Chapter 5
Probabilistic Semantic Classification of Trajectories*

5.1 Introduction

The approaches described in previous chapters are able to classify static observations using a mobile robot. However, mobile robots are dynamic agents that move along different trajectories. When operating in indoor environments, robots usually have a moderate velocity and a relatively continuous movement. That means, that observations obtained by a mobile robot at nearby poses are typically very similar. Furthermore, certain transitions between classes in a trajectory 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 a short distance only. To get from the kitchen to the office, the robot first has to move through a doorway.

In this chapter, we present an approach that takes into account the dependencies between the classification of consecutive poses along a trajectory. In particular, we use a hidden Markov model (HMM) to filter the output of the current classification based on previous ones. In this way, we reduce the number of false classifications during the trajectory.

In addition, this chapter presents new places to be recognized in indoor environments. In particular, we want to recognize corridors, doorways, kitchens, seminar rooms, offices, and laboratories. For this purpose, we equip the robot with an additional camera and extract new features from images that permit us to extend the classification to the new places. The features extracted from the camera images are based on the recognition of objects. As an example, Fig. 5.1 shows an office environment together with some laser and vision data.

The complete approach first classifies the poses of the robot along a trajectory using a probabilistic decision list similar to the one introduced in Sect. 4.3. Then it applies a hidden Markov model to filter the current classification result based on previous ones. As a result, the mobile robot is able to classify the different places it traverses with high confidence.

* The work presented in this chapter originated from a collaboration with Axel Rottmann.

Ó.M. Mozos: Semantic Labeling of Places with Mobile Robots, STAR 61, pp. 57–69.
springerlink.com

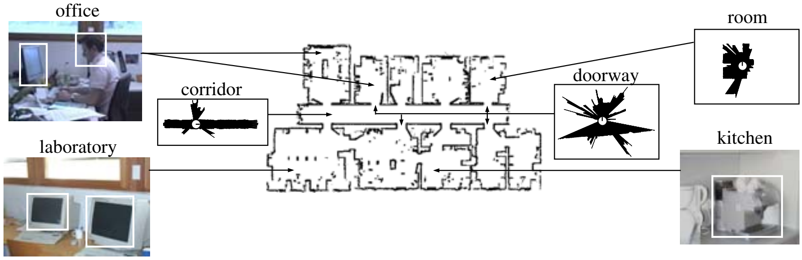


Fig. 5.1 An environment with offices, doorways, a corridor, a kitchen, and a laboratory. Additionally, the figure shows typical observations obtained by a mobile robot at different places.

The rest of the chapter is organized as follows. The next section introduces our modification of the AdaBoost algorithm to include the new weak classifiers for vision features. Section 5.3 describes the complete set of simple features extracted from laser and vision data. The models for the HMM are introduced in Sect. 5.4. In Sect. 5.5, experimental results obtained with this approach are presented. We discuss related work in Sect. 5.6. Finally, we conclude in Sect. 5.7.

5.2 Generalized AdaBoost

The generalized AdaBoost algorithm is a supervised learning algorithm designed to find a binary classifier that discriminates between positive and negative examples (cf. Sect. 2.2.2). AdaBoost boosts the classification performance of a simple learning algorithm by combining a collection of weak classifiers to a stronger one. The final strong classifier takes the form of a weighted combination of weak classifiers in which large weights are assigned to good classification functions whereas poor functions have small weights.

To classify the different places using laser and vision features, two types of weak classifiers are created. The first type is used for laser and vision features and has the form

$$h_j(x) = \begin{cases} +1 \text{ if } p_j f_j(x) < p_j \theta_j \\ -1 \text{ otherwise} , \end{cases} \tag{5.1}$$

where θ_j is a threshold, and p_j is either -1 or $+1$ and represents the direction of the inequality. Note that this form is the same as the one introduced in Sect. 4.2.

The second type of weak classifiers is designed for our set of vision features and has the form

$$h_j(x) = \begin{cases} p_j & \text{if } \theta_j^1 < f_j(x) < \theta_j^2 \\ -p_j & \text{otherwise} , \end{cases} \tag{5.2}$$

here θ_j^1 and θ_j^2 are thresholds delimiting an interval, and p_j is either $+1$ or -1 indicating whether the examples inside the interval are positive or negative. The form of this weak classifier is motivated by the fact that objects appear at different places in different numbers. For example, in an office room we expect more monitors than in the kitchen, but less than in the laboratory. Equation (5.2) is intended to encode this kind of information.

Finally, to learn the final strong classifier we use the algorithm introduced in Sect. 4.2. For the multiple class case, we use the same approach as in Sect. 4.3, and create a probabilistic decision list with binary classifiers.

5.3 Simple Features from Laser and Vision Data

In this chapter, we want to recognize an extended set of indoor places. This set is formed by corridors, doorways, kitchens, offices, laboratories, and seminar rooms. To recognize the additional places we equipped the robot with a laser scan and a camera. In the case of laser observations, the set of features are composed of the list presented in Sect. 3.4 and Sect. 4.4. These features are standard geometrical features used for shape recognition. Furthermore, they are rotational invariant to make the classification of a pose dependent only on the (x,y)-position of the robot and not on its orientation.

In the case of vision, the selection of the features is motivated by the fact that typical objects appear with different probabilities at different places. For example, the probability of detecting a computer monitor is larger in an office than in a kitchen. For each type of object, a vision feature is defined as a function that takes as argument a panoramic vision observation and returns the number of detected objects of this type in it. This number represents the single-valued feature f_j in the weak classifiers in (5.1) and (5.2).

In our case, we consider monitors, coffee machines, soap dispensers, office cupboards, frontal faces, face profiles, full human bodies, and upper human bodies. Example of such objects are shown in Fig. 5.1. The individual objects are detected using classifiers based on the set of Haar-like features proposed in [2].

5.4 Probabilistic Trajectory Classification

The approach described so far is able to classify single observations, but it does not take into account past classifications when determining the type of place at which the robot is currently located. However, whenever a mobile robot moves through an environment, the semantic labels of nearby places are typically identical. Furthermore, certain transitions between classes are unlikely. For example, if the robot is currently in a kitchen, then it is rather unlikely that the robot will end up in an office, given that it moved only a short distance. In many environments, to get from the kitchen to the office, the robot has to move through a doorway.

To incorporate such spatial dependencies between the individual classes, we apply a hidden Markov model and maintain a posterior belief $bel(y_t)$ about the type of place $y_t \in \mathrm{Y}$ the robot is currently at, where Y represents the set of possible semantic labels. The posterior is calculated as

$$bel(y_t) = \eta \, Pr(z_t \mid y_t) \sum_{y_{t-1}} Pr(y_t \mid y_{t-1}, u_{t-1}) bel(y_{t-1}). \tag{5.3}$$

In this equation, η is a normalizing constant ensuring that the left-hand side sums up to one over all classes y_t. To implement this HMM, three components need to be known. First, we need to specify the observation model $Pr(z_t \mid y_t)$, which is the likelihood of getting observation z_t given the actual class is y_t. In this case the observation z_t corresponds to the classification output of the probabilistic decision list. Second, we need to specify the transition model $Pr(y_t \mid y_{t-1}, u_{t-1})$, which defines the probability that the robot moves from class y_{t-1} to class y_t by executing action u_{t-1}. Finally, we need to specify how the belief $bel(y_0)$ is initialized.

In our current system, we choose a uniform distribution to initialize $bel(y_0)$. Furthermore, the classification output z_t is represented by a histogram P_t of probabilities over the set of all classes (cf. Sect. 4.3). In this histogram, the bin $p(k)$ stores the probability that the classified location belongs to the k-th class.

To determine $Pr(z_t \mid y_t)$, we use the KL-divergence between two histograms [1]. The first distribution is the current classification output $z_t = P_t$. The second one is learned from statistics: for each class y, we compute a histogram $\hat{z}_{1:m}(y)$ using m observations recorded within a place belonging to class y (here $m = 50$). This histogram $\hat{z}_{1:m}(y)$ is obtained by averaging out the individual histograms $\hat{z}_1, \ldots, \hat{z}_m$, which are computed according to (4.5). To determine $Pr(z_t \mid y_t)$, we use the KL-divergence $kld(\cdot \parallel \cdot)$ which provides a measure about the similarity of two distributions

$$Pr(z_t \mid y_t) = e^{-kld(z_t \parallel \hat{z}_{1:m}(y_t))} \, . \tag{5.4}$$

To illustrate the computation of the observation likelihood $Pr(z_t \mid y_t)$ consider Fig. 5.2. The first row depicts examples for the histograms $\hat{z}_{1:m}(y)$. The left image in the second row depicts the output z_t of the sequential classifier while the robot was in an office. As can be seen, also the classes doorway and seminar room have a probability significantly larger than zero. This output z_t and the histogram $\hat{z}_{1:m}(y_t)$ is then used to compute $Pr(z_t \mid y_t)$ according to (5.4). The result for all classes is depicted in the right image in the second row. In this image, each bin represents the likelihood $Pr(z_t \mid y_t)$ for the individual classes y_t. As can be seen, the observation likelihood given the robot is in a doorway is close to zero, whereas the likelihood given it is in an office is around 90%, which is actually the correct class.

To realize the transition model $Pr(y_t \mid y_{t-1}, u_{t-1})$, we only consider the two actions $u_{t-1} \in \{Move, Stay\}$. The transition probabilities were learned in a manually labeled environment by running 1000 simulation experiments. In each run, we started the robot at a randomly chosen point and orientation. We then executed a random movement so that the robot traveled between 20 cm and 50 cm forward. These values correspond to typical distances traveled by the robot between two consecutive updates of the HMM. The finally obtained transition probability matrix $Pr(y_t \mid y_{t-1}, u_{t-1})$ for the action *Move* is depicted in Fig. 5.3. As can be seen, the

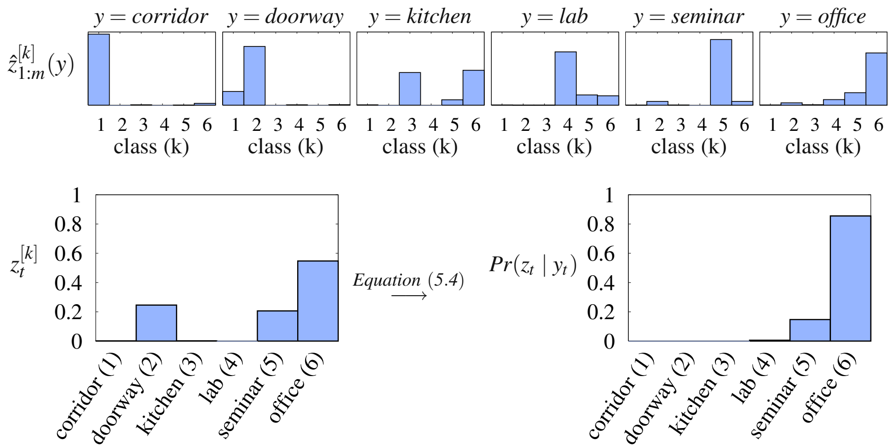


Fig. 5.2 The distributions depicted in the first row show the learned histograms $\hat{z}_{1:m}(y)$ for the individual classes: corridor (1), doorway (2), kitchen (3), lab (4), seminar room (5), and office (6). The left image in the second row depicts a possible classification output z_t. In the right image, each bar represents the corresponding likelihood $P(z_t \mid y_t)$ for the different estimates of y_t.

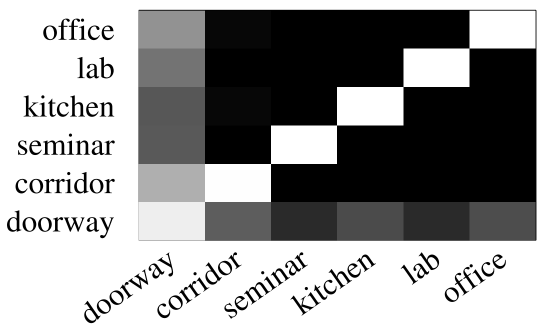


Fig. 5.3 The image depicts probabilities of possible transitions between places in the environment. To increase the visibility, we used a logarithmic scale. Dark values indicate low probability.

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 a room to a doorway is higher than the probability of moving from a room directly to a corridor. This indicates that the robot typically has to cross a doorway first in order to reach a different room. Furthermore, the matrix shows a lower probability of staying in a doorway than staying in the same type of 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.

Table 5.1 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

5.5 Experimental Results

The approach described above has been implemented and tested using simulated and real robot data obtained in the office environment of building 79 at the University of Freiburg. The goal of the experiments is to demonstrate that the presented approach provides a robust classification of places in indoor environments into typical categories. We 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.

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 as shown in Fig. 5.4. Each panoramic view consists of 8 images covering the 360^{o} 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. As example, Fig. 5.5 shows two distributions over the number of coffee machines detected in the database images.

An 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 5.1.

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 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 training classification error also yields good results and at the same time can

Fig. 5.4 The image shows the robot Albert which was used for the experiments. Albert is a B21r robot equipped with a SICK laser range finder and a camera system mounted on a pan/tilt unit.

be computed efficiently. Compared to the optimal order, the classifier generated by this heuristic for an application with six different classes performed on average only 1.3% worse, as demonstrated in [6]. In several situations, the sequence generated by this heuristic turned out to be the optimal one.

5.5.1 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 Fig. 5.6(a). 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 the results.

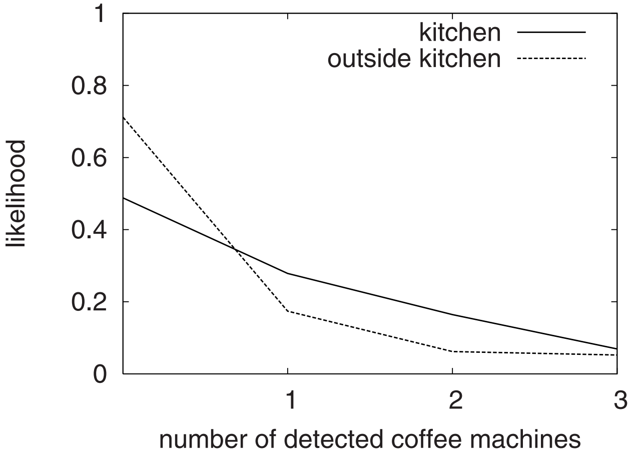
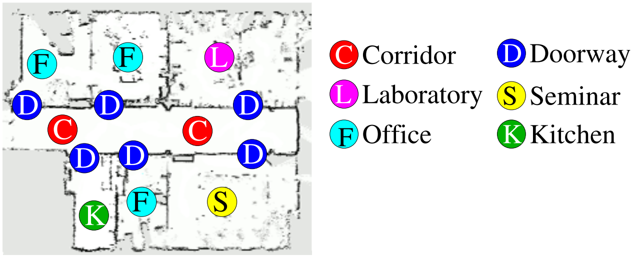


Fig. 5.5 Likelihood of detecting n coffee machines inside and outside a kitchen using Haar-like classifiers.

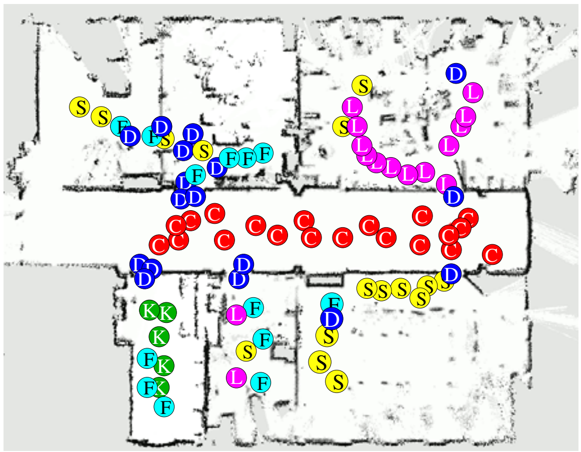
The classification rate of the sequential classifier without applying the HMM is 74.8%. The generated labels are shown in Fig. 5.6(b). If we additionally use the HMM to filter the output of the sequential classifier, the classification rate increases to 83.8%. The labels obtained after applying the HMM are shown in Fig. 5.6(c). As we can see in this example, the model for the HMM encodes the possible transitions and discards the ones with low probability. For example, the wrong office labels that appear in the kitchen (cf. Fig. 5.6(b)) are corrected after the application of the HMM (cf. Fig. 5.6(c)). The reason is that there is a very low probability of going directly from the kitchen to the office according to the model shown in Fig. 5.3. A two-sample t test revealed that the improvement of the resulting classification using an HMM is significant at 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. This improvement is also significant at the $\alpha = 0.01$ level. Both results are shown in Fig. 5.7.

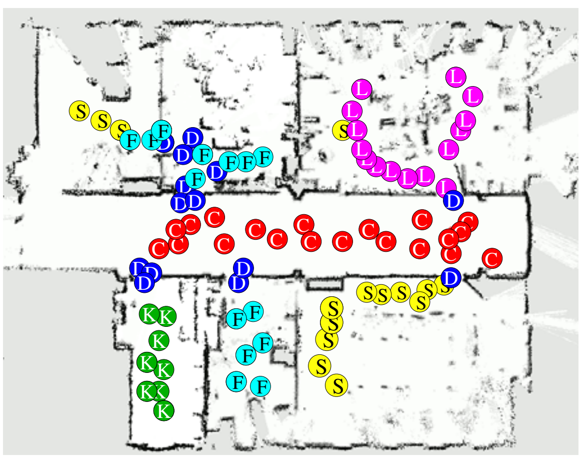
Finally, we studied how the HMM improves the final classification rate according to the output of AdaBoost. For this purpose, we analyzed the improvement of the HMM using different classification rates from AdaBoost. This is achieved by increasing the percentage of weak classifiers used in each binary classifier of the AdaBoost decision list. Here, 100% corresponds to the number of final weak classifiers used in the previous experiment (T in Table 5.1). For example, the classification rate decreases to 60% if only 5% of the final weak classifiers are used. The results are shown in Fig. 5.8. On average, the HMM improves the classification rate by 5.0%.



(a) Ground truth

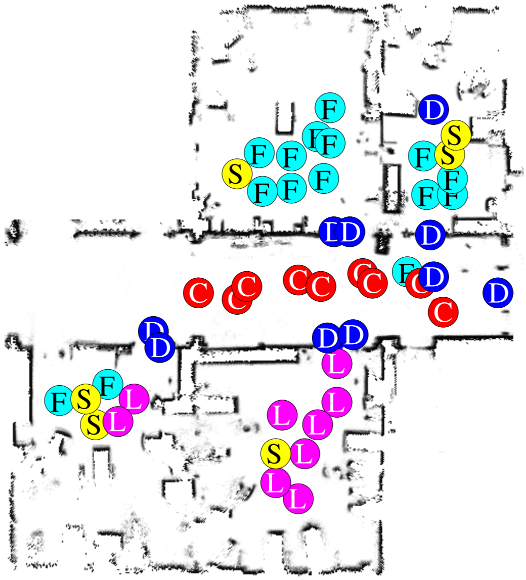


(b) AdaBoost classification

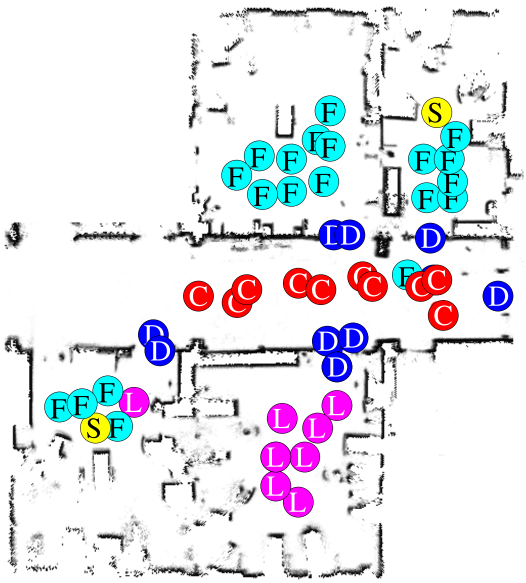


(c) HMM smoothing

Fig. 5.6 (**a**) Ground truth labeling of the individual areas in the environment. (**b**) Classification results obtained for a test set using only the output of the sequential classifier. (**c**) Smoothing of the classification applying the learned HMM.



(a) AdaBoost classification



(b) HMM smoothing

Fig. 5.7 (**a**) Classification without filtering. (**b**) Classification using HMM smoothing.

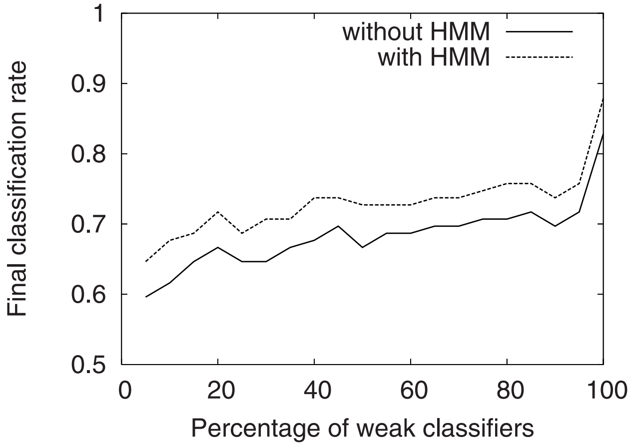


Fig. 5.8 Improvement of the HMM according to the percentage of weak classifiers used in each of the binary AdaBoost classifiers.

5.5.2 Improvement Obtained by Combining Laser and Vision Data

Additionally we analyzed whether the integration of vision and laser data yields any improvements over using only laser. 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 hardly be distinguished from offices using only laser observations. The classification errors obtained by integrating both modalities are summarized in Table 5.2. As can be seen, the combination of laser and vision data yields better results than the classifier relying on laser range data only.

Furthermore, some particular places can be hardly distinguished using only laser. A typical example can be found in the seminar and laboratory rooms. In our office environment, these places have a similar structure, and then similar observations

Table 5.2 Classification error obtained when using only laser data comparing to 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

are obtained using laser scans only. In these cases the addition of visual information improves the separability of the classes. For the seminar room, the error in the classification reduces from 46.9% to 6.3% when using additional vision data. When classifying the laboratory, the error decreases from 34.4% to 3.1%. This reduction in the classification indicates that some rooms are mainly distinguished by the objects found in them, like for example, monitors. These objects cannot be perceived with the laser sensor. A two-sample test indicates that the improvement is significant at the $\alpha = 0.01$ level.

5.6 Related Work

Classifying the places along a trajectory of a mobile robot is a relatively recent area of interest. One of the most known works is the one by Torralba et al. [8], which applies a hidden Markov model to distinguish between the places that a mobile robot traverses. Here, the information about the appearance of images is used to discriminate between different places. In contrast to this approach, the method presented in this chapter uses an additional laser range finder sensor. Moreover, we use the objects detected in the images instead of calculating visual features based on appearance. We classify the places based on their geometrical 2D structure and the objects found in them. In this way, we enable our robot to generalize better when classifying new environments.

Subsequent works analyze the capabilities for distinguishing places along a trajectory using camera images. Pronobis et al. [4] recognize the different places of an office environment using vision. Their approach is based on two kinds of features extracted from the images: interest points descriptors and appearance features. A similar approach is used by Luo et al. [3], but this time applying incremental learning. Also in the work by Spexard et al. [7], rooms are classified according to the appearance of images. In this case the goal of the robot is to recognize rooms previously seen. However, these approaches do not take into account past classifications when calculating the current semantic label.

In a recent study, Pronobis et al. [5] extend their previous work [4] with the additional use of a laser range finder and the set of geometrical features presented in this book. Results show that the laser features improve the generalization of the classifier. However, in [5] no HMM is used to smooth the classification.

For related work about semantic place classification of static poses we refer the reader to Sect. 3.7.

5.7 Conclusion

In this chapter, we have presented an approach to classifying the different poses along a trajectory into semantic classes. The technique uses a combination of simple geometric features extracted from laser range scans as well as features extracted from camera images. It further 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 probabilistic 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 classifying trajectories 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 the approach in which only laser is used.

References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Chichester (1991)
2. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 297–304. Springer, Heidelberg (2003)
3. Luo, J., Pronobis, A., Caputo, B., Jensfelt, P.: Incremental learning for place recognition in dynamic environments. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA (2007)
4. Pronobis, A., Caputo, B., Jensfelt, P., Christensen, H.I.: A discriminative approach to robust visual place recognition. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China (2006)
5. Pronobis, A., Mozos, O.M., Caputo, B.: SVM-based discriminative accumulation scheme for place recognition. In: Proceedings of the IEEE International Conference on Robotics and Automation, Pasadena, California, USA (2008)
6. Rottmann, A.: Bild- und laserbasierte klassifikation von umgebungen mit mobilen robotern. Master's thesis, University of Freiburg, Department of Computer Science (2005) (in German)
7. Spexard, T., Li, S., Wrede, B., Fritsch, J., Sagerer, G., Booij, O., Zivkovic, Z., Terwijn, B., Kröse, B.: BIRON, where are you? - enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)
8. Torralba, A., Murphy, K.P., Freeman, W.T., Rubin, M.A.: Context-based vision system for place and object recognition. In: Proceedings of the International Conference on Computer Vision (2003)