HIDDEN MARKOV MODELS FOR WAVELET IMAGE SEPARATION AND DENOISING

Mahieddine M. ICHIR and Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes (UMR 8506 CNRS-Supélec-UPS)
Supélec, plateau de Moulon, 3 rue Joliot-Curie, 91192 Gif sur Yvette cedex, France
{ichir,djafari}@lss.supelec.fr

ABSTRACT

In this paper, we consider the problem of blind source separation of 2D images under a Bayesian formulation (Bayes-BSS). We transport the problem to the wavelet domain to be able to define appropriate prior distributions for the wavelet coefficients of the unobservable sources: an Independent Gaussians Mixture (IGM) model, a Hidden Markov Tree (HMT) model and Contextual Hidden Markov Field (CHMF) model.

Indeed, we consider a limiting case of the aforementioned prior models to propose a simple procedure for joint source separation and denoising. This procedure shows to be efficient, especially for highly noise observations. Simulation examples and comparisons with standard classical methods are presented to show the performances of the proposed approach.

1. INTRODUCTION

Blind source separation (BSS) finds its applications in many fields of data processing. It consists mainly in recovering *unobservable sources* from a set of their linear and instantaneous mixtures. It is mathematically described by:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{\epsilon}(t), \quad t = 1, \dots, T \tag{1}$$

where $\boldsymbol{A}$ is the mixing matrix, $\boldsymbol{x}(t)$ is an m-column vector of the observed data and $\boldsymbol{s}(t)$ is an n-column vector of the unobservable sources. $\boldsymbol{\epsilon}(t)$ is an m-column vector representing model uncertainties or observation noise (it is assumed to be a zero mean Gaussian process i.i.d. with a covariance matrix $\boldsymbol{R}_\epsilon = \text{diag}\left(\sigma_1^2, ..., \sigma_m^2\right)$).

Independent Component Analysis (ICA) [1] is the most developed solution for BSS. It consists in finding, from the model $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)$ (a noise free direct model), *independent* components that may represent the original unobserved sources. In noisy environments, ICA still identifies the mixing process (mixing matrix) for relatively high signal to noise ratios (provided that the original unobserved sources are independent).

In this paper we consider the Bayesian solution to the BSS problem. Bayesian estimation for BSS has been already addressed by many authors and joint separation and segmentation algorithms of 2D images have been derived as in [2].

In our approach, we transport the problem to the wavelet domain. The latter offers some natural and tractable models for a wide class of signals that can be exploited in a Bayesian formulation. We consider three prior models for the wavelet coefficients based on a two Gaussians mixture distribution: i) an independent model across and within scales, ii) a first order hidden Markov chain model to account for inter scale correlations, iii) a hidden Markov field model to account for both inter and intra scale correlations.

In order to achieve separation in the case of highly noisy observations, a limiting case of the two Gaussians mixture model is considered: the Bernoulli Gaussian (BG) model, where it can be shown that the BG model is equivalent to hard thresholding [3] in wavelet based denoising problems.

This paper is organized as follows: In section 2 we introduce the BSS in the wavelet domain and write the corresponding posteriors of the parameters of interest (the mixing matrix, the noise covariance matrix and the wavelet coefficients of the unobservable sources). We will then describe in details the priors of each one of these parameters in order to write this posterior. A simple and efficient procedure is presented in section 3 to perform source separation in high noisy data. A simulation example and comparisons with an ICA method[4] is presented in section 4 and we conclude in section 5.

2. WAVELET BASED BAYES-BSS

Because of the linearity of the wavelet transform, the BSS model (1) is equivalently written in the wavelet domain:

$$\boldsymbol{w}_x^\lambda = \boldsymbol{A}\boldsymbol{w}_s^\lambda + \boldsymbol{w}_{\boldsymbol{\epsilon}}^\lambda, \quad \lambda = (j, k_j) \tag{2}$$

where $\boldsymbol{w}_x^\lambda$ is a m-column vector of the k_j^{th} wavelet coefficients of the observations $\boldsymbol{x}(t)$ at resolution j (similarly for $\boldsymbol{w}_s^\lambda$ and $\boldsymbol{w}_{\boldsymbol{\epsilon}}^\lambda$). $j = 1, \dots, J$ and $k_j = 1, \dots, 2^{-j}T$. The double index λ means that the BSS problem is described in each wavelet subband[5] separately (having in common the same mixing matrix $\boldsymbol{A}$).

In Bayesian estimation, we need to explicit the *posterior* of the parameters of interest: the unobservable sources, the mixing matrix, the noise covariance matrix and the parameters describing the prior models (commonly called *hyper-parameters*). It is given by:

$$\begin{aligned} p(\boldsymbol{W}_s, \boldsymbol{A}, \boldsymbol{R}_\epsilon, \theta_s | \boldsymbol{W}_x) \propto p(\boldsymbol{W}_x | \boldsymbol{W}_s, \boldsymbol{A}, \boldsymbol{R}_\epsilon) \times ... \\ ... \times \pi(\boldsymbol{W}_s|\theta_s)\, \pi(\boldsymbol{A}|\theta_A)\, \pi(\boldsymbol{R}_\epsilon|\theta_\epsilon)\, \pi(\theta_s)\,, \quad (3) \end{aligned}$$

where $\boldsymbol{W}_x = \bigcup_\lambda \{\boldsymbol{w}_x^\lambda\}$ and $\boldsymbol{W}_s = \bigcup_\lambda \{\boldsymbol{w}_s^\lambda\}$. In equation (3) we assume that $\boldsymbol{W}_s$, $\boldsymbol{A}$ and $\boldsymbol{R}_\epsilon$ are *a priori* independent. The hyperparameters $(\theta_A, \theta_\epsilon)$ defining respectively the priors of $\boldsymbol{A}$ and $\boldsymbol{R}_\epsilon$ are fixed once for all, reducing the number of unknown variables. Only θ_s defining the prior for the wavelet coefficients of the sources will be inferred.

2.1. Mixing matrix prior distribution: $\pi(\boldsymbol{A}|\theta_A)$

The prior distribution of $\boldsymbol{A}$ can be given by a description of the physical mixing process. In this paper, and without loss of generality, we consider the elements of the mixing matrix (a_{ij}) to be Gaussian independent $\mathcal{N}(\mu_{ij}^a, \sigma_a^2)$.

2.2. Noise variance prior distribution: $\pi\big(\sigma_i^2|\theta_\epsilon\big)$

The noise variances σ_i^2 are assigned an Inverse Gamma prior [2] distribution of the form:

$$\pi\big(\sigma_i^2|\alpha_0, \nu_0\big) \propto \frac{1}{\sigma_i^{2(\alpha_0+1)}} \exp\bigg(-\frac{1}{\nu_0 \sigma_i^2}\bigg)\, \mathbb{I}_{\mathbb{R}_+}. \quad (4)$$

2.3. Prior distribution of $\boldsymbol{W}_s$

We will first assume that the sources are a priori mutually independent, thus their wavelet coefficients and write:

$$\pi(\boldsymbol{W}_s|\theta_s) = \prod_{i=1}^{n} \pi(W_{s_i}|\theta_{s_i})\,, \quad (5)$$

where $W_{s_i} = \bigcup_\lambda \{w_{s_i}^\lambda\}$ represents the wavelet coefficients of the source $S_i = \{s_i(1), \dots, s_i(T)\}$. Several properties[6] of the wavelet coefficients motivate the choice of an appropriate prior $\pi(W_{s_i}|\theta_{s_i})$:

P1. Compression: the wavelet transform is a sparse representation of a wide class of signals.

Property **P1** states that, for a large class of signals, the wavelet transform results in a *large* number of *small* coefficients and a *small* number of *large* coefficients. It is statistically described by *peaky* and *heavy tailed* distributions.

One such family of distributions, considered in [5, 7], are the generalized p-Gaussian distributions (gpG). This kind of priors allowed to establish connections between wavelet thresholding and Bayesian MAP estimation[7]. However, such distributions are non conjugate priors presenting optimization difficulties.

Another family of priors that captures efficiently the sparseness property is a two Gaussians mixture [6] of the form:

$$\pi\big(w_{s_i}^\lambda\big) = p_L^{(i)}\, \mathcal{N}(w_{s_i}^\lambda|0, \tau_L^{(i)}) + p_H^{(i)}\, \mathcal{N}(w_{s_i}^\lambda|0, \tau_H^{(i)}), \quad (6)$$

with $\tau_L << \tau_H$ and $p_L = 1 - p_H$. $p_{L/H} = \text{Prob}\big($wav. coeff. $\in$ low/high energy state$\big)$. This prior model classifies the wavelet coefficients into two classes: *low energy* coefficients and *high energy* coefficients.

In order to go further in the description of the prior model, we introduce binary *hidden* random variables associated to the wavelet coefficients of the sources in order to rewrite the distribution (6) under the form:

$$\pi\Big(w_{s_i}^\lambda | z_i^\lambda = q, \tau_q^{(i)}\Big) = \mathcal{N}(w_{s_i}^\lambda|0, \tau_q^{(i)}), \quad (7)$$

where $q \in \{L, H\}$ and $\pi\big(z_i^\lambda = q\big) = p_q$. We will now detail three different possible models for these hidden variables.

2.3.1. *Independent Gaussian Mixture model (IGM)*

A simple model for the wavelet coefficients is an independent model across and within scales:

$$\pi(Z_i) = \prod_\lambda \pi\big(z_i^\lambda\big)\,, \quad (8a)$$

$$\pi(W_{s_i}|Z_i) = \prod_\lambda \pi\Big(w_{s_i}^\lambda | z_i^\lambda = q, \tau_q^{(i)}\Big)\,. \quad (8b)$$

where $Z_i = \bigcup_\lambda \{z_i^\lambda\}$ and $W_{s_i} = \bigcup_\lambda \{w_{s_i}^\lambda\}$. The IGM model is a simple model but lacking local homogeneity. In order to enhance this model, additional properties[6] of the wavelet coefficients are considered. To go further in the description of the next two models, notations have to be fixed in conjunction with Fig. 1:

. $\mathcal{P}_\lambda$ represents a parent (ascendent) node of the node λ, while $\mathcal{C}_\lambda$ represents its set of child nodes (descendents).

. ν_λ is a set of neighboring nodes of λ (in this paper we consider only first order neighboring systems).

Fig. 1 concerns the dyadic wavelet transform of 1D signals, its generalization to 2D images is straightforward.

2.3.2. *Hidden Markov Tree model (HMT)[6]*

The wavelet coefficients have been described to be *persistent*, which means they tend to propagate across scales. This suggests to account for *inter* scale correlations between the hidden variables (Z_i) through a first order Markov chain:

$$p\big(z_i^\lambda = q\big) = \sum_{q'} p\Big(z_i^{\mathcal{P}_\lambda} = q'\Big)\, p\Big(z_i^\lambda = q | z_i^{\mathcal{P}_\lambda}\Big)\,, \quad (9)$$

with $\{q, q'\} \in \{L, H\}$.

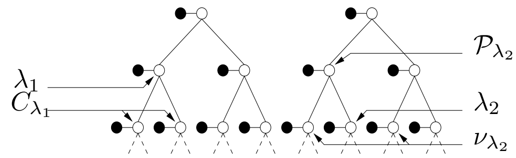


Fig. 1. Quad tree representation of the wavelet coefficients for a 1D signal dyadic wavelet decomposition: • Wavelet coefficients ○ Associated hidden variables.

2.3.3. Contextual Hidden Markov Field model (CHMF)

An additional property enhances a step more the HMT model accounting for intra scale correlations:

P3. Clustering: the wavelet coefficients tend to be locally correlated.

We propose to statistically model this property by a Markov random field on the hidden variables:

$$\pi\big(z_i^\lambda = q\big) \propto \exp\big(\beta_1 \sum_{r\in\nu_\lambda} \delta_{z_i^r}(q) + \beta_2 \sum_{r\in\mathcal{C}_\lambda} \delta_{z_i^r}(q)\big), \quad (10)$$

where β_1 and β_2 are heuristically determined constants.

3. JOINT SOURCE SEPARATION AND DENOISING

In the case of highly noisy observations, and in order to be able to separate the sources, additional information is considered: *a wide class of signals can be well approximated by only a few number of their wavelet coefficients*. In wavelet based denoising, connections between hard/soft wavelet thresholding and Bayesian estimation have been established in [3, 7], especially for the gpG and the Bernoulli-Gaussian (BG) prior model. The BG model is given by:

$$\pi\big(w_{s_i}^\lambda\big) = p_L^{(i)}\delta(0) + (1-p_L^{(i)})\mathcal{N}(w_{s_i}^\lambda|0,\tau_H^{(i)}). \quad (11)$$

It is in fact a limiting case of the two Gaussians mixture model of equation (6). We can easily prove the following:

lemma 1 *The Maximum A Posteriori (MAP) estimate* $\hat{w}_s^\lambda$ *of* w_s^λ *in a denoising problem:* $w_x^\lambda = w_s^\lambda + w_\epsilon^\lambda$*, with* $w_\epsilon^\lambda \sim \mathcal{N}(0,\sigma^2)$*, where the prior distribution of* w_s^λ *is given by equation (6), is:*

$$\hat{w}_s^\lambda = \begin{cases} \frac{\hat{\tau}_L}{\sigma^2} w_x^\lambda, & \textit{if} \quad z^\lambda = L, \\ \frac{\hat{\tau}_H}{\sigma^2} w_x^\lambda, & \textit{if} \quad z^\lambda = H. \end{cases} \quad (12)$$

where $1/\hat{\tau}_q = 1/\sigma^2 + 1/\tau_q$*. If* $(\tau_L \to 0)$*, the prior tends to (11) and the MAP estimate is given by:*

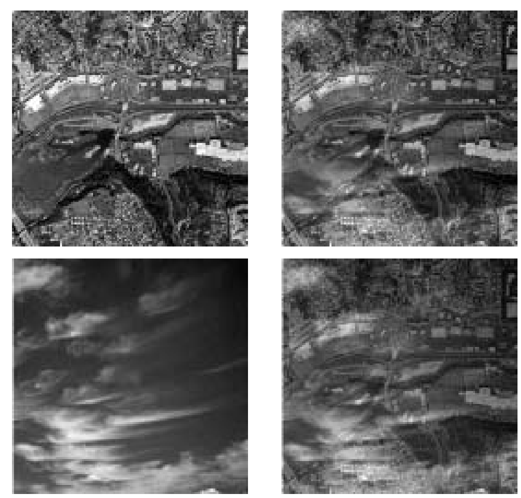
$$\hat{w}_s^\lambda = \begin{cases} 0, & \textit{if} \quad z^\lambda = L, \\ \frac{\hat{\tau}_H}{\sigma^2} w_x^\lambda, & \textit{if} \quad z^\lambda = H. \end{cases} \quad (13)$$

defining a hard thresholding on w_x^λ. ■

The BG prior model of equation (11) is then adopted for the wavelet coefficients, in the case of highly noisy observations in order to implement a joint separation and denoising procedure. The three prior models for the hidden variables described earlier are similarly adopted in that case.

4. SIMULATIONS RESULTS

An MCMC stochastic based algorithm have been implemented for the optimization purpose with adapted sampling procedures for each prior model. Simulations have been made on aerial gray scale images of Fig. 2-a. The mixing matrix $\boldsymbol{A} = [[1,\ .9],[.5,\ 1]]$ and the noise free observations are shown on Fig. 2-b.



a. b.

Fig. 2. a. 128×128 original source images, b. Noise free mixed images with $\boldsymbol{A} = \big[[1,\ .9],[.5,\ .1]\big]$.

For this data set, we have used the "Symmlet" wavelets with 6 vanishing moments, these wavelets are highly symmetrical. The 128×128 observation images of Fig. 2-b have been decomposed up to the 3$^{\text{rd}}$ wavelet scale.

In Fig. 3, the evolution of a normalized L_1 norm error $\delta_s(i) = 10\log_{10}(\|\hat{s}_i - s_i\|_1/\|s_i\|_1)$ in terms of the signal to noise ratio $\text{SNR}_{\text{dB}} = 20\log_{10}(\|[\boldsymbol{A}\boldsymbol{s}]_i\|_2/\sigma)$ is presented for the three different prior models (IGM, HMT and CHMF). For comparison, Fig. 3 shows also the results with a classical ICA method (JADE[4]).

We see from Fig. 3 that the proposed approach outperforms the classical ICA method, at least for this data set. Accounting for inter scale correlations with the HMT model doest not improve the results for the aerial image of Fig. 2-a (top), this is mainly due to its white behaviour on many portion of the image, however, accouting for intra and inter

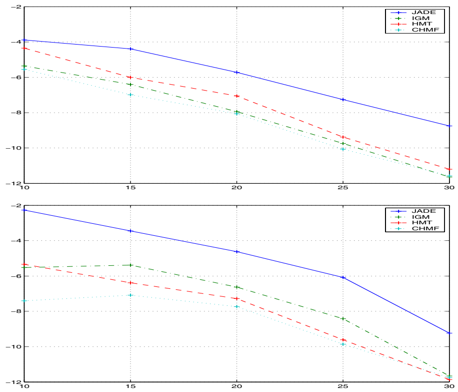


Fig. 3. Δ_s vs. SNR for source 1 (top), source 2 (bottom) obtained with: ICA (continuous line), IGM model (dotted-dashed line), HMT model (dashed line) and CHMF model with $\beta_1 = .6, \beta_2 = .7$ (dotted line).

scale correlations through the CHMF model improves the resulting estimates for both the aerial and the cloud image.

At low SNR's ($\lesssim$ 12dB), the outlined method, with the standard two Gaussians mixture model of equation (6), does not manage to perform any separation. In that context, the Bernoulli Gaussian prior of equation (11) has been considered to jointly separate and denoise. On Fig.4, we show the resulting estimates for a SNR $\approx$ 10dB with the three prior models (IGM, HMT and CHMF).

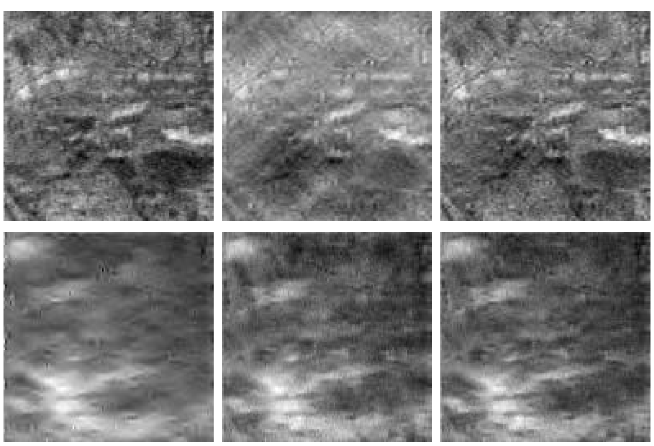


Fig. 4. The obtained separated sources for SNR$\approx$ 10dB with the IGM model (left), HMT model (middel) and CHMF model with $\beta_1 = .6, \beta_2 = .7$ (right).

What we notice, is that the HMT model does not achieve a good separation at such low SNR's for images having a frequency content similar to that of the cloud image of Fig. 2-a (top). However, for the cloud image of Fig. 2-a (bottom), both the HMT and CHMF presents better reconstructed images than that obtained with the IGM model, in the sense that they are less smoothed and contours are well preserved.

5. SUMMARY AND CONCLUSION

In this work, we addressed the problem of blind source separation in the wavelet domain under a Bayesian formulation. We considered three prior models for the wavelet coefficients:
. the Independent two Gaussians mixture model (IGM),
. the hidden Markov tree model (HMT),
. the contextual hidden Markov field model (CHMF).

Simulations have been performed and compared to a classical ICA method, where we see from Fig. 3 that the presented approach outperforms the ICA approach.

A second prior model (the Bernoulli Gaussian model) has been proposed to perform efficient joint source separation and denoising whenever the noise is highly present in the data, this method seems to be a promising approach especially for the Contextual Hidden Markov Filed (CHMF) model.

6. REFERENCES

[1] Pierre Comon, "Independant component analysis, a new concept ?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.

[2] Hichem Snoussi, *Bayesian approach to source separation. Applications in imagery*, Ph.D. thesis, University of Paris–Sud, Orsay, France, september 2003.

[3] Jean-Christophe Pesquet, Hamid Krim, and E. Hamman, "A Bayesian approach to best basis selection," Atlanta, GA, 1996, Int. Conf. Acoust. Specch, Signal Process. ICASSP, pp. 2634–2637.

[4] Jean François Cardoso, "Higher-order contrasts for independant component analysis," in *Neural Computation*, MIT Letters, pp. 157–192. 1999.

[5] Stephane G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[6] Matthiew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, April 1998.

[7] Pierre Moulin and Juan Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, pp. 909–919, April 1999.