Semantic Mapping Using Mobile Robots

Denis F. Wolf and Gaurav S. Sukhatme

***Abstract*— Robotic mapping is the process of automatically constructing an environment representation using mobile robots. We address the problem of semantic mapping, which consists of using mobile robots to create maps that represent not only metric occupancy but also other properties of the environment. Specifically, we develop techniques to build maps that represent activity and navigability of the environment.**

Our approach to semantic mapping is to combine machine learning techniques with standard mapping algorithms. Supervised learning methods are used to automatically associate properties of space to the desired classification patterns. We present two methods, the first based on hidden Markov models and the second on support vector machines. Both approaches have been tested and experimentally validated in two problem domains: terrain mapping and activity-based mapping.

***Index Terms*— Semantic Mapping, Supervised Learning**

I. Introduction

Creating an internal representation (map) of the physical environment is one of the most basic and important capabilities in mobile robotics. Most tasks to be performed by mobile robots requires some type of internal knowledge of the environment. Given the importance of map making in the robotics field, scientists have been actively working on this topic for about two decades, and several mapping techniques have been proposed in the literature over this time [28].

In general, the main focus of the research on mapping has been on representing the geometry of the environment with high accuracy. Although robot-built maps are successfully used for tasks like path planning, navigation, and localization, they are very limited in describing details of the environment other than distinguishing between occupied and empty areas. Virtually any property of space can be represented in a map, but the large majority of the maps built by mobile robots consists of only metric representations of the occupancy of the space. During the mapping task, most mapping techniques neglect a considerable amount of information that describe other aspects of the environment like the navigability, or the nature of the activity that occurs there. The semantic mapping problem consists of using mobile robots to create maps that represent not only metric occupancy but also other properties of the environment.

Semantics consists of assigning meaning to data. Semantic data representation has become a popular research topic that has been explored by scientists in different fields [15] [33] [17] [12] [26]. One of the main purpose to give meaning to information is to make easier the data interpretation. The use of semantics in robotic maps is a method to facilitate the understanding of the data that is used to represent the environment, making easier the sharing of environmental information by robots, people and other machines. It also allows different properties of the space to be represented for richer environmental models. Our approach is based on the use of supervised learning methods to automatically associate properties of space to the desired classification patterns.

In this paper we present semantic mapping approaches for two problems: semantic terrain mapping and semantic activity mapping. The terrain mapping problem consists of creating 3D representations and classifying the terrain according to its navigability. Applications for this approach range from local avoidance of non-navigable areas to path planning. The activity based mapping problem consists of creating two dimensional maps that classifies the environment according to the occupancy of the space by dynamic entities over time. Examples of applications for these maps are path planning and urban traffic modeling.

We propose two semantic mapping approaches, one based on hidden Markov models another based on support vector machines. Both approaches have been applied to both problem domains. A map segmentation algorithm based on Markov random fields has been used to improve the semantic classification. The two semantic mapping approaches have been evaluated based on field experiments.

II. Related Work

Semantic mapping is a very young research topic in mobile robotics. Few papers have been published on this subject and most of them are very recent. The approach presented in [7] suggests an idea to infer high level information from regular robotic maps. The authors claim that the semantic information extracted from the maps can be used for planning and decision making.

In [22] the idea of semantic mapping is presented as part of a human-robot interface. Uniquely identifiable labels are assigned to specific objects of the environment and instructions making use of these labels are given to the robot. In [10] and [27] specific places of indoor environment are labeled based on the presence of key objects in the environment. Computer vision techniques are used to extract the necessary information from images. Similar approach is presented in [16] except by the fact that only range data is used.

The work by [19] present a technique to identify certain objects in the environment based on range information analysis. The bootstrap learning algorithm is used for that. In [1] similar work is presented based on the EM algorithm. In [24] it is presented an approach to extract semantic information from indoor 3D laser maps. This technique is capable of differentiate walls, floor, ceiling, doors, and other key parts

Denis Wolf is with the Computer Science Department at University of Sao Paulo and Gaurav Sukatme is with the Robotic Embedded Systems Laboratory at University of Southern California - USC. Contact author: Denis Wolf - {denis@icmc.usp.br}

of the environment in the map. The mapping approach based on Adaboost, presented in [20] is able to semantically classify different places on indoor environment like room, corridors and doorways based on rage data.

III. Theoretical Background

A. HMM

A hidden Markov model (HMM) consists of a discrete time and discrete space Markovian process that contains some hidden (unknown) parameters and emits observable outputs. The challenge is estimating the hidden parameters based on observable information. This statistical tool is largely used for pattern recognition and is particularly popular for speech recognition. For a complete tutorial see [25].

An HMM can be defined by five elements: a set of possible states (parameters), a set of possible observation symbols, the state transition probability table, the observation symbol probability distribution, and a initial state distribution. For convenience, λ will be used to characterize a HMM model, which consists of the last three elements of the HMM.

Let's assume that O is a observation sequence and each observation symbol in the sequence is denoted by o_n where n means the position of that observation symbol in the sequence ($O = o_1, o_2, ..., o_t$ for a sequence of size t). Let's also assume that Q is a state sequence and each state is denoted by q_n, therefore $Q = q_1, q_2, ..., q_t$. There are three general problems that can be solved with HMMs. The first one consists of given the observation sequence O, and the model λ, how to compute the optimal corresponding state sequence Q (i.e., the state sequence that best explains the observations)? In a nutshell we are trying to maximize the expression $P(Q|O, \lambda)$. The second problem is given the observation sequence O, and the model λ, how to compute $P(O|\lambda)$, the probability of the observations given the model? The third problem consists of how to estimate the model λ to maximize $P(O|\lambda)$, the probability of the observations given the model. In this paper we use the first HMM problem framework to formulate our semantic mapping approach. The solution for this problem can be obtained using the Viterbi Algorithm [8], which is based on dynamic programming techniques.

B. SVM

Support Vector Machines (SVM) is a general class of supervised learning techniques based on statistical learning theory and used for classification and regression problems introduced by [29]. Basically, SVM performs classification by estimating hyperplanes in multidimensional spaces, separating data from different classes. A complete survey is presented in [5]. In order to handle non linear classification problems, the kernel trick proposed by [3] can be used. This idea consists of using a non-linear function to map non linearly separable data to a different Euclidean space where it can be linearly separable.

Four standard kernel types have been used during our experiments: linear, polynomial, radio basis function (RBF), and sigmoid. The package SVM-Light [13] has been used for the SVM learning and classification.

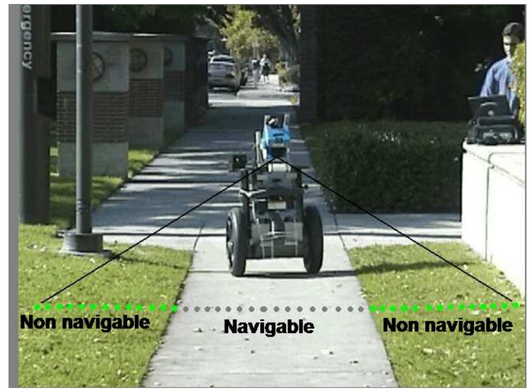


Fig. 1. Robot performing terrain mapping and identifying navigable and non-navigable areas.

IV. Semantic Terrain Mapping

Autonomous navigation is an important capability for a mobile robot. When traversing rough terrain, the robot must have the ability to avoid not only obstacles but also parts of the terrain that are considered not safe for navigation [2]. This is an important problem when one is exploring an unknown terrain. Applications for terrain mapping range from path planning and local obstacle avoidance to detection of changes in the terrain and object recognition [32]. Planetary exploration is an interesting example of practical application for this research topic [21]. In this context, we use a semantic mapping approach based only on information provided by range sensors and the position of the robot builds a 3D map of the terrain and classifies the mapped regions in two semantic categories: navigable and non-navigable.

Our experiments have been performed using ground robots equipped with 2D laser range finders. The range sensors are mounted pitched down on the robots. As the robot moves, the range information generates a 3D point cloud, which models the terrain. We consider flat parts of the terrain such as walkways navigable areas. These areas are considered safe for navigation. Grass and gravel are considered non-navigable (or less desirable) areas. Depending on the application, different types of terrain may be considered navigable and non-navigable. For example, in a planetary exploration context, areas with large rocks that may damage the robot may be considered non-navigable. Classified maps can be very useful for path planning and safe navigation. Even though some times the difference in the roughness of concrete walkways and grass is very small, our approach is capable of semantically classifying them successfully.

A. HMM

The semantic terrain mapping problem can be stated using the HMM framework as follows. The points in our 3D map are the states we are estimating and each scan provided by the range sensor generates a single state sequence Each point can assume one of two possible states: navigable (NA) and non-navigable (NN). The real state of each point is not directly given by the range sensor. Navigable areas in our context will be characterized by flat terrains. On those terrains the points generated by the range sensor are expected to be well aligned, with a minimal variance in altitude. Conversely, the non-navigable areas are characterized by rougher terrains. The 3D points that represent those areas are expected to be not well aligned, with some variance in altitude [31]. Figure 1 shows a robot building a map and identifying navigable and non-navigable parts of terrain at USC campus.

The information provided by the range sensor cannot be used directly as observations in our algorithm, since it is only a measurement of the distance between the sensor and the nearest object in some specific direction. Therefore the data provided by the range sensor are represented as a sequence of points in 3D Cartesian space. Given that sequence of points, the observation for a specific point will be based on the difference in the altitude of that point compared to the altitude of its neighbor points. It is important to notice that instead of having a discrete set of observations for individual states, our approach uses continuous values.

It is possible to use the HMM to learn the model parameters λ (third HMM problem), but in our case, the learning can be done fairly easily through the use of examples. Given a set of range scans, it is possible to manually label the true state for each point on each scan. Given the labeled data, the calculation of the state transition probability distribution NA is straightforward. As we have only two possible states (NA and NN), it consists of counting the number of times that a point labeled NA is followed by a points labeled NA or NN. The same rule applies for points labeled NN. The numbers must be normalized so that the probability distribution sums to 1. Calculating the initial state distribution π is also easy when labeled data are available. The number of states labeled as NA and NN must be counted and the distribution also needs to be normalized.

Based on the labeled data, it is also possible to calculate the variance in the observation for the points classified as NA and NN and use this information as a observation symbol probability distribution. In this case, it is necessary to calculate a mean for the altitude in the points that belong to a specific state. After that it is necessary to calculate the amount of variation in the altitude of those points in comparison to the mean. Figure 2 shows the Gaussian pdfs for the points NA and NN. We can notice that points classified as NA have smaller variance in the relative altitude (observation) compared to the points classified as NN

B. SVM

The terrain mapping problem can also be stated using the SVM framework as follows. Two properties of the space have been used as input of the SVM algorithm: (1) the altitude difference between the 3D point and the robot and (2) the maximum altitude difference of a specific 3D point to its neighbor points (same used in the HMM terrain mapping). As output, the SVM algorithm classifies each point in the 3D terrain map as NA or NN.

As SVM is a supervised learning algorithm, data already classified have to be provided in the learning phase. The same manually labeled points used for the HMM model learning have also been used for the SVM learning. Depending on the kernel choice, some parameters have to be manually set during the learning phase. The cross validation method has been used to tune these parameters.

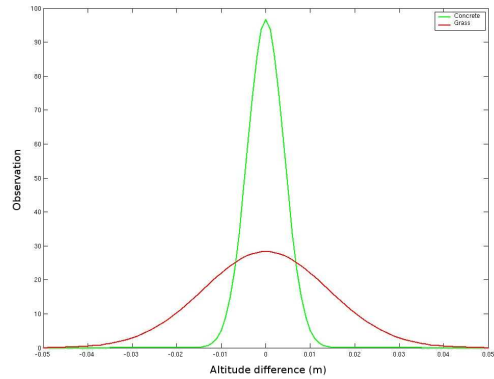


Fig. 2. Gaussian pdfs for the NA and NN points of the walkway terrain. Flat terrain points have smaller variance in the altitude.

V. Activity-Based Semantic Mapping

For the most part, mobile robotics mapping research has been concentrated on static environments. Few mapping algorithms are designed to build maps in presence of dynamic obstacles and, in most cases, the dynamic entities are detected and filtered out. Some mapping approaches, like the one presented in [30], go even further, detecting and representing moving obstacles in the map. There are also robotic algorithms that focus on learning patterns of the activity in the environment [18]. Computer vision researchers have also been addressing the topic of activity modeling for many years [23] [4]. The use of video cameras provide richer information compared to laser range sensors but they also demand considerably more computation to process the data. Usually, cameras have to be placed in high vantage positions to better cover the space to be analyzed. Consequently, they are not suitable to be attached to ground robots to this purpose. Another fact to take into account when monitoring urban spaces is that the lasers preserve the anonymity of the moving entities, which may be a issue for video cameras depending on the situation in which they are used.

We approach the problem of semantic mapping of the environment based on the usage of the space by dynamic entities. In urban environments there are many different types of dynamic entities moving with different properties. For example, some of them are bigger than others; some of them are faster than others. We are particularly interested in obtaining semantic information about environments based on how these moving entities use the space over time.

Our experimental scenario consists of a regular urban environment composed by a street and sidewalks where people, cars, and bicycles move regularly. Based on activity in the environment we try to semantically classify the space in either street (S) or sidewalk (W). The applications for the activity based semantic maps range from path planning to

Fig. 3. Robots mapping the environment according to the activity of the dynamic entities.

traffic modeling. Figure 3 shows two robots collecting activity information in a urban environment.

A. HMM

Our semantic classification for the environment can be stated in the HMM framework as follows. We partition the environment in a two dimensional grid of cells. At the end of the semantic mapping process, each cell is classified in one of the two categories (S or W).

As all the cells are symmetric in the grid, it is possible to organize and address them by columns and rows. Each row of grid cells corresponds to a sequence of hidden states that we are estimating using the HMM framework.

The observations are obtained from the properties of the space sensed by the mobile robots over a determine amount of discrete time. For example, one of the properties of the space is the activity. Every time that a robot detects activity in a portion of the environment, the activity counter on the correspondent grid cell is increased by one. At the end of the data collection period, the amount of activity that occurred in the space represented by each cell is used as observation for the HMM semantic classification.

Due to the fast motion of the dynamic entities and the limited observability of the robots sensors, the robot does not move during the data collection to avoid missing information. The mapped area corresponds to the area covered by the sensor of the robot. In order to decrease the effect of the occlusion and increase the mapped area, two robots have been used during the experiments. The location of the robots is known a priori and it is important to mention that only range information is provided by the sensors.

Four properties have been observed in the environment: activity, occupancy, average size of the dynamic entities, and maximum size of the dynamic entities. 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. For example if a car stops in a determined place, the correspondent grid cells will show high occupancy and low activity. The third property of the environment is the average size of the moving entities that occupied the space during the data acquisition. The fourth property of the environment is the maximum size of the dynamic entities that occupied the space during the data acquisition.

The information about dynamic entities is extracted from the data provided by the range sensors using the algorithm presented in the [30]. The size of the dynamic entities is estimated based on grouping adjacent occupied cells. As in the semantic terrain mapping, HMM model λ is learned based on manually labeled examples.

B. SVM

The activity-based semantic mapping problem can be stated using the SVM framework as follows. The same four properties used in the HMM classification have also been used as input of the SVM algorithm. As output, the SVM algorithm classifies each grid cell of the map into either street (S) or sidewalk (W). An important difference between the HMM and SVM approaches for this problem is that the grid cells with no activity during the experiments have not been considered for classification with the SVM algorithm.

VI. Map Segmentation

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.

Segmentation techniques have been used for many years by the computer vision community [9]. Among several segmentation methods, Markov random fields (MRF) have been extensively used in image processing. For a complete overview of MRF theory see [14].

In the terrain mapping context, it is unlikely that in an area with a large majority of cells NA there are few cells NN. It is also unlikely that among several scans that contain both points labeled NA and NN there is a scan that only contains points labeled NA. The MRF technique make those points agree with their neighbors As a result, NA and NN regions are well defined and clustered.

In order to use MRF as a segmentation tool, we approximate our 3D map to a 2D grid. Each point is projected on the grid based on its x and y coordinates. Each cell in the grid is labeled as NA or NN according to the classification of the points projected on that cell. It is important to notice that not all the cells in the grid have a label, because the distribution of points in the x, y plane is not uniform.

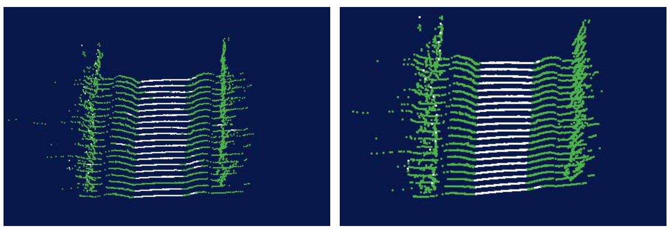
The basic idea of MRF is that the probability distribution for each cell in the grid is specified conditionally on the probability distribution of its neighbor cells. After the application of the filter, all points projected on each cell are labeled with the same label of that grid cell.

Let C_i be a random variable taking the values NA or NN, and denote by $n_i^{(k)}$ $(k = 1, 2, ..., n)$ the number of k neighbors of C_i that are labeled as NA. A simplified MRF model may be specified as:

$$\frac{P(C_i = NA|grid)}{P(C_i = NN|grid)} = exp\Big(\alpha + \sum_{k=1}^{n} (\beta_k n_i^{(k)})\Big)$$

(a) Walkway (b) 3D map of the walkway (c) Doheny Library garden (d) 3D map of the garden

Fig. 4. Real environments and 3D models with the ground truth areas in red color.



(a) HMM semantic classification (b) HMM semantic classification + MRF

Fig. 5. Semantic classification of the Section A of the garden terrain. Some misclassified points are corrected based on the MRF segmentation technique.

where α and β are respectively the prior about the number of NA cells and the importance ratio based on the distance to the cell C_i. As β is increased the chance of each grid cell value agree with the value of its neighbors increases.

The same segmentation idea can be trivially applied to the activity-based mapping problem.

VII. Experimental results

A. Semantic Terrain Mapping

In order to validate our terrain mapping approach, extensive experimental tests have been performed. Our experiments have been done using both an ActiveMedia Pioneer and a Segway RMP robots. Both robots were equipped with SICK laser range finders and a Microstrain IMU. Player [11] has been used to perform the low level control of the robots. On the Pioneer platform, the laser sensor was mounted at 42cm height and a pitch angle of 40^o. On the RMP the laser was mounted at 93cm height and a pitch angle of 35^o, which allowed the robot to map the terrain approximately 1.3m ahead of the robot.

Our experiments have been performed in two different scenarios; a walkway and a garden, both on the USC campus. The walkway scenario is reasonably uniform, with a concrete passage in the center and short grass in both sides. The garden scenario is much more diverse, with grass, bushes, garden seats, and a water fountain. The mapped walkway is approximately 50m long and a complete loop around the mapped of the garden is around 200m.

	Hist 10	Hist 50	HMM	HMM + MRF
Walkway section A	80.0	84.2	97.2%	97.4%
Walkway section B	65.6	77.8	96.2%	98.5%
Garden section A	64.7	72.5	93.1%	97.6%
Garden section B	66.5	67.5	97.8%	98.9%

TABLE I

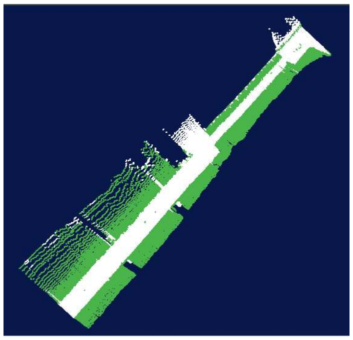
Results of the HMM terrain semantic classification.

In order to verify the accuracy of the HMM classification, we manually labeled 13640 points in the walkway dataset and 12987 points in the garden dataset and used them as groundtruth. For each dataset the labeled points were extracted from two different sections (A and B) of the map to avoid biasing. Approximately 30% of the labeled points has been used in the learning step and the other 70% for the testing. Figure 4 shows pictures of the real environments and the 3D maps before the HMM semantic classification.

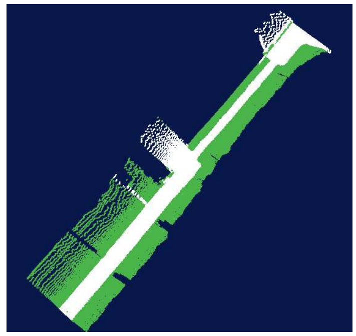
During the experiments, most of the maps have been built when the robots were manually driven with a joystick. But some autonomous navigation experiments using the RMP have also been performed. As our mapping and classification algorithm can be executed in real time, the robot could online use the information about the areas it should avoid and it kept itself in the navigable areas while autonomously mapping the environment. For a better evaluation of the classification results obtained with the HMM method, histogram classification techniques have also been applied to the same data and serve as a reference. Table I show the semantic classification results obtained with histogram with 10 and 50 classes, HMM, and HMM + MRF methods. The histogram classification technique has been experimented with a different number of classes and empirical tests show that the best results were obtained with 50 classes.

Figures 6(a), 6(b), 7(a), and 7(b) show the 3D maps with the semantic classification using HMM on both scenarios. Parts of the terrain colored in white represent the navigable parts of the terrain, while green colored parts corresponds to non-navigable terrain. We can notice that although the grass was very flat in some parts of the terrain, our method could successfully differentiate navigable from non-navigable terrain.

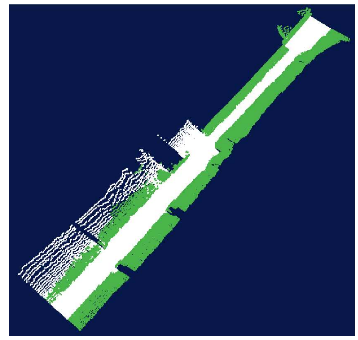
As it can be seen in Figures 6(a) and 7(a), some portions of



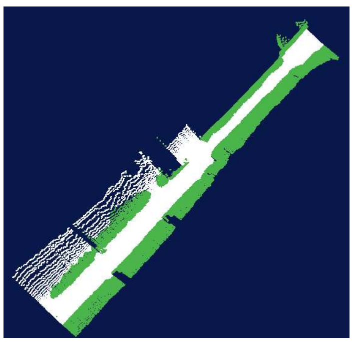
(a) HMM



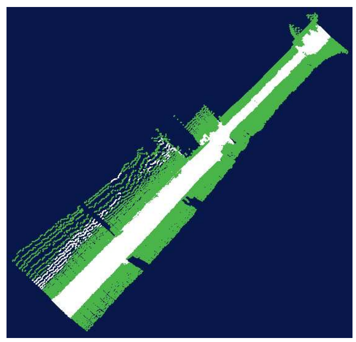
(b) HMM + MRF



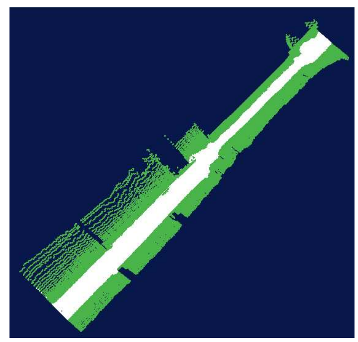
(c) SVM Linear Kernel



(d) SVM Linear Kernel + MRF



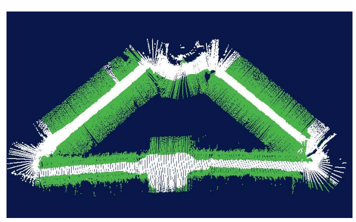
(e) SVM RBF Kernel



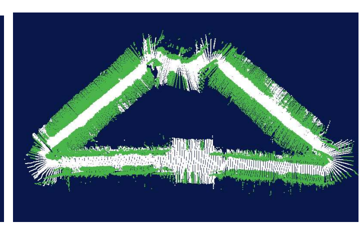
(f) SVM RBF Kernel + MRF

Fig. 6. Semantic classification results for the walkway terrain. The walkway in the center (white color) is correctly identified.

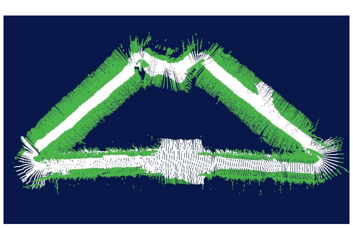
(a) HMM



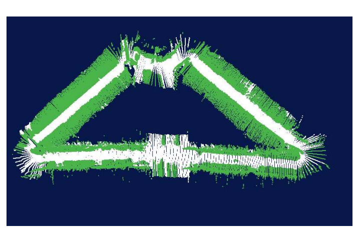
(b) HMM + MRF



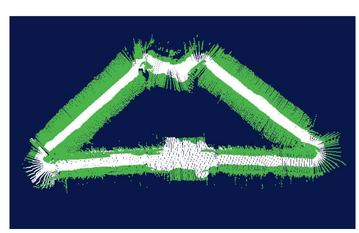
(c) SVM Linear Kernel



(d) SVM Linear Kernel + MRF

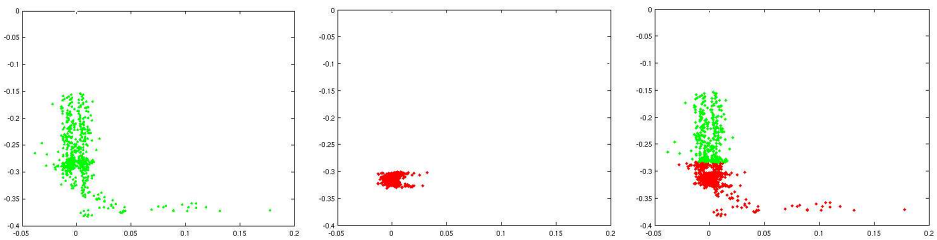


(e) SVM RBF Kernel

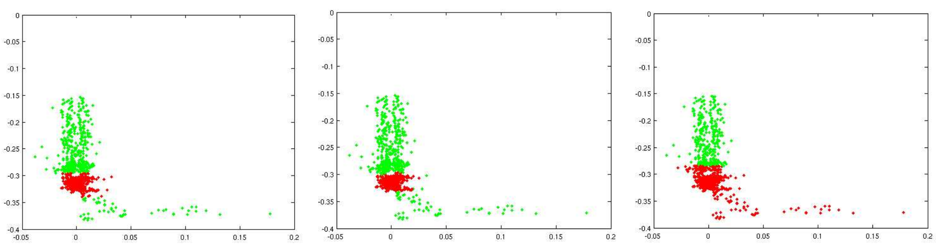


(f) SVM RBF Kernel + MRF

Fig. 7. Semantic classification results for the garden terrain. The walkway in the center (white color) is correctly identified.



(a) Ground truth - NN (b) Ground truth - NA (c) Linear kernel classification



(d) Polynomial kernel classification (e) RBF Kernel classification (f) Sigmoid Kernel classification

Fig. 8. Data points in the property space (NA in green and NN in red).

Kernel	**Walkway section A**		**Walkway section B**	
	SVM	SVM+MRF	SVM	SVM+MRF
Linear	90.91%	90.51%	92.53%	93.18%
Polynomial	97.00%	97.06%	96.82%	97.35%
RBF	97.59%	97.76%	97.82%	98.15%
Sigmoid	90.18%	89.47%	91.85%	92.32%

TABLE II

RESULTS OF THE SVM SEMANTIC CLASSIFICATION FOR THE WALKWAY ENVIRONMENT.

Kernel	**Garden section A**		**Garden section B**	
	SVM	SVM+MRF	SVM	SVM+MRF
Linear	87.75%	87.84%	73.69%	73.78%
Polynomial	92.47%	93.87%	84.14%	85.58%
RBF	94.40%	95.34%	85.48%	91.21%
Sigmoid	47.52%	47.63%	47.20%	43.41%

TABLE III

RESULTS OF THE SVM SEMANTIC CLASSIFICATION FOR THE GARDEN ENVIRONMENT.

the terrain have been misclassified. This happens mostly due to the presence of small obstacles like leaves in the flat parts of the terrain or aligned regions in the rough terrain. Some of those errors were corrected after with MRF segmentation (Figures 6(b) and 7(b)). This improvement in he semantic classification can also be noticed in the results presented in the Table I. Figure 5 shows a closer view of the section A of the garden terrain where the correcting effects of the MRF segmentation can be noticed.

Table II shows the terrain classification results for the Sections A and B of the walkway dataset using the SVM classification approach. As it can be seen, the best semantic classification results have been obtained with the RBF kernel. In almost all cases the MRF segmentation method has contributed with a small improvement in the results. Figures 6 (c) and (e) show the 3D semantic terrain map of the walkway environment using the linear and RBF kernels. It is possible to notice that the MRF segmentation algorithm corrected large part of the misclassified points in the RBF case (Figures 6(d) and (f)). The same did not happen to the linear case due to a larger number of misclassified points.

Table III shows the results of the semantic classification for the garden dataset. The RBF kernel again presents the best classification results. The results obtained in the section A are noticeably better than those obtained in the section B. Differently from the walkway dataset, sections A and B in the garden dataset are very different. The terrain data of the section A contains tall bushes, which are easier to classify than the short grass mapped in the section B. Figure 7 shows the semantic classification results obtained with the linear and RBF kernels, with and without RMF segmentation.

The poor classification obtained by the linear kernel compared to the RBF and polynomial is explained by the fact that the data are not linearly separable. Figure 8 shows the groundtruth and the classification of the walkway data in the property space, where green color represents the NA points and white color the NN points. As it can be noticed in the Figures 8(a) and 8(b) there are some overlapping between the two classes, which prevents to obtain 100% correct classification results.

(a) Real Environment (b) Screenshot of the 2D map. (c) Ground truth map.

Fig. 9. Environment used for the activity based semantic mapping and the space representation created by mobile robots. The orange frame corresponds to the mapped area, the blue lines divide the street from the sidewalks, and the red squares are the position of the robots.

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 IV

RESULTS OF THE HMM ACTIVITY BASED SEMANTIC CLASSIFICATION.

B. *Activity-Based Semantic Mapping*

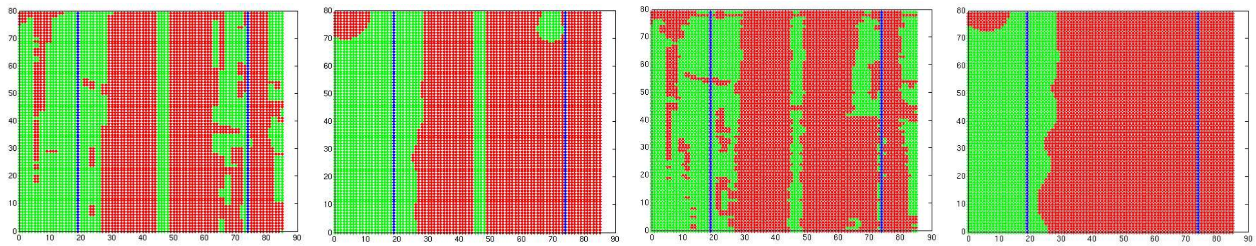
The activity-based semantic mapping approach presented in this section has been tested with experimental data collected on the USC campus. Two Pioneer robots equipped with SICK laser range finders have been positioned in opposite sides of a street to monitor the activity in the region. The area considered for mapping is approximately 16m x 18m with grid cells of 20cm. The ground truth was obtained measuring the width of the street and the sidewalk and each data collection period lasted approximately 15mins with a sampling frequency of 10Hz. One out of 80 rows of grid cells have been manually labeled and used in the learning step of the HMM algorithm. It corresponds to 1.25% of the total grid cells. Figure 9 shows a picture of the environment using for the experiments and a screenshot of the map generated by the robots at a determined instant.

Based on the raw sensor data, the robots are capable to extract the properties of the environment and semantically classify each grid cell in either S or W. Each of the four properties has been individually tested. The results of the HMM classification can be seen in Table IV, which also includes classification results obtained with histogram techniques.

Figure 10 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.

As it can be noticed from the Figure 10, 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 10 (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. After the MRF segmentation, the right sidewalk is entirely classified as part of the street, which is wrong. In fact, based on the data collected during our experiments, it is not possible to determine whether a grid cell belongs to the street or sidewalks just observing its amount of activity or occupancy as it is very similar in both semantic areas.

The semantic classification based on the properties average size and maximum size shows better results, as it can be seen in the Figure 10. We can notice in the Figures 10 (e) and (g) that most area colored in red is between the two blue lines, which matches 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 10 (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

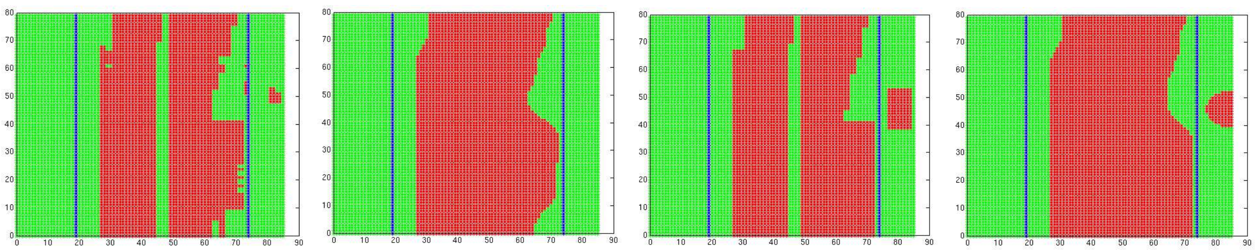


(a) Classification based on activity.

(b) Classification based on activity +MRF.

(c) Classification based on occupancy.

(d) Classification based on occupancy + MRF.



(e) Classification based on average size.

(f) Classification based on average size + MRF.

(g) Classification based on maximum size.

(h) Classification based on maximum size + MRF.

Fig. 10. Activity-based semantic classification based on different properties of the space.

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 V

RESULTS OF THE SVM ACTIVITY-BASED SEMANTIC CLASSIFICATION USING THE FOUR PROPERTIES OF THE SPACE.

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

TABLE VI

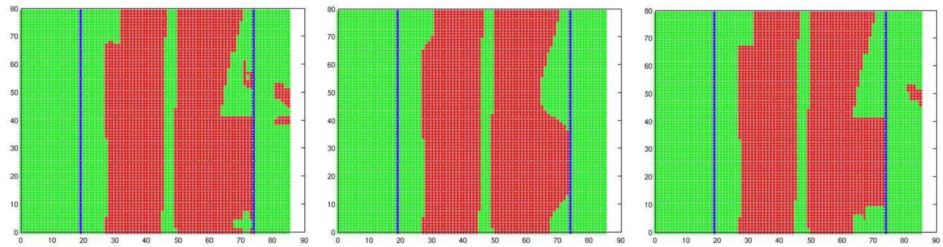
RESULTS OF THE SVM SEMANTIC CLASSIFICATION FOR THE LEARNING AND TESTING DATASETS USING THE PROPERTIES (1) ACTIVITY AND (4) AVERAGE SIZE.

is used for the semantic classification, the effect of the crowd of people is attenuated, as it can be seen in the Figure 10 (e). Figure 10 (f) and (h) shows the classification results after the MRF segmentation.

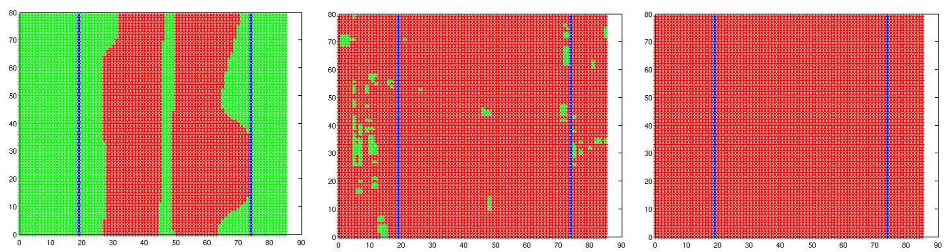
An important difference between the HMM and SVM approaches for this problem is that the grid cells with no activity during the experiments have not been considered for classification with the SVM algorithm. However, they have been classified in the MRF segmentation step. Table V shows the semantic classification results using the four standard kernels and all the four properties of the space. Differently from the terrain classification results, we can notice that the linear kernel obtained the best classification results. Figure 11 shows the classification results for the linear, polynomial, and RBF kernels.

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 to account when one is evaluating the performance of the classifiers.

Differently from the results obtained in the terrain mapping classification, where the RBF kernel obtained the best results, Table V shows that the best classification performance for the activity-based mapping were obtained with the linear kernel. The reason for the poor classification with the RBF kernel can be explained with an analysis of the data presented in the Table VI, which shows the classification results for the learning and complete datasets using the properties only two properties (1) and (4). 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 is possible to visualize the classification results in a 2D graph (Figure 12). As it can be noticed in the Table VI, 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 complete dataset. This suggests an overfitting of the learning dataset and results in a poor classification to the



(a) SVM Linear Kernel (b) SVM Linear Kernel + MRF (c) SVM Polynomial Kernel



(d) SVM Polynomial Kernel + MRF (e) SVM RBF Kernel (f) SVM RBF Kernel + MRF

Fig. 11. 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.

complete dataset. This fact can be confirmed in the Figure 12, which shows the classification in the property space. For the learning dataset, the RBF kernel (Figure 12(e)) obtain very accurate classification compared to the linear kernel (Figure 12(c)). But when the classification is generalized to the complete dataset, the linear kernel (Figure 12(d)) is much more efficient than the RBF kernel (Figure 12(f)). The results obtained with the polynomial kernel were similar to the ones obtained with the linear kernel for the learning dataset, but not as good for the complete dataset.

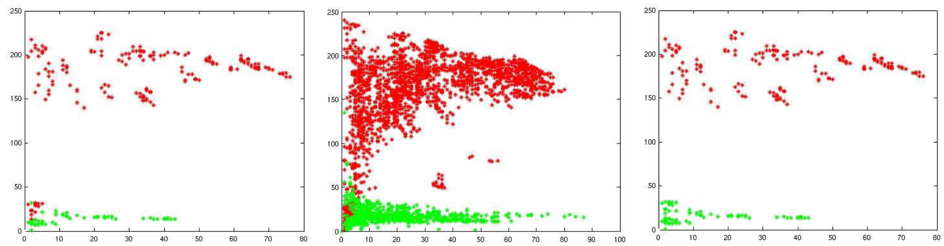
Besides the experiments with all the four properties of the environment, different combinations of properties have also been tested. The results are shown in the Table VII, where the environment properties have been numbered as follows: (1) activity, (2) occupancy, (3) maximum size, and (4) average size. 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 13 show the classification results for the three categories. The results are based only on the linear kernel as it obtained better results than the other kernels for most cases.

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 13(b)). 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 13(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 the HMM classification approach, this property lead to the most accurate results.

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 VII
RESULTS OF THE SVM SEMANTIC CLASSIFICATION USING COMBINATIONS OF THE PROPERTIES OF THE SPACE.

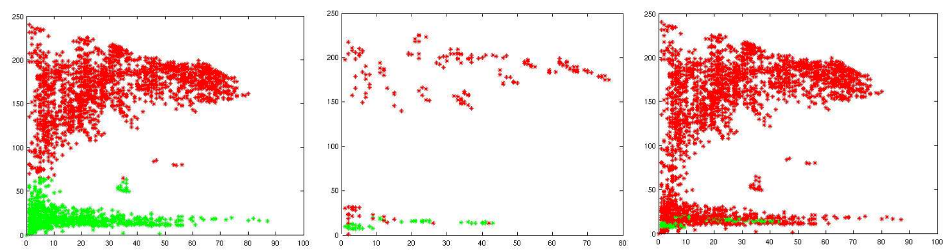
Category B presents good classification results with considerable similarities to Category A, except for a classification error in a small area in the right side of the map. Category B includes all and only property combinations that include maximum size (3) , except for those combinations which



(a) Learning dataset ground truth

(b) Complete dataset ground truth

(c) Learning dataset classified using linear kernel

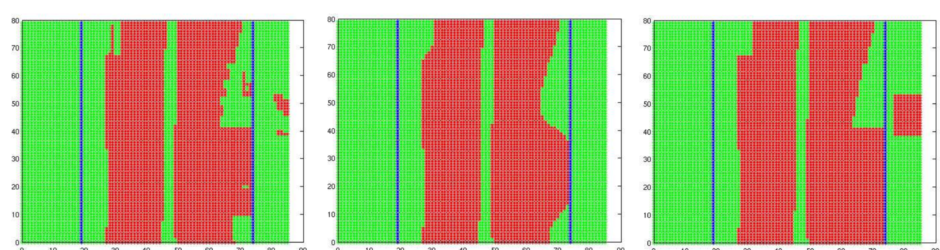


(d) Complete dataset classified using linear kernel

(e) Learning dataset classified using RBF kernel

(f) Complete dataset classified using RBF kernel

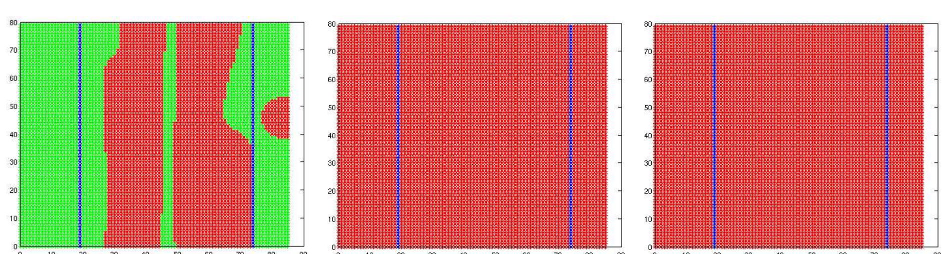
Fig. 12. Results of the SVM semantic classification for the learning and complete datasets using the properties (1) activity and (4) average size.



(a) Category A: SVM

(b) Category A: SVM + MRF

(c) Category B: SVM

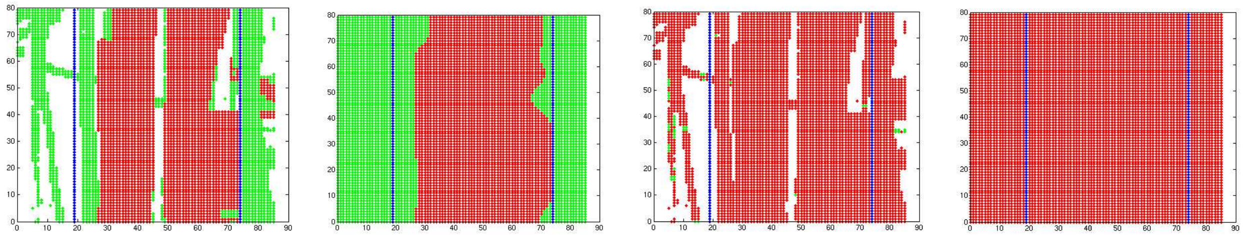


(d) Category B: SVM + MRF

(e) Category C: SVM

(f) Category C: SVM + MRF

Fig. 13. 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.



(a) SVM Linear Kernel

(b) SVM Linear Kernel + MRF

(c) SVM RBF Kernel

(d) SVM RBF Kernel + MRF

Fig. 14. Results of the SVM semantic classification (W in green, S in red, and non used space in white).

Kernel	SVM	SVM+MRF
Linear	86.14%	85.96%
Polynomial	79.93%	80.04%
RBF	76.09%	67.53%
Sigmoid	75.40%	67.47%

TABLE VIII

RESULTS OF THE SVM SEMANTIC CLASSIFICATION EXCLUDING THE NON USED SPACE.

Kernel	Properties	SVM	SVM+MRF
Linear	1,2,3,4	85.37%	84.34%
Linear	1,2,3	85.20%	86.13%
Linear	1,2,4	85.13%	88.93%
Polynomial	1,2,3,4	86.15%	85.62%
Polynomial	1,2,3	87.71%	89.29%
Polynomial	1,2,4	89.17%	95.67%

TABLE IX

RESULTS OF THE SVM MULTI CLASS SEMANTIC CLASSIFICATION.

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. As most of the space around the robot was occupied by moving entities, the range sensors detected a large sized obstacle on that location, which matches with the large obstacles that run in the streets (cars).

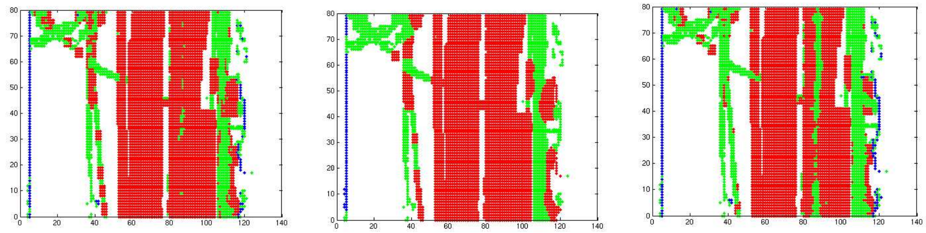
Category C, which includes only properties activity (1) and occupancy (2), presented a the worst results. They could not correctly distinguish between street and sidewalks and wrongly classified the entire space as street.

The results presented in the Table VII 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. The data collected during our experiments shows that the activity and occupancy of the moving entities that occupy the street and the sidewalk is very similar. That is a typical case where there is no association between the input property and desired classification pattern. Both HMM and SVM methods failed when only these to properties were available.

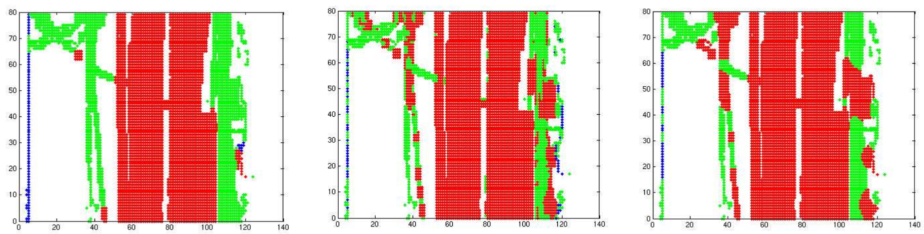
During the experiments there were several of the space which did not register any activity. In the classification results presented above, these areas were also classified as street (S) or sidewalk (W) and as they were present in both S and W areas, they make the learning task harder. If we consider those particular regions as a third class (e.g. non used space), better classification results can be obtained. The classification results are presented in the Table VIII and in the Figure 14, where the MRF segmentation technique is used to estimate the classification of the non used space after the initial SVM classification.

C. Multi-class SVM classification

For all the semantic classification problems presented in this paper, the environment has been divided in two categories: Navigable and non-navigable in the terrain mapping context, and street and sidewalk for the activity based mapping. A natural extension of approaches presented in this paper would be the multi-class semantic classification. In fact the SVM classification algorithm has originally been developed for binary classification, but it has been later extended to cases in which more than two classes [6]. This approach has been tested with the activity-based semantic mapping context. Differently from the experiments previously described where only the information obtained from the dynamic entities has been taken to account, in the multi-class case, the state entities are also considered. The environment has been divided into three classes: street (S), sidewalk (W), and static entities (E). Building structures, trees, and all the other parts of the environment that do not change over time are considered part of E. The main characteristic of the elements in this class are high occupancy and very low activity. The ground truth map can be seen in the Figure 16 where the blue color corresponds the class E. Grid cells with no activities have not been considered for classification. Table IX show the classification results. Only combinations of properties that include the properties 1 and 2 have been presented. All the other property combinations could not classify elements in



(a) Properties 123 (b) Properties 123 with MRF (c) Properties 124



(d) Properties 124 with MRF (e) Properties 1234 (f) Properties 1234 with MRF

Fig. 15. Results of the SVM multi class semantic classification using the polynomial kernel. Most part of the static entities in the environment have been identified

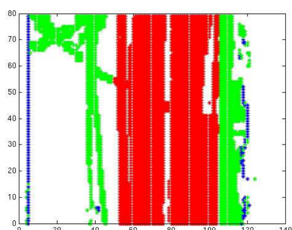


Fig. 16. Multi class classification ground truth (S in red, w in green, and E in blue).

the E class correctly. Only the linear and polynomial kernel obtained reasonable classification results. Figure15 show the classification results with and without MRF segmentation. The best classification results were obtained after the MRF segmentation with the combination of properties 1, 2, and 4. This can be clearly noticed in Figure 15(c).

VIII. Conclusions and Future Work

This paper approached the problem of semantic mapping, which consists of creating robotic maps that go beyond representing the metric structure of the environment. Semantic maps can represent other properties of the environment of the environment, allowing a more complex and complete description of the space. The semantics present in the maps also allow robots to more easily share information and ultimately perform more complex tasks.

Two scenarios have used as a test bed for our semantic mapping approaches: terrain mapping and the activity-based mapping. The terrain mapping problem consists of creating 3D representations and classifying terrain into either navigable or non-navigable areas. The activity-based mapping problem consists of creating two dimensional maps that classify the environment according to the usage of the space by dynamic entities.

Two different approaches for the semantic mapping problem have been presented. The first one is based on hidden Markov models and the second on support vector machines. A fundamental difference between the HMM and SVM 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. In the terrain mapping domain, the data sequences correspond to the 3D points in the range scans. In the activity-based mapping context, the rows in the grid of cells correspond to the data sequences. In order to classify each point in the data sequence, the classification previously made is taken into account. 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, which can be observed in the semantic classification results obtained from field experiments. Another important conclusion obtained from the data analyzed is that not all properties of space lead to the desired classification. In the activity-based mapping problem, both HMM and SVM failed when only the activity and occupancy properties were available.

As semantic mapping is still a very young research topic, there many open problems and interesting directions for new research in the field. Differentiate entities (e.g. people) on a urban environment, understand and possibly predict the behavior of dynamic entities, and improve the complexity of the map representation fusing information provided by other types of sensors (e.g. cameras) are part of the future work to be addressed on this research topic.

IX. Acknowledgments

The authors thank Dieter Fox and Wolfram Burgard for the initial ideas and discussions on terrain mapping using HMM. 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] J. Borenstein and Y. Koren. The vector field histogram - fast obstacle-avoidance for mobile robots. *IEEE Journal of Robotics and Automation*, 7(3):278–288, 1991.
[3] B. E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. *Computational Learing Theory*, pages 144–152, 1992.
[4] N Brand, N Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 994–999, 1997.
[5] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
[6] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2:265–292, 2001.
[7] F. Dellaert and D. Bruemmer. Semantic slam for collaborative cognitive workspaces. In *AAAI Fall Symposium Series*, 2004.
[8] G. D. Forney Jr. The viterbi algorithm. *IEEE*, 61(3):268–278, 1973.
[9] D. A. Forsyth and J. Ponde. *Computer Vision: A Modern Approach*. Prentice Hall, 1st. edition, 2002.
[10] 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. Conference on Intelligent Robots and Systems*, pages 3492–3497, 2005.
[11] B. P. Gerkey, R. T. Vaughan, K. Stoy, A. Howard, G. S. Sukhatme, and M. J. Mataric. Most valuable player: A robot device server for distributed control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1226–1231, 2001.
[12] V. Harmandas, M. Sanderson, and M. D. Dunlop. Image retrieval by hypertext links. In *International Conference on Research and Development in Information Retrieval*, pages 296–303, 1997.
[13] T. Joachims. Svm light - http://svmlight.joachims.org/, Visited in December, 2005.
[14] R. Kindermann and J. L. Snell. *Markov Random Fields and Their Applications*. American Mathematical Society, 1980.
[15] M. D. Klemen, E. R. Weippl, and A. M. Tjoa. The semantic desktop: A semantic personal information management system based on rdf and topic maps. In *Workshop on Ontologies-based techniques for DataBases and Information Systems*, 2005.
[16] B. Limketkai, L. Liao, and D. Fox. Relational object maps for mobile robots. In *International Joint Conference on Artificial Intelligence*, pages 1471–1476, 2005.
[17] Y Liu, N. Lazar, and W. Rothfus. Semantic-based biomedical image indexing and retrieval. In *International Conference on Diagnostic Imaging and Analysis*, pages 18–20, 2002.
[18] 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*, pages 3948–3953, 2005.
[19] J. Modayil and B. Kuipers. Bootstrap learning for object discovery. In *International Conference on Intelligent Robots and Systems*, pages 742–747, 2004.
[20] 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.
[21] NASA. http://ranier.hq.nasa.gov/telerobotics_page/technologies/0240.html, visited in March, 2005.
[22] C. W. Nielsen, B. Ricks, M. A. Goodrich, D. Bruemmer, D. Few, and M. Walton. Snapshots for semantic maps. In *IEEE Conference on Systems, Man, and Cybernetics*, pages 10–13, 2004.
[23] W. Niu, D. H. Han, and Wang Y. F. Human activity detection and recognition for video surveillance. In *IEEE Multimedia and Expo Conference*, pages 719–722, 2004.
[24] A. Nuchter, O. Wulf, K. Lingemann, J. Hertzberg, B. Wagner, and H. Surmann. 3d mapping with semantic knowledge. In *RoboCup International Symposium*, 2005.
[25] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *IEEE*, 77(2):257–286, 1989.
[26] N. Rishe. Efficient organization of semantic databases. In *International Conference on Foundations of Data Organization and Algorithms*, pages 114 – 127, 1989.
[27] 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.
[28] S. Thrun. Robotic mapping: A survey. *Exploring Artificial Intelligence in the New Millenium*, 2002.
[29] V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Nauka, 1979 - English translation: 1982, Springer Verlag.
[30] D. F. Wolf and G. S. Sukhatme. Mobile robot simultaneous localization and mapping in dynamic environments. *Autonomous Robots*, 19(1):53–65, 2004.
[31] D. F. Wolf, G. S. Sukhatme, D. Fox, W. Burgard, Autonomous Terrain Mapping and Classification Using Hidden Markov Models *IEEE International Conference on Robotics and Automation*, pages 2038-2043, 2005.
[32] C. Ye and J. Borenstein. A new terrain mapping method for mobile robots obstacle negotiation. In *UGV Technology Conference at the 2003 SPIE AeroSense Symposium*, pages 52–62, 2003.
[33] H. Zhuge. Retrieve images by understanding semantic links and clustering image fragments. *Journal of Systems and Software archive*, 73(3):455–466, 2004.