SOURCE SEPARATION IN IMAGES VIA MRFS WITH VARIATIONAL APPROXIMATION †

Koray Kayabol[1], *Bulent Sankur*[2], *and Ercan E. Kuruoglu*[3]

[1]Electrical and Electronics Engineering Dept., Istanbul University,
Avcilar, 34320, Istanbul, Turkey
email: kayabol@istanbul.edu.tr
[2]Electrical and Electronics Engineering Dept., Bogazici University,
Bebek, 80815, Istanbul, Turkey
email: bulent.sankur@boun.edu.tr
[3]Instituto di Scienza e Tecnologie dell'Informazione, CNR,
via G. Moruzzi 1, 56124, Pisa, Italy
email: ercan.kuruoglu@isti.cnr.it

ABSTRACT

The problem of source separation in two dimensions is studied in this paper. The problem is formulated in the Bayesian framework. The sources are modelled as MRFs to accommodate for the spatially correlated structure of the sources, which we exploit for separation in 2D. The difficulty of working analytically with general Gibbs distributions is overcome by using an approximate density. In this work, the Gibbs distribution is modelled by the product of directional Gaussians. The sources are estimated by Maximum-a-Posteriori estimation using the approximate density as the prior. At each iteration of the MAP estimation, an annealing schedule is used for approximate density. This annealing schedule aids the algorithm to converge the global extremum. The mixing matrix is found by Maximum Likelihood estimation.

1. INTRODUCTION

The aim of Blind Source Separation (BSS) is to reconstruct L independent sources from K observations which are mixed by different proportions. Some classical applications of source separation include audio separation and biomedical signal separation (e.g fMRI). In this study, we consider image separation problem. An example of image separation problem is found in astrophysics [3].

One of the two main approaches for source separation is learning the mixing matrix A which maximizes the mutual information between the sources. The second approach is maximization of the likelihood of A. These two methods are equivalent [1]. In this study, we follow the second approach to find the mixing matrix A and cast the source estimation problem in a Bayesian framework.

Since the observation noise, the mixing matrix and the model parameters of sources are unknown, BSS problem can be separated into two parts 1) Learning the parameters of observations and sources from observations, 2) Estimation of the sources. The parameters of observations are the mixing matrix A and the covariance matrix Σ of normally distributed zero mean noise. These parameters can be estimated by maximizing the likelihood $p(\mathbf{y}|\mathbf{A},\boldsymbol{\Sigma})$.

Before the identification of sources parameters, it is required to choose source models. In BSS, the separability of the mixed sources depends on the probability distribution of sources which are required to being non-Gaussian. An important generic model for probability density function is Gaussian mixtures which have been used with success in various applications. The parameters of mixture of Gaussians densities are obtained using Expectation-Maximization (EM) method by Attias [2]. Variational EM was used by Miskin [1] for the same purpose.

Most of the studies in source separation literature have not made use of the time (or space) structure in signals and have instead considered them only as data sets. Kuruoglu et al. [3], to be able to exploit the rich information content of the images, have proposed a MRF formulation and have used edge preserving regularizers to model pixel by pixel interactions. The ML estimation is used for parameters of the mixing matrix. Sources and parameters of the mixing matrix are optimized by cycling through the variables. Tonazzini et al. [4] applied mean field approximation to MRF and used EM algorithm to estimate the parameters of the mixing matrix. MRF models generally make use of the Gibbs formulation, i.e. the probability density function expressed in an "energy potential" form. The edge preserving property of the Gibbs distribution arises from the non-convex energy potentials. The resulting non-convexity makes the deterministic optimization difficult and the convergence is not guarantied. In this study, this difficulty is bypassed by approximating the Gibbs distribution by the product of Gaussians. The first order clique potentials at each direction are modeled by a Gaussian. At each pixel location, the product of these directional Gaussians forms its independent pixel density. Consequently, since each pixel is independent, the source image density is obtained by multiplying all these independent densities. This approximation with Gaussians forms a variance field. As the variance field is locally adaptive to edges, the edge preserving property of the Gibbs priors are inherited by the approximate priors. We also use an annealing schedule to ensure to convergence of the algorithm to global or a near global solution. At each step, the approximate Gaussians are shrunk by reducing the temperature. The details of this type of annealing can be found in [11, 12].

†THIS WORK WAS SUPPORTED BY CNR-TUBITAK. PROJECT NUMBER: 104E101.

2. PROBLEM DEFINITION IN THE BAYESIAN FRAMEWORK

The j'th observed image is defined as $y_j(m,n)$, where $(m,n) \in \{1,2,\ldots,M\} \times \{1,2,\ldots,N\}$ are spatial coordinates. It is assumed that this image is formed by the superposition of L independent sources, $s_i(m,n)$, $i = 1,\ldots,L$. The image separation problem is to find L independent sources from K different observation. So the observation model is given as

$$\begin{bmatrix} y_1(m,n) \\ y_2(m,n) \\ \vdots \\ y_K(m,n) \end{bmatrix} = \mathbf{A} \begin{bmatrix} s_1(m,n) \\ s_2(m,n) \\ \vdots \\ s_L(m,n) \end{bmatrix} + V \qquad (1)$$

where $\mathbf{A}$ is a $K \times L$ mixing matrix with elements $a_{j,i}$. V is zero mean noise which is independently occur at every pixel and has Σ covariance matrix.

The $\mathbf{s}_i$ and $\mathbf{y}_j$ are $MN \times 1$ vector representation of source and observation images respectively. In this case, the observation model can be written as $\mathbf{Y} = \mathbf{AS} + \mathbf{V}$

$$\begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_K^T \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{s}_1^T \\ \mathbf{s}_2^T \\ \vdots \\ \mathbf{s}_L^T \end{bmatrix} + \mathbf{V} \qquad (2)$$

Since the sources are assumed to be independent and the joint probability density is factorized as

$$p(\mathbf{S}) = \prod_{l=1}^{L} p(\mathbf{s}_l) \qquad (3)$$

The images provide us with significant structural information, which we would like to include in the separation process in this work using statistical priors. The BSS problem can be modelled in the Bayesian framework and $\mathbf{s}$, $\mathbf{A}$ and $\mathbf{\Sigma}$ can be found by maximizing the posterior density. The MAP estimate in this case become such that

$$\max_{\mathbf{s},\mathbf{A},\mathbf{\Sigma}} p(\mathbf{s},\mathbf{A},\mathbf{\Sigma}|\mathbf{y}) = \max_{\mathbf{s},\mathbf{A},\mathbf{\Sigma}} p(\mathbf{y}|\mathbf{s},\mathbf{A},\mathbf{\Sigma}) p(\mathbf{s}) p(\mathbf{A}) p(\mathbf{\Sigma}) \qquad (4)$$

The MAP estimate given in (4) is separated into two consecutive maximization steps. First of them is Maximum Likelihood (ML) estimation for mixing matrix parameters and noise covariance, and the second one is the MAP estimation for the sources. The prior densities of parameters are assumed to be uniform.

$$\mathbf{s}^{k+1} = \max_{\mathbf{s}} \{p(\mathbf{y}|\mathbf{s},\mathbf{A}^k,\mathbf{\Sigma}^k) p(\mathbf{s})\} \qquad (5)$$

$$\mathbf{A}^{k+1},\mathbf{\Sigma}^{k+1} = \max_{\mathbf{A},\mathbf{\Sigma}} \{p(\mathbf{y}|\mathbf{s}^k,\mathbf{A},\mathbf{\Sigma})\} \qquad (6)$$

One of the difficulties in MAP estimation arises from non-Gaussian prior densities. Maximization task is difficult because the non-Gaussian priors disturb the convexity. In this case , simulated annealing or MCMC based numerical Bayesian methods such as Gibbs sampling can be used. These are relaxation type methods, hence they take a long time.

In this study, the difficult priors are approximated using variational approach by tractable densities. In variational approach, intractable densities are approximated to tractable ones. The chosen tractable priors should be easy to work with analytically and their log should be convex. Let's denote the tractable prior density as $q(\mathbf{s}|\tau)$ where τ is its parameter set. In this case, the steps given in (5) and (6) become

$$\tau^{k+1} = \min_{\tau} D_{KL}(q(\mathbf{s}^k|\tau)||p(\mathbf{s}^k)) \qquad (7)$$

$$\mathbf{s}^{k+1} = \max_{\mathbf{s}} \{p(\mathbf{y}|\mathbf{s},\mathbf{A}^k,\mathbf{\Sigma}^k) q(\mathbf{s}|\tau^k)\} \qquad (8)$$

$$\mathbf{A}^{k+1},\mathbf{\Sigma}^{k+1} = \max_{\mathbf{A},\mathbf{\Sigma}} \{p(\mathbf{y}|\mathbf{s}^k,\mathbf{A},\mathbf{\Sigma})\} \qquad (9)$$

where $D_{KL}(q(\mathbf{s}^k|\tau)||p(\mathbf{s}^k))$ is the Kullback-Leibler (KL) divergence between $p(\mathbf{s}^k)$ and $q(\mathbf{s}^k|\tau)$ and defined as

$$D_{KL}(q(\mathbf{s}^k|\tau)||p(\mathbf{s}^k)) = \int q(\mathbf{s}^k|\tau) \log\left(\frac{q(\mathbf{s}^k|\tau)}{p(\mathbf{s}^k)}\right) d\mathbf{s} \qquad (10)$$

The density q is determined by learning the parameter τ such that KL divergence is minimum. In the following MAP estimate of the sources, we use this tractable density.

3. LEARNING THE PARAMETERS

The likelihood of mixing matrix $\mathbf{A}$ and observation noise $\mathbf{\Sigma}$ can be written over the unobserved variable $\mathbf{s}$ as

$$p(\mathbf{y}|\mathbf{A},\mathbf{\Sigma}) = \int p(\mathbf{y}|\mathbf{s},\mathbf{A},\mathbf{\Sigma}) p(\mathbf{s}) d\mathbf{s} \qquad (11)$$

To find the parameters $\mathbf{A}$ and $\mathbf{\Sigma}$ which maximize the (11) with $\theta = \{\mathbf{A},\mathbf{\Sigma}\}$, logarithm of (11) can be written as

$$\begin{aligned} \log\{p(\mathbf{y}|\theta)\} &= \int \log\{p(\mathbf{y}|\theta)\} p(\mathbf{s}|\mathbf{y},\theta) d\mathbf{s} \qquad (12) \\ &= E_{\mathbf{s}}[\log\left\{\frac{p(\mathbf{y}|\mathbf{s},\theta)p(\mathbf{s})}{p(\mathbf{s}|\mathbf{y},\theta)}\right\}|\mathbf{y},\theta] \end{aligned}$$

At first step, using the parameters, θ', obtained from the previous step, a cost function $C(\theta) = E_{\mathbf{s}}[-\log p(\mathbf{y}|\theta)]$ is constituted by taking expectation. At second step the parameters are found by minimizing this cost function. At each step, EM algorithm converges to minimum of the cost function. According to Jensen's inequality the upper bound is

$$C(\theta) \leq \int -\log\left\{\frac{p(\mathbf{y}|\mathbf{s},\theta)p(\mathbf{s})}{p(\mathbf{s}|\mathbf{y},\theta')}\right\} p(\mathbf{s}|\mathbf{y},\theta') d\mathbf{s} \qquad (13)$$

Analytic calculation of expectation is not possible except the case when the densities are Gaussian. If the densities are modelled as mixture of Gaussians, computational complexity increases as the number of mixtures. A solution to this problem is to use variational approximation methods. The purpose of variational methods is to approximate the intractable densities with tractable ones.

In the case $p(\mathbf{s}|\mathbf{y},\theta)$ an intractable density, its tractable version $q(\mathbf{s}|\mathbf{y},\theta,\tau)$ is used instead of $p(\mathbf{s}|\mathbf{y},\theta)$.

$$
\begin{aligned}
C(\theta) &\leq \int -\log\left\{\frac{p(\mathbf{y}|\mathbf{s},\theta)p(\mathbf{s})}{q(\mathbf{s}|\mathbf{y},\theta,\tau)}\right\} q(\mathbf{s}|\mathbf{y},\theta,\tau)d\mathbf{s} \quad (14)\\
&= -I_O(p(\mathbf{y}|\mathbf{s},\theta)q(\mathbf{s}|\mathbf{y},\theta,\tau);q(\mathbf{s}|\mathbf{y},\theta,\tau))\\
&\quad +D_{KL}(p(\mathbf{s})||q(\mathbf{s}|\mathbf{y},\theta,\tau))\\
&\quad -I_H(q(\mathbf{s}|\mathbf{y},\theta,\tau);q(\mathbf{s}|\mathbf{y},\theta,\tau))
\end{aligned}
$$

where I_O is cross entropy between $p(\mathbf{y}|\mathbf{s},\theta)q(\mathbf{s}|\mathbf{y},\theta,\tau)$ and $q(\mathbf{s}|\mathbf{y},\theta,\tau)$ and I_H is entropy of $q(\mathbf{s}|\mathbf{y},\theta,\tau)$.

This cost function can be solved by using EM algorithm. We divide the EM process into two separate processes. Firstly, D_{KL} term is minimized respect to τ which is given in (7). Secondly using the parameter τ obtained from previous step, I_O is minimized with respect to $\mathbf{A}$ and $\boldsymbol{\Sigma}$. In the expectation step MAP estimate of $\mathbf{s}$ is used instead of $E_q[\mathbf{s}]$. So the algorithm reduces to three steps given in (7-9). Entropy term I_H is not used in optimization process as in the EM algorithm.

4. OBSERVATION MODEL

Since the expectation $E_q[\mathbf{s}]$ and MAP estimate coincide, maximization of cross entropy term $I_O = E_q[\log p(\mathbf{y}|\mathbf{s},\theta)q]$ in equation (14) with respect to $\mathbf{A}$ becomes equal to the ML estimate of $\mathbf{A}$ given previous MAP estimate of $\mathbf{s}$, which is found by maximizing the following density

$$
p(\mathbf{y}|\mathbf{s},\theta) \propto \exp\left\{-\sum_{j=1}^{K}\frac{(\mathbf{y}_j-\sum_{i=1}^{N}a_{j,i}\mathbf{s}_i)^T(\mathbf{y}_j-\sum_{i=1}^{N}a_{j,i}\mathbf{s}_i)}{2\sigma^2}\right\} \quad (15)
$$

The element-wise update equation is found as

$$
a_{k,l} = \frac{1}{\mathbf{s}_l^T\boldsymbol{\Sigma}_N^{-1}\mathbf{s}_l}\left[\mathbf{s}_l^T\boldsymbol{\Sigma}_N^{-1}(\mathbf{y}_k-\sum_{i=1,i\neq l}^{L}a_{k,i}\mathbf{s}_i)\right] \quad (16)
$$

This Iterative Coordinate Decent type update is more stable in convergence than one step update of entire matrix.

5. SOURCES MODELS

The KL divergence between the density of sources $p_{\mathbf{S}}$ and its approximation $q_{\mathbf{S}}$ is minimized to find $q_{\mathbf{S}}$. If this density is chosen to be a parametric density, the optimization is reduced to the determination of its parameter. Sources are modelled to be MRF and the density $p_{\mathbf{S}}$ is chosen to be Gibbs distribution with non-convex energy potential functions. The approximate density $q_{\mathbf{S}}$ is selected to be factor of Gaussians.

MRF is a local statistical image model. It is noted that cliques are formed by only the first order neighbor pixel pairs. The (m,n) and (k,l) represent the coordinates of two adjacent pixels, and $\mathcal{N}_1((k,l))$ represents the set of first order neighbors of (k,l). The entire cliques set can be defined as $\mathscr{C} = \{\{(m,n),(k,l)\}|(m,n)\in\mathcal{N}_1((k,l))\}$.

The potential energy function is expressed as summation of all cliques potentials.

$$
U(s) = \frac{\beta}{2}\sum_{\{(m,n),(k,l)\}\in\mathscr{C}}\rho(s(m,n)-s(k,l)) \quad (17)
$$

where $\rho(.)$ is regularization function which is selected in image restoration to preserve edges. The detailed explanations can be found in [6, 7, 8]. The β is the parameter of the random field.

The probability of event $(S=s)$ is expressed in a Gibbs formulation as

$$
p_s(S=s) = \frac{1}{Z(\beta)}e^{-U(s)} \quad (18)
$$

where $Z(\beta)$ is the partition function to ensure that the total probability equals 1. The density given in (18) can be written in the vector form using the cliques at eight directions as

$$
p_{\mathbf{S}}(\mathbf{s}) = \frac{1}{Z(\beta)}\exp\{-\sum_{d=1}^{8}\beta_d\rho(\mathbf{s}-\mathbf{G}_d\mathbf{s})\} \quad (19)
$$

If the clique differences at each direction is defined as $\mathbf{s}^{(d)} = \mathbf{s}-\mathbf{G}_d\mathbf{s}$, where $\mathbf{G}_d$ is the one pixel shift operator at direction d. The pdf of the image $\mathbf{s}$ is formed by product of pdfs of the edge images $\mathbf{s}^{(d)}$'s.

$$
p_{\mathbf{S}}(\mathbf{s}) = \frac{1}{Z(\beta)}\prod_{d=1}^{8}\exp\{-\beta_d\rho(\mathbf{s}^{(d)})\} \quad (20)
$$

If the same image is modelled as Gaussian at each direction, pdf of the image is expressed as factor of eight Gaussian as

$$
q_{\mathbf{S}}(\mathbf{s}) = \prod_{d=1}^{8}q(\mathbf{s}^{(d)}) \quad (21)
$$

The pdf of a edge image at a particular direction becomes

$$
q(\mathbf{s}^{(d)}) = \frac{1}{(2\pi)^{MN/2}\prod_{i=1}^{MN}\sqrt{\sigma_{d,i}^2}}\exp\{-\frac{1}{2}(\mathbf{s}^{(d)})^T\mathbf{W}_d\mathbf{s}^{(d)}\} \quad (22)
$$

where $\mathbf{W}_d = diag\left\{1/\sigma_{d,i}^2\right\}_{i=1}^{MN}$ is the covariance matrix. This covariance matrix is interpreted as all the pixel in the edge image is independent but has spatially variant variances. This representation is used by Elad in [9] to approximate Gibbs type densities by product of Gaussians. The approximation is done by using first order derivatives as in robust anisotropic diffusion [10]. We use KL divergence measure to approximate to densities.

5.1 Determination of Gaussians Parameters at each Direction

The parameters of directional factor of Gaussians density $q_{\mathbf{S}}(\mathbf{s}|\tau)$ are defined as $\tau = \{\sigma_{d,i}^2\}$ according to (22). The KL divergence is minimized with respect to each $\sigma_{d,i}^2$

$$
\begin{aligned}
D_{KL}(q_{\mathbf{S}}(\mathbf{s})||p_{\mathbf{S}}(\mathbf{s})) &= \int q_{\mathbf{S}}(\mathbf{s})\log\left(\frac{q_{\mathbf{S}}(\mathbf{s})}{p_{\mathbf{S}}(\mathbf{s})}\right)d\mathbf{s} \quad (23)\\
&= E_q\left[-\frac{1}{2}\sum_{d=1}^{8}\sum_{i=1}^{MN}\log\{\sigma_{d,i}^2\}\right.
\end{aligned}
$$

$$-\frac{1}{2}\sum_{d=1}^{8}\sum_{i=1}^{MN}\frac{(s_i^{(d)})^2}{\sigma_{d,i}^2} + \sum_{d=1}^{8}\sum_{i=1}^{MN}\beta_d\rho(s_i^{(d)})+C\Bigg]$$

where $E_q\left[(s_i^{(d)})^2\right]=\sigma_{d,i}^2$, but $E_q\left[\rho(s_i^{(d)})\right]$ is hard to calculate because of the nonlinearity $\rho(.)$. If the $\rho(s_i^{(d)})$ is expanded into Taylor series around the $\tilde{s}_i^{(d)}$ and ignoring the higher order terms, one can write

$$\rho(s_i^{(d)})\approx\rho(\tilde{s}_i^{(d)})+(s_i^{(d)}-\tilde{s}_i^{(d)})\rho'(\tilde{s}_i^{(d)})+(s_i^{(d)}-\tilde{s}_i^{(d)})^2\rho''(\tilde{s}_i^{(d)}) \tag{24}$$

is obtained. The expectation of $\rho(s_i^{(d)})$ is then found as

$$E_q\left[\rho(s_i^{(d)})\right]\approx\rho(\tilde{s}_i^{(d)})-\tilde{s}_i^{(d)}\rho'(\tilde{s}_i^{(d)})+(\sigma_{d,i}^2+(\tilde{s}_i^{(d)})^2)\rho''(\tilde{s}_i^{(d)}) \tag{25}$$

If the approximate expectation is substituted in (23), the variance which maximizes the (23) is obtained as

$$\sigma_{d,i}^2=\frac{1}{\beta_d\rho''(\tilde{s}_i^{(d)})} \tag{26}$$

The sharp discontinuity preserving nonlinear function ρ is chosen as

$$\rho(s_i^{(d)})=\ln\left[1+\frac{(s_i^{(d)})^2}{\delta}\right]. \tag{27}$$

where δ is the threshold parameter.

6. MAP ESTIMATE OF SOURCES

Factor of pdfs of edge images $\mathbf{s}^{(d)}=\mathbf{s}-\mathbf{G}_d\mathbf{s}$ gives the prior of source $\mathbf{s}$.

$$q(\mathbf{s})\propto\exp\{-\sum_{d=1}^{8}(\mathbf{s}-\mathbf{G}_d\mathbf{s})^T\mathbf{W}_d(\mathbf{s}-\mathbf{G}_d\mathbf{s})\} \tag{28}$$

Multiplication of the likelihoods in (15) and priors in (28) produce the posterior. The MAP estimate of sources is found using Newton-Raphson iterations.

7. SIMULATION RESULTS

The variational Bayesian image separation with annealed directionally approximated Gaussians algorithm is summarized in Table 1. We observed that updating sources more frequently than the mixing matrix in a cycle improves the outcomes of the algorithm.

We have worked out two sample problems in order to illustrate the performance of the algorithm. The first example is a mixture of texture images and the second is a mixture of natural image with text images. We compare our algorithm against the background of several other algorithms as listed in the ICALAB Toolbox [5]. Among these we opted for the following popular ones: Fixed Point ICA, JADE, EVD24, and SOBI-RO. The source separation capability of the algorithm is measured with the residual interference in the images, that is, with the Peak Signal to Interference Ratio (PSIR):

$$PSIR=10\log\left(\frac{255}{||\mathbf{s}_l-\hat{\mathbf{s}}_l||^2}\right) \tag{29}$$

and averaged over all the sources.

We first show the performance of the proposed algorithm for a range of signal-to-noise ratios (SNR) ranging from 20 to 50 dB. The SNR is calculated as

$$SNR=\frac{1}{K}\sum_{j=1}^{K}20log\frac{\sum_{i=1}^{L}a(j,i)255}{\sigma} \tag{30}$$

Each algorithm starts with the same initial values. The results are detailed in Table 2. Table 2 also includes the PSIR results of the nearest competitor algorithm in [3]. We can observe that the proposed algorithm is more of a heavy-duty type, that is, it can separate sources under very noisy conditions, while at higher observed image qualities it fares as well as the algorithm in [3]. Second, we compare the chosen algorithms under 5 dB noise contamination. Results are given in Fig. 1. However, we vary the percentage of contaminated pixels from 0% to 100%. We observe that our algorithm outperforms all other algorithms under noise contamination conditions. In the noiseless case, only SOBI-RO algorithm is better, and this by a large margin. These results are illustrated in Figs. 2 and 3. Fig. 2 shows the de-mixing of three texture sources while Fig. 3 is related to the separation of a text source from an image source.

Table 1: The algorithm

1. Initialize $\mathbf{s}$ and $\mathbf{A}$, $T=1$
2. Update $\sigma_{d,i}^2$ and $\mathbf{s}_i$ for M times
3. Update $\mathbf{A}$ for one time
4. $T\longleftarrow 0.96T$
5. Goto 2nd step.

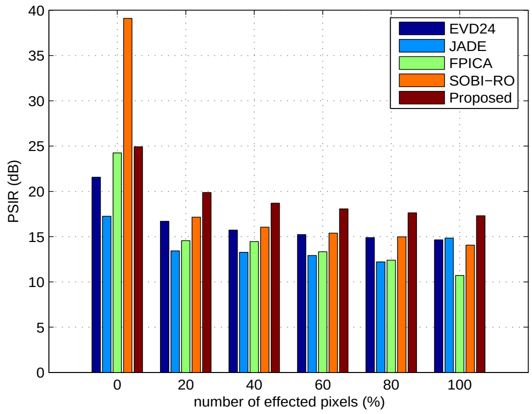

Figure 1: Comparison results for 5 dB noise case. Horizontal axis represents percent of pixel which effected by noise.

Table 2: Comparison with [3].

σ	SNR	[3]	proposed
90.98	15	not converged	14.30
53.34	20	not converged	15.66
30.02	25	not converged	17.54
16.88	30	19.25	19.79
5.33	40	22.94	23.43
1.68	50	24.95	25.81
0	∞	25.69	26.80

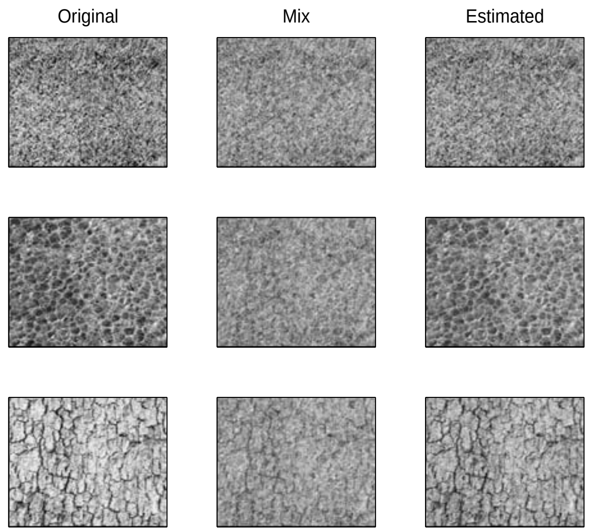

Figure 2: Simulation results 1.

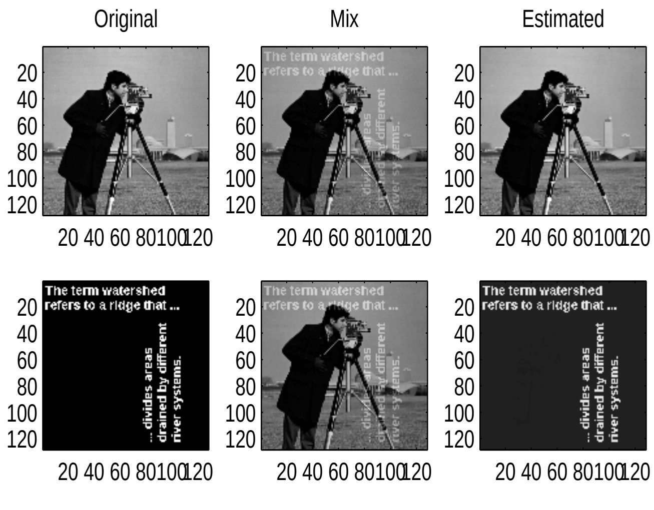

Figure 3: Simulation results 2.

8. CONCLUSIONS

In this study, Gibbs distribution with non-convex potentials is approximated by directional factor of Gaussians. The aim of the approximation is to obtain a tractable prior which should be easy to work analytically and their log should be convex. The approximate density also models the local spatial structure of the image, using second order statistics, in contrast to constant covariance mean field approximation. The proposed method has outperformed all of its competitors under additive noise conditions. Annealing schedule is more appropriate for high SNR values. For low SNR values, slow or no annealing is applied.

REFERENCES

[1] James W. Miskin, *Ensemble Learning for Independent Component Analysis*. PhD Thesis, Address: University of Cambridge, 2000.

[2] H. Attias, “Independent factor analysis,” *Neural Computation*, vol. 11, pp. 803–851, 1999.

[3] E. E. Kuruoglu, A. Tonazzini, and L. Bianchi, “Source separation in noisy astrophysical images modelled by Markov random fields,” in *Int. Conf. on Image Proc. ICIP’04*, Oct., 2004, pp. 24–27.

[4] A. Tonazzini, L. Bedini, and E. Salerno “A Markov model for blind image separation by a mean-filed EM algorithm,” *IEEE Trans. on Image Processing*, vol. 15, pp. 473–482, Feb. 2006.

[5] A. Cichocki, S. Amari, K. Siwek, and T. Tanaka et al., *ICALAB Toolboxes*. http://www.bsp.brain.riken.jp/ICALAB.

[6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Deterministic edge-preserving regularization in computed imaging,” *IEEE Trans. on Image Processing*, vol. 6, pp. 298–311, Feb. 1997.

[7] G. Archer, and D. M. Titterington, “On some Bayesian/regularization methods for image restoration,” *IEEE Trans. on Image Processing*, vol. 4, pp. 989–995, July 1995.

[8] C. Bouman, and K. Sauer, “A generalized Gaussian image model for edge-preserving MAP estimation,” *IEEE Trans. on Image Processing*, vol. 2, pp. 296–310, July 1993.

[9] M. Elad, “On the origin of the bilateral filter and ways to improve it,” *IEEE Trans. on Image Processing*, vol. 11, pp. 1141–1151, Oct. 2002.

[10] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, “Robust anisotropic diffusion,” *IEEE Trans. on Image Processing*, vol. 7, pp. 421–432, Oct. 1998.

[11] K. Rose, “Deterministic annealing for clustring, compression, classification, regression, and related optimization problems,” *Proceedings of the IEEE*, vol. 86, pp. 2210–2238, Nov. 1998.

[12] G. Hua, and Y. Wu, “Variational maximum a posteriori by annealed mean field analysis,” *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 27, pp. 1747–1761, Nov. 2005.