

A Bayesian Approach to Learning 3D Representations of Dynamic Environments

Ralf Kästner, Nikolas Engelhard, Rudolph Triebel, and Roland Siegwart

Abstract We propose a novel probabilistic approach to learning spatial representations of dynamic environments from 3D laser range measurements. Whilst most of the previous techniques developed in robotics address this problem by computationally expensive tracking frameworks, our method performs in real-time even in the presence of large amounts of dynamic objects. The computer vision community has provided comparable methods for learning foreground activity patterns in images. However, these methods generally do not account well for the uncertainty involved in the sensing process. In this paper, we show that the problem of detecting occurrences of non-stationary objects in range readings can be solved online under the assumption of a consistent Bayesian framework. Whilst the model underlying our framework naturally scales with the complexity and the noise characteristics of the environment, all parameters involved in the detection process obey a clean probabilistic interpretation. When applied to real-world urban settings, the results produced by our approach appear promising and may directly be applied to solve map building, localization, or robot navigation problems.

1 Introduction and Related Work

Understanding dynamic properties of the world has become an increasingly popular research topic in mobile robotics. The motivations for this popularity are manifold. The occurrence of moving objects in the robot's sensor range may for example corrupt the localization or map building process [12, 1, 3]. On the other hand, novel

Ralf Kästner, Rudolph Triebel, Roland Siegwart
Autonomous Systems Lab, ETH Zurich, Switzerland
e-mail: [kaestner, triebel]@mavt.ethz.ch, rsiegwart@ethz.ch

Nikolas Engelhard
University of Freiburg, 79110 Freiburg, Germany
e-mail: nikolas.engelhard@informatik.uni-freiburg.de

planning approaches aim at navigating platforms through highly dynamic environments [11, 6]. They therefore strongly rely on robust motion parameter estimates for objects that may potentially interfere with the robot’s trajectory.

A widely common group of methods addressing motion estimation is committed to tracking the displacement of entire point clusters [10]. Whilst such approaches succeed in obtaining a parametric description of the cluster motion, they usually take strong assumptions about the size or shape of objects. In the above scenarios, however, we generally do not want to constrain ourselves with a limited number of object classes. In fact, we seek to detect motion rather than to find explicit motion parameters. Consider therefore a sensor reading that has been introduced by a dynamic object. If, at any later point in time, we acquire another reading that matches the previous observation, we may not care to also answer the difficult question of identity. That is, no matter if the measurement originates from a single or two different objects, we would still want to classify it as being dynamic.

In this paper, we propose a novel approach to the problem of learning 3D representations of dynamic environments from range data. In strong analogy to background modeling in computer vision [8], this problem constitutes a binary classification task. That is, for a series of range observations we seek to estimate whether single measurements originate from a static or a dynamic object. We therefore represent correspondences between measurements and objects using Gaussian mixture distributions [14, 13].

Where standard methods for learning Gaussian mixtures fail due to the non-stationary nature of a dynamic world model [5], we propose an alternative on-line solution that does not make any assumptions about the number of Gaussians in the mixture model but efficiently scales with the complexity of the environment. Even in highly populated settings, our approach is thus capable of distinguishing dynamic from static objects in real-time.

The emphasis of this work strongly lies on the Bayesian formalization of all steps involved in the learning process [8]. In fact, following techniques used in probabilistic change detection [7] our method is strictly governed by the laws of probability, and each effective parameter comes with a clear probabilistic interpretation.

We demonstrate the practicability of our approach in simulation and by experiments involving several urban outdoor scenarios with a diversity of static structures and dynamic objects.

2 Probabilistic Formulation

Our algorithm to learning dynamic environment representations operates on range readings acquired with a nodding 2D laser range scanner that pitches up and down during data acquisition to produce 3D point clouds. This setup has been used frequently in the literature (see, e.g. [15]) and is usually known as the *nodding laser scanner* configuration. Throughout this paper, we define a sensor measurement \mathbf{z}_t as the tuple $(r_t, \vartheta_t, \varphi_t)$, where r_t is the measured range, and ϑ_t and φ_t are the pitch and

yaw angles of the laser beam at time t of the data acquisition. We assume that r_t is affected by Gaussian noise, thus we have $r_t \sim \mathcal{N}(\hat{r}_t, \sigma_r)$, where \hat{r}_t is the true distance between the laser origin and the observed object.

Since we do not account for uncertainty in the acquisition angles ϑ and φ of our nodding range scanner, we represent full sensor sweeps as *range images*. A range image is defined by an $N \times M$ matrix of cells $c_{i,j}$, representing a discretization of the continuous space $\mathcal{S} = [-\vartheta_{max}, \vartheta_{max}] \times [-\varphi_{max}, \varphi_{max}]$, where the pitch and yaw angles ϑ and φ of the laser beam range between predefined boundaries. For each \mathbf{z}_t , we can thus compute a 3D coordinate vector $\mathbf{z}_t^{[i,j]}$, with i and j denoting the indices of the range image cell $c_{i,j}$ that corresponds to r_t .

To formulate our problem mathematically, we furthermore introduce a binary state variable \mathbf{x}_t , which is true if the observation \mathbf{z}_t corresponds to a dynamic object and false otherwise. Our aim is to estimate \mathbf{x}_t at any time step t , given the noisy measurement \mathbf{z}_t acquired at time t . Formally, we therefore want to find $p(\mathbf{x}_t | \mathbf{z}_t)$. Assuming statistical independences between all range image cells, we can then estimate the joint posterior probability of range measurements being caused by dynamic objects $\bar{\mathbf{x}}_t = \{\mathbf{x}_t^{[i,j]}\}$ given the range image $\bar{\mathbf{z}}_t = \{\mathbf{z}_t^{[i,j]}\}$ as

$$p(\bar{\mathbf{x}}_t | \bar{\mathbf{z}}_t) = \prod_{i,j} p(\mathbf{x}_t^{[i,j]} | \mathbf{z}_t^{[i,j]}). \quad (1)$$

To keep the following notations uncluttered, we will drop the superindices $[i, j]$ and perform all further computations only on the cell level. Consequently, the conditional $p(\mathbf{x}_t | \mathbf{z}_t)$ associated with each cell shall from now on be referred to as *cell posterior*.

2.1 Formulation using Gaussian Mixture Models

To infer the binary states \mathbf{x}_t , we use a generative approach: we assume that each observation \mathbf{z}_t was caused by the existence of one of K objects, which can be either dynamic or static. The unobserved distance of each object to the laser origin is modeled using a normal distribution $\mathcal{N}(\mu_k, \sigma_k)$, with mean μ_k , variance σ_k , and $k \in \{1, \dots, K\}$. For each cell, we thus yield a set of model parameters which shall henceforth be denoted $\Theta = \{K, \mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K\}$. To express the fact that \mathbf{z}_t corresponds to object k , we furthermore introduce binary correspondence variables g_t^k , where only one g_t^k can be true for any \mathbf{z}_t .

2.2 Decomposing the Cell Posterior

In an online process, our model changes over time. Let Θ_t therefore represent the set of parameters at acquisition time t . Together with the states \mathbf{x}_t , the model parameters are generally unknown and need to be inferred. In other words, at any given time t we will be facing the question of how to update our latest parameter estimate Θ_{t-1} using the most recent observation \mathbf{z}_t . We therefore want to make the Θ_t and Θ_{t-1} explicit in the probabilistic formulation and rewrite the cell posterior from Eqn. (1) as

$$\begin{aligned} p(\mathbf{x}_t, \Theta_t | \mathbf{z}_t, \Theta_{t-1}) &= p(\mathbf{x}_t | \mathbf{z}_t, \Theta_t, \Theta_{t-1}) p(\Theta_t | \mathbf{z}_t, \Theta_{t-1}) \\ &= p(\mathbf{x}_t | \mathbf{z}_t, \Theta_t) p(\Theta_t | \mathbf{z}_t, \Theta_{t-1}). \end{aligned} \quad (2)$$

Here, the second equality assumes our model Θ_t to provide a complete description of the underlying process.

The above equation constitutes two conditionals that are essential to finding the cell posteriors. The first conditional describes an assignment of binary states \mathbf{x}_t to observations \mathbf{z}_t under the assumption that all model parameters Θ_t are known. It shall therefore be coined as *dynamics likelihood*. The second conditional implies the sought model Θ_t from an observation \mathbf{z}_t and the most recent parameter set Θ_{t-1} . Accordingly, it will be termed the *update rule*.

2.3 The Dynamics Likelihood

We follow the approach presented in [8] and express the dynamics likelihood by marginalization over all correspondence variables g_t^k . We thus obtain

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{z}_t, \Theta_t) &= \frac{p(\mathbf{x}_t, \mathbf{z}_t | \Theta_t)}{p(\mathbf{z}_t | \Theta_t)} \\ &= \frac{\sum_{k=1}^K p(x_t^k, \mathbf{z}_t | g_t^k, \Theta_t^k) p(g_t^k | \Theta_t^k)}{\sum_{k=1}^K p(\mathbf{z}_t | g_t^k, \Theta_t^k) p(g_t^k | \Theta_t^k)} \\ &= \frac{\sum_{k=1}^K p(x_t^k | g_t^k, \Theta_t^k) p(\mathbf{z}_t | g_t^k, \Theta_t^k) p(g_t^k | \Theta_t^k)}{\sum_{k=1}^K p(\mathbf{z}_t | g_t^k, \Theta_t^k) p(g_t^k | \Theta_t^k)}. \end{aligned} \quad (3)$$

Here, we introduce object parameters $\Theta_t^k = \{\mu_t^k, \sigma_t^k\}$ and individual state variables x_t^k expressing if the k -th Gaussian is dynamic or static. We furthermore assume the probability of observing \mathbf{z}_t to be equal for both dynamic and static objects, given the knowledge that \mathbf{z}_t was caused by object k . This results in the conditional independence relation $p(x_t^k, \mathbf{z}_t | g_t^k, \Theta_t^k) = p(x_t^k | g_t^k, \Theta_t^k) p(\mathbf{z}_t | g_t^k, \Theta_t^k)$.

From Eqn. (3), three terms need to be specified further. Starting from the right, we first note that $p(g_t^k | \Theta_t^k)$ is the *prior* probability for \mathbf{z}_t being caused by object k . Usually, $p(g_t^k | \Theta_t^k)$ is named the *weight* of the k -th Gaussian in the mixture and denoted by the symbol w_t^k . Second, the *data likelihood* $p(\mathbf{z}_t | g_t^k, \Theta_t^k)$ is equal to $\mathcal{N}(\mathbf{z}_t; \mu_t^k, \sigma_t^k)$. And, finally, the *dynamics likelihood of Gaussian k* is defined as $p(\mathbf{x}_t^k | g_t^k, \Theta_t^k)$.

2.4 The Update Rule

Just as for the dynamics likelihood, we write the update rule as a marginal over correspondences g_t^k and obtain

$$p(\Theta_t | \mathbf{z}_t, \Theta_{t-1}) = \sum_{k=1}^K p(\Theta_t^k | \mathbf{z}_t, g_t^k, \Theta_{t-1}^k) p(g_t^k | \mathbf{z}_t, \Theta_{t-1}^k) \quad (4)$$

This provides us with two additional terms. The *correspondence likelihood* $p(g_t^k | \mathbf{z}_t, \Theta_{t-1}^k)$ constitutes a statistical law for selecting Gaussian k as an explanation for the occurrence of \mathbf{z}_t , and the *update rule of Gaussian k* is denoted by $p(\Theta_t^k | \mathbf{z}_t, g_t^k, \Theta_{t-1}^k)$.

2.5 The Mixture Weights

As stated above, we assume that each observation \mathbf{z}_t is caused by only one possible object. This corresponds to the *hard assignment* of data points to clusters known from the k -means clustering algorithm. Using this, we can say that all t data points acquired at the discrete time steps $1, \dots, t$ are each assigned to one out of K clusters where each cluster corresponds to a Gaussian. If we denote the number of observations that correspond to cluster k at time t by n_t^k , we can estimate the prior probability of a new observation \mathbf{z}_t to be caused by object k . As this prior is equal to the weight w_t^k , we have

$$p(g_t^k | \Theta_t^k) = w_t^k = \frac{n_t^k}{t} \quad (5)$$

We note that knowing t , n_t^k thus becomes an equivalent representation of the mixture weight w_t^k .

3 Online Estimation of the Mixture Parameters

In this section, we present an online implementation for estimating our model parameters. As shown above, we can compute the posterior probability of an observation \mathbf{z}_t to be caused by a dynamic object in two stages: First, we apply the update rule according to Eqn. (4) and reestimate the mixture parameters Θ_t with respect to the latest model Θ_{t-1} . And second, we use the dynamics likelihood from Eqn. (3) on the updated mixture models in order to infer new state variables \mathbf{x}_t .

3.1 Sequential Parameter Updates

A key contribution of this paper pertains to the central question of how to integrate new observations with an existing cell mixture. In contrast to the approach presented in [14], we seek to strictly govern the sequential update of our mixture models by statistical densities and the laws of probability.

In mathematical terms, finding the optimal parameter assignment for Θ_t given Θ_{t-1} and an observation \mathbf{z}_t is equivalent to maximizing the update probability $p(\Theta_t | \mathbf{z}_t, \Theta_{t-1})$. We therefore want to reconsider the probabilistic update rule for our cell mixtures defined in Eqn. (4). It suggests that in order to maximize $p(\Theta_t | \mathbf{z}_t, \Theta_{t-1})$, we first need to recover an optimal assignment for the correspondence variables g_t^k .

We recall that following our above assumptions, each observation \mathbf{z}_t may only be caused by one possible object. Put differently, we are interested in inferring whether a sensor response \mathbf{z}_t originates from an object $k \in \{1, \dots, K\}$ that is already represented by our mixture or not. Following the proposal in [7], we therefore introduce a joint probability and estimate correspondences with respect to two distinct cases.

Explained An observation \mathbf{z}_t can be explained by the k -th Gaussian in the current mixture model. Consider therefore the range reading r_t associated with \mathbf{z}_t along with the expected measurement noise σ_r . Furthermore, let Θ_{t-1}^k be the parameters of the Gaussian k in the mixture distribution that best explains \mathbf{z}_t . Then our observation model gives rise to the assumption that $p(\mathbf{z}_t | g_t^k) \sim \mathcal{N}(\mu_{t-1}^k, \sigma_{t-1}^k + \sigma_r)$. Note that by summing up the variances of the measurement σ_r and the object representation σ_{t-1}^k , we account for the noise in both models.

Unexplained An observation cannot be explained by any of the K Gaussians in the current mixture model. Without making any specific assumption on how unobserved objects occur within the sensor range, we assume a uniform distribution over the entire beam length. Hence, we define $p(\mathbf{z}_t | g^{new}) \sim \mathcal{U}(0, r_{max})$ where \mathcal{U} denotes the uniform distribution with support in $[0, r_{max}]$, and g^{new} is a new Gaussian explaining \mathbf{z}_t .

Given the above cases, we are able to arrive at a posterior for unexplained observations. We define $p_{new} = p(g^{new})$ and apply Bayes rule to relate $p(g^{new} | \mathbf{z}_t)$ to the likelihood at which observations are generated by unrepresented objects:

$$p(g^{new} | \mathbf{z}_t) = \frac{p(\mathbf{z}_t | g^{new}) \cdot p_{new}}{p(\mathbf{z}_t | g^{new}) \cdot p_{new} + p(\mathbf{z}_t | g_t^k) \cdot (1 - p_{new})} \quad (6)$$

Then, by exploiting the assumption that $p(\mathbf{z}_t | g^{new}) \cdot p_{new}$ is small we can approximate the logarithm of this expression

$$\begin{aligned} \log p(g^{new} | \mathbf{z}_t) &\approx \log \frac{p(\mathbf{z}_t | g^{new}) \cdot p_{new}}{p(\mathbf{z}_t | g_t^k) \cdot (1 - p_{new})} \\ &= \log p(\mathbf{z}_t | g^{new}) + \log p_{new} - \\ &\quad \log p(\mathbf{z}_t | g_t^k) - \log(1 - p_{new}) \end{aligned} \quad (7)$$

Plugging in our model assumptions with respect to the described cases consequently yields

$$\begin{aligned} \log p(g^{new} | \mathbf{z}_t) &\approx -\log r_{max} + \log p_{new} - \log(1 - p_{new}) + \\ &\quad \frac{1}{2} \log 2\pi(\sigma_{t-1}^k + \sigma_r) + \frac{1}{2} \frac{(r_t - \mu_{t-1}^k)^2}{(\sigma_{t-1}^k + \sigma_r)} \end{aligned} \quad (8)$$

It appears beneficial to combine all expressions depending on the range measurement r_t in the above equation. This leaves us with a simple quadratic distance

$$d_k(r_t) = (r_t - \mu_{t-1}^k)^T (\sigma_{t-1}^k + \sigma_r)^{-1} (r_t - \mu_{t-1}^k) \quad (9)$$

Exploiting the assumption that an unexplained observation with a probability of $p(g^{new} | \mathbf{z}_t) > 0.5$ is significant, this distance may then be compared to the following constant threshold

$$\begin{aligned} d_k^{min} &= 2 \log 0.5 + 2 \log r_{max} - 2 \log p_{new} + \\ &\quad 2 \log(1 - p_{new}) - \log 2\pi(\sigma_{t-1}^k + \sigma_r) \end{aligned} \quad (10)$$

We conclude that if $d_k(r) < d_k^{min}$, the k -th Gaussian is a possible explanation for the occurrence of observation \mathbf{z}_t .

The reader may have noticed that the method presented so far only allows for selecting a set of candidate Gaussians from the mixture. We therefore propose to proceed as follows: If we find any Gaussian explaining \mathbf{z}_t , we will arrange a hard assignment of \mathbf{z}_t to the candidate Gaussian k with the lowest distance $d_k(r_t)$. Note that this is equivalent to selecting the object with the highest correspondence likelihood $p(g_t^k | \mathbf{z}_t) = 1 - p(g^{new} | \mathbf{z}_t)$. We then account for maximizing Eqn. (4) by computing the *maximum likelihood* estimate for the k -th Gaussian. A sequential approach exists to finding the maximum likelihood solution for the parameters of a

Algorithm 1: updateMixture(\mathcal{M}_{t-1}, r_t)

Input: Mixture of Gaussians $\mathcal{M}_{t-1} = \{\langle n_{t-1}^1, \theta_{t-1}^1 \rangle, \dots, \langle n_{t-1}^K, \theta_{t-1}^K \rangle\}$
Input: Laser reading r_t
Output: Updated Mixture of Gaussians \mathcal{M}_t

$\mathcal{M}_{cand} \leftarrow \emptyset$
foreach $\langle n_{t-1}^k, \theta_{t-1}^k \rangle \in \mathcal{M}_{t-1}$ **do**
 Evaluate $d_k(r_t)$ and d_k^{min} according to Eqs. 9 and 10
 if $d_k(r_t) < d_k^{min}$ **then**
 | $\mathcal{M}_{cand} \leftarrow \mathcal{M}_{cand} \cup \{\langle n_{t-1}^k, \theta_{t-1}^k \rangle\}$
 end
end
if $\mathcal{M}_{cand} \neq \emptyset$ **then**
 $\langle n_{t-1}^k, \theta_{t-1}^k \rangle \leftarrow \underset{\langle n_{t-1}^k, \theta_{t-1}^k \rangle \in \mathcal{M}_{cand}}{\operatorname{argmin}} d_k(r_t)$
 $\mathcal{M}_t \leftarrow \mathcal{M}_{t-1} \setminus \{\langle n_{t-1}^k, \theta_{t-1}^k \rangle\}$
 $n_t^k \leftarrow n_{t-1}^k + 1$
 $\theta_t^k \leftarrow \operatorname{updateGaussian}(\theta_{t-1}^k, r_t)$
 $\mathcal{M}_t \leftarrow \mathcal{M}_t \cup \{\langle n_t^k, \theta_t^k \rangle\}$
else
 $n_t^{K+1} \leftarrow 1$
 $\theta_t^{K+1} \leftarrow \{\mu_t^{K+1} \leftarrow r_t, \sigma_t^{K+1} \leftarrow \sigma_r\}$
 $\mathcal{M}_t \leftarrow \mathcal{M}_{t-1} \cup \{\langle n_t^{K+1}, \theta_t^{K+1} \rangle\}$
end
collapseMixture(\mathcal{M}_t)

Gaussian distribution that allows new observations \mathbf{z}_t to be processed one at a time. For a detailed discussion of this approach, the interested reader may refer to [2].

In cases where no candidate Gaussian exists, we represent the observed object by introducing a new Gaussian $K + 1$ into the mixture. This Gaussian is then initialized with mean r_t and variance σ_r .

We are now ready to state Alg. 1 for sequential parameter updates. It takes the range measurement r_t associated with a new observation \mathbf{z}_t and the corresponding cell mixture distribution \mathcal{M}_{t-1} as input arguments and in return outputs the updated density \mathcal{M}_t .

Our algorithm makes use of two auxiliary functions. As the name suggests, `updateGaussian` sequentially reestimates mean and variance of the best candidate Gaussian with respect to r_t . The second function of concern is `collapseMixture`. It provides an abstract mechanism for joining Gaussians that share a significant fraction of the state space. To see why this is necessary, the reader may consider the very nature of the mixture density $p(\mathbf{z}_t | \theta_t)$. This density actually constitutes a noise model of the environment with the sensor uncertainty displayed by newly added Gaussians solely acting as a prior. In fact, the variance of each Gaussian may often grow beyond this initial uncertainty, e.g. in order for it to represent a highly scattered surface.

Although alternative approaches exist to collapsing Gaussian mixture models (see e.g. [9], p. 185 ff.), we propose to use a method that widely resembles our

sequential update algorithm. Just as for associating noisy observations with the most likely mixture candidates, we have implemented a very similar algorithm to perform a pair-wise identification of Gaussians obeying the distance threshold from Eq. 10. The advantage of such an implementation is clear at hand: Instead of introducing a new parameter into the update step, we may reuse the prior probability p_{new} for the occurrence of unexplained objects. Hence, the total number K of Gaussians contained in a cell mixture is bound by our model assumptions and does not require artificial clamping.

3.2 Computing the Dynamics Likelihood

Above, we have demonstrated how to update the parameters of a Gaussian mixture distribution that is capable of representing observations in statistically independent range image cells. However, our model does not yet disambiguate between observations caused by dynamic and observations caused by static objects. Or formally speaking, we have not yet provided an estimate for the binary state variables x_t^k associated with each of the Gaussians.

From the dynamics likelihood stated in Eqn. (3), we know that the conditional density $p(x_t^k | g_t^k, \Theta_t^k)$ defines a state labeling strategy and that, in its most general form, this strategy constitutes a weighting of the k -th Gaussian in the mixture distribution.

To compute the dynamics likelihood of Gaussian k , we use the method described in [14]. It is based on the observation that the majority of range readings usually originates from static objects. This means in turn that the objects, to which the most observed data points correspond, are more likely to be static. As we model each hypothetical occurrence of an object with a Gaussian k , and as according to Eqn. (5) the number of observations caused by the k -th object is encoded in the mixture weight w_k , we can estimate $p(x_t^k | g_t^k, \Theta_t^k)$ as follows: First, we sort all weights w_k in descending order. Then, we compute the number K_S of Gaussians that most probably correspond to static objects as

$$K_S = \operatorname{argmin}_l \left\{ \sum_{k=1}^l w_{s(k)} > \rho \right\}. \quad (11)$$

Here, $s(k)$ is the index of the weight w_k after sorting and $\rho \in [0, 1]$ is an environment-dependent parameter that represents a measure for the minimum portion of observations that should be accounted for by static hypotheses. The last $K - K_S$ Gaussians in the sorted mixture are consequently labeled as dynamic. Thus, the lower the value of ρ , the more Gaussians are considered to be dynamic. For $\rho = 0$, only the most evident Gaussian, i.e. the one with the highest count of observations, remains static. Using Eqn. (11), the dynamics likelihood is approximated as

$$p(\mathbf{x}_t^k | g_t^k, \Theta_t^k) = \begin{cases} 1 & \text{if } s(k) > K_S \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

3.3 Computational Complexity

We briefly want to discuss the expected computational costs of the proposed method. We therefore analyze the steps taken in order to update the model parameters Θ_t given a new observation \mathbf{z}_t and for recalculating the state variables \mathbf{x}_t .

Reestimating the cell mixture is largely dominated by the search for the best candidate distribution and by the pair-wise collapsing of Gaussians. As the average number of Gaussians \hat{K} contained in the cells varies with the choice of p_{new} , complexity strongly depends on our prior expectation about the occurrence of world dynamicity. We may however state that the update costs are bound by a recursive collapse of the entire mixture distribution into a single Gaussian density. And hence, Alg. 1 runs in worst-case $O(\hat{K}^2 \log \hat{K})$.

Estimating the states requires an additional sorting of the updated cell mixture which takes additional effort in $O(\hat{K} \log \hat{K})$.

4 Evaluation

In order to evaluate the approach proposed in this paper, we have conducted several experiments based on both simulated and real-world data. In this section, we present our results and some major insights originating from the analysis of these results.

First, we want to consider the influence of the parameter p_{new} on the quality of our on-line mixtures estimates. We will then discuss the performance of our method as applied to range observations from different outdoor settings.

4.1 Simulations and the Influence of p_{new}

To assess the theoretical soundness of our approach, we have initially generated various sets of sample observations from a predetermined ground truth. This ground truth was composed of Gaussian mixture distributions of which the parameters were known. Drawing samples from such mixtures is straightforward, and these samples may directly serve as simulated range measurements to the estimation process.

As a similarity measure between the simulated and the estimated mixture distributions, we have chosen to adapt a variational approximation to the Kullback-Leibler divergence for Gaussian mixture models. This approximation was first proposed in [4] and provides a fairly accurate, closed-form distance function.

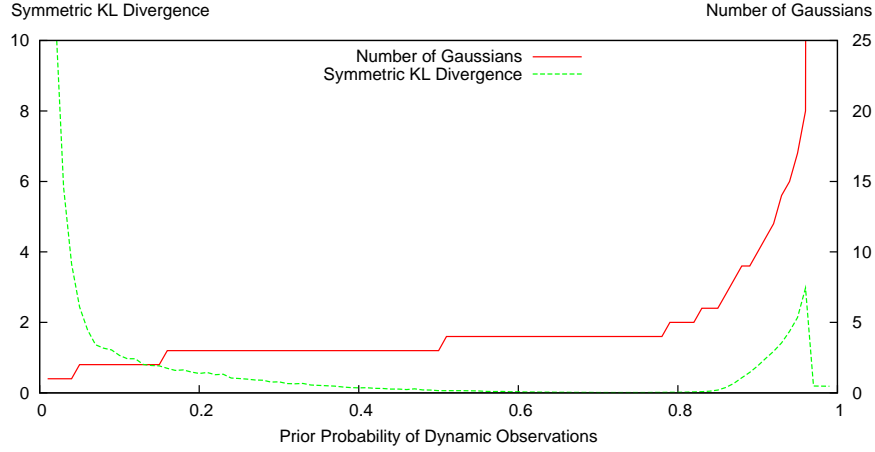


Fig. 1 The influence of the prior probability p_{new} on the size and the quality of the learned mixture distributions. The graphs display the number of estimated Gaussians and the approximated distance to the ground truth.

Consider two Gaussian mixtures \mathcal{M}_a and \mathcal{M}_b with weights w_a^k and w_b^l , and densities \mathcal{N}_a^k and \mathcal{N}_b^l , respectively. Then, the variational approximation of the distance $d_{var}(\mathcal{M}_a||\mathcal{M}_b)$ between \mathcal{M}_a and \mathcal{M}_b is given by

$$d_{var}(\mathcal{M}_a||\mathcal{M}_b) = \sum_k w_a^k \log \frac{\sum_{k'} w_a^{k'} e^{-d_{KL}(\mathcal{N}_a^k||\mathcal{N}_a^{k'})}}{\sum_l w_b^l e^{-d_{KL}(\mathcal{N}_a^k||\mathcal{N}_b^l)}} \quad (13)$$

where $d_{KL}(\mathcal{N}_k||\mathcal{N}_l)$ denotes the Kullback-Leibler divergence between normal densities \mathcal{N}_k and \mathcal{N}_l .

The symmetric form of the approximated distance shall then be defined as

$$d_{sym}(\mathcal{M}_a||\mathcal{M}_b) = \frac{d_{var}(\mathcal{M}_a||\mathcal{M}_b) + d_{var}(\mathcal{M}_b||\mathcal{M}_a)}{2} \quad (14)$$

One of the major objectives of this evaluation was to examine the influence of the prior probability p_{new} on the resulting estimates. In the course of our analysis, we have therefore repeatedly sampled 1000 data points from a ground truth consisting of 4 Gaussians with varying mean and variance parameters. Each of those sample sets was then used to infer a new mixture under the assumption of different values for p_{new} . The number of Gaussians in the resulting mixtures and the approximated distance between the ground truth and the estimates are illustrated in Fig. 1.

A deeper investigation of the graphs in this figure reveals that it can be divided into four major parts: The Gaussians in the estimates tend to associate easily for very small values of p_{new} . Under these conditions, the number of independent state hy-

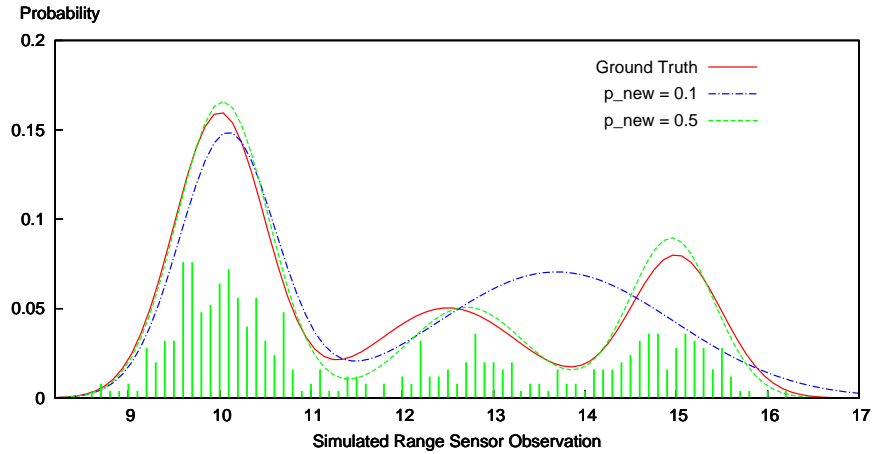


Fig. 2 The influence of p_{new} by example. A small prior probability causes estimates to generalize over the ground truth. But for a wide parameter range, a fairly accurate regression of the sample distribution is achieved.

potheses is underestimated, but the model might excellently generalize over ground truth distribution with many individual densities. An example of this behavior is depicted in Fig. 2.

For slightly higher values of p_{new} up to a probability of about 0.8, the number of Gaussians then remains near constant, whereas the resulting mixture evidently achieves a fairly accurate regression of the sample distribution. This insight is also supported by Fig. 2 where we witness a tight fit between ground truth and estimate.

Assuming values above 0.8, p_{new} allows Gaussians to only associate if they are very close with respect to our probabilistic distance. This consequently leads to an overfitting behavior, resulting in a higher number of individual densities and decreasing similarity.

If p_{new} converges towards a probability of 1.0, individual Gaussians cease to associate. The scenario culminates in a trivial situation where each sample is represented by a single Gaussian. In those cases, the distance becomes negligible, but the mixture model completely explodes in complexity.

Our analysis on the influence of p_{new} on learning mixtures from simulated data advocates important insights pertaining to real-world processes. In order for the estimator to disambiguate between static and dynamic objects, it is important to adjust p_{new} according to the following criterion: If we expect very little motion in the environment, we will choose small priors such that the model generalizes well even for noisy readings of still surfaces. For widely dynamic environments, we will assume higher probabilities p_{new} to allow for larger numbers of individual and strongly discriminating hypotheses.

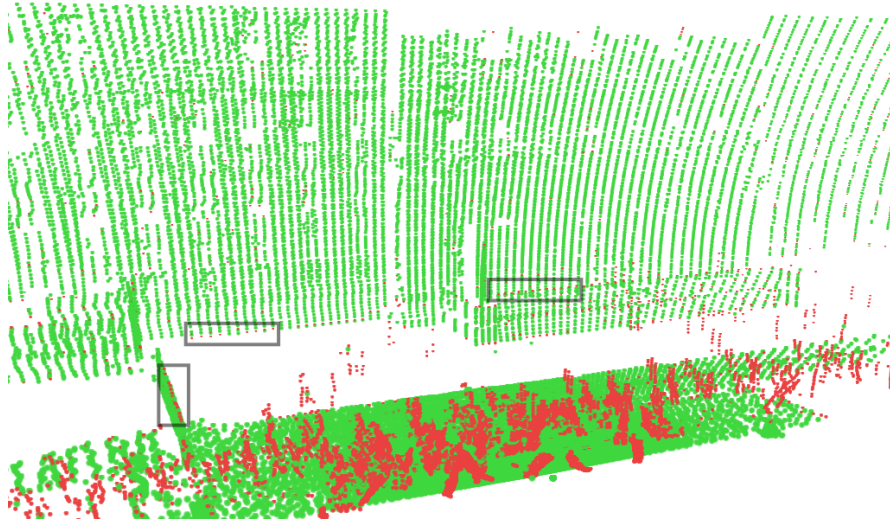


Fig. 3 Estimation results from an urban scene with pedestrians moving in front of a building. The approach robustly segments dynamic objects (red) from static background surfaces (green). The gray boxes mark problematic edges that cannot be explained by the proposed model.

4.2 Outdoor Experiments

To evaluate its real-world performance, we have furthermore applied our method to several data sets of outdoor scenarios with varying dynamic properties. Such scenarios involved different background structures as well as a selection of dynamic objects that usually move within urban settings. Amongst these objects, we found pedestrians, cars, trams, and cyclists. We have fixed our nodding range sensor with a typical pitch range of about 45 degrees in positions overlooking extensive areas. The pitch frequency was usually adjusted to about 0.5 Hz. Hence, moving objects appear slightly distorted.

Fig. 3 shows an exemplary outdoor scene and the learned environment representation. The data set used in order to produce these results is composed of continuous scans over a time period of several minutes. For the purpose of visualization, we have decided to obtain a pointcloud representation of the cell mixtures. We therefore depict the mean of a static Gaussian by a green point. Accordingly, dynamic hypotheses are marked in red.

By qualitative visual investigation, we have found that the depicted outdoor results display a robust segmentation of static and dynamic objects. Unfortunately, providing a labeled ground truth explaining our data sets is a difficult challenge that limits the feasibility of a quantitative analysis. Instead, we show an application of our approach to create a 3D grid-based Multi-level Surface (MLS) map [16] in Fig. 4. As we can see, our approach to detect and remove dynamic objects reduces the number of obstacles in the map, represented as non-traversable map cells.

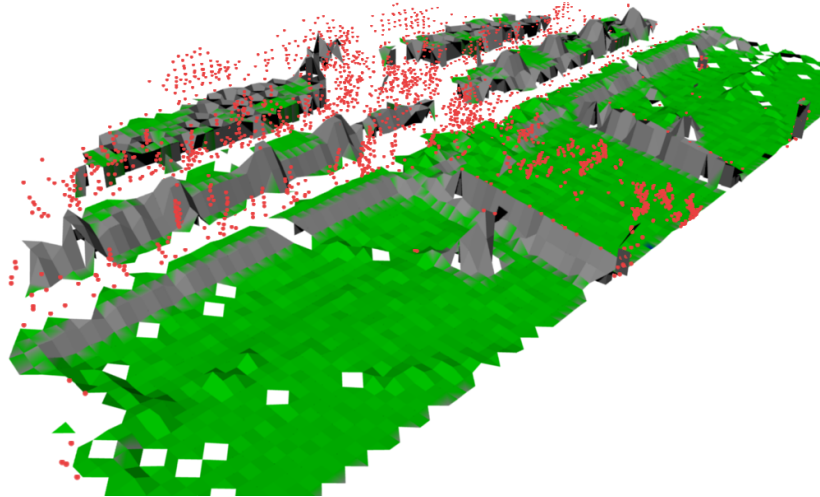


Fig. 4 Multi-level Surface map created from a different urban scene. The dynamic objects have been removed using the approach presented in this paper. Green cells are classified as traversable. The laser readings that have been classified as dynamic are overlaid to the map as red points.

The attentive observer may have noticed that some of the static regions of our estimates contain small numbers of outliers. The regions of concern are specifically characterized by edges lying in the measurement plane of the range sensor. For those edges, the laser beam has an equally distributed chance of observing a foreground or a background surface. Problematic edges hence constitute a model discontinuity which cannot be explained by our current noise assumptions. Addressing such discontinuities, we therefore propose an alternative implementation of $p(\mathbf{x}_t | \mathbf{z}_t, \theta_t)$ that takes into account the spatial neighborhood between cell models [13].

5 Conclusions

We have presented a novel approach to the difficult problem of detecting dynamic objects from range measurements. As opposed to previous work in the field, our method takes very few assumptions about the structure of the environment. Nevertheless, our estimation algorithm is strictly governed by statistical models and the laws of probability. The outdoor results produced within the scope of this paper appear promising and may directly serve as input to a variety of high-level approaches, such as map building or object tracking.

6 Acknowledgements

This work was funded within the EU Projects EUROPA-FP7-231888 and BACS-FP6-IST-027140.

References

1. P. Biber and T. Duckett. Dynamic maps for long-term operation of mobile service robots. In *Proc. of Robotics: Science and Systems (RSS)*, 2005.
2. C. Bishop et al. *Pattern Recognition and Machine Learning*. Springer New York:, 2006. pages 94-97.
3. W. Burgard, C. Stachniss, and D. Hahnel. Mobile robot map learning from range data in dynamic environments. *Springer Tracts in Advanced Robotics*, 35, 2007.
4. J. Hershey and P. Olsen. Approximating the Kullback Leibler divergence between Gaussian mixture models. In *Proc. of The International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 317–320, 2007.
5. S. Hou and A. Galata. Robust estimation of Gaussian mixtures from noisy input data. In *Proc. of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.
6. B. Jensen, R. Philippsen, and R. Siegwart. Motion detection and path planning in dynamic environments. In *Workshop Proceedings Reasoning with Uncertainty in Robotics, International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.
7. R. Kaestner, S. Thrun, M. Montemerlo, and M. Whalley. A non-rigid approach to scan alignment and change detection using range sensor data. In *Field and Service Robotics, Springer Tracts in Advanced Robotics*, volume 25/2006, pages 179–194. Springer Press, 2006.
8. D. Lee, J. Hull, and B. Erol. A Bayesian framework for Gaussian mixture background modeling. In *Proc. of The IEEE International Conference on Image Processing*, volume 3, pages 973–976, 2003.
9. U. Lerner. *Hybrid Bayesian Networks for Reasoning about Complex Systems*. PhD thesis, Stanford University, 2002.
10. M. Luber, K. Arras, C. Plagemann, and W. Burgard. Classifying dynamic objects: An unsupervised learning approach. *Robotics: Science and Systems IV*, page 270, 2009.
11. N. Roy, W. Burgard, D. Fox, and S. Thrun. Coastal navigation: Mobile robot navigation with uncertainty in dynamic environments. In *IEEE International Conference on Robotics and Automation*, pages 35–40. Citeseer, 1999.
12. D. Schulz and W. Burgard. Probabilistic state estimation of dynamic objects with a moving mobile robot. *Robotics and Autonomous Systems*, 34(2-3):107–115, 2001.
13. Y. Sheikh and M. Shah. Bayesian object detection in dynamic scenes. In *Proc. of The IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, page 74, 2005.
14. C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.
15. H. Surmann, A. Nüchter, and J. Hertzberg. An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. *Journal of Robotics and Autonomous Systems (JRAS)*, 45(3-4), 2003.
16. R. Triebel, P. Pfaff, and W. Burgard. Multi-level surface maps for outdoor terrain mapping and loop closing. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2006.