
Activity-based Semantic Mapping of an Urban Environment

Denis F. Wolf and Gaurav S. Sukhatme

Robotic Embedded Systems - University of Southern California
denis|gaurav@robotics.usc.edu

We address the problem of semantic mapping using mobile robots. We focus on the problem of mapping activity as a precursor to automatically classifying, modeling and ultimately understanding the usage of space in a typical urban outdoor environment. We propose and compare two methods for activity mapping - one based on hidden Markov models and the other based on support vector machines. Both approaches estimate high level properties of space based on low level sensor data using supervised learning to associate features to desired classification patterns.

1 Introduction

We are interested in the development of automated techniques for classifying, modeling and ultimately understanding the usage of space in a typical urban outdoor environment. Urban space is used by entities that vary in size and speed (e.g. cars vs. people), locations differ because entities visit them at different frequencies and spend varying amounts of time occupying them (cafeterias vs. classrooms). This overall vision presents many challenges including (but not limited to) segmenting, detecting, registering, and identifying objects in multiple sensor streams; modeling activity; inferring intent, purpose, plans; and finally, developing a common means of representing and visualizing the results.

In this paper we address a vertical slice through the overall problem (i.e. all the way from sensor streams recorded by robots, to a final activity map) with an emphasis on an experimental comparison of two techniques (HMM and SVM) to classifying activity in an outdoor urban setting. Since useful maps are more than simply occupancy measures, we see such **semantic mapping** as a logical extension of the immense amount of research on mobile robot-based map making. This is particularly the case since robot map learning can now cope with large-scale urban environments [13] and dynamic environments [1, 12]. We also note that there is recent related work in semantic

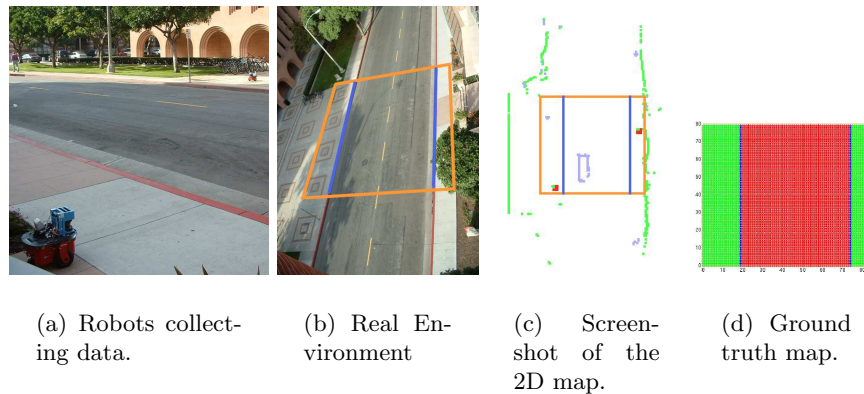


Fig. 1. Environment used for the activity-based semantic mapping and the representation created by the robots. The orange frame corresponds to the mapped area and the blue lines divide the street from the sidewalks.

mapping, specifically object maps [1, 3, 5], place classification [10], activity-based models [6] and the extraction of semantic information from indoor 3D laser maps [7, 8]. Most semantic mapping approaches in the literature look for specific previously known features in the environment to perform classification. We use supervised learning techniques to automatically associate spatial features to desired classification patterns.

2 Semantic Mapping Approach

Our approach to semantic mapping is to combine machine learning techniques with standard mapping algorithms. We present two methods, the first based on hidden Markov models (HMM) and the second on support vector machines (SVM). Our experimental scenario consists of an urban environment; a street and sidewalks on a university campus where people, cars, and bicycles move regularly. Based on range readings obtained from laser range finders on two robots we semantically classify space into either 'street' or 'sidewalk' (Figure 1).

An HMM consists of a discrete time and discrete space Markovian process that contains some hidden (unknown) parameters and emits observable outputs. The challenge is to estimate the hidden parameters based on observable information (see [9] for a tutorial). SVMs are a general class of supervised learning techniques based on statistical learning theory [11]. They are used for classification and regression problems (see [2] for a survey).

We model the environment as a two dimensional grid of cells. At the end of the semantic mapping process, each cell is classified into one of the two

categories S (street) or W (sidewalk). In the HMM framework, each row of cells is considered as a state sequence to be classified while in the SVM approach, each grid cell is classified individually.

We compute four properties of the environment based on range data, and use them as input in both the HMM and SVM methods. These four properties - activity, occupancy, average size of the dynamic entities, and maximum size of the dynamic entities - are defined as follows. Activity is detected every time a certain place in the environment is occupied (by a dynamic entity) and becomes free or vice-versa. Occupancy occurs when a certain location of the environment is occupied by a dynamic entity. The third property of the environment is the average size of the moving entities that occupied the space during the data acquisition process. The fourth property of the environment is the maximum size of the dynamic entities that occupied the space during the data acquisition process. The computation of these four properties from range data is described in [12].

It may happen that parts of the map are not correctly classified due to sensor noise or other reasons. When those errors occur in small parts of the map (considered noise), segmentation techniques can be used to fix them. We used a segmentation method based on Markov random fields (MRF), which has been extensively used in image segmentation. For a complete overview of MRF theory see [4].

3 Experimental Results

We tested the two approaches with experimental data collected on our campus. Two Pioneer robots equipped with SICK laser range finders were positioned on opposite sides of a street to monitor activity in an area approximately 16m x 18m. We used square grid cells of width 20 cm. Each data collection period was approximately 15 mins with a (range) sampling frequency of 10Hz. The mapped areas were manually measured in order to obtain ground truth data to use during the learning steps of the HMM and SVM algorithms, and to evaluate the semantic classification results. Approximately 3% of the total grid cells were used for learning. In order to correct occasional classification errors, a map segmentation technique based on Markov random fields (MRF) was applied to the classified data.

Each of the four properties has been individually tested. The results of the HMM classification can be seen in Table 1, which also includes classification results obtained with standard histogram techniques for a comparison. Figure 2 shows the semantic classification results with and without the use of MRF segmentation algorithm. Parts of the map colored in light green corresponds to the W areas, red colored areas corresponds to the S areas. The two blue lines are the ground truth and the space between them corresponds to the street, while the side spaces correspond to the sidewalks.

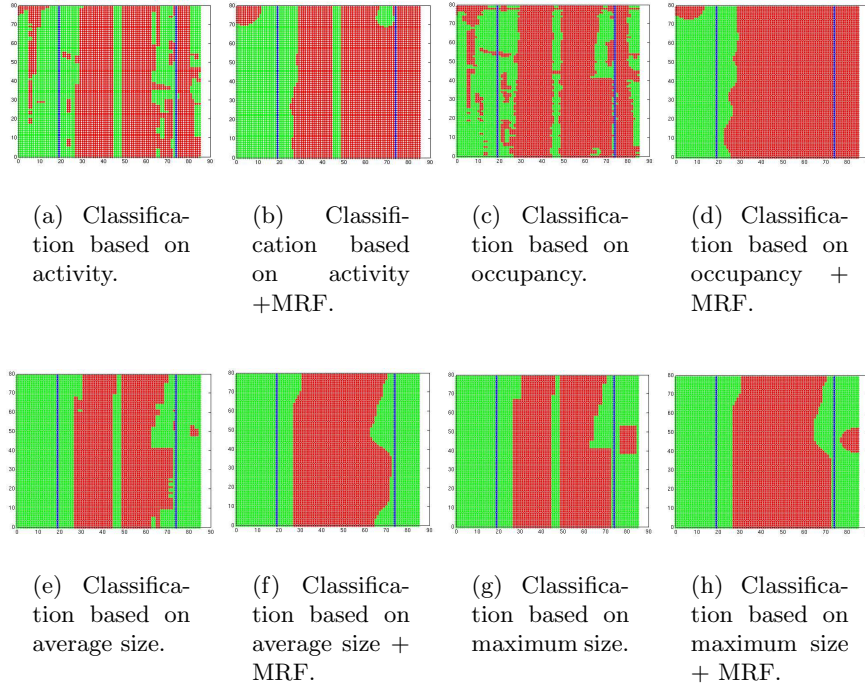


Fig. 2. Activity-based semantic classification based on different properties of the space using HMM.

Property	Hist 10	Hist 50	HMM	HMM + MRF
Activity	53.09%	51.77%	65.00%	69.87%
Occupancy	52.64%	47.32%	69.78%	76.73%
Average size	55.97%	54.11%	78.20%	83.01%
Maximum size	46.93%	31.70%	78.26%	82.72%

Table 1. Results of the HMM activity based semantic classification.

As it can be noticed from the Figure 2, the properties activity and occupancy cannot correctly differentiate the street from the sidewalks. The two wide red lines in the center of the blue lines in the Figures 2 (a) and (c) do correspond to the used parts of the streets, but when the semantic classification algorithm tries to generalize the learned information, it also classifies the most active parts of the sidewalks as S .

The semantic classification based on the properties average size and maximum size shows better results. We can notice in the Figures 2 (e) and (g) that most area colored in red is between the two blue lines, which matches

Kernel	SVM	SVM+MRF
Linear	79.96%	80.12%
Polynomial	78.88%	79.19%
RBF	66.69%	65.12%
Sigmoid	65.12%	65.12%

Table 2. Results of the SVM activity-based semantic classification using the four properties of the space.

the ground truth information. Some parts of the space between the two blue lines are misclassified as W . It happens because this area that is close to the sidewalks or in the center of the street is not used by cars. In the right side of the Figure 2 (g) it is possible to see a red area in the place that corresponds to the sidewalk. The explanation for this misclassification is that during the experiments, in a specific moment, a crowd of people stopped in front of the robot, which was placed in the location. As most of the space around the robot was occupied by moving entities, the range sensors detected a large sized obstacle on that location. When the average size of the moving entities that occupied that space is used for the semantic classification, the effect of the crowd of people is attenuated, as it can be seen in the Figure 2 (e). Figure 2 (f) and (h) shows the classification results after the MRF segmentation.

For the SVM classification, four standard kernels have been used: linear, polynomial, radial basis function(RBF), and sigmoid. Table 2 shows the semantic classification results using the four standard kernels and all the four properties of the space. It is interesting to notice that although the Sigmoid kernel classified every grid cell as S (which is obviously wrong), it still got 65.12% correct classification results. It happens because the ground truth data indicates that a larger part of the map is indeed supposed to be classified as S . In this case, the visual results have to be taken into account when one is evaluating the performance of the classifiers. Figure 3 shows the classification results for the linear, polynomial, and RBF kernels.

The best classification performance for the activity-based mapping were obtained with the linear kernel (Table 2). The reason for the poor classification with the RBF kernel can be explained with an analysis of the data presented in the Table 3, which shows the classification results for the learning and testing datasets using the properties only two properties: activity and average size. The reasons that only these two properties of the space have been chosen for the analysis are that the classification results are very similar to the ones obtained with the four properties, and it allows us to visualize the classification results in a 2D graph (Figure 4).

As it can be noticed in the Table 3, the classification results for the learning dataset using the RBF kernel are better than the ones obtained with the linear kernel. But the same performance is not obtained with the testing dataset. This suggests an overfitting of the learning dataset and results in a poor

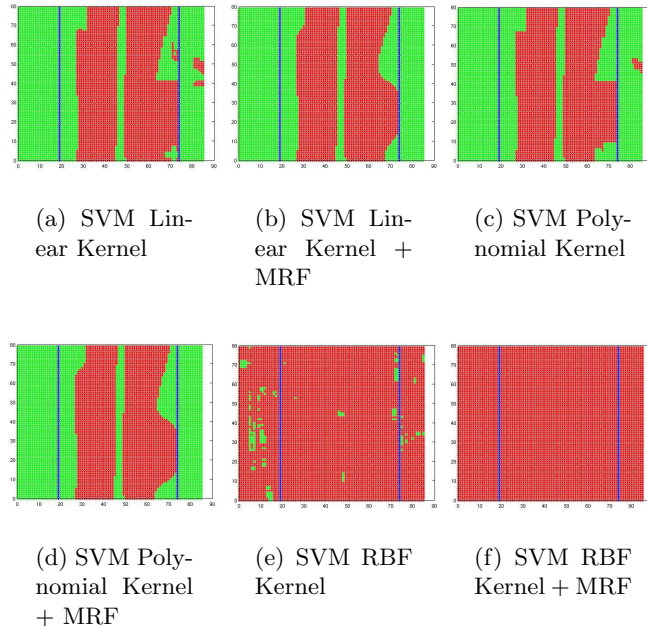


Fig. 3. Results of the SVM semantic classification (W in green and S in red). Different from the RBF kernel, the linear kernel correctly distinguish the street from the sidewalks.

classification to the testing dataset. This fact can be confirmed in the Figure 4, which shows the classification in the property space. For the learning dataset, the RBF kernel (Figure 4(c)) obtain very accurate classification compared to the linear kernel (Figure 4(b)). But when the classification is generalized to the testing dataset, the linear kernel (Figure 4(e)) is much more efficient than the RBF kernel (Figure 4(f)).

Kernel	Learning dataset	Testing dataset
Linear	92.25%	79.78%
Polynomial	92.25%	77.57%
RBF	97.06%	69.27%

Table 3. Results of the SVM semantic classification for the learning and testing datasets using the properties (1) activity and (4) average size.

Besides the experiments with all the four properties of the environment, different combinations of properties have also been tested. The results are

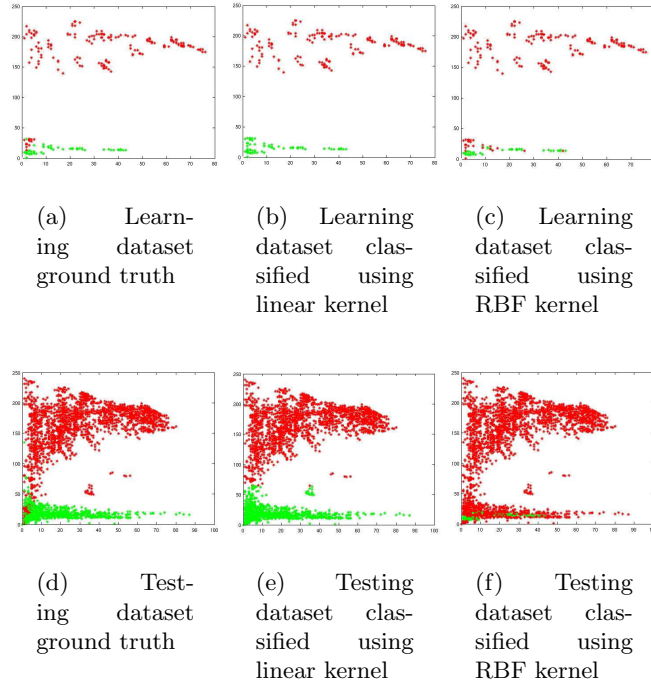


Fig. 4. Results of the SVM semantic classification for the learning and testing datasets using the properties (1) activity and (4) average size.

shown in the Table 4, where the environment properties have been numbered as follows: (1) activity, (2) occupancy, (3) maximum size, and (4) average size. We grouped all the classification results in three categories such that the difference between results in the same category is very minor. Analyzing property combinations that belong to each group it is possible to notice how each property contributes to the classification results. Figure 5 show the classification results for the three categories.

Category *A* presents the most accurate results. Notice that after the MRF segmentation, all the space corresponding to the sidewalk have been correctly classified (Figure 5(d)). There are some classification errors in the space that corresponds to the street, mainly in the region close to the blue line. This can be explained with the fact that almost no activity happens in that region. In most cases, cars (which characterize the streets) occupy the center of the map. In the Figure 5(a) it is even possible to notice the space that divides the two lanes of the street, which is also not used by most of the cars. All the property combinations that belongs to the *A* category (and only these) present the average size (4) as a property. Similar to the results obtained with

Properties	SVM	SVM+MRF	Category
1	65.12%	65.12%	C
2	65.12%	65.12%	C
3	78.53%	78.88%	B
4	79.64%	79.77%	A
1,2	65.12%	65.12%	C
1,3	79.20%	78.88%	B
1,4	79.78%	79.88%	A
2,3	79.40%	79.19%	B
2,4	79.85%	79.90%	A
3,4	79.48%	79.41%	A
1,2,3	79.14%	79.01%	B
1,2,4	79.87%	79.90%	A
1,3,4	79.87%	79.90%	A
2,3,4	79.81%	79.96%	A

Table 4. Results of the SVM semantic classification using combinations of the properties of the space.

the HMM classification approach, this property lead to the most accurate results.

Category *B* presents good classification results with considerable similarities to Category *A*, except for a classification error in a small area in the right side of the map. Category *B* includes all and only property combinations that include maximum size (3), except for those combinations which includes the property 4 (which are classified as *A*). This results are also similar the ones obtained with the HMM approach using maximum size as a property. This can be explained due to the fact that at a specific moment, a crowd of people stopped in front of the robot, which was placed in that location.

Category *C*, which includes only properties activity (1) and occupancy (2), presented a the worst results. They could not correctly distinguish between street and sidewalks and wrongly classified the entire space as street.

The results presented in the Table 4 suggests how each property of the space contributes to the classification. The properties average size and maximum size lead to reasonable results while occupancy and activity do not provide enough information to correctly differentiate the environment into street and sidewalk. Both HMM and SVM methods failed when only these two properties were available.

4 Conclusion

We have shown an experimental comparison of HMM and SVM techniques to classify urban space based on properties inferred from range data. Our approaches are capable of estimating high level properties of space based on

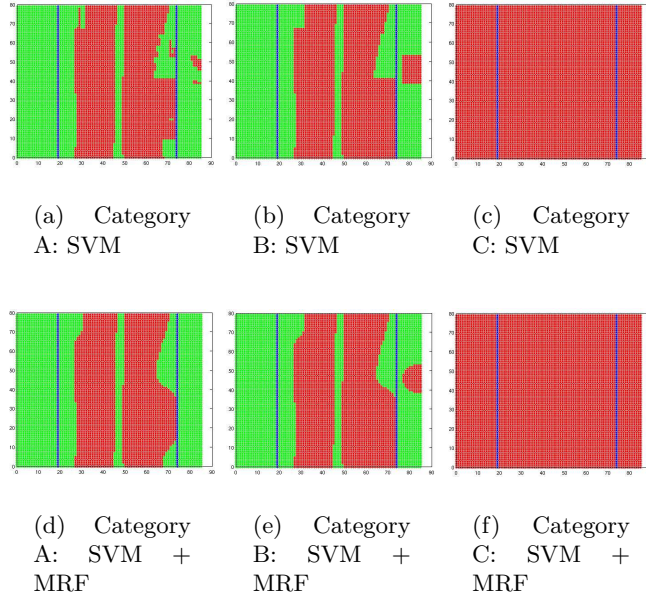


Fig. 5. Results of the SVM semantic classification for the three categories. Results of the categories A and B have correctly distinguished the street and the sidewalks.

low level sensor data. This is done using supervised learning techniques to automatically associate spatial features to the desired classification patterns.

A fundamental difference between the two presented semantic classification methods is that in the HMM approach each data sequence is considered at once, while in the SVM algorithm each point is individually classified. This characteristic does not necessarily lead to better classification results; it all depends on the nature of the data to be classified. In most cases, when the data are divided into well defined clusters, the HMM method tend to be more efficient. The SVM approach is theoretically considered better for non-clustered data, exploring the effect of locality. Another important difference between these two learning methods is the fact that the SVM can handle several input properties while only one can be used in our particular implementation of HMM. For most experiments performed in this paper, the classification results of the two methods are very similar and noticeably better than the standard histogram-based classification algorithm. Another conclusion obtained from the experimental results is that not all properties of space lead to the desired classification. Both HMM and SVM failed when only the activity and occupancy properties were available for classification.

5 acknowledgements

This work is supported in part by NSF grants IIS-0133947, CNS-0509539, CNS-0331481, CCR-0120778, and by the grant 1072/01-3 from CAPES-BRAZIL.

References

1. D. Anguelov, R. Biswas, D. Koller, B. Limketkai, S. Sanner, and S. Thrun. Learning hierarchical object maps of non-stationary environments with mobile robots. In *Annual Conference on Uncertainty in Artificial Intelligence*, pages 10–17, 2002.
2. C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
3. C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *IEEE/RSJ International. Conf. on Intelligent Robots and Systems*, page 34923497, 2005.
4. R. Kindermann and J. L. Snell. *Markov Random Fields and Their Applications*. American Mathematical Society, 1980.
5. B. Limketkai, L. Liao, and D. Fox. Relational object maps for mobile robots. In *International Joint Conference on Artificial Intelligence*, pages 1471–1476, 2005.
6. A. Lookingbill, D. Lieb, D. Stavens, , and S. Thrun. Learning activity-based ground models from a moving helicopter platform. In *IEEE International Conference on Robotics and Automation*, 2005.
7. O. M. Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using adaboost. In *IEEE International Conference on Robotics and Automation*, pages 1742–1747, 2005.
8. A. Nuchter, O. Wulf, K. Lingemann, J. Hertzberg, B. Wagner, and H. Surmann. 3d mapping with semantic knowledge. In *RoboCup International Symposium*, 2005.
9. L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *IEEE*, 77(2):257–286, 1989.
10. A. Rottmann, O. Martinez Mozos, C. Stachniss, and W. Burgard. Place classification of indoor environments with mobile robots using boosting. In *National Conference on Artificial Intelligence (AAAI)*, pages 1306–1311, 2005.
11. V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Nauka, 1979 - English translation: 1982, Springer Verlag.
12. D. F. Wolf and G. S. Sukhatme. Mobile robot simultaneous localization and mapping in dynamic environments. *Autonomous Robots*, 19(1):53–65, 2004.
13. M. Montemerlo and S. Thrun. A Multi-Resolution Pyramid for Outdoor Robot Terrain Perception. *Proceedings of the AAAI National Conference on Artificial Intelligence*, 2004.