



Clustering Quality Measures for Point Cloud Segmentation Tasks

Jakub Walczak and Adam Wojciechowski^(✉)

Institute of Information Technology, Łódź University of Technology, Łódź, Poland
adam.wojciechowski@p.lodz.pl

Abstract. This paper presents improved weighted measures for a point cloud segmentation quality evaluation. They provide more reliable and intuitive appraisal as well as more representative classification characteristics. The new measures are compared with the existing ones: based on classification, and based on information theory. The experiments and measures evaluation were performed for the recently outstanding fresh planes segmentation method. Experiments results showed that newly elaborated measures provide a researcher with distinguished information about segmentation output. This paper introduces recommendations for quality measures adjustment to a particular planar fragments detection problem, what implies contributions for effective development of such methods.

Keywords: Point cloud segmentation · Quality measures
Plane detection

1 Introduction

Point cloud segmentation, being usually a step preceding semantic analysis of a set, is a task of high interest of many researchers. Besides cases like recognizing structures within a point cloud [23], hand pose and face segmentation [10–12], big data visualisation [13] or urban and architectural reconstruction [14], segmentation may be successfully used for compression purposes [22] thanks to storing objects by means of mathematical formulas instead of thousands of 3D points. Certainly, these are just a few out of plenty of possibilities where 3D segmentation contributes.

Beyond any doubt, any algorithm, by definition, is meant to produce a good or expected results in the context of presumed characteristics. These characteristics, or rather measures, allowed us to clearly compare algorithms and understand their flaws. That is why a quality measure should be carefully chosen prior to actual problem and method definition.

In this paper, we introduced two novel measures for segmentation quality assessment: weighted classification statistics (WCS) - improved version of the ordinary classification statistics (OCS), and planarity statistics (PS) as an indicator of planes extraction quality. Additional contribution of this paper is evaluation of new and existing measures for selected model-based method of planes

detection (as a reference a recent Li et al. method [3] has been chosen) and proposal of recommendations concerning particular measures.

2 Related Works

2.1 Planes Segmentation Methods

Among all methods aiming at segmentation of point clouds, Nguyen and Le identified five main categories, namely: edge-based, region-based, attributes-based, graph-based, and model-based methods [15]. Current studies focus mainly on model-based group, especially on the approaches based on random sample consensus (RANSAC). Below, we review above groups briefly to emphasize the main differences and to justify why recently the most effective Normal Distribution Transform cell-based RANSAC method, by Li et al. [3], was opted for our experiments.

Edge-based methods are constituted by methods employing gradient- or line-fitting algorithms for locating interesting points. Bhanu et al. proposed range images processing for edges extraction [1]. On the other hand, Sappa and Devy [19] relied on binary edge map generation with scan-lines approximation in orthogonal directions. These methods are characterized by sensitivity both to existing noise and non-uniform point distribution.

Region-based algorithms use greedy approach to examine similarity or distinctiveness in a limited vicinity. Rabbani et al. [18] applied simple region growing approach taking into account surface normal and points' connectivity. On the other hand, Xiao et al. [24] proposed subregion based region growing, where each subregion is considered as a planar one or not, based on Principal Component Analysis (PCA) or KLT [25]. Region-based methods, suffer from dependence on seed point selection as well as from the fact, that decision is made locally, and it may be not correct from the global point of view.

Generally, segmentation methods based on attributes make use of clustering algorithms built onto extracted attributes. Mean-shift clustering [6], hierarchical clustering [7], contextual processing [8] or statistically supported clustering [9] are cases in point. Limitations of this group are clustering methods constraints themselves. For *k-means* clustering, number of clusters need to be known in advance. On the other hand, for other methods, high noise sensitivity or time complexity involved with multidimensionality may occur.

Among the algorithms based on graphs, one may find the method making use of colour information added to laser scans [21] or the method of vicinity graph presented in [2]. Clearly, graph-based methods may need a complex preprocessing phase, like training [15].

Most of methods employing a model makes use of RANSAC algorithm. Its main advantage is inherent robustness against outlying data, unlike the other methods. Currently, it appears in many variations. Schnabel et al. [20] used RANSAC for efficient detection of planes, cylinders and spheres. They evaluated their method by means of correctly detected regions. Oehler et al. [16] combined RANSAC with Hough transform [17] to increase quality of planes detection.

Here, the authors identified number of true positives (TP) with 80% overlap region in order to evaluate their algorithm. Xu et al. [5] evaluated usability of different functions for weighted RANSAC in terms of their completeness (Eq. 4), correctness (Eq. 3), and quality (Eq. 5). Calvo et al. [33] engaged k-means strategy together with RANSAC to search for predefined model (like cube or pyramid) in a point cloud. They evaluated this algorithm by comparing an average angular deviation between normal vectors detected in a cloud and those known from a reference set. Finally, Li et al. [3] introduced an improved RANSAC-based method for uniform space division, which, following authors, presents itself as the most efficient state-of-art solution. Thus it was selected as a point cloud reference segmentation method.

The algorithm proposed by Li et al. relies on space division into cells of fixed size, whose dimensions have to be tuned for point cloud specifically. Cells are then classified either as planar ones or not. The decision is made on the grounds of eigenvalues, being the output of PCA procedure. If the ratio of the two greatest eigenvalues is lower than assumed threshold te , a cell is said to be a planar one (Eq. 1). For the dataset Room-1 [31], the authors assumed $te = 0.01$, whereas for Room-2 dataset [31], they took $te = 0.02$. The formula for te was determined by the authors empirically, taking into account that it is influenced by the cell size s and the noise level ϵ (Eq. 2).

$$\frac{\lambda_1}{\lambda_2} < te. \quad (1)$$

where λ_1 and λ_2 are, respectively the greatest, and the middle eigenvalue.

$$\left(\frac{\epsilon}{s}\right)^2 < te < 0.04. \quad (2)$$

Subsequently, the authors performed plane segmentation procedure, called NDT-RANSAC. In short, it consists of RANSAC-like examination of an individual planar cell. From each planar cell, the minimal set of three points is randomly taken to construct a hypothetical plane which may differ from that obtained with PCA. Having calculated a plane parameters, the rest of planar cells are compared to the current one, in terms of normal vector angular deviation and relative shift between objects. The plane, obtained by merging coherent patches is then refined taking into account its consensus set. The authors used verification measures based on confusion matrix analysis and they claim the quality of their segmentation procedure exceeds 88.5% of correctness (Eq. 3) and 85% of completeness (Eq. 4).

2.2 Current Quality Measures

Classification-Based Measures. Many current researchers, appraise their methods with confusion matrix analysis, treating the segmentation task as a kind of classification problem [3–5]. This kind of assessment relies on calculation of three basic measures using for classification evaluation:

correctness (also referred to as precision, Eq. 3), completeness (known also as recall or sensitivity, Eq. 4), and quality (Eq. 5) according to maximum overlapping technique introduced by Awrangjeb and Fraser [32].

$$correctness = \frac{\|TP\|}{\|TP\| + \|FP\|} \tag{3}$$

$$completeness = \frac{\|TP\|}{\|TP\| + \|FN\|} \tag{4}$$

$$quality = \frac{\|TP\|}{\|TP\| + \|FN\| + \|FP\|} \tag{5}$$

where $\|\cdot\|$ states for a cardinality of a set; TP, FP, FN states, respectively, for: true positives, false positives, and false negatives.

These measures do require unambiguous correspondence finding among groups of reference clustering \mathcal{R}_i , such that $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \dots, \mathcal{R}_n\}$ ($\bigcup_{i=1}^n \mathcal{R}_i = \mathcal{D}$) and clustering being appraised \mathcal{O}_j , where $\mathcal{O} = \{\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \dots, \mathcal{O}_m\}$ ($\bigcup_{j=1}^m \mathcal{O}_j = \mathcal{D}$). Both clusterings are built over the dataset \mathcal{D} . This correspondence is usually determined by searching patches that overlap the most [32], namely we look for the corresponding clusters \mathcal{O}_j and \mathcal{R}_i where:

$$\arg \max_j \frac{\|\mathcal{O}_j \cap \mathcal{R}_i\|}{\|\mathcal{O}_j\|} = \arg \max_i \frac{\|\mathcal{O}_j \cap \mathcal{R}_i\|}{\|\mathcal{R}_i\|}.$$

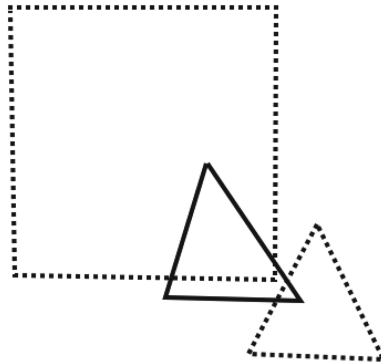


Fig. 1. A case of improper correspondence finding with maximum- overlapping method

Nevertheless these metrics consider solely number of clusters classified as TP, FP or FN. They do not take into account cardinality of overlapping regions, hence these statistics may be easily far-fetched and results might not be reliable. Depending on the presumed overlapping threshold (if any), these measures fail in case of high ratio of FP or FN, or highly unexpected output (see Fig. 1). In Fig. 1, we may clearly see that, using maximum overlapping strategy, a solid-line triangle will be associated with a dashed rectangle rather than its actual counterpart.

Micro- and Macro- averaging. The other approach for clustering quality assessment, or its variation applicable for any measures, are techniques being widely used in machine learning domain for multi-label classification, called micro- and macro- averaging. They aggregate intermediate results into global information. Manning et al. [29] defined macro-averaging as arithmetical average across classes and micro- averaging as weighted mean of multi-label classification measures. Hence, we may formally define macro-averaging as the arithmetic mean of the values and micro-averaged completeness, correctness and quality as in the Eqs. 7 and 8 [30].

$$correctness = \frac{\sum ||TP||}{\sum (||TP|| + ||FP||)} \quad (6)$$

$$completeness = \frac{\sum ||TP||}{\sum (||TP|| + ||FN||)} \quad (7)$$

$$quality = \frac{\sum ||TP||}{\sum (||TP|| + ||FP|| + ||FN||)} \quad (8)$$

where TP, FN, FP states respectively for true positives, false negatives, and false positives.

Variation of Information. Besides correctness, completeness, and quality relying solely on the binary decision: does cluster correspond or not, there is another group of methods for clustering comparison [34] making use of fuzzy correspondences, defined by information theory and information entropy of Shannon [26] (Eq. 9).

$$H(\mathcal{R}) = - \sum_{i=1}^n \frac{||\mathcal{R}_i||}{||\mathcal{D}||} \log_2 \left(\frac{||\mathcal{R}_i||}{||\mathcal{D}||} \right) \quad (9)$$

where \mathcal{R}_i is the cluster being considered and $||\mathcal{R}_i||$ stands for its cardinality.

One of measures constructed on the notion of information entropy, is mutual information $I(\cdot, \cdot)$. Mutual Information of two random variables \mathcal{O} and \mathcal{R} ($I(\mathcal{O}, \mathcal{R})$), as Gelfand and Yaglom [27] defined, is the amount of information of \mathcal{O} contained within the variable \mathcal{R} and may be represented with Eq. 10.

$$I(\mathcal{O}, \mathcal{R}) = \sum_{j \leq m} \sum_{i \leq n} \frac{||\mathcal{R}_i \cap \mathcal{O}_j||}{||\mathcal{D}||} \log \frac{||\mathcal{R}_i \cap \mathcal{O}_j|| \cdot ||\mathcal{D}||}{||\mathcal{O}_j|| \cdot ||\mathcal{R}_i||} \quad (10)$$

Mutual Information evaluates many-to-many relations, unlike classification-based measures relying on one-to-one correspondences. Hence Mutual Information provides insight into segmentation result by means of many-to-many relations. Another method, inspired by information entropy was introduced by Meilă. She derived measure of Variation of Information (VoI) dedicated for comparing clusterings. Meilă [28] defined VoI as loss and gain of information while switching from clustering \mathcal{O} to the clustering \mathcal{R} .

$$VoI(\mathcal{O}, \mathcal{R}) = H(\mathcal{O}) + H(\mathcal{R}) - 2I(\mathcal{O}, \mathcal{R}) \quad (11)$$

where $H(\cdot)$ is an information entropy (Eq. 9) and $I(\cdot, \cdot)$ is Mutual Information (Eq. 10).

Although, VoI is not directly dependent on the number of points $\|\mathcal{D}\|$, value of Variation of Information is constrained with upper bounds according to (12). The best possible value of VoI, for the perfect clustering ($\mathcal{O} = \mathcal{R}$) is zero. Moreover, it is true metric, unlike Mutual Information.

$$VoI_{max}(\mathcal{O}, \mathcal{C}) \leq \log \|\mathcal{D}\| \quad (12)$$

Superiority of VoI over classification- based measures is the fact that no one-to-one correspondence has to be found a priori. It produces reliable results even in case of significant clusters' granulation and partial overlapping. Hence it may be thought of as fuzzy decision about clusters correspondences rather than binary: *yes* or *no*.

3 New Quality Measures

Since we focus on evaluation of planarity detection methods, by example of Li et al. [3] approach, four distinguished measures were used.

1. Variation of Information - (VoI)
2. ordinary classification statistics (used in [3]) - (OCS)
3. micro- and macro- weighted classification statistics in terms of overlapping size (see Subsect. 3.1) - (WCS)
4. micro- and macro- averaged planarity statistics (see Subsect. 3.2) - (PS)

Two last of them, weighted classification statistics (WCS) and planarity statistics (PS), are newly derived ones as none of the reviewed authors exploited them and it became the contribution of this paper.

3.1 Weighted Classification Statistics

Weighted classification statistics (WCS) measure is influenced by the number of common part between a reference cluster and the best fitting resulting cluster not being associated yet. In the Fig. 2 we may see correspondence found between a reference cluster (dashed border) and an output cluster (solid line rectangle). The correspondence is found with maximum common-part strategy. In that image (Fig. 2), TP is the cardinality of the inner white region, strips signify region whose cardinality is said to be FP, and the number of points belonging to grey region pose the number of FN.

To clearly state it, for each reference cluster $R_i \in \mathcal{R}$ the set of remaining output clusters is searched to identify the cluster $\mathcal{O}_j \in \mathcal{O}$ whose the largest number of points lies within \mathcal{R}_i . Having found one-to-one correspondence between the reference \mathcal{R}_i and the output cluster \mathcal{O}_j , TP, FP, and FN are calculated. True positives are thought of as the common part between corresponding clusters. The number of false positives (Eq. 13) is calculated as a difference between sum

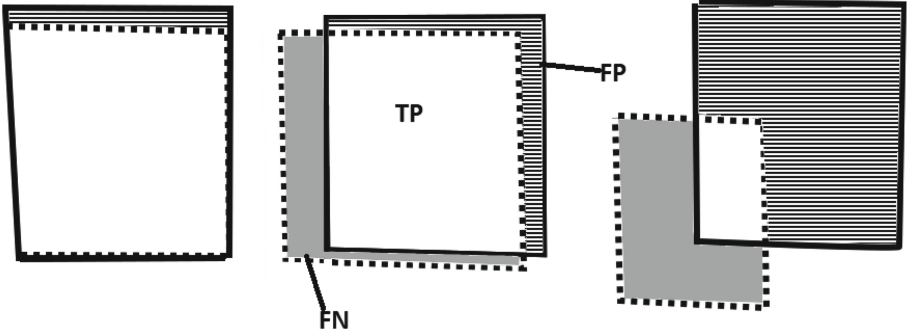


Fig. 2. Idea of WCS measure. The dashed lines state for a reference cluster, solid line borders identify output cluster. White region is said to be TP, grey area- FN, and striped region- FP

of cardinalities of the output clustering $\mathcal{O}_j \in \mathcal{O}'$ (assuming some clusters of \mathcal{O} may be rejected $\mathcal{O}' \subset \mathcal{O}$) and the number of TP. False negatives (Eq. 14) are calculated as a difference of the whole set cardinality $||\mathcal{D}||$ and the number of TP. This way of calculating local clustering characteristics gives us appraisal of an individual result influenced by overlapping size.

Contrary to the OCS, a significance of correspondence in the case presented in the Fig. 1 will be properly diminished with respect to the size of overlapping part. In the Fig. 2 one may see, that value of WCS will vary much whereas OCS may still indicate the same values.

$$||FP|| = \bigcup_{\mathcal{O}_j \in \mathcal{O}'} ||\mathcal{O}_j|| - ||TP|| \quad \text{where } \mathcal{O}' \subset \mathcal{O} \quad (13)$$

$$||FN|| = ||\mathcal{D}|| - ||TP|| \quad (14)$$

3.2 Planarity Statistics

Planarity statistics (PS) measure supplies the estimation of actual planar fragments detection without penalty for division of fragments which constitute the one actual plane (Fig. 3). Penalty is put only for those points of an output cluster which exceed a reference plane and those of an output cluster which do not have their counterpart in a reference one. It identifies TP, FP, and FN as in the Fig. 3, where inner white region describes TP, strips signify FP, and gray region- FN. Everything under the condition that overlapping part of a single output cluster has to be at the level of, at least, 50% to consider a part of a cluster as TP.

4 Results and Discussion

Experiments were carried out for the dataset Room-1 [31] down-sampled to the size of 121,988 points, and for Room-2 [31] built of 292,997 points.

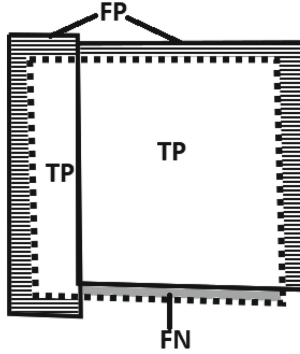


Fig. 3. Idea of planarity statistics (PS) appraisal. The dashed lines state for a reference cluster, solid line borders identify output clusters

Two reference clusterings, for each room dataset, were manually labelled. The detailed clusterings identified as much planar fragments as it was perceptually justified, whereas general clusterings, resembled the reference clusterings of Li et al. [3].

For Room-1, the first, more exact and detailed clustering, including significantly more planar clusters, contained 73 groups of points, whereas the second one - general, only 10 clusters. For Room-2, detailed clustering had 32 groups of points, and general included only 7 clusters. Experiment configuration was like that used in the Li's method [3] (see Table 1) and reflected semantic characteristics of considered rooms. Detailed clusterings reflected all perceptually perceived planar fragments, whereas general, less detailed, clusterings reflected mainly the largest planar fragments. Different number of clusters, though possible, were not considered.

Table 1. Values of parameters of the method by Li et al. [3] used during experiments

Parameter	Value for Room-1	Value for Room-2
Cell size	0.5	0.5
te	0.01	0.02
max. plane shift	0.08	0.076
max. angle deviation between normals	15°	15°

Analysing quality measures, applied for classification tests performed for the planarity detection procedure introduced in [3], we may notice how considered measures reflected selected aspects of the datasets classification procedure. As the output of our experiments, we have examined four classification quality measures: Variation of Information (VoI), ordinary classification statistics (OCS), and those introduced in this paper: weighted classification statistics (WCS), and

planarity statistics (PS). Experiments were conducted for the general clustering of the reference sets (see Table 2) - similar to that used by Li et al., and for the detailed clustering of the reference sets (see Table 4). We have also compared values of OCS and WCS for general clusterings with switched values of te - clustering method parameter, as to evaluate usability of the proposed measures (Table 3).

Table 2. Values of measures for Room-1 and Room-2 datasets for Li’s method [3] obtained by applying general clustering and suggested values of te ($te = 0.01$ for Room-1 and $te = 0.02$ for Room-2)

Measure		Value for Room-1	Value for Room-2	
VoI		1.01	0.70	
OCS	Correctness	90.00%	44.44%	
	Completeness	75.00%	57.14%	
	Quality	69.23%	33.33%	
WCS	Correctness	Micro-averaged	85.68%	75.12%
		Macro-averaged	74.50%	74.73%
	Completeness	Micro-averaged	95.61%	81.36%
		Macro-averaged	89.03%	80.12%
	Quality	Micro-averaged	82.44%	64.10%
		Macro-averaged	72.21%	71.49%
PS	Correctness	Micro-averaged	92.02%	91.96%
		Macro-averaged	86.69%	91.43%
	Completeness	Micro-averaged	82.95%	59.56%
		Macro-averaged	98.25%	99.97%
	Quality	Micro-averaged	77.39%	56.62%
		Macro-averaged	85.18%	91.42%

Values of OCS for general clustering of reference sets are close to these, presented by Li et al. [3]. On the other hand, the OCS values for detailed clustering of reference sets (see Table 4) differ much, but it is clear, since we provide much more detailed clustering. We may think that all output clusters have found their reference counterpart. But there are many planar fragments skipped by the algorithm [3], like a chair or a suitcase. Hence some reference clusters do not have their corresponding clusters in the output, what obviously leads to low completeness indicator. On the other hand, WCS values seems to vary a lot from OCS (Table 4). Correctness values of WCS, lower than that of OCS, point out that corresponding reference and output clusters do not overlap perfectly and may be shifted with respect to each other. On the other hand, very high completeness values inform us about low number of FN. This might be interpreted in such a way that larger planes found proper counterpart and smaller reference clusters were not associated.

Table 3. Values of measures for Room-1 and Room-2 datasets for Li's method [3] obtained by applying general clustering and switched values of te ($te = 0.02$ for Room-1 and $te = 0.01$ for Room-2)

Measure		Value for Room-1	Value for Room-2
VoI		1.34	2.16
OCS	Correctness	100.0%	66.67%
	Completeness	66.67%	57.14%
	Quality	66.67%	44.44%
WCS	Correctness	Micro-averaged	84.90%
		Macro- averaged	74.68%
	Completeness	Micro-averaged	86.80%
		Macro-averaged	84.58%
	Quality	Micro-averaged	75.19%
		Macro-averaged	69.42%
PS	Correctness	Micro-averaged	92.17%
		Macro-averaged	89.12%
	Completeness	Micro-averaged	72.61%
		Macro-averaged	93.81%
	Quality	Micro-averaged	68.40%
		Macro-averaged	83.67%

Table 4. Values of measures for Room-1 and Room-2 datasets for Li's method [3] obtained by applying detailed reference clustering

Measure		Value for Room-1	Value for Room-2
VoI		1.69	2.78
OCS	Correctness	80.00%	55.55%
	Completeness	10.96%	15.62%
	Quality	10.67%	13.88%
WCS	Correctness	Micro-averaged	70.85%
		Macro-averaged	67.24%
	Completeness	Micro-averaged	99.69%
		Macro- averaged	99.02%
	Quality	Micro-averaged	70.70%
		Macro-averaged	67.13%
PS	Correctness	Micro-averaged	83.78%
		Macro-averaged	81.46%
	Completeness	Micro-averaged	71.62%
		Macro-averaged	99.94%
	Quality	Micro-averaged	62.91%
		Macro-averaged	81.43%

High values of PS for Room-1, both micro- and macro- averaged, suggest that planar fragments, mainly, do not contain many disturbing points from other planar patches. For Room-2 we may see poorer values of micro-averaged completeness and quality of PS. From this, we may suspect that the Li's method [3] finds only basic planes in Room-2 like a wall, a ceiling, or a floor.

Let us compare values of measures between Tables 2 and 3, where values of te were switched. Values of OCS indicate increase of correctness and fall of completeness and quality. These values indicate that, approximately, one more correct correspondence was found (TP) but number of undetected reference clusters (FN) grew. WCS values give us more information. For Room-1, micro- and macro- averaged correctness remain virtually equal, what suggests that actually numbers of TP and FP have not changed much- the same correspondences were found. On the other hand, fluctuations in WCS completeness and quality tell us that number of FN increased for larger fragments (clusters of more points). For Room-2 we may see slightly different tendency. Values of correctness and quality increased and the completeness remained the same. This would suggest that $te = 0.01$ for Room-2 suits better than $te = 0.02$. However, values of OCS and PS indicate something opposite. Actually, fewer points from larger and smaller clusters were grouped correctly.

Having analysed results presented in the Tables 2 and 4, several conclusions may be withdrawn. First of all, ordinary classification statistics might be useful for tasks of clusters counting, for instance, how many roofs terrestrial laser scan contains or how many potential walls we have in our indoor scan. OCS provides quantitative evaluation of an output clustering. On the other hand, if we would like to know how well our resulting clusters fit a reference set, we need a qualitative measure, supported by WCS. This measure allows a researcher to construct a method that focuses on maximizing overlapping parts for major clusters, whereas penalty put onto undetected small regions is accordingly smaller. This measure may be valuable for compression purposes, where we expect an algorithm to reduce size, most of all of the greatest regions, preserving at the same time the highest possible quality. Measuring PS, in turn, gives us insight into the process of space division. Regardless to the approach we use for space division, either hierarchical or uniform, it has to supply sufficiently small patches that only one plane is contained therein. Planarity statistics let us appraise whether space was divided enough to enable then the proper segments aggregation.

Comparing values of both - general and detailed clusterings (respectively, Tables 2 and 4), one may noticed that values of OCS differ a lot, when number of reference clusters has changed. The opposite tendency we may see for our measures: WCS and PS, which indicate quantitative evaluation, which regardless to the reasonable changes of number of reference clusters, point out similar assessment of the method.

5 Conclusions

In this paper, we presented two new measures used for sophisticated assessment of planes detection methods, namely: weighted classification statistics and planarity statistics. These measures were compared with two, the most popular ones. The results of the performed experiments indicate benefits of the introduced measures, because they focus on different aspects of segmentation and supplement classical approach. Whereas OCS may sometimes indicate better results, WCS clearly suggest deterioration of clustering. Planarity statistics show how good planes we found, regardless to the fact how many of them were considered. Thanks to that, we may estimate if partition is sufficient to cover each plane. On the other hand, weighted classification statistics provide us with the information how big regions were found correctly. Since each measure has its own application, we provided also recommendations concerning using particular measures for dedicated purposes.

References

1. Bhanu, B., Lee, S., Ho, C., Henderson, T.: Range data processing: representation of surfaces by edges. In: *Proceedings - International Conference on Pattern Recognition*, pp. 236–238. IEEE Press, New York (1986)
2. Golovinskiy, A., Funkhouser, T.: Min-cut based segmentation of point clouds. In: *2009 IEEE 12th ICCV Workshops, Kyoto*, pp. 39–46 (2009)
3. Li, L., Yang, F., Zhu, H., Li, D., Li, Y., Tang, L.: An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells. *Remote Sens.* **9**(5), 433 (2017)
4. Wang, Y., et al.: Three-dimensional reconstruction of building roofs from airborne LiDAR data based on a layer connection and smoothness strategy. *Remote Sens.* **8**(5), 415 (2016)
5. Xu, B., Jiang, W., Shan, J., Zhang, J., Li, L.: Investigation on the weighted RANSAC approaches for building roof plane segmentation from LiDAR point clouds. *Remote Sens.* **8**(1), 5 (2015)
6. Liu, Y., Xiong, Y.: Automatic segmentation of unorganized noisy point clouds based on the Gaussian map. *Comput. Aided Des.* **40**(5), 576–594 (2008)
7. Lu, X., Yao, J., Tu, J., Li, K., Li, L., Liu, Y.: Pairwise linkage for point cloud segmentation. *ISPRS Ann. Photogrammetry Rem. Sens. Spat. Inf. Sci.* **3**(3), 201–208 (2016)
8. Romanowski, A., Grudzien, K., Chaniecki, Z., Wozniak, P.: Contextual processing of ECT measurement information towards detection of process emergency states. In: *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, pp. 291–297. IEEE (2013)
9. Wosiak, A., Zakrzewska, D.: On integrating clustering and statistical analysis for supporting cardiovascular disease diagnosis. In: *Annals of Computer Science and Information Systems*, pp. 303–310. IEEE Press, Lodz (2015)
10. Półrola, M., Wojciechowski, A.: Real-time hand pose estimation using classifiers. In: *Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594*, pp. 573–580. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33564-8_69

11. Staniucha, R., Wojciechowski, A.: Mouth features extraction for emotion classification. In: 2016 Federated Conference on Computer Science and Information Systems, pp. 1685–1692. IEEE Press, Gdansk (2016)
12. Forczmanski P., Kukharev G.: Comparative analysis of simple facial features extractors. *J. Real-Time Image Process.* **1**(4), 239–255 (2007)
13. Skuza, M., Romanowski, A.: Sentiment analysis of Twitter data within big data distributed environment for stock prediction. In: Federated Conference on Computer Science and Information Systems, pp. 1349–1354. IEEE (2015)
14. Martinović, A., Knopp, J., Riemenschneider, H., Gool, L.V.: 3D all the way: semantic segmentation of urban scenes from start to end in 3D. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4456–4465. IEEE Press, Boston (2015)
15. Nguyen, A., Le, B.: 3D point cloud segmentation: a survey. In: 013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM), pp. 225–230. IEEE Press, Manila (2013)
16. Oehler, B., Stueckler, J., Welle, J., Schulz, D., Behnke, S.: Efficient multi-resolution plane segmentation of 3D point clouds. In: Jeschke, S., Liu, H., Schilberg, D. (eds.) ICIRA 2011. LNCS (LNAI), vol. 7102, pp. 145–156. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25489-5_15
17. Chmielewski, L.J., Orłowski, A.: Hough transform for lines with slope defined by a pair of co-primes. *Mach. Graph. Vis.* **22**(1/4), 17–25 (2013)
18. Rabbani, T., van den Heuvel, F.A., Vosselman, G.: Segmentation of point clouds using smoothness constraint. *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.* **36**, 248–253 (2006)
19. Sappa, A.D., Devy, M.: Fast range image segmentation by an edge detection strategy. In: 3-D Digital Imaging and Modeling, pp. 292–299. IEEE Press, Quebec (2001)
20. Schnabel, R., Wahl, R., Klein, R.: Efficient RANSAC for point-cloud shape detection. *Comput. Graph. Forum* **26**(2), 214–226 (2007)
21. Strom, J., Richardson, A., Olson, E.: Graph-based segmentation for colored 3D laser point clouds. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2131–2136. IEEE Press, Taipei (2010)
22. Vaskevicius, N., Birk, A., Pathak, K., Schwertfeger, S.: Efficient representation in 3D environment modeling for planetary robotic exploration. *Adv. Robot.* **24**(8–9), 1169–1197 (2010)
23. Vosselman, G., Gorte, G.H., Sithole, G., Rabbani, T.: Recognizing structures in laser scanner point cloud. *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.* **36**(8), 33–38 (2003)
24. Xiao, J., Zhang, J., Adler, B., Zhang, H., Zhang, J.: Three-dimensional point cloud plane segmentation in both structured and unstructured environments. *Robot. Auton. Syst.* **61**(12), 1641–1652 (2013)
25. Puchała, D.: Approximating the KLT by maximizing the sum of fourth-order moments. *IEEE Sig. Process. Lett.* **20**(3), 193–196 (2013)
26. Shannon, C.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948)
27. Gelfand, I.M., Yaglom, A.M.: Calculation of the Amount of Information about a Random Function Contained in Another Such Function. American Mathematical Society, Washington (1959)
28. Meilă, M.: Comparing clusterings by the variation of information. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT-Kernel 2003. LNCS (LNAI), vol. 2777, pp. 173–187. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45167-9_14

29. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press, Cambridge (2008)
30. Sebastiani, F.: Machine learning in automated text categorization. *ACM Comput. Surv.* **34**(1), 1–47 (2002)
31. Rooms UZH Irchel Dataset. <http://www.ifi.uzh.ch/en/vmml/research/datasets.html>
32. Awrangjeb, M., Fraser, C.S.: An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LIDAR data. *IEEE J. Sel. Topics App. Earth Observ. Remote Sens.* **7**(10), 4184–4198 (2014)
33. Saval-Calvo, M., Azorin-Lopez, J., Fuster-Guillo, A., Garcia-Rodriguez, J.: Three-dimensional planar model estimation using multi-constraint knowledge based on k-means and RANSAC. *Appl. Soft Comput.* **34**, 572–586 (2015)
34. Wagner, S., Wagner, D.: *Comparing Clusterings - An Overview*. Universität Karlsruhe (TH), Karlsruhe (2007)