2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Performance of Histogram Descriptors for the Classification of 3D Laser Range Data in Urban Environments

Jens Behley, Volker Steinhage and Armin B. Cremers

Abstract— **The selection of suitable features and their parameters for the classification of three-dimensional laser range data is a crucial issue for high-quality results. In this paper we compare the performance of different histogram descriptors and their parameters on three urban datasets recorded with various sensors—sweeping SICK lasers, tilting SICK lasers and a Velodyne 3D laser range scanner. These descriptors are 1D, 2D, and 3D histograms capturing the distribution of normals or points around a query point. We also propose a novel histogram descriptor, which relies on the spectral values in different scales. We argue that choosing a larger support radius and a z-axis based global reference frame/axis can boost the performance of all kinds of investigated classification models significantly. The 3D histograms relying on the point distribution, normal orientations, or spectral values, turned out to be the best choice for the classification in urban environments.**

I. INTRODUCTION

The interpretation of sensor readings is a fundamental ability required by autonomous mobile robots to fulfill different tasks in changing urban environments, such as exploration, localization, and navigation. In this context we are not only interested in different object categories, such as pedestrians, cars, or obstacles in general, but also in surface properties, such as flat, rough or slippery, to deduce drivability characteristics. Particularly the classification of three-dimensional laser range data by rotating laser rangefinders recently received increasing interest, as nowadays devices are available, which offer complete 360° laser range scans of the environment in a fraction of a second.

Especially in outdoor and urban environments, laser rangefinders are the method of choice to generate depth information. Other active or passive sources, such as stereo vision or the yet very popular depth estimation by projection of structured infrared light, e.g., Microsoft's Kinect, do not provide the depth accuracy at long ranges or are just not applicable in an outdoor scenario. A special characteristic of three-dimensional laser range data is the distance-dependent sampling rate of objects, i.e., we get a very dense point cloud of objects and surfaces near the laser range sensor, but only a sparsely sampled point cloud at long range.

We are particularly interested in a point-wise classification, as we do not only want to classify distinct objects with well-defined boundaries, but also surfaces with less clearly defined boundaries, such as ground, vegetation, and tree canopies. Hence, we cannot exploit the range data in terms of first generating a segmentation and then classifying the segments [1], or even use tracking information to segment dynamic objects of interest [2].

Generally, in a classification approach we usually have two components—the *data* and the *model*. In the last years, much scientific work concentrated on the development of more expressive models, e.g., based on Conditional Random Fields [3], [4], [5], similarity-preserving hashing [6], or stacked classification [7]. Nonetheless, to attain robust and high-quality classification results, we also have to consider the data part—the features extracted from the point clouds, which enable the model to predict a label. The classification model and the features are two sides of the same coin: a more complex model can compensate for insufficient features, and better features can compensate for a too simplistic model. A linear classifier with features capable to linearly separate the data ideally should be as effective as a highly non-linear and complex classifier with very simple features.

More precisely, we are interested in investigating and answering the following questions: (1) What do we expect from features to get a robust and high-quality result? (2) Which features are in this sense suitable to classify laser range data in an urban environment? And (3) which parameters are required to attain high-quality classification results?

In this contribution, we answer these questions for the point-wise classification in urban environments. We will investigate the usage of histogram-based local descriptors for this purpose, as they showed promising results in other application areas, such as the matching of laser scans [8] and object recognition [9]. However, we have to cope with a very different situation compared to the mentioned application areas, as we encounter highly varying sampling of objects and surfaces. Thus, we are interested in descriptors relying on the radius neighborhood, i.e., all neighbors in a certain radius, rather than descriptors using the k-nearest neighborhood.

In computer vision and computer graphics several studies on the quality of descriptors for matching and object recognition were presented [10], [11]. Rusu *et al.* [12] evaluated the proposed Point Feature Histograms with different classifiers and segmentation approaches—SVMs with different kernels, k nearest neighbors and k-means with different distance metrics. Still to our best knowledge this is the first thorough investigation of these descriptors in the context of classification of three-dimensional laser range data.

The paper is organized as follows. In Section II we

J. Behley, V. Steinhage, and A. B. Cremers are with the Department of Computer Science III, University of Bonn, 53117 Bonn, Germany.
{behley,steinhag,abc}@iai.uni-bonn.de,

briefly introduce the evaluated histogram-based descriptors. In Section III we discuss the different reference frames that are used. Section IV describes the methodology of the performance evaluation, the datasets, and the classification approaches. In the following Section V we discuss the experimental results. Finally, Section VI gives a conclusion and presents future work.

II. HISTOGRAM DESCRIPTORS

As it is usual in the computer vision community, we will term a feature a *descriptor* when it tries to generate a discriminative description of a laser point rather than only specifying its shape properties. We can further differentiate between *histogram descriptors* and *signature descriptors* [13]. The former ones maintain a histogram of neighboring points or their properties, and the latter calculate some single value of the neighboring points—the *support*. In case of histograms we need a *reference axis* or *reference frame* in which we determine the bin index of the property we want to measure.

Over the last years, a variety of descriptors for matching of point clouds [8], [13], object recognition [9], [14] and classification [15], [5], [4] representing properties of the surrounding surface of one point were proposed. In this section, we briefly introduce histogram descriptors used in some recent work, which emerged to be a good choice for a descriptive representation of laser points in terms of shape and geometry.

We have some special requirements on descriptors for point-wise classification of three-dimensional laser range data. In shape retrieval or object recognition applications, we want to discriminate between different instances of a category of objects, e.g., we want a description that enables us to distinguish between different types of cars. This is often not necessary in classification as we only want to distinguish between different classes or categories, but not single instances. In addition, this description should result in well-separated clusters in the feature space. Another requirement is the robustness to partial occlusions, e.g., scan shadows by other objects. Lastly, we are looking for features that are robust against very different sampling rates of the objects. This is encountered seldom in both shape retrieval applications, where we find similar sampling rates in the database, and indoor object recognition applications, as there we usually encounter near range scans.

The following descriptors discussed in the following paragraph fulfill most of the mentioned requirements. All descriptors are local features following the taxonomy of Tangelder *et al.* [11] as they represent the local neighborhood of a point instead of determining a global description of the whole segmented object. Thus we get a local representation, which is robust against partial occlusions and independent of a given segmentation. As all descriptors use a radius neighborhood $\mathcal{N}_p^\delta$, i.e., all points in a radius δ around the query point $\mathbf{p}$ instead of using the k-nearest neighbors $\mathcal{N}_p^k$, we get a sampling invariant representation by a proper normalization of the feature vectors. The normalization constant will be denoted by η and determined separately for every feature vector. We empirically determined that a normalization by the maximal entry of a feature vector is superior to a normalization by the sum of all entries.

In the following, we will use $\mathbf{r} \in \mathbb{R}^3$ to denote the reference axis and $\mathbf{R} \in \mathbb{R}^{4\times4}$ for the reference frame used to determine the histogram indices.

A. Histogram of Normal Orientations

Triebel *et al.* [15] used a normal histogram storing the angle between the reference axis $\mathbf{r}$ and the normal of a neighboring point $\mathbf{n}_q, \mathbf{q} \in \mathcal{N}_p^\delta$. The histogram descriptor $\mathbf{h} \in \mathbb{R}^b$ with b entries is defined as follows:

$$h_i = \eta \left| \left\{ \mathbf{q} \;\middle|\; \cos\left(\frac{i\cdot\pi}{b}\right) \le \mathbf{r}\cdot\mathbf{n}_q < \cos\left(\frac{(i+1)\cdot\pi}{b}\right) \right\} \right| \tag{1}$$

Regions with a strong curvature result in a uniformly distributed histogram, while flat areas lead to a peaked histogram. The histogram is parameterized by the number of bins b and the size of the support region δ.

B. Spin Images

The spin image [9] by Johnson *et al.* is a very prominent histogram descriptor and is used in several retrieval, matching or classification approaches [16], [7]. A spin image is calculated by spinning a grid around the reference axis $\mathbf{r}$, where the grid cells collect or 'count' the neighboring points $\mathbf{q} \in \mathcal{N}_p^\delta$. An entry of the spin image $\mathbf{SI} \in \mathbb{R}^{b\times b}$ with indexes (i, j) is calculated using the distance to the line defined by $\mathbf{p} + \lambda\cdot\mathbf{r}$ with parameter $\lambda \in \mathbb{R}$, and the distance to the plane originating in $\mathbf{p}$ and normal $\mathbf{r}$. The local coordinates (α, β) in respect to the reference axis are given by $\alpha = \|\mathbf{n_p} \times (\mathbf{q} - \mathbf{p})\|$ and $\beta = \mathbf{n_p}\cdot(\mathbf{q} - \mathbf{p})$ The indexes (i, j) in the image are calculated from α and β by $i = \lfloor \rho^{-1}\cdot\alpha \rfloor$ and $j = \lfloor \rho^{-1}\cdot(\frac{1}{2}\delta - \beta)\rfloor$, where $\rho = \delta b^{-1}$ is the grid resolution of the spin image. The spin images are parameterized by the size b (width and height of the spin image) and the radius of the support δ.

C. Distribution Histogram

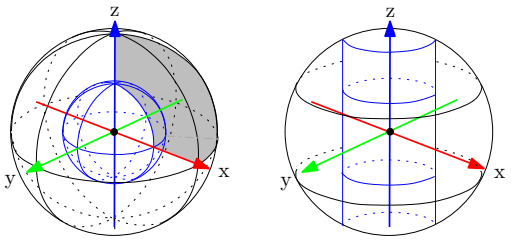
The distribution histogram by Anguelov *et al.* [3] tries to capture the shape around a point in a cube defined by the reference frame $\mathbf{R} \in \mathbb{R}^{4x4}$.

In order to transform a neighboring point $\mathbf{q} \in \mathcal{N}_p^\delta$, the reference frame is inverted, i.e., $\mathbf{q}' = \mathbf{R}^{-1}\mathbf{q}$.

The distribution histogram $\mathbf{h} \in \mathbb{R}^{b\times b\times b}$ is defined as follows:

$$\mathbf{h}_{i,j,k} = \eta \left| \left\{ \mathbf{q}' \;\middle|\; \left\lfloor \frac{3}{2}\cdot\left(\frac{q'}{d} - \mathbf{1}\right)\right\rfloor = \begin{pmatrix} i \\ j \\ k \end{pmatrix} \right\} \right|, \tag{2}$$

where $\mathbf{1} \in \mathbb{R}^3$ denotes the vector that contains only ones.



(a) (b)

Fig. 1. Subdivisions used by (a) the SHOT descriptor and (b) the spectral histogram. The inner sectors or shells are depicted in blue. In (a) one sector of the SHOT descriptor is highlighted in light gray.

D. Signature of Histograms of Orientations (SHOT)

Recently, Tombari *et al.* [13] proposed to use a combination of histograms and signatures. Their descriptor subdivides the space around the query point into sectors (see Figure 1). Then, for every sector a histogram of normal orientations between the neighboring point inside the sector and the query point is calculated.

More precise, the histogram index i of a neighboring point $\mathbf{q} \in \mathcal{N}_p^\delta$ inside a sector is calculated by $\frac{1}{2}(1 + \mathbf{r} \cdot \mathbf{n}_q)b$, where b is the number of bins in the histogram. To reduce quantization errors, a neighboring point also contributes to histograms in neighboring sectors of the subdivision using a quadrilinear interpolation.

The authors suggested to use 8 azimuth divisions, 2 elevation divisions and 2 radial divisions for the subdivision. The remaining parameters of interest are the radius of the support region δ, and the number of bins in the sector histograms b.

E. Spectral Histogram

Motivated by the results of experiments with spectral shape signatures, we propose to use a Spectral Histogram. Similar to the SHOT descriptor, we calculate for every sector of a subdivision three signature values based on spectral values of points falling inside the sector.

The eigenvalues of the covariance matrix of N points $\{p_1, \ldots, p_N\}$ defined by $\mathbf{C} = \frac{1}{N}\sum_i (\mathbf{p}_i - \bar{\mathbf{p}})^T(\mathbf{p}_i - \bar{\mathbf{p}})$, $\mathbf{p} \in \mathbb{R}^3$ with $\bar{\mathbf{p}} = N^{-1}\sum_i p$, encode the general distribution of the points within the support radius. Let $\lambda_0 \leq \lambda_1 \leq \lambda_2$ be the eigenvalues of the covariance matrix $\mathbf{C}$ and $\hat{\lambda}_i = \lambda_i/\lambda_2$ the normalized eigenvalues. A measure of pointness then is defined by $\hat{\lambda}_0$, surfaceness by $\hat{\lambda}_1 - \hat{\lambda}_0$, and linearness by $\hat{\lambda}_2 - \hat{\lambda}_1$ [17], [4].

We subdivide the space around a point in different shells and slices, as shown in Figure 1. Let s be the number of slices, r the number of radial shells, and $\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z$ be the base vectors of the reference frame. Then we add to the local covariance of sector (i, j) the point $\mathbf{q}' = \mathbf{R}^{-1}\mathbf{q}$, if $i = \lfloor \frac{1}{2\delta}(q'^{(3)} + \delta)l \rfloor$ and $j \leq \lfloor \frac{1}{\delta}\|\mathbf{q}' \times \mathbf{v}_z\|s \rfloor$. Hence, every shell collects all points up to its radius.

For every radial shell in every slice, we get a different scale of the point distribution. The descriptor is rotation invariant around the z axis of the reference frame.

III. Reference Frame and Reference Axis

The only question left is the choice of the reference frame or the reference axis, which are needed to calculate the indices in the histograms. We evaluated two canonical choices—the *local reference frame* based on eigenvectors and a *global reference frame* based on the global z-axis.

The local reference frame $\mathbf{R}_{\text{local}} \in \mathbb{R}^{4\times4}$ of a point $\mathbf{p}$ is based on the normalized eigenvectors $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2$ of the covariance matrix of neighboring points $\mathbf{q} \in \mathcal{N}_p^\delta$. From the eigenvectors with eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2$ we can build the following homogeneous transformation:

$$\mathbf{R}_{\text{local}} = \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 & \mathbf{v}_0 & \mathbf{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

A local reference axis $\mathbf{r}_{\text{local}} \in \mathbb{R}^3$ is given by $\mathbf{v}_0$, which corresponds to the point normal $\mathbf{n}_p$ of point $\mathbf{p}$.

The global reference frame can be constructed using the global z-axis denoted by $\mathbf{z}$. We decided to use the normal $\mathbf{n}_p$ to get a rotation invariant reference frame. Following from this we get the global reference frame $\mathbf{R}_{\text{global}} \in \mathbb{R}^{4\times4}$:

$$\mathbf{R}_{\text{global}} = \begin{bmatrix} \frac{(\mathbf{n_p}\times\mathbf{z})\times\mathbf{z}}{\|(\mathbf{n_p}\times\mathbf{z})\times\mathbf{z}\|} & \frac{\mathbf{n_p}\times\mathbf{z}}{\|\mathbf{n_p}\times\mathbf{z}\|} & \frac{\mathbf{z}}{\|\mathbf{z}\|} & \mathbf{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

Thus, only the rotation about the z-axis of the reference frame depends on local information.

Tombari *et al.*[13] proposed to use a weighted version of the covariance for determining the eigenvectors, thus getting a more stable reference frame in point clouds with clutter and also a disambiguation scheme of the directions of the eigenvectors based on the point density. We applied this weighted covariance and the disambiguation scheme only with the SHOT descriptor.

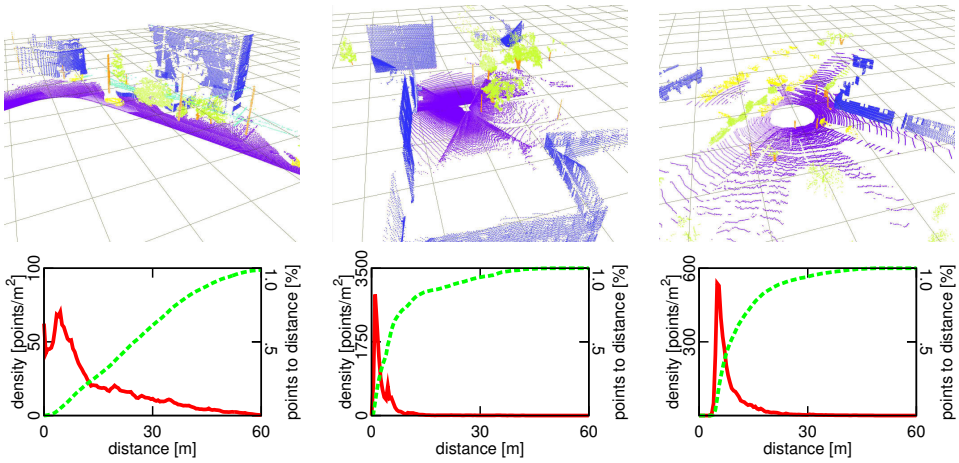
IV. Experimental Setup

A. Datasets

We evaluated the performance of the previously introduced descriptors on datasets generated by three common 3D laser rangefinder setups—a pan-tilting 2D laser rangefinder, 2D sweeping laser rangefinders, and a Velodyne HDL64-E laser rangefinder [18]. Figure 2 depicts example scans from these datasets and the distribution of laser returns for the specific setup.

The first dataset was recorded at the University of Freiburg, Germany, using a SICK LMS laser rangefinder mounted on a pan-tilt unit. The point clouds were manually labeled[1] as pavement, sidewalk, lawn, pole, shrub, bush, foliage, tree trunk, building facade, window, door, bicycle, and car. For the evaluation we only used a subset of these classes and combined subclasses into more general classes: (1) *ground* consisting of pavement, sidewalk, and lawn, (2) *vegetation* containing shrub, foliage, and bushes, (3) *facades* subsuming building facades, doors, and windows, (4) *poles* combined with tree trunks.

[1]The registered laser range scans with the robot odometry are available at http://ais.informatik.uni-freiburg.de/projects/datasets/fr360/. The labels can be downloaded at http://www.iai.uni-bonn.de/∼behley/data/.



(a) (b) (c)

Fig. 2. Laser range data of (a) the Pittsburgh, (b) the Freiburg, and (c) the Wachtberg dataset used in the evaluation. The different labels are colored as follows: purple = *ground*, blue = *building facades*, green = *vegetation*, orange = *poles*, yellow = *vehicles*, light-blue = *wire*. Below every scan is the distribution of laser returns per distance to the laser scanner depicted. The red/solid curve depicts the number of laser points per m^2 at this distance. The green/dashed curve is the fraction of laser returns up to the distance.

We chose these more general classes because they contain the surfaces and objects most relevant for outdoor applications. Furthermore, the distinction of pavement, sidewalk and lawn is often only possible by using contextual knowledge. Since we want to investigate the performance of local approaches, we decided to combine these different ground classes into a single class. The same argument holds also for bulding facades, windows, and doors. Poles are especially interesting because they allow to reveal registration errors and thus can be useful to assess the perfomance of SLAM approaches.

The second dataset was acquired on the campus of the Carnegie Mellon University in Pittsburgh with a Jeep equipped with SICK laser scanners facing sideways and a laser scanner mounted on the front of the vehicle. The dataset contains the same labels as the Freiburg dataset, but we additionally use *vehicles* and *wire* like Xiong *et al.*[7]. The dataset was filtered and registered to get a complete point cloud, and chunks of approximately 100.000 laser points were extracted.

The last dataset was recorded at the Fraunhofer FKIE in Wachtberg, Germany, using a Velodyne HDL-64E S2 laser range scanner mounted on an Opel Vectra. We also manually labeled the dataset with the classes *ground*, *vegetation*, *facades*, *vehicles*, and *poles*.

All three datasets show different characteristics. Figure 2 depicts the point density and the number of laser points per square meter. In case of the Pittsburgh dataset, we find homogeneous sampling of the surfaces and nearly linear increase of points per distance (green/dashed curve in the plots). The Freiburg and Wachtberg dataset in contrast show a significant drop in the sampling rate at larger distances. As the Velodyne HDL-64 rotates in order to generate a full 360° scan, we also can see a ring pattern with points in the same ring much closer to each other than points in different rings.

B. Evaluation Criterion

Let $\mathcal{X}^\star = \{(\mathbf{x}_i, y_i^\star)\}, |\mathcal{X}^\star| = M$ be a set of test instances with ground truth labels $y_i^\star$, and $\hat{y}_i = f(x_i)$ the predicted label of a classifier f trained on a separate set $\mathcal{X} = \{(\mathbf{x}_i, y_i)\}, |\mathcal{X}| = N, \mathcal{X} \cap \mathcal{X}^\star = \emptyset$.

The class-wise precision p_k of a class or label k is given by the ratio of correctly classified instances $\{(\mathbf{x}_i, y_i^\star)\}$ of the test set and all instances classified as class k:

$$p_k = \frac{|\{(\mathbf{x}_i, y_i^\star) \in \mathcal{X}^\star | \hat{y}_i = k \wedge y_i^\star = k\}|}{|\{(\mathbf{x}_i, y_i^\star) \in \mathcal{X}^\star | \hat{y}_i = k\}|} \tag{5}$$

The class-wise recall r_k is given by the ratio of correctly classified instances and all instances with reference label k:

$$r_k = \frac{|\{(\mathbf{x}_i, y_i^\star) \in \mathcal{X}^\star | \hat{y}_i = k \wedge y_i^\star = k\}|}{|\{(\mathbf{x}_i, y_i^\star) \in \mathcal{X}^\star | y_i^\star = k\}|} \tag{6}$$

The F_1 measure is defined as the average over the class-wise precisions p_k and recalls r_k:

$$F_1 = \frac{1}{K} \sum_k \frac{2 \cdot p_k \cdot r_k}{p_k + r_k}. \tag{7}$$

Thus, we are independent of the actual number of instances. Furthermore, the F_1 measure penalizes high precision and low recall.

C. Classification Approaches

We evaluated the descriptor performance using three different classification approaches ranging from a simple linear model to a more complex one based on Conditional Random Fields.

Logistic Regression is the baseline approach, which is a binary linear classifier. For multi-class classification with k classes, we train k binary logistic regressions one-vs.-the-rest, resulting in k binary logistic models for each class. To infer the class of an unseen feature vector $\mathbf{x}$, we assign the class i such that it maximizes $P(y = i|x)$.

The second classification approach, the Spectrally Hashed Logistic Regression (SHLR)[6], is an extension of the logistic regression that uses the principle of semantic hashing.

The idea of semantic hashing is to learn a hash function h that results in similar hash codes $h(\mathbf{x}_i)$, $h(\mathbf{x}_j)$ for two different, but similar feature vectors $\mathbf{x}_i$, $\mathbf{x}_j$; the hamming distance of the codewords is small, if the euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is also small. Spectral Hashing [19] has shown to be an effective choice for this purpose, especially for small codes.

In SHLR, the learned hash function clusters the space of feature vectors $\mathbf{x_i}$ into similar vectors sharing the same codeword. These similar vectors are used to learn a local logistic regression model, which turns out to be far more efficient than comparing all feature vectors inside a hash bin to a query feature vector. To classify an unseen laser point, the hash code of its feature vector is calculated and the logistic model of this codeword is used to assign a class.

Both local classifiers are compared with a collective classification approach using Conditional Random Fields (CRF)—the Functional Max-Margin Markov Networks (FM^{3}N) [4]. These approaches try to find the MAP assignment to $P(y_{1:N}|\mathbf{x}_{1:N})$, which maximizes the likelihood of the joint assignment of classes to Y_i given the features $\mathbf{X}_i$:

$$P(y_{1:N}|\mathbf{x}_{1:N}) = \frac{1}{Z(X_{1:N})} \prod_{y_i \in \mathcal{V}} \phi(y_i, x_i) \prod_{(y_i,y_j)\in\mathcal{E}} \psi(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j), \quad (8)$$

where the underlying graph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is given by vertices y_i representing random variables for labels and $\mathbf{x}_i$ for feature vectors of every laser point, and edges $(y_i, y_j) \in \mathcal{E}$ between them, if there exist a direct dependency. $Z(X_{1:N})$ is a normalizer depending on the feature vectors—the partition function. For more details on probabilistic graphical models, we refer to [20].

The edges are given by the k-nearest neighbors of a laser point $\mathbf{p}_i$, i.e., $(y_i, y_j) \in \mathcal{E}$, if $\mathbf{p}_j \in \mathcal{N}^k_{\mathbf{p}i}$. As proposed by Xiong *et al.* [7], we use a 'similarity' edge potential $\psi(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j)$ computed from the node features $\mathbf{x}_i, \mathbf{x}_j$: $\psi(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) = \mathbf{1}\{y_i = y_j\} \cdot \exp(-w_i^T \cdot \mathbf{x_{ij}})$, where the k-th entry of $\mathbf{x}_{ij} \in \mathbb{R}^m$ is calculated as $\mathbf{x}_{ij}^{(k)} = \left(1 + \mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)}\right)^{-1}$, i.e., the more similar the node features, the larger the entries in the edge feature $\mathbf{x_{ij}}$. $\mathbf{1}\{s\}$ denotes the indicator function that returns 1 if statement s is true, and 0 otherwise.

V. RESULTS AND DISCUSSION

We performed a 5-fold cross-validation for all datasets using the histogram descriptors. For this purpose, we selected 5 representative and non-overlapping 360° laser range scans from every dataset. We only evaluated a subset of parameters with the CRFs, which allowed us to store the networks with node and edge potentials in memory. Hence, we were able to evaluate the CRFs using large descriptors in reasonable time. Table I shows the evaluated feature parameters.

TABLE I

PARAMETERS OF THE DESCRIPTORS. VALUES USED FOR NODE AND EDGE POTENTIALS ARE BOLD.

Descriptor	Parameter values
Normal Histogram	$b = \{\mathbf{5}, \mathbf{10}, \mathbf{15}\}$
Distribution Histogram	$b = \{\mathbf{3}, \mathbf{5}, 7\}$
Spin Image	$b = \{\mathbf{5}, \mathbf{10}, 20\}$
SHOT	$b = \{\mathbf{5}, \mathbf{10}, 15\}$
Spectral Histogram	$l = \{\mathbf{3}, \mathbf{5}, 7\}$, $s = 5$

Implementation details. The logistic regression uses the implementation of LibLinear [21], where we used a regularizer weight of $C = 1.0$ and a intercept of $\beta = 1.0$ for all experiments. The SHLR uses 8 bits and maximal hamming distance of 4 for the search of a matching codeword. The FM^{3}N implementation[2] was adapted to our needs and used only pair-wise potentials with linear regressors. We choose 50 iterations to learn linear regressors with learning rate of 0.1. All descriptors were implemented using C++. For the SHOT descriptor we adapted the available implementation in the Point Cloud Library(PCL). For every laser range scan we used an octree to determine the nearest neighbors, and the normals of a scan were estimated using PCA [22] on a neighborhood of 0.6 m. All experiments were performed on an Intel Xeon X5550 with 2.67 GHz and 12 GB memory.

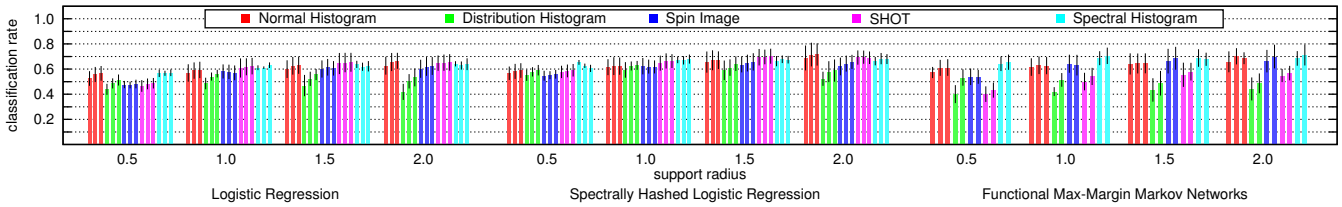
Reference frame. We evaluated the influence of different reference frames and axes on the performance. Tombari *et al.* [13] showed that the reference frame and its stability significantly affect the performance in the matching task. Is this also valid with the classification of three-dimensional laser range data?

Figures 3, 4, and 5 depict the performance using the local reference frame. Comparing these results with the global reference frame-based descriptors, we observe a significant improvement over the local reference frame. Particularly the Distribution Histogram is strongly affected by the choice of the reference frame. The normal histogram only shows a minor effect when changing the reference frame.

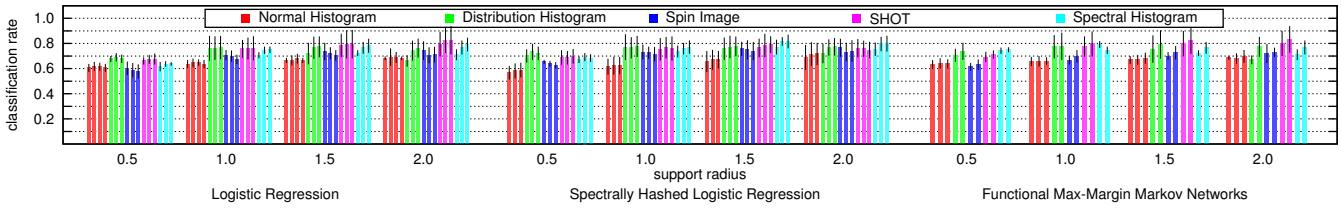
The global reference frame is more stable than the local reference frame, as the eigenvalues are affected by clutter. However, in the presented datasets we only observe man-made objects and scatter by vegetation, which both tend to have a vertical orientation. Thus, the z-axis is a better choice in this application, as we do not have arbitrary rotations of objects in the dataset. In addition, we get more discriminative histograms of ground and facades, which also improves the performance of the classification approaches.

Support radius. The diagrams also show a significant increase in classification performance of all approaches with an increase of the support radius. An increasing radius of the histograms also includes more contextual information of the vicinity. For instance, all classifiers performed better in classifying cars when the support radius is increased. This is an unexpected result as cars are rather complex objects

[2]Available at http://www.cs.cmu.edu/∼vmr/software/.



(a)



(b)

Fig. 3. Results on the Freiburg dataset. In (a) we used a local reference frame or axis, in (b) the global reference frame was used. Multiple bars for a feature result from different numbers of bins.

TABLE II

PRECISION AND RECALL OF THE DIFFERENT APPROACHES ON THE FREIBURG DATASET.

C	Feature	b	δ	F_1	ground	facade	pole	vegetation
SHLR	Normal Histogram	10	2.0	71.6	97.9/95.0	87.0/85.2	20.2/50.0	80.4/78.4
	Distribution Histogram	7	1.5	77.7	98.9/99.0	89.2/84.2	48.5/64.9	81.5/76.5
	Spin Image	5	2.0	76.5	97.9/99.0	84.3/80.4	51.4/72.5	78.7/69.8
	SHOT	15	1.5	**79.1**	98.6/98.8	89.5/88.5	38.4/69.1	88.8/78.9
	Spectral Histogram	7	1.5	**81.6**	98.5/98.5	92.5/89	50.4/67.5	89.8/82.8
FM3N	Normal Histogram	15	2.0	69.7	98.4/94.9	95.6/87.8	5.65/10.4	83.1/86.5
	Distribution Histogram	5	1.5	**79.2**	98.9/99.1	92.3/84.1	53.1/65.9	80.3/82.1
	Spin Image	10	2.0	72.9	98.9/99.2	85.3/82.1	27.8/60.6	83.8/76.1
	SHOT	10	2.0	**83.3**	99.1/99.3	94.3/89.8	61.4/67.7	85.1/87.2
	Spectral Histogram	3	1.0	**79.2**	99.0/98.3	92.1/81.0	46.7/77.1	78.0/80.7

with both flat and curved surfaces. However, vehicles are usually parked/driven on flat ground, which seems to be the discriminating property of cars in the Pittsburgh dataset. So if the classifier recognizes something that resides on flat ground, this is most likely a vehicle. Thus, we can indirectly encode—to some extent—the context in features for a local classifier. This is also achieved in collective approaches by a more complex model, or can be directly incorporated by stacking with multiple logistic regressions [7].

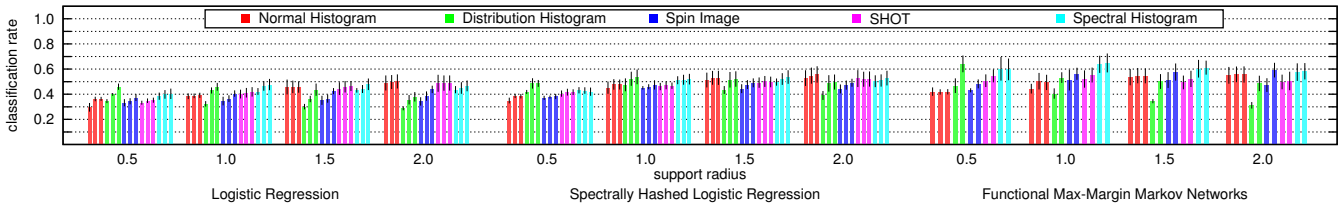
In the Wachtberg dataset, this contextual information does not always help, as there are shrubs/bushes on flat lawns. Hence, in this dataset the classes *vehicle* and *vegetation* are more often confused with other classes than in the Pittsburgh dataset.

We also experimented with larger radii than 2.0 m, but these turned out to entail no significant improvement, or even decreased the perfomance of the different classification approaches.

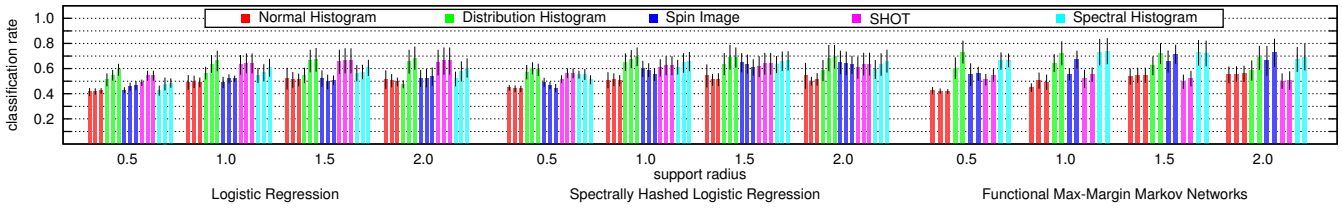
Number of Bins. Another factor is the number of bins for the different histograms. We used different numbers of bins per dimension, i.e., 3, 5, and 7 bins. We see a profound difference of the overall accuracy using different numbers of bins—especially with a larger radius of the histograms. We find a degradation of the 3 bins histograms with increasing radius, as one bin gets too large to capture the structure of the different classes. The influence of the number of bins is more severe with the normal histogram than with the others. We can see a significant drop in performance with an increasing radius of the normal histogram. The resolution of the histogram gets too coarse to capture the differences between the classes. The distribution histogram with global reference frame is less affected by the number of bins, but by inspecting the results visually, we conclude that a better performance is achieved with more bins.

Class-wise Performance. Tables II, III and IV show the precision and recall of the classes using descriptors with a global reference frame. We only show the best results with every descriptor using SHLR and the FM^{3}N.

The class-wise precision and recall reveal the deficiency of the different descriptors. Generally, the classes *ground*, *facades*, and *vegetation* could be well distinguished from the other classes. *Poles*, *vehicles*, and *wire* turned out to be the more challenging classes. Especially, the class *wire* in the Pittsburgh dataset was mostly confused with vegetation. The class *wire* was mostly assigned to wires from utility poles, i.e., these wires are only sparsely sampled by laser points. Thus, wires look similar to vegetation, where many laser beams just pass through the foliage.



(a)



(b)

Fig. 4. Results on the Pittsburgh dataset. In (a) we used a local reference frame or axis, in (b) the global reference frame was used. Multiple bars for a feature result from different numbers of bins.

TABLE III

PRECISION AND RECALL OF THE DIFFERENT APPROACHES ON THE PITTSBURGH DATASET.

C	Feature	b	δ	F_1	wire	pole	ground	vegetation	facade	vehicle
SHLR	Normal Histogram	5	2.0	54.3	21.8/47.5	2.3/29.2	98.8/98.0	96.0/91.0	88.6/79.3	22.3/35.4
	Distribution Histogram	7	2.0	**69.5**	22.8/43.4	53.1/66.2	98.8/99.5	96.3/86.1	80.9/82.4	65.2/72.2
	Spin Image	5	1.5	64.7	27.3/44.0	59.0/73.5	98.5/99.2	92.1/78.5	67.7/71.6	48.1/59.8
	SHOT	10	2.0	63.6	14.4/32.8	18.7/46.2	99.1/99.2	96.5/90.4	89.0/83.0	70.4/68.2
	Spectral Histogram	7	1.5	66.2	35.9/41.7	51.5/68.8	99.1/99.2	92.1/84.1	78.9/75.9	46.9/58.3
FM3N	Normal Histogram	15	2.0	55.9	0.0/0.0	0.7/20.0	99.5/97.9	97.1/90.0	82.2/80.2	66.7/62.1
	Distribution Histogram	5	0.5	**73.2**	32.5/50.0	69.9/66.1	99.2/99.4	97.2/89.6	84.1/85.6	52.6/73.7
	Spin Image	10	2.0	**72.6**	29.5/54.7	65.3/69.0	98.9/99.4	96.4/87.2	74.4/87.8	71.4/69.5
	SHOT	10	1.0	55.2	45.5/41.9	42.6/50.4	99.1/91.4	98.3/85.3	12.1/58.7	48.8/71.9
	Spectral Histogram	5	1.0	63.8	29.8/36.1	14.1/54.0	99.5/99.4	96.3/87.2	87.7/84.9	63.5/75.8

In the Wachtberg dataset, vehicles were nearly always confused with vegetation. Here the front part of a car is similar to lower bushes and shrubs. In all datasets the windows in the facades and the induced sparsity of laser returns lead to a classification as vegetation.

In case of a linear classifier, the SHOT and the Distribution Histogram using the global reference frame seem to be a good choice. A radius δ of $1.5\,\mathrm{m}$ to $2.0\,\mathrm{m}$ was needed to attain good classification results. With more complex models, such as SHLR and FM^{3}N, the Spin Image and the Spectral Histogram also were effective representations of the different classes. The normal histograms turned out to be insufficient for the classification of poles and also vehicles.

VI. CONCLUSIONS AND FUTURE WORK

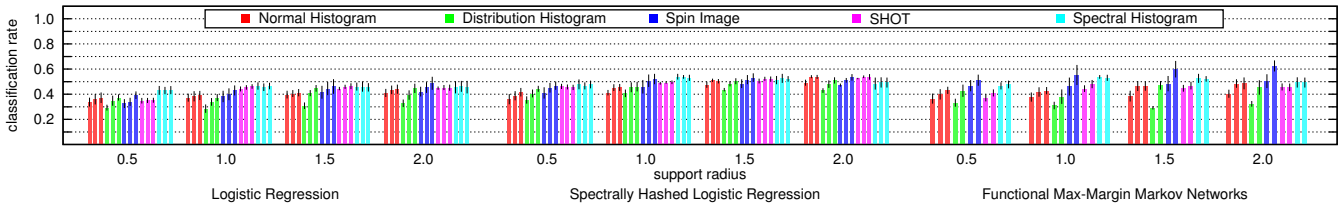
The choice of suitable features dramatically influences the performance of classification approaches. In this contribution we evaluated several histogram-based descriptors for the classification of three-dimensional laser range data in an outdoor scenario. We evaluated the descriptors on three datasets acquired with different state-of-the-art sensor setups.

From the presented results we can draw several conclusions. First, we showed that a proper choice of the reference frame can significantly improve the performance of all evaluated classification approaches. Here a z-axis-based reference frame was superior to the usual normal-based reference frame. Second, we could show that the performance strongly correlates to the support radius. Choosing a larger support radius improved the descriptiveness significantly. Last, we showed that the influence of the number of bins is not as strong as the influence of the support radius. In all experiments the 3D histograms—Distribution Histogram, SHOT, and Spectral Histogram—were the descriptors that resulted in the best performance.

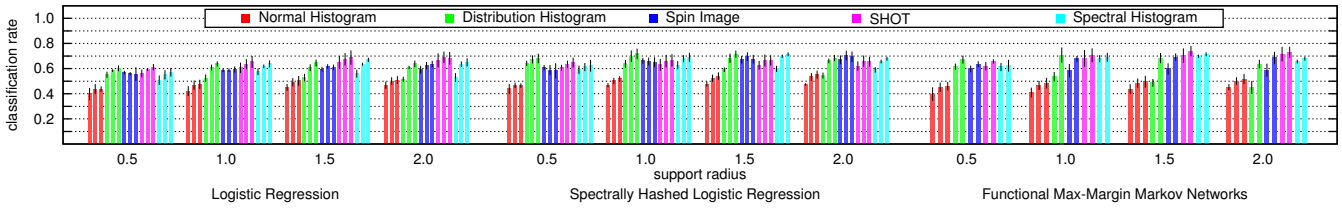
An interesting avenue for future work is the investigation of segment descriptors using the k-nearest neighborhood for objects with well-defined boundaries, such as cars or people. Another interesting issue is the combination with other shape properties and the investigation of suitable combinations.

VII. ACKNOWLEDGMENTS

Thanks to Daniel Munoz for providing hints on their FM^{3}N implementation. We also want to thank Samuele Salti and Federico Tombari for making the SHOT available in the PCL. Thanks to Dirk Schulz, Achim Königs, Frank Höller, Timo Röhling, and Ansgar Tessmer at the Fraunhofer FKIE. Special thanks to Daniel Seidel, Dominik A. Klein, Florian Schöler, and Jenny Balfer for fruitful discussions.



(a)



(b)

Fig. 5. Results on the Wachtberg dataset. In (a) we used a local reference frame or axis, in (b) the global reference frame was used. Multiple bars for a feature result from different numbers of bins.

TABLE IV

PRECISION AND RECALL OF THE DIFFERENT APPROACHES ON THE WACHTBERG DATASET.

C	Feature	b	δ	F_1	vehicle	ground	facade	pole	vegetation
SHLR	Normal Histogram	15	2.0	55.3	20.8/32.1	91.7/87.4	92.2/88.8	1.75/16.4	73.2/70.5
	Distribution Histogram	7	1.5	**71.5**	52.7/62.2	93.2/92.4	81.6/81.1	48.5/72.6	76.2/70.5
	Spin Image	10	2.0	70.4	53.2/61.2	96.0/92.1	71.4/70.9	53.2/76.8	72.9/70.9
	SHOT	10	1.5	66.5	48.0/63.7	94.3/88.2	89.9/85.6	21.7/60.4	72.0/72.7
	Spectral Histogram	7	1.5	**71.4**	47.2/63.2	95.3/92.1	78.3/77.4	53.4/74.5	76.3/72.0
FM3N	Normal Histogram	15	2.0	51.7	4.7/27.6	95.2/84.8	95.8/86.8	0.2/20.0	72.7/72.4
	Distribution Histogram	5	1.0	70.8	30.9/65.8	97.7/91.2	84.3/74.2	60.7/75.9	79.0/79.0
	Spin Image	10	2.0	68.5	26.5/64.0	98.3/92.9	79.9/78.4	47.0/78.1	81.3/73.6
	SHOT	10	1.5	**73.9**	54.6/67.8	97.6/89.2	92.3/89.5	49.1/59.6	75.3/80.7
	Spectral Histogram	5	1.5	68.6	36.4/60.7	96.7/90.5	79.2/75.1	51.6/68.5	72.6/71.9

REFERENCES

[1] M. Himmelsbach, T. Luettel, and H.-J. Wuensche, "Real-time Object Classification in 3D Point Clouds Using Point Feature Histograms," in *IROS*, 2009, pp. 994–1000.

[2] A. Teichman and S. Thrun, "Tracking-Based Semi-Supervised Learning," in *RSS*, 2011.

[3] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng, "Discriminative Learning of Markov Random Fields for Segmentation of 3D Scan Data," in *CVPR*, 2005, pp. 169–176.

[4] D. Munoz, J. A. D. Bagnell, N. Vandapel, and M. Hebert, "Contextual Classification with Functional Max-Margin Markov Networks," in *CVPR*, 2009, pp. 975–982.

[5] A. Agrawal, A. Nakazawa, and H. Takemura, "MMM-classification of 3D Range Data," in *ICRA*, 2009.

[6] J. Behley, K. Kersting, D. Schulz, V. Steinhage, and A. B. Cremers, "Learning to Hash Logistic Regression for Fast 3D Scan Point Classification," in *IROS*, 2010, pp. 5960–5965.

[7] X. Xiong, D. Munoz, J. A. Bagnell, and M. Hebert, "3-D Scene Analysis via Sequenced Predictions over Points and Regions," in *ICRA*, 2011, pp. 2609–2616.

[8] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *ICRA*, 2009, pp. 3212–3217.

[9] A. Johnson and M. Hebert, "Using spin images for effcient object recognition in cluttered 3D scenes," *TPAMI*, vol. 21, no. 5, pp. 433–449, 1999.

[10] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *TPAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[11] J. W. Tangelder and R. C. Veltkamp, "A survey on content based 3d shape retrieval methods," *J. of Mult. Tools and Appl.*, vol. 39, no. 3, pp. 441–471, 2008.

[12] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, "Learning Informative Point Classes for the Acquisition of Object Model Maps," in *ICARCV*, 2008.

[13] F. Tombari, S. Salti, and L. D. Stefano, "Unique Signatures of Histograms for Local Surface Description," in *ECCV*, 2010, pp. 356–369.

[14] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "Point Feature Extraction on 3D Range Scans Taking into Account Object Boundaries," in *ICRA*, 2011, pp. 2601–2608.

[15] R. Triebel, K. Kersting, and W. Burgard, "Robust 3D Scan Point Classification using Associative Markov Networks," in *ICRA*, 2006, pp. 2603–2608.

[16] A. Patterson, P. Mordohai, and K. Daniilidis, "Object Detection from Large-Scale 3D Datasets using Bottom-up and Top-down Descriptors," in *ECCV*, 2008, pp. 553–566.

[17] G. Medioni, M.-S. Lee, and C.-K. Tang, *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.

[18] Velodyne Lidar Inc., "High Definition Lidar HDL-64E S2 Datasheet," http://www.velodyne.com/lidar/products/specifications.aspx.

[19] Y. Weiss, A. Torralba, and R. Fergus, "Spectral Hashing," in *NIPS*, 2009, pp. 1753–1760.

[20] D. Koller and N. Friedman, *Probabilistic Graphical Models*. MIT Press, 2009.

[21] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A Library for Large Linear Classification," *JMLR*, vol. 9, pp. 1871–1874, 2008.

[22] K. Klasing, D. Althoff, D. Wollherr, and M. Buss, "Comparison of Surface Normal Estimation Methods for Range Sensing Applications," in *ICRA*, 2009, pp. 3206–3211.