

# H-RANSAC: A HYBRID POINT CLOUD SEGMENTATION COMBINING 2D AND 3D DATA

\*A. Adam, E. Chatzilari, S. Nikolopoulos, I. Kompatsiaris

Information Technologies Institute, Center of Research and Technology Hellas, Thessaloniki, Greece-  
{katadam, ehatzi, nikolopo, ikom}@iti.gr

Commission II/ WG 3

**KEY WORDS:** point cloud, point cloud segmentation, image segmentation, SfM data, RANSAC, building segmentation

## ABSTRACT:

In this paper, we present a novel 3D segmentation approach operating on point clouds generated from overlapping images. The aim of the proposed hybrid approach is to effectively segment co-planar objects, by leveraging the structural information originating from the 3D point cloud and the visual information from the 2D images, without resorting to learning based procedures. More specifically, the proposed hybrid approach, H-RANSAC, is an extension of the well-known RANSAC plane-fitting algorithm, incorporating an additional consistency criterion based on the results of 2D segmentation. Our expectation that the integration of 2D data into 3D segmentation will achieve more accurate results, is validated experimentally in the domain of 3D city models. Results show that H-RANSAC can successfully delineate building components like main facades and windows, and provide more accurate segmentation results compared to the typical RANSAC plane-fitting algorithm.

## 1. INTRODUCTION

3D segmentation is the task of partitioning a point cloud into consistent regions. Typically, it relies 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 features are not able to separate different objects belonging to the same surface (in our case the same plane). On the other hand, 2D segmentation may be able to separate such coplanar objects, by exploiting their differences in visual appearance. It is obvious that 3D and 2D data contain complementary information that, if effectively combined, can provide more accurate segmentation results. However, until recently, most of the existing works address the 3D segmentation problem as a detection of structural objects, by adopting three main methodological steps; a) model-fitting (i.e. detecting objects of specific shape such as planes), b) region-growing (i.e. starting from a seed point, grow the region with neighboring points and stop when a certain criterion is met) and c) clustering-based (i.e. cluster the points into objects based on a set of extracted features).

However, the aforementioned methodologies typically rely only on the structural information of the 3D point cloud, discarding completely the visual information. This visual information is becoming more and more available with the popularity of obtaining 3D point clouds through photogrammetry, combining multiple overlapping images. In fact, the use of highly accurate and dense 3D point clouds resulting from multiple overlapping images using Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) algorithms has become an extremely attractive solution for 3D documentation. The advantage of these datasets is that they contain both 3D and 2D information for the scene, with an explicit structural connection between the two spaces.

Motivated by the above, in this paper we propose a hybrid point cloud segmentation algorithm that jointly considers the structural

information of 3D representations and the visual information of 2D images. The proposed hybrid segmentation method, H-RANSAC, aims to successfully segment co-planar objects, that cannot be segmented based solely on geometric features. More specifically, H-RANSAC extends the well-known RANSAC algorithm, RANSAC (Fischler and Bolles, 1981), in this case used for plane-based segmentation, by further optimizing its results using the output of k-means-based segmentation applied on 2D images. RANSAC is an iterative model-fitting algorithm, aiming to robustly estimate the parameters of a mathematic model. In our case, we initially use RANSAC to compute the parameters of a mathematical model describing a plane. Subsequently, we use this model on the 3D point cloud to identify the computed plane. H-RANSAC, on the other hand, integrates the 2D segmentation output in each iteration, by first identifying the most favorable 2D image (i.e. the image that depicts the whole plane and diminishes occlusions and radial distortions) from the overlapping images. Then, it applies k-means segmentation to the selected 2D image and finally refines the 3D segmentation so that all points belonging to the same 3D plane correspond to pixels that belong to the same 2D object in the segmented image.

We validate the benefit of the proposed hybrid segmentation approach in the domain of 3D building segmentation, that is becoming more and more popular due to advancements in the field of UAVs and mobile mapping systems. More specifically, we show experimentally that H-RANSAC can successfully achieve building delineation, i.e. the segmentation of buildings into its individual parts such as main facade, windows, etc., while the typical 3D segmentation based on RANSAC cannot distinguish between these objects. In addition, it is important to note that one great advantage of H-RANSAC compared to learning-based segmentation methodologies, is that the segmentation of co-planar objects is achieved without any prior knowledge or annotated data for training, which is particularly useful for analyzing city models that typical come without any ground-truth information.

The rest of the paper is organized as follows. Section 2 reviews

\*Corresponding author

the related work, and focuses on the methodologies combining 3D and 2D segmentation. Section 3 describes in detail the proposed hybrid procedure. The evaluation protocol of H-RANSAC consisting of two publicly available datasets and three evaluation metrics is described in Section 4. The experimental results are outlined and discussed in Section 5. Finally, Section 6 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 segmentation approach for co-planar objects that jointly considers structural and visual characteristics of the objects, without using any prior knowledge or annotated data for training. Related methodologies working on point cloud segmentation mainly rely on the geometrical features of the scene, while a limited number of works, combine visual and geometrical features. 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.

### 2.1 Point cloud segmentation based on geometry

The typical approaches for point cloud segmentation that may be applied irrespectively of the capturing device and the point cloud generation technique (e.g. LIDAR, terrestrial laser scanner devices, point clouds sourcing from SfM algorithms), solely rely on the point cloud geometry. Such methods may be categorized into three different groups; namely model-fitting, region-growing and clustering approaches, and are explained 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 is capable of detecting both well-defined shapes as well as arbitrary shapes, while RANSAC is capable of robustly computing the parameters of a given mathematical model (i.e. plane) and thus detecting surfaces that can be modeled in mathematical terms. Comparisons between HT and RANSAC have shown that both algorithms are very sensitive to parameter tuning, with RANSAC being able to achieve better segmentation accuracies (Tarsha-Kurdi et al., 2007). Based on HT, the authors of (Vosselman et al., 2004) review plane, sphere and cylinder detection from city models, while, in (Rabani and Van Den Heuvel, 2005) the automatic detection of cylinders in point clouds is examined. Moreover, in (Tarsha-Kurdi et al., 2007) comparative results for plane-based detection, in the context of extracting roofs from buildings, are presented. Concerning RANSAC-based algorithms, the authors of (Boulaassal et al., 2007) extract facades from point clouds by manually tuning the parameters of RANSAC. Other works extend RANSAC by incorporating some extra steps in order to develop a more efficient methodology. For example, (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 used an improved RANSAC methodology to detect roof-tops in point clouds sourcing from airborne laser scanners. After segmentation, 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 aforementioned studies have shown that RANSAC is robust to high noise and

outliers. However, it is based on the strict assumption that all the points belonging to the same pre-defined surface (e.g. plane, sphere, etc.) belong to the same object, which is not true for all types of objects (e.g. cannot distinguish windows from facades).

Region-growing techniques were first introduced in (Besl and Jain, 1988). In this work, the mean Gaussian curvature was computed and resulted in a coarse segmentation, which was afterwards refined through an iterative region-growing procedure based on a variable order bivariate surface-fitting. Also deploying region-growing for 3D segmentation, the authors of (Gorte, 2002) proposed to use Triangulated Irregular Networks (TINs) as seeds and merged triangles with similar angles within a pre-defined distance. Furthermore, in (Vieira and Shimada, 2005), points across edges were eliminated with the help of a curvature threshold. Afterwards, noise reduction was performed and the remaining points served as seeds. In a similar context, the authors of (Nurunnabi et al., 2012) selected the points with the least curvature as seeds and the criterion for merging points was the computation of the normal. Overall, the efficiency of region growing methodologies for point cloud segmentation is strongly correlated with successful seed selection and merging criteria and these methodologies have not shown to be robust enough.

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 and afterwards applying clustering to the points' feature vectors. A typical example is (Biosca and Lerma, 2008), where the point cloud was segmented with the help of fuzzy clustering. For each point features were extracted (including the height difference of a point relative to its neighbors, the unit normal vector of the tangent plane and the distance of the tangent plane to the origin) 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 so as to extract the centers of the clusters and use them afterwards 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 and do not require the selection of seed points. However, their efficiency depends on the quality of the extracted features.

### 2.2 Hybrid segmentation approaches

So far, there is only a limited number of 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 in both domains. 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, in order to identify and classify different coral reefs. In a similar vein, the authors of (Aijazi et al., 2016) transformed the 3D point cloud (object space) into an elevation image, with the help of mathematical morphology, in order to segment urban objects, such as sidewalks and facades of buildings.

More closely related to H-RANSAC are the works that refine the segmentation by incorporating information from the 2D and the 3D space. In this direction, 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 segmentation. Then, image segmentation in the CIE LAB color

space followed, while preventing any merging to occur across depth discontinuities. Segmentation both in 3D object and 2D image space was based on graph theory. 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 was employed to them in order to refine the 2D segmentation results. The hybrid segmentation was conducted by applying a region-growing algorithm based on both structural and visual features. Finally, the authors of (Omializarandi and Saadatseresht, 2013) have also worked on identifying roof segments using this combined segmentation. First, they detect planar segments deploying a robust-fitting approach and afterwards, they adopt a region growing methodology both in the 3D space based on the direction of surface normals, and in the 2D image space based on intensity values.

In a similar vein, there have been also methods combing 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. 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). Finally, concerning solely facade parsing, different algorithms are presented in (Martinovic et al., 2015), (Gadde et al., 2017) and in (Jampani et al., 2015). These approaches aim at semantically segmenting the different parts of a facade into doors, windows, balconies etc, with the help of annotated data combining 2D and 3D features with learning-based approaches.

Contrary to H-RANSAC, these approaches incorporate learning-based approaches and rely on annotated data, while H-RANSAC attempts 3D segmentation without any prior knowledge or annotated data. The most similar approaches to H-RANSAC are the ones that do not require manual annotations. As seen from above, these methods typically rely on region-growing segmentation with features originating from both spaces. Compared to what has been proposed so far, this paper presents an extended version of RANSAC, which eliminates the main drawback of a baseline segmentation method based solely on RANSACs robust plane parameter computation (i.e. mis-segmenting co-planar segments) by incorporating visual information from the 2D space. The greatest advantage of H-RANSAC is that objects differing both in texture and in geometry can be separated. This is particularly useful in building delineation, where structural components of buildings are typically co-planar.

### 3. METHODOLOGY

H-RANSAC, consists of the following steps (Figure 1). First, different planes are computed with the help of RANSAC (Fischler and Bolles, 1981). The planes are extracted from the point cloud that has been generated from the overlapping images. RANSAC is an iterative algorithm, where in each iteration aims to compute the parameters of a given mathematical model, ensuring that the solution presents the maximum inliers (i.e. in our case as inliers we consider the points assigned to the plane based on the distance between the points and the plane). In finding the plane with maximum inliers, RANSAC identifies a set of candidate planes by

executing the same algorithm multiple times with different initial starting conditions. In order to incorporate the information from the 2D space, H-RANSAC modifies the candidate planes based on the 2D segmentation results. More specifically, for each candidate plane, H-RANSAC selects the most favorable view of the plane from the overlapping images, assuring that it depicts the whole plane, while radial distortions and occlusions are minimized. 2D segmentation based on k-means is applied on the RGB and spatial features of the optimal image and the plane is updated based on a consistency check. This check ensures that all points belonging to the same plane, also belong to the same segmented object in the 2D space. Then, the plane with the maximum number of inliers from the modified candidate planes is selected and removed from the point cloud. This process is repeated iteratively, starting with RANSAC based segmentation on the remaining point cloud, until there are no points left. Finally, after the 3D point cloud has been segmented, H-RANSAC refines the segmentation output by dismissing the small planes with points less than a predefined threshold and assigns them to neighboring objects.

#### 3.1 Point cloud segmentation

For the detection of planes in the 3D point cloud, RANSAC forms multiple random subsets of three points which define the candidate planes. For each subset of three points, the distance of the remaining points from the plane is computed and the points with a distance lower than a predefined threshold are considered as inliers.

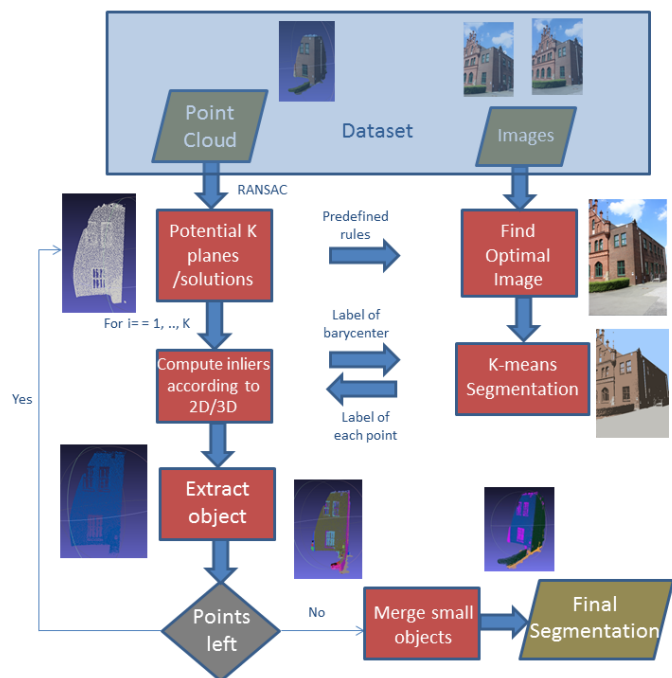


Figure 1. The flowchart of H-RANSAC

#### 3.2 Image selection

Since the point cloud has resulted from overlapping images, each 3D region of interest is depicted in more than one image. Our objective is to select the most favorable image for 2D segmentation, i.e. the one that has the best chances of detecting the desired object. For this, the most favorable image is selected so as to depict the whole identified plane, while radial distortions and occlusions

are minimized. Towards this goal, first we have to identify all the images that depict the whole examined plane.

In order to ensure that the whole plane is depicted, the 3D bounding box of the plane is computed and the four corners are projected onto the images through equation 1, using the exterior orientation of each frame. Equation 1 defines the well-established relationship between the 2D and 3D space, known as the collinearity condition. If the projected points fall within the dimensions of the image, the 3D region of interest is depicted in the image and the specific image is selected as a candidate for the image segmentation procedure.

$$\begin{bmatrix} X - X_o \\ Y - Y_o \\ Z - Z_o \end{bmatrix} = \lambda \cdot \mathbf{R} \cdot \begin{bmatrix} x - x_o \\ y - y_o \\ -c \end{bmatrix} \quad (1)$$

where  $X, Y, Z$  refer to the 3D coordinates and  $x, y$  to the 2D coordinates.  $X_o, Y_o, Z_o$  represent the reference center of the 3D space and are equal to the linear parameters of the exterior orientation of each image, while  $x_o, y_o, c$  stand for the principal point coordinates and the focal length of the camera respectively. Finally,  $\lambda$  refers to the image scale and  $\mathbf{R}$  to the rotation matrix.

After having identified the set of candidate images depicting the 3D object, it is important to select the most favorable view with respect to 2D segmentation, by overcoming the problem of unfavorable perspectives (i.e. extremely oblique views). The selection of the most favorable image is based on two rules. The first one depends on the distance between the projected center of gravity of the 3D object and the center of the image. More specifically, the preferred images are those for which the aforementioned distance is minimized, in order to reduce the effects of radial distortion. Furthermore, the most suitable view has to be as vertical as possible, to diminish occlusions. Thus, the angle between the camera axis and the normal of the examined planar object has to be minimized.

### 3.3 Image segmentation

Subsequently, the step of 2D segmentation is performed using k-means clustering. k-means is an unsupervised way of clustering, initially proposed by (MacQueen, 1967), which classifies the input data into k classes, according to the distances between them. While clustering is mainly performed based on RGB values, in our case, both RGB and spatial features are employed to ensure spatial proximity of the clusters. Thus, each pixel is associated with five features in total consisting of the RGB values and the x,y location of each pixel, forming the following feature vector.

$$\mathbf{f} = [r \quad g \quad b \quad x \quad y] \quad (2)$$

The pixels are clustered around centroids  $\mu_i \forall i = 1 \dots k$ , which are obtained by minimizing the objective function:

$$V = \arg \min \sum_{i=1}^k \sum_{f_j \in S_i} \|\mathbf{f}_j - \mu_i\|^2 \quad (3)$$

where there are  $k$  clusters  $S_i, i = 1, 2, \dots, k$  and  $\mu_i$  is the centroid or mean point of all the points  $f_j \in S_i$

### 3.4 Merging and refinement

A final refinement is applied to the 3D segmentation results in order to merge small objects that comprise of few points. To this end, the known DB-SCAN algorithm is applied (Density-based clustering), as in (Ester et al., 1996). With the application of DB-SCAN multiple clusters for each object are computed. Subsequently, distances are measured between the different cluster centers and small segments (comprised of points fewer than a predefined threshold) are assigned to the object that the closest cluster belongs to. The application of DB-SCAN algorithm is a crucial step towards the refinement of the final outcome, as an object with a uniform label might comprise of multiple not spatially connected segments.

### 3.5 H-RANSAC

Finally, H-RANSAC is presented in Algorithm 1.

---

#### Algorithm 1 H-RANSAC

---

- 1: **while** Point cloud has remaining points **do**
  - 2:     Extract candidate planes based on RANSAC (k candidate planes)
  - 3:     **for**  $i = 1 : k$  **do**
  - 4:         Compute the barycenter  $b_i$  and project it to the most favorable image (Eq. 1).
  - 5:         Identify the most favorable image (Section 3.3)
  - 6:         Remove the 3D points from the plane that are not in the 2D segment containing the projection of the barycenter
  - 7:          $Inliers_i = \#points$  in the plane
  - 8:     **end for**
  - 9:     Select the plane with the most inliers and remove it from the point cloud
  - 10: **end while**
  - 11: Apply DBSCAN and merge small objects
- 

## 4. EXPERIMENTAL PROTOCOL

### 4.1 Dataset

For the assessment of the proposed approach the ISPRS / EuroSDR Benchmark for Multi-Platform Photogrammetry<sup>1</sup> and the Gerrard-hall theater dataset<sup>2</sup> have been used. A subset of 20 and 99 images were selected respectively for each dataset. 3D models have been formed using Pix4D mapper<sup>3</sup> and are depicted in Figures 2(a) and 3(a) respectively. Even though the aim of creating the ISPRS dataset was to assess the accuracy and reliability of methodologies for calibration and image orientation computation, as well as the success of different dense matching algorithms, this benchmark was deemed as one of the most appropriate for evaluating the results of our methodology, as it is widely established and was captured with state-of-the art sensor data. Similarly, the aim of creating the Gerrard-hall theater dataset was to assess the performance of indoor and outdoor alignment methods. However, the dataset was selected for evaluation purposes, as it consists of multiple overlapping images depicting a building, which constitute the dataset appropriate to measure the success of our proposed methodology.

<sup>1</sup>[http://www2.isprs.org/commissions/com1/icwg15b/benchmark/Benchmark\\_Aim.html](http://www2.isprs.org/commissions/com1/icwg15b/benchmark/Benchmark_Aim.html)

<sup>2</sup><http://www.cvg.ethz.ch/research/indoor-outdoor/>

<sup>3</sup><https://pix4d.com/>

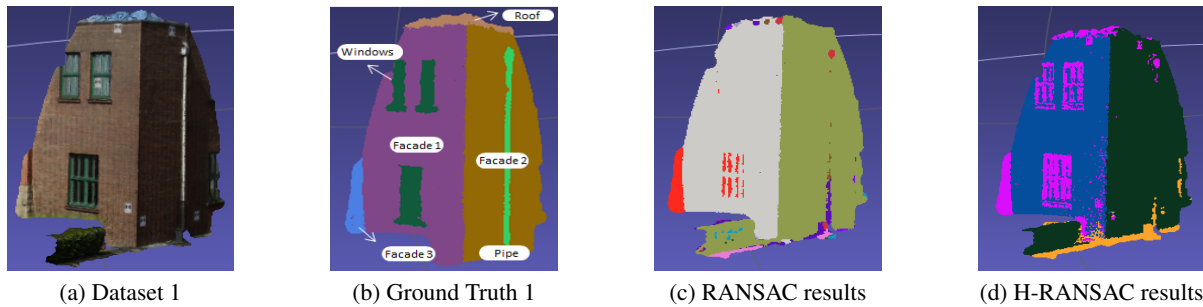


Figure 2. ISPRS / EuroSDR Benchmark for Multi-Platform Photogrammetry: a) original point cloud, b) segments obtained after manual segmentation, c) segments obtained using only geometry features, d) segments obtained using both geometry and visual features

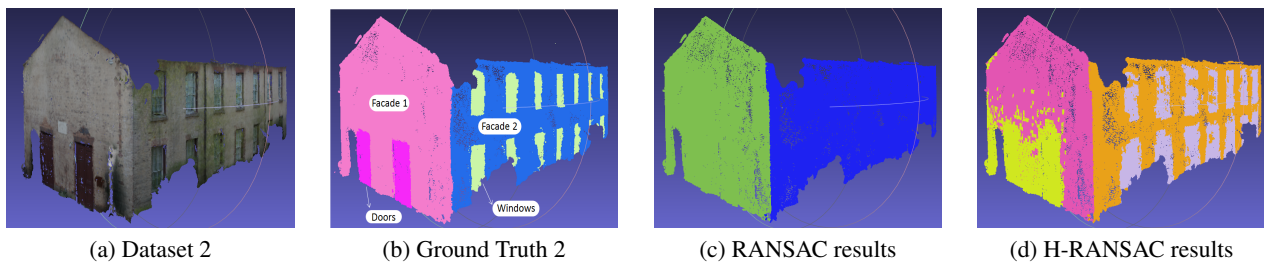


Figure 3. Gerrard-hall theater dataset: a) original point cloud, b) segments obtained after manual segmentation, c) segments obtained using only geometry features, d) segments obtained using both geometry and visual features

## 4.2 Evaluation metrics

The success of the proposed segmentation methodology is quantified by comparing the resulting segments with reference data, i.e. ground truth resulting from manual segmentation, with the help of various evaluation metrics. Several metrics have been used for 3D segmentation, including Global Consistency Error (GCA) and Local Consistency Error (LCE), as proposed in (Martin et al., 2001). Consistency error is also introduced in (Chen et al., 2009) along with other metrics such as Cut Discrepancy, Rand Index and Hamming Distance. While (Vo et al., 2015) and (Yan et al., 2014) use Precision, Recall and F1-measure, as originally introduced in the field of information retrieval (Manning, 1995).

In line with (Vo et al., 2015) and (Yan et al., 2014) we have decided to use Precision, Recall and F1 measure in our experiments. Precision refers to the percentage of correctly retrieved elements, as it results from the number of True Positives (TP), divided by the sum of True Positives (TP) and False Positives (FP):

$$precision = \frac{|TP|}{|TP| + |FP|} \quad (4)$$

In the context of measuring 3D segmentation results, precision may be declared as follows. Let  $m_j$  be a segmented object in the ground truth dataset and  $a_i$  its extracted corresponding object from the proposed methodology. Any point  $p_k \in (a_i \cap m_j)$  is counted as True Positive and any point  $p_k \in (a_i \setminus m_j)$  is counted as a False Positive. In other words, the false positives represent all points in segment  $a_i$  that do not have a corresponding point in the reference segment  $m_j$ .

Recall is the percentage of the ground truth elements that were correctly retrieved and is sensitive to the existence of ground truth segments that were not successfully detected. In mathematical terms, recall is expressed as:

$$recall = \frac{|TP|}{|TP| + |FN|} \quad (5)$$

Where TP refers to the True Positives and FN to the False Negatives. As explained with  $m_j$  being the ground-truth segmented object and  $a_i$  its corresponding object extracted from the proposed methodology, FN is calculated as the number of points where  $p_k \in (m_j \setminus a_i)$ . Thus, False Negatives are the points in the reference segment  $m_j$  without a correspondence in  $a_i$ .

Finally, the F1-score combines the two afore-mentioned measures and may be used as a unique index, showing the capacity of the proposed methodology. F1-measure is computed from Equation (6):

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (6)$$

## 4.3 Ground truth data

The ground truth data was segmented manually, in order to assess the success of the proposed hybrid approach. In generating the ground truth segmentations we have used the convention of clearly identifying the different structural components of a building, such as windows, doors, main facades, roofs, pipes, etc. In Figures 2(b) and 3(b), the manually segmented structural components are depicted, for Datasets 1 and 2 respectively. These ground truth segments were used compare the performance of RANSAC and H-RANSAC, with the help of Precision, Recall and F1-score, as defined in Section 4.2.

## 4.4 Identification of corresponding object

From the above, it is clear that in order to assess the efficiency of the proposed methodology, it is necessary to identify the

corresponding objects between the algorithm-produced and the manually-generated segments. Thus, a pair for segments  $(a_i, m_j)$  are identified as corresponding if  $m_j$  mostly overlapped  $a_i$  and  $a_i$  mostly overlapped  $m_j$ .

## 5. EXPERIMENTAL RESULTS

This section presents quantitative and qualitative results, comparing H-RANSAC with a baseline consisting exclusively of the RANSAC plane-fitting segmentation algorithm. It is important to note that performance is measured on the basis of successfully segmenting the different parts of the building facade, without assigning any semantic labels on the identified segments. However, in order to make the presentation of our results easier for the reader, we refer to the different segments of the scene with their semantic labels.

### 5.1 Qualitative results

Figure 2 presents the results for Dataset 1, consisting of the original dataset 2(a), the ground truth segmentation 2(b), the results of RANSAC plane-fitting segmentation 2(c) and the results of H-RANSAC 2(d). RANSAC threshold was set through trial and error over the available data, making sure to avoid over-segmentation. It is evident from Figure 2(c) that the RANSAC based approach failed to separate the important structural components of the buildings, i.e. the windows, as they are almost coplanar with the main facades of the buildings. Moving on to the results of H-RANSAC, our hybrid procedure has successfully segmented the structural components of the buildings, since the windows have been successfully separated from the main building.

With respect to Dataset 2, results are depicted in Figure 3, consisting of the original dataset 3(a), the generated ground truth 3(b), as well as the results of RANSAC plane-fitting 3(c) and H-RANSAC 3(d) segmentation. The RANSAC based segmentation has segmented Dataset 2 into the two main facades, failing to delineate the doors and the windows of the building. On the other hand, after visual inspection of Figure 3(d), it is evident that H-RANSAC successfully delineated parts of the structural components of the building (i.e. the doors, the windows and the two main facades).

### 5.2 Quantitative Results

Quantitative results are presented in Table 1 (Dataset 1) and Table 2 (Dataset 2) for both the H-RANSAC and RANSAC methods. Since there is no one-to-one correspondence between the ground truth segments and the segments generated by RANSAC or H-RANSAC, some ground truth segments may not have a corresponding part in the segments produced by the algorithms. In these cases, values for the evaluation metrics are not available, and are marked with (-).

Concerning the results tabulated in Table 1 and with respect to the facade 1 of the building, the RANSAC based segmentation presents lower precision, as expected from the visual inspection, since a lot of points belonging to the windows are segmented as points of the facade 1 (False Positives). Contrary, H-RANSAC appears to have identified less FP, as it has successfully segmented bigger parts of the windows and thus achieves higher precision for facade 1. On the other hand, the recall of RANSAC plane-based detection is higher than the hybrid procedure, since

H-RANSAC has segmented larger segments from the facade 1, some of them being False Negatives.

The performance of both algorithms is rather similar for facade 2, while most differences are observed in the case of windows. H-RANSAC achieves better segmentation of windows and thus presents a much higher recall with only marginal reduction of the precision. This is reflected in the value of F1-measure that is considerably higher than RANSAC. Furthermore, with respect to the other structural components, H-RANSAC has not segmented the water pipe, the roof and the facade 3 as they have been assigned to larger objects of the scene, during the merging step. However, as we can see in Table 1 the scores achieved by RANSAC for these structural components are particularly low, which makes them inappropriate for making any comparisons. Summarizing the results obtained for Dataset 1, we can see that both algorithms perform similarly in detecting facades 1 and 2 and very poorly (not suitable for comparisons) for roof, water pipe, and facade 3. Their main difference derives from the ability to segment windows, which constitute one of the most important structural component of city buildings.

In a similar vein, the results for Dataset 2 are presented in Table 2. Concerning the two main facades, H-RANSAC achieves higher precision than RANSAC, as the latter returns a lot of False Positives (i.e. points belonging to the doors and windows that are assigned to the two main facades). On the other hand, the RANSAC based segmentation achieves a higher recall rate for the two main facades, similarly to Dataset 1, since H-RANSAC is characterized by more False Negatives. As also observed in Dataset 1, the main advantage of H-RANSAC is the successful segmentation of doors and windows. More specifically and with respect to F1-scores, H-RANSAC and RANSAC perform similarly concerning the segmentation of the main facades. On the other hand, H-RANSAC outperforms the conventional plane-fitting method when referring to the main structural components of the building, such as windows and doors. Indeed, H-RANSAC succeeds in delineating the doors and the windows with an F1-score of 0.7656 and 0.5991 respectively, compared to a total failure of the baseline approach.

Table 1. Evaluation metrics for Dataset 1

	H-RANSAC			RANSAC		
	Precision	Recall	F1	Precision	Recall	F1
Facade 1	0.9164	0.8146	0.8625	0.8208	0.9054	0.8610
Facade 2	0.6798	0.9768	0.8017	0.6876	0.9952	0.8133
Windows	0.5027	0.6736	0.5757	0.5364	0.1880	0.2784
Roof	-	-	-	0.1311	0.0375	0.0583
Water pipe	-	-	-	0.1768	0.1489	0.1617
Facade 3	-	-	-	-	-	-

Table 2. Evaluation metrics for Dataset 2

	H-RANSAC			RANSAC		
	Precision	Recall	F1	Precision	Recall	F1
Facade 1	0.7999	0.7075	0.7509	0.5661	0.9634	0.7134
Windows	0.7085	0.8238	0.7656	-	-	-
Facade 2	0.9990	0.6650	0.7988	0.8233	0.9676	0.8896
Doors	0.4329	0.9722	0.5991	-	-	-

## 6. CONCLUSIONS

This paper proposes a novel approach for segmenting point clouds that result from multiple overlapping images. H-RANSAC performs segmentation incorporating information from both

the 3D and 2D space and combines the results in the final outcome. By leveraging geometrical and visual features, objects belonging to the same mathematical surface (in our case planes), can be successfully delineated in different segments. The overall effectiveness of H-RANSAC has been evaluated in the domain of city buildings and facade segmentation. More specifically, doors and windows constitute a typical challenge in building delineation, as they are coplanar with building main facades. Experimental results, outlined in Section 5, have proved the effectiveness of H-RANSAC in successfully segmenting facades into different parts without resorting to learning-based approaches. Concerning the limitations of our approach and our intentions for future improvements, in this experimental study we have been mainly concerned with small-scale objects that typically fit in the view of a single image. However, this may not be the case for large-scale objects that extend across the view of two or more images. To overcome this shortcoming, we aim to create and use panoramic images covering large-scale objects that will be generated by performing automatic stitching of multiple images. Moreover, in order to avoid trial and error in setting the RANSAC threshold on the distance of the points considered as inliers, we plan to investigate how to automatically set this value by applying different thresholds and examining the residuals of each threshold. Finally, H-RANSAC could be easily extended for segmenting objects belonging to spheres, curves and other mathematical surfaces.

#### ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union H2020 Horizon Programme (2014-2020) under grant agreement 665066, project DigiArt (The Internet Of Historical Things And Building New 3D Cultural Worlds).

#### 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.
- 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.
- 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.
- Chen, X., Golovinskiy, A. and Funkhouser, T., 2009. A benchmark for 3d mesh segmentation. In: *Acm transactions on graphics (tog)*, Vol. 28number 3, ACM, p. 73.
- 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.
- Ester, M., Kriegel, H.-P., Sander, J. and Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Knowledge Discovery and Data Mining*, Vol. 96number 34, pp. 226–231.
- 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.
- 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.
- 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 Moorec, K., 2014. 3d reconstruction, segmentation and classification of corals from aerial images final report: Cs 231a.
- 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.
- 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.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1number 14, Oakland, CA, USA., pp. 281–297.
- Manning, D., 1995. Introduction. In: *Introduction to Industrial Minerals*, Springer, pp. 1–16.
- Martin, D., Fowlkes, C., Tal, D. and Malik, J., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, Vol. 2, IEEE, pp. 416–423.
- 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.
- Nurunnabi, A., Belton, D. and West, G., 2012. Robust segmentation in laser scanning 3d point cloud data. In: *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on*, IEEE, pp. 1–8.
- 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.

- 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.
- 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.
- 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.
- 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.
- Vo, A., Truong-Hong, L., Laefer, D. and Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp. 88–100.
- 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.
- Yan, J., Shan, J. and Jiang, W., 2014. A global optimization approach to roof segmentation from airborne lidar point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 94, pp. 183–193.