

Applying Probabilistic Mixture Models to Semantic Place Classification in Mobile Robotics

Cristiano Premebida, Diego R. Faria, Francisco A. Souza, Urbano Nunes

Abstract—In this paper a study is made of the problem of classifying scenarios, in terms of semantic categories, based on data gathered from sensors mounted on-board mobile robots operating indoors. Once the data are transformed to feature space, supervised classification is performed by a probabilistic approach called Dynamic Bayesian Mixture Models (DBMM). This approach combines class-conditional probabilities from supervised learning models and incorporates *past* inferences. In this work, several experiments on multi-class semantic place classification are reported based on publicly available datasets. Such experiments were conducted in a such way that generalization aspects are emphasized, which is particularly important in real-world applications. Benchmark results show the effectiveness and competitive performance of the DBMM method, in terms of classification rates, using features extracted from 2D range data and from a RGB-D (Kinect) sensor.

I. INTRODUCTION

In order to perceive and understand the surrounding environment, a mobile robot primary depends on data acquired from on-board sensors, being cameras (mono, stereo), laser-scanners (2D, 3D) and RGB-D the most common employed sensors. Among the various tasks that can be performed by a robot, the ability to build a consistent map of the environment and to estimate the robot’s pose is of particular interest [1], [2]. To build a map, a mobile robot moves along a trajectory while collecting data from sensors and uses such sensory information to perceive the environment and construct the map concurrently. Maps can be represented, for example, in terms of metric coordinates or semantic descriptions; the latter is particularly useful in graph-based SLAM. The ability to understand and associate semantic terms such as “corridor” or “office” with places gives a much more intuitive idea of the location of the mobile robot in complement to metric coordinates. The process of applying a semantic term to some place of the map is also known as semantic mapping [3], [4], [5], [6].

Semantic classification of places has been an active research topic in robotics, and also in the machine vision community, wherein the majority of methods use features extracted from images (depth [6], monocular [7], [8], omnidirectional [9], [10]), from laserscanners [4], [11], [12], or a combination of vision and laser-range features as in [5]. Higher-level features can be explored, as in [2] where spatial semantics and objects are used in a Bayesian framework. In terms of pattern recognition techniques, discriminative or

generative solutions have both been used in this research topic namely, Boosting techniques [13], Support Vector Machine (SVM) [9], [11], Naive Bayes classification [7], transfer learning [8], logistic regression [12], Bayesian [2], dynamic time warping and bag-of-words [10], conditional random fields (CRF) [5], and combination of techniques *e.g.*, CRF+SVM as in [4].

In this work, place categorization using 2D laserscanner is particularly emphasized, motivated by three reasons: (1) this is an active sensor modality which is very robust against illumination changes (as shown in the results reported in Section IV); (2) laserscanners are extensively used in robotic applications in academia and industry (guaranteeing safe navigation); (3) most of the range-based features can be directly extrapolated to 3D lasers. Regarding the classification algorithm, a probabilistic approach that follows the principles of a dynamic Bayesian network [14] is used. This approach, with a mixture of probabilistic models in its structure, is named Dynamic Bayesian Mixture Models (DBMM) [15]. This method is tested and compared with a state-of-the-art classifier (a SVM) in terms of classification performance on laserscanner based datasets; nevertheless, experiments using RGB-D are also reported. Concerning experimental evaluation, three benchmark databases (detailed in [16], [17], [18] respectively) with data collected from mobile robots navigating in indoor scenarios were used for experiments.

The contributions of this paper to the problem of place classification in mobile robotics applications are: (i) a general probabilistic approach that combines different class-conditional probabilities, allowing flexibility and improvement in classification performance; (ii) *a-posteriori* outputs are smoothed by means of past probabilities incorporated in the DBMM model with the purpose of mitigating eventual abrupt changes in the likelihood response; (iii) this work reports thorough experiments highlighting generalization aspects, which is particularly important in real-world applications.

The remainder of this paper is organized as follows. A brief review of mixture models and a description of the proposed DBMM method are presented in Section II. Datasets are described in Section III. Experimental results are reported and discussed in Section IV, emphasizing experiments in semantic place categorization using 2D-laserscanners. Finally, Section V concludes this paper.

II. DYNAMIC BAYESIAN MIXTURE MODELS: DBMM

The DBMM is formulated in terms of a joint probability that takes into account the observed feature vector X , the set

The authors are with the Department of Electrical and Computer Engineering, Institute of Systems and Robotics, University of Coimbra, Portugal. {cpremebida, diego, urbano}@isr.uc.pt. The third author is also with the Dept. of Electrical Engineering, Univ. Federal of Ceara, Brazil. {alexandre}@dee.ufc.br.

of classes C , and past inferences D . The concept of mixture models is used and integrated into our dynamic Bayesian process where conditional probabilities from different base classifiers are weighted into a single probabilistic output. The next subsections present the general structure of the DBMM and how the weights are computed in order to assign confidence to the base classifiers.

A. Mixture Models

A mixture model is here understood as a weighted combination of component probabilities, assumed independently distributed, that were modeled according to base classifiers (BCs). Considering the set of classes C and the class-conditional probabilistic outputs from n base classifiers $P_i(X|C)_{i=1,\dots,n}$, the general mixture model outputs a weighted probability $P(X|C)$ as follows:

$$P(X|C) = \sum_{i=1}^n w_i P_i(X|C), \quad (1)$$

where n is the number of base classifiers and w_i is the weight associated to a given probabilistic output $P_i(X|C)$ obtained by a supervised classifier. The weights, that sum to one $\sum_{i=1}^n w_i = 1$, were estimated by an Entropy-based measure as confidence level, as explained in Section II-C.

B. The DBMM structure

The DBMM structure is composed of a mixture model that outputs $P(X|C)^t$ (1), the *a-priori* class probabilities $P(C)^t$, and past inferences, as illustrated in Fig. 1. The past inferences enters into the DBMM in the form of a time-based element D^T that specifies a set of prior states given a time slice T i.e., $D^T = \{D^{t-1}, D^{t-2}, \dots, D^{t-T}\}$. In the DBMM the general joint probability is expressed as:

$$P(D^T, X, C) = \prod_{k=t-1}^{t-T} P(D^k | \bigwedge_{j=k-1}^{t-T} D^j, X, C) \times P(X|C) \times P(C), \quad (2)$$

where the chain-decomposition of D^T will be considered to be conditionally independent given (X, C) e.g., $P(D^{t-1} | D^{t-2}, X, C) = P(D^{t-1} | D^{t-2})$. The joint probability (2) can be rewritten as $P(C, X, D^T)$ and, after decomposition and arrangement, the posterior $P(C|X, D^T)$ can be expressed as $P(C|X, D^T) = \beta P(D^T, X, C)$, where β is a normalization term to ensure a valid probability. For a specific time instant t , the DBMM model assumes the form:

$$P(C|X, D^T)^t = \beta \prod_{k=t-1}^{t-T} P(D^k | D^{k-1}) \times \sum_{i=1}^n w_i P_i(X^t | C) \times P(C)^t \quad (3)$$

where, for instance, the first element inside the multiplication operator takes the value of the previous posterior i.e., $P(D^{t-1} | D^{t-2}) \leftarrow P(C|X, D^T)^{t-1}$, and so on.

C. Assigning Weights using Entropy

In the DBMM framework we use Entropy H , from information theory, as a confidence level to estimate the weights w that will be used to combine the mixture of classifiers.

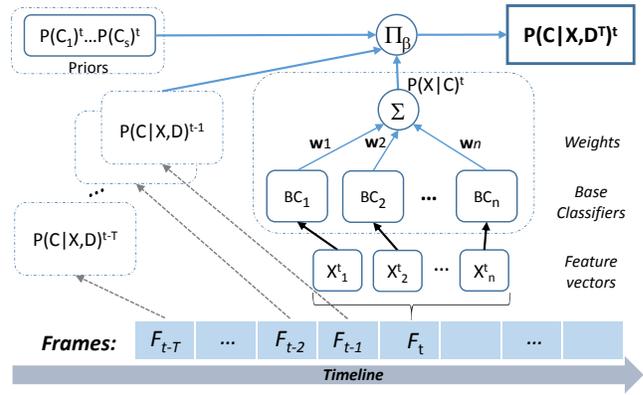


Fig. 1. Illustrative representation of the DBMM approach, where the posterior $P(C|X, D^T)^t$ depends on the priors $P(C)$, the combined likelihood $P(X|C)$, the time-conditional probabilities (past inferences), and the normalization β .

Considering a training set comprising the likelihoods delivered by the set of base classifiers, Entropy is computed as follows:

$$H(P_i(A)) = - \sum_i P_i(A) \log(P_i(A)), \quad (4)$$

where, in our case, $P_i(A) = P_i(C|X)$ represents the conditional probability given the model of a i^{th} classifier and the set of features $X = \{x_1, x_2, \dots, x_m\}$. From the learning stage (i.e., using a training set), the set of likelihoods assumes the form of a probability density function (pdf). Knowing H , the weight w_i for each i^{th} classifier is estimated as the distribution inversely proportional to Entropy as follows:

$$w_i = \frac{1 - \left(\frac{H_i}{\sum_{i=1}^n H_i} \right)}{\sum_i \left(1 - \left(\frac{H_i}{\sum_{i=1}^n H_i} \right) \right)}, \quad i = \{1, \dots, n\}, \quad (5)$$

where H_i is the value of Entropy resultant from (4). The denominator in (5) guarantees that $\sum_i w_i = 1$. This weighting strategy will smooth the base classifier's response by continuously multiplying its classification belief by the correspondent weight.

III. BENCHMARK DATASETS

In this Section, the semantic place labeling datasets used in the experiments are briefly described (these datasets are publicly available on the Web). The first dataset was collected from three indoor environments of buildings 52, 79 and 101 at the University of Freiburg (named FR52, FR79 and FR101 respectively), and are composed by the classes "corridors", "rooms", "doorways", while FR101 has an additional class "hall". The second dataset considered here is one of the three sub-datasets that composes the COLDB database [17], namely we used the Saarb-dataset because of two reasons: it has a greater number of classes than others COLDB's sub-sets, and it provides laserscanner data with adequate resolution while Ljubljana-dataset has laser data but with low resolution. The RGBD-based dataset, detailed in [18], is composed of a set of vision-based features extracted from a Kinect sensor and is characterized by five classes. Table I shows a summary

TABLE I SUMMARY OF FRs, COLD-SAARB AND RGB-D DATASETS.

Dataset	Classes	#Fea.	N# of examples
FR52	corridor, rooms doors	50	Train: (7396, 43957, 1112) Test: (6093, 32214, 743)
FR79	corridor, rooms doors	50	Train: (16687, 43624, 1449) Test: (15516, 50180, 701)
FR101	corridor, rooms doors, hall	50	Train: (858, 9005, 262, 18095) Test: (931, 6795, 243, 26688)
COLD Saarb	9 classes*	50*	N# of examples depends on the experiments*
RGBD	room, corr., kit., labs, office	397	(252, 279, 191, 357, 149) N# depends on a random 10-fold

(*) See Section III-B for details.

of these datasets, where the numbers of examples of each class is given in parentheses.

A. Datasets (FR52, FR79 and FR101) and features

In the FRs datasets, and also in the COLD-Saarbrücken database (Section III-B), the mobile robot is equipped with a 2D laserscanner covering a field of view around the robot. A given sensor measurement, $z = \{b_0, \dots, b_m\}$, contains a set of laser-beams b_i , where $0 \leq i \leq m$. Each beam b_i consists of a pair $b_i = (\alpha_i, d_i)$, where α_i is the angle of the beam relative to the robot and d_i is the length of the beam. Mozos *et al.* [13] have proposed the use of a set of geometrical features extracted from range scans (the so called *B* and *P*-features). These features are invariant to robot rotation, making them dependent on the robot position and not on its orientation. Thus, for each incoming laser-scan, the *B*-features are calculated using the raw beams in z , while the *P*-features are calculated from a polygonal approximation $P(z)$ of the area covered by z . The vertexes of the closed polygon $P(z)$ correspond to the Cartesian coordinates of the end-points of the laser-beams in z : $P(z) = \{(d_i \cos \alpha_i, d_i \sin \alpha_i) \mid i = 0, \dots, m + 1\}$.

In this work, we actually employed a subset of the *B* & *P*-features. The components of our feature vector are summarized in Table II showing the correspondence with the *B* & *P*-features proposed in [13]. We noticed that the feature B.3 has constant value, and hence it was not considered, while feature P.6 - which corresponds to coefficients 2 to 100 of Fourier transform - tends to induce overfitting problems and thus it was discarded.

B. COLD Saarbrücken dataset

The COLD-Saarb sequences were acquired under different weather and illumination conditions (designated by Cloudy, Night, Sunny), and across a time span of two/three days [9]. The Saarb-set has 9 classes: “Corridor”, “Terminal room”, “Robotic lab”, “1-person office”, “2-persons office”, “Conference room”, “Printer area”, “Kitchen”, and “Bath room”. Two paths were followed by the robot during data acquisition, the Standard and the Extended paths; moreover, sequences of the dataset were annotated as portions A and B: the main difference is that those parts annotated as “A” do not have sequences under “Sunny” condition (see [9] for

TABLE II FEATURE COMPONENTS AND THEIR CORRESPONDENCE WITH *B* & *P*-FEATURES AS ORIGINALLY PRESENTED IN [13]

Features	Description
$x_{1,2,3,4}$	correspond to features B.1, B.2, B.5, B.6; correspond to feature B.7 with thresholds
x_5, \dots, x_{24}	0.5m to 10.0m (0.5m steps);
$x_{25,26,27,28}$	correspond to features B.11, B.12, B.13, B.14; correspond to feature B.15 with thresholds
x_{29}, \dots, x_{38}	0.1m to 1.0m (0.1m steps);
x_{39}	corresponds to feature B.16;
$x_{40,41,42}$	correspond to features P.1, P.2, P.3;
$x_{43,44}$	correspond to features P.4, P.5;
$x_{45,46,47}$	correspond to features P.7, P.8, P.9;
$x_{48,49}$	correspond to features P.16, P.17;
x_{50}	corresponds to feature P.15;

more details). This dataset provides, among mono and Omni-image frames, raw laser scans with FOV=180° and 0.5° of resolution *i.e.*, each laser scan has 361 points. In order to allow a consistent analysis, we opted to extract the same *B* & *P* features (see Table II) to compose a classification set with 50 features.

C. Kyushu University Kinect place recognition dataset

This dataset contains RGB-D images acquired by a Kinect sensor mounted onboard a mobile platform. Five classes are considered: “corridor”, “kitchen”, “labs”, “office” and “room”. This dataset was recorded from a mobile robot driving through different buildings of the Kyushu University [18]. To perform semantic place categorization, histograms of local binary pattern (LBP) were extracted from grayscale and depth images in order to compose a feature vector as described in [18]. Moreover, the dimensionality of each histogram was further reduced by selecting a subset of their LPB’s using a uniformity measurement introduced in [19]. A threshold (θ) is used to control such uniformity measurement, where lower values of θ produce histograms with reduced dimensionality.

IV. EXPERIMENTS AND PERFORMANCE EVALUATION

In this Section, the DBMM method is evaluated in terms of classification performance on the semantic place labeling datasets described in Section III. In the FRs datasets, classification performance is assessed by applying $F_{measure} = 2 \frac{Pr \cdot Re}{Pr + Re}$, calculated on the testing part of the datasets, where *Pr* and *Re* denote precision and recall respectively. Unlike in [13], [16] where accuracy is used as performance measure, here we primary adopted $F_{measure}$ ¹ because all datasets have unbalanced classes. On the other hand, and to follow the same methodology described in [18], average accuracy is used to assess performance in the RGB-D dataset. In the COLD-Saarb bar charts are used to report classification rates, enabling a direct comparison with the results in [9].

The experiments are firstly conducted using a SVM (using libSVM²) and the DBMM within the FR’s scenarios (intra-

¹In this paper the values of $F_{measure}$ are presented in percentage.

²<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

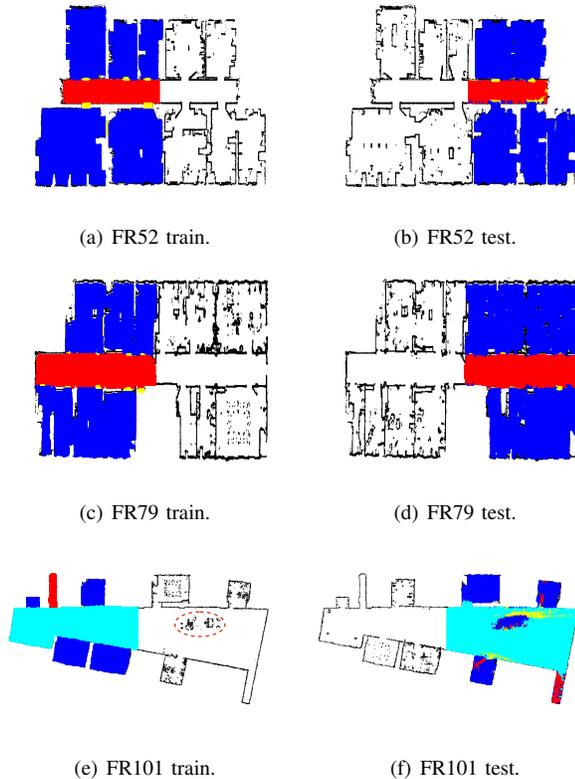


Fig. 2. Maps exhibiting the DBMM classification results for the FR52, FR79 and FR101 buildings; — Corridors (in red); — Rooms (in blue); — Doorways (in yellow); — Hall (in cyan). See text for details.

buildings), and their results are analyzed and discussed in the Section IV-A. Seeking to verify the performance on “unseen” environments, in Section IV-B we report results in terms of generalization capabilities (cross-building). Secondly, and following the methodology in [9], Section IV-C brings two series of experiments on the Saarb-dataset: (1) where differences in illumination conditions and categorical aspects are considered, and (2) different paths (Standard and Extended) are interchanged between training and testing sets, which allows analysis in terms of generalization capacity. The experiments are concluded with Section IV-D, where the results using the RGBD dataset are presented.

A. Intra-building semantic classification

In a first experiment, a SVM was trained having as input normalized values of the features from Table II, and the results regarding the FR52, FR79 and FR101 testing sets are summarized in Table III. The results are reported by feature type, *i.e.*, features of type *B* (left), *P* (center) and in the last column all features are used; $\bar{F}_{measure}$ denotes the average value of the $F_{measure}$ for all classes. From the results shown in Table III, one can conclude that the subset of P-type features has a good discriminative capacity (see values in bold for FR79 and 101), keeping low complexity: only 11 features. On the other hand, in the FR52 scenario the combination of B&P features achieved better results. Concerning the class “Doors”, the results are significantly lower in comparison

TABLE III EVALUATION OF A SVM (TESTING SET)

Perf.	fea.B (#n = 39)			fea.P (#n = 11)			B & P (#n = 50)		
	FR52	FR79	FR101	FR52	FR79	FR101	FR52	FR79	FR101
$\bar{F}_{measure}$	71.20	68.17	56.00	65.29	69.07	61.47	72.08	68.50	59.37
$Re_{(corr)}$	94.86	99.12	35.77	96.44	98.56	69.92	97.67	99.35	53.28
$Re_{(room)}$	99.86	94.46	93.35	99.78	100.0	92.88	99.90	94.61	92.38
$Re_{(door)}$	13.06	23.82	7.00	0.0	4.99	14.40	12.11	25.25	11.52
$Re_{(hall)}$	-	-	93.73	-	-	89.82	-	-	93.81

TABLE IV SVM AND DBMM PERFORMANCE ON THE FR DATASETS

Performance	SVM			DBMM		
	FR52	FR79	FR101	FR52	FR79	FR101
$\bar{F}_{measure}$	72.08	69.07	61.47	70.92	75.75	63.75
$Re_{(corr)}$	97.67	98.56	69.92	96.04	98.85	71.64
$Re_{(room)}$	99.90	100.0	92.88	99.96	100.0	94.61
$Re_{(door)}$	12.11	4.99	14.40	10.63	17.26	15.23
$Re_{(hall)}$	-	-	93.81	-	-	91.17

with other classes; the most likely explanation is because “Doors” is much less frequent than others classes (see Table I) and thus the classifier tends to favor the majority classes.

The DBMM is used in a second experiment, where the mixture model inside the DBMM contains two base classifiers ($n = 2$ in Fig. 1) as class-conditional probabilities *i.e.*, BC_1 is a SVM using the 39 *B*-features ($X_1 = fea.B$) and BC_2 is another SVM using the 11 *P*-features ($X_2 = fea.P$). This strategy keeps the complexity of the DBMM consistent with the SVM used in the first experiment. Classification results achieved by the proposed DBMM are shown in Table IV, as well as the best results (in terms of $\bar{F}_{measure}$) obtained by the SVM are also presented to make comparison easier. The results indicate the effectiveness of the DBMM, obtaining best generalization in two scenarios (FR79 and 101) while keeping the complexity *i.e.*, the amount of features, equivalent to the benchmarked SVM.

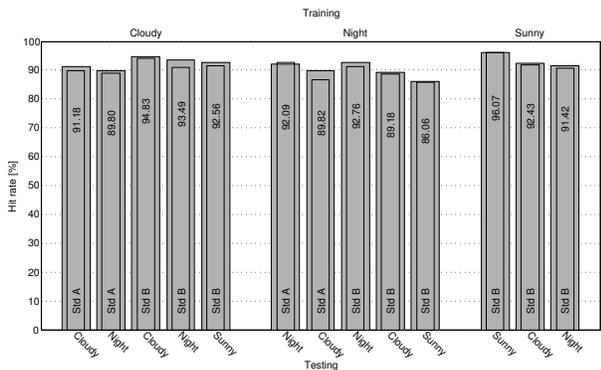
Figure 2 exhibits the maps of the buildings, where each place is indicated by its respective color. The training sets are shown in the left part of Fig. 2, while the semantic outputs using the DBMM algorithm are illustrated in the right part. Notice that the classifier had difficulties in classifying “Doorways” (in yellow): the lower number of examples and the appearance on narrow FOV are explanations for these results. Moreover, in the highlighted region at the testing part of FR101 the performance suffered a degradation, meaning that the number of false positives at that region increased (*i.e.*, *Pr* decreased).

B. Cross-building semantic classification

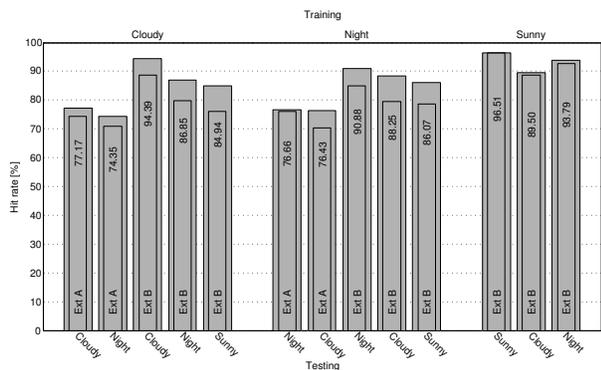
Looking to verify the performance in unseen environment *i.e.*, the generalization capability of the methods in ‘new’ scenarios, in this Section the classification methods are trained in one dataset and tested in other set (for that purpose, buildings 52 and 79 were considered). The models were trained using the dataset collect in the FR52 building and then tested on the dataset acquired at FR79 building, and *vice-versa*. Results are summarized in Table V. With respect

TABLE V EVALUATION OF SVM AND DBMM ON CROSS-BUILDING

Perf.	Trained in FR52 tested in FR79		Trained in FR79 tested in FR52	
	SVM	DBMM	SVM	DBMM
$\bar{F}_{measure}$	74.86	75.61	65.47	67.64
$Re_{(corr.)}$	94.17	93.73	92.84	97.52
$Re_{(room)}$	99.92	99.98	86.30	88.70
$Re_{(door)}$	27.58	31.77	45.82	44.64



(a) Standard (Std) sequence



(b) Extended (Ext) sequence

Fig. 3. Results are presented per illumination condition in terms of overall hit rate calculated over all classes. Each bar shows the sequence (Std,Ext) and the portion (A,B) used in the testing data *e.g.*, StdA denotes Standard sequence of the portion A.

to the first case, the DBMM method obtained the best value of $\bar{F}_{measure}$ due to a better balance between precision and recall, in particular for class “Doors”. In the second case (right part of Table V), the best $\bar{F}_{measure}$ result was again obtained by the DBMM but, this time, the class “Corridor” contributed more. Overall, results exhibit a balanced performance between the cases.

C. Experiments on COLD-Saarb

This Section starts with experiments on the Standard and Extended parts, and the portions “A” and “B”, in accordance with the experiments carried out in [9]. The main goal of these experiments is to evaluate place recognition under varying illumination and categories (classes) settings. Essentially, the training and testing sets differ as function of the weather and illumination conditions (Cloudy, Night, Sunny). Here, as in the experiments using the FRs dataset, the mixture

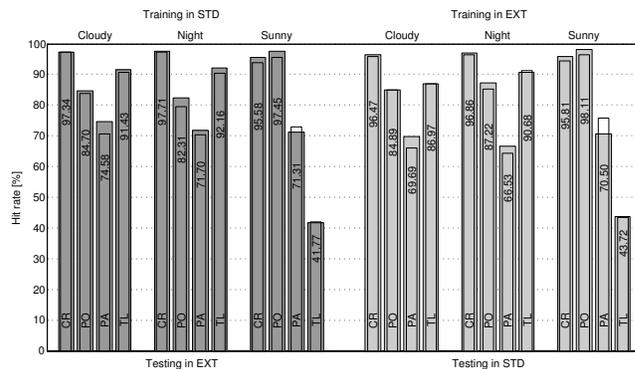


Fig. 4. Hit rate performance, per class, considering non-correlated scenarios: trained on the Standard-set and tested on the Extended-set (bar charts in the left), and *vice-versa* (right part). Four categories were tested: corridor (CR), person-office (PO), printer-area (PA), and toilet-room (TL).

model used in the DBMM structure was built with two SVM classifiers, learned with B and P features respectively.

The results shown in Fig. 3 were grouped according to the sequence type (Standard or Extended), the portions of the scenario (‘A’ or ‘B’), and the weather/illumination conditions used for training and testing. As in [9], bottom axes indicate the conditions used for testing, while top axes refer the conditions on training. These results allow us to conclude that DBMM achieved an average classification performance of 91.67% and 85.82%, which surpass SVM (narrow bars depicted within the gray-bars), and also the solution in [9]. Furthermore, our solution using laser-based features demonstrates to be robust under variations of illumination.

Figure 4 shows recall, or hit rate, for the four classes that are all available in the two paths followed by the robot (sequences Std and Ext). This experiment explores the situation where the classification method is trained and tested in sequences whose conditions are substantially different and therefore, it allows us to study cross-dataset generalization. In general, the performance on classes ‘CR’ and ‘PO’ are very good, while the accuracy on classes ‘PA’ and ‘TL’ degrades in some conditions; an explanation is that the training and testing set are very unbalanced. This is particularly evident, for instance, in the “STD-Sunny” subset where the training set has (821,364,325,828) elements while the testing set “EXT” (which includes all sub-sets) has (9726,2346,1841,3292) elements. In general, DBMM achieved better results than SVM. Further discussions are provided in Section IV-E.

D. RGB-D place recognition

In the experiments on laserscanner-based datasets, only SVMs were used as base classifier, however in the experiments on the RGB-D dataset an Artificial Neural Network (ANN) classifier is also employed. As consequence, the DBMM was trained with two heterogeneous base classifiers, one SVM (BC_1) and one ANN (BC_2), having both classifiers the same number of features $\#Fea=397$ and the same feature vector *i.e.*, $X_1 = X_2$. To evaluate the classification performance in this dataset, this paper follows the same

TABLE VI EVALUATION ON THE RGB-D DATASET.

	ANN	SVM	DBMM
$\theta = 2$	89% \pm 0.086	89% \pm 0.085	91% \pm 0.079
$\theta = 4$	92% \pm 0.074	91% \pm 0.076	94% \pm 0.061
$\theta = 8$	92% \pm 0.090	91% \pm 0.078	93% \pm 0.075

procedure described in [18]: each testing set was created by randomly selecting one place from each category, while the remaining places are used for training purpose. This methodology assures that testing sets contain examples from places that do not appear in the training part.

Experiments using three subsets of LBP's features were performed based on values of thresholds θ applied as function of an uniformity measurement [18]. For $\theta = \{2, 4, 8\}$, the aforementioned procedure is repeated 10 times and the average accuracy is reported, as well as the standard deviation. Table VI shows the results using the 10-fold random-validation strategy and averaging the performance across the folds for the models trained and tested separately. According the results from Table VI, the proposed DBMM ranks as the best approach.

E. Discussion

Experiments on the FR datasets are very detailed, and the goals were to study the potentialities of solutions using 2D-laser based features and to conduct systematic analysis of the DBMM's performance in comparison with SVM. The latter was chosen because SVM is one of the most common and effective method used in this field. Regarding the first experiments on the COLD-Saarb dataset, which contains more classes than FRs and RGB-D datasets, the DBMM achieved a higher classification rate relative to the image-based solution in [9]. The second experimental part, summarized in Fig. 4, constitutes a new experiment conducted on this dataset. Finally, the experiments on the RGB-D dataset had the purpose of exploring the DBMM with a more complex feature-set and using heterogeneous base classifiers in its mixture model. Classification performance was compared with a SVM and an ANN (table VI), and the implemented DBMM obtained higher accuracy and lower deviations.

V. CONCLUSION

In this study we carried out many experiments on semantic place recognition, in the context of mobile robotics applications, to assess the state of the art in this topic. Particular focus was devoted to publicly available datasets using 2D-laserscanner. In addition to the experimental part and analysis, this paper contributes with a probabilistic framework (called DBMM), that uses mixture models and temporal-inference in its structure. From the several experiments carried out on semantic place recognition, the DBMM demonstrated to be a very promising approach. In summary, the results reported in this work show that the DBMM is a very competitive method, reaching better classification

results than benchmarked solutions, with interesting characteristics: (i) DBMM supports general probabilistic class-conditional models; (ii) dynamic information in the form of temporal/prior inferences can be incorporated; (iii) diversity on the base-classifiers are suitable to be controlled in a direct manner. DBMM allows fast implementation and direct probabilistic interpretation.

ACKNOWLEDGMENT

This work has been supported by the FCT project "AMSHMI2012-RECI/EEIAUT/0181/2012" and project "ProjB: Diagnosis and Assisted Mobility - Centro-07-ST24-FEDER-002028" with FEDER funding, programs QREN and COMPETE. We also thank the reviewers for their valuable comments and suggestions.

REFERENCES

- [1] M. Milford, "Vision-based place recognition: how low can you go?" *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 766–789, 2013.
- [2] S. Vasudevan and R. Siegwart, "Bayesian space conceptualization and place classification for semantic maps in mobile robotics," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 522–537, jun 2008.
- [3] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *ICRA, IEEE*, 2012.
- [4] L. Shi, S. Kodagoda, and G. Dissanayake, "Application of semi-supervised learning with voronoi graph for place classification," in *IROS, IEEE/RSJ*, 2012.
- [5] L. Shi, S. Kodagoda, and M. Piccardi, "Towards simultaneous place classification and object detection based on conditional random field with multiple cues," in *IROS, IEEE/RSJ*, 2013.
- [6] H. Jung, O. Martinez Mozos, Y. Iwashita, and R. Kurazume, "Indoor place categorization using co-occurrences of LBPs in gray and depth images from RGB-D sensors," in *Emerging Security Technologies (EST), 2014 Fifth International Conference on*, Sept 2014.
- [7] J. Wu, H. Christensen, and J. Rehg, "Visual place categorization: Problem, dataset, and algorithm," in *IROS, IEEE/RSJ*, 2009.
- [8] G. Costante, T. Ciarfuglia, P. Valigi, and E. Ricci, "A transfer learning approach for multi-cue semantic place recognition," in *IROS, IEEE/RSJ*, 2013.
- [9] M. Ullah, A. Pronobis, B. Caputo, J. Luo, R. Jensfelt, and H. Christensen, "Towards robust place recognition for robot localization," in *ICRA, IEEE*, 2008.
- [10] L. Yuan, K. C. Chan, and C. George Lee, "Robust semantic place recognition with vocabulary tree and landmark detection," in *IROS's Workshop, IEEE*, 2011.
- [11] P. Sousa, R. Araujo, and U. Nunes, "Real-time labeling of places using support vector machines," in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*, June 2007.
- [12] L. Shi, S. Kodagoda, and G. Dissanayake, "Laser range data based semantic labeling of place," in *IROS, IEEE/RSJ*, 2010.
- [13] O. M. Mozos, *Semantic Place Labeling with Mobile Robots*. Springer, Tracts in Advanced Robotics (STAR), 2010.
- [14] V. Mihajlovic and M. Petkovic, "Dynamic Bayesian Networks: A State of the Art," *Technical reports series, TR-CTIT-34*, 2001.
- [15] D. Faria, C. Premebida, and U. Nunes, "A probabilistic approach for human everyday activities recognition using body motion from RGB-D images," in *RO-MAN, IEEE*, 2014.
- [16] O. M. Mozos, A. Rottmann, R. Triebel, P. Jensfelt, and W. Burgard, "Supervised semantic labeling of places using information extracted from sensor data," *Robotics and Autonomous Systems*, vol. 55, pp. 391–402, 2007.
- [17] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *The International Journal of Robotics Research (IJRR)*, vol. 29, no. 2-3, pp. 298–320, 2010.
- [18] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 5, pp. 6695–6711, 2012.
- [19] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, no. 7, pp. 971–987, 2002.