



# Geometry and Colour Based Classification of Urban Point Cloud Scenes Using a Supervised Self-Organizing Map

ERIC KENNETH MATTI & STEPHAN NEBIKER, Muttenz, Schweiz

**Keywords:** terrestrial laser scanning, fast point feature histogram, self-organizing maps, HSV colour space

**Summary:** This paper presents a robust approach for directly labelling textured 3D points within complex urban scenes. The approach is primarily based on the specific exploitation of various colour and geometry based point features, namely by calculating the HSV colour values, a fast point feature histogram (FPFH), and the zenith angle to the surface normal per point. The geometrical point features are thereby calculated over two different levels of neighbourhood regions in order to accommodate point density variation. This results in a 71-dimensional feature vector per point, which is used as input for a supervised point classification using a previously trained self-organizing map (SOM). Investigations of the proposed method show, that a 3D point cloud of a real complex urban laser scanning scene can be classified with good to very good accuracy for the object classes “road”, “building façade” and “vegetation” but with an inferior performance for the class “tree trunk / branch”.

**Zusammenfassung:** *Geometrie- und farbbasierte Punktwolkenklassifizierung von urbanen Laserscanningszenen mittels überwachter SOM-Klassifikation.* In diesem Beitrag wird eine robuste Methode zur direkten Klassifizierung texturierter 3D-Punkte innerhalb komplexer urbaner Laserscanningszenen vorgestellt. Die Methode basiert auf farblichen und geometrischen Charakteristiken der Punkte. So wird für jeden Punkt sein jeweiliger HSV-Farbwert, sein Fast Point Feature Histogramm und sein Zenitwinkel zur Flächennormalen berechnet. Um dabei variable Punktdichten zu berücksichtigen, werden die geometrischen Punktcharakteristika für zwei unterschiedlich große Nachbarschaftsregionen berechnet. Daraus resultiert ein 71-dimensionaler Featurevektor pro Punkt als Input für die überwachte Punktklassifikation mittels einer trainierten Self-Organizing Map (SOM). Untersuchungen des vorgeschlagenen Verfahrens an einer realen komplexen urbanen Laserscanningszene zeigen auf, dass die Methode in der Lage ist, Punkte nach den Objektklassen Straße, Vegetation und Gebäudefassade mit einer guten bis sehr guten Klassifikationsgenauigkeit zu klassifizieren, jedoch eine geringere Qualität bei der Baumstamm/-ast-Klassifikation erreicht wird.

## 1 Introduction

The urban space can nowadays be effectively represented by means of dense textured 3D point clouds derived from the recordings of mobile or non-mobile terrestrial laser scanning and image based systems. Apart from the use of such rich point clouds for interactive 3D visualization and measuring purposes (NEBIKER et al. 2010), point clouds are widely used as a basis to manually, semi-automatically or at

best automatically derive explicit 3D models composed of points, lines, surfaces and volumes.

With recent progress in large scale terrestrial surveying, e.g. by means of mobile or portable TLS (Terrestrial Laser Scanning) systems (VENNEGEERTS 2010), the automatic derivation of non-geometrical and geometrical object information has become increasingly important in order to efficiently exploit the recorded scenes, such as in order to create various

products, such as 3D city models, or to further understand a recorded scene. Although dense textured 3D point clouds have high geometric and radiometric information content, the automatic robust derivation of further semantic or geometric information and models is not easily achieved. Furthermore, the complex structure of different objects, their arrangement within the world as well as point cloud properties, e.g. outliers in measurements, variations in point density and occlusions, have resulted in the development of various specialized object detection and labelling methods. Thus, among others, algorithms have been developed for extracting geometric primitives (SCHNABEL et al. 2007, HOHMANN et al. 2009, WAN & SHARF 2012), for directly classifying single points (LIM & SUTER 2007, MUNOZ et al. 2009, RUSU et al. 2009, BRODU & LAGUE 2012, SHAPOVALOV et al. 2010, XIONG et al. 2011, NIEMEYER et al. 2012), for detecting objects (PATTERSON et al. 2008, GOLOVINSKIY et al. 2009, STEDER et al. 2009, TEICHMAN et al. 2011, MONNIER et al. 2012, VELIZHEV et al. 2012) and for segmenting point clouds, e.g. by the means of region-growing algorithms (ZHAN et al. 2009), graph-based algorithms (LAI et al. 2009, SCHOENBERG et al. 2010, STROM et al. 2010) or clustering algorithms (JIANG 2004, LIU et al. 2007, BIOSCA & LERMA 2008, ZHUANG et al. 2008).

In this paper, a method for directly labelling 3D points is presented, which exploits a combination of geometric and radiometric properties. It enables the classification of textured points within a complex urban laser scanning scene according to a finite number of disjoint object classes. The method is primarily based on the specific calculation of various colour and geometry based point features, namely the calculation of the HSV colour values, a fast point feature histogram (FPFH) and the zenith angle of the surface normal per point. Similar to BRODU & LAGUE (2012) the geometrical point features are calculated over two different levels of neighbourhood regions in order to consider point density variations. In doing so, the FPFH values allow to determine if a point is part of a surface, sphere, cylinder, edge or corner, whereas the zenith angle of the surface normal is calculated in order to receive an absolute inclination measure of the underlying surface. In addition to deriving

these geometrical point features, the RGB colour values per point are transformed into the HSV colour space in order to receive decorrelated and thus robust colour values. The resulting feature vector is used as input for a point-based classification, using a previously trained self-organizing map (SOM).

In the following section, related work regarding the direct classification of points, also known as point labelling, is presented. After presenting our method and the used algorithms in section 3, investigation results, namely the analysis of the expressiveness of the used point features and the obtainable classification accuracy, are presented in section 4. In the last section, conclusions and an outlook on possible future optimization of the proposed method are given.

## 2 Related Work

Object detection within point clouds can be viewed as the process of assigning different distinct point subsets to an object class of the real world. In order to distinguish these object classes, a set of geometrical or non-geometrical features characterizing the properties of points belonging to these classes can be analysed. Thus, by calculating features for each point and using a classification method which labels a point according to its features one can implicitly create the point subsets belonging to different object classes. According to BIOSCA & LERMA (2008), the local point features used should represent the surrounding area as well as possible. In addition, the choice of features also greatly depends on the objects to be distinguished from one another.

In particular, LIU et al. (2007) and ZHUANG et al. (2008) both utilize the 3D position, surface normal, mean and Gaussian local curvature for their different segmentation methods, namely a fuzzy-clustering- and SOM-based segmentation. In presence of a textured point cloud, ZHAN et al. (2009) and SAREEN et al. (2010) both present a region-growing segmentation method solely based on the colour values of each point, whereas SCHOENBERG et al. (2010) and STROM et al. (2010) use the colour values in addition to the surface normal per point in order to achieve a more robust result

of their graph-based segmentation. In addition to these rather simple point features, there has been considerable work done on making local shape 3D features more discriminative and robust. Amongst others, PATTERSON et al. (2008) use spin images (JOHNSON & HEBERT 1999) and extended Gaussian images (HORN 1984) for detecting cars, RUSU et al. (2009) use fast point feature histograms (FPFH) in order to assign points to different surfaces, edges or corners, BRODU & LAGUE (2012) present a multi-scale eigenvalue-based feature in order to characterize the 1D/2D/3D properties of the local scene at each point and at different scales while FLITTON et al. (2012) present a novel 3D extension to the visual cortex model, previously used in 2D object recognition (SERRE et al. 2005, MUTCH & LOWE 2008), for recognizing various objects in 3D volumetric imagery. Furthermore, inspired by the SIFT (LOWE 2004) image feature, SKELLY & SCLAROFF (2007) present the rotation invariant feature transform (RIFT) and TOMBARI et al. (2010) the Signature of Histograms of Orientations (SHOT) feature, both showing promising results for the problem of identifying corresponding points. Finally, we refer to TOMBARI et al. (2010) and ALEXANDRE (2012) for a comparative evaluation and presentation of further 3D point features.

Besides the point feature choice, various classification methods are proposed for directly classifying single points. On the one hand, point independent classification schemes using a linear discriminant analysis (BRODU & LAGUE 2012), a support vector machines (GOLOVINSKIY et al. 2009, RUSU et al. 2009, BRODU & LAGUE 2012) or a random forest (GOLOVINSKIY et al. 2009, SHAPOVALOV et al. 2010) classifier are proposed. On the other hand, classification schemes that consider point context, such as conditional random fields (LIM & SUTER 2007, NIEMEYER et al. 2012, RUSU et al. 2009), high-order associative Markov network (MUNOZ et

al. 2009) as well as non-associative Markov network (SHAPOVALOV et al. 2010) classifiers or stacked 3-D parsing using a simple K-class logistic regression classifier (XIONG et al. 2011) are proposed, in order to generally achieve a higher classification performance than non-point-context considering approaches.

### 3 Method

#### 3.1 Overview

Our system takes an unclassified textured point cloud as input. As a result, it creates a classified point cloud, in which each point is labelled according to the object class it is most likely to belong to. The system proceeds in three steps, as outlined in Fig. 1. First, we manually extract a set of points for each object class to be detected from the given input point cloud, thus yielding a set of training data. Then the local point features are calculated for both, the training and the test data, namely the HSV colour values  $\{h_q, s_q, v_q\}$ , the surface normals  $\{n_q^{(1)}, n_q^{(2)}\}$ , the zenith angles of the surface normal  $\{z_q^{(1)}, z_q^{(2)}\}$  and the 33 features of the fast point feature histogram  $\{fpfh_{33_q}^{(1)}, fpfh_{33_q}^{(2)}\}$  for each point  $p_q$  and the two neighbourhood regions (1) and (2). Finally, the SOM classifier is trained and used to classify each point of the point cloud using the feature vector  $[h_q, s_q, v_q, z_q^{(1)}, z_q^{(2)}, fpfh_{33_q}^{(1)}, fpfh_{33_q}^{(2)}]$  having a dimension of 71. The different steps and the prototype implementation are described in detail in the following sections.

#### 3.2 Transformation of the RGB Colour Values to HSV Colour Values

Based on the works of ZHAN et al. (2009) and SAREEN et al. (2010), which both present a re-

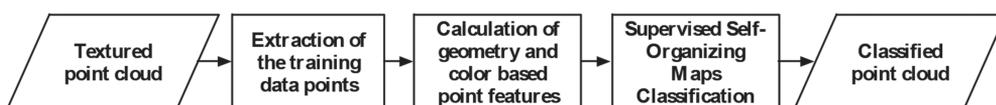


Fig. 1: Schematic representation of the proposed point classifications steps.

gion-growing segmentation method solely using colour values in order to achieve better results compared to geometry-based methods for point clouds with a high degree of outliers and areas of occlusion, as well as on the work of STROM et al. (2010), who also propose to use colour values in addition to surface normals in order to ensure a more robust segmentation performance, we use point colour values to include a non-geometrical local point feature during classification. Due to the high correlation between the colour values in the RGB colour space, the recorded RGB values per point are transformed into the HSV colour space. In this model a colour is primarily characterized by the hue (H) parameter, whereas the saturation (S) and value (V) represent variations of the same hue (H). Thus, resulting in a colour characterization primarily based on a single parameter.

### 3.3 Calculation of the Surface Normals and their Zenith Angles

In order to calculate the zenith angles of the surface normals and the Fast Point Feature Histograms, the surface normal of each point is calculated. The surface normal  $n_q$  of a point  $p_q$  is calculated according to the principal component analysis (PCA) method originally proposed by BERKMANN & CAELLI (1994), by calculating the covariance matrix  $C_q$  according to (1) using the  $k$ -nearest neighbours and by determining its eigenvector  $n_0 = n_q$  corresponding to the smallest eigenvalue  $\lambda_0$ .

$$c_q = \frac{1}{k} \sum_{j=1}^k (p_j - \bar{p})(p_j - \bar{p})^T \quad (1)$$

However, since there is no mathematical way to solve for the sign of  $n_0$ , the inward or outward orientation of the normals with respect to the underlying surface is unknown. Thus, in order to solve this problem, we orient all normals  $n_q$  consistently towards the sensor viewpoint  $v$ , as proposed by RUSU (2009), by determining the sign of  $n_q$  so that the (2) is satisfied.

$$n_q \cdot (v - p_q) > 0 \quad (2)$$

In order to obtain an absolute geometrical local point feature, enabling inference of the absolute inclination of the underlying surface, the zenith angle  $z_q$  of the surface normal  $n_q$  is calculated according to (3).

$$z_q = \arccos \left( \frac{n_q \cdot [0 \ 0 \ 1]^T}{|n_q|} \right) \quad (3)$$

### 3.4 Calculation of the Fast Point Feature Histograms (FPFH)

As a relative geometrical local point feature the fast point feature histogram, as presented in RUSU et al. (2009) and RUSU (2009), is used. The purpose of this feature is to encode local geometrical properties by generalizing the mean curvature around a query point  $p_q$  using a multi-dimensional histogram. The feature is invariant to the pose of the underlying surface and copes very well with different sampling densities or noise (RUSU 2009). In addition to this robustness the FPFH has good discriminating properties. According to the comparative study of ALEXANDRE (2012) the FPFH performs well compared to other state of the art 3D features in object and category recognition and has shown to yield excellent results in classifying points (RUSU et al. 2009). The calculation of the FPFH for a point  $p_q$  based on its neighbouring points  $p_k$  is presented in detail in Alg. 1 and Fig. 2.

1. Calculation of the simplified point feature histogram (SPFH) for point  $p_q$ :

a) Definition of the points  $p_k$  constituting the neighbourhood  $p_q$ , either by taking all points within a certain range or by taking the  $n$ -nearest neighbours (Fig. 2a).

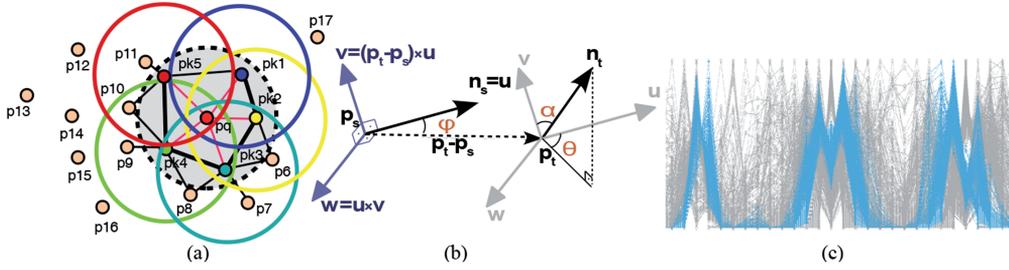
b) Calculation of the  $k$  angles  $(\alpha, \phi, \theta)$  between point  $p_q$  and  $p_k$ :

By defining a DARBOUX  $uvw$  coordinate frame in one of the points  $p_q$  or  $p_k$  (Fig. 2b)

according to,  $u = n_s$ ,  $v = u \times \frac{(p_t - p_s)}{\|p_t - p_s\|_2}$ ,

$w = u \times v$  with

if  $\arccos(n_q p_{kq}) \leq \arccos(n_k p_{qk})$ ,  $p_{kq} = p_q - p_k$ , then  $p_s = p_q$ ,  $n_s = n_q$  and  $p_t = p_k$ ,  $n_t = n_k$  else  $p_s = p_k$ ,  $n_s = n_k$  and  $p_t = p_q$ ,  $n_t = n_q$



**Fig. 2:** Illustration of the calculation of the FPFH for point  $p_q$  adapted from Rusu 2009) in (a) and (b) as well as a parallel coordinate plot of the FPFH's for multiple tree trunk / branch points (highlighted in blue) in (c).

and calculating the angles  $(\alpha, \phi, \theta)$  according to  $\alpha = v \cdot n_t, \phi = u \cdot \frac{(p_t - p_s)}{\|p_t - p_s\|_2},$   
 $\theta = \arctan(w \cdot n_t, u \cdot n_t)$

- c) Creation of the histogram for the  $k$  angles  $(\alpha, \phi, \theta)$ .
2. Calculation of the simplified SPFH for all neighbouring points  $p_k$  as in step 1.
3. Calculation of the FPFH for point  $p_q$  according to:

$$FPFH(p_q) = SPFH(p_q) + \frac{1}{k} \sum_{i=1}^k \frac{1}{w_k} \cdot SPFH(p_k)$$

with  $w_k =$  Distance between point  $p_q$  and  $p_k$ .

**Alg. 1:** Calculation of the FPFH for point  $p_q$  (according to Rusu 2009).

### 3.5 Supervised Self-Organizing Map Classification

In order to implement a precise as well as fast 3D point classification method a self-organizing map (SOM) was used. In contrast to supervised methods, a SOM, in principle being an unsupervised clustering method, demands significantly less computational resources and scales better to large learning problems (HAYKIN 1999). Yet, as shown by VASIGH & KOMPANY-ZAREH (2013), it can achieve a similar classification performance as a more powerful support vector machines classification. SOM can also be used for online learning, i.e. the model can be adapted incrementally as

new training samples are made available. The choice of using a self-organizing map was additionally motivated by works of JIANG (2004) and LIU et al. (2007), which both achieved good results in segmenting point clouds using a SOM, even for noisy data.

The self-organizing map, a competitive self-learning neural network, developed by KOHONEN since 1981, is mainly used to project data of an  $n$ -dimensional feature space onto a two-dimensional regular grid of nodes / neurons. By doing so, the original information is compressed, while the most relevant topological and metrical relationships of primary data patterns are preserved. Thus, entities that lie near to one another within the  $n$ -dimensional feature space will also come to lie close to one another in the SOM (KOHONEN 2001). On the one hand, a SOM can be used for clustering datasets, whereas on the other hand, by applying a slight modification to the algorithm, it can also be used for the classification of datasets. The so-called supervised SOM classification, in principle still an unsupervised learning method, in addition to the  $n$ -dimensional feature vectors takes into account their class labels during training. This allows the resulting trained neuron weights to be assigned to a class and, consequently, a test sample to be classified (see following Alg. 2).

Input data:

- Set of training vectors  $x_{train}^{(i)} = [x^{(i)}, y^{(i)}]$

with vector  $x^{(i)} \in \mathbb{R}^n$  containing the  $N$  point feature values and  $y^{(i)} \in \mathbb{R}^k$  corresponding to the unit label vector with its  $k^{\text{th}}$  component

- set to one if  $x^{(i)}$  belongs to the object class  $k$  and the rest of its components set to zero.
- Set of vectors  $x_{test}^{(ii)} \in \mathbb{R}^n$ .
1. Random or linear initialization of the weight vector  $m^{(j)} \in \mathbb{R}^{n+k}$  associated with each neuron  $j$ .
  2. Training phase of the SOM:
 

For each training step  $t$  from 1 to  $\max\_iterations$ :

    - a) Random selection of a training vector  $x_{train}(t)$  from the set of training vectors.
    - b) Comparison of  $x_{train}(t)$  with all neuron weight  $m^{(j)}$ , in order to determine the best matching neuron  $c$  according to the chosen similarity measure (i.e., the Euclidean distance).
    - c) Modification of the weight vector  $m^{(c)}$  of the best matching neuron  $c$  and the weight vectors  $m^{(j)}$  of the neurons  $j$  within a circular neighbourhood of radius  $\sigma(t)$  of the best matching neuron  $c$  according to:
 
$$m^{(j)}(t+1) = m^{(j)}(t) + h_{cj}(t)[x_{train}(t) - m^{(j)}(t)]$$
 with  $h_{cj} = \alpha(t) \cdot e^{-d_{cj}^2/2\sigma^2(t)}$ : Gaussian neighbourhood function,  $\alpha(t) \in [0, 1]$ : learning rate,  $d_{cj} \in \mathbb{Z}$ : grid distance between neuron  $c$  and respective neuron  $j$  and  $\alpha(t) \in [0, \max\_som\_size]$ : neighbourhood distance.  $\alpha(t)$  and  $\sigma(t)$  are monotonically decreasing functions over time  $t$  in our case linear functions with negative slope (section 4).
  3. Classification phase using the trained SOM:
 

Each test vector  $x_{test}^{(ii)}$  is compared to the  $x$ -part (the first  $N$  components corresponding to the features) of all trained neuron weights  $m^{(j)}$ , in order to determine the best matching neuron  $c$  according to the chosen similarity measure. The point's class membership is then given by the index of the largest component of the  $y$ -part (the last  $K$  components corresponding to the label vector) of  $m^{(c)}$ .

**Alg. 2:** Supervised SOM classification (according to KOHONEN 2001).

### 3.6 Prototype Implementation

A first prototype of the proposed labelling method was implemented by using Python as

a wrapper for calculating the point features using the point cloud library (RUSU & COUSINS 2011) and for performing the supervised SOM classification using Matlab's SOM toolbox (ALHONIEMI et al. 2012).

## 4 Investigations

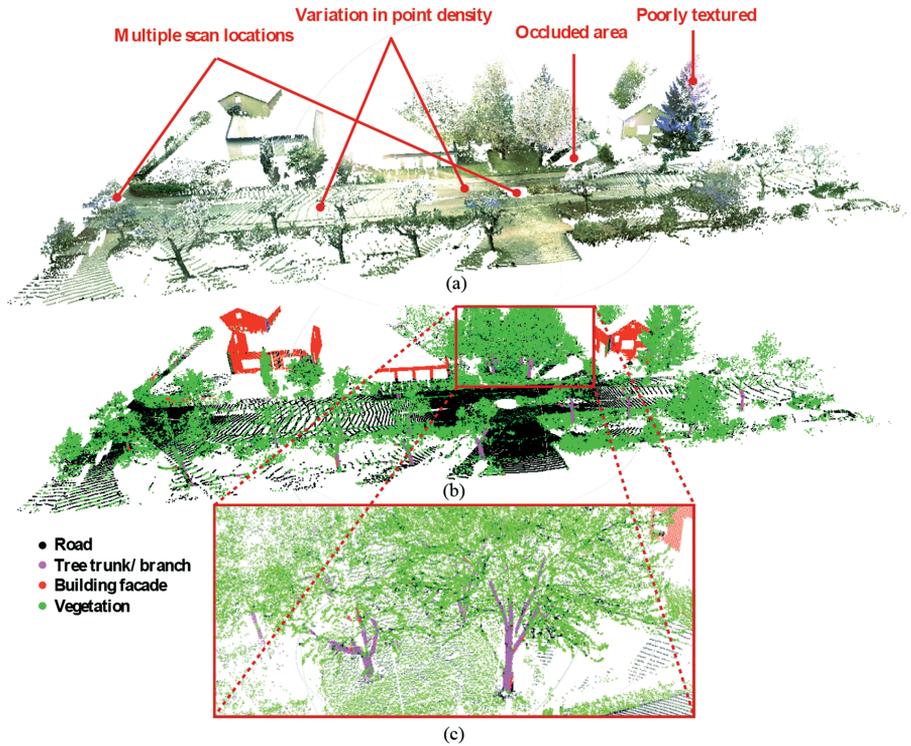
### 4.1 Test, Training and Reference Datasets

For the investigation of the expressiveness of the point features and of the overall classification accuracy, a point cloud of an urban region, recorded from two locations with a Leica ScanStation 2, was used. The recorded scene has the dimension of  $100 \text{ m} \times 70 \text{ m} \times 15 \text{ m}$ , contains over 1 million points and is shown in Fig. 3a. Apart from a high inter and intra object class variance within the scene, e.g. for vegetation, the point cloud contains variations in point density, occluded areas and, due to using the on-board capabilities of the Leica ScanStation 2, is generally poorly textured (see Fig. 3a).

Based upon this unclassified point cloud the training and the reference datasets were created. The training dataset was created by manually extracting small regions corresponding to the object classes "building façade", "road", "vegetation", and "tree trunk / branch". In this process regions of different point densities per object class were selected in order to account for any inter-class variation of the derived point features. The final training dataset consists of 100,000 points. The point count per object class is equal to 10% of the number of points of that class in the reference (see Tab. 2). The reference data was created by manually labelling the rest of the entire unclassified point cloud according to the object classes to be detected.

### 4.2 Expressiveness of the used Point Features

During the feature design phase various graphical exploratory data analysis methods were used, in order to primarily determine whether the selected point features allow



**Fig. 3:** Original textured point cloud (a) with examples of typical problems such as variations in point density, occluded areas and poor texturing. Classified point cloud (b) and a detail (c).

for a distinction between the selected object classes to be detected and which of the features are the most expressive ones. First, a matrix of pairwise scatter plots of the four point features  $h_q$ ,  $s_q$ ,  $v_q$  and  $z_q$  of training data subsets with similar point density was generated. Fig. 4 illustrates that the HSV colour values allow a partial separation of the object classes “building façade” and “road” from the classes “tree trunk / branch” and “vegetation”, but the zenith angle of the surface normal needs to be taken into account in order to separate the object classes “tree trunk / branch” and “vegetation”, overlapping in the HSV colour space. Furthermore, it can be seen, that the zenith angles of the vegetation class have a high variance and that therefore some points nevertheless come to lie within the region mainly occupied by points belonging to the object class “tree trunk / branch”. The reason for this high zenith angle variance is that the vegetation training sample contains both points of

the ground vegetation and leaves on the trees, with the latter vegetation subclass having an especially high zenith angle variance.

For further investigation of the expressiveness of the calculated point features, the values of the 37-dimensional feature vector  $[h_q, s_q, v_q, z_q, fpf, h_{33}]$  for a training data subset with similar point density were visualized using an interactive and dynamic grand tour visualization (ASIMOV 1985, BUJA & ASIMOV 1986). This multi-dimensional visualization technique generally maps the values of an n-dimensional feature space using a linear combination to a lower dimensional feature space. Thus, it is a useful technique for visually analysing the distribution of and for identifying clusters within multivariate data. In order to perform an optimal visual cluster analysis, we projected the values of the 37-dimensional feature vectors into the orthogonal coordinate frame defined by their principal components and determined the projection parameters us-

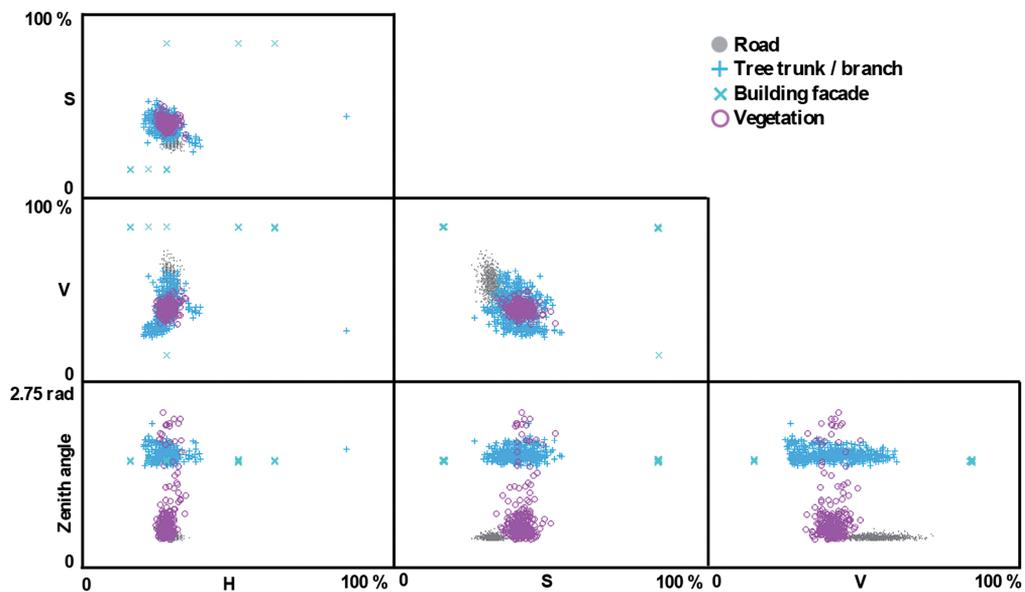


Fig. 4: Scatterplot matrix of the HSV colour values and the zenith angle of the surface normals (using 32 nearest neighbours for the surface normal calculation).

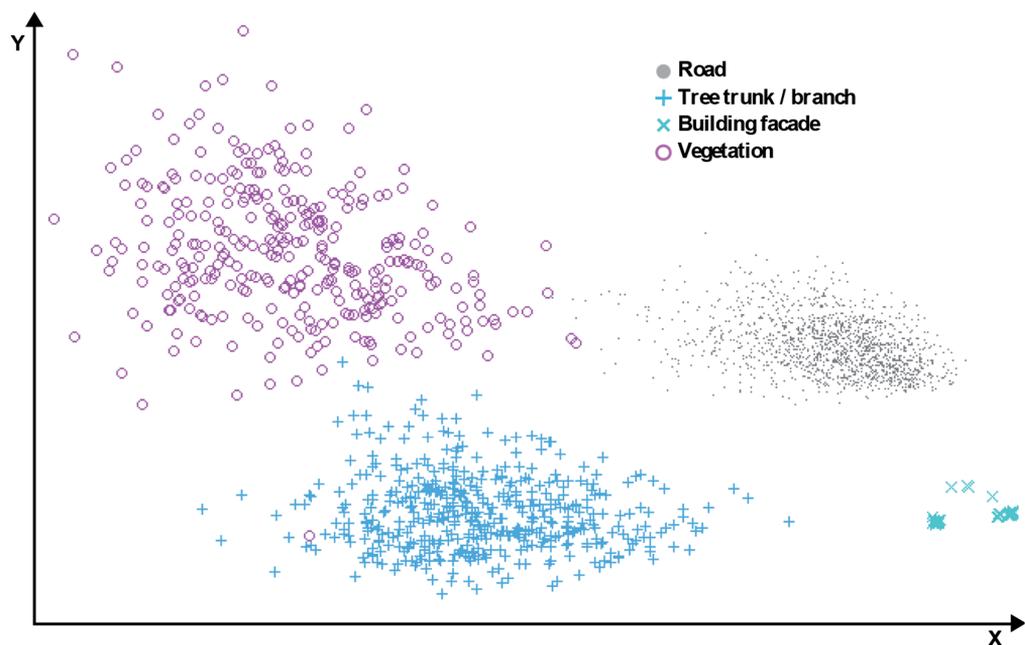


Fig. 5: Grand Tour Visualization of the HSV colour, the zenith angle of the surface normals and the FPFH values per point (using 32 nearest neighbours for the surface normal and FPFH calculation).

ing a Linear Discriminant Analysis (LDA) (COOK et al. 1995, COOK & SWAYNE 2007). Fig. 5 shows that by using the HSV colour, zenith angle and FPFH values, all the points belonging to each object class form a distinct cluster, which is different to the case in Fig. 4, when FPFH were not used. This provides a visual indication that the selected point features are expressive enough to classify the points according to the examined object classes and under the assumption of similar point density.

### 4.3 Classification Accuracy

In order to numerically quantify the achievable classification accuracy for various parameterizations of the proposed point labelling method, a manually labelled reference dataset was created (see section 4.1 for further

details). This enables the calculation of confusion matrices and various quality measures by comparing each point's object class membership resulting from the classification with its actual membership. All of the following classification results are based upon using the supervised SOM classification functionality of Matlab's SOM toolbox (ALHONIEMI et al. 2012). The linear initialized, 100 x 100 sized SOM of hexagonal lattices was thereby trained according to Alg. 2 using a two-phased training scheme and the parameters presented in Tab. 1.

Tab. 2 shows the resulting classification accuracy of a single-scale approach using the HSV, the zenith angle and FPFH values per point as input for the supervised SOM classification. The geometrical features were calculated using the 32 nearest neighbouring points. Tab. 2 shows that the precision for all classes, except for the class "tree trunk / branch", lies

**Tab. 1:** Parameterization of the rough (a) and fine-tuning (b) phase of the SOM training.

Parameter	Value
Max. iterations	15 x number of neurons <sup>(a)</sup> , 60 x number of neurons <sup>(b)</sup>
$\alpha(t)$ (Learning rate)	Linear decreasing function from 0.5 to 0.05 <sup>(a)</sup> and 0.05 to 0.0 <sup>(b)</sup>
$\sigma(t)$ (Neighbourhood radius)	Linear decreasing function from $m/8$ to $m/32$ <sup>(a)</sup> and $m/32$ to 1 <sup>(b)</sup> , with $m$ = maximal SOM side length / SOM size

**Tab. 2:** Confusion matrix and classification accuracy using both colour and geometrical features composed of the HSV values as well as the single set of zenith angles and FPFH values calculated using 32 nearest neighbours.

		Reference data				Total amount of points	Precision (%)
		Road	Tree trunk / branch	Building façade	Vegetation		
Classification	Road	597,217	4,578	15,040	127,935	744,770	80
	Tree trunk / branch	1,903	6,937	238	1,183	10,261	68
	Building façade	25	14	23,131	87	23,257	99
	Vegetation	25,151	12,075	1,264	298,864	337,354	89
Total amount of points		624,296	23,604	39,673	428,069	1,115,642	
Recall (%)		96	29	58	70		
Overall classification accuracy (%)		83					

above 80%. In contrast, the recall for all classes is generally unsatisfying, with only the points of the class “road” having a recall rate of 96%. The analysis of this poor classification result shows that the majority of the misclassifications occur in areas where the point density is sparse and hence the consideration of only 32 neighbouring points while calculating the geometrical features is not sufficient.

Tab. 3, in contrast, shows the classification accuracy achieved in two experiments involving geometrical features obtained from two neighbourhoods. The results shown in bold font were achieved using the HSV values, the two sets of zenith angle values using 8 and 128 nearest neighbouring points and the two sets of FPFH values using 8 and 32 neighbouring points. The results shown in regular font were obtained by using the same point features, but neglecting the HSV values. The results show that the precision and recall rate for all classes, except for the class “tree trunk / branch”, now are above 91% and 87%, respectively. Furthermore, the results show that the additional consideration of colour values results in a more accurate classification, especially for the geometrically similar road and ground vegetation

points. Analysing the relatively poor classification result for the object class “tree trunk / branch” it was found that the majority of misclassifications are due to the challenging problem of correctly distinguishing between leaves and branches for each point while creating the reference dataset (see Fig. 3c). Hence, the classification results for this object class can be considered as too pessimistic. Comparing our best classification results presented in Tab. 3 to other terrestrial point cloud labeling methods with comparable object classes, we achieve a nine percent lower and two percent higher overall classification accuracy than LIM & SUTER (2007) and MUNOZ et al. (2009), respectively. As these results were achieved on different datasets, this comparison is not conclusive. Nevertheless, it shows that a similar classification performance is achieved.

In summary, the proposed point labelling method, using both colour and multi-scale geometrical point features, allows for a moderately accurate classification of points belonging to object class “tree trunk / branch”, while achieving good to very good classification results for the object classes “road”, “building façade” and “vegetation”. Additionally,

**Tab. 3:** Confusion matrix and classification accuracy using both colour and geometrical features composed of the HSV values, the two sets of zenith angles using 8 and 128 nearest neighbours and the two sets of FPFH values using 8 and 32 nearest neighbours (bold font). In comparison, the values in regular font show the results obtained without the colour features.

		Reference data				Total amount of points	Precision (%)
		Road	Tree trunk / branch	Building façade	Vegetation		
Classification	Road	600,257 558,519	4,200 3,993	1,520 1,402	52,203 70,192	658,180 634,106	91 88
	Tree trunk / branch	413 488	10,086 9,885	127 761	4,094 3,225	14,720 14,359	69 69
	Building façade	331 603	1,206 909	36,876 36,719	1,347 1,462	39,760 39,693	93 93
	Vegetation	23,295 64,686	8,112 8,817	1,150 791	370,425 353,190	402,982 427,484	92 83
Total amount of points		624,296	23,604	39,673	428,069	1,115,642	
Recall (%)		96 89	43 42	93 93	87 83		
Overall classification accuracy (%)		91 86					

the presented classification results show two properties of the used features that result in a better overall classification. Firstly, the HSV colour values support a correct classification of geometrically similar object classes and secondly, by calculating the geometrical point features over two levels of neighbourhood regions, point density variations are accounted for.

## 5 Conclusion

This paper presents an approach for classifying textured points within a complex urban point cloud according to a distinct set of object classes. By using the HSV colour values as well as the zenith angle and the Fast Point Feature Histogram (FPFH) values, calculated over two different local neighbourhoods, as point features as an input to a supervised SOM classification, we were able to classify points according to the object classes “road”, “tree trunk / branch”, “building façade” and “vegetation”. Our investigations show that the selected 3D point features are sufficiently expressive for labelling points according to the examined object classes under the assumption of constant point density. Secondly, by using a supervised SOM classifier, the features allow the classification of a poorly textured point cloud affected by occlusion and density variations with a good overall classification accuracy of 91%, but with a relatively poor recall for the challenging object class “tree trunk / branch”. Our work further supports the known fact that the features used for classification have a large influence on achieving optimal results and thus must be designed according to the object classes to be distinguished from one another. In our application we thus selected point features allowing the separation of various objects of different colours, absolute surface inclinations, and shapes within a point cloud of varying point density. Future work will focus on the further development of the proposed method, on combining different features to a multi-scale / multi-modal point feature in dependency of the classification task, by including an approach for automatically determining the size and amount of required neighbourhood regions, e.g. DEMANTKÉ et al.

2011, as well as the features needed for optimal classification.

## References

- ALEXANDRE, A., 2012: 3D Descriptors for Object and Category Recognition: A Comparative Evaluation. – IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal.
- ALHONIEMI, E., HIMBERG, J., PARHANKANGAS, J. & VESANTO, J., 2012: SOM Toolbox. – <http://www.cis.hut.fi/somtoolbox/> (24.4.2012).
- ASIMOV, D., 1985: The Grand Tour: A Tool for Viewing Multidimensional Data. – *SIAM Journal on Scientific Computing* **6** (1): 128–143.
- BERKMANN, J. & CAELLI, T., 1994: Computation of Surface Geometry and Segmentation using Covariance Techniques. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16** (11): 1114–1116.
- BIOSCA, J.M. & LERMA, J.L., 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): 84–98.
- BRODU, N. & LAGUE, D., 2012: 3D Terrestrial Lidar Data Classification of Complex Natural Scenes using a Multi-Scale Dimensionality Criterion: Applications in geomorphology. – *ISPRS Journal of Photogrammetry and Remote Sensing* **68** (0): 121–134.
- BUJA, A. & ASIMOV, D., 1986: Grand Tour Methods: An Outline. – *Computing Science and Statistics* **17**: 63–67.
- COOK, D., BUJA, A., CABRERA, J. & HURLEY, C., 1995: Grand Tour and Projection Pursuit. – *Journal of Computational and Graphical Statistics* **4** (3): 155–172.
- COOK, D. & SWAYNE, D.F., 2007: *Interactive and Dynamic Graphics for Data Analysis*. – With R and Ggobi, Springer.
- DEMANTKÉ, J., MALLET, C., DAVID, N. & VALLET, B., 2011: Dimensionality Based Scale Selection in 3D LIDAR Point Clouds. – *ISPRS Workshop Laser Scanning*: 97–102, Calgary, Alberta, Canada.
- FLITTON, G., BRECKON, T.P. & MEGHERBI, N., 2012: A 3D Extension to Cortex Like Mechanisms for 3D Object Class Recognition. – *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 3634–3641, Los Alamitos, CA, USA.
- GOLOVINSKIY, A., KIM, V.G. & FUNKHOUSER, T., 2009: Shape-Based Recognition of 3D Point Clouds in Urban Environments. – *International*

- Conference on Computer Vision (ICCV): 2154–2161, Kyoto, Japan.
- HAYKIN, S., 1999: *Neural Network*. – Second Edition, MacMillan College Publishing Company, London, UK.
- HOHMANN, B., KRISPEL, U., HAVEMANN, S. & FELLNER, D., 2009: Cityfit: High-Quality Urban Reconstructions by Fitting Shape Grammars to Images and Derived Textured Point Clouds. – 3D Virtual Reconstruction and Visualization of Complex Architectures, Trento, Italy.
- HORN, B.K.P., 1984: Extended Gaussian images. – *IEEE* **72** (12): 1671–1686.
- JIANG, B., 2004: Extraction of Spatial Objects from Laser-Scanning Data using a Clustering Technique. – XXth ISPRS Congress, Technical Commission III: 219–224, Istanbul, Turkey.
- JOHNSON, A. & HEBERT, M., 1999: Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21** (5): 433–449.
- KOHONEN, T., 2001: *Self-Organizing Maps*. – Third Edition, Springer, Heidelberg, Germany.
- LAI, Y.-K., HU, S.-M., MARTIN, R.R. & ROSIN, P.L., 2009: Rapid and Effective Segmentation of 3D Models using Random Walks. – *Computer Aided Geometric Design* **26** (6): 665–679.
- LIM, E.H. & SUTER, D., 2007: Conditional Random Field for 3D Point Clouds with Adaptive Data Reduction. – International Conference on Cyberworlds: 404–408, Hannover, Germany.
- LIU, X.-M., ZHONG, S.-S. & BAI, X.-L., 2007: A Modified SOFM Segmentation Method in Reverse Engineering. – Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing: 570–573, Qingdao, China.
- LOWE, D.G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. – *International Journal of Computer Vision* **60** (2): 91–110.
- MONNIER, F., VALLET, B. & SOHEILIAN, B., 2012: Trees Detection from Laser Point Clouds Acquired in Dense Urban Areas by a Mobile Mapping System. – XXIIInd ISPRS Congress, Technical Commission III: 245–250, Melbourne, Australia.
- MUNOZ, D., VANDAPEL, N. & HEBERT, M., 2009: On-board Contextual Classification of 3-D Point Clouds with Learned High-order Markov Random Fields. – *IEEE International Conference on Robotics and Automation*: 4273–4280, Piscataway, NJ, USA.
- MUTCH, J. & LOWE, D., 2008: Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields. – *International Journal of Computer Vision* **80** (1): 45–57.
- NEBIKER, S., BLEISCH, S. & CHRISTEN, M., 2010: Rich Point Clouds in Virtual Globes – A New Paradigm in City Modeling? *Computers, Environment and Urban Systems* **34** (6): 508–517.
- NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U., 2012: Conditional Random Fields for LIDAR Classification in Complex Urban Areas. – XXI-Ind ISPRS Congress, Technical Commission III: 263–268, Melbourne, Australia.
- PATTERSON, A., MORDOHAJ, P. & DANILIDIS, K., 2008: Object Detection from Large-Scale 3D Datasets using Bottom-up and Top-down Descriptors. – *ECCV 2008: 10th European Conference on Computer Vision*: 553–566, Marseille, France.
- RUSU, R.B., 2009: *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. – Ph.D. Computer Science Department, Technische Universität München, Germany.
- RUSU, R.B., HOLZBACH, A., BLODOW, N. & BEETZ, M., 2009: Fast Geometric Point Labeling using Conditional Random Fields. – *IEEE/RSJ International Conference on Intelligent Robots and Systems*: 3212–3217, St. Louis, MO, USA.
- RUSU, R.B. & COUSINS, S., 2011: 3D is here: Point Cloud Library (PCL). – *IEEE International Conference on Robotics and Automation (ICRA)*: 1–4, Shanghai, China.
- SAREEN, K., KNOFF, G. & CANAS, R., 2010: Rapid Clustering of Colorized 3D Point Cloud Data for Reconstructing Building Interiors. – *ISOT 2010 International Symposium on Optomechatronic Technologies*: 1–6, Toronto, ON, Canada.
- SCHNABEL, R., WAHL, R. & KLEIN, R., 2007: Efficient RANSAC for Point-Cloud Shape Detection. – *Computer Graphics Forum* **26** (3): 1–12.
- SCHOENBERG, J., NATHAN, A. & CAMPBELL, M., 2010: Segmentation of Dense Range Information in Complex Urban Scenes. – *IEEE/RSJ International Conference on Intelligent Robots and Systems*: 2033–2038, Taipei, Taiwan.
- SERRE, T., WOLF, L. & POGGIO, T., 2005: Object recognition with features inspired by visual cortex. – *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2**: 994–1000.
- SKELLY, L.J. & SCLAROFF, S., 2007: Improved Feature Descriptors for 3D Surface Matching. – *SPIE Conference on Two- and Three-Dimensional Methods for Inspection and Metrology V* 6762.
- SHAPOVALOV, R., VELIZHEV, A. & BARINOVA, O., 2010: Non-Associative Markov Networks for 3D Point Cloud Classification. – *ISPRS Technical Commission III Symposium, PCV 2010 – Photogrammetric Computer Vision and Image Analysis*: 103–108, Saint-Mandé, France.

- STEDER, B., GRISETTI, G., VAN LOOCK, M. & BURGARD, W., 2009: Robust On-line Model-based Object Detection from Range Images. – IEEE/RSJ International Conference on Intelligent Robots and Systems: 4739–4744, St. Louis, MO, USA.
- STROM, J., RICHARDSON, A. & OLSON, E., 2010: Graph-Based Segmentation for Colored 3D Laser Point Clouds. – IEEE/RSJ International Conference on Intelligent Robots and Systems: 2131–2136, Taipei, Taiwan.
- TEICHMAN, A., LEVINSON, J. & THRUN, S., 2011: Towards 3D Object Recognition via Classification of Arbitrary Object Tracks. – IEEE International Conference on Robotics and Automation (ICRA): 4034–4041, Shanghai, China.
- TOMBARI, F., SALTI, S. & STEFANO, L., 2010: Unique Signatures of Histograms for Local Surface Description. – 11th European Conference on Computer Vision (ECCV): 356–369, Hersonissos, Greece.
- VASIGHI, M. & KOMPANY-ZAREH, M., 2013: Classification Ability of Self Organizing Maps in Comparison with Other Classification Methods. – MATCH Communications in Mathematical and in Computer Chemistry **70**: 29–44.
- VELIZHEV, A., SHAPOVALOV, R. & SCHINDLER, K., 2012: Implicit Shape Models for Object Detection in 3D Point Clouds. – XXIIInd ISPRS Congress, Technical Commission **III**: 79–184, Melbourne, Australia.
- VENNEGEERTS, H., 2010: Objektraumgestützte kinematische Georeferenzierung für Mobile-Mapping-Systeme. – Ph.D. Fakultät für Bauingenieurwesen und Geodäsie, Leibniz Universität Hannover, Germany.
- WAN, G. & SHARF, A., 2012: Grammar-Based 3D Facade Segmentation and Reconstruction. – Computers Graphics **36** (4): 216–223.
- XIONG, X., MUNOZ, D., BAGNELL, J.A. & HEBERT, M., 2011: 3-D Scene Analysis via Sequenced Predictions over Points and Regions. – IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China.
- ZHAN, Q., LIANG, Y. & XIAO, Y., 2009: Color-Based Segmentation of Point Clouds. – Laserscanning **09**: 248–252, Paris, France.
- ZHUANG, J., LIU, X. & HOU, X., 2008: The Fuzzy Clustering Algorithm Based on Weighted Distance Measures for Point Cloud Segmentation. – 2008 Second International Symposium on Intelligent Information Technology Application (IITA): 51–54, Shanghai, China.

## Address of the Authors:

ERIC KENNETH MATTI & Prof. Dr. STEPHAN NEBIKER, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Institute of Geomatics Engineering, CH-4132 Muttenz, Tel.: +41-61-467-4242, Fax: +41-61-467-4460, e-mail: {eric.matti}{stephan.nebiker}@fhnw.ch

Manuskript eingereicht: Mai 2013  
Angenommen: Februar 2014