Pattern Recognition Letters
January 2011, Volume 32, Issue 2, Pages 153-158
http://dx.doi.org/10.1016/j.patrec.2010.10.001

Object recognition using proportion-based prior information: Application to fisheries acoustics

R. Lefort[a, b, *], R. Fablet[b] and J.-M. Boucher[b]

[a] Ifremer/STH (French Research Institute for Exploitation of the Sea), Technople Brest Iroise, 29280 Plouzane, France
[b] Institut Telecom/Telecom Bretagne, UMR CNRS Lab-Sticc, Universit de Europenne de Bretagne Technople Brest-Iroise – CS 83818, 29238 Brest Cedex, France

*: Corresponding author : R. Lefort, email address : riwal.lefort@telecom-bretagne.eu

Abstract:

This paper addresses the inference of probabilistic classification models using weakly supervised learning. The main contribution of this work is the development of learning methods for training datasets consisting of groups of objects with known relative class priors. This can be regarded as a generalization of the situation addressed by Bishop and Ulusoy (2005), where training information is given as the presence or absence of object classes in each set. Generative and discriminative classification methods are conceived and compared for weakly supervised learning, as well as a non-linear version of the probabilistic discriminative models. The considered models are evaluated on standard datasets and an application to fisheries acoustics is reported. The proposed proportion-based training is demonstrated to outperform model learning based on presence/absence information and the potential of the non-linear discriminative model is shown.

Research highlights

► Weakly supervised learning deals with prior annotation of objects in images. ► Classification model must be assessed by using probabilities. ► Reported results promote discriminative.

Keywords: Weakly supervised learning; Generative classification model; Discriminative classification model

1. Introduction

In object recognition and classification, most of the research effort was initially dedicated to supervised learning, i.e. solving classification problems for which a dataset of labelled objects exists. Unsupervised learning has also encountered a great interest when no prior knowledge on object classes

is available in the training dataset (Hinton and Sejnowski (1999)). In a number of applications however the training set comprises some information on object classes but not all objects may be labelled. Typical examples, often referred to semi-supervised or partially-supervised cases, involve training sets in which only a subset of the objects are labelled (Chapelle et al. (2006)). Semi-supervised learning generally relies on an initial supervised learning followed by a refinement of the classifiers from non-labelled samples in the training set. Weakly-supervised learning involves a more general situation: the prior information provided in the training set is given as a subset of potential classes for each object. Examples of weakly-supervised learning can be drawn from image analysis. The training object dataset may be built from a set of labelled images, each label corresponding to an object class, such that a particular object in a given image may be associated with several possible classes (Vasconcelos et al. (2006), Fergus et al. (2007), Crandall and Huttenlocher (2006), Bishop and Ulusoy (2005)). Similar situations are reported in the management of large document databases, as several classes, corresponding to different concepts, may be associated with each document in the training dataset (Gosselin and Cord (2006)).

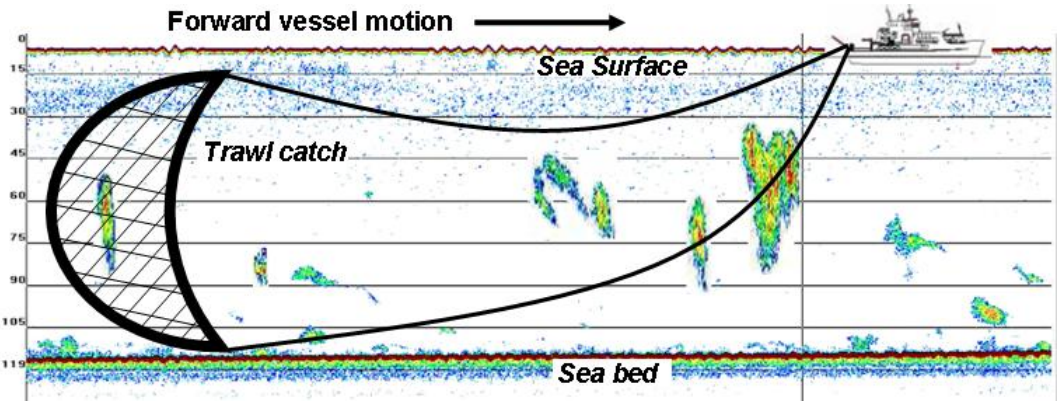


Figure 1: *The sonar echo sounder placed under a vessel acquires echograms. In an echogram, fish school aggregations of sardina, anchovy and horse mackerel may be observed. Basic descriptors are extracted (length, height, depth, energy, etc, of the school) for species discrimination. If associated with trawling, the analysis of trawl catches provides the species proportion in the training echograms.*

In this paper, contrary to previous approaches (Fergus et al. (2007), Cran-

dall and Huttenlocher (2006), Weber et al. (2000), Xie and Perez (2004), Bishop and Ulusoy (2005)), we do not restrict our interest to training information given as the presence/absence of a class in an image or a set of objects. We assume that class priors are available for each training image or object cluster. Data labelling derives from different priors, i.e. samples are not necessary equally fuzzy in terms of label. Furthermore, experiments are carried out with equal fuzzy labelling or with different fuzzy labelling, the objective being to assess performances of classifiers given the complexity of the training data. We consider that a label has noise when samples are nearly equally fuzzy and that a label has few noises when one prior is dominating among classes.

Weakly supervised training schemes are applied to fisheries acoustics. In fisheries acoustics, the estimation of fish stock biomass (Scalabrin and Mass (1993)) requires to carry out a species-based classification of the fish schools observed in the echograms acquired by an echosounder (MacLennan and Simmonds (1992)) (Figure 1). In this case, the training data is built by a set of echograms associated with trawling data. As the analysis of trawl catches provides relative species proportions, each training image and the corresponding set of segmented fish schools are assigned to this relative proportion of each class, i.e. each species. Similar examples may be encountered in image analysis applications, especially regarding remote sensing applications (Anderson et al. (2008), Descle et al. (2006)).

In this paper, weakly supervised learning from proportion-based information is investigated. Following works by Bishop and Ulusoy (2005), two categories of models are considered: generative classifier and discriminative probabilistic classifier. In addition, a novel Fisher-based learning of the discriminative models and the introduction of non-linear conditional models are shown to bring significant improvement in terms of classification performances. The paper is organized as follows. Section 2 presents the problem statement and notations while sections 3 and 4 describe the probabilistic classification models. Experiments and performances of these models are tested using standard datasets and a fisheries acoustics dataset in section 5. Concluding remarks and perspectives are given in section 6.

2. Problem statement

We assume that the training set is provided as a set of images containing segmented objects along with the relative proportion of each class within each

image. Formally, let us denote by k the image index and by $\{x_{kn}\}$ the feature vector of the object indexed by n contained in image k. For image k, the associated global information is provided as the relative proportions of the different classes in image k. Let us denote by $\pi_k = \{\pi_{ki}\}$ these relative class proportions, i being the class. Depending on the application, proportions might be computed with respect to the relative occurrences of objects for each class or with respect to physical characteristics of the objects (e.g., surface, energy). For instance, in the considered application to fisheries acoustics, these relative class proportions are computed as the relative acoustic energy of each class.

We further introduce the following notations. y_{kn} is the class vector of the object n in image k and $y_{kn} = i$ indicates that this object is associated with class i. The aim is to evaluate the probability for object n in image k to be assigned to class i knowing feature vector x_{kn} and model parameters Θ: $p\left(y_{kn} = i|x_{kn}, \Theta\right)$. Θ is assessed in the training step. The subsequent sections detail the chosen models and parameterizations for Θ as well as the associated learning schemes.

3. Generative model (GM)

The generative model proposed by Bishop and Ulusoy (2005) that deals with presence/absence case, is extended in this paper to the proportion-based case. Given $\Theta = \{\rho_{i1} \ldots \rho_{iM}, \mu_{i1} \ldots \mu_{iM}, \Sigma_{i1}^2 \ldots \Sigma_{iM}^2\}$ the parameters of a Gaussian mixture model:

$$p\left(x|y = i, \Theta\right) = \sum_{m=1}^{M} \rho_{im} \mathcal{N}(x|\mu_{im}, \Sigma_{im}^2) \tag{1}$$

$\mathcal{N}(x|\mu_{im}, \Sigma_{im}^2)$ is the normal distribution with mean μ_{im} and covariance matrix: Σ_{im}^2. The learning of model parameters Θ is then stated as a probabilistic inference issue. For proportion training data set of the form $\{x_k, \pi_k\}_k$, and considering that $\pi_{ki} = p\left(y_{kn} = i\right)$, a maximum likelihood criterion can be derived:

$$\hat{\Theta} = \arg\max_{\Theta} p(\pi|x, \Theta) = \arg\max_{\Theta} \prod_{k} p(\pi_k|x_k, \Theta) \tag{2}$$

The EM (Expectation-Maximization) (Dempster et al. (1977)) procedure is exploited to solve for this estimation. When considering proportion-based

training data, the proportion data is regarded as a class prior for each image, such that the E-step is modified to take into account this prior as follows:

$$p\left(y_{kn}=i|x_{kn},\Theta^{c}\right)=\frac{\pi_{ki}p\left(x_{kn}|y_{kn}=i,\Theta^{c}\right)}{\sum_{l}\pi_{kl}p\left(x_{kn}|y_{kn}=l,\Theta^{c}\right)} \tag{3}$$

Θ^{c} refers to current parameters. Given these posterior likelihoods, the M-step is similar to the one by Bishop and Ulusoy (2005).

Let us stress that we consider diagonal covariance matrix for the multi modal Gaussian model. Once the learning step is performed i.e. parameter Θ is estimated, object classification resorts to selecting the most likely class according to posterior likelihood (3).

There are various discriminative optimizations for Gaussian mixture models such as Maximum Mutual Information (MMI). This one allows to removed from training dataset samples that are fuzzed in terms of features by calculating their contribution in probability density function (Yang and Zwolinski (2001)), or it allows to estimate model parameters that maximize the mutual information (Bahl et al. (1986)). In this work, MMI is not used because the criterion does not take into account prior information. Nevertheless, the development of new optimizations that mix MMI criterion and prior knowledge should be considered in future work.

4. Discriminative model

In this section, discriminative models, both linear and non-linear ones, are considered.

4.1. Linear discriminative model (LDM)

Discriminative models are stated as an explicit parameterization of the classification likelihood:

$$p\left(y=i|x,\Theta\right)\propto F(\langle\omega_{i},x\rangle+b_{i}) \tag{4}$$

where $<\omega_{i},x>+b_{i}$ is the distance to the separation hyperplane between class i and the other classes. The hyperplane equation is $<\omega_{i},x>+b_{i}=0$ in the feature space. Model parameters Θ are given by ω_{i} and b_{i}. F is an

increasing function, typically an exponential or continuous stepwise function. Hereafter, F will be chosen to be an exponential function:

$$p\left(y=i|x,\Theta\right)=\frac{\exp(\langle\omega_i,x\rangle+b_i)}{\sum_l\exp(\langle\omega_l,x\rangle+b_l)} \tag{5}$$

4.2. Non linear discriminative model (NLDM)

We propose an extension to a non-linear discriminative model using a kernel approach (Schlkopf and Smola (2002)). Prior to model training, a non-linear mapping of the feature space is carried out. We consider a kernel function K_e associated with a non-linear mapping Φ of the original feature space to a new space, such that K_e is the dot product in the mapped feature space: $K_e(x_1,x_2)=\langle\Phi(x_1),\Phi(x_2)\rangle$. A non-linear discriminative model can then be defined as a probabilistic classifier in the mapped feature space:

$$p\left(y=i|x,\Theta\right)\propto\exp\left(\langle w_i,\Phi(x)\rangle+b_i\right) \tag{6}$$

where w_i is the vector normal to the separation hyperplane in the mapped space. Depending on the chosen kernel, the direct parameterization of w_i may not be obvious. Consequently, we exploit a kernel Principal Component Analysis (PCA) to exhibit a parameterization from the subspace, spanned by the training data:

$$w_i=\sum_{p=1}^{N_{PCA}}w_{I,p}*B_p \tag{7}$$

where N_{pca} is the number of PCA basis and $\{Bp\}$ the orthonormal PCA basis in the mapped feature space. Exploiting PCA for dimensionality reduction into the mapped space, mapped feature vector $\Phi(x)$ is approximated as:

$$\Phi(x)=\sum_{p=1}^{N_{PCA}}\alpha_p(x)*B_p \tag{8}$$

where $\alpha_p(x)$ is the projection of $\phi(x)$ onto $\{Bp\}$. Basis $\{Bp\}$, issued from the diagonalization of the matrix $\{K_e(x_i,x_j)\}_{i,j}$ are determined as linear combinations of $\{\Phi(x_i)\}$, $B_p=\sum_i\beta_{p,i}*\Phi(x_i)$, and $\alpha_p(x)$ are given by: $\alpha_p(x)=\langle\Phi(x),B_p\rangle=\sum_i\beta_{p,i}K_e(x,x_i)$. Considering the projection of model parameter ω_i onto the PCA basis, $p\left(y=l|x,\Theta\right)$ can be rewritten as:

$$\log p\left(y=i|x,\Theta\right)\propto\sum_{p,l}\omega_{i,p}\cdot\beta_{p,l}K(x,x_l)+b \tag{9}$$

where vector $\{\omega_{i,p}\}$, such that $\omega_{i,p} = \langle \Phi(\omega_{i,p}), B_p \rangle$, is the actual model parameter of dimension N_{PCA} for each class. It should be stressed that the expression of the conditional likelihood for the non-linear model is similar to the expression of the linear model: it amounts to replacing original feature vector x by feature vector $\{\sum_l \beta_{p,l} K_e(x, x_l)\}_p$ of dimension N_{PCA}.

4.3. Training scheme

In section 4.1 and section 4.2, we present the mathematical formalization of the discriminative model. In this section, the learning step is introduced, in particular, we present a criterion that learns parameters $\Theta = \{\omega_i, b_i\}_i$.

Given an initialization for Θ, model parameters Θ are optimized according to a minimization criterion based on Battacharrya divergence (Bhattacharyya (1943)). It consists in determining parameters $\hat{\Theta}$ that minimize the error between the known proportion π_k and the estimated proportion $\hat{\pi}_k(\Theta)$:

$$\hat{\Theta} = \arg\min_{\Theta} \sum_k D(\hat{\pi}_k(\Theta), \pi_k) \tag{10}$$

Among the different distances between likelihood functions, the Battacharrya distance is chosen (Bhattacharyya (1943)): $D(\hat{\pi}_k(\Theta), \pi_k) = \frac{1}{I} \sum_i \sqrt{\hat{\pi}_k(\Theta) \cdot \pi_k}$ where I is the number of classes. The classification likelihoods (5) and (6) give the estimated proportion $\hat{\pi}_k(\Theta)$.

The general idea of the criterion (10) is that the ideal classifier, i.e. the classifier that finds all classes of instances, will produce the same class proportions as for the training dataset. If training data are good classified, the class proportion in training images must be the same than before the classification.

Given an initial parameter estimate, the minimization of criterion (10) is achieved using a gradient descent. In this paper, the initialization of the parameters are done by setting coefficients $\{\omega_i, b_i\}_i$ to 1, as a uniform initialization, or by using the Fisher-based criterion introduced in the next section.

4.4. Fisher-based Model

A Fisher-based discriminative model can be used as an initialization for the previous model based on the optimization criterion 10. For the sake of simplicity, we hereafter consider a two-class case. Fisher discrimination (Fisher (1936)) amounts to maximizing ratio between inter-class and intra-class variances:

$$\hat{\omega} = \arg\max_{\omega} \left\{ \frac{\left(\omega^T (m_1 - m_2)\right)}{\omega^T \left(\Sigma_1 + \Sigma_2\right) \omega} \right\} \tag{11}$$

where m_1 and Σ_1 are the mean and variance of the first class, and m_2 and Σ_2 are the mean and variance of the second class. The estimate is given by $\hat{\omega} = (\Sigma_1 + \Sigma_2)^{-1}(m_1 - m_2)$.

Fisher discrimination is applied to weakly supervised learning based on the estimation of class mean and variance for known object class priors. Formally, considering a one-versus-all strategy, for a given class i, mean m_1 and variance Σ_1 are estimated as:

$$m_1 = \frac{\sum_{k}^{K} \sum_{n}^{N(k)} \pi_{ki} x_{kn}}{\sum_{k}^{K} \sum_{n}^{N(k)} \pi_{ki}}, \tag{12}$$

$$\Sigma_1 = \frac{\sum_{k}^{K} \sum_{n}^{N(k)} \pi_{ki} (x_{kn} - m_1)(x_{kn} - m_1)^T}{\sum_{k}^{K} \sum_{n}^{N(k)} \pi_{ki}}, \tag{13}$$

m_2 and Σ_2 are computed replacing π_k by $(1-\pi_k)$. This procedure is employed as an estimation for model parameters ω_i or only as an initialization of the gradient-based minimization of criterion (10).

5. Experiments and performances

5.1. Simulation procedure

For evaluation purposes, a groundtruthed set of objects is considered. We apply a random sampling of this candidate object set to form training images as random object subsets. This random sampling procedure is carried out according to target class proportions which determine the complexity of the mixture. Depending on these target proportions, a given training image may comprise objects from one (i.e. proportions equalling zero or one) to all classes (i.e. all proportions values are non-zero). It might be noted that several specific cases may be encountered: supervised training sets for which binary target relative proportions lead training images involving only one object class, unsupervised training sets for which target relative proportions are uniform, and semi-supervised training sets being a combination of the

two previous situations. The overall test procedure, including the generation of the training and test images, the training and the evaluation of the correct classification rate and of the proportion estimation error on the test dataset, is repeated one hundred times to evaluate classification performance.

5.2. Data sets

Four datasets are considered. The first dataset, D1, is sampled from two-dimensional Gaussian mixtures. 1000 samples are generated for each class. Each class is characterized by a mixture of two Gaussian modes $\mathcal{N}(\mu_{im}, \Sigma_{im}^2)$ with mode proportions $\rho_i = \begin{pmatrix} 0.6 & 0.4 \end{pmatrix}$. The parameters of the Gaussian modes are as follows:

$\mu_{11} = \begin{pmatrix} 0 & 1 \end{pmatrix}$,
$\mu_{12} = \begin{pmatrix} 0 & 4 \end{pmatrix}$,
$\mu_{21} = \begin{pmatrix} \sqrt{3}/2 & -1/2 \end{pmatrix}$,
$\mu_{22} = \begin{pmatrix} 2\sqrt{3} & -2 \end{pmatrix}$,
$\mu_{31} = \begin{pmatrix} -\sqrt{3}/2 & -1/2 \end{pmatrix}$,
$\mu_{32} = \begin{pmatrix} -2\sqrt{3} & -2 \end{pmatrix}$,
and $\Sigma_{im} = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$.

Two other data set, D2 and D3, have been considered in numerous comparison. D2 comes from the Waveform Database Generator (Breiman et al. (1984)). It contains 3 classes, each sample being characterized by 21 continuous attributes. Each class comprises 1660 samples. D3 comes from the UCI Repository of Machine Learning Databases (Blake and Merz (1998)). Containing 7 classes of objects with 19 continuous attributes, it allows us to simulate images containing 7 classes mixture. Each class comprises 330 samples.

The fourth database D4 is a set of fish schools, as described in Figure 1, that have been labelled by experts. The school dataset is composed of four school classes corresponding to different fish species: sardina (179 instances), anchovy (478 instances), horse mackerel (667 instances) and blue whiting (95 instances). It was built from schools observed in echograms corresponding to trawl catches with only one species (Scalabrin et al. (1996)). 19 school features are used: solid geometry and energy for each school (Scalabrin et al. (1996)). The estimated proportion in echogram k for a given species i is defined as follows:

$$\hat{\pi}_{ki}(\Theta) = \frac{\sum_{n=1}^{N(k)} E_{kn} p\left(y_{kn} = i | x_{kn}, \Theta\right)}{\sum_{i'} \sum_{n=1}^{N(k)} E_{kn} p\left(y_{kn} = i' | x_{kn}, \Theta\right)} \tag{14}$$

where E_{kn} is the energy of fish school n in echogram k and $p\left(y_{kn} = i | x_{kn}, \Theta\right)$ the posterior classification likelihood.

5.3. Results

Global performance. Global classification performances are shown in Table 1 for the four datasets. The following notations are used for the classification models: GM (generative model in section 3), F-LDM (Fisher-based linear discriminative model in section 4.4), UO-LDM (optimization procedure as seen in section 4.3 with a uniform initialization), FO-LDM (optimization procedure with a Fisher-based initialization), and the non-linear version of LDM, i.e. F-NLDM, UO-NLDM, and FO-NLDM. The mean correct classification rate is shown as a function of the complexity of the mixture from one class per training image (supervised learning) to 3 or 4 or 7 species per training image respectively for datasets D1, D2, D3, and D4. We define the classification rate as the mean of the correct classification rate over classes. The best classification rate for each dataset is in bold. A procedure is considered as outperforming the others when it has a better classification rate and when it is more robust with regard to the complexity of training object mixture. Ideal situation would lead to a steady and high classification rate when the number of class per image increases.

Overall, the higher the number of species in the training images, the lower the classification performance. This is expected as model training is more difficult when training instances can be assigned to several classes, compared to the supervised case. Results indicate that the non-linear discriminative Fisher-based model (F-NLDM) outperforms the other models for datasets D2, D3, and D4 except for 7-class mixtures with D3 and 4-class mixtures with D4. F-LDM and FO-LDM are the best models with dataset D1.

Regarding the GM, results can be compared to the other models for supervised learning situation but the performances fall down when the number of class in the mixtures is rising. Actually, for dataset D1 it falls from 79.4% with 2-class mixture to 57.3% with 3-class mixture, for D2 from 79.5% with

Class number per image		1	2	3	4	5	6	7
D1	GM	81.7%	79.4%	57.3%				
	F-LDM	**82.1%**	**81.8%**	**81.9%**				
	FO-LDM	**82.1%**	**81.8%**	**81.9%**				
	UO-LDM	81.5%	81.3%	81.2%				
	F-NLDM	81.5%	80.3%	77.2%				
	FO-NLDM	81.7%	81.5%	81.3%				
	UO-NLDM	33.3%	33.3%	33.3%				
D2	GM	80.8%	79.5%	58.3%				
	F-LDM	85.9%	85%	83.7%				
	FO-LDM	85.9%	85%	83.7%				
	UO-LDM	33.3%	33.3%	33.3%				
	F-NLDM	**86.4%**	**85.5%**	**84.1%**				
	FO-NLDM	85.7%	85.3%	83.9%				
	UO-NLDM	33.3%	33.3%	33.3%				
D3	GM	83.7%	83.6%	84.4%	83.7%	83.8%	83.1%	75.1%
	F-LDM	87%	83%	82%	79.5%	82.7%	76.6%	67.3%
	FO-LDM	69.9%	76.1%	77.3%	77.2%	76.9%	77.5%	78%
	UO-LDM	28.3%	13.2%	13.2%	13.2%	13.2%	13.2%	13.2%
	F-NLDM	**89.7%**	**89.2%**	**89.5%**	**89.1%**	**89.1%**	**89.1%**	85.9%
	FO-NLDM	88.3%	88.5%	88.3%	88.5%	88.7%	88.2%	**86.7%**
	UO-NLDM	85.9%	35.3%	13.7%	14.3%	15.8%	14.6%	15%
D4	GM	66.9%	52%	51.2%	47.9%			
	F-LDM	67.6%	68.6%	64.2%	**57.7%**			
	FO-LDM	67.5%	68.6%	62.3%	56.9%			
	UO-LDM	61%	60.2%	56.3%	56.9%			
	F-NLDM	**69.9%**	**71.7%**	**65.9%**	56.2%			
	FO-NLDM	69.8%	70.3%	61.9%	52.9%			
	UO-NLDM	53.6%	52.7%	48.7%	49.9%			

Table 1: *Classification performance as a function of the complexity of the training data. The rate of correct classification is reported as a function of the proportion complexity of the training dataset, from supervised learning to 7-class mixtures. Reported results include the classification performance of the generative model (GM), the Fisher-based linear discriminative model (F-LDM), the optimization procedure with a Fisher-based initialization (FO-LDM), the optimization procedure with a uniform initialization (UO-LDM), and the non-linear version of LDM, i.e. F-NLDM, UO-NLDM, and FO-NLDM.*

2-class mixture to 58.3% with 3-class mixture, for D3 from 83.1% with 6-class mixture to 75.1% with 7-class mixture, and for D4 from 66.9% with supervised learning to 52% with 2-class mixtures. The lack of robustness with regard to mixture complexity can be explained by different reasons. The high number of parameters to be assessed, compared to discriminative models, makes the model sensitive to complex data, i.e. it is inherent to the EM procedure when high dimensional data are considered. Besides, when feature distributions partially overlap classes, the unsupervised estimation of

class models is too complex and classification performances are affected.

Discriminative models do not only outperform the GM in terms of classification rate but they are also proven robust. NLDM outperforms the LDM for D2, D3 and D4. Among the different training methods, the Fisher-based procedure almost always outperform the optimization criterion (10). The optimization-based model is strongly dependent on the initialization. This is shown by UO-LDM and UO-NLDM results that consider coefficients that are uniformly initialized to one. In this case, classification rate reaches 33% with D1 and D2, and around 15% with D3 when the initialization with the Fisher criterion resorts to 81.3% for D1, 83.9% for D2 and 86.7% for D3. In most of case, the gradient-based optimization of criterion 10 from the initialization given by the Fisher procedure does not bring significant improvements.

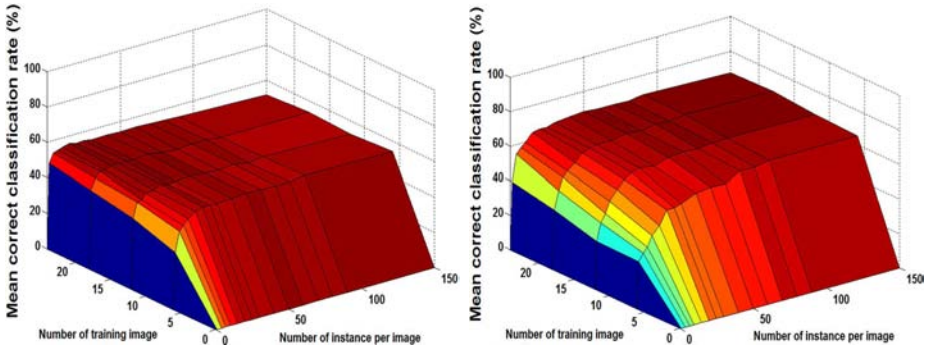
Overall, the best trade-off between the performance and the complexity of the training methods is the Fisher-based estimation. The choice of the non-linear kernel is relevant, as for datasets D2, D3, and D4, when objects classes are poorly discriminated in the original feature space.

Proportion complexity π_k	$\begin{pmatrix}0.63\\0.33\\0.03\end{pmatrix}$	$\begin{pmatrix}0.53\\0.33\\0.2\end{pmatrix}$	$\begin{pmatrix}0.46\\0.33\\0.25\end{pmatrix}$	$\begin{pmatrix}0.41\\0.33\\0.25\end{pmatrix}$	$\begin{pmatrix}0.36\\0.33\\0.29\end{pmatrix}$	$\begin{pmatrix}0.33\\0.33\\0.33\end{pmatrix}$
Prop GM	73.2%	64.4%	52.1%	46.8%	39.6%	34.2%
Pres GM	33.1%	33.3%	33.5%	33.3%	33.3%	33.3%
Prop FO-LDM	**81.3%**	**81.2%**	**80.23%**	**70.2%**	**56.2%**	**41.3%**
Pres FO-LDM	33.3%	33.3%	33.3%	33.3%	33.3%	33.3%
Prop FO-NLDM	80.3%	79.2%	78.2%	69.2%	47%	36.9%
Pres FO-NLDM	33.3%	33.3%	33.3%	33.3%	33.3%	33.3%

Table 2: *Comparison between proportion-based vs. presence/absence training for different kinds of object mixtures in the training image set considering D1. The mean correct classification rate is calculated for three-class mixtures in the training images. "Prop" refers to proportion mixture while "Pres" refers to presence/absence data. Results are reported for GM, FO-LDM and FO-NLDM.*

Proportion vs Presence/absence. The improvement brought by considering proportion-based training compared to presence/absence training (Bishop and Ulusoy (2005)) is shown in Table 2 for D1. The correct classification rate is shown as a function of the complexity of the training mixtures from the case in which one class predominates ($\pi_k = \begin{pmatrix}0.03 & 0.33 & 0.63\end{pmatrix}$) to the case with equal class proportions ($\pi_k = \begin{pmatrix}0.33 & 0.33 & 0.33\end{pmatrix}$). Note that the latter case corresponds to an unsupervised situation. This explains the fact that the mean correct classification rate equals about 33% in the presence/absence

case whatever the proportion mixture or the model. These results show that the knowledge of prior proportion-based information can greatly improve object recognition compared to presence/absence data. Non-surprisingly, the classification performances tend to decrease if the training tends to the unsupervised case (i.e. equal relative class proportions). It should be noted that for real applications (dataset D4) the training dataset is expected to contain a variety of mixture complexities among training images so that relevant classification performances can be reached.



(a) (b)

Figure 2: *Mean rate of correct classification for the synthetic object dataset D1 as a function of both the number of training images and the number of objects per training image. The GM (a) and the FO-LDM (b) are considered. Results are achieved from two-class training images.*

Image consistency. A key question arises regarding learning. Should there be lots of training images containing few objects or is the number of objects per training image decisive to achieve a correct classification rate? Figure 2 partially addresses this issue for dataset D1 with two-class training images: the evolution of the mean correct classification rate for the GM and the LDM is represented as a function of the number of training images and of the number of objects per training image. We first notice that for both models the maximal rate of classification is rapidly reached with the synthetic D1 data set. For the GM, the maximal correct classification rate is reached for 5 training images containing at least 5 objects. For the LDM, it is reached for

5 training images containing at least 20 objects. It seems that the primary condition to learn a reliable model is that the total number of objects in all training images must be high enough. Two cases may be advised: either lots of images with not many objects in the images or not many images comprising lots of objects. For the two scenarios, if the total number of objects is sufficient the maximal rate of classification should be reached.

5.4. Summary

The quantitative evaluation has been carried out using four datasets: a synthetic two-dimensional object feature dataset, two standard datasets and a fish school dataset. These dataset have been voluntarily chosen because they are different, so if the same trend is observed for all datasets, results can be generalized.

Reported results first show that the GM is outperformed by the proposed discriminative models which are more robust when the complexity of the training dataset increases. The proposed Fisher-based procedure for learning discriminative models is shown to greatly improve classification performances and robustness. We also showed the improvement brought by considering proportion-based training data compared to the presence/absence case, and then, we evaluated the effects of the number of training images and the number of the objects in training images on the classification performances.

These results were expected. Actually, last years discriminative models have shown their abilities to outperform generative models but in a supervised learning schemes (Schlkopf and Smola (2002)). In this paper, these results are confirmed and extended to the weakly supervised learning.

6. Conclusion

The development of reliable methods for object classification and recognition in images is an active area of research. In this paper, a probabilistic method is proposed to address weakly supervised learning with training information provided as the relative proportions of object classes in images. An application to fisheries acoustics data has been considered to learn fish school classifier in acoustic echograms from the estimation of the relative species in trawl catches. Our contribution is three-fold: the extension of generative and discriminative probabilistic models for proportion-based learning, a non-linear extension of the discriminative model and an efficient Fisher-based procedure for discriminative models.

In the condition of the proposed weakly supervised learning, quantitative experiments have shown that discriminative models are more robust and more accurate than generative model. Furthermore, regarding the application to fisheries acoustics, reported performances are compliant with an application to operational data.

Future work may investigate two methodological aspects: pursuing the development of classifiers for weakly supervised learning with an emphasis on high-dimensional feature space, and the introduction of contextual information in the classification of objects within a given image (Lefort et al. (2009)). Regarding the application to fisheries acoustics, future work will also include the use of new features issued from new multi beam sensors and the application to operational survey data.

Anderson, J., Holliday, D., Kloser, R., Reid, D., Simard, Y., 2008. Acoustic seabed classification: current practice and future directions. ICES Journal of Marine Science 65(6), 1004–1011.

Bahl, L., Brown, P., de Souza, P., Mercer, R., 1986. Maximum mutual information estimation of hidden markov model parameters for speech recognition. Int. Conf. on Acoustics Speech and Signal Processing., 49–52.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by probability distributions. Bull. Calcutta Maths. Soc. 35, 99–109.

Bishop, C. M., Ulusoy, I., 2005. Generative versus discriminative methods for object recognition. Conf. IEEE. on Computer Vision and Pattern Recognition 2, 258–265.

Blake, C., Merz, C., 1998. Uci repository of machine learning databases.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Chapman & Hall.

Chapelle, O., Schlkopf, B., Zien, A., 2006. Semi-supervised learning. MIT Press.

Crandall, D., Huttenlocher, D., 2006. Weakly supervised learning of part-based spatial models for visual object recognition. European Conf. on Computer Vision.

Dempster, A., Laird, N., Rubin, D., 1977. Likelihood from incomplete data via the em algorithm. Journal of the Royal Statistic Society Series B, 39(1), 1–38.

Descle, B., Bogaert, P., Defourny, P., 2006. Object-based method for automatic forest change detection. Remote Sensing of Environment 102, 1–11.

Fergus, R., Perona, P., Zissermen, A., 2007. Weakly supervised scale-invariant learning of models for visual recognition. International journal of computer vision 71(3), 273–303.

Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 179–188.

Gosselin, P., Cord, M., 2006. Feature-based approach to semi-supervised similarity learning. Pattern Recognition 39, 1839–1851.

Hinton, G., Sejnowski, 1999. Unsupervised learning: foundations of neural computation. MIT Press.

Lefort, R., Fablet, R., Karoui, I., Boucher, J., 2009. Combining image-level and object-level inference for weakly supervised object recognition. application to fisheries acoustics. Proc. Int. Conf. IEEE. on Image Processing, ICIP'09.

MacLennan, D. N., Simmonds, E. J., 1992. Fisheries acoustics. Chapman & Hall.

Scalabrin, C., Diner, N., Weill, A., Hillion, A., Mouchot, M.-C., 1996. Narrowband acoustic identification of monospecific fish shoals. ICES Journal of Marine Science 53 (2), 181–188.

Scalabrin, C., Mass, J., 1993. Acoustic detection of the spatial and temporal distribution of fish shoals in the bay of biscay. Aquatic Living Resources 6, 269–283.

Schlkopf, B., Smola, A., 2002. Learning with kernels. The MIT Press.

Vasconcelos, M., Carneiro, G., Vasconcelos, N., 2006. Weakly supervised top-down image segmentation. Conf. on Computer Vision and Pattern Recognition.

Weber, M., Welling, M., Perona, P., 2000. Unsupervised learning of models for recognition. European Conf. on Computer Vision 1, 18–32.

Xie, L., Perez, P., 2004. Slightly supervised learning of part-based appearance models. Conf. on Computer Vision and Pattern Recognition Workshop 6, 107.

Yang, Z., Zwolinski, M., 2001. Mutual information theory for adaptative mixture models. IEEE Trans. on Pattern Analysis and Machine Learning. 23(4), 396–403.