Unsupervised Discovery of Object Classes from Range Data using Latent Dirichlet Allocation

Felix Endres   Christian Plagemann   Cyrill Stachniss   Wolfram Burgard

Abstract— Truly versatile robots operating in the real world have to be able to learn about objects and their properties autonomously, that is, without being provided with carefully engineered training data. This paper presents an approach that allows a robot to discover object classes in three-dimensional range data in an unsupervised fashion and without a-priori knowledge about the observed objects. Our approach builds on Latent Dirichlet Allocation (LDA), a recently proposed probabilistic method for discovering topics in text documents. We discuss feature extraction, hypothesis generation, and statistical modeling of objects in 3D range data as well as the novel application of LDA to this domain. Our approach has been implemented and evaluated on real data of complex objects. Practical experiments demonstrate, that our approach is able to learn object class models autonomously that are consistent with the true classifications provided by a human. It furthermore outperforms unsupervised method such as hierarchical clustering that operate on a distance metric.

I. INTRODUCTION

Home environments, which are envisioned as one of the key application areas for service robots, typically contain a variety of different objects. The ability to distinguish objects based on observations and to relate them to known classes of objects therefore is important for autonomous service robots. The identification of objects and their classes based on sensor data is a hard problem due to the varying appearances of the objects belonging to specific classes. In this paper, we consider a robot that can observe a scene with a 3D laser range scanner. The goal is to perform

- unsupervised learning of a model for object classes,
- consistent classification of the observed objects, and
- correct classification of unseen objects belonging to one of the known object classes.

Figure 1 depicts a typical point cloud of a scene considered in this paper. It contains four people, a box, and a balloon-like object. The individual colors of the 3D data points illustrate the corresponding object classes that we want our algorithm to infer.

An important distinction between different approaches to object detection and recognition is the way the objects or classes are modeled. Models can be engineered manually, learned from a set of labeled training data (supervised learning) or learned from unlabeled data (unsupervised learning). While the former two categories have the advantage that detailed prior knowledge about the objects can be included easily, the effort for manually building the model or labeling
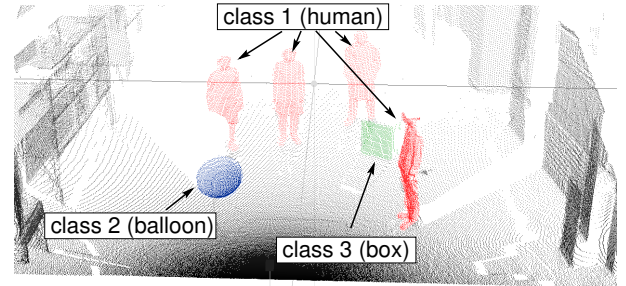
Fig. 1: Example of a scene observed with a laser range scanner mounted on a pan-tilt unit. Points with the same color resemble objects belonging to the same class (best viewed in color).

a significant amount of training data becomes infeasible with increasing model complexity and larger sets of objects to identify. Furthermore, in applications where the objects to distinguish are not known beforehand, a robot needs to build its own model, which can then be used to classify the data.

The contribution of this paper is a novel approach for discovering object classes from range data in an unsupervised fashion and for classifying observed objects in new scans according to these classes. Thereby, the robot has no a-priori knowledge about the objects it observes. Our approach operates on a 3D point cloud recorded with a laser range scanner. We apply Latent Dirichlet Allocation (LDA) [2], a method that has recently been introduced to seek for topics in text documents [9]. The approach models a distribution over feature distributions that characterize the classes of objects. Compared to most popular unsupervised clustering methods such as k-means or hierarchical clustering, no explicit distance metric is required. To describe the characteristics of surfaces belonging to objects, we utilize spin-images as local features that serve as input to the LDA. We show in practical experiments on real data that a mobile robot following our approach is able to identify similar objects in different scenes while at the same time labeling dissimilar objects differently.

II. RELATED WORK

The problem of classifying objects and their classes in 3D range data has been studied intensively in the past. Several authors introduced features for 3D range data. One popular free-form surface descriptor are spin-images, which have been applied successfully to object recognition problems [13; 12; 14; 15]. In this paper, we propose a variant of spin-images that—instead of storing point distributions of the surface— stores the angles between the surface normals of points, which we found to yield better results in our experiments.

An alternative shape descriptor has been introduced by [18]. It relies on symbolic labels that are assigned to regions. The symbolic values, however, have to be learned from a labeled training set beforehand. Stein and Medioni [19] present a point descriptor that, similar to our approach, also relies on surface orientations. However, it focuses on the surface normals in a specific distance to the described point and models their change with respect to the angle in the tangent plane of the query point. Additional 3D shape descriptors are described in [5] and [6].

A large amount of work has focused on supervised algorithms that are trained to distinguish objects or object classes based on a labeled set of training data. For example, Anguelov *et al.* [1] and Triebel *et al.* [20] use supervised learning to classify objects and associative Markov networks to improve the results of the clustering by explicitly considering relations between the class predictions. In a different approach, Triebel *et al.* [21] use spin-images as surface descriptors and combine nearest neighbor classification with associative Markov networks to overcome limitations of the individual methods. Another approach using probabilistic techniques and histogram matching has been presented by Hetzel *et al.* [10]. It requires a complete model of the object to be recognized, which is an assumption typically not fulfilled when working on 3D scans recorded with a laser range finder. Ruhnke *et al.* [17] proposed an approach to reconstructing full 3D models of objects by registering several partial views. The work operates on range images from which small patches are selected based on a region of interest detector.

In addition to the methods that operate on 3D data, much research has also focused on image data as input. A common approach to locate objects in images is the sliding window method [4; 7]. Lampert *et al.* [16] proposed a new framework that allows to efficiently find the optimal bounding box without applying the classification algorithm explicitly to all possible boxes. Another prominent supervised detector is the face detector presented by Viola and Jones [22]. It computes Haar-like features and applies AdaBoost to learn a classifier.

In the domain of unsupervised classification of text documents, several models that greatly surpass mere counting of words have been proposed. These include probabilistic latent semantic indexing (PLSI) [11] and Latent Dirichlet Allocation [2], which both use the co-occurrence of words in a probabilistic framework to group words into topics. In the past, LDA has also been applied successfully to image data. In contrast to text documents [9], images often contain data of many different categories. Wang and Grimson [23], therefore, first perform a segmentation before applying LDA. Bosch *et al.* [3] used PLSI for unsupervised discovery of object distributions in image data. As shown in [8], LDA supersedes PLSI and it has been argued that the latter can be seen as a special case of LDA, using a uniform prior and maximum a posteriori estimation for topic selection. Fritz and Schiele [7] propose the sliding window approach on a grid of edge orientations to evaluate topic probabilities on subsets of the whole image. While the general approach of these papers
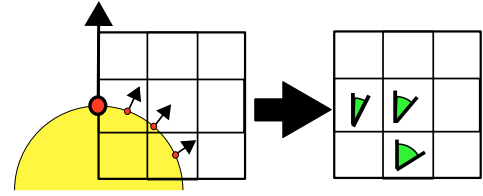

Fig. 2: Variant of spin-images used to compute a surface signature: the 3D object structure (yellow circle) is rotated around the surface normal of a query point (large red point) and a grid model accumulates the average angular distances between the surface normal at the query point and those of the points falling into the grid cells (small red points).

is related to ours, to the best of our knowledge the algorithm described in this paper is the first to apply LDA on laser range data and which addresses the specific requirements of this domain.

III. DATA PRE-PROCESSING AND LOCAL SHAPE FEATURES

As most approaches to object detection, identification, and clustering, we operate on local features computed from the input data. Our primary focus lies on the description of *shape* as this is the predominant feature captured in 3D range data. However, real-world objects belonging to the same class do not necessarily have the same shape and vice versa. Humans, for example, have a significant variability in shape. To deal with this problem, we model classes of objects as distributions of local shape features.

In the next sections, we first describe our local feature used to represent the characteristics of surfaces and after than, we address the unsupervised learning problem to estimate the distributions over local features.

A. Representation and Data Pre-processing

Throughout this work, we assume our input data to be a point cloud of 3D points. Such a point cloud can be obtained with a 2D laser range finder mounted on a pan-tilt unit, a standard setting in robotics to acquire 3D range data. An example point cloud recorded with this setup is shown in the motivating example in Figure 1 on the first page of this paper.

As in nearly all real world settings, the acquired data is affected by noise and it is incomplete due to perspective occlusions. The segmentation of range scans into a set of objects and background structure is not the key focus of this work. We therefore assume a ground plane as well as walls that can be easily extracted and assume the objects to be spatially disconnected. This allows us to apply a spatial clustering algorithm to create segments containing only one object.

B. Local Shape Descriptors

For characterizing the local shape of an object at a query point, we propose to use a novel variant of spin-images [12]. Spin-images can be seen as small raster images that are aligned to a point such that the upwards pointing vector of the raster image is the surface normal of the point. The image is then virtually rotated around the surface normal, "collecting" the

neighboring points it intersects. To account for the differences in data density caused by the distance between sensor and object, the spin-images are normalized.

To actually compute a normal for each data point, we compute a PCA using all neighboring points in a local region of 10cm. Then, the direction of the eigenvector corresponding to the smallest eigenvalue provides a comparably stable but smoothed estimate of the surface normal.

We have developed a variant of spin-images that does not count the points "collected" by the pixels of the raster image. Instead, we compute the average angle between the normal of the query point for which the spin-image is created and the normals of all collected points. See Figure 2 for an illustration. The average between the normals is then discretized to obtain a discrete feature space, as required in the LDA approach. As we will show in our experiments, this variant of spin-images provides better results, since they contain more information about the shape of the object.

IV. PROBABILISTIC TOPIC MODELS FOR OBJECT SHAPE

After segmenting the scene into a finite set of scan segments and transforming the raw 3D input data to the discrete feature space, the task is to group similar segments to classes and to learn a model for these classes. Moreover, we aim at solving the clustering and modeling problems simultaneously to achieve a better overall model. Inspired by previous work on topic modeling in text documents, we build on Latent Dirichlet Allocation for the unsupervised discovery of object classes from feature statistics.

Following this model, a multinomial distribution is used to model the distribution of discrete features in an object class. Analogously, another multinomial distribution is used to model the mixture of object classes which contribute to a scan segment. In other words, we assume a generative model, in which (i) segments generate mixtures of classes and (ii) classes generate distributions of features.

Starting from a prior distribution about these latent (i.e., hidden) mixtures, we update our belief according to the observed features. To do this efficiently, we express our prior $P(\theta)$ as a distribution that is conjugate to the observation likelihood $P(y \mid \theta)$. $P(\theta)$ being a conjugate distribution to $P(y \mid \theta)$ means that

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{\int P(y \mid \theta)P(\theta)\,d\theta} \tag{1}$$

is in the same family as $P(\theta)$ itself. For multinomial distributions, the conjugate prior is the Dirichlet distribution, which we explain in the following.

A. The Dirichlet Distribution

The Dirichlet distribution is a distribution over multivariate probability distributions, i.e., a distribution assigning a probability density to every possible multivariate distribution. For the multinomial variable $\mathbf{x} = \{x_1, \ldots, x_K\}$ with K exclusive states x_i, the Dirichlet distribution is parameterized by a vector $\alpha = \{\alpha_1, \ldots, \alpha_K\}$. If $\alpha_i = 1$ for all i, the Dirichlet distribution
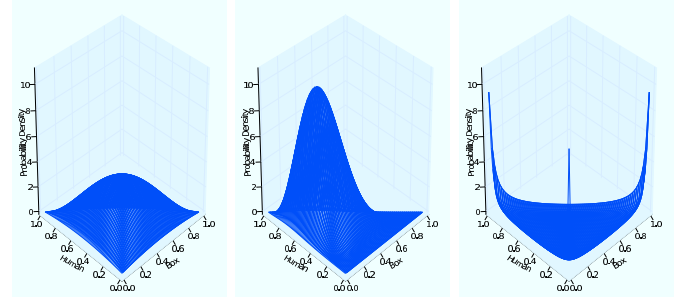

Fig. 3: Three Dirichlet distributions. On the left for the parameter vector $\alpha = \{2,2,2\}$, in the middle for $\alpha = \{3,6,3\}$ and on the right for $\alpha = \{0.1,0.1,0.1\}$.

is uniform. One can think of $(\alpha_i - 1)$ for $\alpha_i \in \mathbb{N}^{>0}$ as the number of observations of the state i. The Dirichlet distribution can be calculated as

$$f(\mathbf{x}) = \underbrace{\frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}}_{\text{Normalization}} \prod_{i=1}^{K} x_i^{\alpha_i-1}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function and where the elements of $\mathbf{x}$ have to be positive and sum up to one.

Consider the following example: let there be three object classes "human", "box", and "chair" with a Dirichlet prior parameterized by $\alpha = \{2,2,2\}$. This prior assigns the same probability to all classes and hence results in a *symmetric* Dirichlet distribution. A 3D Dirichlet distribution $Dir(\alpha)$ can be visualized by projecting the the manifold where $\sum \alpha_i = 1$ to the 2D plane, as depicted in the left plot of Figure 3. Here the third variable is given implicitly by $\alpha_3 = 1 - \alpha_1 - \alpha_2$. Every corner of the depicted triangle represents the distributions where only the respective class occurs and the center point represents the uniform distribution over all classes. Now consider an observation of one human, four boxes, and a chair. By adding the observation counts to the elements of α, the posterior distribution becomes $Dir(\{5,8,5\})$ which is shown in the middle plot in Figure 3. The same result would of course occur when calculating the posterior using Eq. (1).

However choosing the values of α_i larger than 1 favors distributions that represent mixtures of classes, i.e. we expect the classes to occur together. To express a prior belief that either one or the other dominates we need to choose values smaller than 1 for all α_i. The shape of the distribution then changes in a way that it has a "valley" in the middle of the simplex and peaks at the corners. This is depicted in the right plot in Figure 3. In our setting, where a Dirichlet distribution is used to model the distribution of object classes, such a prior would correspond to the proposition that objects are typically assigned to one (or only a few) classes.

The calculation of the *expected probability distribution* over the states and can be performed easily based on α. The expected probability for x_i is given by

$$\mathbb{E}[x_i] = \frac{\alpha_i}{\sum_{i'}\alpha_{i'}}. \tag{3}$$

B. Latent Dirichlet Allocation

Latent Dirichlet allocation is a fully generative probabilistic model for semantic clustering of discrete data, which was developed by Blei *et al.* [2]. In LDA, the input data is assumed to be organized in a number of discrete data sets—these correspond to scan segments in our application. The scan segments contain a set of discretized features (a spin image for every 3D point). Obviously, a feature can have multiple *occurrences* since different 3D data points might have the same spin image. Often, the full set of data (from multiple scans) is referred to as "corpus". A key feature of LDA is that it does not require a distance metric between features as most approaches to unsupervised clustering do. Instead, LDA uses the co-occurrence of features in scan segments to assign them probabilistically to classes—called *topics* in this context.

Being a generative probabilistic model, the basic assumption made in LDA is that the scan segments are generated by random processes. Each random process represents an individual topic. In this work, we distinguish topics using the index j and scan segments are indexed by d. A random process generates the features in the segments by sampling them from its own specific discrete probability distribution $\phi^{(j)}$ over the features. A segment can be created by one or more topics, each topic having associated a distinct probability distribution over the features.

To represent the mixture of topics in a segment d, a multinomial distribution $\theta^{(d)}$ is used. For each feature in the segment, the generating topic is selected by sampling from $\theta^{(d)}$. The topic mixture $\theta^{(d)}$ itself is drawn from a Dirichlet distribution once for every segment in the corpus. The Dirichlet distribution represents the prior belief about the topic mixtures that occur in the corpus, i.e., whether the segments are generated by single topics or from a mixture of many topics. We express the prior belief with respect to the topic distribution using the Dirichlet parameter vector α.

Griffiths and Steyvers [9] extended LDA by additionally specifying a Dirichlet prior $Dir(\beta)$ on the conditional distributions $\phi^{(j)}$ over the features. This prior is useful in our application since it enables us to model a preference for selecting few characteristic features of a topic.

C. Learning the Model

In this section, we describe how to find the assignments of topics to 3D data points in range scans following the derivation of Griffiths and Steyvers [9]. Given the corpus $\mathbf{w} = \{w_1, w_2, ...w_n\}$ as the set of all feature occurrences, where each occurrence w_i belongs to exactly one scan segment. We are then looking for the most likely topic assignment vector $\mathbf{z} = \{z_1, z_2, ...z_n\}$ for our data $\mathbf{w}$. Here, each z_i is an index referring to topic j that generated w_i. Hence, we seek to estimate the probability distribution $P(\mathbf{z} \mid \mathbf{w})$. Based on $P(\mathbf{z} \mid \mathbf{w})$, we can then obtain the most likely topic assignment for each 3D data point. Using Bayes rule, we know that

$$P(\mathbf{z} \mid \mathbf{w}) = \frac{P(\mathbf{w} \mid \mathbf{z})P(\mathbf{z})}{P(\mathbf{w})}. \quad (4)$$

Unfortunately, the partition function $P(\mathbf{w})$ is not known and cannot be computed directly because it involves T^N terms, where T is the number of topics and N is the number of feature occurrences.

A common approach to approximate a probability distribution, for which the partition function $P(\mathbf{w})$ is unknown, is Markov chain Monte Carlo (MCMC) sampling. MCMC approximates the target distribution $P(\mathbf{z} \mid \mathbf{w})$ by randomly initializing the states of the variables—here the topic assignments. Subsequently, it samples new states using a Monte Carlo transition function leading to the target distribution. Therefore, the target distribution has to be the equilibrium distribution of the transition function. The transition function obeys the Markov property, i.e., it is independent of all states but the last. In our approach, we use Gibbs sampling as the transition function where the new state (the topic assignment) for each feature occurrence is sampled successively.

Gibbs sampling requires a *proposal distribution* to generate new states. Therefore, the next section describes how to obtain an appropriate proposal distribution for our problem.

D. Computing the Proposal Distribution for Gibbs Sampling

The proposal probability distribution over the possible topic assignments of a feature occurrence is calculated conditioned on the current assignments of the other feature occurrences. A new topic assignment is then sampled from this proposal distribution.

For estimating $P(\mathbf{z} \mid \mathbf{w})$, we successively sample from the distribution in the numerator on the right hand side of Eq. (4) the topic assignment z_i for each feature occurrence w_i given the topics of all other features. The distribution over the topics for sampling z_i is given by

$$P(z_i = j \mid \mathbf{z_{-i}}, \mathbf{w}) = \frac{\overbrace{P(w_i|z_i = j, \mathbf{z}_{-i}, \mathbf{w}_{-i})}^{\text{likelihood of } w_i} \overbrace{P(z_i = j|\mathbf{z}_{-i})}^{\text{prior of } z_i}}{\sum_{j=1}^{T} P(w_i|z_i = j, \mathbf{z}_{-i}, \mathbf{w}_{-i})P(z_i|\mathbf{z}_{-i})}. \quad (5)$$

In Eq. (5), $\mathbf{w}_{-i}$ denotes the set $\mathbf{w}$ without w_i and $\mathbf{z_{-i}}$ the corresponding assignment vector. We can express the conditional distributions in the nominator of Eq. (5) by integrating over ϕ and θ, where ϕ denotes the feature distribution of all topics and θ denotes the topic distribution for each scan segment.

The likelihood of w_i in Eq. (5) depends on the probability of the distribution of topic j over features, so we need to integrate over all these distributions $\phi^{(j)}$:

$$P(w_i = w \mid z_i = j, \mathbf{z}_{-i}, \mathbf{w}_{-i}) =$$
$$\int \underbrace{P(w_i = w \mid z_i = j, \phi^{(j)})}_{\phi_w^{(j)}} \underbrace{P(\phi^{(j)} \mid \mathbf{z}_{-i}, \mathbf{w}_{-i})}_{\text{posterior of } \phi^{(j)}} d\phi^{(j)} \quad (6)$$

Since the Dirichlet distribution is conjugate to the multinomials (to which $\phi^{(j)}$ belongs to), this posterior can be computed easily from the prior and the observations by adding the observations to the respective elements of the parameter vector β of the prior (see also Section IV-A). As a result, we obtain a Dirichlet posterior with parameter vector $\beta + n_{-i,j}^{(w)}$

where the elements of $n_{-i,j}^{(w)}$ are the number of occurrences of feature w assigned to topic j by the assignment vector $\mathbf{z}_{-i}$.

The first term on the right hand side of Eq. (6) is the probability for feature w under the multinomial $\phi^{(j)}$ and the second term denotes the probability of that multinomial. Therefore, solving this integral results in computing the expectation of $\phi_w^{(j)}$ which is the probability of w under $\phi^{(j)}$. According to Eq. (3), this expectation can be easily computed. The probability that an occurrence w_i is feature w is

$$P(w_i = w \mid z_i = j, \mathbf{z}_{-i}, \mathbf{w}_{-i}) = \mathbb{E}(\phi_w^{(j)}) = \frac{n_{-i,j}^{(w)} + \beta_w}{\sum_{w'} n_{-i,j}^{(w')} + \beta_{w'}}. \quad (7)$$

In the same way, we integrate over the multinomial distributions over topics θ, to find the prior of z_i from Eq. (5). With d_i being the index of the scan segment to which w_i belongs, we can compute the probability of a topic assignment for feature occurrence w_i as:

$$P(z_i = j \mid \mathbf{z}_{-i}) = \int \underbrace{P(z_i = j \mid \theta^{(d_i)})}_{\theta_j^{(d_i)}} \underbrace{P(\theta^{(d_i)} \mid \mathbf{z}_{-i})}_{\text{posterior of } \theta^{(d_i)}} d\theta^{(d_i)} \quad (8)$$

Let $n_{-i,j}^{(d_i)}$ be the number of features in the scan segment d_i that are assigned to topic j. Then, analogous to Eq. (7), the expected value of $\theta_j^{(d_i)}$ can be calculated by adding $n_{-i,j}^{(d_i)}$ to the elements of the parameter vector α of the prior:

$$P(z_i = j \mid \mathbf{z}_{-i}) = \mathbb{E}(\theta_j^{(d_i)}) = \frac{n_{-i,j}^{(d_i)} + \alpha_j}{\sum_{j'} n_{-i,j'}^{(d_i)} + \alpha_{j'}} \quad (9)$$

Combining the results of Eq. (7) and (9) in Eq. (5), we obtain the *proposal distribution* for the sampling of z_i as

$$P(z_i = j \mid \mathbf{z}_{-i}, \mathbf{w}) \quad \propto \quad \frac{n_{-i,j}^{(w)} + \beta_w}{\sum_{w'} n_{-i,j}^{(w')} + \beta_{w'}} \frac{n_{-i,j}^{(d_i)} + \alpha_j}{\sum_{j'} n_{-i,j'}^{(d_i)} + \alpha_{j'}} . \quad (10)$$

Eq. (10) is the proposal distribution used in Gibbs sampling to obtain next generation of assignments.

After a random initialization of the Markov chain, a new state is generated by drawing the topic for each feature occurrence successively from the proposal distribution. From these samples, the distributions θ and ϕ can be estimated by using the sampled topic assignments $\mathbf{z}$.

Note that in our work, we restrict the Dirichlet priors to be symmetric. This implies that all topics and all features have the same initial prior occurrence probability. As a result, we only have to specify only value for the elements of the parameter vectors α and β which we denote by $\hat{\alpha}$ and $\hat{\beta}$. This leads to:

$$\phi_j^{(w)} \sim \frac{n_j^{(w)} + \hat{\beta}}{\left(\sum_{w'} n_j^{(w')}\right) + W\hat{\beta}} \qquad \theta_j^{(d)} \sim \frac{n_j^{(d)} + \hat{\alpha}}{\left(\sum_{j'} n_{j'}^{(d)}\right) + T\hat{\alpha}} \quad (11)$$

where T is the number of topics and W the number of features.

To summarize, we explained how to compute the proposal distribution in Eq. (10) used in Gibbs sampling during MCMC. The obtained samples can then be used to estimate the distributions ϕ and θ. Due to our restriction to symmetric priors, only two parameters $(\hat{\alpha}, \hat{\beta} \in \mathbb{R})$ have to be specified.

E. Unsupervised Topic Discovery and Classification of Newly Observed Objects

This section briefly summarizes how the components presented so far are integrated to perform the unsupervised discovery of object classes and the classification when new observations are made.

First of all, we preprocess the data according to Section III-A to extract the scan segments which correspond to objects in the scene and for which we aim to learn a topic model. For each data point in a scan segment, we compute our feature, a variant of the spin-image, according to Section III-B to describe the surfaces characteristics.

For the discovery of topics, we then compute the feature distributions ϕ of the object classes as well as the topic mixtures θ for the scan segments using MCMC as described in the previous section. The learned distributions θ denote a probabilistic assignment of objects to topics.

Class inference, that is, the classification of objects contained in *new* scenes can be achieved using the feature distribution ϕ. In this case, ϕ and θ can be used to compute the proposal distribution directly and are not updated.

Note that the approach presented here does not automatically determine the number of object classes. This is similar to other unsupervised techniques such as k-means clustering or EM-based Gaussian mixture models in which the number of object classes is assumed to be known. We experimentally evaluated settings in which the number of topics was higher or lower than the number of manually assigned classes in the data set. Our observation was that a higher number of topics leads to the detection of shape classes such as "corner", "edge", or "flat surface" and that the objects are modeled as mixtures of those.

F. The Influence of the Dirichlet Priors $\hat{\alpha}$ and $\hat{\beta}$

Two hyperparameters $\hat{\alpha} \in \mathbb{R}$ and $\hat{\beta} \in \mathbb{R}$ need to be provided as the input to the presented approach. They define the prior distributions for the mixture of object classes in a data set and for the mixture of features in an object class respectively.

As briefly discussed in Section IV-A, choosing $\hat{\alpha}$ larger than one favors the occurrence of many topics in each scan segment, while lower values result in less topics per scan segment. Similarly, the lower the hyperparameter $\hat{\beta}$ for the Dirichlet distribution over the features, the stronger the preference for fewer features per topic and unambiguous ones. Due to the segmentation in the preprocessing step, we assume that there are only few topics per scan segment and thus a low value for the hyperparameter is favored in this setting. For $\hat{\beta}$ holds: On the one hand different objects can yield the same individual features (yet in distinct distributions). On the other hand, we expect features to be related to specific topics.

From this intuitions about the Dirichlet parameters, a high performance can be expected if both parameters are selected between zero and one. This could be confirmed experimentally and the results are given in Section V-D, where we analyze the influence of the hyperparameters on manually labeled data sets.
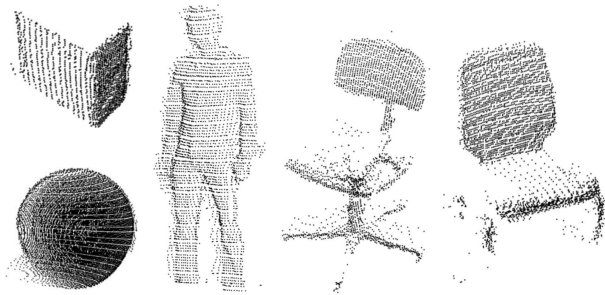
Fig. 4: Example point cloud segments of Corpus-A (box, balloon) and Corpus-B (box, balloon, human, swivel chair, chair)

V. Experimental Evaluation

In this section, we present experiments carried out to evaluate our approach on recorded data. All results are based on scans of real scenes acquired with an ActivMedia pioneer robot equipped with a SICK LMS range finder mounted on a Schunk pant-tilt unit. No simulator was involved in the evaluation.

The goal of the evaluation is to answer the following questions: (i) Are the proposed local shape features in conjunction with the topic model approach expressive enough to represent real-world objects? (ii) Is the approach able to discover object classes from unlabeled point clouds and are these classifications consistent with human-provided class labels? (iii) How does our LDA-based approach compare to conceptually simpler approaches for unsupervised clustering? (iv) How sensitive is the proposed algorithm w.r.t to the choice of parameters for the feature extraction step as well as of the Dirichlet priors?

A. Test Data

For the experimental evaluation, we prepared and re-arranged indoor scenes containing five different object types: balloons, boxes, humans, and two types of chairs. In total, we recorded 51 full laser-range scans containing 121 object instances. The first part of this data set is termed *Corpus-A*. It contains 31 object instances of low geometric complexity (different boxes and balloons). The second and larger part comprising of 82 object instances, *Corpus-B*, additionally contains complex and variable shapes of chairs and humans. See Figure 4 for examples of such object segments represented as 3D point clouds.

The data was acquired and pre-processed as described in Section III-A. Some difficulties, inherent in 3D data recorded in this way, should be pointed out: Only one side of an object can be recorded and non-convex objects typically occlude themselves partially. Objects were scanned from different view points and thus different parts are observed. Different objects of the same class were scanned (different humans, different chairs, etc.). Metal parts, such as the legs of chairs, reflect the laser beams and, thus, are invisible to the sensor. Finally, local shape features extracted from the scans of humans are highly diverse compared to the simpler objects.

Figure 5 shows typical classification results achieved by our algorithm when applied to entire scans in three example
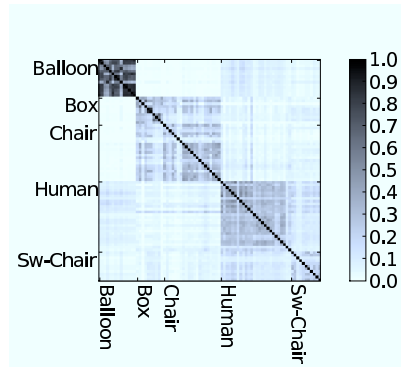

Fig. 7: Visualization of the confusion matrix of classification based on matching spin-image histograms.

scenes. Here, the points are color-coded according to their class assignments (elements of Corpus-A on the left and Corpus-B in the middle and on the right). The labels assigned to the individual points are taken from a sample of the posterior distribution $P(\mathbf{z} \mid \mathbf{w})$ as generated during the clustering process. It can be seen that the point labels are almost perfectly consistent within each object segment and, thus, the maximum likelihood class assignment per segment is unambiguous.

In addition to that, Figure 6 gives a visual impression of the topics assigned by our approach to the 82 scan segments of Corpus-B. The labels in this diagram show the true object class. Each color in the diagram denotes one topic and the ratios of colors denote for each object segment the class assignment weight. As the diagram shows, except of one chair, all objects are grouped correctly when using the maximum likelihood assignment.

We furthermore analyzed the runtime requirements of our approach, disregarding the time for pre-processing and the computation of the spin images. In Corpus-B (82 objects from 39 different 3D scans, 300 000 spin image in total), it took less than 20 s to learn the topic distributions via MCMC and to classify the objects. Thus, the computation time per 3D scan is around 500 ms which is faster than the time needed to record a 3D scan.

B. Clustering by Matching Shape Histograms

In order to compare our LDA-based approach to an unsupervised clustering technique, we implemented hierarchical clustering (HC) using the similarity between spin-image histograms as the distance metric. In this implementation, we build a feature histogram for each object segment by counting the occurrences of the individual spin-images from the (finite) spin-image dictionary (see. Section III-B). To compare two scan segments, we first normalize their histograms to sum up to one over all bins. Among the popular measures for comparing histograms, namely histogram intersection [10], χ^2 distance, and the Kullback Leibler divergence (KL-D), histogram intersection appeared to provide the best results in our domain. This is due to the fact that the χ^2 distance and the KL-D are heavily influenced by features with few or no occurrences—an effect that can be observed frequently in our data sets. The quantitative results comparing LDA to
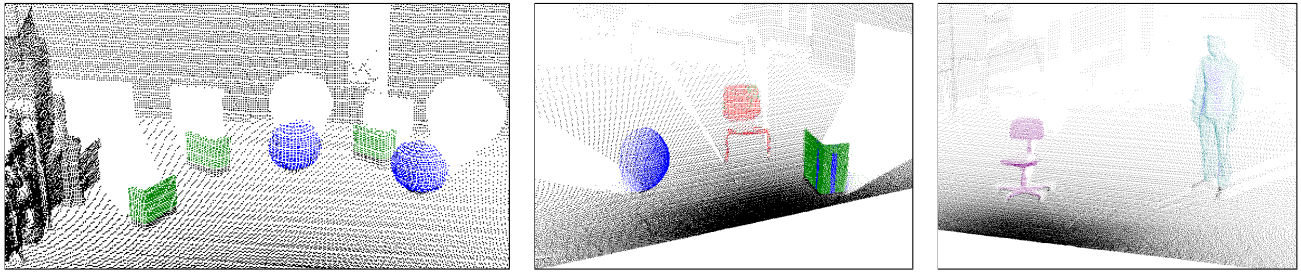
Fig. 5: Example classification results on test scans from Corpus-A (left) and Corpus-B (middle and right). The detected object classes are colored according to the LDA-assigned shape model.
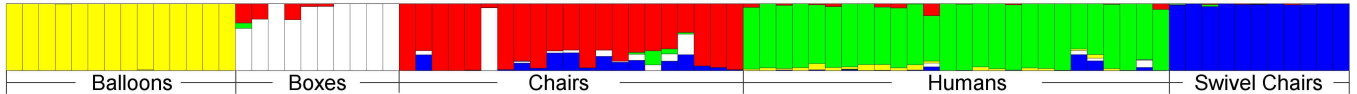

Fig. 6: Resulting topic mixtures θ for 82 segments of Corpus-B computed via LDA (the labels were not provided to the system).

HC are given in Table I. As can be seen for the simpler setting of Corpus-A, HC gives acceptable results but is still outperformed by LDA. In the more complex setting of Corpus-B, however, HC was not able to find a good clustering of the scene. In multiple runs using different setups, we found that the difference is statistically significant.

Figure 7 visualizes the similarity matrix between scan segments obtained using histogram intersection. Due to their rather uniform shape, balloons can be well distinguished from other objects. Objects with a more complex shape, however, are confused easily. This indicates that approaches working only based on such a distance metric are likely operate less accurately in more complex scenes. In contrast to that, LDA considers distributions of features and their dependencies and therefore perform substantially better.

C. Parameters of the Spin-Image Features

In this experiment, we analyzed the difference of the clustering performance when the regular spin-images (referred to as "Type 1") and our variant (referred to as "Type 2") is used. We also investigated the influence of the parameters used to create the features. These parameters are (i) the support distance, i.e., the size of the spinning image, (ii) the grid resolution, and (iii) the discretization of the stored values.

To compare the two alternative types of spin images, we collected statistics measuring the LDA clustering performance on a labeled test set, integrating over the three feature parameters. That way, we analyzed 10 780 different parameter settings—each for regular spin-images and for our variant. Figure 8 shows the results of this experiment as a histogram. The higher the bars on the right hand side of the histogram, the better the results. As can be seen, our approach outperforms
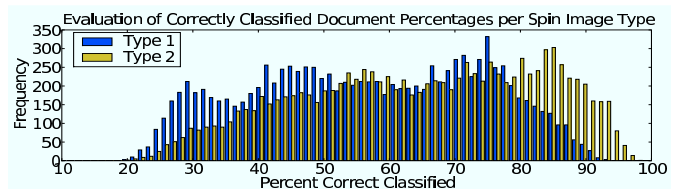
TABLE I: Summary of the classification results on the test data sets. The percentages give the average correct classifications achieved by hierarchical clustering (HC) and the proposed model based on LDA.

Data set	No. of scenes	No. of segments	HC	LDA
Corpus-A	12	31	94.84%	99.89%
Corpus-B	39	82	71.19%	90.38%

Fig. 8: Classification using standard spin-image features ("Type 1" shown in blue) generally labels less documents correctly than classification upon the features we proposed ("Type 2", yellow).
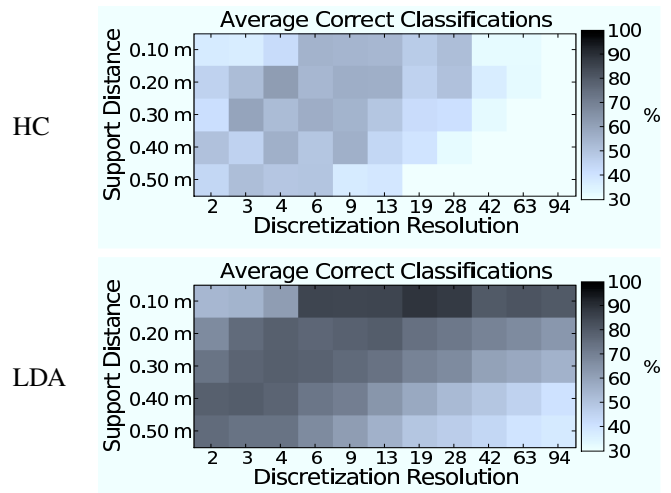

Fig. 9: Classification accuracy on Corpus-B for different discretization resolutions and respect to support distances for HC (top) and LDA (bottom).

regular spin-images.

In addition to that, we computed the clustering performance of our approach and HC for a wide variety of feature parameters using Corpus-B. Figure 9 shows the results for HC and LDA. Again, our approach clearly outperforms HC. The broad spread of high classification rates over the range of parameters demonstrates that the results presented in the previous section were not caused by selecting feature parameters that were suboptimal for HC.

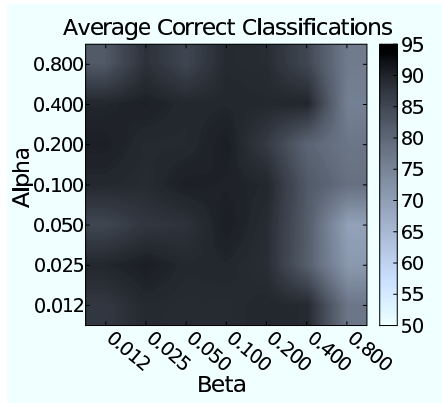We observe that for smaller support distances, a higher dis-

Fig. 10: Evaluation of classification accuracy for various values of alpha and beta.

cretization resolutions work well and vice versa. The intuition for this finding is that feature distributions with a large support and a very accurate discretization have overly detailed features, that do not match the distributions of other segments well.

The best results in our setting are obtained for features with a discretization resolution between 5 and 27 and a rather short support distance. In conclusion we see, that choosing such parameters for the feature generation, we can achieve over 90 % correct classifications (compare lower plot in Figure 9).

D. Sensitivity of the Dirichlet Priors

We furthermore evaluated how sensitive our approach is with respect to the choice of the parameters $\hat{\alpha}$ and $\hat{\beta}$ for the Dirichlet priors. Figure 10 depicts the average classification rates for varying parameters. In this plot, we integrate over the three feature parameters in a local region around the values determined in the previous experiment to illustrate how robust LDA performs. As can be seen from Figure 10, determining the hyperparameters is not a critical task since the performance stays more or less constant when varying them. Good values for $\hat{\alpha}$ lie between 0.1 and 0.8 and between 0.1 and 0.3 for $\hat{\beta}$. In these ranges, we always achieved close-to-optimal classification accuracies on labeled test sets.

VI. CONCLUSION

In this paper, we presented a novel approach for discovering object classes from laser range data in an unsupervised fashion. We use a feature-based approach that applies a novel variant of spin-images as surfaces representations but is not restricted to this kind of features. We model object classes as distributions over features and use Latent Dirichlet Allocation to learn clusters of 3D objects according to similarity in shape. The learned feature distributions can subsequently be used as models for the classification of unseen data. An important property of our approach is that it is unsupervised and does not need labeled training data to learn the partitioning.

We carried out experiments using 3D laser range data acquired with a mobile robot. Even for datasets containing complex objects with varying appearance such as humans, we achieve a robust performance with over 90% correctly

grouped objects. We furthermore demonstrate that our approach clearly outperforms unsupervised clustering approaches such as hierarchical clustering. Not only does LDA achieve higher classification accuracy throughout the entire parameter range, it is also less sensitive to the choice of parameters.

REFERENCES

[1] 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 *Proc. of the Conf. on Comp. Vision and Pattern Recognition (CVPR)*, pages 169–176, 2005.

[2] D.M. Blei, A.Y. Ng, M.I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.

[3] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via plsa. In *In Proc. ECCV*, pages 517–530, 2006.

[4] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proc. of the ACM Int. Conf. on Image and Video Retrieval*, pages 401–408, 2007.

[5] B. Bustos, D.A. Keim, D. Saupe, T. Schreck, and D.V. Vranić. Feature-based similarity search in 3d object databases. *ACM Comput. Surv.*, 37(4):345–387, 2005.

[6] R.J. Campbell and P.J. Flynn. A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, 81(2):166–210, 2001.

[7] M. Fritz and B. Schiele. Decomposition, discovery and detection of visual categories using topic models. In *Proc. of the Conf. on Comp. Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[8] M. Girolami and A. Kabán. On an equivalence between PLSI and LDA. In *Proc. of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 433–434, 2003.

[9] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proc Natl Acad Sci U S A*, 101 Suppl 1:5228–5235, 2004.

[10] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3d object recognition from range images using local feature histograms. In *Proc. of the Conf. on Comp. Vision and Pattern Recognition (CVPR)*, pages 394–399, 2001.

[11] T. Hofmann. Probabilistic latent semantic indexing. In *Proc. of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 50–57, 1999.

[12] A. Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 1997.

[13] A. Johnson and M. Hebert. Recognizing objects by matching oriented points. Technical Report CMU-RI-TR-96-04, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 1996.

[14] A.E. Johnson and M. Hebert. Surface matching for object recognition in complex three-dimensional scenes. *Image and Vision Computing*, 16:635–651, 1998.

[15] A.E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:433–449, 1999.

[16] C.H. Lampert, M.B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[17] M. Ruhnke, B. Steder, G. Grisetti, and W Burgard. Unsupervised learning of 3d object models from partial views. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009. To appear.

[18] S. Ruiz-Correa, L.G. Shapiro, and M. Meila. A new paradigm for recognizing 3-d object shapes from range data. *Computer Vision, IEEE International Conference on*, 2:1126, 2003.

[19] F. Stein and G. Medioni. Structural indexing: Efficient 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):125–145, 1992.

[20] R. Triebel, K. Kersting, and W. Burgard. Robust 3d scan point classification using associative markov networks. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006.

[21] R. Triebel, R. Schmidt, O. Martinez Mozos, and W. Burgard. Instace-based amn classification for improved object recognition in 2d and 3d laser range data. In *Proc. of IJCAI*, pages 2225–2230, 2007.

[22] P. Viola and M.J. Jones. Robust real-time object detection. In *Proc. of IEEE Workshop on Statistical and Theories of Computer Vision*, 2001.

[23] X. Wang and E. Grimson. Spatial latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, volume 20, 2007.