UNSUPERVISED RESTORATION IN GAUSSIAN PAIRWISE MIXTURE MODEL

Stéphane Derrode and Wojciech Pieczynski

École Centrale Marseille,
Institut Fresnel (CNRS UMR 6133),
38, rue F. Joliot-Curie,
F-13451 Marseille cedex 20, France.
stephane.derrode@centrale-marseille.fr
www.fresnel.fr/perso/derrode

Télécom SudParis,
CITI department (CNRS UMR 5157),
9, rue Charles Fourier,
F-91011 Evry cedex, France.
wojciech.pieczynski@it-sudparis.eu
www-public.int-evry.fr/˜pieczyn

ABSTRACT

The idea behind the Pairwise Mixture Model (PMM) we propose in this work is to classify simultaneously two sets of observations by introducing a joint prior between the two corresponding classifications and some inter-dependence between the two observations. We address the bayesian restoration of PMM using either MPM or MAP criteria, and an EM-based parameters estimation algorithm by extending the work done for classical Mixture Model (MM). Systematic experiments conducted on simulated data shows the effectiveness of the model when compared to the MM, both in supervised and unsupervised contexts.

1. INTRODUCTION

An important problem in signal processing consists in restoring an unobservable process $\mathbf{x} = \{x_n\}_{n\in\mathbb{N}}$ from an observed process $\mathbf{y} = \{y_n\}_{n\in\mathbb{N}}$. In the so-called probabilistic Mixture Model (MM), observations are assumed mutually independent and the restoration problem is to classify the data into a finite set of K classes. Coupled models appear when trying to co-analyse two series of observations, *ie* $\mathbf{y}^1$ and $\mathbf{y}^2$, to get two classifications denoted by *ie* $\mathbf{x}^1$ and $\mathbf{x}^2$ [6, 1].

In the field of image processing, such models can be of interest in several situations:

- The joint segmentation of two co-registered satellite radar and optical images. The number of observed classes can be different for the two modalities since the technologies to capture earth ground information are radically different. This problem can also be faced in medical imaging, *e.g.* T1, T2 and Proton-Density modalities from magnetic resonance imaging.
- The joint segmentation of two images with the same modality but some part of one image being hidden (missing data). This is the case when trying to compare two optical satellite images from the same site, but one of them has been acquired under cloudy conditions. Change detection after a major natural disaster also requires the co-segmentation of one old image of reference with a recent image showing the impact of, *e.g.*, earthquake, tornado or eruption damages.

The "Coupled Mixture Model" (CMM) in [6] was developed to co-analyse transcriptomic and proteomics sequences by mean of a joint prior on the mixtures components. In this work, we extend the CMM to get a "Pairwise Mixture Model" (PMM) by adding an inter-dependence between observations themselves, resulting in a rich mixture of bi-dimensional densities. We limit ourself to Gaussian densities, but the model can be easily parameterized with non-Gaussian distributions, *e.g.* with copula-based models [5, 2].

This paper addresses the bayesian restoration of PMM using either MPM or MAP criteria, and an EM-based [3] parameters estimation algorithm by extending the work done for classical MM. Section 2 presents the PMM and establishes two Bayesian decision rules. Section 3 derives an EM-based parameters estimation procedure. Section 4 deals with the simulation and restoration of data following a PMM. A systematic comparison between MM, CMM and PMM is performed, in both supervised and unsupervised contexts. Section 5 gives some segmentation results from a pair of images. Conclusion and further work are depicted in Section 6.

2. PAIRWISE MIXTURE MODEL

Let $\mathbf{y} = \{y_1,\dots,y_N\}$ denotes a set of N observed data ($y_n \in \mathbb{R}$). Let $\mathbf{x} = \{x_1,\dots,x_N\}$ denotes the result of classifying $\mathbf{y}$-data into a finite set of classes ($\Omega = \{1,\dots,K\}$).

In the classical probabilistic mixture model [7], one assumes that data y_n are realizations of mutually independent random variables Y_n with the same mixture distribution

$$f_{MM}(y_n) = \sum_{k=1}^{K} \pi_k\, f_k(y_n). \tag{1}$$

Each $f_k(.) = p(.|k)$ is a class k conditional probability density function (pdf) and priors $\pi_k = p(x_n = k)$ are such that $\sum_{k=1}^{K} \pi_k = 1$.

This model can be interpreted by assuming that observed data are distributed in K classes, sometimes called mixture components. The variable $X_n \in \Omega$ is an essential element of the problem which has the disavantage of not being able to be observed in practice (x_n is a missing or hidden data).

2.1 Pairwise mixture model

In a PMM, we consider two sets of N observed data $\mathbf{y}^1 = \left\{y_n^1\right\}_{n\in[1,N]}$ and $\mathbf{y}^2 = \left\{y_n^2\right\}_{n\in[1,N]}$. Let $\mathbf{x}^1 = \left\{x_n^1\right\}_{n\in[1,N]}$ and $\mathbf{x}^2 = \left\{x_n^2\right\}_{n\in[1,N]}$ denote the corresponding classifications into K and L classes respectively, *ie* $\forall n \in [1,N]$, $x_n^1 \in \Omega^1 = \{1,\dots,K\}$ and $x_n^2 \in \Omega^2 = \{1,\dots,L\}$. We also note $\mathbf{y}_n = \left(y_n^1, y_n^2\right)'$ and $\mathbf{x}_n = \left(x_n^1, x_n^2\right)'$ for latter use.

In this model we assume that pairwise data $\mathbf{y}_n$ are realizations of mutually independent random vectors $\mathbf{Y}_n$ with the

Table 1: Parameters used to draw mixtures in Fig. 2 and for simulations in Section 4.

	(k,l)	$\gamma_{k,l}$	$\mu_{k,l}^{(1)}$	$\sigma_{k,l}^{2(1)}$	$\mu_{k,l}^{(2)}$	$\sigma_{k,l}^{2(2)}$	$\rho_{k,l}$
PMM	$(1,1)$	0.14	0	2	1	2	0.8
	$(1,2)$	0.56	0.4	2	-0.6	2	0.3
	$(2,1)$	0.06	1.6	2	0.7	2	0.3
	$(2,2)$	0.24	2	2	-1	2	0.7
CMM	$(1,1)$	0.14	0	2	1	2	0
	$(1,2)$	0.56	0	2	-1	2	0
	$(2,1)$	0.06	2	2	1	2	0
	$(2,2)$	0.24	2	2	-1	2	0

same (finite) mixture distribution

$$f_{PMM}\left(y_n^1, y_n^2\right) = \sum_{k=1}^{K} \sum_{l=1}^{L} \gamma_{k,l}\, f_{k,l}\left(y_n^1, y_n^2\right), \qquad (2)$$

where $\gamma_{k,l} = p\left(x_n^1 = k, x_n^2 = l\right)$ is the joint prior with $\sum_{k=1}^{K}\sum_{l=1}^{L} \gamma_{k,l} = 1$. The dependence graph for such a model is given in Fig. 1(d). All 2D pdf $f_{k,l}(.,.)$ are assumed Gaussian in this work. Hence the set of PMM parameters is given by $\Theta = \left\{\Theta_{k,l}\right\}_{(k,l)\in\Omega^1\times\Omega^2}$ with $\Theta_{k,l} = \left(\gamma_{k,l}, \mu_{k,l}^{(1)}, \sigma_{k,l}^{(1)}, \mu_{k,l}^{(2)}, \sigma_{k,l}^{(2)}, \rho_{k,l}\right)$, where μ and σ denote the mean and standard deviation of Gaussian margins and ρ their correlation coefficient.

This model is different from a 2D mixture model (or vectorial MM) in which $\mathbf{y}_n$ is used to estimate one classification x_n only, *cf* Fig. 1(a). It is also more general than the product of two independent MMs defined by

$$f_{IMMs}\left(y_n^1, y_n^2\right) = \sum_{k=1}^{K} \pi_k^{(1)}\, f_k^{(1)}\left(y_n^1\right) \sum_{l=1}^{L} \pi_l^{(2)}\, f_l^{(2)}\left(y_n^2\right), \qquad (3)$$

and whose dependence graph is depicted in Fig. 1(b).

Remark: Let $\forall (k,l) \in \Omega^1 \times \Omega^2$, $\rho_{k,l} = 0$, *ie* $f_{k,l}\left(y_n^1, y_n^2\right) = f_{k,l}^{(1)}\left(y_n^1\right)\, f_{k,l}^{(2)}\left(y_n^2\right)$. Assuming further $f_{k,l}^{(1)}(.) = f_k^{(1)}(.)$ and $f_{k,l}^{(2)}(.) = f_l^{(2)}(.)$, we get the coupled mixture model (CMM) studied in [6]

$$f_{CMM}\left(y_n^1, y_n^2\right) = \sum_{k=1}^{K} p(k)\, f_k^{(1)}\left(y_n^1\right) \sum_{l=1}^{L} p(l\,|k)\, f_l^{(2)}\left(y_n^2\right). \qquad (4)$$

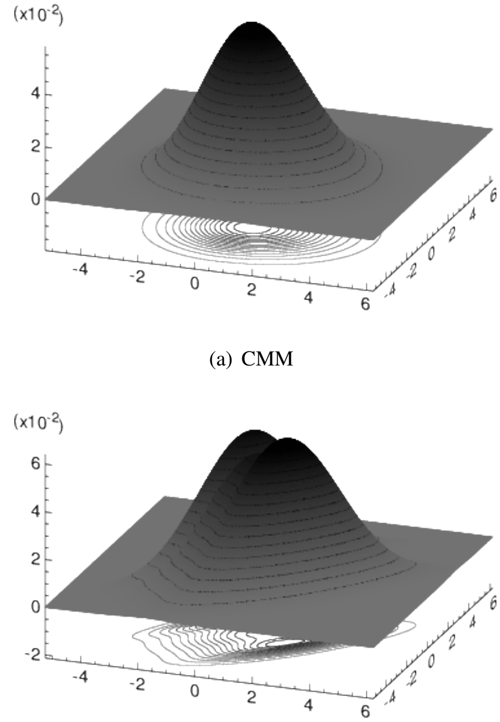
whose dependence graph is shown in Fig. 1(c).

Fig. 2 shows an example of $K = L = 2$ mixtures obtained from the CMM and PMM models with the parameters reported in Table 1.

2.2 Bayesian restoration in PMM

The estimation of $\mathbf{x}_n$ from $\mathbf{y}_n$ is done by performing a decision rule s. Such a rule is characterized by a cost function L which measures the error between $\mathbf{x}_n$ and its estimate $\hat{\mathbf{x}}_n$. The Bayesian estimator $\hat{\mathbf{x}}_n = s(\mathbf{y}_n)$ is the one that minimizes the mean cost:

$$\begin{aligned} \hat{\mathbf{x}}_n &= \arg\min_{\tilde{\mathbf{x}}_n\in\Omega^1\times\Omega^2} \mathbb{E}\left[L\left(\mathbf{x}_n, \tilde{\mathbf{x}}_n\right)|\mathbf{y}_n\right] \\ &= \arg\min_{\tilde{\mathbf{x}}_n\in\Omega^1\times\Omega^2} \sum_{\mathbf{x}_n\in\Omega^1\times\Omega^2} L\left(\mathbf{x}_n, \tilde{\mathbf{x}}_n\right) p\left(\mathbf{x}_n|\mathbf{y}_n\right) \end{aligned}$$



(a) CMM

(b) PMM

Figure 2: Example of CMM and PMM mixtures with parameters given in Table 1.

Two "0-1" cost functions L_1 and L_2 can be considered

$$\begin{aligned} L_1\left(\mathbf{x}_n, \tilde{\mathbf{x}}_n\right) &= \mathbb{1}_{x_n^1\neq\tilde{x}_n^1} + \mathbb{1}_{x_n^2\neq\tilde{x}_n^2} \\ L_2\left(\mathbf{x}_n, \tilde{\mathbf{x}}_n\right) &= \mathbb{1}_{\mathbf{x}_n\neq\tilde{\mathbf{x}}_n} = \mathbb{1}_{x_n^1\neq\tilde{x}_n^1 \text{ or } x_n^2\neq\tilde{x}_n^2} \end{aligned}$$

where $\mathbb{1}_c = 1$ if condition c is true, else 0. L_1 refers to the Maximum Posteriori Mode (MPM) and L_2 to the Maximum A Posteriori (MAP). After simple calculations, the two rules write:

$$MPM: \quad \begin{cases} \hat{x}_n^1 = \arg\max_{k\in\Omega^1} p\left(x_n^1 = k\,|\mathbf{y}_n\right) \\ \hat{x}_n^2 = \arg\max_{l\in\Omega^2} p\left(x_n^2 = l\,|\mathbf{y}_n\right) \end{cases} \qquad (5)$$

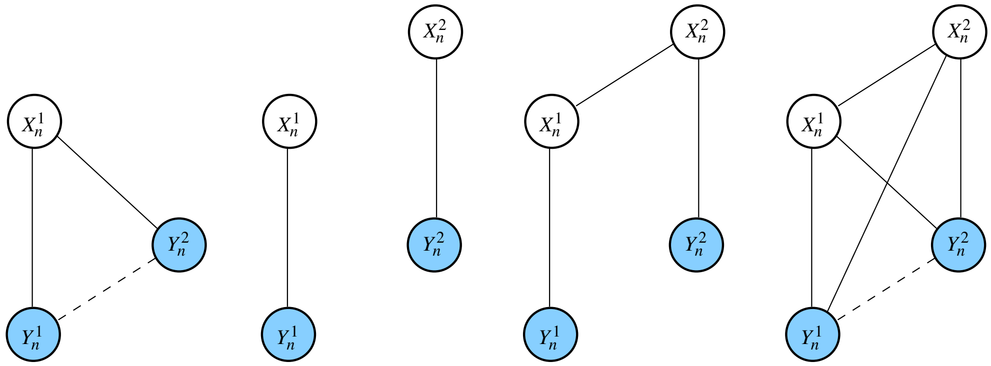
$$MAP: \quad \hat{\mathbf{x}}_n = \arg\max_{(k,l)\in\Omega^1\times\Omega^2} p\left(\mathbf{x}_n = (k,l)\,|\mathbf{y}_n\right) \qquad (6)$$

where $p\left(x_n^1 = k\,|\mathbf{y}_n\right) = \sum_{l=1}^{L} p\left(\mathbf{x}_n = (k,l)\,|\mathbf{y}_n\right)$ and $p\left(x_n^2 = l\,|\mathbf{y}_n\right) = \sum_{k=1}^{K} p\left(\mathbf{x}_n = (k,l)\,|\mathbf{y}_n\right)$. In the classical mixture model, the distinction between the MAP and MPM criteria does not exist.

3. PARAMETERS ESTIMATION

The aim of this section is to derive an EM-based parameters estimation algorithm [3, 7] for the Gaussian pairwise mixture model. Denoting Θ the true value of parameter θ that should be estimated and $\Theta^{(\ell)}$ an estimation of θ at iteration ℓ, EM-based estimation [4] consists in computing

$$\Theta^{(\ell)} = \arg\max_{\Theta} Q\left(\Theta^{(\ell-1)}, \Theta\right), \qquad (7)$$



(a) 2D MM (b) IMMs (c) CMM (d) PMM

Figure 1: Dependence graphs for the (a) 2D, (b) product, (c) coupled and (d) pairwise mixture models.

with

$$Q\left(\Theta^{(\ell-1)},\Theta\right)=\mathbf{E}\left[\underbrace{\ln H\left(\mathbf{x}^1,\mathbf{x}^2,\mathbf{y}^1,\mathbf{y}^2\right)}_{\mathscr{H}}\middle|\mathbf{y}^1,\mathbf{y}^2;\Theta^{(\ell-1)}\right]$$

recursively until convergence.

The joint log-likelihood of observations and hidden states writes

$$\begin{aligned}\mathscr{H} &= \sum_{n=1}^{N}\ln\gamma_{k,l}\,f_{k,l}\left(y_n^1,y_n^2\right)\\ &= \sum_{n=1}^{N}\ln\sum_{k=1}^{K}\sum_{l=1}^{L}\gamma_{k,l}\,f_{k,l}\left(y_n^1,y_n^2\right)\mathbb{1}_{\mathbf{x}_n=(k,l)}\\ &= \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L}\ln\left(\gamma_{k,l}\,f_{k,l}\left(y_n^1,y_n^2\right)\right)\mathbb{1}_{\mathbf{x}_n=(k,l)}\end{aligned}$$

Taking the expectation of $\mathscr{H}$ with respect to $\mathbf{x}^1$ and $\mathbf{x}^2$, we can write $\xi_n^{(\ell-1)}(k,l)=\mathbf{E}\left[\mathbb{1}_{\mathbf{x}_n=(k,l)}\middle|\mathbf{y}^1,\mathbf{y}^2;\Theta^{(\ell-1)}\right]$ in the following way

$$\begin{aligned}\xi_n^{(\ell-1)}(k,l) &= \sum_{i=1}^{K}\sum_{j=1}^{L}\mathbb{1}_{\mathbf{x}_n=(k,l)}p\left(\mathbf{x}_n=(i,j)\middle|\mathbf{y}^1,\mathbf{y}^2;\Theta^{(\ell-1)}\right)\\ &= p\left(\mathbf{x}_n=(k,l)\middle|\mathbf{y}^1,\mathbf{y}^2;\Theta^{(\ell-1)}\right)\\ &= p\left(\mathbf{x}_n=(k,l)\middle|y_n^1,y_n^2;\Theta^{(\ell-1)}\right)\\ &= \frac{p\left(\mathbf{x}_n=(k,l),\mathbf{y}_n;\Theta^{(\ell-1)}\right)}{p\left(\mathbf{y}_n;\Theta^{(\ell-1)}\right)}\end{aligned}$$

so that

$$\xi_n^{(\ell-1)}(k,l)=\frac{\gamma_{k,l}^{(\ell-1)}\,f_{k,l}^{(\ell-1)}(y_n^1,y_n^2)}{\sum_{i=1}^{K}\sum_{j=1}^{L}\gamma_{i,j}^{(\ell-1)}\,f_{i,j}^{(\ell-1)}(y_n^1,y_n^2)}. \tag{8}$$

Finally, we get

$$Q\left(\Theta^{(\ell-1)},\Theta\right)= \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L}\ln\left(\gamma_{k,l}^{(\ell-1)}\,f_{k,l}^{(\ell-1)}\left(y_n^1,y_n^2\right)\right)\xi_n^{(\ell-1)}(k,l).$$

According to (7), the next estimation is obtained by maximizing Q. Using standard calculations and following the approach developed for classical mixture models, we get

$$\gamma_{k,l}^{(\ell)} = \frac{1}{N}\sum_{n=1}^{N}\xi_n^{(\ell-1)}(k,l) \tag{9}$$

$$\mu_{k,l}^{(\ell)} = \frac{\sum_{n=1}^{N}\xi_n^{(\ell-1)}(k,l)\,\mathbf{y}_n}{\sum_{n=1}^{N}\xi_n^{(\ell-1)}(k,l)} \tag{10}$$

$$\mathbf{\Gamma}_{k,l}^{(\ell)} = \frac{\sum_{n=1}^{N}\xi_n^{(\ell-1)}(k,l)\,\left(\mathbf{y}_n-\mu_{k,l}^{(\ell)}\right)\,\left(\mathbf{y}_n-\mu_{k,l}^{(\ell)}\right)'}{\sum_{n=1}^{N}\xi_n^{(\ell-1)}(k,l)} \tag{11}$$

The entire algorithm for unsupervised data classification according to the PMM is sketched in Appendix A.

4. SIMULATION AND RESTORATION OF PMM

Given all model parameters, the simulation of pairwise data (y^1,y^2) and (x^1,x^2) such that (2) holds can be performed according to the following 4-steps procedure:

- draws x^1, according to $p\left(x^1\right)=\sum_{l=1}^{L}\gamma_{x^1,l}$;
- draws x^2, according to $p\left(x^2\left|x^1\right.\right)=\frac{\gamma_{x^1,x^2}}{p(x^1)}$;
- draws y^1, according to Gaussian margin $f_{x^1,x^2}^{(1)}\left(y^1\right)=p\left(y_1\left|x^1,x^2\right.\right)$;
- draws y^2, according to conditional Gaussian distribution $p\left(y^2\left|x^1,x^2,y^1\right.\right)$.

Table 2: Error rates for experiment in Section 4.1.

	MPM		MAP	
	$\mathbf{x}^1$	$\mathbf{x}^2$	$\mathbf{x}^1$	$\mathbf{x}^2$
PMM	17.46	14.04	17.71	13.91
CMM	22.66	17.94	23.10	18.38

Table 3: Parameters estimated by EM/PMM for experiment in Section 4.2. These values should be compared to the ones in Table 1, row one.

	(k,l)	$\gamma_{k,l}$	$\mu_{k,l}^{(1)}$	$\sigma_{k,l}^{2(1)}$	$\mu_{k,l}^{(2)}$	$\sigma_{k,l}^{2(2)}$	$\rho_{k,l}$
PMM	$(1,1)$	0.15	-0.02	2.01	0.97	2.02	0.95
	$(1,2)$	0.53	0.35	1.91	-0.61	2.08	0.44
	$(2,1)$	0.05	1.75	2.28	1.07	1.01	0.66
	$(2,2)$	0.27	1.94	1.93	-0.99	1.77	0.87

Iterating simulations N times, we get two series of observations $\mathbf{y}^1$ and $\mathbf{y}^2$, and their corresponding classifications $\mathbf{x}^1$ and $\mathbf{x}^2$. The latter are used for comparison with supervised and unsupervised estimates, by mean of classification error rates. For experiments we simulated $N = 10000$ pairwise samples using PMM parameters in Table 1 (first row).

4.1 Supervised restoration

We restored the two series of data using both MPM (5) and MAP (6) Bayesian criteria assuming first a PMM and then a CMM, with parameters given in Table 1. Error rates are reported in Table 2.

As expected the error rates are smaller when the PMM model is considered for restoration: the CMM is not able to capture the complexity of the PMM model.

4.2 Unsupervised restoration

In this experiment, we assume that model parameters are unknown and must be estimated before classification. Using the algorithm sketches in App. A with a number of EM iterations set to $\pounds = 1500$, we get the estimates reported in Table 3. The log-likelihood evolution against iterations is reported in Fig. 3. Evolution of coefficients $\gamma_{k,l}$ is plotted in Fig. 4. EM is converging very slowly. Nevertheless, estimated parameters are close to the true parameters, which produces error rates similar to the ones obtained from supervised classification, see Table 4.

5. APPLICATION TO JOINT IMAGES SEGMENTATION

In this section, the PMM model is used to segment jointly a couple of synthetic images with different number of classes.

Table 4: Error rates for experiment in Section 4.2.

	MPM		MAP	
	$\mathbf{x}^1$	$\mathbf{x}^2$	$\mathbf{x}^1$	$\mathbf{x}^2$
PMM	17.57	14.05	17.56	14.10

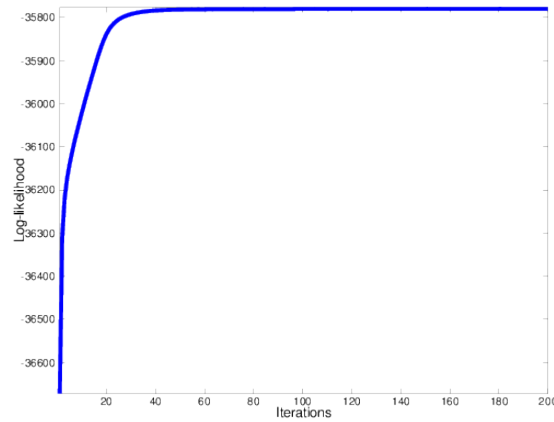


Figure 3: Log-likelihood evolution for experiment in Section 4.2.

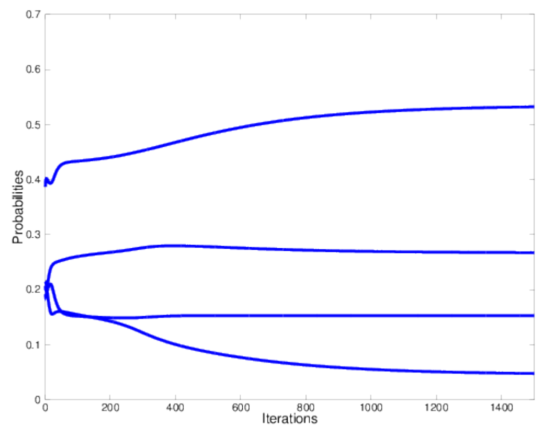


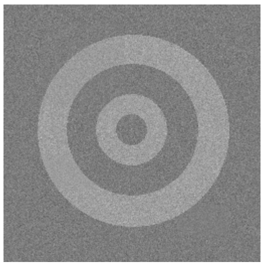
Figure 4: Evolution of $\gamma_{k,l}$ for $(k,l) \in \Omega^1 \times \Omega^2$ for experiment in section 4.2. Expected values are 0.06, 0.14, 0.24 and 0.56.

Fig. 5 shows original and noisy images alterated with Gaussian noises whose mean vectors have been set to $\mu_{1,1} = (125,120)$, $\mu_{1,2} = (127,162)$, $\mu_{1,3} = (126,136)$, $\mu_{2,1} = (157,120)$, $\mu_{2,2} = (159,160)$, and $\mu_{2,3} = (155,141)$, and covariance matrices to $\Gamma_{1,1} = \Gamma_{2,2} = \begin{pmatrix} 110 & 88 \\ 88 & 110 \end{pmatrix}, \Gamma_{1,2} = \Gamma_{2,1} = \begin{pmatrix} 50 & 22 \\ 22 & 50 \end{pmatrix}, \Gamma_{1,3} = \Gamma_{2,3} = \begin{pmatrix} 80 & 0 \\ 0 & 80 \end{pmatrix}$ which correspond to correlations of $\rho = 0.90$, $\rho = 0.44$ and $\rho = 0$.

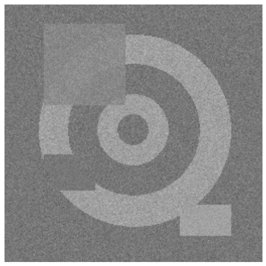
Fig. 6 shows the two independent unsupervised segmentation results obtained with a classical mixture model with 2 and 3 classes respectively. The unsupervised PMM segmentation of the two noisy images are reported in Fig. 7 for the MPM criterion only. Error rates are reported under each classification. Each time the PMM shows some improvements with respect to the IMMs.

(a) $K = 2$-classes image

(b) $L = 3$-classes image

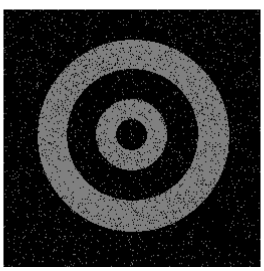


(c)

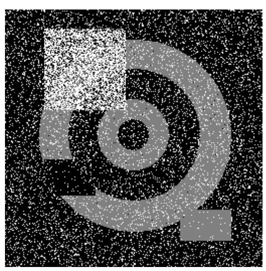


(d)

Figure 5: Original and noisy images used in Section 5.



(a) $K = 2$ - $\tau = 4.45\%$



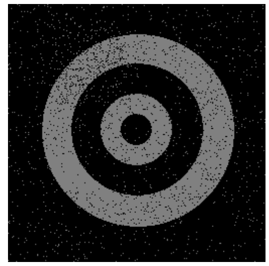
(b) $L = 3$ - $\tau = 15.09\%$

Figure 6: Result of unsupervised image segmentation using two independent MMs with $K = 2$ and $L = 3$ classes (τ: error rate).

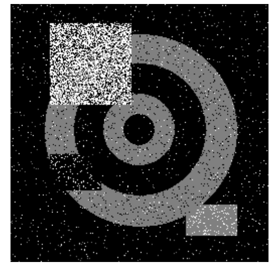
6. CONCLUSION

This work describes an algorithm for the joint classification of two series of observations with a bidimensional mixture model we called "Pairwise Mixture Model". This PMM is able to take into account the inter-dependence between states and between observations, resulting in a very rich mixture of components. We proposed an EM-based estimation procedure and two Bayesian criterion for restoration. Numerous experiments on simulated data and sythetic images confirm the interest of the model with respect to simpler models (independent MM and CMM [6]), even if the number of parameters to be estimated increases.

Next step will be to include a structure of Markov chain for $\mathbf{x}^1$ and for $\mathbf{x}^2$ in a way to extend the "Coupled Hidden Markov Model" (CHMM) in [1] in the same way this PMM extends the CMM. Another interesting work to develop concerns the replacement of Gaussian distributions with parametric models constructed from copulas [5].



(a) $K = 2$ - $\tau = 3.05\%$



(b) $L = 3$ - $\tau = 7.55\%$

Figure 7: Result of unsupervised image segmentation using the PMM with the MPM criterion (τ: error rate).

REFERENCES

[1] M. Brand. Coupled hidden Markov models for complex action recognition. In *Technical Report 405*, MIT Media Lab Perceptual Computing, June 1997.

[2] N. Brunel, W. Pieczynski, and S. Derrode. Copulas in vectorial hidden Markov chains for multicomponent image segmentation. In *Proc. of the ICASSP'05*, Philadelphia, USA, March 19-23 2005.

[3] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. of the Royal Statsitical Society, series B*, 39:1–38, 1977.

[4] G. McLachlan and T. Khrishnan. *The EM algorithm and extensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 2008.

[5] R. B. Nelsen. *An introduction to copulas*. Springer. Lecture Notes in Statistics, New York, USA, 1999.

[6] S. Rogers *et al*. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*, 24(24):2894–2900, 2008.

[7] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical analysis of finite mixture distributions*. Wiley Series in Probability and Statistics. John Wiley, 1985.

A. ALGORITHM

Require: $\mathbf{y}^1$, $\mathbf{y}^2$, K and L

1. Initialization of parameters at $\ell = 0$:

Classify $\mathbf{y}^1$ and $\mathbf{y}^2$ separately using a standard technique such as the k-means algorithm.

Use standard empirical estimators to get $\Theta^{(0)}$.

2. EM estimation.

for $\ell = 1$ to $\mathcal{L}$ **do**

Compute *a posteriori* probabilities $\xi_n^{(\ell-1)}(.,.)$ from (8).

Compute *a priori* probabilities $\gamma_{.,.}^{(\ell)}$ from (9).

Compute data-driven parameters $\mu_{.,.}^{(\ell)}$ et $\boldsymbol{\Gamma}_{.,.}^{(\ell)}$ from (10) and (11).

end for

3. Classification from $\Theta^{(\mathcal{L})}$

Compute MPM and MAP using (5) and (6).