Semantic Place Classification of Indoor Environments with Mobile Robots using Boosting

Axel Rottmann **Óscar Martínez Mozos** **Cyrill Stachniss** **Wolfram Burgard**
University of Freiburg, Department of Computer Science, D-79110 Freiburg, Germany
{rottmann | omartine | stachnis | burgard@informatik.uni-freiburg.de}

Abstract

Indoor environments can typically be divided into places with different functionalities like kitchens, offices, or seminar rooms. We believe that such semantic information enables a mobile robot to more efficiently accomplish a variety of tasks such as human-robot interaction, path-planning, or localization. This paper presents a supervised learning approach to label different locations using boosting. We train a classifier using features extracted from vision and laser range data. Furthermore, we apply a Hidden Markov Model to increase the robustness of the final classification. Our technique has been implemented and tested on real robots as well as in simulation. The experiments demonstrate that our approach can be utilized to robustly classify places into semantic categories. We also present an example of localization using semantic labeling.

Introduction

In the past, many researchers have considered the problem of building accurate metric or topological maps of the environment from the data gathered with a mobile robot. Only a few approaches considered the problem of integrating semantic information into a map. Whenever robots are designed to interact with their users, semantic information about places can be important. For a lot of applications, robots can improve their service if they are able to recognize places and distinguish between them. A robot that possesses semantic information about the type of the places can easily be instructed, for example, to go to the kitchen.

In this paper, we address the problem of semantic classification of the environment using range finder data and vision features. Indoor environments, like the one depicted in Figure 1, can typically be divided into areas with different functionalities such as seminar rooms, office rooms, corridors, or kitchens. Some of these places have a different structure and others can be distinguished due to their furniture. For example, the bounding box of a corridor is usually longer than the one of a room. Furthermore, a coffee machine is typically located in the kitchen.

The key idea of this paper is to classify the position of the robot based on objects extracted from the camera images and the scan obtained from the range sensor. Examples for typical observations obtained in an office environment at different locations are shown in Figure 1.

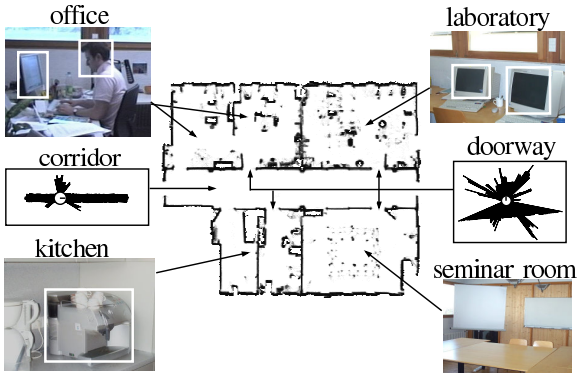

Figure 1: An environment with offices, doorways, a corridor, a kitchen, a laboratory, and a seminar room. Additionally, the figure shows typical laser and vision observations obtained by a mobile robot at different places.

Our approach uses the AdaBoost algorithm (Schapire & Singer 1999) to boost simple features extracted from laser and vision data, which on their own are insufficient for a reliable categorization of places, to a strong classifier. Since the original AdaBoost algorithm provides only binary decisions, we determine the best decision list of binary classifiers. To reduce the number of outliers during the classification, we apply a Hidden Markov Model (HMM), which filters the current classification result based on previously calculated labels. Experimental results given in this paper illustrate that the resulting classification system can determine the type of a place with a recognition rate of more than 87%.

Related Work

In the past, several authors considered the problem of adding semantic information to places. Buschka & Saffiotti (2002) describe a virtual sensor that is able to identify rooms from range data. Also Koenig & Simmons (1998) use a preprogrammed routine to detect doorways from range data. Althaus & Christensen (2003) use line features to detect corridors and doorways. Some authors also apply learning techniques to localize the robot or to identify distinctive

states in the environment. For example, Oore, Hinton, & Dudek (1997) train a neural network to estimate the location of a mobile robot in its environment using the odometry information and ultrasound data. Kuipers & Beeson (2002) apply different learning algorithms to learn topological maps of the environment.

Additionally, learning algorithms have been used to identify objects. For example, Anguelov *et al.* (2002; 2004) apply the EM algorithm to cluster different types of objects from sequences of range data. Treptow, Masselli, & Zell (2003) use the AdaBoost algorithm to track a ball without color information in the context of RoboCup. In a recent work, Torralba *et al.* (2003) use Hidden Markov Models for learning places from image data.

Compared to these approaches, our algorithm does not require any pre-defined routines for extracting high-level features. Instead, it uses the AdaBoost algorithm to boost simple features to strong classifiers for place categorization. Our approach is also supervised, which has the advantage that the resulting semantic labels correspond to user-defined classes.

In our previous work (Martínez Mozos, Stachniss, & Burgard 2005), we presented an approach which also applies the AdaBoost algorithm to classify places based on laser range data only. In this paper, we extend this approach by also taking into account vision features. This allows to distinguish between a greater variety of places, especially such with a similar geometric structure. Furthermore, we apply an HMM to filter the output of the resulting classifier which yields more robust classification results.

Boosting

Boosting is a general method for creating an accurate strong classifier by combining a set of weak classifiers. The requirement to each weak classifier is that its accuracy is better than a random guessing. In this work, we will use the AdaBoost algorithm originally introduced by Freund & Schapire (1995). The input to this algorithm is a set of labeled training examples $(x_n, y_n), n = 1, \ldots, N$, where each x_n is an example and $y_n \in \{-1, +1\}$ is a value indicating whether x_n is negative or positive respectively. In a series of rounds $t = 1, \ldots, T$, the algorithm repeatedly selects a weak classifier $h_t(x)$ using a distribution D_t over the training examples. The selected weak classifier is expected to have a small classification error in the training data. The idea of the algorithm is to modify the distribution D_t increasing the weights of the most difficult training examples on each round. The final strong classifier H is a weighted majority vote of the T best weak classifiers.

Throughout this work, we will use the approach presented by Viola & Jones (2001) in which the weak classifiers depend on single-valued features f_j. Two kinds of weak classifiers are created. The first type is used for laser and vision features and has the form

$$h_j(x) = \begin{cases} +1 & \text{if } p_j f_j(x) < p_j \theta_j \\ -1 & \text{otherwise,} \end{cases} \quad (1)$$

where θ_j is a threshold and p_j is either -1 or $+1$ and thus represents the direction of the inequality. We designed a second type of weak classifiers only for our set of vision features. These classifiers have the form

$$h_j(x) = \begin{cases} p_j & \text{if } \theta_j^1 < f_j(x) < \theta_j^2 \\ -p_j & \text{otherwise,} \end{cases} \quad (2)$$

where θ_j^1 and θ_j^2 define an interval and p_j is either $+1$ or -1 indicating whether the examples inside the interval are positive or negative.

For both types of weak classifiers, the output is $+1$ or -1 for depending on whether the classification is positive or negative. In each round $t = 1, \ldots, T$, each weak classifier $h_j(x)$ determines the optimal values for its respective parameters (p_j,θ_j) or $(p_j,\theta_j^1,\theta_j^2)$, such that the number of misclassified training examples is minimized. Then the one with smallest error is selected. The final AdaBoost algorithm is shown in Table 1 in the generalized form given by Schapire & Singer (1999) and modified for the concrete task of this work.

The approach described so far is able to distinguish between two classes of examples, namely positives and negatives. In typical indoor environments, however, we have to deal with more than two types of places. As proposed by Martínez Mozos, Stachniss, & Burgard (2005), we create a sequential multi-class classifier using $K - 1$ binary classifiers, where K is the number of classes we want to recognize. Each element in the sequence determines if an example belongs to one specific class. If the classification is positive, the example is assigned the corresponding class. Otherwise, the example is passed to the next element in the sequence.

Features from Vision and Laser Data

In this section, we describe the features used to create the weak classifiers in the AdaBoost algorithm. Our robot is equipped with a 360 degrees field of view laser sensor and a camera mounted on a pan/tilt unit. Each laser observation consists of 360 beams and each vision observation consists of 8 images which form a panoramic view. Examples of laser scans obtained in a doorway and in a corridor as well as images taken in different places are shown in Figure 1.

Each training example for the AdaBoost algorithm consist of one laser observation l, one vision observation v and its classification y. Thus, the set of training examples is given by

$$E = \{(l, v, y) \mid y \in Y = \{\text{Office}, \text{Corridor}, \ldots\}\}, \quad (3)$$

where Y is the set of classes.

In our current system, we follow the approach of our previous work (Martínez Mozos, Stachniss, & Burgard 2005) and use single-valued features extracted from laser and vision data. In the case of laser data, we extract a variety of simple geometric properties from the range scans such as the area covered by the scan or the average distance of consecutive beams. In the case of vision, the selection of the features is motivated by the fact that typical objects appear at different places with different probabilities. For example, the probability of finding a computer monitor in an office is larger than finding one in a kitchen. For each type of object, a vision feature is defined as a function that takes

Table 1: Generalized version of AdaBoost for place categorization.

- Input: Set of examples $(x_1, y_1), \ldots, (x_N, y_N)$, where $y_n = +1$ for positive and $y_n = -1$ for negative.
- Let l and m be the number of positive and negative examples respectively. Initialize weights $D_1(n) = \frac{1}{2l}, \frac{1}{2m}$ depending of the value of y_n.
- For $t = 1, \ldots, T$:
  1. Normalize the weights $D_t(n)$:
  $$D_t(n) = \frac{D_t(n)}{\sum_{i=1}^{N} D_t(i)}$$
  2. Train each weak classifier h_j using distribution D_t.
  3. For each classifier h_j calculate:
  $$r_j = \sum_{i=1}^{N} D_t(i) y_i h_j(x_i)$$
  where $h_j(x_i) \in \{-1, +1\}$.
  4. Choose the classifier h_j that maximizes $|r_j|$ and set $(h_t, r_t) = (h_j, r_j)$.
  5. Update the weights:
  $$D_{t+1}(n) = D_t(n) \exp(-\alpha_t y_n h_t(x_n))$$
  where $\alpha_t = \frac{1}{2} \log(\frac{1+r_t}{1-r_t})$.
- The final strong classifier is given by:
$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

as argument a panoramic vision observation and returns the number of detected objects of this type in it. In our current system, we consider the following types of objects: "monitor on", "monitor off", "coffee machine", "office cupboard", "frontal face", "face profile", "full human body", and "upper human body". Typical instances of these objects in our environment are shown in Figure 1. The different objects are detected using classifiers based on Haar-like features as proposed by Lienhart, Kuranov, & Pisarevsky (2003).

Probabilistic Place Recognition

The approach described so far is only able to classify single observations and does not take into account past classifications when determining the class of the current observation. In the particular domain, however, observations obtained at nearby places are typically identical. Furthermore, certain transitions between classes are rather unlikely. For example, if the classification of the current pose is "kitchen", then it is rather unlikely that the classification of the next pose is "office" given the robot moved only a short distance. To get from the kitchen to the office, the robot has to move through a doorway first.

To utilize these dependencies between the individual

Figure 2: Probabilities of possible transitions between places in the environment. To increase the visibility, we used a logarithmic scale. Dark values indicate low probability.

classes, we use a Hidden Markov Model (HMM) and maintain a posterior $Bel(\xi_t)$ about the type of the room $\xi_t \in Y$ the robot is currently in

$$Bel(\xi_t) = \alpha P(z_t|\xi_t) \sum_{\xi_{t-1}} P(\xi_t|\xi_{t-1}, u_{t-1}) Bel(\xi_{t-1}). \quad (4)$$

In this equation, α is a normalizing constant ensuring that the left-hand side sums up to one over all ξ_t. To implement such an HMM, three components need to be known. First, we need to define the observation model $P(z_t|\xi_t)$, which is the likelihood that the classification output is $z_t \in Y$ given the actual class is ξ_t. Second, we need to specify the transition model $P(\xi_t|\xi_{t-1}, u_{t-1})$, which corresponds to the probability that the robot moves from class ξ_{t-1} to class ξ_t by executing action u_{t-1}. Finally, we need to define how the belief $Bel(\xi_0)$ is initialized.

In our current system, we choose a uniform distribution to initialize $Bel(\xi_0)$. To determine the quantity $P(z_t|\xi_t)$, we generated statistics about the output of the sequential multi-class classifier given the robot was at a place corresponding to ξ_t. To realize the transition model $P(\xi_t|\xi_{t-1}, u_{t-1})$, we only consider the two actions $u_{t-1} \in \{MOVE, STAY\}$. The transition probabilities were estimated by running 1,000 simulation experiments, in which we started the robot at a randomly chosen point and orientation in the environment and commanded it to move 20-50cm forward. This value corresponds to the distance typically traveled by the robot between two consecutive updates of the HMM. The finally obtained transition probability matrix $P(\xi_t|\xi_{t-1}, u_{t-1})$ for the action $MOVE$ is depicted in Figure 2. As can be seen, the probability of staying in a place with the same classification is higher than the probability of changing the place. Moreover, the probability of moving from one room to a doorway is higher than the probability of moving from a room to a corridor. This indicates that the robot must first cross a doorway in order to reach a different room. Furthermore, the matrix shows a lower probability of staying in a doorway than moving into a room. This is due to the fact that a doorway is usually a small area in which the robot never rests for a longer period of time.

Experimental Results

The approach described above has been implemented and tested using simulated and real robot data obtained in our office environment. The goal of the experiments is to demonstrate that our approach provides a robust classification of places in indoor environments into typical categories. We

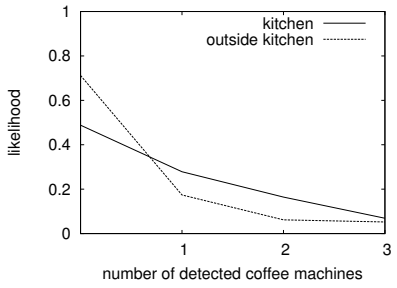

Figure 3: Likelihood of detecting n coffee machines inside and outside a kitchen using Haar-like classifiers.

Table 2: Number T of weak classifiers and training error for the individual binary classifiers.

Binary Classifier	T	Training error [%]
lab	440	0.99
corridor	165	2.02
doorway	171	2.10
kitchen	68	2.46
seminar	334	2.58
office	288	7.31

furthermore describe results indicating that the filtering of the classification output using an HMM significantly increases the performance of the overall approach. Additionally, we analyze the benefits of using vision features for the classification. Finally, we present an application example which illustrates that the semantic classification results can be used to speed-up global localization of a mobile robot.

To train the classifier used throughout the experiments, we used 38,500 training examples. For each training example, we simulated the laser observations given an occupancy grid map of the environment. To generate the features extracted from vision data, we used 350 panoramic views recorded with our B21r robot, which is equipped with a SICK laser range finder and a camera system mounted on a pan/tilt unit. Each panoramic view consists of 8 images covering the 360 degrees field of view around the robot. For each simulated laser scan, we then randomly drew a panoramic view from those corresponding to the type of the current place and used the vision features extracted from this view. Figure 3 shows two distributions over the number of coffee machines detected in the database images.

One important parameter of the AdaBoost algorithm is the number T of weak classifiers used to form the final strong binary classifier. For each strong binary classifier, we performed several experiments with up to 500 weak classifiers and analyzed the classification error. The number T of weak classifiers used to carry out the experiments has then been determined as the minimum in the error function. The resulting numbers T of weak classifiers used to form the strong binary classifiers and the classification errors of the finally obtained strong classifiers on the training data are given in Table 2.

In our current system, we determine the optimal sequence of strong binary classifiers by considering all possible sequences of strong binary classifiers. Although this approach

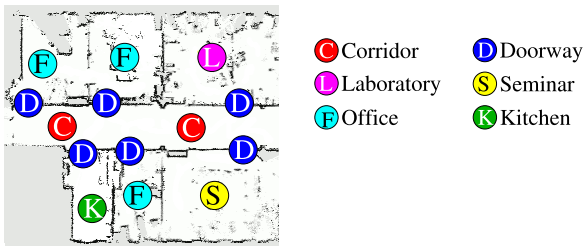

Figure 4: Ground truth labeling of the individual areas in the environment.

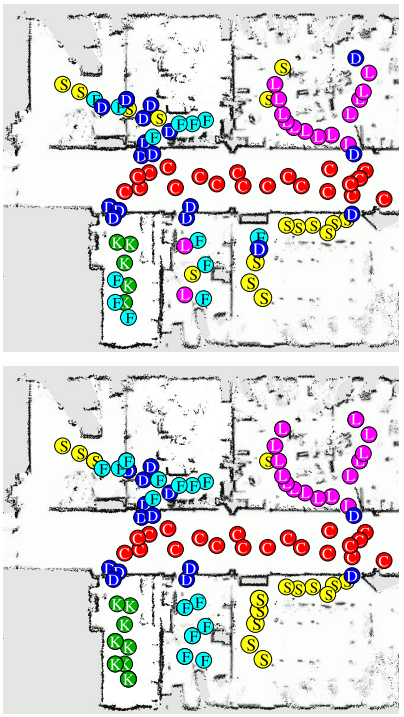

Figure 5: Typical classification obtained for a test set using only the output of the sequential classifier (top) and in combination with the HMM (bottom).

is exponential in the number of classes, the actual number of permutations considered is limited in our domain due to the small number classes. In practice, we found out that the heuristic which sorts the classifiers in increasing order according to their classification rate also yields good results and at the same time can be computed efficiently. In several situations, the sequence generated by this heuristic turned out to be the optimal one.

Classifying Places along Trajectories

The first experiment is designed to demonstrate that the classifier learned from the training data in combination with the HMM can be used to robustly classify observation sequences acquired with a mobile robot in a real office environment. This environment contains six different types of places, namely offices, doorways, a laboratory, a kitchen, a seminar room, and a corridor. The ground truth for the different places in this environment is shown in Figure 4.

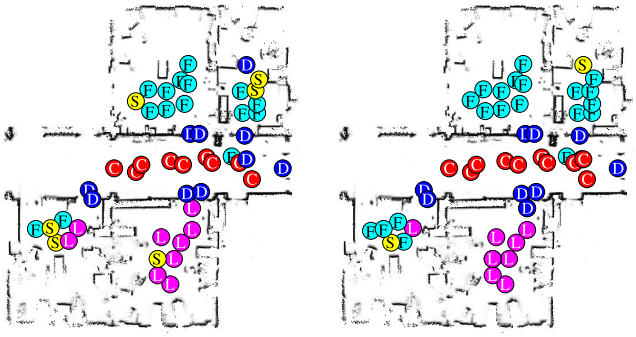

Figure 6: Classification obtained without (left) and with HMM filtering (right) for a different part of the building.

We steered our robot through the environment and collected laser and image data along its trajectory. We then calculated the classification output without and with the HMM filtering and compared this to the ground truth information.

The classification rate of the sequential classifier without applying the HMM is 74.8%. The labels generated are shown in the upper image of Figure 5. If we additionally use the HMM to filter the output of the sequential classifier, the classification rate increases to 83.8%. The labels obtained after applying the HMM are shown in the lower image of Figure 5. A two-sample t test revealed that the improvements are significant on the $\alpha = 0.01$ level. This illustrates that by using the HMM the overall classification rate can be improved seriously.

A second experiment was carried out using test data from a different part of the same building. We used the same sequential classifier as in the previous experiment. Whereas the sequential classifier yields a classification rate of 77.19%, the HMM generated the correct answer in 87.72% of all cases (see Figure 6). This improvement is also significant on the $\alpha = 0.01$ level.

Improvement Obtained by Combining Laser and Vision Data

Additionally we analyzed whether the integration of vision and laser data yields any improvements over our previous approach (Martínez Mozos, Stachniss, & Burgard 2005). To perform this experiment, we trained AdaBoost only with the three classes office, corridor, and doorway, because the other classes kitchen, seminar room, and lab can only hardly be distinguished from offices using proximity data only. The classification obtained by integrating both modalities is summarized in Table 3. As can be seen, the combination of laser and vision data yields better results than the classifier only relying on laser range data.

We furthermore evaluated how much the vision information improves the classification rate for classes that can only hardly be distinguished using laser data only. A typical example are seminar and laboratory rooms which have a similar structure and therefore cause similar laser range scans. For the seminar room the classification error decreased from 46.9% to 6.3%, and for the laboratory it decreased from 34.4% to 3.1%. This serious reduction of the classification error is due to the fact, that both rooms can mainly be distinguished by objects like monitors, which cannot be perceived with the laser scanner. A two-sample t test revealed that this improvement is significant on the $\alpha = 0.01$ level.

Table 3: Classification error obtained when using only laser data or both laser and vision data.

Sequential Classifier	Error [%] laser	Error [%] laser & vision
corridor-doorway	3.21	1.87
doorway-room	3.74	2.67
doorway-corridor	3.21	2.14
room-corridor	1.60	1.34
corridor-room	1.60	1.34
room-doorway	1.60	1.60
average	2.50	1.83

Localization Using Place Recognition

The last experiment is designed to illustrate how semantic information about places can be used to improve the localization of a mobile robot in its environment. In this experiment, we used an ActivMedia Pioneer II robot. Note that the laser data is only fed into the AdaBoost and not used for metric localization.

To estimate the pose x_t at time t of the robot, we used the popular Monte-Carlo localization approach (Dellaert *et al.* 1998), which applies a recursive Bayesian scheme to maintain a posterior about x_t given the map m of the environment, the odometry information $u_{0:t-1}$, and the observations $z_{1:t}$

$$p(x_t \mid m, z_{1:t}, u_{0:t-1}) = \eta \cdot p(z_t \mid m, x_t) \cdot p(x_t \mid m) \cdot \int_{x'} p(x_t \mid x', u_{t-1}) \cdot p(x' \mid m, z_{1:t-1}, u_{0:t-2})\, dx' . \quad (5)$$

In our application, m is a occupancy grid map, in which each cell also stores the assigned semantic class. As observations $z_{1:t}$, we use the output of the sequential classifier and determine the quantity $p(z_t \mid m, x_t)$ as $p(z_t \mid \xi_t)$, where ξ_t is the class assigned to x_t in m. Additionally, we weight the particles inversely proportional to the occupancy probability at x_t in m.

Figure 7 illustrates the evolution of two particle sets over time. In the first row, the semantic information was available whereas in the second row only the odometry information was used. Both filters were initialized with a uniform distribution with $10,000$ particles and the robot initially was located in the second left office, north of the corridor. Therefore, particles located in office received higher importance weights compared to the other samples. Whereas the approach utilizing semantic information converges quickly to the correct solution, the particle filter that solely relies on the pose information $p(x_t \mid m)$ ultimately diverged.

Conclusions

In this paper, we presented a novel approach to classify different places in the environment into semantic classes. Our technique uses a combination of simple geometric features

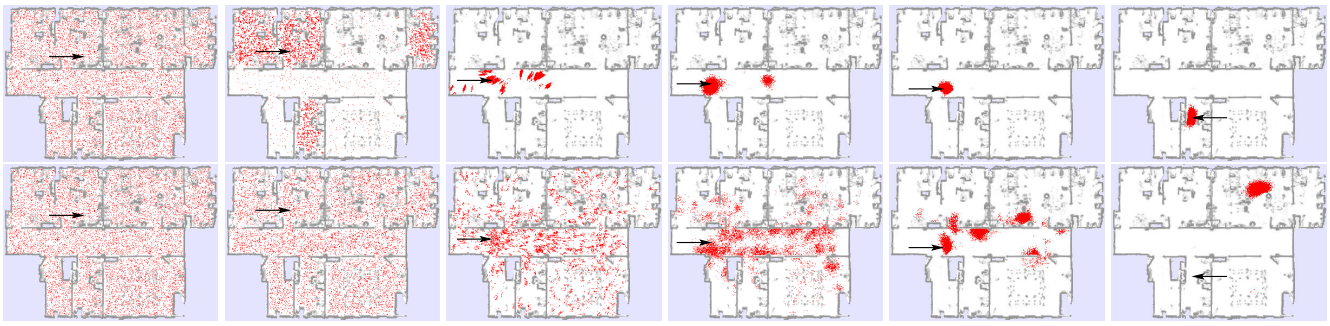

Figure 7: Global localization using semantic information and odometry (first row) compared to an approach using the odometry information only (second row). The images in one same column depict the corresponding filter at the same time. The arrow indicates the ground truth position. As can be seen, semantic information can be used to speed up global localization.

extracted from a laser range scans as well as features extracted from camera images. It applies the AdaBoost algorithm to form a strong classifier. To distinguish between more than two classes, we use a sequence of binary classifiers arranged in a decision list. To incorporate the spatial dependency between places, we apply a Hidden Markov Model that is updated upon sensory input and movements of the robot.

Our algorithm has been implemented and tested using a mobile robot equipped with a laser range finder and a camera system. Experiments carried out on a real robot as well as in simulation illustrate that our technique is well-suited to classify places in indoor environments. The experiments furthermore demonstrate that the Hidden Markov Model significantly improves the classification performance. Additional experiments revealed that the combination of vision and laser data increases the robustness and at the same time allows to distinguish between more classes compared to previous approaches. We believe that semantic information can be utilized in a variety of applications. As an example, we presented a experiment illustrating that semantic labels can be used to speed up global localization.

Acknowledgment

This work has partly been supported by the German Science Foundation (DFG) under contract number SFB/TR-8 (A3) and by the EC under contract number FP6-004250-CoSy.

References

Althaus, P., and Christensen, H. 2003. Behaviour coordination in structured environments. *Advanced Robotics* 17(7):657–674.

Anguelov, D.; Biswas, R.; Koller, D.; Limketkai, B.; Sanner, S.; and Thrun, S. 2002. Learning hierarchical object maps of non-stationary environments with mobile robots. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Anguelov, D.; Koller, D.; E., P.; and Thrun, S. 2004. Detecting and modeling doors with mobile robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*.

Buschka, P., and Saffiotti, A. 2002. A virtual sensor for room detection. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 637–642.

Dellaert, F.; Fox, D.; Burgard, W.; and Thrun, S. 1998. Monte carlo localization for mobile robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*.

Freund, Y., and Schapire, R. 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory (Eurocolt)*.

Koenig, S., and Simmons, R. 1998. Xavier: A robot navigation architecture based on partially observable markov decision process models. In Kortenkamp, D.; Bonasso, R.; and Murphy, R., eds., *Artificial Intelligence Based Mobile Robotics: Case Studies of Successful Robot Systems*. MIT-Press. 91–122.

Kuipers, B., and Beeson, P. 2002. Bootstrap learning for place recognition. In *Proc. of the Nat. Conf. on Artificial Intelligence (AAAI)*.

Lienhart, R.; Kuranov, A.; and Pisarevsky, V. 2003. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM, 25th Pattern Recognition Symposium*.

Martínez Mozos, O.; Stachniss, C.; and Burgard, W. 2005. Supervised learning of places from range data using adaboost. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*.

Oore, S.; Hinton, G.; and Dudek, G. 1997. A mobile robot that learns its place. *Neural Computation* 9(3):683–699.

Schapire, R. E., and Singer, Y. 1999. Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.* 37(3):297–336.

Torralba, A.; Murphy, K.; Freeman, W.; and Rubin, M. 2003. Context-based vision system for place and object recognition. In *Proc. of the Int. Conf. on Computer Vision (ICCV)*.

Treptow, A.; Masselli, A.; and Zell, A. 2003. Real-time object tracking for soccer-robots without color information. In *Proc. of the Europ. Conf. on Mobile Robots (ECMR)*.

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