A Bayesian Approach for Place Recognition

Fabio T. Ramos   Ben Upcroft   Suresh Kumar   Hugh F. Durrant-Whyte

ARC Centre of Excellence for Autonomous Systems
Australian Centre for Field Robotics
The University of Sydney
Sydney, NSW 2006, Australia
{f.ramos,b.upcroft,suresh,hugh}@acfr.usyd.edu.au

Abstract

This paper presents a robust place recognition algorithm for mobile robots. The framework proposed combines nonlinear dimensionality reduction, nonlinear regression under noise, and variational Bayesian learning to create consistent probabilistic representations of places from images. These generative models are learnt from a few images and used for multi-class place recognition where classification is computed from a set of feature-vectors. Recognition can be performed in near real-time and accounts for complexity such as changes in illumination, occlusions and blurring. The algorithm was tested with a mobile robot in indoor and outdoor environments with sequences of 1579 and 3820 images respectively. This framework has several potential applications such as map building, autonomous navigation, search-rescue tasks and context recognition.

1 Introduction

Localisation in complex environments is one of the main challenges for autonomous navigation. Currently, this task is performed by solving the simultaneous localisation and map building problem (SLAM) [7]. The main issue of this approach is the identification of landmarks in unstructured environments - they have to facilitate detection and association. On the other hand, rather than navigating using relative coordinates, humans have an abstract notion of distance but are still able to recognise where they are in the space. This ability is provided mainly from visual information associated with an internal (map) representation that, rather than localising the person with distances from objects, localises it based on the appearance of that scene. The same approach is used here for robot localisation in two different situations, indoor and outdoor environments. Places are learnt and identified from images obtained with a camera fixed in a mobile robot. The aim is thus to learn a multi-class classifier to label new images as the robot navigates. No other extra information such as a topological map is provided.

Place recognition and localisation from images are not new in robotics. Some previous approaches use image histograms and topological maps for classification [21, 14]. Others use invariant features such as [11, 22]. Place recognition was also shown to improve the performance of object recognition when both are performed simultaneously [20]. In another approach, image features were used to estimate the position of the robot for map building problems [17, 18, 5]. The novelty of this work compared to the others lies on the proposed method that allows robust recognition from few training images (usually 3 to 10 per place) and without the need of a map. Because the robot needs fewer training images, a quick learning exploration is enough for further localisation tasks. In addition, the framework was tested in both indoor and outdoor environments proving to be robust for practical applications.

In the solution proposed, the world is interpreted as a set of places. Each place has a probabilistic representation learnt from images and localisation is performed in near real-time by evaluating the responses of each model given a new image. The place recognition task is treated as a Bayesian learning problem in a space of *essential features*. Initially, training images are divided into small patches that constitute a high dimensional set. The dimensionality of this set is then reduced with nonlinear and neighbourhood preserving techniques to create a low dimensional set. These two sets are used to learn a mixture of linear models for the nonlinear regression, from points in the high to the low dimensional space. Points is this low dimensional space constitute the set of *essential features* and are used in the next step where the variational approximation for Bayesian learning is computed to create a probabilistic density for each place. Recognition is performed by computing the log-likelihood of an entire image over each place model. This approach was tested with sequences of images obtained by a mobile robot operating under different conditions - moving objects, changes in illumination, different viewpoints, occlusions, outdoor and indoor environments - and proved to be robust for localisation.

2 Approach

This paper focuses on a classification procedure to map images to labels. Each label corresponds to a place learnt from a set of images. The learning algorithm is supervised as every image in the training set has an assigned label. Thus, given a training set of n pairs $(\mathbf{I}_i, p_i)$, where $\mathbf{I}_i$ is the *i*th image and p_i is the label of that particular image, the algorithm has to generate a model able to classify new images regardless whether they come from different view points or are partially occluded.

Images are represented as sets of patches of the same size that, when combined, recover the original image. Thus an image $\mathbf{I}_i$ is represented as a set of m patches $\{I_{i,1}, \dots, I_{i,m}\}$. Each patch is convolved with a sequence of Gabor wavelets to quantify texture. This convolution has also a natural interpretation as it provides a good approximation of natural processes for spectral decomposition that occurs in the primary visual cortex. Each patch now has a feature-vector representation $x_{i,j} = \phi(I_{i,j}) \in \mathbb{R}^D$. The dimension D is usually intractable for direct density estimation. Hence, dimensionality reduction techniques are applied to extract the *essential* information of each patch and represent them in a lower dimensional space. However, this reduction should preserve important characteristics of the original space such as keeping the neighbourhood of points unchanged. This ensures that patches representing trees and grass for example are situated nearby and not mixed up with other patches even when lying in a nonlinear surface.

2.1 Neighbourhood-Preserving Dimensionality Reduction

Dimensionality reduction is one of the techniques that can manage the amount of information robotics application face. In this work, a nonlinear technique, Isomap [19], is applied to reduce the dimensionality of image patches into a feasible number where further statistical learning methods can be used. As opposed to principal components analysis (PCA) [8] and multidimensional scaling (MDS) [4], Isomap has the desired property of preserving the neighbourhood of points in the low dimensional manifold.

2.2 Noisy Non-Linear Regression

Isomap and indeed most nonlinear dimensionality reduction algorithms are inherently deterministic algorithms that do not provide a measure of uncertainty of underlying states of high dimensional observations. In addition, Isomap does not output a model or function to directly compute the low dimensional coordinates of new observations, thus requiring k-neighbours based algorithms that can be cumbersome in real-time applications.

An alternative solution is to learn a generative model $p(x|y)$, where x is the feature-vector in the high dimensional space and y is its low dimensional representation. This model encapsulates the uncertainties inherent in the inference of low dimensional points from noisy high dimensional observations. It can be learnt in a supervised manner to derive compact mappings that generalise over large portions of the input and embedding space. The input-output pairs of Isomap can serve as training data for an invertible function approximator in order to learn a parametric mapping between the two spaces. Once the model is learnt, patches of new images can have their low dimensional representation according to the manifold of the training set. This is the key point to make real-time recognition since the essential features of a new image can be quickly computed from the model by making probabilistic inferences.

Given the results of Isomap, the parameters of a mixture of linear models $p(x, y, s)$ can be learnt through Expectation Maximisation (EM) [6], where s is a hidden discrete variable representing the weights of the components. Mixture of linear models are similar to mixtures of factor analysers, that are commonly used to perform simultaneous clustering and local dimensionality reduction [9]. The only differences are that the low dimensional variable y is observed (through Isomap), not hidden, and the Gaussian distributions $p(y|s)$ have nonzero mean vectors ν_s and full covariance matrices Σ_s. Learning when the variable y is observed seems to discover a solution of better quality than in the opposite situation, the conventional mixture of factor analysers [16]. The discrete hidden variable s introduced in the model physically represents a specific neighbourhood on the manifold over which a mixture component is representative. This representation conveniently handles highly nonlinear manifolds through the capability to model the local covariance structure of the data in different areas of the manifold.

The result of the inference process in this model is a mixture of Gaussians with means μ, covariances Σ and weights s. To make it feasible for Bayesian learning this mixture is collapsed so as to have a single mean, which will be used as a training point, and a covariance matrix which will be used in the initialisation of the hyperparameters.

2.3 VBEM for Mixtures of Gaussians

The training data now represented with its essential features in the low dimensional space is used to learn a generative model for each place. This problem is formulated in a Bayesian framework where the model selection task consists of calculating the posterior distribution over a set of models (which in this case will be mixtures of Gaussians with different numbers of components) given the prior knowledge and the dataset. Denoting $\mathbf{s}$ for the hidden variable representing the weights, $\mathbf{y}_i$ for the observations of a place i, θ for the parameters of a model M, the *marginal likelihood* $p(\mathbf{y}_i|M)$ is the key expression in the Bayesian formulation for model selection. It represents an average of how good a particular model fits the observations over all possible parametrisation, convoluted by the prior. This quantity permits the comparison of different models given the data by having the Occam's Razor property[12]. However, analytical solutions are intractable so that the idea of the variational Bayesian approach is to approximate the marginal likelihood with a lower bound by using variational calculus techniques [10, 13, 2].

Introducing a free distribution q over $\mathbf{s}$ and θ, with $\int \sum_{\mathbf{s}} q\left(\mathbf{s},\theta\right) d\theta = 1$, and applying the Jensen's inequality [3], it is possible to compute a lower bound on the log of the marginal likelihood. Maximising this lower bound with respect to the free distribution $q\left(\mathbf{s},\theta\right)$ is analytically intractable. A better strategy is to factorise this free distribution to yield a variational approximation in which $q\left(\mathbf{s},\theta\right) \approx q_{\mathbf{s}}\left(\mathbf{s}\right) q_{\theta}\left(\theta\right)$:

$$\ln p\left(\mathbf{y}_i \mid M\right) \geq \int \sum_{\mathbf{s}} q_{\mathbf{s}}\left(\mathbf{s}\right) q_{\theta}\left(\theta\right) \ln \frac{p\left(\mathbf{s},\mathbf{y}_i,\theta \mid M\right)}{q_{\mathbf{s}}\left(\mathbf{s}\right) q_{\theta}\left(\theta\right)} d\theta \tag{1}$$

$$= \mathcal{F}_M\left(q_{\mathbf{s}}\left(\mathbf{s}\right), q_{\theta}\left(\theta\right), \mathbf{y}_i\right) \tag{2}$$

The quantity $\mathcal{F}_M$ is a functional of the free distributions $q_{\mathbf{s}}\left(\mathbf{s}\right)$ and $q_{\theta}\left(\theta\right)$ and is known as the negative free energy. The variational Bayesian algorithm iteratively maximises $\mathcal{F}_M$ with respect to the free distributions until the function reaches a stationary value.

An interesting implementation of VBEM can use conjugate priors that are analytically tractable and easy to interpret. Thus, Dirichlet, Normal and Wishart multivariate distributions [15] are used as priors over weights, means and covariances. They are denoted as $\mathcal{D}\left(\pi;\lambda\right)$, $\mathcal{N}\left(\mathbf{x};\mu,\mathbf{\Sigma}^{-1}\right)$ and $\mathcal{W}\left(\Gamma;\alpha,\mathbf{B}\right)$ and are functions of their hyperparameters. Also, a multivariate Student-t distribution $\mathcal{S}\left(\mathbf{x};\rho,\mathbf{\Lambda},\omega\right)$ is used to represent the predicted density.

Assuming a particular model M, the Gaussian mixture model has S components, where each component has weight given by π_s, mean μ_s and covariance $\mathbf{\Gamma}_s$. The set of parameters can be written as $\theta = \{\pi,\mu,\mathbf{\Gamma}\}$ where $\pi = \{\pi_1,\pi_2,\ldots,\pi_S\}$, $\mu = \{\mu_1,\mu_2,\ldots,\mu_S\}$ and $\mathbf{\Gamma} = \{\mathbf{\Gamma}_1,\mathbf{\Gamma}_2,\ldots,\mathbf{\Gamma}_S\}$.

Given these parameters and the model, the likelihood of an observation $y_{i,j}$ in a d-dimensional space can be written as

$$p\left(y_{i,j} \mid \theta, M\right) = \sum_{s=1}^{S} p\left(s_n = s \mid \pi\right) p\left(y_{i,j} \mid \mu_s, \mathbf{\Gamma}_s\right), \tag{3}$$

where each component is a Gaussian with $p(y_{i,j} \mid \mu_s, \mathbf{\Gamma}_s) = \mathcal{N}(y_{i,j}; \mu_s, \mathbf{\Gamma}_s)$ and $p(s_n = s \mid \pi)$ is a multinomial distribution representing the probability of the observation $y_{i,j}$ be associated with component s.

The prior over the parameters is given by

$$p(\theta \mid M) = p(\pi) \prod_s p(\mathbf{\Gamma}_s) p(\mu_s \mid \mathbf{\Gamma}_s) \tag{4}$$

where the weight prior is a symmetric Dirichlet $p(\pi) = \mathcal{D}(\pi; \lambda_0 \mathbf{I})$, the prior over each covariance matrix is a Wishart $p(\mathbf{\Gamma}_s) = \mathcal{W}(\Gamma; \alpha_0, \mathbf{B}_0)$ and the prior over the means given the covariance matrices is a multivariate normal $p(\mu_s \mid \mathbf{\Gamma}_s) = \mathcal{N}(\mu_s; \mathbf{m}_0, \beta_0 \mathbf{\Gamma}_s)$. The joint likelihood of the data, assuming the samples are i.i.d., can be computed as

$$p(\mathbf{y_i}, \mathbf{s} \mid \theta, M) = \prod_{n=1}^{N} p(s_n = s \mid \pi) p(y_{i,n} \mid \mu_s, \mathbf{\Gamma}_s) \tag{5}$$

where $\mathbf{y}_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,n}\}$ and $\mathbf{s} = \{s_1, s_2, \ldots, s_S\}$.

The variational approximation for the log marginal likelihood leads to the following free densities q:

- For the covariance matrices, $q(\mathbf{\Gamma}) = \prod_s q(\mathbf{\Gamma}_s)$ with $q(\mathbf{\Gamma}_s) = \mathcal{W}(\Gamma; \alpha_s, \mathbf{B}_s)$;
- For the means, $q(\mu \mid \mathbf{\Gamma}) = \prod_s q(\mu_s \mid \mathbf{\Gamma}_s)$ with $q(\mu_s \mid \mathbf{\Gamma}_s) = \mathcal{N}(\mathbf{x}; \mathbf{m}_0, \beta_s \mathbf{\Gamma}_s)$;
- For the mixing coefficients, $q(\pi) = \mathcal{D}(\pi; \lambda)$, where $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_s\}$;
- For the hidden variable $\mathbf{s}$, $q(\mathbf{s}) = \prod_s q(s_s)$.

Taking the functional derivatives of the free energy with respect to the free densities q produces the update rules of VBEM for the mixture of Gaussians case. In the VBE-Step the weights of the hidden variable are calculated and in the VBM-Step parameters and hyperparameters are updated. These rules are omitted here for brevity but can be found in [1].

Once parameters and the model were obtained, the predictive density for a particular patch $p(y' \mid \mathbf{y_i}, M)$ has a close-form solution of a mixture of Student-t distributions,

$$p(y' \mid \mathbf{y_i}, M) = \sum_{s=1}^{S} \bar{\pi}_s \mathcal{S}(\mathbf{y}'; \rho_s, \mathbf{\Lambda}_s, \omega_s), \tag{6}$$

with $\omega_s = \alpha_s + 1 - d$ degrees of freedom, where the means are$\rho_s = \mathbf{m}_s$ and the covariances are $\mathbf{\Lambda}_s = ((\beta_s + 1) / \beta_s \omega_s) \mathbf{B}_s$. The weights are computed based on the hyperparameters of the Dirichlet distribution with $\bar{\pi}_s = \lambda_s / \sum_{s'} \lambda_{s'}$.

2.3.1 Heuristic for Searching

VBEM allows direct model comparison by evaluating the free energy function of different models. In the case of mixtures of Gaussians this model can be a single component or a mixture with hundreds of components. Theoretically, there is no limit for the number of components and the search for the best model can be cumbersome. To cope with this problem, an heuristic based on birth and death of components is used. This heuristic seems to be appropriate for robotics problems since simpler models are evaluated before more complex ones, which decreases the computational complexity.

The heuristic used here has the same stopping and splitting criteria as in [2] for mixtures of factor analysers. The selection for splitting is based on the component with the smallest individual free energy. The search ends when all existent components were divided but none of those divisions resulted in free energy improvement.

2.4 Multi-Class Classification

As opposed to most classification problems where the input is a single feature-vector, in this approach the whole set of patches of an image is used. Each patch has equal contribution to the final classification decision and it is evaluated under the different models representing the places. The idea is to compute the log-likelihood of a set of image patches for every model learnt. The log-likelihood with the largest value is the final decision of the classification. Thus, the label of an image i is the label corresponding to the place model that maximises the expression:

$$M^* = \arg\max_M \sum_{j=1}^{m} \log p\left(y_{i,j} \mid M\right) \tag{7}$$

The computation for the log-likelihood in selecting the model that best explains the set of patches can be quickly computed. Also, it is possible to include more models, allowing sequential learning implementations. This is one of the demands for autonomous navigation as the robot visits new places, representations of them should be incorporated and correlated with the current knowledge.

3 Implementation

The whole framework was implemented in Matlab and tested with a Pioneer 2-AT. Images were obtained with 320x240 of resolution in 24-bits colour. Patches have a size of 5x5 pixels and were convolved by 4 Gabor wavelets resulting in a input space of 175 dimensions. Each image is thus a set of 3072 patches equally distributed, and of the same size. Isomap is then used to reduce the dimensionality to 5 and create the training data. The intrinsic dimensionality was also estimated by Isomap through the evaluation of the residual variance.

Learning is performed offline with labelled images from the above set. The training images were selected to give a multi-view perspective of the place. They were, however, taken from a particular position inside the place so as to verify how the algorithm generalises them to multiple positions. In the indoor experiment for example, if an office has a rectangular shape, 4 training images are taken close to the walls but the algorithm should still be able to recognise the place when observing it from the centre.

The variational Bayesian learning starts searching for the best model from a single-component model and follows the heuristic of birth and death as described before. The covariance obtained from the mixture of linear models is used to initialised the parameters α_0 and $\mathbf{B}_0$ of the Wishart distribution.

When testing the algorithm, the whole set of 3072 patches of an image is used. The process takes about 1 second per image in a Pentium M 1.7GHz which comprises Gabor convolutions, inference in the mixture of linear models and log-likelihood computation for each model learnt. Further implementations may use a subset of patches sampled from the original set to reduce the classification time. This, however, may decrease the accuracy which characterises an accuracy-time trade-off.

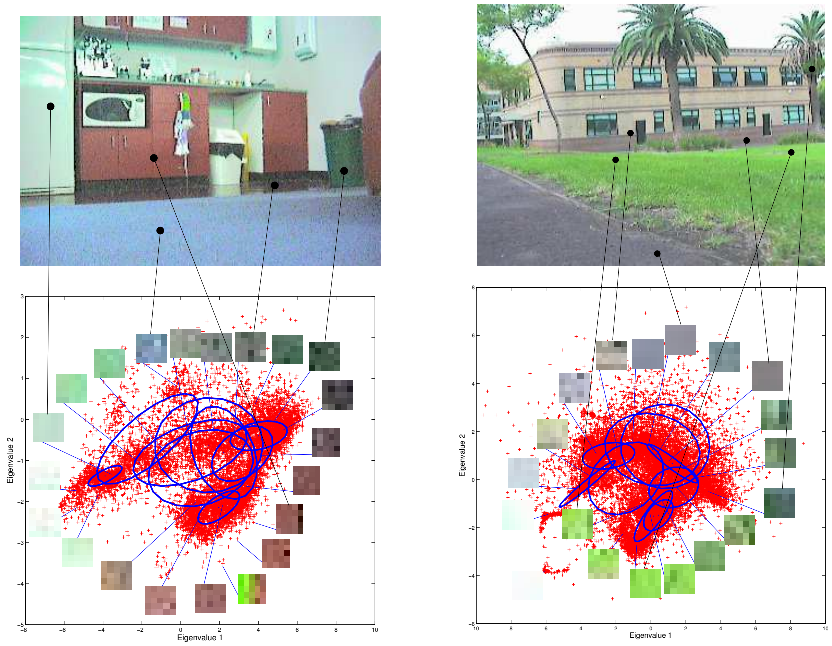
4 Experiments[1]

Two different experiments were performed to evaluate the algorithm in different conditions - indoor and outdoor environments. In both experiments, there were people walking by the places and sometimes occluding the robot's view. In the outdoor experiment, there were also cars and bicycles moving in some places which make the problem more difficult since the environment is dynamic.

4.1 Indoor dataset

The indoor dataset consists of 55 training images of 9 different places - each place has 5 to 9 training images. The test set has 1579 sequence images obtained by the robot when navigating inside the lab. The classes are {kitchen, seminar room, student cubicle 1, corridor 1, student cubicle 2, corridor 2, corridor

[1] Videos with the experiments are available at: http://www.acfr.usyd.edu.au/people/postgrads/ftozeto



a) Kitchen model. b) ACFR Park model.

Figure 1: Generative models learnt for kitchen (indoor) and ACFR Park (outdoor). Points are plotted on the direction of the two largest eigenvalues of the essential features (the total dimensionality of this space is 5). Ellipses correspond to the covariance matrices of the components learnt with VBEM. The association between the patches and their location in the real scene is also indicated.

3, research office and professor office}. The generative model for the kitchen is depicted in Figure 1a. It shows the covariance matrices learnt through VBEM from the essential features. The correlation between the patches and their real position is also indicated.

Table 1a shows the precision and recall results for this 9-class problem. The classification was accurate considering that in many occasions the robot was very close to walls or other objects, in situations where even for a human, the classification would be difficult. Also, some places are very alike, for example, the corridors and the student cubicles. These places can be distinguished from few objects such as paintings in the case of the corridors and books, computers on the desks of the student cubicles. However, both objects were not observed by the robot since they are in a higher position, away from the robot view. Figure 2 shows the training images for the kitchen model. Even not having a training image with a wider view of the kitchen, the classifier was able to recognise the kitchen from the image of Figure 1a. This generalisation is one of the main properties of the algorithm.

4.2 Outdoor dataset

The outdoor dataset consists of 57 training images of 11 different places at University of Sydney, with each place having 3 to 8 training images. The test set has 3820 images obtained from a half-kilometre journey around the University of Sydney. The classes are {ACFR front, ACFR park, Eng. Building, Eng. Road, Eng. Carpark, Mech. Building, ACFR carpark, ACFR road, garage(indoor) and office(indoor)}. The generative model for the class ACFR park is shown in Figure 1b. Also annotated is the correlation between patches and their real location.

Table 1b presents the precision and recall results which are in general better than the indoor dataset. The most difficult problem of the outdoor dataset was to distinguish between the two carparks. When the robot was very close to a car, it was not able to have a more general view of the place which resulted in

Figure 2: The training images used for the kitchen model. The model learnt from them was able to recognise wider views of the place such as the image in Figure 1a.

Place Name	Precision	Recall
Kitchen	81.82	76.60
Seminar Room	68.07	81.82
Student Cubicle 1	52.86	48.05
Corridor 1	74.96	84.09
Student Cubicle 2	62.96	46.12
Corridor 2	40.00	62.50
Corridor 3	100	15.13
Researcher Office	73.01	90.16
Professor Office	62.96	71.83

a) Indoor results.

Place Name	Precision	Recall
ACFR Front	71.04	87.44
ACFR Park	83.19	71.05
Eng. Building	65.40	77.31
Eng. Road	91.43	45.43
Eng. Carpark	26.17	87.10
Mech. Building	64.47	17.63
Mech. Corridor	90.68	65.54
ACFR Carpark	55.24	76.40
ACFR Road	87.00	74.34
Garage	23.94	87.18
Office	99.47	61.84

b) Outdoor results.

Table 1: Precision and recall results for the indoor and outdoor datasets.

classifying the image as the other carpark. Also, “Mech. Building” and “Eng. Building” are physically in the same building and the limits of where one starts and the other finish are not very clear.

5 Conclusions

The framework here proposed has three main contributions: it shows that mixture of linear models can be used as a tool for nonlinear regression problems with noise such as Isomap mappings; it demonstrates how variational Bayesian learning with a free-energy heuristic can choose the right number of components of a mixture of Gaussians; it shows how the log-likelihood can be applied to multi-class problems where classification is given from a set of samples rather than from a single point.

Patches of images and images themselves are treated here as independent and identically distributed. In the case of patches, further implementations can include the positions of each patch as additional dimensions in the feature-vector. Also, spatial relations among them can be included in more sophisticated relational statistical models. This, however, should preserve the main benefits of the model such as learning from few images and real-time classification.

Most of the wrong classifications took place when the robot was close to a wall or object that occluded a wider view of the scene. This problem could be avoided if a topological map of the environment were encoded in a hidden Markov model to constrain the search to fewer places. In future works, algorithms for learning this topological map will be investigated as well as how to integrate them in the framework.

Acknowledgements

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales (NSW) State Government.

References

[1] H. Attias. A variational bayesian framework for graphical models. In *Proceedings of Neural Information Processing Systems 12*, Cambridge, MA, USA, 2000. MIT Press.

[2] M. J. Beal. *Variational Algorithms for Approximate Bayesian inference*. PhD thesis, The Gatsby Computational Neuroscience Unit, University College London, May 2003.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, New York, 1991.

[4] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman and Hall, 1994.

[5] A. Davison and D. Murray. Simultaneous localisation and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:865–880, 2002.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–37, 1977.

[7] M. G. Dissanayake, P. Newman, S. Clark, and H. F. Durrant-Whyte. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17:229–241, 2001.

[8] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, New York, second edition, 2001.

[9] Z. Ghahramani and G. E. Hinton. The EM algorithm for mixtures of factor analyzers. Technical report, Department of Computer Science, University of Toronto CRG-TR-96-1, 1996.

[10] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

[11] J. Kosecka and F. Li. Vision based topological Markov localization. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2004.

[12] D. J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.

[13] T. P. Minka. Using lower bounds to approximate integrals. *Informal notes available at http://www.stat.cmu.edu/minka/papers/learning.html*, 2001.

[14] T. Mitchell and F. Labrosse. Visual homing: a purely appearance-based approach. In *In Proc. of Towards Autonomous Robotic Systems*, Colchester, UK, 2004.

[15] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, USA, 1982.

[16] L. K. Saul and S. T. Roweis. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4(Jun):119–155, 2003.

[17] S. Se, D. Lowe, and J. Little. Global localization using distintive visual features. In *Proc. of International Conference on Intelligent Robots and Systems*, pages 226–231, Lausanne, Switzerland, 2002.

[18] R. Sim and G. Dudek. Learning environmental features for pose estimation. *Image and Vision Computing*, 19(11):733–739, 2001.

[19] J. Tenenbaum, V. DeSilva, and J. C. Langford. A global geometric framework for nonlinear dimensionality redution. *Science*, 290:2319–2323, 2000.

[20] A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context-based vision system for place and object recognition. In *Proceedings of International Conference of Computer Vision*, 2003.

[21] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of ICRA 2000*, volume 2, pages 1023–1029, April 2000.

[22] J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization for mobile robots using an image retrieval system based on invariant features. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.