paper presents a 3D point cloud semantic segmentation approach that jointly considers structural and visual characteristics of the objects. Concerning the related works, many of them focus on point cloud segmentation and do not assign the segments to specific semantic labels. These methodologies mainly rely on the geometrical features of the scene, while a limited number of works combine visual and geometrical features for simple scene segmentation without any reference on the semantics. Finally, there are some works that deploy machine learning techniques, combining visual and structural features to achieve a more accurate semantic segmentation. In the following, Section 2.1 presents the traditional segmentation methodologies relying on geometrical features, and Section 2.2 presents the hybrid approaches combining visual and geometrical characteristics. Finally, Section 2.3 presents the most related works to our proposed methodology, that focus on 3D semantic segmentation of scenes, exploiting both 3D and 2D features.

2.1 Point cloud segmentation based on geometry

The typical methodologies for point cloud segmentation rely on the geometric properties of the point cloud and thus can be applied irrespective of the source of the point cloud (e.g. different capturing devices such as LIDAR and terrestrial laser scanner devices or point clouds sourcing from SfM algorithms). These methods may be roughly categorized into three different groups; namely, model-fitting, region-growing and clustering approaches. Brief examples of the aforementioned techniques are described in the following.

Most of the algorithms belonging to model-fitting approaches are based on either the Hough Transform (HT) (Ballard, 1981) or the Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981) approach. HT was firstly introduced as a method of detecting well-defined shapes, but was soon extended to arbitrary shapes as well, while, RANSAC is capable of robustly computing the parameters of a given mathematical model (i.e. plane) and thus detecting only the surfaces that can be modeled in mathematical terms.

Concerning the HT related methods, relative works include (Voselman et al., 2004), where the authors review plane, sphere and cylinder detection from city models, while, in (Rabbani and Van Den Heuvel, 2005) the automatic detection of cylinders in point clouds is examined. Comparative results for plane-based detection, in the context of extracting roofs from buildings, are presented in (Tarsha-Kurdi et al., 2007). Among more methodologies that rely on RANSAC, we find the work of (Boulaassal et al., 2007), where the authors extract facades from points and also extend RANSAC by incorporating some extra steps in order to achieve better segmentation results. Moreover, (Schnabel et al., 2007) extracts planes, cylinders and spheres, after computing normal vectors for each point, in order to perform the assignment in a candidate shape. Furthermore, (Chen et al., 2014) also improved

RANSAC to detect roof-tops in point clouds sourcing from airborne laser scanners. More specifically, firstly they performed segmentation and afterwards a Voronoi-based primitive boundary extraction algorithm was introduced in order to extract the primitives of each rooftop. Finally, (Li et al., 2017) proposed the clustering of points into planar and non-planar, according to the extracted normals.

The second category for the simple geometry-based segmentation of 3D scenes, relies on region-growing techniques, as first introduced in (Besl and Jain, 1988). In this work, the mean Gaussian curvature was computed and resulted in a coarse segmentation. The initial segmentation was afterwards refined through an iterative region-growing procedure based on a variable order bivariate surface-fitting. The authors of (Gorte, 2002) also employed region-growing for 3D segmentation, since they proposed to use Triangulated Irregular Networks (TINs) as seeds and merged triangles with similar angles within a predefined distance. Furthermore, in (Vieira and Shimada, 2005) points across edges were eliminated with the help of a curvature threshold and noise reduction was performed. The final remaining points served as seeds for the region-growing. From studying all the aforementioned works, it is obvious that the efficiency of region growing methodologies for point cloud segmentation is strongly correlated with successful seed selection and merging criteria.

The final category refers to methods which approach the segmentation problem through clustering, by assigning each point in the 3D point cloud with a set of features. Among them, there is the work of (Biosca and Lerma, 2008), where the point cloud was segmented with the help of fuzzy clustering. More specifically, for each point features were extracted, clusters were formed and subsequently the results were refined by assigning unlabeled points to the nearest cluster. Combining clustering and region-growing the authors of (Dorninger and Nothegger, 2007) first use hierarchical clustering in four dimensional feature space in order to compute the center of the clusters. These centers serve as seeds. Similar regions were merged if the normal distance between the examined point and the seed cluster was within a predefined threshold. Compared to region-growing, clustering methodologies are more robust. However, their efficiency depends on the quality of the extracted features.

2.2 Hybrid segmentation approaches

Concerning the hybrid segmentation procedures, combining both 2D and 3D spaces, to our knowledge, relative literature is rather limited. There are quite a few studies that have combined segmentation in the 2D and 3D space, either by projecting the 3D space onto an image and applying segmentation in the 2D space, or by deploying segmentation both in the 3D and 2D domain and afterwards combine them. With respect to the first category, the authors of (Kabaria et al., 2014) proposed to project the point cloud on a gray scale image and segment the 2D image with the help of an active contour segmentation algorithm. Similar to this work, the authors of (Aijazi et al., 2016) transformed the 3D point cloud (object space) into an elevation image in order to segment urban objects. Moreover, the authors of (Hickson et al., 2014) first identified the object boundaries in the 3D space where depth discontinuities occurred and used this information as an initial coarse segmentation, to be refined. Afterwards, image segmentation in the CIE LAB color space followed, while preventing any merging to occur across depth discontinuities. In (Vetrivel et al., 2015), the authors proposed a work flow for building detection from point clouds sourcing from UAV images, which firstly

detected planar segments in the point clouds and defined 3D Regions of Interest (ROIs). The 3D ROIs were afterwards identified in the images and hybrid segmentation (by deploying region growing both in the 2D and the 3D space) was employed to them in order to refine the 2D segmentation results. Finally, the authors of (Omidalizarandi and Saadatseresht, 2013) have also worked on identifying roof segments using this combined segmentation.

2.3 Hybrid methodologies for facade parsing

Closer to our work, there have been also methods combining information from the 2D and 3D spaces aiming at semantic segmentation. These methods rely on learning-based approaches and require labelled training data. In this direction, (Landrieu et al., 2017) perform ensemble learning, based on 3D features from the point cloud and 2D features sourcing from the projection of points into a horizontally oriented plane. (Ladický et al., 2012), also aims at semantic segmentation of 3D scenes, combining both depth maps and appearance features. In (Serna and Marcotegui, 2014), elevation images are used and final results are obtained by the re-projection of images into the 3D point cloud. Objects are classified with the help of Support Vector Machines (SVMs).

Moreover, there are also some works, trying to achieve an accurate segmentation of city models and more specifically concern facade parsing, such as (Gadde et al., 2017) and (Jampani et al., 2015). These approaches aim at segmenting the different parts of a facade and at assigning to them a semantic label such as door, window, balcony etc. with the help of annotated data. The methodologies most related to ours are (Martinovic et al., 2015) and (Riemenschneider et al., 2014). The first forms a hybrid vector, comprising of both 2D and 3D features to be inserted into the classification scheme. 2D features (RGB and Lab values) are extracted as the mean values among all images in which the point to be classified is visible, while 3D features such as height and depth of point as well as the vector normals are directly extracted from the mesh. Finally, in a similar vein to our research, but working on meshes instead of point clouds, is (Riemenschneider et al., 2014) which selects the best view for each triangle mesh to be classified and uses this image for the 2D feature extraction.

3. METHODOLOGY

The problem at hand points to a traditional multi class classification problem; given a set of joint geometrical and visual characteristics, we try to estimate correct the class of the object. Our contribution concerns the formulation of a heuristic rule for optimal image identification.

3.1 Feature extraction

The aim of this work is to perform semantic segmentation on point clouds of urban scenes, exploiting both 2D and 3D features. Our novelty focuses on proposing an innovative optimal view selection technique for the 2D features extraction, when point clouds instead of meshes are to be classified. When the optimal view is selected, i.e. the one that has the best chances of classifying correctly the desired object, occlusions are diminished and the overall accuracy of semantic segmentation improves.

For the optimal view selection, first all 3D points X are projected, via the projection matrices P to the 2D space, according to Equation 1 and images in which the points are visible are selected:

$$X = P \cdot x \quad (1)$$

where X refers to the 3D world point and x to the 2D image point. Finally, P refers to the projection matrix.

Secondly, around each point in the 3D space a rectangular patch is formed on the plane of X , with predefined dimensions. The coordinates of the four corners forming the rectangular patch are known, and are projected via Equation 1, to the different images. From all images in which the point is visible, only those where the whole rectangle is depicted (i.e. all the four corners are projected within the image) are selected. Afterwards, the area of the projected patch in each image is computed and the optimal image is the one ensuring the maximum area of the rectangle (i.e. the image with the maximum number of pixels forming the projected patch). Using this heuristic rule for the optimal image selection, we are eliminating very oblique and small-scale views for the 2D feature extraction step.

When the optimal view is identified, features are extracted. Concerning the geometric (3D) features, the depth (d), normalised height (h) and vector normal (n) for each point are computed, directly from the point cloud. On the other hand, texture features are extracted from the optimal image and include RGB and Lab values of the projected point. A hybrid vector:

$$V = [d \quad h \quad n \quad \text{RGB} \quad \text{Lab}] \quad (2)$$

with dimensions 11×1 is formed for each point and afterwards inserted into a Random Forest (RF) and an AdaBoost classifier.

3.2 Employed Machine Learning Techniques

In order to tackle the complexity of input feature space, random forest classifier is employed. RF is a decision trees ensemble method, which has high accuracy in a variety of fields. Decision trees, as a predictive models, map observations about an item to conclusions about the item's target value. In classification tree structures, leaves represent class labels and branches represent conjunctions of features leading to those class labels. Each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with a feature are labeled with each of the possible values of the feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes. Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items (Rokach and Maimon, 2005). Different algorithms use different metrics for measuring "best". These generally measure the homogeneity of the target variable within the subsets. These metrics are applied to each candidate subset, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split.

Adaptive boosting (AdaBoost) is an ensemble learning algorithm, which is more resistant to over-fitting, but it is often sensitive to noisy data and outliers (Boyarshinov and Magdon-Ismael, 2007). AdaBoost creates a strong learner (a classifier that is well-correlated to the true classifier) by iteratively adding weak learners (a classifier that is only slightly correlated to the true classifier). During each round of training, a new weak learner is added to the ensemble, and a weighting vector is adjusted to focus on examples that were mis-classified in previous rounds. The result is a classifier that has higher accuracy than the weak learners classifiers.

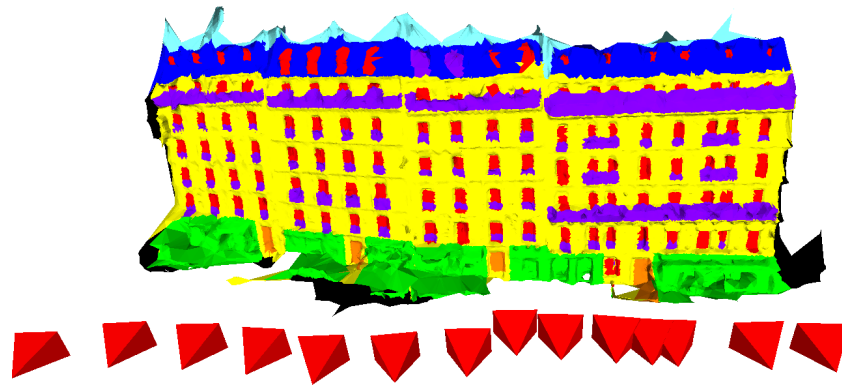


Figure 1. The ground-truth annotation of RueMongue 2014 Dataset, visualising in different colors the semantic classes

4. EXPERIMENTAL PROTOCOL

4.1 Dataset

For the assessment of the proposed approach RueMongue 2014¹, a freely available dataset was employed (Figure 1), consisting of 428 images capturing urban scenes. SfM and MVS techniques have been deployed to acquire the point cloud, thus image data are available along with the point cloud. Pictures and meshes are also available and have been used for semantic segmentation of urban scenes. Our proposed methodology aims at semantic segmentation of point clouds, incorporating both 2D and 3D features, thus only the images and the points clouds are used. The dataset contains eight different semantic classes, namely sky, balcony, facade, shop, window, door, shop and a void class, along with a train/test split and an evaluation protocol. An instance of the used dataset, is depicted in Figure 1 visualising the ground-truth annotation of the different semantic classes. More specifically, class shop is visualised with green color, balcony with magenta, sky with light blue, roof with dark blue, window with red, door with orange and main facade with green. Points that belong to void class are visualised with black color.

4.2 Experimental Results

This section presents quantitative results to assess the efficiency of our proposed methodology. In particular, two baseline procedures have also been implemented for comparison with our proposed technique, performing semantic segmentation separately at the 3D and the 2D space (e.g. forming a vector of exclusively the aforementioned 3D and 2D features, respectively). Moreover, another hybrid baseline procedure, where a hybrid vector containing the aforementioned 3D and 2D features (the 2D features are extracted from a random image) has also been implemented and compared with our approach employing the optimal image selection procedure. The mean accuracy metric has been employed used for the evaluation. Experimental results are presented for the two employed machine learning techniques, namely Random Forests and Adaboost. Results are tabulated in Table 1.

After a close examination of Table 1, one can observe that the combination of 2D and 3D features resulted into a more accurate semantic segmentation of the 3D scene, which is in accordance with the literature. Moreover, our initial assumption, i.e., a more sophisticated image selection procedure would lead to a more detailed semantic segmentation result has been validated. In particular, the fusion of 2D and 3D features resulted into an overall

¹<http://varcity.ethz.ch/3dchallenge/>

mean accuracy equal to 50.1% for the baseline procedure and a mean accuracy equal to 60.0% for the proposed procedure, when Random Forests are used.

Furthermore, after a close look at the quantitative results, one can observe that image selection is a highly crucial step towards scene semantic segmentation since for the cases of employing separately 2D and 3D features. More specifically, 2D features alone resulted into relative higher accuracy rates when the optimal image was selected, i.e., 57.2% accuracy compared to 49.9% when only 3D features were employed. On the other hand, when a random image was selected for 2D feature extraction, 3D features seem to be more important as they lead to 49.9% accuracy, compared to 48.6% achieved with the 2D features.

Finally, it is also important to note here that more sophisticated machine learning techniques, such as Conditional Random Fields (CRFs), as in (Riemenschneider et al., 2014), could be integrated into the whole procedure to achieve higher accuracy rates. However, this is out of the scope of our research, as the aim of this particular work is to explore whether the optimal image selection is crucial towards the semantic segmentation of the scene.

5. CONCLUSIONS

The aim of this work is to propose a new methodology for optimal image selection and to explore whether this procedure is beneficial towards the semantic segmentation of the scene. Segmentation is performed incorporating information from both the 3D and 2D space, forming a hybrid vector that is afterwards inserted into two machine learning techniques. 2D features are extracted from the optimal image which is selected upon predefined criteria. Experimental results have shown that by leveraging geometrical and visual features the overall effectiveness of 3D semantic segmentation is improved, as it has been evaluated in the domain of city buildings and facade segmentation. Moreover, the novel procedure for the optimal view selection for 2D feature extraction has been validated. As shown by the experimental results outlined in Section 4.2, the expected efficiency of applying the proposed heuristic rule for the optimal selection has been demonstrated. What is important to note here is that he have used traditional and basic machine learning techniques such as RFs and Adaboost, in order to explore whether the optimal view selection could achieve results of higher accuracy. Thus, concerning future work, we aim to integrate more innovative machine learning techniques into the whole pipeline, such as deep learning techniques.

Table 1. Classification accuracy when only 2D, only 3D and combined 2D + 3D features are used for the baseline and the optimal image case, both for Random Forests and AdaBoost classifiers.

	AdaBoost (random image)	RF(random image)	AdaBoost (optimal image)	RF (optimal image)
2D features	39.3%	48.6%	46.5%	57.2%
3D features	40.2%	49.9%	40.2%	49.9%
2D+3D features	47.2%	50.1%	54.9%	60.0%

ACKNOWLEDGEMENTS

This research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship Grant (GA. no. 139230/2/22-8-2017).

REFERENCES

- Aijazi, A. K., Serna, A., Marcotegui, B., Checchin, P. and Trassoudaine, L., 2016. Segmentation and classification of 3d urban point clouds: Comparison and combination of two approaches. In: *Field and Service Robotics*, Springer, pp. 201–216.
- Badami, I., Tom, M., Mathias, M. and Leibe, B., 2017. 3d semantic segmentation of modular furniture using rjmc. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 64–72.
- Ballard, D. H., 1981. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition* 13(2), pp. 111–122.
- Besl, P. and Jain, R., 1988. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10(2), pp. 167–192.
- Biosca, J. M. and Lerma, J., 2008. Unsupervised robust planar segmentation of terrestrial laser scanner point clouds based on fuzzy clustering methods. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(1), pp. 84–98.
- Boulaassal, H., Landes, T., Grussenmeyer, P. and Tarsha-Kurdi, F., 2007. Automatic segmentation of building facades using terrestrial laser data. In: *ISPRS Workshop on Laser Scanning and SilviLaser*, pp. 65–70.
- Boyarsinov, V. and Magdon-Ismail, M., 2007. Efficient optimal linear boosting of a pair of classifiers. *IEEE transactions on neural networks* 18(2), pp. 317–328.
- Chen, D., Zhang, L., Mathiopoulos, P. T. and Huang, X., 2014. A methodology for automated segmentation and reconstruction of urban 3-d buildings from als point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, pp. 4199–4217.
- Dorninger, P. and Nothegger, C., 2007. 3d segmentation of unstructured point clouds for building modelling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35(3/W49A), pp. 191–196.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.
- Gadde, R., Jampani, V., Marlet, R. and Gehler, P., 2017. Efficient 2d and 3d facade segmentation using auto-context. *IEEE transactions on pattern analysis and machine intelligence*.
- Gorte, B., 2002. Segmentation of tin-structured surface models. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* 34(4), pp. 465–469.
- Hantang Liu, Jialiang Zhang, J. Z. S. C. H. H., 2017. Deepfacade: A deep learning approach to facade parsing. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 2301–2307.
- Hickson, S., Birchfield, S., Essa, I. and Christensen, H., 2014. Efficient hierarchical graph-based segmentation of rgbd videos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 344–351.
- Huang, J. and You, S., 2016. Point cloud labeling using 3d convolutional neural network. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 2670–2675.
- Jampani, V., Gadde, R. and Gehler, P. V., 2015. Efficient facade segmentation using auto-context. In: *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, IEEE, pp. 1038–1045.
- Kabaria, H., Goelb, A. and Moore, K., 2014. 3d reconstruction, segmentation and classification of corals from aerial images final report: Cs 231a.
- Kahler, O. and Reid, I., 2013. Efficient 3d scene labeling using fields of trees. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3064–3071.
- Ladicky, L., Sturges, P., Russell, C., Sengupta, S., Bastanlar, Y., Clocksin, W. and Torr, P. H., 2012. Joint optimization for object class segmentation and dense stereo reconstruction. *International Journal of Computer Vision* 100(2), pp. 122–133.
- Landrieu, L., Raguét, H., Vallet, B., Mallet, C. and Weinmann, M., 2017. A structured regularization framework for spatially smoothing semantic labelings of 3d point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 132, pp. 102–118.
- Lawin, F. J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F. S. and Felsberg, M., 2017. Deep projective 3d semantic segmentation. In: *International Conference on Computer Analysis of Images and Patterns*, Springer, pp. 95–107.
- Li, L., Yang, F., Zhu, H., Li, D., Li, Y. and Tang, L., 2017. An improved ransac for 3d point cloud plane segmentation based on normal distribution transformation cells. *Remote Sensing* 9(5), pp. 433.
- Martinovic, A., Knopp, J., Riemenschneider, H. and Van Gool, L., 2015. 3d all the way: Semantic segmentation of urban scenes from start to end in 3d. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4456–4465.
- Munoz, D., Bagnell, J. A. and Hebert, M., 2012. Co-inference for multi-modal scene analysis. In: *European Conference on Computer Vision*, Springer, pp. 668–681.
- Omidiazarandi, M. and Saadatseresht, M., 2013. Segmentation and classification of point clouds from dense aerial image matching. *The International Journal of Multimedia & Its Applications* 5(4), pp. 33.
- Qi, C. R., Su, H., Mo, K. and Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE 1(2), pp. 4.

- Rabbani, T. and Van Den Heuvel, F., 2005. Efficient hough transform for automatic detection of cylinders in point clouds. *Journal of Photogrammetry and Remote Sensing* 3, pp. 60–65.
- Riemenschneider, H., Bódis-Szomorú, A., Weissenberg, J. and Van Gool, L., 2014. Learning where to classify in multi-view semantic segmentation. In: *European Conference on Computer Vision*, Springer, pp. 516–532.
- Rokach, L. and Maimon, O., 2005. Top-down induction of decision trees classifiers—a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35(4), pp. 476–487.
- Schnabel, R., Wahl, R. and Klein, R., 2007. Efficient ransac for point-cloud shape detection. In: *Computer graphics forum*, Vol. 26number 2, Wiley Online Library, pp. 214–226.
- Serna, A. and Marcotegui, B., 2014. Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, pp. 243–255.
- Song, S. and Xiao, J., 2014. Sliding shapes for 3d object detection in depth images. In: *European conference on computer vision*, Springer, pp. 634–651.
- Tarsha-Kurdi, F., Landes, T. and Grussenmeyer, P., 2007. Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. In: *ISPRS Workshop on Laser Scanning and SilviLaser*, Vol. 36, pp. 407–412.
- Tchapmi, L. P., Choy, C. B., Armeni, I., Gwak, J. and Savarese, S., 2017. Segcloud: Semantic segmentation of 3d point clouds. *arXiv preprint arXiv:1710.07563*.
- Valentin, J. P., Sengupta, S., Warrell, J., Shahrokni, A. and Torr, P. H., 2013. Mesh based semantic modelling for indoor and outdoor scenes. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, IEEE, pp. 2067–2074.
- Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G., 2015. Segmentation of uav-based images incorporating 3d point cloud information. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(3), pp. 261.
- Vieira, M. and Shimada, K., 2005. Surface mesh segmentation and smooth surface extraction through region growing. *Computer aided geometric design* 22(8), pp. 771–792.
- Vosselman, G., Gorte, B., Sithole, G. and Rabbani, T., 2004. Recognising structure in laser scanner point clouds. *International archives of photogrammetry, remote sensing and spatial information sciences* 46(8), pp. 33–38.