Monocular Range Sensing: A Non-Parametric Learning Approach

Christian Plagemann Felix Endres Jürgen Hess Cyrill Stachniss Wolfram Burgard

***Abstract*— Mobile robots rely on the ability to sense the geometry of their local environment in order to avoid obstacles or to explore the surroundings. For this task, dedicated proximity sensors such as laser range finders or sonars are typically employed. Cameras are a cheap and lightweight alternative to such sensors, but do not directly offer proximity information. In this paper, we present a novel approach to learning the relationship between range measurements and visual features extracted from a single monocular camera image. As the learning engine, we apply Gaussian processes, a non-parametric learning technique that not only yields the most likely range prediction corresponding to a certain visual input but also the predictive uncertainty. This information, in turn, can be utilized in an extended grid-based mapping scheme to more accurately update the map. In practical experiments carried out in different environments with a mobile robot equipped with an omnidirectional camera system, we demonstrate that our system is able to produce proximity estimates with an accuracy comparable to that of dedicated sensors such as sonars or infrared range finders.**

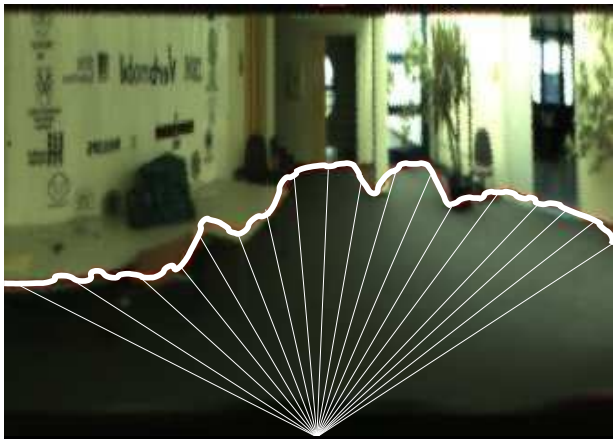


Fig. 1. Our approach estimates proximity information from a single image after having learned how visual appearance is related to depth.

I. Introduction

Cameras have become popular sensors in the robotics community. Compared to proximity sensors such as laser range finders, they have the advantage of being cheap, lightweight, and energy efficient. The drawback of cameras, however, is the fact that due to the projective nature of the image formation process, it is not possible to sense depth information directly. From a geometric point of view, one needs at least two images taken from different locations to recover the depth information analytically. An alternative approach, that requires just one monocular camera and that we follow in this work, is to learn from previous experience how visual appearance is related to depth. Such an ability is also highly developed in humans who are able to utilize monocular cues for depth perception [22].

As a motivating example, consider Figure 1, which depicts the (warped) image of an office environment. Overlayed in white, we visualize the most likely area of free space that is predicted by our approach. We explicitly do not try to estimate a depth map for the whole image, as for example Saxana *et al.* [18]. Rather, we aim at learning the function that, given an image, maps measurement directions to their corresponding distances to the closest obstacles. We believe that such a function can be utilized to solve various tasks of mobile robots including local obstacle avoidance, localization, mapping, exploration, or place classification.

The authors are with the University of Freiburg, Department of Computer Science, Georges-Koehler-Allee 79, 79110 Freiburg, Germany {plagem, endres, hess, stachnis, burgard} @ informatik.uni-freiburg.de

In this paper, we formulate the range estimation task as a supervised regression problem, in which the training set is build by acquiring images of the environment as well as proximity data provided by a laser range finder. We discuss how appropriate visual features can be extracted from the images using algorithms for edge detection and dimensionality reduction. We apply Gaussian processes as the learning framework in our proposed system, since this technique is able to model non-linear functions, offers a direct way of estimating uncertainties for its predictions, and it has proven successful in previous work involving range functions [15].

The paper is organized as follows. After discussing related work, we formalize our problem and introduce the proposed learning framework in Section III. In Section IV we then discuss appropriate visual features and how they can be extracted from images. Section V presents the experimental evaluation of our algorithm as well as an application to the mapping problem. Finally, we conclude in Section VI and give an outlook to future research.

II. Related Work

The problem of recovering geometric properties of a scene from visual measurements is one of the fundamental problems in computer vision and is also frequently addressed in the robotics literature. Stereo camera systems are widely used to estimate the missing depth information that single cameras cannot provide directly. Stereo systems either require a careful calibration to analytically calculate depth using geometric constraints or, as Sinz *et al.* [20] demonstrated,

can be used in combination with non-linear, supervised learning approaches to recover depth information. Often, sets of features such as SIFT [12] are extracted from two images and matched against each other. Then, the feature pairs are used to constrain the poses of the two camera locations and/or the point in the scene that corresponds to the image feature. In this spirit, the motion of the camera is considered by [5], [21]. Sim and Little [19] present a stereo-vision based approach to the SLAM problem, which also includes the recovery of depth information. Their approach contains both the matching of discrete landmarks as well as dense grid mapping using vision cues.

An active way of sensing depth using a single monocular camera is known as *depth from defocus* [8] or *depth from blur*. Corresponding approaches typically adjust the focal length of the camera and analyze the resulting local changes in image sharpness. Torralba and Oliva [24] present an approach for estimating the mean depth of full scenes from single images using spectral signatures. While their approach is likely to improve a large number of recognition algorithms by providing a rough scale estimate, the spatial resolution of their depth estimates does not appear to be sufficient for the problem studied in this paper. Dahlkamp *et al.* [3] learn a mapping from visual input to road traversability in a self-supervised manner.

The problem dealt with in this paper, is closely related to the work of Saxena *et al.* [18], who utilize Markov random fields (MRFs) for reconstructing dense depth maps from single monocular images. An alternative approach that predicts 2D range scans based using reinforcement learning techniques has been presented by Michels et al. [13]. Compared to these methods, our Gaussian process formulation provides the predictive uncertainties for the depth estimates directly, which is not straightforward to achieve in an MRF model. Hoiem *et al.* developed an approach to monocular scene reconstruction based on local features combined with global reasoning [11]. Whereas Han and Zhu presented a Bayesian method for reconstructing the 3D geometry of wire-like objects in simple scenes [10], Delage *et al.* introduced an MRF model on orthogonal plane segments to recover the 3D structure of indoor scenes [6].

As mentioned above, one potential application of the approach described in this paper is to learn occupancy grid maps. This type of maps and an algorithm to update such maps based on ultrasound data has been introduced by Moravec and Elfes [14]. In the past, different approaches to learn occupancy grid maps from stereo vision have been proposed [23], [17]. If the positions of the robot are unknown during the mapping process, the entire task turns into the so-called simultaneous localization and mapping (SLAM) problem. Vision-based techniques have been proposed by Elinas *et al.* [7] and Davison *et al.* [5] to solve this problem. In contrast to the mapping approach presented in this paper, these techniques mostly focus on landmark-based representations.

III. Learning Depth from Monocular Vision Features

The goal of this work is to learn the relationship between visual input and the extent of free space around the robot. By using a regular range sensors, it is comparably easy to acquire training data for this task, so that we can formulate the problem as a supervised learning problem. Figure 2 (a) depicts the configuration of our robot used for data acquisition. An omnidirectional camera system (catadioptric with a parabolic mirror) is mounted on top of a SICK laser range finder. This setup allows the robot to perceive the full surrounding area at every time step as the two example images in Figure 2 (b) and (c) illustrate. The omnidirectional images can be mapped directly to the laser scans, since both measurements can be represented in a common, polar coordinate system. Note that our approach is not restricted to omnidirectional cameras in principle. However, the correspondence between range measurements and omnidirectional images is a more direct one and the field of view is considerably larger compared to standard perspective optics.

A. A Gaussian Process Model for Range Functions

We extract for every viewing direction α a vector of visual features $\mathbf{x}$ from the images and phrase the problem as learning the range function $f(\mathbf{x}) = y$ that maps the visual input $\mathbf{x}$ to distances y. We learn this function in a supervised manner using a training set $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ of observed features $\mathbf{x}_i$ and corresponding laser range measurements y_i. If we place a Gaussian process (GP) prior [16] on the non-linear function f, i.e., we assume that all range samples y indexed by their corresponding feature vectors $\mathbf{x}$ are jointly Gaussian distributed, we obtain

$$f(\mathbf{x}^*) \quad \sim \quad \mathcal{N}(\boldsymbol{\mu}, \sigma) \tag{1}$$

with

$$\boldsymbol{\mu} \quad = \quad \mathbf{k}^{*T}(K + \sigma_n^2 I)^{-1}\mathbf{y} \tag{2}$$

$$\sigma \quad = \quad k(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}^{*T}(K + \sigma_n^2 I)^{-1}\mathbf{k}^* \tag{3}$$

as the predictive distribution for the range function at new query points $\mathbf{x}^*$. Here, K denotes the $n \times n$-dimensional covariance matrix constructed as $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ using a covariance function k, which is parameterized by a set of hyper-parameters $\boldsymbol{\theta}$. The term $\mathbf{y}$ denotes the vector of given target values from the training set, $\mathbf{k}^*$ stands for the vector of covariances between the new query point $\mathbf{x}^*$ and the training points with $\mathbf{k}_i^* = k(\mathbf{x}^*, \mathbf{x}_i)$. Finally σ_n denotes a global noise parameter. In this work, we apply the often-used squared exponential covariance function

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \cdot \exp\left(-\frac{1}{2\ell^2}|\mathbf{x}_p - \mathbf{x}_q|\right), \tag{4}$$

which depends on the Euclidian distance between points $\mathbf{x}_p$ and $\mathbf{x}_q$ as well as on the amplitude parameter σ_f^2 and the length-scale ℓ. These parameters as well as the noise parameter σ_n in Eq. (2) and (3) can be learned from data. Starting from an initial guess, we apply conjugate gradient-based optimization to find the values for $\{\ell, \sigma_f^2, \sigma_n^2\}$ that

(a) (b) (c)

Fig. 2. The left diagram depicts our experimental setup: the training set was recorded using a mobile robot equipped with an omnidirectional camera (monocular camera with a parabolic mirror) as well as a laser range finder. The next two images illustrate two typical omnidirectional images recorded at the University of Freiburg (b) and at the DFKI in Saarbruecken (c).

minimize the negative log marginal likelihood of the GP model.

A particularly useful property of Gaussian processes for our application is the availability of the predictive uncertainty (see Eq. (3)) at every query point. This means, that visual features $\mathbf{x}^*$, which lie close to points $\mathbf{x}$ of the training set result in more confident predictions than features, which fall into a less densely sampled region of feature space.

IV. Features in Omnidirectional Images

The part of an omnidirectional image which is most strongly correlated with the distance to the nearest obstacle in a certain direction α is the strip of pixels oriented in the same direction and going from the center of the image to its margins. With the type of camera used in our experiments, such strips have a dimensionality of 420 (140 pixels, each having a *hue*, *saturation*, and a *value* component). In order to make these strips easier accessible to filter operators, we warp the omnidirectional images (e.g., see Figure 2 (b) and (c)) into panoramic views (e.g., see Figure 5 (a)), such that angles in the polar representation now correspond to column indices in the panoramic one. This transformation allows us to replace complicated image operations in the polar domain by easier and more robust ones. In the following, we describe several ways of extracting useful low-dimensional feature vectors $\mathbf{x}$ from the 420-dimensional image columns, which can then be directly used to index the training and test targets in the GP framework.

1) Unsupervised Dimensionality Reduction: As a classic way of reducing the complexity of a data set, we applied the principle component analysis (PCA) to the raw 420-dimensional pixel vectors that are contained in the columns of the panoramic images. The PCA is implemented using eigenvalue decomposition of the covariance matrix of the training vectors. It yields a linear transformation which brings the input vectors into a new basis such that their dimensions are now ordered by the amount of data-set variance they carry. In this way, we can truncate the vectors to a few components without losing a large amount of information. The diagram in Figure 3 depicts the relative amount of variance that is explained for two different data sets when truncating the transformed data vectors after a certain number of components. In the experiments reported below, we trained the PCA on 600 input images and retained the first six principle components. Our experiments revealed that the *value* channel of the visual input is more important than *hue* and *saturation* for our task. The GP model learned with these 6-dimensional features is termed *PCA-GP* in the experimental section.

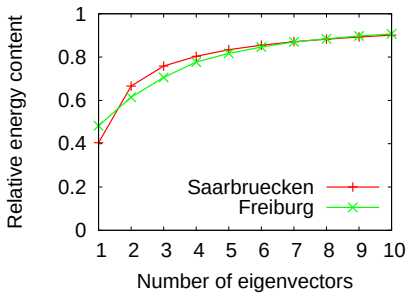


Fig. 3. The amount of variance explained by the the first principle components (eigenvectors) of the pixel columns in the two data sets.

2) Edge-based Features: The PCA is an unsupervised method that does not take into account prior information that might be available about the task to be solved – in this case, the fact that distances to the closest obstacles are to be predicted. Driven by the observation that, especially in indoor environments, there is a strong correlation between the extent of free space and the presence of horizontal edge features in the panoramic image, we also assessed the potential of edge-type features in our approach.

Laws' convolution masks [4] provide an easy way of constructing local feature extractors for discretized signals. The idea is to define three basic convolution masks

- $L_3 = (1, 2, 1)^T$ (Weighted Sum: Averaging),
- $E_3 = (-1, 0, 1)^T$ (First difference: Edges),
- $S_3 = (-1, 2, -1)^T$ (Second difference: Spots),

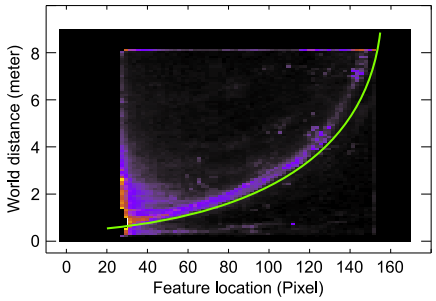


Fig. 4. Feature histogram for *Laws5+LMD* edge features. Each cell in the histogram is indexed by the pixel location of the edge feature (x-axis) and the length of the corresponding laser beam (y-axis). The optimized (parametric) mapping function that is used as a benchmark in our experiments is overlaid in green.

each having a different effect on (1-dimensional) patterns, and to construct more complex filters by a combination of the basic masks. In our application domain, we obtained good results with the (2-dimensional) directed edge filter $E_5 L_5^T$, which is the outer product of E_5 and L_5. Here, E_5 is a convolution of E_3 with L_3 and L_5 denotes L_3 convolved with itself. After filtering with this mask, we apply an optimized threshold to yield a binary response. This feature type is denoted as *Laws5* in the experimental section. As another well-known feature type, we applied the $E_3 L_3^T$ filter, i.e., the Sobel operator, in conjunction with Canny's algorithm [2]. This filter yields binary responses at the image locations with maximal grey-value gradients in gradient direction. We denote this feature type as *Laws3+Canny* in Section V. For both edge detectors, *Laws5* and *Laws3+Canny*, we search along each image column for the first detected edge. This pixel index then constitutes the feature value.

To increase the robustness of the edge detectors described above, we applied *lightmap damping* as an optional preprocessing step to the raw images. This means that, in a first step, a copy of the image is converted to gray scale and strongly smoothed with a Gaussian filter, such that every pixel represents the brightness of its local environment. This is referred to as the *lightmap*. The brightness of the original image is then scaled with respect to the lightmap, such that the *value* component of the color is increased in dark areas and decreased in bright areas. In the experimental section, this operation is marked by adding *+LMD* to the feature descriptions.

All parameters involved in the edge detection procedures described above were optimized to yield features that lie as close as possible to the laser end points projected onto the omnidirectional image using the acquired training set. For our regression model, we can now construct 4-dimensional feature vectors $\mathbf{x}$ consisting of the Canny-based feature, the Laws5-based feature, and both features with additional preprocessing using lightmap-damping. Since every of these individual features captures slightly different aspects of the visual input, the combination of all in what we call the *Feature-GP* yields more accurate predictions than any single one.

As a benchmark for predicting range information from edge features, we also evaluated the individual features directly. For doing so, we fitted a parametric function to training samples of feature-range pairs. This mapping function computes for each pixel location of an edge feature the length of the corresponding laser beam. The diagram in Figure 4 depicts the feature histogram for the *Laws5+LMD* features from one of our test runs that was used for the optimization. The color of a cell (x, y) in this diagram encodes the relative amount of features that were extracted at the pixel location x (measured from the center of the omnidirectional image) and that have a corresponding laser beam with a length of y in the training set. The optimized projection function is overlayed in green.

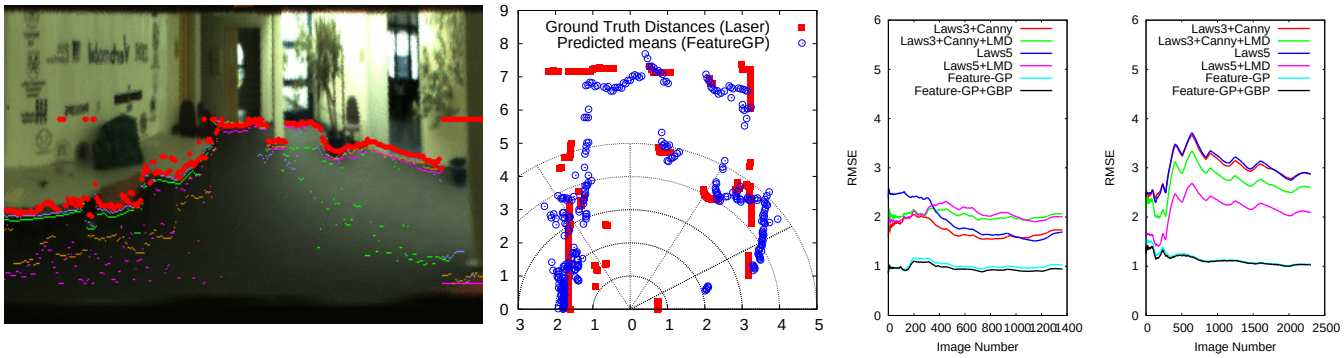
V. Experiments

The experiments presented in this section are designed to evaluate how well the proposed system is able to estimate range data from single monocular camera images. We document a series of different experiments: First, we evaluate the accuracy of the estimated range scans using the individual edge features directly, the *PCA-GP*, and the *Feature-GP*, which constitutes our regression model with the 4 edge-based vision features as input dimensions. Then, we illustrate how these estimates can be used to build grid maps of the environment. We also evaluated, whether applying the GBP model [15] as a post-processing step to the predicted range scans can further increase the prediction accuracy. The GBP model places a Gaussian process prior on the range function (rather than on the function that maps features to distances) and, thus, also models angular dependencies. We denote these models by *Feature-GP+GBP* and *PCA-GP+GBP*.

The two data sets used for the experiments have been recorded using a mobile robot equipped with a laser scanner, an omnidirectional camera, and odometry sensors at the AIS lab at the University of Freiburg (Figure 2 (b)) and at the DFKI lab in Saarbrücken (Figure 2 (c)). The two environments have quite different characteristics – especially in the visual aspects. While the environment in Saarbrücken mainly consists of solid, regular structures and a homogeneously colored floor, the lab in Freiburg exhibits many glass panes, an irregular, wooden floor, and challenging lighting conditions.

A. Accuracy of Range Predictions

We evaluated eight different system configurations, each on both test data sets. Table I summarizes the average RMSE (root mean squared error) obtained for the individual scenarios. The error is measured as the deviation of the range predictions using the visual input from the corresponding laser ranges recorded by the sensor. The first four configurations, referred to as C1 to C4, apply the optimized mapping functions for the different edge features (see Figure 4). Depending on the data, the features provide estimates with an RMSE of between 1.7 m and 3 m. We then evaluated the configurations C5 and C6 which use the four edge-based features as inputs to a Gaussian process model as described



(a) (b) (c) (d)

Fig. 5. (a) Estimated ranges projected back onto the camera image using the feature detectors directly (small dots) and using the *Feature-GP* model (red points). (b) Prediction results and the true laser scan at one of the test locations. The evolution of the root mean squared error (RSME) for the individual images of the Saarbrücken (c) and Freiburg (d) data sets.

in Section III to learn the mapping from the feature vectors to the distances. The learning algorithm was able to perform range estimation with an RMSE of around 1 m. Note that we measure the prediction error relative to the recorded laser beams rather than to the true geometry of the environment. Thus, we report a conservative error estimate that also includes errors due to reflected laser beams contained in the test set. To give a visual impression of the prediction accuracy of the *Feature-GP*, we give a typical laser scan and the mean predictions in diagram (b) of Figure 5.

TABLE I

AVERAGE ERRORS OBTAINED WITH THE DIFFERENT METHODS

Configuration	RMSE on test set Saarbrücken	Freiburg
C1: Laws5	1.70m	2.87m
C2: Laws5+LMD	2.01m	2.08m
C3: Laws3+Canny	1.74m	2.87m
C4: Laws3+Canny+LMD	2.06m	2.59m
C5: Feature-GP	1.04m	1.04m
C6: Feature-GP+GBP	**1.03m**	**0.94m**
C7: PCA-GP	1.24m	1.40m
C8: PCA-GP+GBP	1.22m	1.41m

As configuration C7, we evaluated the *PCA-GP* approach that does not require engineered features, but rather works on the low-dimensional representation of the raw visual input computed using the PCA. The resulting 6-dimensional feature vector is used as input to the Gaussian process model. With an RMSE of 1.2 m to 1.4 m, the *PCA-GP* outperforms all four engineered features, but is not as accurate as the *Feature-GP*. For configurations C6 and C8, we predicted the ranges per scan using the two different methods and additionally applied the GBP model [15] to incorporate angular dependencies between the predicted beams. This post-processing step yields slight improvements compared to the original variants C5 and C7.

Figure 5 (a) depicts an example images showing the predictions based on the individual vision features and the *Feature-GP*. It can be clearly seen from the image, that the different edge-based features model different parts of the range scan well. The *Feature-GP* fuses these unreliable estimates to achieve high accuracy on the whole scan. The result of the *Feature-GP+GBP* variant for the same situation is given in Figure 1. The evolution of the RMSE for the different methods over time is given in Figures 5 (c) and (d). As can be seen from the diagrams, the prediction using the *Feature-GP* model outperforms the other techniques and achieves a near-constant error rate.

B. Application to Mapping

Our approach can be applied to a variety of robotics tasks such as obstacle avoidance, localization, or mapping. To illustrate this, we show how to learn a grid map of the environment from the predictive range distributions. Compared to occupancy grid mapping where one estimates for each cell the probability of being occupied or free, we use the so called *reflection probability maps*. A cell of such a map models the probability that a laser beam passing this cell is reflected or not. Reflection probability maps, which are learned using the so called *counting model*, have the advantage of requiring no hand-tuned sensor model such as occupancy grid maps (see [1] for further details). The reflection probability m_i of a cell i is given by $m_i = \alpha_i/(\alpha_i + \beta_i)$ where α_i is the number of times an observation hits the cell, i.e., ends in it, and β_i is the number of misses, i.e., the number of times a beam has intercepted a cell without ending in it. Since our GP approach does not estimate a single laser end point, but rather a full (normal) distribution $p(z)$ of possible end points, we have to integrate over this distribution. More precisely, for each grid cell c_i, we update the cell's reflectance values according to the predictive distribution $p(z)$ according to the following formulas:

$$\alpha_i \leftarrow \alpha_i + \int_{z \in c_i} p(z)\, dz \tag{5}$$

$$\beta_j \leftarrow \beta_i + \int_{z > c_i} p(z)\, dz\ . \tag{6}$$

Note that for perfectly accurate predications, the extended update rule is equivalent to the standard formula stated above.

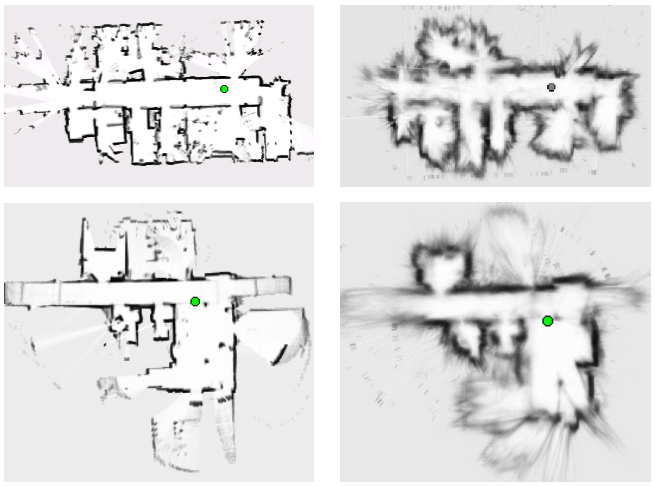


Fig. 6. Maps of the Freiburg AIS lab (top row) and DFKI Saarbrücken (bottom row) using real laser data (left) and the predictions of the *Feature-GP* (right).

We applied this extended reflection probability mapper to the trajectories and range predictions that resulted from the experiments reported on above. Figure 6 gives the laser-based maps using a standard mapper (left column) and the extended mapper using the predicted ranges (right column) for both environments (Freiburg on top and Saarbrücken below). In both cases, it is possible to build an accurate map, which is comparable to maps obtained with infrared proximity sensors [9] or sonars [23].

VI. Conclusions

We presented a novel approach for predicting range functions from single images recorded with a monocular camera. Our model is based on a Gaussian process model for regression, utilizing edge-based features extracted from the image or, alternatively, using the PCA to find a low-dimensional representation of the visual input in an unsupervised manner. Both models outperform the optimized individual features. We showed in experiments with a real robot that the range predictions are accurate enough to feed them into an extended mapping algorithm for predictive range distributions and that the resulting maps are comparable to maps obtained with infrared or sonar sensors.

In future research we would like to evaluate alternative techniques for dimensionality reduction, especially those taking the actual task into account (supervised PCA, LDA) or others that are directly integrated into the GP framework. Furthermore, we would like to evaluate our approach in other robotics tasks, such as exploration or place classification.

Acknowledgments

We would like to thank Kristian Kersting for the fruitful discussions and Andrzej Pronobis and Jie Luo for creating the data sets. This work has partly been supported by the EC under contract numbers FP6-004250-CoSy, FP6-IST-034120, and FP6-IST-045144, by the DFG under contract number SFB/TR-8, and by the German Ministry for Education and Research (BMBF) through the DESIRE project.

References

[1] W. Burgard, C. Stachniss, and D. Haehnel. *Autonomous Navigation in Dynamic Environments*, volume 35 of *STAR Springer tracts in advanced robotics*, chapter Mobile Robot Map Learning from Range Data in Dynamic Environments. Springer Verlag, 2007.

[2] F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 679–714, 1986.

[3] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G.R. Bradski. Self-supervised monocular road detection in desert terrain. In *Proc. of Robotics: Science and Systems (RSS)*, 2006.

[4] E. R. Davies. Laws texture energy in texture. In *Machine Vision: Theory, Algorithms, Practicalities*. Acedemic Press, 1997.

[5] A. Davision, I. Reid, N. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29(6), 2007.

[6] E. Delage, H. Lee, and A.Y. Ng. Automatic single-image 3d reconstructions of indoor manhattan world scenes. In *Proceedings of the 12th International Symposium of Robotics Research (ISRR)*, 2005.

[7] P. Elinas, R. Sim, and J. J. Little. σSLAM: Stereo vision SLAM using the rao-blackwellised particle filter and a novel mixture proposal distribution. In *Proc. of ICRA*, 2006.

[8] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):406–417, 2005.

[9] Y.S. Ha and H.H. Kim. Environmental map building for a mobile robot using infrared range-finder sensors. *Advanced Robotics*, 18(4):437–450, 2004.

[10] F. Han and S.-C. Zhu. Bayesian reconstruction of 3d shapes and scenes from a single image. In *IEEE Intern. Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis (HLK*, page 12, Washington, DC, USA, 2003.

[11] D. Hoiem, A.A. Efros, and M. Herbert. Recovering surface layout from an image. *IJCV*, 75(1), October 2007.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[13] J. Michels, A. Saxena, and A.Y. Ng. High speed obstacle avoidance using monocular vision and reinforcement learning. In *ICML*, pages 593–600, 2005.

[14] H.P. Moravec and A.E. Elfes. High resolution maps from wide angle sonar. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 116–121, St. Louis, MO, USA, 1985.

[15] C. Plagemann, K. Kersting, P. Pfaff, and W. Burgard. Gaussian beam processes: A nonparametric bayesian measurement model for range finders. In *Proc. of Robotics: Science and Systems (RSS)*, 2007.

[16] C.E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[17] K. Sabe, M. Fukuchi, J.-S. Gutmann, T. Ohashi, K. Kawamoto, and T. Yoshigahara. Obstacle avoidance and path planning for humanoid robots using stereo vision. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, New Orleans, LA, USA, 2004.

[18] A. Saxena, S.H. Chung, and A.Y. Ng. 3-d depth reconstruction from a single still image. *Intern. Journal of Computer Vision (IJCV)*, 2007.

[19] R. Sim and J. J. Little. Autonomous vision-based exploration and mapping using hybrid maps and Rao-Blackwellised particle filters. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2082–2089, 2006.

[20] F. Sinz, J. Quinonero-Candela, G. Bakir, C. Rasmussen, and M. Franz. Learning depth from stereo. In *26th DAGM Symposium*, 2004.

[21] H. Strasdat, C. Stachniss, M. Bennewitz, and W. Burgard. Visual bearing-only simultaneous localization and mapping with improved feature matching. In *Fachgespräche Autonome Mobile Systeme (AMS)*, 2007.

[22] G. Swaminathan and S. Grossberg. Laminar cortical mechanisms for the perception of slanted and curved 3-D surfaces and their 2-D pictorical projections. *J. Vis.*, 2(7):79–79, 11 2002.

[23] S. Thrun, A. Bücken, W. Burgard, D. Fox, T. Fröhlinghaus, D. Hennig, T. Hofmann, M. Krell, and T. Schimdt. Map learning and high-speed navigation in RHINO. In D. Kortenkamp, R.P. Bonasso, and R. Murphy, editors, *AI-based Mobile Robots: Case studies of successful robot systems*. MIT Press, Cambridge, MA, 1998.

[24] A. Torralba and A. Oliva. Depth estimation from image structure. *IEEE Transactions on Pattern Analysis and Machine Learning*, 2002.