

# Dynamic Bayesian Network for semantic place classification in mobile robotics

Cristiano Premebida · Diego R. Faria · Urbano Nunes

Received: Nov,2015 / Accepted: July,2016.

**Abstract** In this paper, the problem of semantic place categorization in mobile robotics is addressed by considering a time-based probabilistic approach called Dynamic Bayesian Mixture Model (DBMM), which is an improved variation of the Dynamic Bayesian Network (DBN). More specifically, multi-class semantic classification is performed by a DBMM composed of a mixture of heterogeneous base classifiers, using geometrical features computed from 2D laser-scanner data, where the sensor is mounted on-board a moving robot operating indoors. Besides its capability to combine different probabilistic classifiers, the DBMM approach also incorporates time-based (*dynamic*) inferences in the form of previous class-conditional probabilities and priors. Extensive experiments were carried out on publicly available benchmark datasets, highlighting the influence of the number of time-slices and the effect of additive smoothing on the classification performance of the proposed approach. Reported results, under different scenarios and conditions, show the effectiveness and competitive performance of the DBMM.

**Keywords** Semantic place recognition · Dynamic Bayesian Network · Artificial intelligence

## 1 Introduction

The capability of perceiving and understanding complex, dynamic and unstructured environments is essential for intelligent robots to be introduced in our daily life. However,

a mobile robot primary depends on the sensory information from on-board sensors, such as cameras (mono, stereo), laserscanners (2D, 3D) and RGB-D data. Despite the possibility of different sensory perception, there is still a world of sensory uncertainty to deal with. The ability to build a consistent map of the environment and to estimate the pose of the robot is one of the various tasks that can be performed by a robot, as in Milford (2013) and Posner et al (2009). In order to build a map, the sensory information plays an important role to perceive the environment to construct the map concurrently, allowing a mobile robot to move along the trajectory while the data arrives from the sensors. Nevertheless, in the case a map is provided, approaches based on semantic localisation can be explored as in (McManus et al 2015).

Most of the maps in mobile robotics are represented as a combination of metrical and topological data structures (Werner et al 2012). For path and task planning, the representation of maps has to be simplified and adapted to the scenario where the robot has to deal with. Maps based on semantic descriptions are useful, for instance, in graph-based SLAM (Hong et al 2015). The capacity of reasoning on sensor data to associate semantics to a specific place of an indoor environment, such as “corridor” or “office”, provides more intuitive idea of the mobile robot location in complement to metric values. The process of semantic place recognition, or categorization (Jung et al 2016), incorporated in a map building process is known as semantic mapping (Shi et al 2012; Pronobis and Jensfelt 2012; Jung et al 2014; Shi et al 2013).

Robotics and machine vision communities have been involved for many years in the problem of semantic classification of places, as summarized in the recent survey of Kostavelis and Gasteratos (2015), and many solutions were proposed using different sensors and techniques. Moreover, existing datasets like in Martinez-Gomez et al (2015) and Pronobis et al (2010), have been an important contribution

---

C. Premebida and U. Nunes are with the Department of Electrical and Computer Engineering at the University of Coimbra, and with the Institute of Systems and Robotics. D. Faria is with the System Analytics Research Institute, Aston University, Birmingham, UK.  
Address: University of Coimbra, Polo II, Coimbra, Portugal.  
Tel.: +351-239-796214  
E-mail: {cpremebida,urbano}@isr.uc.pt; d.faria@aston.ac.uk

to the progress in this field. Regarding camera sensors, the work of Jung et al (2014) explores information from a depth-camera, while a monocular camera is employed in (Wu et al 2009; Costante et al 2013), and an omnidirectional one in (Ullah et al 2008; Yuan et al 2011). Laserscanners are used by Shi et al (2012, 2010), and information from both sensors is explored in (Rogers and Christensen 2012; Shi et al 2013). On the other hand, when it comes to pattern recognition level, both discriminative and generative solutions have been largely used in this research area, namely Boosting techniques (Mozos 2010), support vector machine (SVM) (Ullah et al 2008), Bayesian classifier (Vasudevan and Siegwart 2008), Naive Bayes classification (Wu et al 2009), logistic regression (Shi et al 2010), transfer learning (Costante et al 2013), dynamic time warping and bag-of-words (Yuan et al 2011), conditional random fields (CRF) (Rogers and Christensen 2012; Shi et al 2013), and combination of techniques e.g., CRF+SVM as in (Shi et al 2012).

In this work, semantic place categorization is addressed with focus on a probabilistic approach for classification using 2D laserscanner data. The fact of emphasizing laserscanner data is due to three reasons: (1) this is an active sensor modality which is very robust against illumination changes, as shown in the results reported in (Premebida et al 2015); (2) laserscanners are broadly 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. The classification algorithm addressed here follows the principles of a dynamic Bayesian network (Mihajlovic and Petkovic 2001) but, since its structure incorporates a mixture of probabilistic models, it is named Dynamic Bayesian Mixture Models (DBMM) (Faria et al 2014).

This paper, which is an extension of (Premebida et al 2015), brings contributions to the problem of place classification in mobile robotics applications as follows: (i) a general expression for the DBMM in the form of a finite product of *past* (time-based) class-conditional probabilities and priors, allowing a direct interpretation of the number of time-slices in the DBMM structure; (ii) *a-posteriori* outputs are smoothed by means of ‘additive smoothing’ incorporated in the DBMM model with the purpose of mitigating eventual close-to-zero class-conditional probabilities; (iii) this work reports thorough experiments highlighting generalization capacities in different scenarios and conditions, which is particularly important in real-world applications. DBMM is extensively evaluated in terms of classification performance on two benchmark databases (detailed in (Ullah et al 2008) and (Pronobis et al 2010) respectively) with laserscanner data collected from mobile robots navigating in indoor scenarios.

The remainder of this paper is organized as follows. A brief review of DBN, the mixture models, and the weighting strategy are given in Sect. 2. A description of the proposed

DBMM method is presented in Sect. 3. Datasets using 2D-laserscanners are described in Sect. 4. Experimental results are reported and discussed in Section 5, emphasizing the effect of the number of time-slices and nodes on the DBMM performance, as well as the additive smoothing on the prior distributions. Finally, Sect. 6 concludes this paper.

## 2 Preliminaries

This section starts with a brief review of the DBN, followed by the basic formulations w.r.t. semantic place recognition problem. The concept of finite mixture models is then described in the sequel, and this section concludes with the weighting approach used to combine a finite set of base classifiers into the mixture model. The developments described in this section will serve as a basis for the DBMM formulation in Section 3.

### 2.1 Brief review of DBN

A Dynamic Bayesian Network (DBN)<sup>1</sup> is a generalization of a Bayesian Network (BN) where temporal relationship between state-variables in a BN is explicitly modeled. DBNs follow the same principles of BNs, where the nodes represent a set of random variables and the arcs (or links) represent the direct and acyclic dependencies between the nodes. Denoting by  $X = \{X_1, \dots, X_m\}$  the set of  $m$  random variables, represented by the nodes in a BN, a DBN with  $T$  time-slices expresses the dynamic behavior by the time step variable  $t$ :

- Previous time-slices:  $\{\mathbf{X}^{t-1}, \mathbf{X}^{t-2}, \dots, \mathbf{X}^{t-T}\}$
- Current time step:  $\mathbf{X}^t = \{X_1^t, X_2^t, \dots, X_n^t\}$
- Subsequent time-slices:  $\{\mathbf{X}^{t+1}, \mathbf{X}^{t+2}, \dots, \mathbf{X}^{t+T}\}$

The temporal relationships between the nodes, called inter-slice or temporal arcs, can include the same variable over time e.g.,  $X_i^t \rightarrow X_i^{t+1}$ , and different variables over time e.g.,  $X_i^t \rightarrow X_j^{t+1}$  (Korb and Nicholson 2010). Usually DBNs are built in a such way that a node at one time-slice affects only the node ahead i.e.,  $X_i^{t-1} \rightarrow X_i^t$ , however a network with multiple time-connected arcs can be built. In the case where the arcs connect only the current time nodes with the previous nodes, such condition is said to follow the first-order Markov assumption (Korb and Nicholson 2010).

Conditional probabilities are used to model the dependencies between the nodes, both intra (same time step) and inter-slice relationships (previous and/or subsequent slices). For example, given a node  $X_i^t$  with intra-slice parents  $C_1^t$  and  $C_2^t$  and inter-slice parents  $X_i^{t-1}$  and  $C_1^{t-1}$ , this probabilistic relationship is expressed as  $P(X_i^t | C_1^t, C_2^t, X_i^{t-1}, C_1^{t-1})$ .

<sup>1</sup> Also known, according to (Korb and Nicholson 2010), as dynamic belief network, probabilistic temporal network, or dynamic causal probabilistic network.

## 2.2 Formulation of the problem

Considering a set of nodes  $C$ , the evidence about such nodes  $\mathbf{X}$ , and for the current time step  $t$ , inference is posed as a Bayesian problem of the form  $P(C^t|\mathbf{X}^t) \propto P(\mathbf{X}^t|C^t)P(C^t)$ . In semantic place recognition, we formulate the problem in terms of  $P(C^t, X^t)$ , where  $C^t$  is the set of classes of interest and  $X^t$  denotes evidence obtained from sensor input signals. Here,  $X^t$  enters into the network in the form of a feature vector, calculated from the raw-input measurement, conditioned to the parameters  $\Theta$  of a given classification/learning model. The joint probability, with some rigor, is  $P(C^t, X^t, \Theta)$ , but, since  $\Theta$  represents non time-varying parameters of a given supervised classifier trained in advance,  $\Theta$  is omitted for sake of conciseness.

A simple dynamic network with two nodes  $X$  and  $C$ , and considering just the current time step, is modeled by the joint probability  $P(X^t, C^t)$ . Nevertheless, to solve the problem of semantic place classification for the instant  $t$ , the probability of interest - given by  $P(C^t|X^t)$  - can be easily obtained by the Baye's formula  $P(C^t|X^t) = P(X^t|C^t)P(C^t)/P(X^t)$ , where the class-conditional probability  $P(X^t|C^t)$  comes from a probabilistic-based classifier and the *a-priori*  $P(C^t)$  can be estimated recursively as detailed in Sect. 3.1.

## 2.3 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  $\sum_i w_i = 1$ , were estimated by an Entropy-based measure as confidence level, as explained in the sequel.

## 2.4 Assigning Weights using Entropy

There are numerous techniques one can use in the estimation of a finite set of weights to combine classifiers. Here, we use Entropy  $H$ , from information theory, as a confidence level to estimate the weights  $w$  that will be used to compose the mixture of classifiers. Considering a training set com-

prising the normalized likelihoods delivered by the set of base classifiers, Entropy is computed as follows:

$$H_i(P_i(\cdot)) = - \sum_{j=1}^m P_j(\cdot) \log(P_j(\cdot)), \quad (2)$$

where, in our case,  $P_i(\cdot) = P_i(C|\Theta, X)$  represents the class-conditional probability given the model  $\Theta$  of a  $i^{th}$  classifier and the set of features  $X$ ; simply denoted by  $P_i(C|X)$ . From the learning stage using a training set, the likelihoods from the BCs are properly normalized in order to obtain actual probabilities to be used in (2); the summation operates only on the set of correctly classified examples, of size  $m$ . Knowing  $H_i$ , the weight  $w_i$  for each  $i^{th}$  classifier is estimated as being inversely proportional to Entropy as follows:

$$w_i = \frac{1 - \left( \frac{H_i}{\sum_{j=1}^N H_j} \right)}{(N-1)}, i = \{1, \dots, N\}, N > 1, \quad (3)$$

where  $H_i$  is the value of Entropy resultant from (2). The denominator in (3) guarantees that  $\sum_{i=1}^N w_i = 1$ . This weighting strategy will smooth the base classifier's response by continuously multiplying its classification belief by the correspondent weight.

## 3 Dynamic Bayesian Mixture Models: DBMM

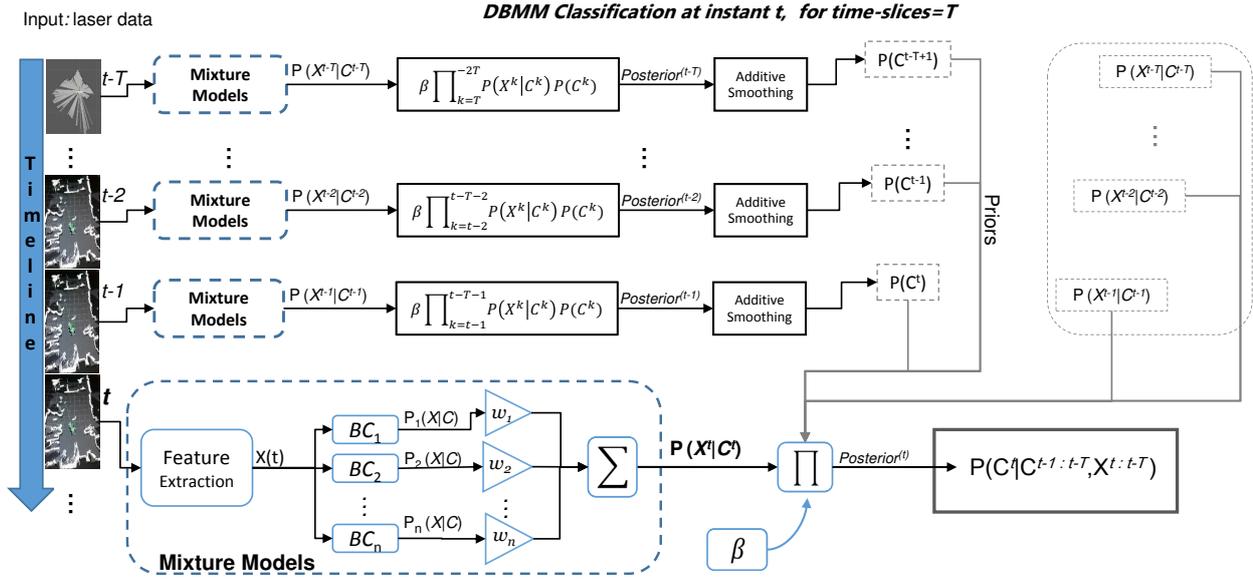
The DBMM is formulated in the same way a DBN except that the mixture models part is integrated into the network. In other words, in the DBMM network different base classifiers are weighted, resulting in a combined expression for  $P(X|C)$  as in (1). For the problem of interest, the DBMM is formulated in terms of the current time  $t$ , and the set of finite and previous (*past*) time-slices  $(t-1, \dots, t-T)$ . This section ends with a technique, called *additive smoothing*, used to prevent the undesirable situation where the prior for some of the classes tends to be very close to zero.

### 3.1 The DBMM structure

The DBMM structure is composed of the mixture probabilistic outputs  $P(X|C)$  (1) and the *a-priori* class probabilities  $P(C)$ , on a time basis, as illustrated in Fig. 1. The time-based order  $T$  specifies the number of time slices. The DBMM works according to a finite size sliding "window" approach of time slices (Faria et al 2014). Basically, as the inference process moves forward to the next time step  $\rightarrow (t+1)$ , the oldest time slice is dropped off the network.

In its simplest form i.e., for one time-slice  $T = 1$ , the structure of a DBMM with nodes  $X$  and  $C$  is modeled by the joint probability given by

$$P(X^t, X^{t-1}, C^t, C^{t-1}) = P(X^t|X^{t-1}, C^t, C^{t-1})P(X^{t-1}|C^t, C^{t-1}) \times P(C^t|C^{t-1})P(C^{t-1}). \quad (4)$$



**Fig. 1** Illustrative representation of the DBMM approach with  $T$  time-slices, where  $k = t, \dots, t - T$ . The posterior depends on the priors  $P(C^k)$ , the combined probabilities from the base-classifiers  $P(X^k|C^k)$ , and the normalization  $\beta$ .

More generally, for  $T$  time steps, the joint-probability is expressed as

$$P(X^t, \dots, X^{t-T}, C^t, \dots, C^{t-T}) = \prod_{k=t}^{t-T} P(X^k | \bigwedge_{j=k-1}^{t-T} X^j, \bigwedge_{k=t}^{t-T} C^k) \times \prod_{k=t}^{t-T} P(C^k | \bigwedge_{j=k-1}^{t-T} C^j) P(C^{t-T}). \quad (5)$$

To obtain the posterior probability, the quantity of interest here, the product rule can be used as

$$P(C^t | C^{t-1}, X^t, X^{t-1}) = \beta^{-1} P(X^t | X^{t-1}, C^t, C^{t-1}) \times P(X^{t-1} | C^t, C^{t-1}) P(C^t | C^{t-1}) P(C^{t-1}), \quad (6)$$

where, for  $nc$  classes, the normalization is ensured by  $\beta = \sum_{i=1}^{nc} P(X^t | X^{t-1}, C_i^t, C_i^{t-1}) P(X^{t-1} | C_i^t, C_i^{t-1}) P(C_i^t | C_i^{t-1}) P(C_i^{t-1})$ . To make the problem more tractable, two assumptions are considered. First,  $X$  is considered to be independent of previous  $X$ -nodes i.e.,  $P(X^t | X^{t-1}, C^t, C^{t-1}) = P(X^t | C^t, C^{t-1})$ . Secondly, the nodes are not conditionally dependent of later (future) nodes e.g.,  $P(X^{t-2} | C^t, C^{t-1}, C^{t-2}) = P(X^{t-2} | C^{t-2})$ . As consequence, the transition probabilities between classes reduces to the probability of the current-time class

$$P(C^t | C^{t-1}) = \underbrace{P(C^{t-1} | C^t) P(C^t)}_{P(C^{t-1})} / P(C^{t-1}) = P(C^t), \quad (7)$$

as shown for  $T = 1$ . Finally, and to reinforce the network “memory”, previous posterior probabilities become new (current) priors e.g., for  $T = 2$ , it is considered that  $P(C^t) \leftarrow P(C^{t-1} | C^{t-2}, X^{t-1}, X^{t-2})$ .

According to the developments given above, a general expression for a DBMM with  $T$  time-slices can be obtained as follows

$$P(C^t | C^{t-1}, \dots, C^{t-T}, X^t, \dots, X^{t-T}) = \beta \prod_{k=t}^{t-T} P(X^k | C^k) P(C^k) \quad (8)$$

where, for instance, the current prior element takes the value of the previous posterior i.e.,  $P(C^t) \leftarrow P(C^{t-1} | C^{t-2}, \dots, X^{t-2}, X^{t-1})$ ,

$X^{t-1}, \dots, X^{t-T-1}$ ), and so on. Finally, and knowing that  $P(X^k | C^k)$  is actually a mixture of probabilities as indicated earlier in (1), the explicit expression for the DBMM with  $T$  time-slices, after dropping the normalization factor  $\beta$ , assumes the form

$$P(C^t | C^{t-1:t-T}, X^{t:t-T}) \propto \prod_{k=t}^{t-T} (\sum_i w_i P_i(X^k | C^k)) P(C^k). \quad (9)$$

In summary, the *a-posteriori* of classes, given the current and past parent nodes, is proportional to the product of the weighted conditional probabilities, given by (1), and the priors. Finally, for  $T = 0$ , expressions (8) and (9) define a DBMM with just the current time step.

### 3.2 Additive smoothing for the prior probabilities

The structure of the DBMM, described in Sect. 2.2 and summarized in (9), assigns the values of the current-time posterior probabilities to the *a-priori* probabilities that will be used in the next time-slice. This is an effective technique that precludes specifying the prior distribution in advance, and conversely allowing the priors to be estimated sequentially; this strategy can be referred as ‘conjugate prior’ estimation, as discussed in (Duda. et al 2001). Nevertheless, in such sequential class estimation problem it may happen that the probability of a given class become unacceptably close to zero. This problem can be solved by ‘additive smoothing’, which is carried out by adding a term ( $\alpha$ ) to the prior distribution.

The technique used here to avoid close-to-zero values on priors is called Lidstone smoothing (Chen and Goodman

1996), that consists on adding a term  $\alpha < 1$  to the prior  $P(C)$ , expressed by

$$\hat{P}(C_i) = \frac{P(C_i) + \alpha}{P(C_i) + \alpha \cdot nc}, i = 1, \dots, nc \quad (10)$$

where  $\hat{P}(C_i)$  is the smoothed prior,  $nc$  is the number of classes, and  $\alpha$  is the smoothing parameter. We performed the experiments in semantic place classification considering values of  $\alpha$  within the interval  $[0, 0.1]$ . The value of  $\alpha$  has to be specified to prevent zero-probabilities ( $\alpha > 0$ ) but also to maintain the prior distribution over the classes as close as possible to the distribution before the additive parameter. This condition is guaranteed by small values, thus we limited  $\alpha$  to be smaller than 0.1.

## 4 Datasets

In this section, the semantic place labeling datasets named IDOL2 (hereafter just IDOL) and COLD, which are used in our experiments, are briefly described (more details are provided in (Pronobis et al 2010) and (Luo et al 2007)). The first dataset consists of 24 data sequences, collect using two mobile robots, and is characterized by five semantic-classes; further details are given in Sect. 4.1. Regarding the second dataset, actually in this work we have considered one of its three sub-datasets, namely the Saarb-COLD was used in our experiments because of two reasons: it has a greater number of classes than others COLD’s sub-sets, and it provides laserscanner data while Ljubl-dataset doesn’t<sup>2</sup>.

### 4.1 IDOL dataset

The Image Database for rObot Localization (IDOL) (Luo et al 2007) comprises 24 sequences with data from a monocular camera, laserscanner and odometry system, collected using two mobile robot platforms (a PeopleBot and a PowerBot robots; see <http://www.cas.kth.se/IDOL/>). Semantic places are represented by five indoor categories: “1-person office” (1pO), “2-persons office” (2pO), “Corridor” (CR), “Kitchen” (KT), and “Printer area” (PR). Each robot was manually driven through the indoor environments while acquiring data at 5 fps. The data sequences were collected under varying illumination conditions and during different time periods. The total of 24 data sequences are the result of 4 sequences, per mobile robot, recorded under the 3 weather/illumination conditions (sunny, cloudy, night). Of these 4 sequences, the first two were acquired during January and February (with a time span of 2 weeks), and the remaining two sequences were recorded during June and July (again, with a time span of 2 weeks). The time interval between the sequences pairs

**Table 1** IDOL Database: recording conditions

| RobotA (PowerBot)  |       |       |       |       |       |
|--------------------|-------|-------|-------|-------|-------|
| Cloudy             |       | Night |       | Sunny |       |
| ID                 | Month | ID    | Month | ID    | Month |
| 1                  | Feb   | 5     | Feb   | 9     | Feb   |
| 2                  | Feb   | 6     | Feb   | 10    | Feb   |
| 3                  | Jun   | 7     | Jun   | 11    | Jun   |
| 4                  | Jul   | 8     | Jun   | 12    | Jun   |
| RobotB (PeopleBot) |       |       |       |       |       |
| Cloudy             |       | Night |       | Sunny |       |
| ID                 | Month | ID    | Month | ID    | Month |
| 13                 | Jan   | 17    | Jan   | 21    | Feb   |
| 14                 | Jan   | 18    | Jan   | 22    | Feb   |
| 15                 | Jul   | 19    | Jun   | 23    | Jun   |
| 16                 | Jul   | 20    | Jun   | 24    | Jun   |

is approximately of 6 months. Therefore, the dataset covers a wide range of variations introduced by illumination and weather conditions, presence or absence of people, furniture/objects relocated, viewpoint differences, etc. Table 1 summarizes the IDOL dataset where, per each robot, there are 12 data sequences divided into 3 groups according to the illumination conditions, with each group having 4 sequences.

### 4.2 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 (Ullah et al 2008). 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 a mobile robot during data acquisition, the Standard (STD) and the Extended (EXT) 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 (Ullah et al 2008) for 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.

### 4.3 Laser-based features

In both datasets i.e., IDOL and COLD-Saarbrücken, the mobile robots used to record data were equipped, besides other sensors and instruments, with 2D laserscanners mounted on-board the robots. In the experiments carried out in this work, only laser-based features were used as basis for the supervised learning algorithms. In particular, a subset of 50 components from the geometrical features proposed in (Mozos

<sup>2</sup> Freib-dataset has laser data indeed, but with low resolution.

2010) (the so called  $B$  and  $P$ -features) were employed in our experiments. The aforementioned  $B\&P$  features are computed from the raw laser scans, where the  $B$ -features are calculated using the laser-beams and the  $P$ -features are calculated from a polygonal approximation of the area covered by the laser-scan. The components of the feature vector used in this work are detailed in Table II of (Premebida et al 2015). The reasons for using the  $B\&P$  features are twofold: to allow a fair comparison with previous works that use laser data and to demonstrate that with a low complexity feature vector, of only 50 elements, it is possible to achieve very good performance.

## 5 Experiments and performance evaluation

In this section the classification performance of the DBMM, applied on the semantic place labeling datasets described in Sect. 4, is evaluated in terms of (i) the number of time-slices, and (ii) the effect of the additive parameter on the priors. For the IDOL dataset, we also evaluate the influence of combining (mixture) base-classifiers in the DBMM structure. The overall 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. We primarily adopted  $F_{measure}$ <sup>3</sup> because all datasets have unbalanced classes.

The mixture model of the network (denoted hereafter by BMM) is composed of 3 BCs i.e.,  $N = 3$  in (1), where  $BC_1$  is a SVM using linear-kernel and usual parameters,  $BC_2$  is a MLP Neural Network with 10 hidden nodes, and  $BC_3$  is another lin-SVM using a margin parameter  $C = 100$ . The implementations use libSVM<sup>4</sup> and the Neural Network Toolbox of Matlab. All BCs are learned using the same training set, and the outputs are normalized in order to delivery probabilistic estimates.

The experiments are firstly conducted on the IDOL dataset, seeking to verify the classification performance of the mixture model against the base classifiers. In a first experiment, temporal relationship inside the DBMM structure is not considered i.e., the classification depends only on the response from the mixture of BCs. Secondly, a series of experiments using the DBMM for increasing number of  $T$  is carried out and the results are reported in Sect. 5.1. Finally, Sect. 5.2 brings the experiments on the Saarb-dataset where different paths and locations are interchanged between training and testing sets.

**Table 3** Results on IDOL for the BCs and the mixture model, in terms of  $F_{measure}$ .

|                     | $BC_1$ | $BC_2$ | $BC_3$ | $BMM$ |
|---------------------|--------|--------|--------|-------|
| <i>Exp.1</i>        | 90.7   | 92.9   | 93.0   | 93.7  |
| <i>Exp.2</i>        | 87.7   | 86.1   | 88.2   | 88.4  |
| <i>Exp.3</i>        | 83.7   | 79.8   | 82.4   | 84.7  |
| <i>Exp.4</i>        | 84.4   | 80.6   | 83.5   | 85.5  |
| $\hat{F}_{measure}$ | 86.6   | 84.9   | 86.8   | 88.1  |

### 5.1 Experiments on IDOL dataset

The experiments performed on IDOL follow, essentially, the same methodology described in (Pronobis et al 2010) but, we opted to conduct the most challenging experiments reported in (Pronobis et al 2010) (thus, the experiments *under stable illumination conditions* were not performed here). In summary, four experiments are carried out as follows:

1. Exp.1 (under varying illumination conditions and close in time), performed separately for each robot.
2. Exp.2 (under varying illumination conditions and distant in time), performed separately for each robot.
3. Exp.3 (recognition across robot platforms, same illumination conditions).
4. Exp.4 (recognition across robot platforms, different illumination conditions).

The last two experimental runs (Exp.3 and 4) were carried out to assess the generalization performance in very challenging conditions. Exp.3 follows similar methodology as reported in (Pronobis et al 2006), while Exp.4 is an additional experimental case presented here. Table 2 summarizes these four experiments in terms of training and testing sets.

As described in Sect. 2.3, the class-conditional probability output of the DBMM is a weighted combination of BCs. We begin by evaluating the framework with non-sequential (time) decision (referred as BMM), in order to assess the effect of the weighting strategy and to compare the results with the BCs. Classification results achieved by weighting the BCs, as well as the results from the BCs, are shown in Table 3. The results indicate the effectiveness of combining a set of classifiers into the mixture model using the method described in Sect. 2.4.

For the ‘dynamic part’, which is of particular interest in this work, the DBMM is evaluated in terms of the number of time-slices and as function of the smoothing parameter  $\alpha$ . The impact on the  $F_{measure}$  of incorporating time-based inferences is shown in Fig. 2, for each of the four experiments listed above (Exp.1,2,3,4), where  $\alpha$  varies from 0 to 0.1. The experiments were conducted for  $T$  in the interval  $[0, \dots, 4]$ . Notice that when  $T = 0$  the ‘dynamic’ behavior of the DBMM depends only on the current time step (this is in accordance with the convention adopted in Sect. 3.1). In all

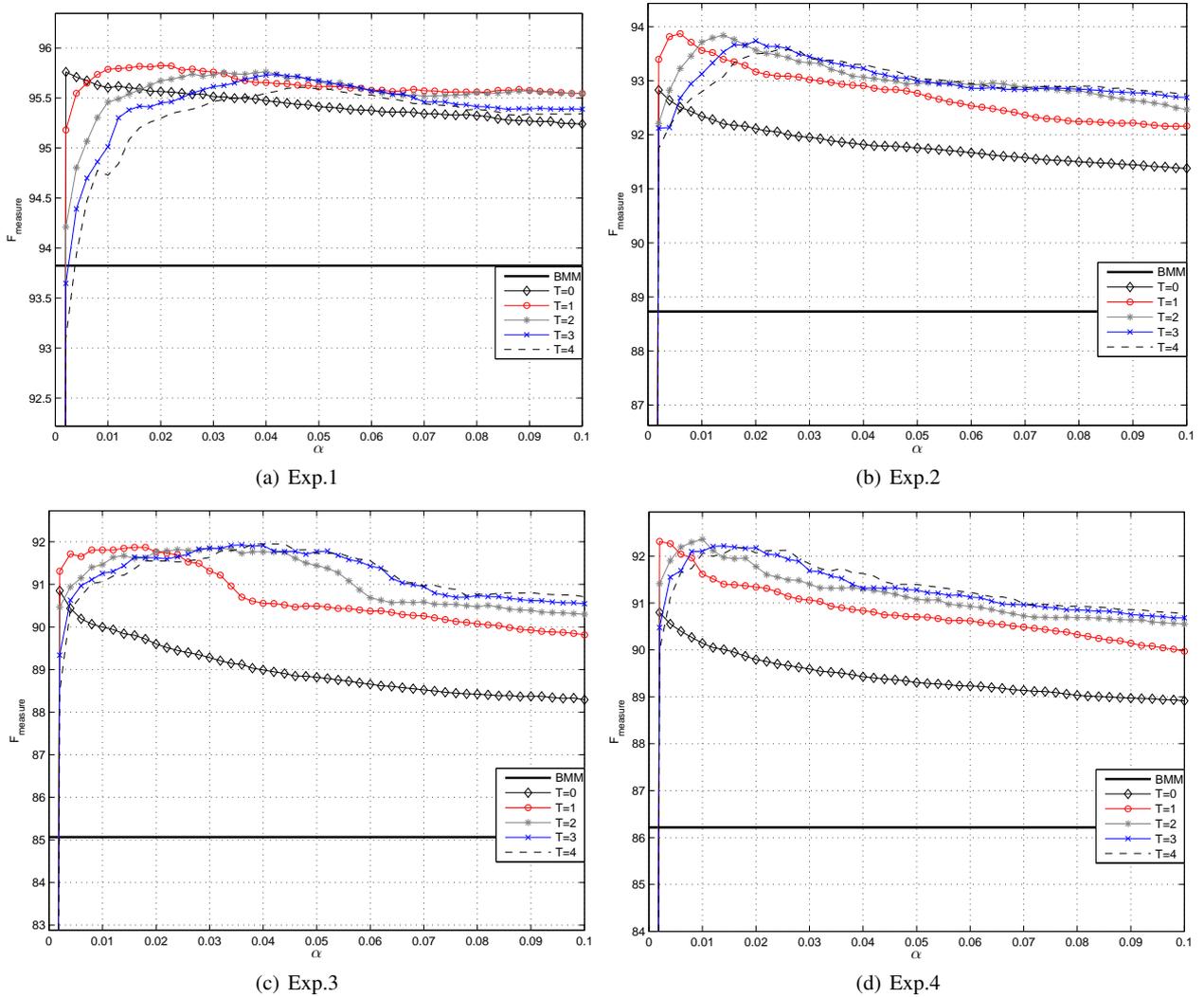
<sup>3</sup> In this paper the values of  $F_{measure}$  are presented in percentage.

<sup>4</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

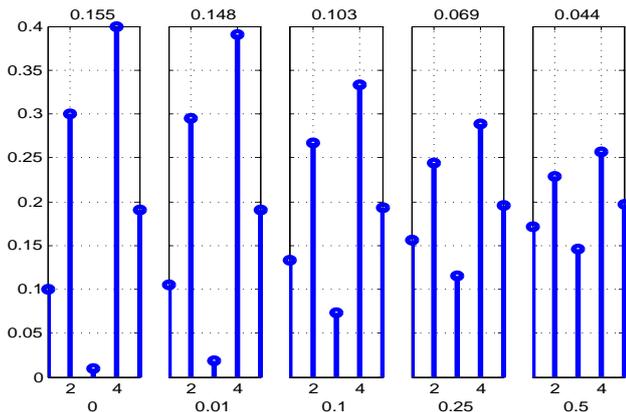
**Table 2** Experiments on IDOL dataset. Sequences ID, between brackets, are from Table 1.

| Exp.1                      |         | Exp.2                      |         | Exp.3         |               | Exp.4                      |                       |
|----------------------------|---------|----------------------------|---------|---------------|---------------|----------------------------|-----------------------|
| Train $\rightleftharpoons$ | Test    | Train $\rightleftharpoons$ | Test    | Train         | Test          | Train $\rightleftharpoons$ | Test                  |
| {1,2}                      | {5,6}   | {1,2}                      | {7,8}   | {1,2,3,4}     | {13,14,15,16} | {1,2,3,4}                  | {17,...,24}           |
| {1,2}                      | {9,10}  | {1,2}                      | {11,12} | {5,6,7,8}     | {17,18,19,20} | {5,6,7,8}                  | {13,...,16,21,...,24} |
| {3,4}                      | {7,8}   | {3,4}                      | {5,6}   | {9,10,11,12}  | {21,22,23,24} | {9,10,11,12}               | {13,...,20}           |
| {3,4}                      | {11,12} | {3,4}                      | {9,10}  |               |               |                            |                       |
| {5,6}                      | {1,2}   | {5,6}                      | {3,4}   | {13,14,15,16} | {1,2,3,4}     | {13,14,15,16}              | {5,...,12}            |
| {5,6}                      | {9,10}  | {5,6}                      | {11,12} | {17,18,19,20} | {5,6,7,8}     | {17,18,19,20}              | {1,...,4,9,...,12}    |
| {7,8}                      | {3,4}   | {7,8}                      | {1,2}   | {21,22,23,24} | {9,10,11,12}  | {21,22,23,24}              | {1,...,8}             |
| {7,8}                      | {11,12} | {7,8}                      | {9,10}  |               |               |                            |                       |
| {9,10}                     | {1,2}   | {9,10}                     | {3,4}   |               |               |                            |                       |
| {9,10}                     | {5,6}   | {9,10}                     | {7,8}   |               |               |                            |                       |
| {11,12}                    | {3,4}   | {11,12}                    | {1,2}   |               |               |                            |                       |
| {11,12}                    | {7,8}   | {11,12}                    | {5,6}   |               |               |                            |                       |

$\rightleftharpoons$  means that training and testing sets are interchanged. Exp.1 and Exp.2 are here performed for the RobotA.



**Fig. 2** Evolution of the  $F_{measure}$ , per values of  $\alpha$  and  $T = [0, \dots, 4]$ , shown for the four experiments on the IDOL dataset as described in Sect. 5.1. The legends indicates the DBMM for different values of time-slices. These curves clearly demonstrate improvement on the performance of the DBMM when the ‘dynamic’ part is considered. Here, the legend *BMM* indicates a DBMM without time steps.



**Fig. 3** An example of the influence of  $\alpha$  on a given  $P(C)$ . Distributions are shown, from left to right, for increasing values of  $\alpha$ . Additionally, standard deviation is provided in the top of each subplot.

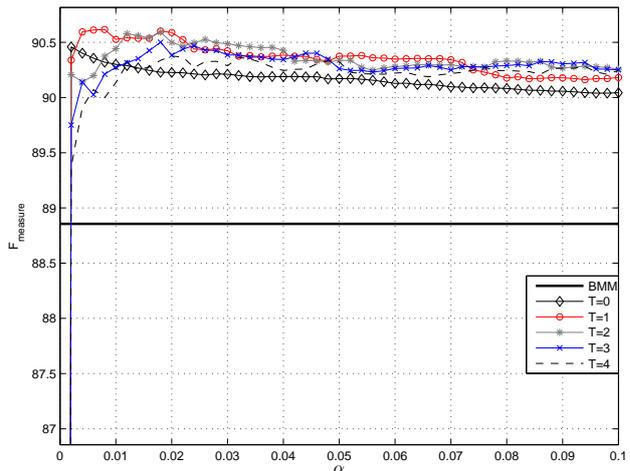
cases, and for any  $\alpha > 0.01$ , the classification performance when temporal relationship is taken into account improved significantly in comparison with the case of non-temporal integration (indicated by BMM). The plots in Fig. 2 show that the performance on all the four experiments is improved when time-slices are taken into consideration.

As expected, performance drops as the additive parameter increases. This happens because the prior distribution becomes to lose its definiteness due to the uniform “bias” induced by  $\alpha$ . This can be seen as follows: let  $P(C) = (0.1, 0.3, 0.01, 0.4, 0.19)$  be a given prior distribution for five classes, and let consider the additive term as  $\alpha = (0, 0.01, 0.1, 0.25, 0.5)$ , applying normalization to guarantee the total probability mass is unity, Fig. 3 illustrates the effect of an additive term on a prior distribution.

Regarding the curves presented in Fig. 2, for  $T = (2, 3, 4)$  the classification performance have approximate behavior, while for  $T = 1$  the response tends to follow the previous cases but with higher peaks (although for a short period) in most of the experiments. Finally, for  $T = 0$  the DBMM reach the peak shortly at  $\alpha > 0$  and then it follows a monotonic decreasing function with average classification error higher than the DBMMs with  $T > 0$ . Further discussion is provided in Sect. 5.3.

## 5.2 Experiments on COLD-Saarb

In Sect. 4.2 the COLD-Saarb dataset is concisely described, while detailed information can be found in (Ullah et al 2008). In (Premebida et al 2015), and in accordance with the experiments carried out in (Ullah et al 2008), exhaustive experimental results were reported for different conditions of illumination and portions (“A” and “B”), and separately for STD and EXT sequences. Additionally, experiments involving both sequences were also reported. This work concentrates on this last experimental part, where  $F_{measure}$  is used to



**Fig. 4** Results on the COLD-Saarb as addressed in Sect. 5.2.

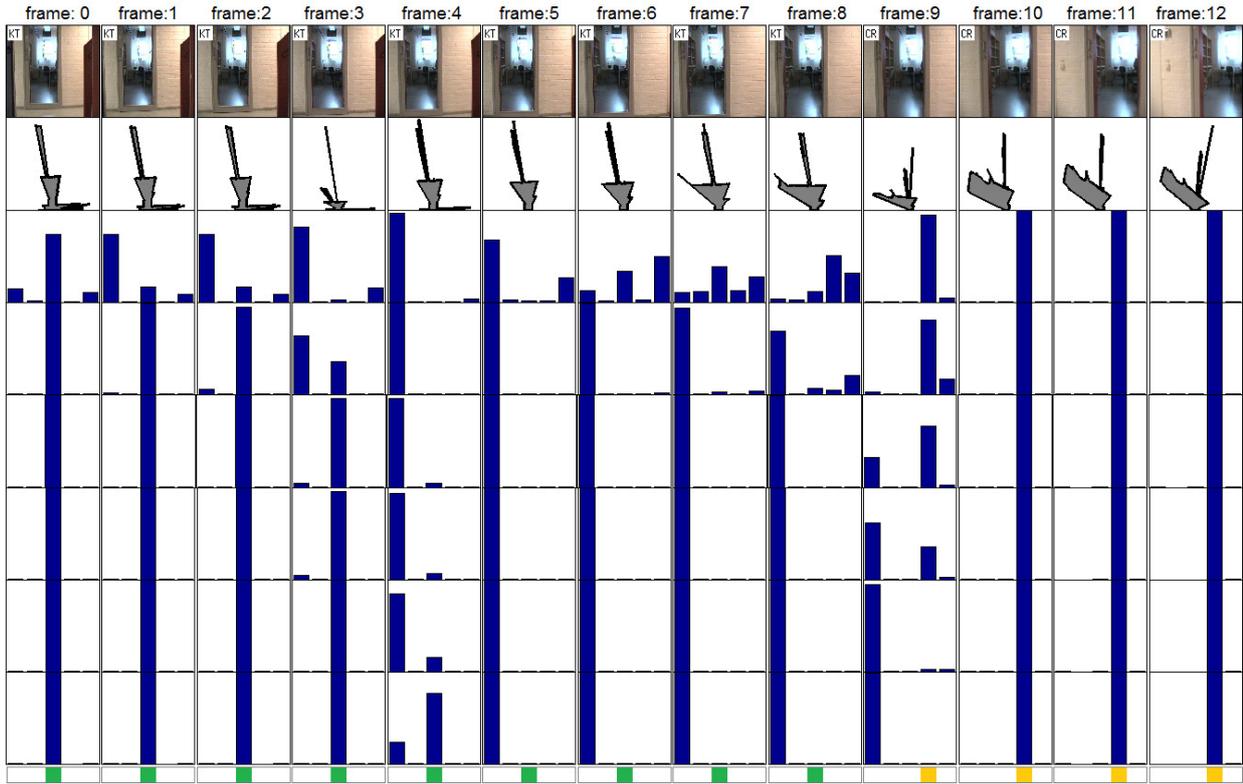
assess classification performance considering STD vs EXT, alternating between training and testing.

Some of the classes in COLD-Saarb are labeled only for the EXT path and some only for one of the portions A or B (more details in Table I of (Ullah et al 2008)). Therefore, for the experiments presented in this section the classes that are present in both sequences (EXT and STD) have been considered, they are: “Corridor” (CR), “Printer area” (PA), “Bath room/Toilette” (TL), and “Person office” (PO); here, PO assembles the classes “1-person office” and “2-persons office”.

Figure 4 shows  $F_{measure}$  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. The results, approximately proportional to the behavior on the IDOL dataset, show that when time-slices are integrated into the system the performance is much better than a solution without dynamic nodes.

## 5.3 Discussion

Experiments on the IDOL and COLD-Saarb datasets were primarily conducted to evaluate the DBMM’s performance with regard to (i) the mixture models and (ii) the number of time-slices. The first experimental results, summarized in Table 3, indicate a better performance when combining classifiers in a DBN framework. The second round of experiments gave a clear indication of the effect when time-based nodes (states) are used in the system. As explained in sec. 3.2, in order to obtain consistent results a term was added to the prior probabilities. Based on the reported results on both datasets, provided in Figs. 2 and 4, an ‘optimal’ value of  $\alpha$  is



**Fig. 5** Classification results for a short sequence of an indoor scenario, extracted from IDOL dataset, using laser-based features (see Sect. 4.3) and the DBMM with increasing number of time slices. The first row depicts images captured by an onboard camera; the second row shows the laser scans; the third row provides the results of a DBMM without time-slices, and the subsequent rows show classification probabilities for DBMMs with  $T = 0, 1, 2, 3, 4$  respectively. The color bars at the bottom of the figure indicates the ground-truth label: *green* indicates ‘kitchen’ (KT), and *yellow* denotes ‘corridor’ (CR). In this work the image frames are not used in the classification i.e., they are shown for illustrative purposes.

not the same for all values of  $T$  neither for all experiments. However, it is clear that  $\alpha$  should be small,  $0 < \alpha < 0.05$ .

The approximate value of  $\alpha$  with the highest  $F_{measure}$ , according to the average results in the experiments on IDOL, are  $\alpha = (0.002, 0.011, 0.023, 0.028, 0.032)$ . Results of the DBMM with these values of  $\alpha$  and for  $T = (0, \dots, 4)$  are presented in Table 4, where the average performances of the DBMM with  $0 < T \leq 4$  are similar; this allows us to conclude that a DBMM with  $T = 1$  is a reasonable choice for the best network, under the assumption that it will require less computational effort and lower complexity than for  $T > 1$ .

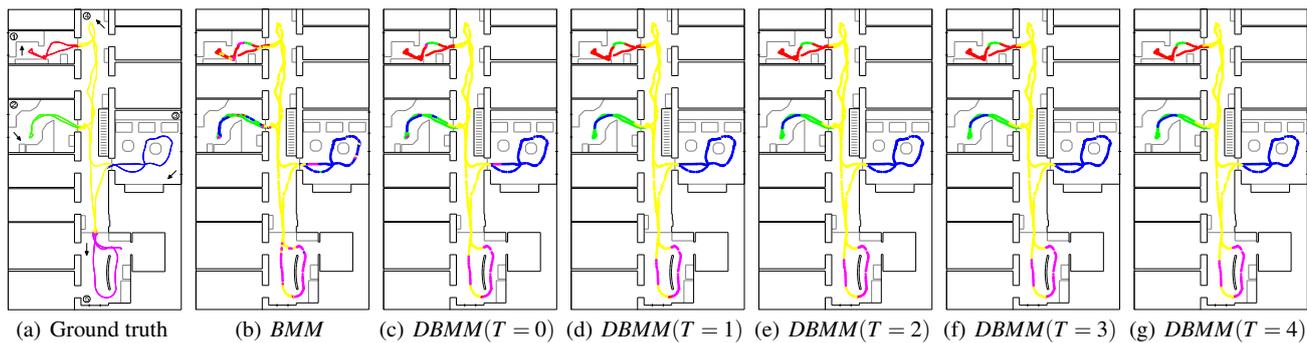
Figure 5 shows classification results on a part (only 13 frames) of a sequence from the IDOL dataset. The results illustrate the ‘temporal’ behavior of the DBMM for increasing number of time-slices. The third row shows the results for the approach without time steps (i.e., BMM), which presents more variations in the response than the DBMMs with incorporated time-slices. Conversely, as the number of time-slices increases, the sequential response becomes less sensible to variations on the scene, but the ‘latency’ of the DBMM is more evident. In the case shown in Fig. 5, from frame 5 to 8 all approaches, except BMM, fail to classify correctly; frame

4 was successfully classified only by  $DBMM^{(T=4)}$ , while on frame 9 the  $DBMM^{T=2,3,4}$  did not perform well.

Finally, Fig. 6 provides results along the path driven by the robot in one of the testing sequences in IDOL dataset. The map of the environment is divided in five places: ① (1pO: office); ② (2pO: office); ③ (KT: kitchen); ④ (CR: corridor); ⑤ (PR: printer office). As can be seen from Figs. 6(b-g), as the number of slices  $T$  increases the classification response becomes more stable and, therefore, the occurrence of changing between categories tends to decrease. In terms of transition errors, from one place to another, the passage from place ④ to ⑤ is often a cause of misclassification. Further classification errors occur in place ① and ②, often misclassified as 2pO and KT respectively. These detailed results are from Exp.3, which is the most challenging experiment conducted in this paper.

## 6 Conclusion

We have introduced an effective form of the Dynamic Bayesian Network (DBN), modeled as a sequential classification network, for the semantic place recognition problem in the scope of mobile robotics. Based on the Dynamic Bayesian Mixture



**Fig. 6** Maps showing the classification results obtained by the DBMM approach. The colors encode the categories for each frame along the path: ■ - 1pO; ■ - 2pO; ■ - KT; ■ - CR; ■ - PR. These results, for every robot position, were obtained in a testing (unseen) sequence from Exp.3.

**Table 4** Results on IDOL, averaged for the classes.

|                     | <i>BMM</i> | $T = 0$ | $T = 1$ | $T = 2$ | $T = 3$ | $T = 4$ |
|---------------------|------------|---------|---------|---------|---------|---------|
| <i>Exp.1</i>        | 93.7       | 95.6    | 95.7    | 95.6    | 95.5    | 95.4    |
| <i>Exp.2</i>        | 88.4       | 92.7    | 93.4    | 93.4    | 93.4    | 93.2    |
| <i>Exp.3</i>        | 84.7       | 90.6    | 91.6    | 91.6    | 91.7    | 91.6    |
| <i>Exp.4</i>        | 85.5       | 90.3    | 91.0    | 91.1    | 91.5    | 91.3    |
| $\bar{F}_{measure}$ | 88.1       | 92.3    | 92.9    | 92.9    | 93.0    | 92.9    |

Model (DBMM), introduced by Faria et al (2014) and applied to the place classification problem in (Premebida et al 2015), in this paper we present a general expression of the DBMM in terms of a finite set of time-slice nodes, also valid for a DBN, modeled as a product of *past-posteriors* and priors probabilities. Extensive experiments using datasets from publicly available repository were carried out to assess the performance of the DBMM on semantic place classification. Additionally, this paper brings evidence of the impact of additive smoothing on the DBMM network’s performance.

From the several experimental results reported in this work, the DBMM demonstrated to be a very promising approach, with interesting characteristics: (i) DBMM supports general probabilistic class-conditional models; (ii) dynamic information in the form of priors and *past-inferences* can be easily incorporated; (iii) DBMM enables the combination of a diversity of base-classifiers. In conclusion, from this study we learned that the proposed method can be successfully applied in sequential (time-based) multi-class place recognition problems, being a very powerful solution to be followed due its low complexity, faster implementation and its direct probabilistic interpretation.

**Acknowledgements** This work was partially supported by the Portuguese Foundation for Science and Technology (FCT) and FEDER through COMPETE 2020 under grants RECI/EEI-AUT/0181/2012 (AMSHMI12) and UID/EEA/00048/2013.

## References

- Chen SF, Goodman J (1996) An empirical study of smoothing techniques for language modeling. In: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, Stroudsburg, PA, USA., ACL ’96, pp 310–318.
- Costante G, Ciarfuglia T, Valigi P, Ricci E (2013) A transfer learning approach for multi-cue semantic place recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 2122–2129.
- Duda RO, Hart PE, Stork DG (2001) Pattern classification. John Wiley & Sons, Inc, NY.
- Faria D, Premebida C, Nunes U (2014) A probabilistic approach for human everyday activities recognition using body motion from RGB-D images. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN., pp 732–737.
- Hong S, Kim J, Pyo J, Yu SC (2015) A robust loop-closure method for visual SLAM in unstructured seafloor environments. Autonomous Robots pp 1–15.
- Jung H, Martinez Mozos O, Iwashita Y, Kurazume R (2014) Indoor place categorization using co-occurrences of LBPs in gray and depth images from RGB-D sensors. In: Fifth International Conference on Emerging Security Technologies (EST).
- Jung H, Mozos OM, Iwashita Y, Kurazume R (2016) Local n-ary patterns: a local multi-modal descriptor for place categorization. Advanced Robotics 30(6):402–415
- Korb KB, Nicholson AE (2010) Bayesian Artificial Intelligence, Second Edition, 2nd edn. CRC Press, Inc., Boca Raton, FL, USA.
- Kostavelis I, Gasteratos A (2015) Semantic mapping for mobile robotics tasks: A survey. Robotics and Autonomous Systems 66:86–103.
- Luo J, Pronobis A, Caputo B, Jensfelt P (2007) Incremental learning for place recognition in dynamic environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), San Diego, CA, USA, pp 721–728.
- Martinez-Gomez J, Garcia-Varea I, Cazorla M, Morell V (2015) ViDRILo: The visual and depth robot indoor localization with objects information dataset. The International Journal of Robotics Research
- McManus C, Upcroft B, Newman P (2015) Learning place-dependant features for long-term vision-based localisation. Autonomous Robots 39(3):363–387.
- Mihajlovic V, Petkovic M (2001) Dynamic Bayesian Networks: A State of the Art. Technical reports series, TR-CTIT-34
- Milford M (2013) Vision-based place recognition: how low can you go? The International Journal of Robotics Research 32(7):766–789.

- Mozos OM (2010) *Semantic Place Labeling with Mobile Robots*. Springer, Tracts in Advanced Robotics (STAR).
- Posner I, Cummins M, Newman P (2009) A generative framework for fast urban labeling using spatial and temporal context. *Autonomous Robots* 26(2-3):153–170.
- Premevida C, Faria D, Souza F, Nunes U (2015) Applying probabilistic mixture models to semantic place classification in mobile robotics. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Pronobis A, Jensfelt P (2012) Large-scale semantic mapping and reasoning with heterogeneous modalities. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp 3515–3522.
- Pronobis A, Caputo B, Jensfelt P, Christensen HI (2006) A discriminative approach to robust visual place recognition. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp 3829–3836.
- Pronobis A, Mozos OM, Caputo B, Jensfelt P (2010) Multi-modal semantic place classification. *The International Journal of Robotics Research (IJRR)* 29(2-3):298–320.
- Rogers J, Christensen H (2012) A conditional random field model for place and object classification. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp 1766–1772.
- Shi L, Kodagoda S, Dissanayake G (2010) Laser range data based semantic labeling of places. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp 5941–5946.
- Shi L, Kodagoda S, Dissanayake G (2012) Application of semi-supervised learning with voronoi graph for place classification. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp 2991–2996.
- Shi L, Kodagoda S, Piccardi M (2013) Towards simultaneous place classification and object detection based on conditional random field with multiple cues. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp 2806–2811.
- Ullah M, Pronobis A, Caputo B, Luo J, Jensfelt R, Christensen H (2008) Towards robust place recognition for robot localization. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp 530–537.
- Vasudevan S, Siegwart R (2008) Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *Robotics and Autonomous Systems* 56(6):522–537.
- Werner F, Sitte J, Maire F (2012) Topological map induction using neighbourhood information of places. *Autonomous Robots* 32(4):405–418.
- Wu J, Christensen H, Rehg J (2009) Visual place categorization: Problem, dataset, and algorithm. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp 4763–4770.
- Yuan L, Chan KC, George Lee C (2011) Robust semantic place recognition with vocabulary tree and landmark detection. In: *ASP-AVS-11 Workshop in conjunction with the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.