Watermark extraction using blind image separation and sparse representation

Abdeldjalil Aissa El Bey, Amina Saleem, Karim Abed-Meraim, Azzedine Beghdadi

WATERMARK EXTRACTION USING BLIND IMAGE SEPARATION AND SPARSE REPRESENTATION

Abdeldjalil Aïssa-El-Bey[1], *Amina Saleem*[2], *Karim Abed-Meraim*[3,4], *Azzedine Beghdadi*[2]

[1]TELECOM Bretagne, S&C department, Technopôle Brest-Iroise, 29285 Brest, France
[2]Université Paris 13, L2TI, 99 Av. J.B. Clément, 93430, Villetaneuse, France
[3]TELECOM ParisTech, TSI department, 37-39 rue Dareau 75014 Paris, France
[4]University of Sharjah, ECE department, 27272 Sharjah, U.A.E.

ABSTRACT

This paper proposes a new image watermarking technique, which employs an Iterative sparse blind separation algorithm (ISBS) and a new sparsity based BSS algorithm (New ISBS algorithm or NISBS) for watermark extraction. The ISBS algorithm performs the separation based on the optimization of an ℓ_p norm based contrast function. The new sparsity based BSS algorithm employs a smooth approximation of the absolute value function as the cost function. The watermark embedding of the images is performed in the spatial domain. The results of the simulations demonstrate the validity and performance of these two techniques for watermark extraction.

1. INTRODUCTION

Digital watermarking technology has evolved as an important technology in the recent years. The basic principle of most watermarking method involves the application of small, pseudo-random changes to the selected coefficients in the spatial or transform domain. Most of the watermark detection schemes use some kinds of correlating detector to verify the presence of the embedded watermark [1, 2]. Digital image watermarking has its applications in copy rights protection, data tracking and monitoring [1]. Blind source separation (BSS) is an important area of research in signal and image processing. The BSS problem can be solved using sparse representations of the source signals. Solution for the blind separation of image sources using sparsity include the wavelet-transform domain method in [2] and the method in [3] using projection onto sparse dictionaries and the iterative Blind source separation algorithm presented in [4]. This paper introduces a new image watermark extraction technique based on the iterative sparse blind separation algorithm (ISBS) and a new ISBS algorithm. The ISBS algorithm employs an ℓ_p norm based contrast function for blind signal separation. When the images are sparse or sparsely representable, a smooth approximation of the absolute value function is a good choice for the cost function. The NISBS algorithm proposed in this paper employs the modeling of the distributions of sparse images using a family of convex smooth functions. The ISBS algorithm proposed in [4] and the NISBS algorithm presented in this paper are shown to be more efficient than other existing techniques in the literature and both lead to improved separation quality with lower computational cost. The performance of the proposed algorithms is compared to the performance of other ICA algorithms using the objective image quality measure inspired by the Human Visual System (HVS) proposed in [5]. It is shown that the ISBS and NISBS algorithms perform better in terms of PSNR-WAV. The new watermarking technique is presented with the principal assumptions of (i) image source sparsity, (ii) instantaneous mixtures and (iii) the same number of mixtures and sources (three mixtures and three sources). In our proposed BSS based method, we do not have restrictions on the mixing process as well as the mixing coefficients. The paper is organized as follows. Section 2 presents the data model and assumptions of our system. The blind watermark extraction system using the sparsity based algorithms is described in Section 3. The simulations and the performance of the algorithms is discussed in Section 4. The conclusions are drawn in Section 5.

2. DATA MODEL AND ASSUMPTIONS

A generic watermark embedding system consists of the inputs which are the original data f_1 , the watermark signal f_2 and an optional public or secret key f_3 each of size (m_f, n_f). The key is used to enforce the security, that is, to prevent unauthorized party from recovering and manipulating the watermark. The proposed image watermarking system uses a watermark f_2 and a secret key f_3 for the purpose of conducting two levels of security, by using the special images as the watermark and the key, with the same size as the original image f_1, to be embedded. Both the watermark f_2 and the key f_3 are inserted in the spatial domain of the original image f_1. The watermarked image g_1 is a linear mixture of the original image, key and watermark. That is,

$$g_1(m,n) = f_1(m,n) + a\ f_2(m,n) + b\ f_3(m,n) \tag{1}$$

where a and b are the weighting coefficients [6]. To assure the identifiability of BSS model, it is required that the number of observed linear mixture inputs is at least equal to or larger than the number of independent sources. For the proposed watermark extraction scheme, at least three linear mixtures of the three independent sources are needed. Using the key image f_3 and with the help of original image f_1, two more mixed images are generated by adding them into the watermarked image

$$g_2(m,n) = c\, g_1(m,n) + d\ f_3(m,n) \tag{2}$$
$$g_3(m,n) = k\, g_1(m,n) + l\ f_1(m,n) \tag{3}$$

where $\{c, d, k, l\}$ are arbitrary real numbers. The latter mixtures can be modeled by the following linear system:

$$\boldsymbol{g}(m,n) = \boldsymbol{A}\boldsymbol{f}(m,n) \tag{4}$$

where, $\boldsymbol{f}(m,n) = [f_1(m,n), \cdots, f_N(m,n)]^T$ is a $N \times 1$ (with $N = 3$) image source vector consisting of the stack of corresponding pixels of source images, $\boldsymbol{A}$ is the $M \times N$ full column rank mixing matrix (here, $M = N = 3$), $\boldsymbol{g}(m,n) =$

$[g_1(m,n), \cdots, g_M(m,n)]^T$ is an $M \times 1$ vector of mixture image pixels and the superscript T denotes the transpose operator. The purpose of blind image separation is to find a separating matrix, i.e. a $N \times M$ matrix $\boldsymbol{B}$ such that $\widehat{\boldsymbol{f}}(m,n) = \boldsymbol{B}\boldsymbol{g}(m,n)$ is an estimate of original images.

3. BLIND WATERMARK EXTRACTION

As shown in [3,7], exploiting the sparsity of some representations of the original images afford us to achieve the BSS problem. Indeed, the mixture destroys or 'reduces' the sparsity of the considered signals that is restored after source separation. Reversely, it is shown in [3, 7] that restoring (maximizing) the sparsity leads to the desired source separation. Based on this, we propose in the sequel a two-step BSS solution consisting in a linear pre-treatment that transforms the original sources into sparse signals followed by a BSS algorithm that minimizes the cost function of the transformed image mixtures using natural gradient technique.

3.1. Image pre-treatment

The algorithms proposed in this article are efficients for separating sparse sources. For some signals, one can assume that the spatial or temporal representation is naturally sparse, whereas for natural scenes, this assumptions falls down. We propose to make the image sparse by simply taking into account its Laplacian transform:

$$\boldsymbol{\mathcal{F}} = \nabla \boldsymbol{f} = \frac{\partial^2 \boldsymbol{f}}{\partial x^2} + \frac{\partial^2 \boldsymbol{f}}{\partial y^2} \quad , \tag{5}$$

or, in discrete form

$$\begin{aligned}\boldsymbol{\mathcal{F}}(m,n) \quad = \quad & \boldsymbol{f}(m+1,n) + \boldsymbol{f}(m-1,n) + \boldsymbol{f}(m,n+1) \\ & + \boldsymbol{f}(m,n-1) - 4\boldsymbol{f}(m,n) \ .\end{aligned}$$

Our motivation for choosing this transformation is two fold. First the Laplacian transform is a sparse representation of the image since it acts as an edge detector which provides a two-level image, the edges and the homogeneous background. Second, the Laplacian is a linear transformation. This latter property is 'critical' since the separation matrix estimated to separate the image mixtures is the same to separate the mixture of Laplacian images:

$$\boldsymbol{\mathcal{G}} = \frac{\partial^2 \boldsymbol{Af}}{\partial x^2} + \frac{\partial^2 \boldsymbol{Af}}{\partial y^2} = \boldsymbol{A\mathcal{F}} \tag{6}$$

where $\boldsymbol{\mathcal{G}}$ is the Laplacian transform of the mixtures. In the literature, some other linear transformations were proposed in order to make the image sparse, including the projection into a sparse dictionary [7]. In Figure 1, the original cameraman image is displayed as well as its Laplacian transform and their respective histograms that clearly show the sparsity of the latter.

In the pre-treatment phase, we also propose an optional whitening step which aims to set the mixtures to the same energy level. Furthermore, this procedures reduces the number of parameters to be estimated. More precisely, the whitening step is applied to the Laplacian image mixtures before using our separation algorithm. The whitening is achieved by applying a $N \times M$ matrix $\boldsymbol{Q}$ to the Laplacian image mixtures in such a way $\mathrm{Cov}(\boldsymbol{Q\mathcal{G}}) = \boldsymbol{I}$ in the noiseless case, where $\mathrm{Cov}(\cdot)$ stands for the covariance operator. As shown in [8], $\boldsymbol{Q}$ can be computed as the inverse square root of the noiseless covariance matrix of the Laplacian image mixtures (see [8] for more details). In the following, we apply our separation algorithm on the whitened data:

$$\boldsymbol{\mathcal{G}}_w(m,n) = \boldsymbol{Q\mathcal{G}}(m,n).$$

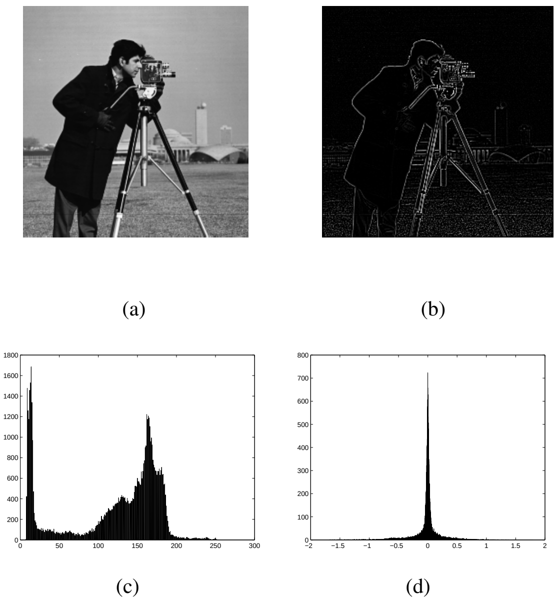

Fig. 1. (a) Original image, (b) Laplacian transform, (c) Original image histogram, (d) Sparse Laplacian transform histogram

3.2. Sparsity-based BSS algorithm

In this section, we propose an iterative algorithm for the separation of sparse signals, namely the ISBS for Iterative Sparse Blind Separation algorithm. It is well known that Laplacian image transform is characterized by its sparsity property in the spatial domain [9, 10]. This property can be measured by the ℓ_p norm where $0 \leq p < 2$. More specifically, one can define the following sparsity based contrast function,

$$G_p(\boldsymbol{\mathcal{F}}) = \sum_{i=1}^{N} \left[\mathcal{J}_p(\mathcal{F}_i)\right]^{\frac{1}{p}} \tag{7}$$

where

$$\mathcal{J}_p(\mathcal{F}_i) = \frac{1}{m_f n_f} \sum_{m=1}^{m_f} \sum_{n=1}^{n_f} |\mathcal{F}_i(m,n)|^p \ . \tag{8}$$

The algorithm finds a separating matrix $\boldsymbol{B}$ such as,

$$\boldsymbol{B} = \arg\min_{\boldsymbol{B}} \{\mathbb{G}_p(\boldsymbol{B})\} \tag{9}$$

where

$$\mathbb{G}_p(\boldsymbol{B}) \triangleq G_p(\boldsymbol{\mathcal{H}}) \tag{10}$$

and $\boldsymbol{\mathcal{H}}(m,n) \triangleq \boldsymbol{B\mathcal{G}}_w(m,n)$ represents the estimated image sources Laplacian. The approach we choose to solve (9) is inspired from [11]. It is a block technique based on the processing of $m_f n_f$ observed image pixels and consists in searching the minimum of the sample version of (9). Solutions are obtained iteratively in the form:

$$\boldsymbol{B}^{(k+1)} \quad = \quad (\boldsymbol{I} + \boldsymbol{\epsilon}^{(k)})\boldsymbol{B}^{(k)} \tag{11}$$

$$\boldsymbol{\mathcal{H}}^{(k+1)}(m,n) \quad = \quad (\boldsymbol{I} + \boldsymbol{\epsilon}^{(k)})\boldsymbol{\mathcal{H}}^{(k)}(m,n) \ . \tag{12}$$

At iteration k, a matrix $\boldsymbol{\epsilon}^{(k)}$ is determined from a local linearization of $G_p(\boldsymbol{B\mathcal{G}}_w)$. It is an approximate Newton technique with the benefit that $\boldsymbol{\epsilon}^{(k)}$ can be very simply computed (no Hessian inversion) under the additional assumption that $\boldsymbol{B}^{(k)}$ is close to a separating

matrix. This procedure is illustrated in the following.
At the $(k+1)^{th}$ iteration, the proposed criterion (8) can be developed as follows:

$$\mathcal{J}_p(\mathcal{H}_i^{(k+1)}) = \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,n_f}\left|\mathcal{H}_i^{(k)}(m,n) + \sum_{j=1}^{N}\epsilon_{ij}^{(k)}\mathcal{H}_j^{(k)}(m,n)\right|^p$$

Under the assumption that $\boldsymbol{B}^{(k)}$ is close to a separating matrix, we have

$$|\epsilon_{ij}^{(k)}| \ll 1$$

and thus, a first order approximation of $\mathcal{J}_p(\mathcal{H}_i^{(k+1)})$ is given by:

$$\mathcal{J}_p(\mathcal{H}_i^{(k+1)}) \approx \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,n_f}|\mathcal{H}_i^{(k)}(m,n)|^p + p\sum_{j=1}^{N}\epsilon_{ij}^{(k)}\left(|\mathcal{H}_i^{(k)}(m,n)|^{p-1}\text{sign}\left(\mathcal{H}_i^{(k)}(m,n)\right)\mathcal{H}_j^{(k)}(m,n)\right) \quad (13)$$

where sign$(\cdot)$ represents the sign value operator. Using equation (13), equation (7) can be rewritten in more compact form as:

$$\mathbb{G}_p\left(\boldsymbol{B}^{(k+1)}\right) = \mathbb{G}_p\left(\boldsymbol{B}^{(k)}\right) + Tr\left(\boldsymbol{\epsilon}^{(k)}\boldsymbol{\mathcal{R}}^{(k)T}\boldsymbol{D}^{(k)}\right) \quad (14)$$

where $Tr(\cdot)$ is the matrix trace operator, the ij^{th} entry of matrix $\boldsymbol{\mathcal{R}}^{(k)}$ is given by:

$$\mathcal{R}_{ij}^{(k)} = \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,n_f}|\mathcal{H}_i^{(k)}(m,n)|^{p-1}\text{sign}\left(\mathcal{H}_i^{(k)}(m,n)\right)\mathcal{H}_j^{(k)}(m,n)\,.$$

and

$$\boldsymbol{D}^{(k)} = \left[\text{diag}\left(\mathcal{R}_{11}^{(k)},\ldots,\mathcal{R}_{NN}^{(k)}\right)\right]^{\frac{1}{p}-1}\,. \quad (15)$$

Using a gradient technique, $\boldsymbol{\epsilon}^{(k)}$ can be written as:

$$\boldsymbol{\epsilon}^{(k)} = -\mu\boldsymbol{D}^{(k)}\boldsymbol{\mathcal{R}}^{(k)}\,, \quad (16)$$

where $\mu > 0$ is the descent step. Replacing (16) into (14) leads to,

$$\mathbb{G}_p\left(\boldsymbol{B}^{(k+1)}\right) = \mathbb{G}_p\left(\boldsymbol{B}^{(k)}\right) - \mu\|\boldsymbol{D}^{(k)}\boldsymbol{\mathcal{R}}^{(k)}\|^2\,, \quad (17)$$

so μ controls the decrement of the criterion. Now, to avoid the algorithm's convergence to the trivial solution $\boldsymbol{B} = \boldsymbol{0}$, one normalizes the outputs of the separating matrix to unit-power, i.e. $\rho_{\mathcal{H}_i}^{(k+1)} \triangleq E\left(|\mathcal{H}_i^{(k+1)}(m,n)|^2\right) = 1,\ \forall\, i$. Using first order approximation, this normalization leads to:

$$\epsilon_{ii}^{(k)} = \frac{1-\rho_{\mathcal{H}_i}^{(k)}}{2\rho_{\mathcal{H}_i}^{(k)}}\,. \quad (18)$$

The final estimated separation matrix $\boldsymbol{B} = \boldsymbol{B}^{(\mathcal{K})}\boldsymbol{Q}$ is applied to the image mixtures $\boldsymbol{g}$ to obtain an estimation of the original images. $\mathcal{K}$ denotes here the number of iterations that can be either chosen a priori or given by a stopping criterion of the form $\|\boldsymbol{B}^{(k+1)} - \boldsymbol{B}^{(k)}\| < \delta$ where δ is a small threshold value.

3.3. New sparsity-based BSS algorithm

When the images are sparse or sparsely representable, a smooth approximation of the absolute value function might be a better choice for the cost function [12]. We, therefore, focus our attention on modeling distributions of sparse images using a family of convex smooth functions and propose the following cost function :

$$\widetilde{G}_\lambda(\boldsymbol{\mathcal{F}}) = \sum_{i=1}^{N}\widetilde{\mathcal{J}}_\lambda(\mathcal{F}_i) \quad (19)$$

with

$$\widetilde{\mathcal{J}}_\lambda(\mathcal{F}_i) = \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,n_f}|\mathcal{F}_i(m,n)| - \lambda\log\left(1 + \frac{|\mathcal{F}_i(m,n)|}{\lambda}\right) \quad (20)$$

where λ is a positive smoothing parameter [12]. The algorithm finds a separating matrix $\boldsymbol{B}$ such as,

$$\boldsymbol{B} = \arg\min_{\boldsymbol{B}}\left\{\widetilde{\mathbb{G}}_\lambda(\boldsymbol{B})\right\} \quad (21)$$

where

$$\widetilde{\mathbb{G}}_\lambda(\boldsymbol{B}) \triangleq \widetilde{G}_\lambda(\boldsymbol{\mathcal{H}})\,. \quad (22)$$

In the same way as the ISBS algorithm presented in the previous section, the solutions are obtained iteratively as describe by the equations (11) and (12). Therefore, at the $(k+1)^{th}$ iteration, the proposed criterion (20) can be developed as follows :

$$\widetilde{\mathcal{J}}_\lambda(\mathcal{H}_i^{(k+1)}) = \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,nf}\left|\mathcal{H}_i^{(k)}(m,n) + \sum_{j=1}^{N}\epsilon_{ij}^{(k)}\mathcal{H}_j^{(k)}(m,n)\right| - \lambda\log\left(1 + \frac{\left|\mathcal{H}_i^{(k)}(m,n) + \sum_{j=1}^{N}\epsilon_{ij}^{(k)}\mathcal{H}_j^{(k)}(m,n)\right|}{\lambda}\right)\,.$$

Under the assumption that $\boldsymbol{B}^{(k)}$ is close to a separating matrix, we have

$$|\epsilon_{ij}^{(k)}| \ll 1$$

and thus, by using a first order approximation of $\widetilde{\mathcal{J}}_\lambda(\mathcal{H}_i^{(k+1)})$, we can rewrite equation (19) in more compact form as:

$$\widetilde{\mathbb{G}}_\lambda\left(\boldsymbol{B}^{(k+1)}\right) = \widetilde{\mathbb{G}}_\lambda\left(\boldsymbol{B}^{(k)}\right) + Tr\left(\boldsymbol{\epsilon}^{(k)}\widetilde{\boldsymbol{\mathcal{R}}}^{(k)T}\right) \quad (23)$$

where the ij^{th} entry of matrix $\widetilde{\boldsymbol{\mathcal{R}}}^{(k)}$ is given by:

$$\widetilde{\mathcal{R}}_{ij}^{(k)} = \frac{1}{m_f n_f}\sum_{m,n=1}^{m_f,nf}\frac{\text{sign}\left(\mathcal{H}_i^{(k)}(m,n)\right)\mathcal{H}_j^{(k)}(m,n)}{\lambda + |\mathcal{H}_i^{(k)}(m,n)|}\,.$$

Using a gradient technique, $\boldsymbol{\epsilon}^{(k)}$ can be written as:

$$\boldsymbol{\epsilon}^{(k)} = -\mu\widetilde{\boldsymbol{\mathcal{R}}}^{(k)}\,, \quad (24)$$

Replacing (24) into (23) leads to,

$$\widetilde{\mathbb{G}}_\lambda\left(\boldsymbol{B}^{(k+1)}\right) = \widetilde{\mathbb{G}}_\lambda\left(\boldsymbol{B}^{(k)}\right) - \mu\|\widetilde{\boldsymbol{\mathcal{R}}}^{(k)}\|^2\,. \quad (25)$$

Now, to avoid the algorithm's convergence to the trivial solution $\boldsymbol{B} = \boldsymbol{0}$, one normalizes the outputs of the separating matrix to unit-power (see equation (18)).

4. SIMULATION

Simulation experiments are conducted to demonstrate the feasibility of the proposed BSS method for watermark extraction. All simulations are carried on 256×256 parrot, secret key and watermark images. The algorithms are developed on MATLAB environment. Illustrative result for watermark extraction is shown in Fig. 2. where

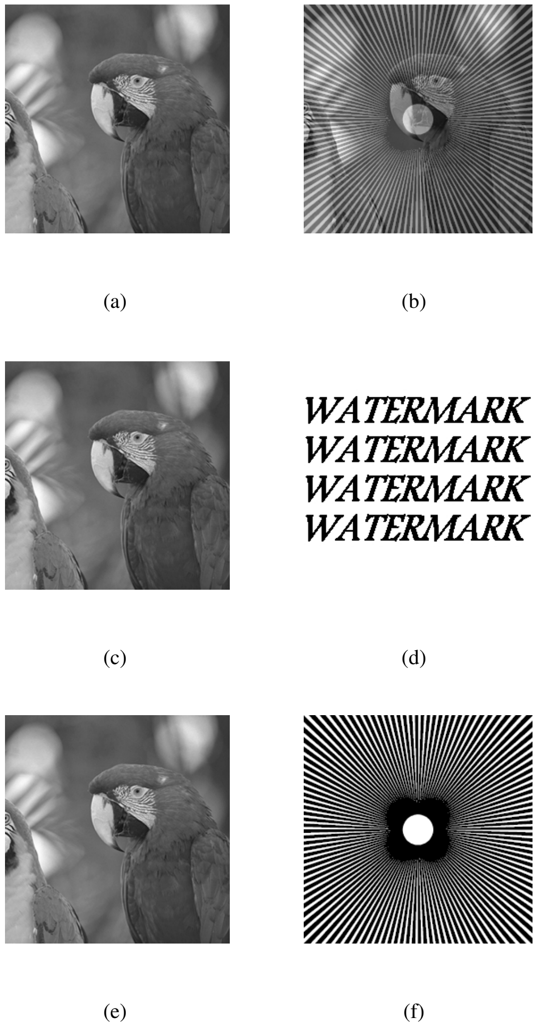

Fig. 2. (a) watermarked image, (b) and (c) generated mixture images, (d) extracted watermark, (e) extracted image, (f) extracted key.

we represent the watermarked image, generated mixture images, extracted watermark, extracted image and extracted key by the proposed algorithm. The mixture coefficients are $a = 2 \cdot 10^{-3}$, $b = 5 \cdot 10^{-3}$, $c = 1$, $d = 2$, $k = 1$, $l = 1$. The performance of watermark extraction is evaluated by an objective image quality measure inspired from the Human Visual System (HVS) properties and developed in [5]. It is called $PSNR{-}WAV$ for Peak Signal to Noise Ratio based on Wavelet decomposition. Table 1 shows the PSNR-WAV between the original and extracted images for the example described in Fig. 2. We compare the performance of the proposed algorithms to the ICA algorithm. It is clearly shown that our algorithms (ISBS and New ISBS) perform better in terms of PSNR-WAV.

5. CONCLUSION

In this paper, a new watermarking technique based on BSS algorithm using sparsity property of images has been proposed. The proposed technique consists in a sparsification of the natural observed mixtures followed by a blind separation of the original images. The

BSS Algorithm	PSNR-WAV (dB)		
	Parrot	Key	Watermark
ICA	28.76	15.85	22.36
ISBS	33.12	19.32	25.67
NISBS	40.70	28.92	34.49

Table 1. Performance evaluation : Comparison of the PSNR-WAV for ICA, ISBS and NISBS algorithms

sparsification is simply the Laplacian transform and has a low computational cost. The separation is performed using an iterative algorithm based on the minimizing of the sparsity cost function of the Laplacian image.

6. REFERENCES

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.

[2] K. Stefan and F. A. P. Petitcolas Eds., *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, 2000.

[3] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Separation of reflections via sparse ICA," in *Proc. IEEE International Conference on Image Processing ICIP*, September 2003, vol. 1, pp. 313–316.

[4] W. Souidene, A. Aïssa-El-Bey, K. Abed-Meriam, and A. Beghdadi, "Blind image separation using sparse representation," in *Proc. IEEE International Conference on Image Processing ICIP*, September 2007, vol. 3, pp. 125–128.

[5] A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," in *Proc. International Symposium on Signal Processing and Its Applications ISSPA*, Paris, France, July 2003, vol. 1, pp. 485–488.

[6] D. Yu, F. Sattar, and K. K. Ma, "Watermark detection and extraction using independent component analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 92–104, 2002.

[7] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[8] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transaction on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

[9] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, Wiley & Sons, Ltd., UK, 2003.

[10] M. Zibulevsky, "Sparse source separation with relative Newton method," in *Proc. ICA*, Apr. 2003, pp. 897–902.

[11] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE T-SP*, vol. 45, no. 7, pp. 1712–1725, July 1997.

[12] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Blind deconvolution of images using optimal sparse representations," *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 726–736, June 2005.