A Natural Feature Representation for Unstructured Environments

Fabio Tozeto Ramos, *Member, IEEE*, Suresh Kumar, Ben Upcroft, and Hugh Durrant-Whyte

***Abstract*—This paper addresses the long-standing problem of feature representation in the natural world for autonomous navigation systems. The proposed representation combines Isomap, which is a nonlinear manifold learning algorithm, with expectation maximization, which is a statistical learning scheme. The representation is computed off-line and results in a compact, nonlinear, non-Gaussian sensor likelihood model. This model can be easily integrated into estimation algorithms for navigation and tracking. The compactness of the model makes it especially attractive for deployment in decentralized sensor networks. Real sensory data from unstructured terrestrial and underwater environments are used to demonstrate the versatility of the computed likelihood model. The experimental results show that this approach can provide consistent models of natural environments to facilitate complex visual tracking and data-association problems.**

***Index Terms*—Dimensionality reduction, feature extraction, field robotics, Isomap, probabilistic representation.**

I. Introduction

THIS PAPER addresses the long-standing issue of natural feature extraction and representation in unstructured worlds for autonomous navigation. A unified framework based on information theory and probabilistic learning is proposed for terrestrial, underwater, and aerial applications. The stochasticity of the computed natural feature models enables easy integration into conventional estimation algorithms for navigation and tracking. This results in rich characterizations of unstructured environments in terms of color, texture, and other sensory properties to facilitate the design of robust autonomous navigation systems.

The feature-extraction algorithm used is based on concepts of information theory to extract novel features from the sensory space. Novelty is informally defined to correspond to features with a low probability of occurrence and, thus, high information content [1]. The objective of effective feature selection is to identify regions in the image with unique sensory properties. The frequency of occurrence of the properties can be quantified through property histograms and the feature-selection problem is addressed by working with the least likely features.

The integration of natural features described by rich color, texture, reflectivity, and other sensory properties with nonlinear filtering schemes requires a probabilistic model. This paper combines the concepts of nonlinear dimensionality reduction (NLDR) with statistical learning algorithms to compute probabilistic representations of natural features.

Probabilistic modeling of visual features has been reported in the modeling of texture using Markov random fields with the maximum entropy principle as described in [2]. Lee *et al.* [3] present a generative model based on independent components analysis that provides a linear and non-Gaussian framework for feature representation. In the work of Karklin *et al.* [4], a hierarchical probabilistic framework is used for detection of higher order statistical structure in natural imagery. A key limitation of these models is that they do not necessarily preserve the inherent similarities and distinctions in the original visual data. This minimizes their utility in classical estimation and data-association tasks.

More recently, a technique to learn both an invariant mapping and the mapping function with convolutional neural networks was proposed in [5]. The learning procedure relies on complex nonconvex optimization methods to compute the embeddings. However, a main limitation is the lack of a probabilistic representation which makes this technique unsuitable for data-fusion tasks.

The main contribution of this paper is the combination and a real-time implementation of feature extraction and selection, NLDR, and statistical learning techniques for large-scale perception problems in unstructured environments. The probabilistic representation of natural features may be integrated within existing nonlinear filtering algorithms [6].

The proposed framework for feature extraction and representation includes an off-line learning phase where a statistical natural feature model is computed from environment-specific training data. Isomap, which is an NLDR algorithm, is used in conjunction with expectation–maximization (EM) to compute a generative statistical model in the off-line phase of the proposed framework. The online implementation of the scheme performs natural feature extraction, followed by probabilistic model inference to classify the selected features with respect to the generated model. The complete framework is shown in Fig. 1.

The natural feature extraction algorithm used in the online phase of the implementation is detailed in Section II. Section III describes the Isomap and EM algorithms that are used in the off-line computation of a generative natural feature model. Section IV describes the model inference methodology to

Manuscript received March 1, 2007; revised April 15, 2008. First published December 9, 2008; current version published December 30, 2008. This paper was recommended for publication by Associate Editor G. Sukhatme and Editor L. Parker upon evaluation of the reviewers' comments. This work was supported in part by Australian Research Council (ARC) and in part by the New South Wales State Government under the ARC Centre of Excellence Programme.

F. T. Ramos, S. Kumar, and H. Durrant-Whyte are with the ARC Centre of Excellence for Autonomous Systems, Australian Centre for Field Robotics, University of Sydney, Sydney, NSW 2006, Australia (e-mail: f.ramos@acfr.usyd.edu.au; suresh@acfr.usyd.edu.au; b.upcroft@acfr.usyd.edu.au; hugh@acfr.usyd.edu.au).

B. Upcroft is with the University of Queensland, Brisbane, Qld 4072, Australia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2008.2007933

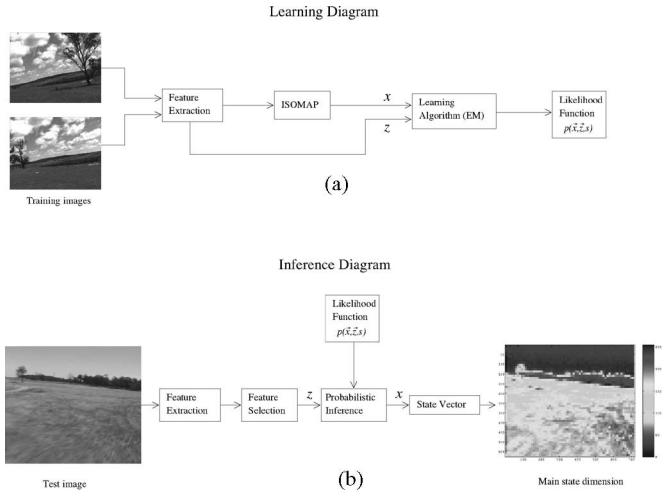


Fig. 1. Block diagram depicting the off-line model generation and online model inference phases in the proposed feature-extraction and representation framework.

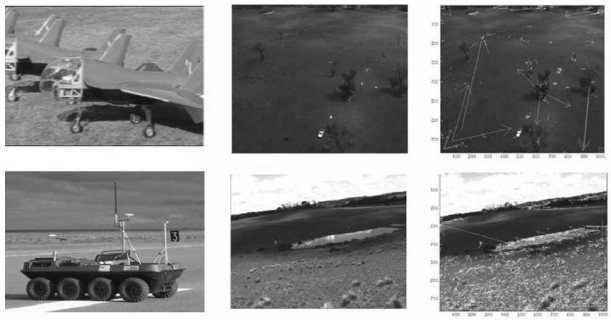


Fig. 2. Autonomous vehicles in unstructured environments. Conventional feature extractors such as SIFT extract too many features in natural imagery, as shown in the pictures on the right. The tails of the white lines depict the SIFT-extracted features. These features are not useful in autonomous mapping applications, where the main objective is to extract unique physical features, such as trees, bush, and lakes that exist in the environment. This paper uses a frequentist approach to feature extraction to detect unique natural features.

enable online classification of natural features. Real applications involving natural imagery acquired by autonomous ground and underwater vehicles are presented in Section V.

II. Sensor Data Preprocessing—Natural Feature Extraction

The extraction of stable natural features in unstructured dynamic worlds has been a persistent research question at the intersection of robotics, computer vision, and machine learning. Visual feature extraction is inherently complex due to the variability in sensor data on account of factors such as observer viewpoint, ambient illumination conditions, and occlusions. Interest point feature detectors [7] and scale-invariant feature transforms (SIFT) [8] are quite successful at extracting reliable features in structured environments, but fail to select a compact feature set in natural imagery.

Fig. 2 illustrates the complexity of natural imagery acquired by unmanned terrestrial and aerial vehicles. Interest point feature detection schemes are not guaranteed to extract a compact set of physical features such as trees, shrubs, bush, or lakes within these images. These methods may, however, have significant potential in addressing other pressing issues in robotics such as loop closure. The utility of these methods in autonomous mapping of physical natural features seems limited.

Recent study [9] suggests that robust segmentation of unstructured imagery could be achieved by a linear combination of relevant visual cues such as color, texture, color opponency, motion, and spatial proximity. In particular, it has been shown that a rigorous treatment of texture in natural imagery is required to select meaningful features.

The feature selection challenge is quite distinct from the detection of boundaries in natural images that is addressed by [9]. A frequentist approach is adopted in this paper to select regions of natural imagery with the most unique color and texture. The information content of image regions is assumed to be inversely proportional to the frequency of occurrence of the region properties in the image. If p is the probability of occurrence of an image property, the Shannon information content is computed as $\log(\frac{1}{p})$.

A. Information Content in Color Space

The information content of image regions in color space I_C is computed based on the *joint* probability of occurrence of the red, green, and blue intensities at each pixel within the image. Multidimensional histograms are used to compute the joint probabilities at the pixel locations. The Shannon information content at a pixel $\log(\frac{1}{p})$ is computed from the joint statistics.

B. Information Content in Texture Space

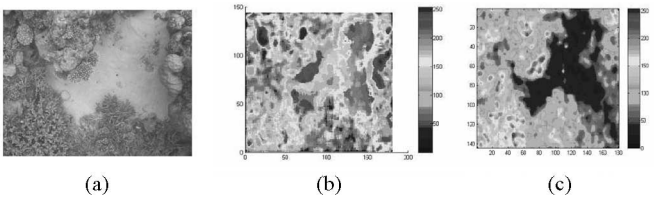
Spectral decomposition through Gabor filtering is used in this paper to quantify texture as it provides a good approximation of natural processes in the primary visual cortex [10]. The input image is convolved with a precomputed filter-bank at multiple scales and orientations. A histogram of the sum of the amplitudes for each of the scales and orientations is computed to represent the probability of texture occurrence. The pixel-wise Shannon information content in terms of the texture I_T can then be calculated.

C. Feature Extraction

A simple linear combination of the color and texture space information contents is used to compute the net information content of features with respect to both visual cues. In general, $I = \alpha \times I_C + \beta \times I_T$, where $\alpha + \beta = 1$. In a practical implementation, equally weighted linear combinations have been found to perform reliably. The net information content is appropriately thresholded to yield the most informative features with respect to color and texture.

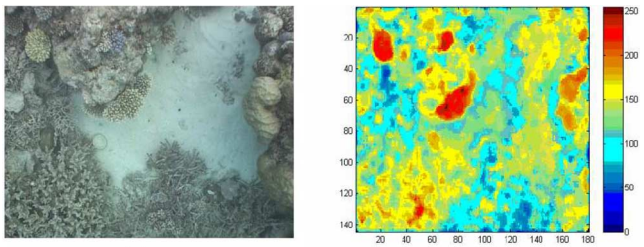
D. Example

The entire feature extraction process is illustrated through a simple example image from an underwater sequence (Fig. 3). A 2-D color histogram of the raw red and green intensities in the



(a) (b) (c)

Fig. 3. (a) Sample image from underwater data sequence. (b) Information content in color and (c) texture spaces.



(a) (b)

Fig. 4. (a) Original image and (b) mutually informative feature set. The features that are maximally informative with respect to both the color and texture cues are highlighted in red. They exhibit significant contrast with respect to their local neighborhoods and are expected to persist in the environment.

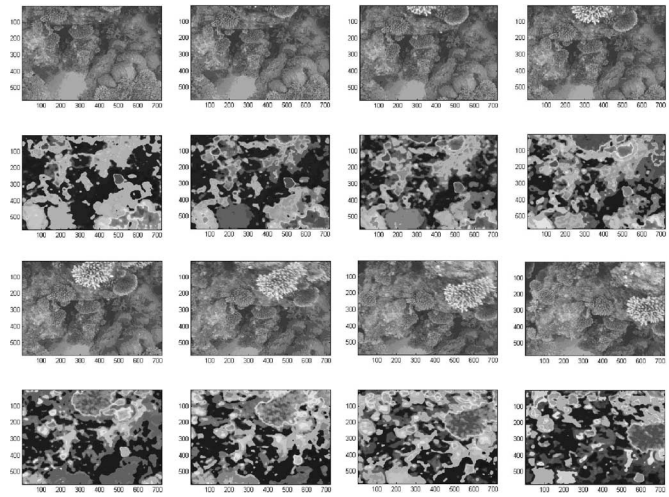


Fig. 5. Stability of the feature selection scheme over 100 images from the reef data. A sample of eight frames is shown to demonstrate the consistency of the promoted features. Features with the highest contrast displayed in red physically correspond to regions with distinctive hue and texture. Note that only features with fine texture are extracted. Although the blue sandy area in the lower left of the image is quite distinct from the other areas in the image, this is not deemed to be a distinctive feature through lack of texture, as in most natural scenes, textureless patches commonly correspond to large expansive regions that are not useful for characterizing unstructured imagery.

image was computed and the information content of each pixel was computed as $\log(\frac{1}{p})$ [see Fig. 3(b)]. In order to promote corals with rich texture content into the feature space, the raw image was convolved with Gabor wavelets at four scales and six orientations. The histogram of the resultant amplitudes of the response I_T is shown in Fig. 3(c).

The maximal mutually informative (MMI) subspace is computed from an equally weighted linear combination of the information contents in color and texture space (Fig. 4, right). The maximally relevant features (color-coded red) exhibit significant contrast with respect to their neighborhoods and permit convenient feature selection through the choice of an appropriate mutual information threshold.

The stability of the feature selection algorithm was tested on a sequence of 100 images from an existing underwater data sequence and the MMI subspace was used to compute the feature space as described earlier, except for the fact that the color space description used a hue histogram as opposed to any of the raw color components. This choice was preferred in order to enable robust feature selection over varying illumination states. The results are shown in Fig. 5. The promoted features are consistent through the tracked sequence and are clearly perceived as regions with distinctive hue and texture.

III. Model Generation

Information-theoretic concepts can be used to extract features with unique properties within the sensory space. However, each such feature is potentially set in a very high-dimensional space that is not readily amenable to simple interpretation and reasoning tasks. The development of compact and useful representations of natural features in unstructured dynamic worlds is critical to the development of robust perception systems.

Eigenvector methods such as principal component analysis (PCA) and its numerous variants provide theoretically optimal natural feature representations from a data compression standpoint. They are, however, unable to provide neighborhood preserving representations that are crucial in classification tasks. This limitation has motivated the development of various NLDR methodologies such as kernel PCA [11], Isomap [12], Laplacian eigenmaps [13], and locally linear embedding [14]. Most NLDR techniques presume that the data lies on or in the vicinity of a low-dimensional manifold and attempt to map the high-dimensional data into a single low-dimensional, global coordinate system. The Isomap algorithm is adopted in this paper to provide a low-dimensional description of high-dimensional features, primarily because it estimates the intrinsic dimensionality of the manifold in addition to the underlying states.

A. *Theoretical Aspects of the Isomap Method*

The Isomap method [12] formulates NLDR as the problem of finding a Euclidean feature space embedding a set of observations that attempts to explicitly preserve their intrinsic metric structure; the metric structure is quantified as the geodesic distances between the points along the manifold.

The Isomap method starts out assuming that the data $\mathbf{X}$ lies on an unknown manifold embedded in the high-dimensional observation space and attempts to reconstruct an implicit mapping $f : \mathbf{X} \rightarrow \mathbf{Y}$ that transforms the data to a low-dimensional Euclidean feature space $\mathbf{Y}$ that optimally preserves the distances between the observations as measured along geodesic paths on the manifold. Significant steps in the Isomap algorithm are summarized next.

B. Nearest-Neighbor Computation

Neighboring points on the manifold are determined based on the input space distances $d_X(i,j)$ between pairs of points $i, j \in \mathbf{X}$. Each input point is connected to adjacent points based either on the K-nearest neighbors or all points within a fixed distance ϵ from the point under consideration. The neighborhood relations are expressed as a weighted graph G over the data points with edges of weight $d_X(i,j)$ between neighboring points.

C. Computation of Geodesic Distances

The length of a path in G is defined as the sum of the link weights along the path. The shortest path lengths d_G^{ij} between two nodes i and j in the graph G are computed through the Floyd's algorithm [15] that generally scales as $O(N^3)$ or the Dijkstra algorithm [16] that scales as $O(N^2 \log(N))$, where N is the number of data points.

D. Graph Embedding Through Classical Multidimensional Scaling

Classical multidimensional scaling (MDS) is a technique that computes a low-dimensional configuration of high-dimensional input that best preserves pairwise distances between input points [17]. The key innovation in the Isomap method is to apply MDS to the pairwise geodesic distances on the manifold so as to compute a low-dimensional and neighborhood-preserving configuration. The coordinate vectors $y_i \in \mathbf{Y}$ are chosen to minimize the cost function $E = \|\tau(d_G) - \tau(d_Y)\|_{L^2}$, where d_Y is the matrix of output space distances, and the norm is the matrix L^2 norm $\sqrt{\sum_{i,j}(\tau(d_G) - \tau(d_Y))_{ij}^2}$, $\tau = \frac{1}{2}HSH$ is an operator that converts distances into inner products, $H_{ij} = \delta_{ij} - (1/N)$ is the centering matrix, and $S_{ij} = (d_G^{ij})^2$ is the matrix of squared geodesic distances. The global minimum of the cost function is computed by setting the output space coordinates y_i to the top k eigenvectors of $\tau(d_G)$. Introducing a unit row vector $P = [111\ldots\ldots1]$ with N columns and an $N \times N$ translation matrix Q with constant entries $Q_{ij} = (\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} S_{ij})$, the $N \times N$ matrix $\tau(d_G)$ is expressed as

$$\tau(d_G) = \frac{1}{2}\left(S - \left(\Sigma_{j=1}^{N} S_{ij}\right)\frac{P}{N} - P^T\frac{\left(\Sigma_{i=1}^{N} S_{ij}\right)}{N} + \frac{Q}{N^2}\right) \quad (1)$$

E. Landmark Isomap Method

The landmark Isomap method was designed to overcome the significant computational burden involved in the Dijkstra algorithm and subsequent eigensolution of a full symmetric matrix incurred in global Isomap. The theoretical description of landmark Isomap presented here closely adheres to the implementation provided by the original authors [18]. In this method, the elements of a small random subset n_L of the total number of data points N are designated as landmarks. The particular landmarks chosen are not critical but an adequate distribution of landmarks over the manifold is important to ensure the general shape of the manifold is preserved. It was found that a uniform sampling of at least 100 landmarks was required for an accurate embedding. The distance matrix d_G^{ij} now corresponds only to the distances between the landmarks and is of size $n_L \times n_L$ (as opposed to $N \times N$ in global Isomap). The cost of the Dijkstra algorithm correspondingly reduces to $O(n_L \times N \times \log(N))$.

F. Computation of Low-Dimensional Embedding

MDS is applied to the now smaller distance matrix to compute a low-dimensional embedding of the landmarks. The low-dimensional embedding is obtained by computing the eigenvectors of the inner product matrix $B_n = -H_n \Delta H_n/2$, $H_n = \delta_{ij} - (1/n_L)$ and Δ is a matrix of squared distances between the landmarks. Introducing a unit row vector with n_L columns $P^* = [11\ldots1]$ and an $n_L \times n_L$ translation matrix Q^* with constant entries $Q_{ij}^* = (\Sigma_{i=1}^{n_L}\Sigma_{j=1}^{n_L}\Delta_{ij})$, the $n_L \times n_L$ matrix B_n can be expressed as

$$B_n = \frac{1}{2}\left(\Delta - \left(\Sigma_{j=1}^{n_L}\Delta_{ij}\right)\frac{P^*}{n_L} - P^T\frac{\left(\Sigma_{i=1}^{n_L}\Delta_{ij}\right)}{n_L} + \frac{Q^*}{n_L \times n_L}\right). \quad (2)$$

This low-dimensional coordinate of the landmark in l-dimensional space is designated as $L = (\sqrt{(}\lambda_1)v_1^T, \sqrt{(}\lambda_2)v_2^T, \ldots.., \sqrt{(}\lambda_l)v_l^T)^T$, where λ_i are the eigenvalues and v designate the eigenvectors.

G. Nonlandmark Extensions

If $\triangle_n$ is the column-wise mean of Δ, a nonlandmark point $\mathbf{y}$ can be embedded into the l-dimensional space as

$$\mathbf{y} = (1/2)L^{\sharp}(\triangle_n - \triangle_y) \quad (3)$$

where $L^{\sharp}$ is the pseudoinverse transpose of L and $\triangle_y$ is the vector of squared distances between the candidate point and the landmarks. Thus, the remaining $N - n_L$ nonlandmark points can be embedded into the l-dimensional space.

The embedding computed by the landmark Isomap method is consistent with that computed by classical MDS at the landmark locations and is an estimate at the nonlandmark locations. If the distance matrix between all the points and the landmarks can be represented exactly by a Euclidean configuration in R^l, and the landmarks are chosen such that their affine span in that configuration is l-dimensional, the estimate at the nonlandmark locations is accurate up to a rotation and translation.

H. Generative Model

The Isomap algorithm and, indeed, most NLDR algorithms are inherently deterministic algorithms. They do not provide a measure of the uncertainty of the underlying states of high-dimensional observations. The integration of the low-dimensional states computed by Isomap into a nonlinear filtering framework requires the definition of a generative likelihood model $p(\mathbf{z}|\mathbf{x})$, where $\mathbf{z}$ and $\mathbf{x}$ are the observation and state spaces, respectively. This likelihood model encapsulates the uncertainties inherent in the inference of a low-dimensional state from noisy high-dimensional observations. The incorporation of natural feature states within a non-Gaussian and nonlinear

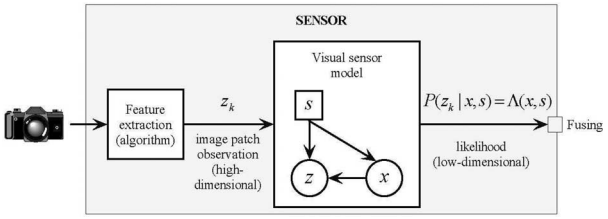


Fig. 6. Graphical model for computation of parametric models from NLDR algorithms. An arrow directed into a node depicts a dependency on the originating node. The discrete hidden variable s represents a specific neighborhood on the manifold. The feature extraction and Isomap algorithms supply the input data for learning.

filter is expected to significantly enhance data association as the low-dimensional appearance states and kinematic variables are complementary.

Methods from supervised learning can be used to derive compact mappings that generalize over large portions of the input and embedding space. The input–output pairs of Isomap can serve as training data for an invertible function approximation in order to learn a parametric mapping between the two spaces.

Given the results of Isomap, a probabilistic model of the joint distribution $p(\mathbf{z}, \mathbf{x}, s)$ can be learned through the EM algorithm [19]. The joint distribution can be used to map inputs to outputs and vice versa by computing the expected values $E[\mathbf{z}|\mathbf{x}]$ and $E[\mathbf{x}|\mathbf{z}]$. The joint distribution is similar to a mixture of factor analyzers that is commonly used in the machine learning community to perform simultaneous clustering and local dimensionality reduction [20]. The only differences are that the low-dimensional variable $\mathbf{x}$ is observed and not hidden, and the Gaussian distributions $p(\mathbf{x}|s)$ have nonzero mean vectors ν_s and full covariance matrices Σ_s. Learning when the variable $\mathbf{x}$ is observed seems to discover a solution of better quality than in the opposite situation [21]. The graphical model in Fig. 6 depicts the assumed dependencies. The discrete hidden variable s introduced in the model physically represents a specific neighborhood 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. Although this model is used here for learning Isomap mappings, its range of applications is quite wide. Basically, all linear or nonlinear regression problems under noise can be addressed using this scheme. The only constraint is that the number of samples has to be large enough for a reasonable likelihood approximation.

The complete three-step generative model can now be summaried based on the assumed dependencies (4)–(6). The joint probability distribution of all the random variables in the graphical model is expressed as

$$p(\mathbf{z}, \mathbf{x}, s) = p(\mathbf{z} \mid \mathbf{x}, s) p(\mathbf{x} \mid s) p(s) \tag{4}$$

where the dependencies are given by

$$p(\mathbf{z} \mid \mathbf{x}, s) = \frac{1}{(2\pi)^{D/2} |\Psi_s|^{1/2}} \times \exp\left\{ -\frac{1}{2} [\mathbf{z} - \Lambda_s \mathbf{x} - \mu_s]^T \Psi_s^{-1} [\mathbf{z} - \Lambda_s \mathbf{x} - \mu_s] \right\} \tag{5}$$

$$p(\mathbf{x} \mid s) = \frac{1}{(2\pi)^{d/2} |\Sigma_s|^{1/2}} \times \exp\left\{ -\frac{1}{2} [\mathbf{x} - \nu_s]^T \Sigma_s^{-1} [\mathbf{x} - \nu_s] \right\}. \tag{6}$$

I. Parameter Estimation

In this model, the set of parameters $\theta = \{p(s), \nu_{\mathbf{s}}, \mu_{\mathbf{s}}, \Sigma_s, \Psi_s, \Lambda_s\}$ that need to be estimated from the observed high- and low-dimensional spaces are the prior probabilities $p(s)$ that follow a multinomial distribution, the mean vectors $\nu_{\mathbf{s}}$ and $\mu_{\mathbf{s}}$, the full covariance matrix Σ_s, the diagonal covariance matrix Ψ_s, and the loading matrices Λ_s. The EM algorithm performs parameter estimation by maximizing the log-likelihood of the data given the model and the set of parameters. The observable parameters in the graphical model are denoted as $\{\mathbf{z}_n, \mathbf{x}_n\}_{n=1}^N$, where N is the number of samples. EM iteratively maximizes the log-likelihood of the observations w.r.t. θ

$$\mathcal{L} = \sum_{n=1}^{N} \log \sum_{i=1}^{M} p(\mathbf{z}_n, \mathbf{x}_n, s_i \mid \theta) \tag{7}$$

where M is the number of mixtures considered in the model. Since direct maximization over the aforesaid expression is hard to be calculated analytically, an auxiliary distribution $q(s_i)$ over the hidden variable is introduced as

$$\mathcal{L} = \sum_{n=1}^{N} \log \sum_{i=1}^{M} q(s_i) \frac{p(\mathbf{z}_n, \mathbf{x}_n, s_i \mid \theta)}{q(s_i)}. \tag{8}$$

Maximizing $\mathcal{L}$ with respect to $q(s)$ is equivalent to minimizing the Kullback–Leibler divergence between the free distribution $q(s)$ and the posterior probability $p(s \mid \mathbf{z}_n, \mathbf{x}_n, \theta)$. This is done by setting $q(s) = p(s \mid \mathbf{z}_n, \mathbf{x}_n, \theta)$.

Thus, for each iteration, EM alternates between the expectation step, where the posterior probability of s given the observations is computed through

$$p(s \mid \mathbf{z}_n, \mathbf{x}_n) = \frac{p(\mathbf{z}_n \mid \mathbf{x}_n, s) p(\mathbf{x}_n \mid s) p(s)}{\sum_{s'} p(\mathbf{z}_n \mid \mathbf{x}_n, s') p(\mathbf{x}_n \mid s') p(s')} \tag{9}$$

and the maximization step, where this posterior is used to reestimate the parameters. The algorithm continues until the difference between the log-likelihood of two iterations is smaller than a given threshold. The update rules for the maximization step are presented next [14].

Defining $\gamma_{sn} = p(s \mid \mathbf{z}_n, \mathbf{x}_n)$ and $\omega_{sn} = (\gamma_{sn} / \sum_{n'})$, the updates are

$$\nu_s \leftarrow \sum_n \omega_{sn} \mathbf{x}_n \tag{10}$$

$$\Sigma_s \leftarrow \sum_n \omega_{sn} \left[\mathbf{x}_n - \nu_s\right] \left[\mathbf{x}_n - \nu_s\right]^T \tag{11}$$

$$\Lambda_s \leftarrow \sum_n \omega_{sn} \mathbf{z}_n \left(\mathbf{x}_n - \nu_s\right)^T \Sigma_s^{-1} \tag{12}$$

$$\mu_s \leftarrow \sum_n \omega_{sn} \left[\mathbf{z}_n - \Lambda_s \mathbf{x}_n\right] \tag{13}$$

$$\Psi_s \leftarrow \sum_n \omega_{sn} \left[\mathbf{z}_n - \Lambda_s \mathbf{x}_n - \mu_s\right] \left[\mathbf{z}_n - \Lambda_s \mathbf{x}_n - \mu_s\right]^T \tag{14}$$

$$p\left(s\right) \leftarrow \frac{\sum_n \gamma_{sn}}{\sum_{s'n'} \gamma_{s'n'}}. \tag{15}$$

Once the parameter estimation is completed, the joint distribution (4) is fully characterized, and a likelihood model $p(\mathbf{x}, s|\mathbf{z})$ can be computed by making an observation in the high-dimensional space. The likelihood is expressed as a Gaussian mixture model, and can thus be easily integrated within a nonlinear, non-Gaussian filtering scheme [6].

J. Choice of a Suitable Number of Mixture Components

The choice of an optimal number of mixture components is crucial to the accurate computation of a likelihood model. An excessive number of mixture components increases online computation costs and overfits the training data. An inadequate number of mixture components fails to model the training data accurately. A rigorous model selection approach has been described in the framework of variational inference [22].

A simpler approach is adopted in this paper that adequately addresses overfitting concerns. The number of mixture components is chosen so that the resulting probabilities $p(s)$ are always greater than a predefined threshold. If the computed probabilities $p(s)$ are all significantly greater than the threshold, the model is refined. If any of the mixture components results in probabilities smaller than the threshold, the model is coarsened through a reduction in the number of components.

IV. Model Inference

The most common inference in this model is the evaluation of the posterior $p(\mathbf{x}, s|\mathbf{z} = \mathbf{z}_i)$. This posterior represents the probability of the low-dimension vector given an observation in the high-dimensional space. It can be used to represent the beliefs of an observation and can be further passed to a filtering scheme. From the joint distribution, $p(\mathbf{x}, s|\mathbf{z} = \mathbf{z}_i)$ is calculated as

$$p(\mathbf{x}, s|\mathbf{z} = \mathbf{z}_i) = \frac{p(\mathbf{z} = \mathbf{z}_i \mid \mathbf{x}, s) p(\mathbf{x} \mid s) P\left(s\right)}{\sum_{s'} \int p(\mathbf{z} = \mathbf{z}_i \mid \mathbf{x}, s') p(\mathbf{x} \mid s') p\left(s'\right) d\mathbf{x}}. \tag{16}$$

A general method to solve this equation is to transform it into canonical forms and compute the posterior in terms of the canonical characteristics as described in [23]. However, inferences in a simple structure as the one considered in this paper can be computed more efficiently by deriving the specific formulas in the moment form.

The joint distribution $p(\mathbf{z}, \mathbf{x}|s)$ can be recovered by multiplying $p(\mathbf{z} \mid \mathbf{x}, s) p(\mathbf{x} \mid s)$

$$p\left(\begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mu_s + \Lambda_s \nu_s\\ \nu_s\end{bmatrix}, \begin{bmatrix}\Psi_s + \Lambda_s \Sigma_s^T \Lambda_s^T & \Lambda_s \Sigma_s\\ \Sigma_s^T \Lambda_s^T & \Sigma_s\end{bmatrix}\right) p(s) \tag{17}$$

where T denotes transpose. It follows from this joint distribution that the conditional $\mathbf{x}$ given an observation $\mathbf{z} = \mathbf{z}_i$ is a multidimensional Gaussian with mean

$$\mu_{\mathbf{x}|s,\mathbf{z}=\mathbf{z}_i} = E\left[\mathbf{x} \mid s, \mathbf{z} = \mathbf{z}_i\right] = \nu_s + \left(\Sigma_s^{-1} + \Lambda_s^T \Psi_s^{-1} \Lambda_s\right)^{-1} \times \left(\Lambda_s^T \Psi_s^{-1}\right) \left(\mathbf{z}_i - \mu_s - \Lambda_s \nu_s\right) \tag{18}$$

and covariance

$$\Sigma_{\mathbf{x}|s,\mathbf{z}=\mathbf{z}_i} = E\left[\mathbf{x}\mathbf{x}^T \mid s, \mathbf{z} = \mathbf{z}_i\right] = \Sigma_s - \Sigma_s^T \Lambda_s^T \left(\Psi_s + \Lambda_s \Sigma_s^T \Lambda_s^T\right)^{-1} \Lambda_s \Sigma_s. \tag{19}$$

Since Ψ_s is a diagonal matrix and Σ_s is assumed to be nonsingular, the inverse of the aforesaid expression can be efficiently computed by using the matrix inversion lemma: $(\Psi_s + \Lambda_s \Sigma_s^T \Lambda_s^T)^{-1} = \Psi_s^{-1} - \Psi_s^{-1} \Lambda_s \left(\Sigma_s^{-1} + \Lambda_s^T \Psi_s^{-1} \Lambda_s\right)^{-1} \Lambda_s^T \Psi_s^{-1}$.

Weights can be computed by marginalizing the joint probability $p(\mathbf{z}, \mathbf{x}|s)$ over $\mathbf{x}$ to obtain

$$p\left(s \mid \mathbf{z} = \mathbf{z}_i\right) = \frac{p(\mathbf{z} = \mathbf{z}_i \mid s) p\left(s\right)}{\sum_{s'} p(\mathbf{z} = \mathbf{z}_i \mid s') p\left(s'\right)} \tag{20}$$

where

$$p\left(\mathbf{z} = \mathbf{z}_i \mid s\right) = \frac{1}{(2\pi)^{D/2} \mid \Psi_s + \Lambda_s \Sigma_s^T \Lambda_s^T \mid^{1/2}} \times \exp\left\{ -\frac{1}{2} \left(\mathbf{z}_i - \mu_s - \Lambda_s \nu_s\right)^T \left(\Psi_s + \Lambda_s \Sigma_s^T \Lambda_s^T\right)^{-1} \times \left(\mathbf{z}_i - \mu_s - \Lambda_s \nu_s\right) \right\}. \tag{21}$$

A. Integration Within a Bayesian Filtering Framework

It is significant to note that the results of inference are available in terms of multidimensional means and covariance matrices for each component of the mixture of Gaussians (4)–(6). Thus, the deterministic results of Isomap are transformed into a generative likelihood model of natural features that can be passed to a filtering algorithm.

In a filtering framework, Bayes theorem provides an incremental and recursive probabilistic method for combining high-dimensional visual observations $\mathbf{Z}^k$ of a state $\mathbf{x}_k$, at time t_k,

with a prior belief of the state $p(\mathbf{x}_{k-1})$. Features are extracted in real time from incoming natural imagery and are represented as a conditional probability distribution $p(\mathbf{x}_k, s_k|\mathbf{z} = \mathbf{z}_k)$, and the resultant combination is a revised posterior distribution on the state

$$p\left(\mathbf{x}_k, s_k|\mathbf{Z}^k\right) = \frac{p\left(\mathbf{z} = \mathbf{z}_k|\mathbf{x}_k, s_k\right) p\left(\mathbf{x}_k, s_k|\mathbf{Z}^{k-1}\right)}{p\left(\mathbf{z} = \mathbf{z}_k|\mathbf{Z}^{k-1}\right)} \quad (22)$$

where $\mathbf{Z}^k = \left\{\mathbf{z}_k, \mathbf{Z}^{k-1}\right\}$ is the set of high-dimensional visual observations from all nodes in the decentralized sensor network. The representation of the visual likelihoods as a Gaussian mixture model simplifies this update step into an algebraic computation of the product of two Gaussian mixture models.

V. Example Datasets

This section presents some experimental results using the feature extraction, statistical learning, and inference techniques with data from unstructured terrestrial and underwater environments. The probabilistic model can be used off-line to learn the model parameters consisting of the means, loading, and covariance matrices of the constituent conditional Gaussian distributions. This model can be used to instantiate visual feature likelihoods in *near-real-time* from raw images for use in conjunction with classical estimation algorithms for navigation and tracking.

A. Autonomous Ground Vehicle (AGV)—Inference of Underlying Visual States

A sample of about 9000 high-dimensional points physically representing colors and textures of typical objects in a natural environment such as sky, trees, bush, and grass was selected from a sequence of images acquired from a camera mounted on a ground vehicle at the Marulan test facility operated by the Australian Centre for Field Robotics. Texture information was included in the high-dimensional input space by convolving 11×11 pixel patches with a bank of Gabor wavelets [10] at two scales and two orientations, resulting in an input space dimensionality of 847. Isomap was used to compute a low-dimensional embedding of the training data and the intrinsic dimensionality of the manifold was estimated to be 3 through an examination of the largest eigenvalues of the scatter matrix representing the manifold [24]. The EM algorithm was then used to learn the parameters of the generative model (4)–(6). The learnt model was subsequently used to infer the low-dimensional states within a typical test image that was acquired in the same environment.

The top two eigenvectors of the computed low-dimensional embedding and the components of the mixture model (represented by the covariance ellipses centered on the means) learnt through EM are shown in Fig. 7. It is readily observed that image patches corresponding to blue skies are grouped on the left side, those representing bush are on the bottom right, while grass and transitional patches are grouped between the two extremes. The results of inference on a test image in terms of the means of the eigenvectors scaled to gray-scale limits (0–255) for the top four states are shown in Figs. 8–10.

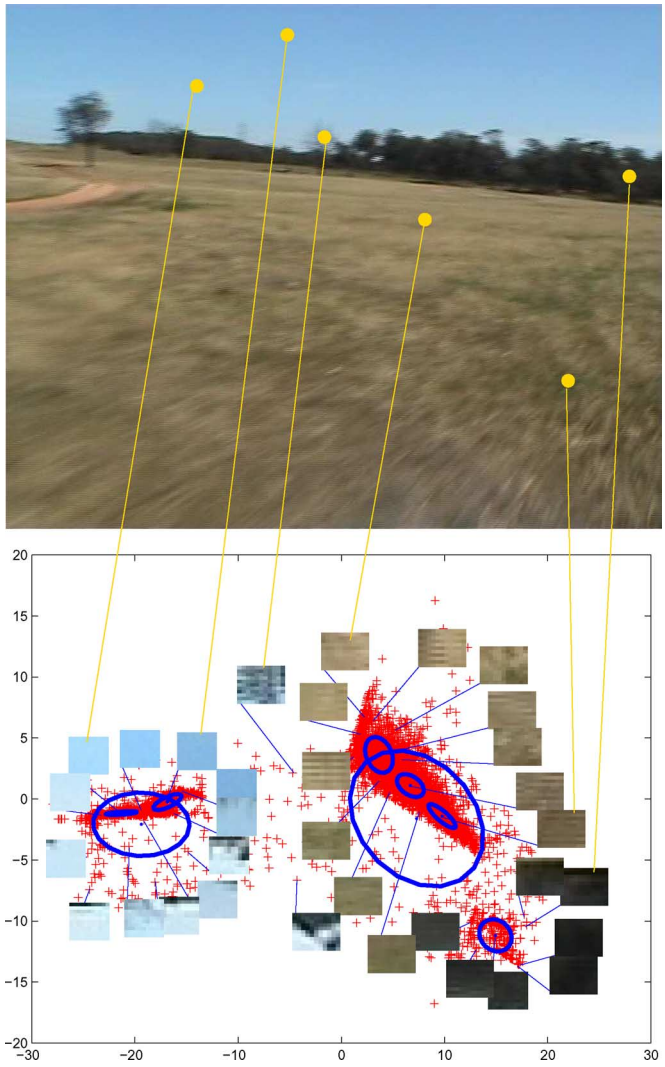


Fig. 7. (Top) Sample image acquired by the AGV and (bottom) low-dimensional embedding of randomly sampled high-dimensional image patches used as the training set. Ellipses represent the covariance matrices of the mixture model learnt through EM. Typical colors and textures in the environment are captured in the low-dimensional representation.

B. Unmanned Underwater Vehicle (UUV)—Inference of Underlying Visual States

A sample of about 17 000 high-dimensional points physically representing colors and textures of typical objects in an underwater environment such as beach sand and corals was selected from a sequence of images [25] acquired from a camera mounted onto the UUV Oberon, operated by the Australian Centre for Field Robotics.

Texture information that is vital in the characterization of the corals was included in the high-dimensional input space by convolving pixel patches with Gabor wavelets [10] and Isomap was used to compute a low-dimensional embedding of the training data and the intrinsic dimensionality of the manifold was estimated to be 3 in this case. The learning algorithm was used to learn a parametric model and the learnt model was used to infer the low-dimensional states of a typical image acquired by the

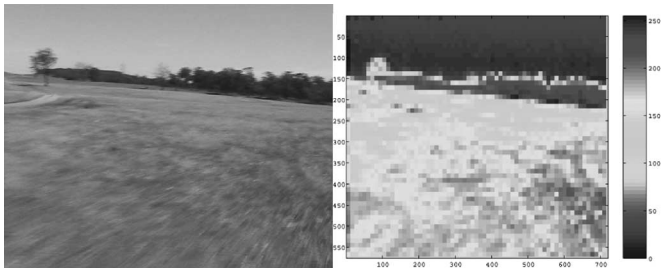


Fig. 8. Contour of the inferred means of the top eigenvector (right) on each 11×11 image patch. This state enables a clear discrimination of the sky (dark blue, range ≈0–50) from all other visual groups in the image.

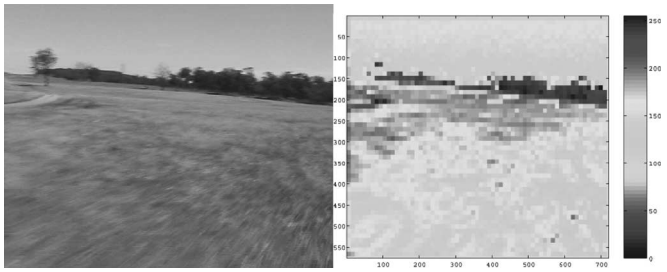


Fig. 9. Contour of inferred means of the second eigenvector. This state allows separation of the bush (range ≈0–50) from the grass and the tracks in the scene.

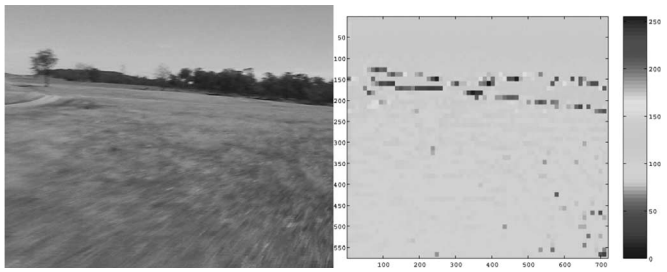


Fig. 10. Contour of inferred means of the third eigenvector. This state allows discrimination between the bush (≈125) and the tracks (≈150). Contours of the bush as well as the tracks are also apparent.

UUV. The learned manifold is shown in Fig. 11, and the results of inference are shown in Figs. 12 and 13.

C. Discussion

Each of the plots depicting the low-dimensional states must be interpreted as a contour plot of the respective states in the image plane. It is significant that every image patch consists of 847 correlated observations in the sensory space, while only a few uncorrelated states are sufficient to capture the similarities (or differences) between the patches after state inference.

The inferred low-dimensional states are reasonable in that similar high-dimensional image patches (such as those corresponding to sky, grass, trees, bush, sand, or corals) are assigned similar low-dimensional states, as is to be expected from a parametric model of neighborhood-preserving manifold learning algorithms. Each inferred low-dimensional state enables some degree of discrimination between important objects in the scene

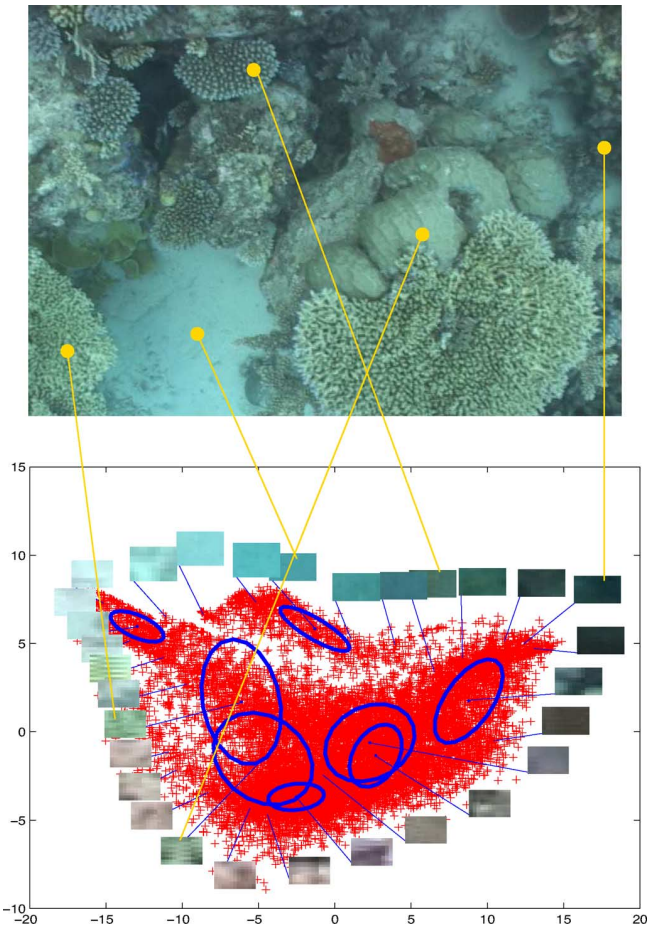


Fig. 11. (Top) Sample image acquired by the UUV and (bottom) low-dimensional embedding of randomly sampled high-dimensional image patches used as the training set. Ellipses represent the covariance matrices of the mixture model learned through EM. Typical colors and textures in the environment are captured in the low-dimensional representation.

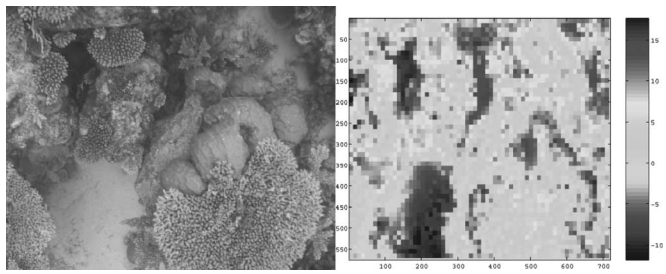


Fig. 12. Contour of the inferred means of the top eigenvector on each 11×11 image patch. This state physically corresponds to the brightness of the patches.

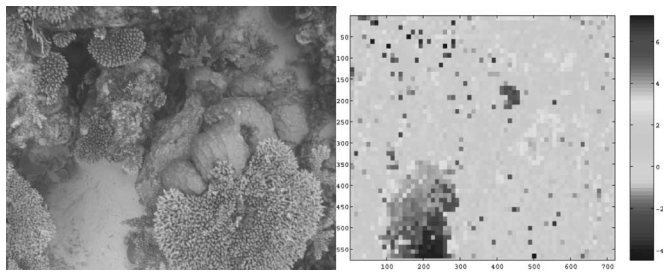


Fig. 13. Contour of the inferred means of the second eigenvector. This state seems to be correlated to the hue of the image patches.

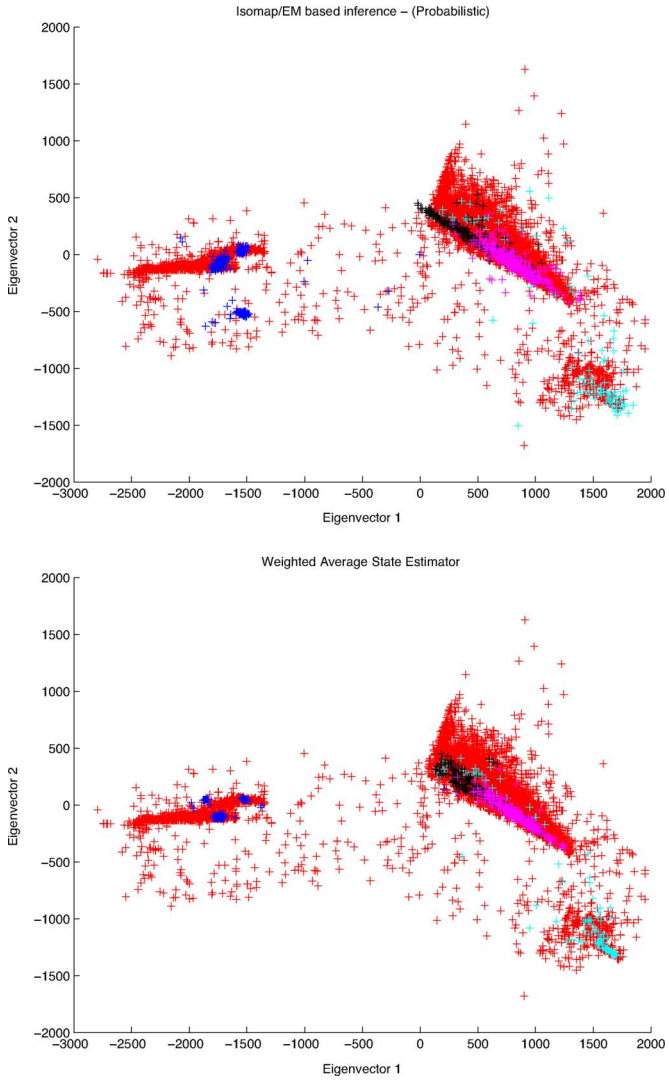


Fig. 14. (Top) Inferred low-dimensional states and (bottom) weighted average state estimate. The weighted average estimator assumes that the low-dimensional state of a test sample is the weighted average of the k nearest high-dimensional neighbors of the sample, with the weights being inversely proportional to the high-dimensional distances. The trained manifold is represented in red and corresponding test samples are similarly color coded in each figure.

such as the tree, bush, tracks, and the sky in case of the AGV or sand and corals in case of the UUV.

The accuracy of the inferred low-dimensional states is qualitatively evaluated in Fig. 14. The inferred means of the computed likelihoods of image patches in test images are overlaid on the learnt manifold (color-coded red). They are compared to results from an estimator that assumes that the low-dimensional state of a test sample is the weighted average of the k nearest neighbors ($k$nn) of the sample. The normalized weights are assumed to be inversely proportional to the high-dimensional distances. It is observed from the color coding in these figures that there is good qualitative agreement between the results of inference and a $k$nn based estimator. The probabilistic inference is more robust due to the inherent stochasticity that encodes uncertainty in the inferred states.

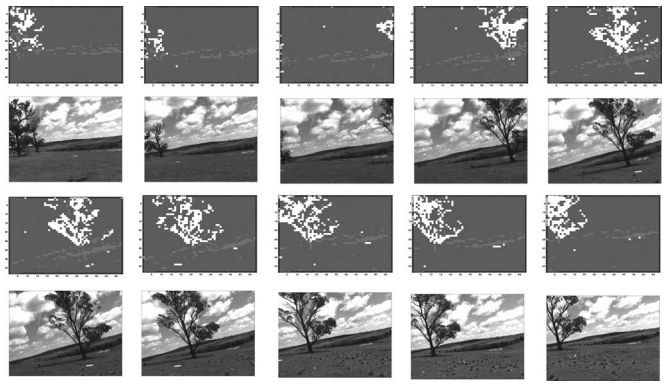


Fig. 15. Tracking of bush and tree in ten different frames of the sequence. Green patches correspond to tree while magenta patches correspond to bush. Note that bush and trees are very similar in terms of color and texture. The proposed approach is able to distinguish them based on a nonlinear combination of color and texture properties.

VI. Natural Feature Tracking

This section investigates the applicability of the proposed natural feature representation scheme to tracking and data association in an unstructured terrestrial environment. A sequence of 1100 images was acquired from the AGV shown in Fig. 2. The natural feature representation model is used in an attempt to track and discriminate between the bush and trees through the image sequence. This is a challenging problem as the bush and trees present a remarkably similar appearance.

A. Representation Model

A set of 20 000 image patches of size 11×11 was randomly sampled from the sequence and used as input to learn the probabilistic natural feature model consisting of 32 mixture components. The input space dimensionality was 847 corresponding to three color intensities and four texture dimensions as reported in Section V.

B. General Observations From Likelihood Instantiation

The learnt natural feature model was used to infer the low-dimensional visual states of test images as described in Section IV. For this experiment, no feature selection is performed, the whole image is divided into patches, and inference is computed for all patches. Numerical experiments consistently demonstrated that the region of the manifold representing bush-like image patches is completely explained by the mixture component numbered 8 (of the 32 learned components). This is evidenced by negligible mixture weights (20) in the other components of the inferred likelihood. Similarly, mixture components numbered 2, 7, 23, 27 were always associated with tree-like patches.

C. Data Association and Segmentation

The regions selected by the feature extractor are subdivided into patches that form high-dimensional observations for the generative model. The magnitudes of the weights of the instantiated likelihoods are used to identify bush or tree-like patches. Fig. 15 demonstrates that image patches are correctly

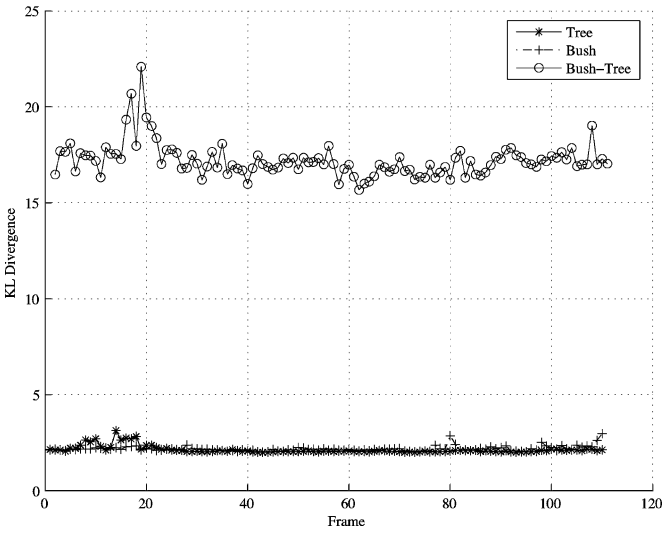


Fig. 16. (Blue line) KL divergence between "tree" patches in two frames. (Red dot-dash line) KL divergence between "bush" patches in two frames and KL divergence between "bush" and "tree" in the same frame. Every tenth frame from the entire image sequence is used in these computations.

associated to their real world identities despite the ambient changes and occlusions inherent in the image sequence. It also shows that a reasonable segmentation of tree and bush can be directly obtained by using this approach.

The average likelihood of all bush (as evidenced by the dominance of the mixture component numbered 8) and tree-like image patches (mixture components 2, 7, 23, 27) within a single frame is computed. The discriminative capability of the likelihood is explored by computing the Kullback–Leibler (KL) divergence [26] between similar (tree–tree, bush–bush) and distinct (tree–bush) natural features. The KL divergence provides a relative measure of similarity between general probability distributions and is depicted for every tenth frame in the entire image sequence in Fig. 16.

Fig. 16 shows that the KL divergence between physically distinct features is about six times larger than that between the same features in the selected frames. Thus, the proposed natural feature representation significantly aids data association through a comparison of visual likelihoods conditioned on the intrinsic coordinates on the low-dimensional manifold.

D. Discussion

The unsupervised association of specific mixture components to tree and bush-like image patches is quite remarkable and is a direct consequence of the proposed natural feature representation. The choice of a nonlinear manifold learning scheme (i.e., Isomap) for the sensor data compression ensures that similar natural features are positioned in close proximity on the learnt manifold. This translates into the fact that specific mixture components are quite likely to represent physically similar natural features. The second important aspect that allows superior data association is the choice of an optimal number of mixture components. An inadequate number of mixture components would result in representations where distinct natural features could be explained by the same mixture components. An excessive number of mixture components could lead to an overconstrained model that may not be representative of a broad class of visual data [22]. An appropriate number of mixture components (32) was chosen in this scenario through numerical experiments on test image sequences.

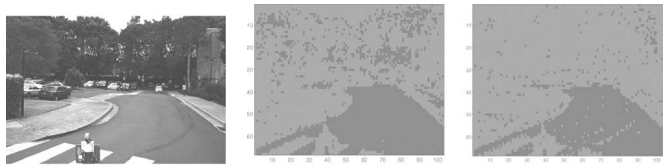


Fig. 17. Results for KNN road segmentation. (a) Original image. (b) Segmentation on the original high-dimensional space. (c) Segmentation on the low-dimensional space.

VII. Road Detection

In order to evaluate the proposed approach in navigation tasks, an experiment was performed for the detection of roads in urban environments. Ten high-resolution images (1134×746) were manually segmented from a dataset consisting of a 2-km trajectory in the university campus. These ten images were selected to consider different asphalts and illumination conditions found in the campus.

A. Representation Model

As in the previous experiment, the images were divided into 11×11 nonoverlapping patches with texture computed with Gabor filter convolutions. The total input dimensionality was 847 corresponding to three color intensities and four texture dimensions.

B. Classification Results

Direct classification in the original space is computationally very expensive given the dimensionality of the data. As an alternative, the proposed approach can be used to provide a low-dimensional and compact representation of the data for the classifier. For this experiment, the *K*-nearest neighbor (KNN) classifier [24] was used given its simplicity and capacity to model highly nonlinear decision boundaries. The number of nearest neighbors was 7, selected by performing cross validation. The performance of KNN is then compared by applying it to both the original 847-D space and the 5-D space obtained with the proposed approach. Fig. 17 shows an image of the dataset and the segmentation obtained in both cases.

Results are presented as the average of a tenfold cross validation. For each cross-validation case, nine images are used for training and the remaining for testing. The average 10-fold cross-validation result for KNN applied to the original space was 85.21% of correct classification. When KNN is applied to the compact low-dimensional space, the classification obtained was 88.70%.

C. Discussion

In addition to the improvement of about 3.5% in the classification results, the low-dimensional representation provides a substantial improvement in terms of speed. While KNN takes about an hour to classify an image in the original space, the same classifier takes only around 5 s to classify the whole image with low-dimensional features. The segmentation involves the classification of 7004 patches, which in KNN requires searching the K-nearest neighbors from the training data. These results demonstrate that the low-dimensional representation is able to preserve most of the information content necessary to perform classification. Note that road segmentation in urban environments can be quite a difficult task considering that colors and textures of the roads can be very similar to the footpath or even buildings.

VIII. Conclusion

The combination of nonparametric manifold learning algorithms with statistical learning strategies leads to a consistent description of natural features in unstructured environments. While the entire learning procedure can be incorporated in the training phase of these models that is performed off-line, inference can be performed in real-time on any extracted features to compute likelihoods for the natural features as a Gaussian mixture model. Natural features can thus be fully integrated within existing non-Gaussian, nonlinear filtering algorithms through the likelihood model so that tasks of estimation and data association are significantly enhanced through a combination of kinematic and visual states.

The experiments show the potential of the proposed framework for different problems of perception in unstructured environments. The probabilistic representation correctly associates patches representing coral, sand, sky, bush, tree, and grass in different places of the low-dimensional space. The tracking and data association experiments demonstrate how the proposed approach can significantly facilitate these tasks.

Although the model is sufficient for perception of natural features where colors and textures are in general good descriptors of the objects, it does not incorporate shape information that is more appropriate for human-made objects. Furthermore, given the nature of the maximum likelihood estimator, prior information is not considered. An interesting avenue of research in this vein would be the combination of the proposed approach with spatial statistics in a Bayesian learning framework where prior information can be regarded together with knowledge coming from the dataset.

References

[1] D. J. MacKay, *Information Theory, Learning and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[2] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 107–126, 1998.

[3] T. Lee, T. Wachtler, and T. J. Sejnowski, "Relations between the statistics of natural images and the response properties of cortical cells," *Vis. Res.*, vol. 42, pp. 2095–2103, 2002.

[4] Y. Karklin and M. Lewicki, "Learning higher order structure in natural images," *Comput. Neural Syst.*, vol. 14, pp. 483–499, 2003.

[5] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. 2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 1735–1742.

[6] B. Upcroft, B. Douillard, T. Kaupp, M. X. Ridley, M. Ridley, L.-L. Ong, S. Kumar, T. Bailey, F. Ramos, S. Sukkarieh, and H. Durrant-Whyte, "Non-Gaussian state estimation in an outdoor decentralised sensor network," in *Proc. IEEE Conf. Decision Control*, Dec. 13–15, 2006, vol. 45, pp. 366–372.

[7] C. Tomasi and T. Kanade, "Detection and tracking of point features," Sch. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-91-132, 1991.

[8] D. G. Lowe, "Object detection from local scale invariant features," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[9] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.

[10] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, 1987.

[11] B. Scholkopf, A. J. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1998.

[12] J. Tenenbaum, V. DeSilva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality redution," *Sci.*, vol. 290, pp. 2319–2323, 2000.

[13] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Univ. Chicago, Dept. Comput. Sci., Tech. Rep., 2002.

[14] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Sci.*, vol. 290, pp. 2323–2326, 2000.

[15] I. Foster, *Designing and Building Parallel Programs*. Reading, MA: Addison-Wesley, 1995.

[16] E. W. Dijkstra, "A note on two problems in connexion to graphs," *Numer. Math.*, vol. 1, pp. 269–271, 1959.

[17] T. Cox and M. Cox, *Multidimensional Scaling*. London, U.K.: Chapman & Hall, 1994.

[18] J. Tenenbaum, V. DeSilva, and J. C. Langford. (2000) [Online]. Available: http://isomap.stanford.edu

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–37, 1977.

[20] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," Dept. Comput. Sci., Univ., Toronto, Toronto, ON, Canada, Tech. Rep. CRG-TR-96-1, 1996.

[21] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of nonlinear manifolds," Univ. Pennsylvania, Dept. Comput. Sci., Tech. Rep., 2002.

[22] M. Beal, "Variational algorithms for approximate Bayesian inference" Ph.D. dissertation, Univ. Cambridge, U.K., 2003.

[23] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag, 1999.

[24] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[25] S. Williams and I. Mahon, "Simultaneous localization and mapping on the Great Barrier Reef," in *Proc. IEEE Conf. Robot. Autom.*, 2004, pp. 1771–1776.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

Fabio Tozeto Ramos (M'07) received the B.Sc. and M.Sc. degrees in mechatronics engineering from the University of Sao Paulo, Sao Paulo, Brazil, in 2001 and 2003, respectively, and the Ph.D. degree from the University of Sydney, Sydney, N.S.W., Australia, in 2007.

Since 2007, he has been a Senior Research Fellow with the ARC Centre of Excellence for Autonomous Systems, Australian Centre for Field Robotics, University of Sydney. He received the Australian Research Council (ARC) Postdoctoral Fellowship in 2008. His current research interests include statistical learning techniques for dimensionality reduction, stochastic processes modeling, and object recognition with applications in robotics and mining.

Dr. Ramos received the Best Paper Award at the International Conference on Intelligent Robots and Systems in 2005.

Suresh Kumar received the B.Tech degree in engineering from the Indian Institute of Technology, Chennai, India, in 1992 and the M.S. and Ph.D. degrees in computational engineering mechanics from the State University of New York, Buffalo, in 1994 and 1997, respectively.

He is currently with the ARC Centre of Excellence for Autonomous Systems, Australian Centre for Field Robotics, University of Sydney, Sydney, N.S.W., Australia. His current research interests include finite and boundary element methods for the solution of partial differential equations, inverse problems (acoustics, EEG, electrocardiography), and machine learning methods for data representation and interpretation.

Ben Upcroft received the Undergraduate degree (with honors) in science and the Ph.D. degree in ultracold atomic physics, both from the University of Queensland, Queensland, Australia.

He held a Postdoctoral position in robotics in 2003 with the ARC Centre of Excellence for Autonomous Systems, School of Aerospace, Mechatronics, and Mechanical Engineering, University of Sydney, Sydney, N.S.W., Australia. He has run major industrial projects involving autonomous aircraft, offroad and urban vehicles, and network communications. He is currently a Senior Lecturer in mechatronics with the University of Queensland, Brisbane, Qld., Australia. His current research interests include computer-vision-aided localization and navigation for autonomous ground vehicles.

Hugh Durrant-Whyte received the B.Sc. degree in nuclear engineering from the University of London, London, U.K., in 1983 and the M.S.E. and Ph.D. degrees, both in systems engineering, from the University of Pennsylvania, Philadelphia, in 1985 and 1986, respectively.

From 1987 to 1995, he was a University Lecturer in engineering science at the University of Oxford, Oxford, U.K. He was also a Fellow of Oriel College, Oxford. Since 1995, he has been a Professor of mechatronic engineering with the ARC Centre of Excellence for Autonomous Systems, Australian Centre for Field Robotics, University of Sydney, Sydney, N.S.W., Australia. He received Australian Research Council Federation Fellowships in 2002 and 2007. He is the author or coauthor of more than 350 research papers. His current research interests include robotics and sensor networks.

Prof. Durrant-Whyte was the recipient of numerous awards and prizes.