Robust Blind Image Watermarking with Independent Component Analysis: A Embedding Algorithm

Juan José Murillo-Fuentes and Rafael Boloix-Tortosa

Dept de teoria de la Señal y Comunicaciones, Universidad de Sevilla,
Paseo de los Descubrimientos sn, Sevilla 41092, Spain*
murillo@esi.us.es
http://viento.us.es/~murillo

Abstract. The authors propose a new solution to the blind robust watermarking of digital images. In this approach we embed the watermark into the independent components of the image. Since independent components are related to the edges of the image, this method has a little perceptual impact on the watermarked image. Besides, we exploit the orthogonality of independent components and spread-spectrum generated watermarks in the blind extraction of the watermark. As extraction algorithm we use a simple matched filter. We also improve this novel method with standard techniques such as perceptual masking and holographic properties. Some experiments are included to illustrate the good performance of the algorithm against compression, cropping, filtering or quantization based attacks.

1 Introduction

Robust Watermarking (RW) of digital images [1] is one common solution to protect owners rights. It consists of embedding another signal or mark into the to be protected host image. We aim the watermark to be detected after severe attacks. In addition, our watermark is designed to be transparent to the user and we do not use the host image at detection, i.e., this is a invisible blind RW approach.

In some RW approaches we embed the mark in the spatial domain. On the other hand, we have methods working in a transform domain, such as the DCT or the DWT. ICA has been recently applied to digital watermarking following two main approaches. On the one hand, we have those approaches based on the mixture of the host image, or some transform domain coefficients, and the watermark [2, 3, 4]. In these methods, ICA is applied at detection to extract the watermark. On the other hand, based on the original results in [5], we have methods based on ICA as a transform domain where to embed the watermark

* Thanks to Spanish goverment for funding TIC-2003-03781.

J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.): IWANN 2005, LNCS 3512, pp. 1100–1107, 2005.

[6, 7]. The authors in [6] develop a non-blind approach focusing mainly in the detection stage by using non-linear techniques. In this paper we focus on simple techniques in the embedding stage [7] such as spread-spectrum watermark, perceptual masking and holographic methods, to greatly improve the blind method in [5].

2 ICA in Image Processing

2.1 Independent Component Analysis

Independent component analysis (ICA) [8] consists of projecting a set of components onto another statistically independent set. In the simple ICA, the l entries of a sample t of a column vector sequence $\boldsymbol{x}_t$ are projected into a space of l components $\boldsymbol{y}_t$ as statistically independent as possible. This projection is represented by an $l \times l$ matrix $\boldsymbol{B}$.

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t \tag{1}$$

Very much literature has been devoted to ICA algorithms. We will use here batch algorithms that minimize the marginal entropies ME of the outputs [9]. These algorithms have a good performance and are easy to use compared to gradient based methods.

2.2 Application to Image Processing

There are two common applications of ICA to image processing. On the one hand, we may assume we have l linear mixtures of l images. Therefore, we simply need to reshape each mixture of images into a vector and then apply ICA to separate them as in equation (1). The methods in [2, 3, 4] are based on this approach. On the other hand, we have other approaches where only one image is involved. These methods first decompose an image into components $\boldsymbol{x}_t$ to later apply ICA [10]. Afterwards, any image processing technique may be applied to these, so computed, independent components (IC) [11]. Notice that each row of $\boldsymbol{B}$ provides one independent component (an entry of vector $\boldsymbol{y}_t$). Therefore, if we reshape each row of $\boldsymbol{B}$ into a matrix, we obtain a set of 2-dimensional basis functions. These basis functions, also regarded as patches or features, are closely related to well-localized and oriented Gabor filters [8]. Some other authors suggest these basis functions to be the edges of the image [10], [12], or even to model the receptive fields of the primary visual cortical neurons [8]. An analysis of the image ICA components shows that many independent components are sparse distributed and that only some basis functions are needed to represent the image. Besides, the probability of independent components having small amplitudes is high, but large amplitudes occurs as well [8]. In [11] these features are used to compress or encode an image. Basic compression algorithms exploit these ideas as they retain only the independent components with larger energy. In addition, the authors in [11, 13] show that groups of images with similar features may be restored from a common set of basis (rows of matrix $\boldsymbol{B}$). Particularly, the projections $\boldsymbol{B}$ computed for one image can be used in the processing of another one of the same class (text images, natural scenes,...).

The IC of an image $\boldsymbol{I}$ can be computed as follows. Assume matrix $\boldsymbol{I}$ be a gray-scale image of size $n \times m$.This matrix can be divided into $k \times k$ blocks or patches [8] $\boldsymbol{C}_{p,q}$ to reshape them into vectors $\boldsymbol{x}_t^I$ where $t = (p-1) \cdot m/k + q$. The rows of $\boldsymbol{x}_t^I$ may be then projected onto $l = k^2$ independent components,

$$\boldsymbol{y}_t^I = \boldsymbol{B}\boldsymbol{x}_t^I \qquad t = 1, \ldots, mn/k^2 \tag{2}$$

In Fig. 1 we show the spatial basis functions computed for $k = 8$ and Lena Image (256×256). Each row matrix of the separating matrix $\boldsymbol{B}$ was reshaped into a 8×8 image. They were arranged row wise in descending order of energy, i.e., those basis functions (rows of $\boldsymbol{B}$) providing independent components with larger variance are located at the top rows. The top-left corner basis allows to represent the DC component of every 8×8 patch of the image. Notice also that the first rows are the basis functions to build the borders of the image. Besides, the last ones provide low energy components, details of the image.

3 Watermarking with ICA

Having the previous ideas in mind, we propose a new algorithm as follows. Firstly, the edges of the images are the candidate areas where to embed the watermark if we aim it to be imperceptible. Since some of the ICA basis functions are the edges of the images, by embedding the watermark in the associate IC we improve the invisibility of the mark. Hence, at a given PSNR, our watermark will be more imperceptible than the one embedded by using, e.g., DCT coefficients. Secondly, we may use a common public set of basis or rather use, following the ideas in [11, 13], our own private ICA projection. Hence, we fulfill one of the Kerckhoffs [14] principles, even if the attackers know we embedded the watermark using ICA they still do not know the exact projections. In addition, this privacy of the embedded watermark can be improved if we recall about IC tending to be sparse. In Fig. 2 we depict the power spectral density (PSD) for the independent components number 1 (Fig 1a) , 5 (Fig 1b) and 9 (Fig 1c) computed for the image of Lena with $k = 3$. As the IC have been arranged in descending order of energy, we have the DC component in Fig. 2a. IC number 5 in Fig. 2b has a white noise-like frequency response. Finally, the last IC has a high frequency response. In this paper we propose to use spread-spectrum watermark, with flat frequency response, to be added to the middle IC. This way we improve the robustness against any frequency based watermark filtering and most important, we ensure a low cross correlation between the IC and the watermark allowing blind detection by using a simple matched filter. Finally, the spread-spectrum watermark can be generated using a circular convolution with the bits of copyright information. This way we embed every bit into every pixel of the image, improving the robustness against some attacks such as cropping. We next describe the algorithm in detail.

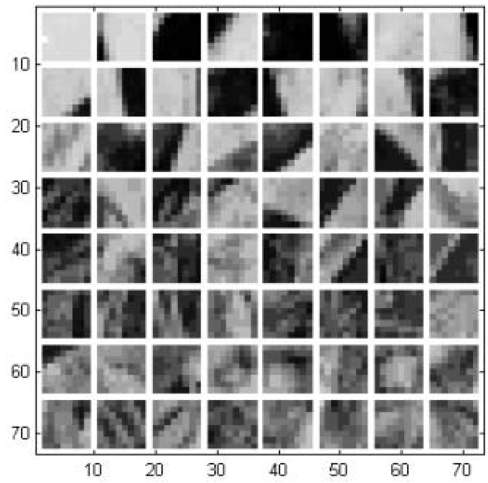

Fig. 1. Example of basis functions for $k = 8$

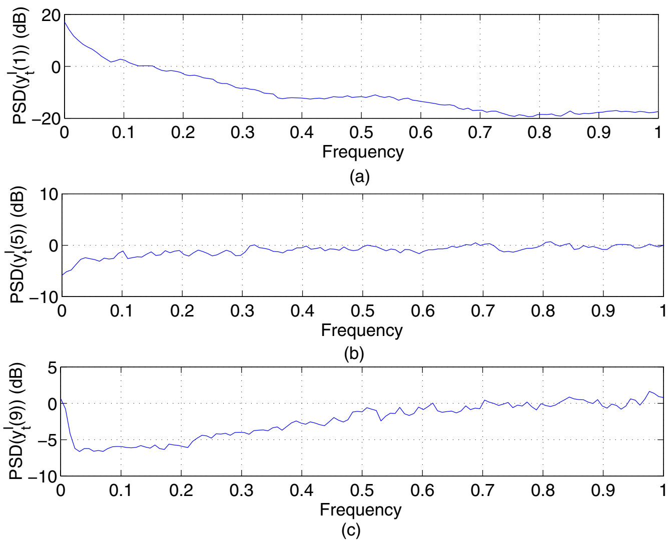

Fig. 2. Power spectral density for independent components number 1 (a), 5 (b) and 9 (c) of the Lena image for k=3

3.1 Embedding

In Fig. 3 we include a scheme for a ICA based RW algorithm. We first will describe the basic steps of the embedding method to later discuss on its particular features. The embedding method yields

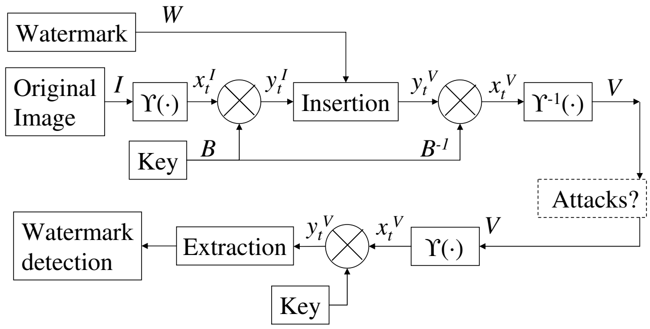

Fig. 3. ICA based robust watermarking algorithm

Algorithm 1: Embedding.

1 *Image to column vectors.* Compute the components $\boldsymbol{x}_t^I$ of the $n \times m$ host image $\boldsymbol{I}$ using $k \times k$ blocks.
2 *ICA components.* Compute its IC, $\boldsymbol{y}_t^I = \boldsymbol{B}\boldsymbol{x}_t^I$, using an ICA projection $\boldsymbol{B}$, the key of the insertion method.
3 *Insertion.* Compute the IC of the marked image, $\boldsymbol{y}_t^V$, by updating $\boldsymbol{y}_t^I$ with the watermark, $\boldsymbol{W}$.
4 *Restoration.* Restore the (watermarked) image $\boldsymbol{V}$ from components $\boldsymbol{x}_t^V = \boldsymbol{B}^{-1}\boldsymbol{y}_t^V$.

As already dicussed, we propose the ICA projection $\boldsymbol{B}$ computed for an image of the same kind as a private key. We use this matrix $\boldsymbol{B}$ in the watermarking of a group of (e.g. same owner's) images. Regarding the watermark, in [5] we used another image as watermark. In this paper we embed a spread spectrum mark, i.e., a message (the copyright information) "modulated" by means of spread spectrum techniques (SS) [1], hiding every bit of the message over the entire image ("holographic" property). Hence we endow the method with robustness against cropping and better synchronization properties [15]. We propose to design the watermark to have the size of one component, $n/k \times m/k$. This watermark is computed as the circular convolution of a key-dependent pseudorandom image $\boldsymbol{P}$ and an image containing the bits of the message $\boldsymbol{Q}$

$$\boldsymbol{W} = \boldsymbol{P} \otimes \boldsymbol{Q} \tag{3}$$

Let's $\boldsymbol{M}$ be a $p \times p$ matrix whose pixels are the bits of the message. We define matrix $\boldsymbol{Q}$ as follows

$$\boldsymbol{Q}(i,j) = \sum_{rs} M(r,s)\delta(i - r \cdot n_r/2, j - s \cdot n_c/2) \tag{4}$$

where $n_r = n/(k \cdot p)$ and $n_c = m/(k \cdot p)$. Therefore, matrix $\boldsymbol{Q}$ is a zero valued matrix except for the bits of the message, located at the center of each $n_r \times n_c$ block. Once we have the watermark, we perform a perceptual masking based on edge detection [16] to improve the invisibility of the watermark. Now we have the watermark ready to be embedded.

By arranging the IC components $y_t(i)$, $i = 1, ..., k^2$ in descending order of magnitude we have the low frequency, the medium frequency and the very high-frequency coefficients, in this ordering. Next, we reshape the watermark into a row vector, y_t^W, and add it to the first r host image IC, $\boldsymbol{y}_t^V$. Similarly to other frequency transform watermarking algorithms [1], we propose to embed the watermark in the r most significant IC, excluding the first one. The first IC is important since it is the low pass component and it is the more robust one to compression. However, since it is not orthogonal to spread-spectrum watermarks we cannot easily blindly extract the mark by using simple detectors. We update the host IC as follows

$$y_t^V(h) = y_t^I(h) + \alpha_h y_t^W \quad h = 2, \ldots, r \tag{5}$$

where α_h is a scaling factor to control the perception of the watermark. Other techniques such as the multiplicative or exponential approaches are possible [1]. In [5] we proposed a replacement of high-frequency components instead. However, these components can be easily removed by a simple, e.g., compression of the watermarked image.

3.2 Detection

In the detection of the watermark $\boldsymbol{W}$ from the (attacked) watermarked image $\boldsymbol{V}$ we go back on the steps of the embedding Algorithm 1, as illustrated in Fig. 3. The watermark detection yields

Algorithm 2: Detection.

1 *Watermarked image to column vectors.* Compute the components $\boldsymbol{x}_t^V$ of the watermarked image $\boldsymbol{V}$ by dividing it in $k \times k$ patches.
2 *ICA components.* Compute the IC, $\boldsymbol{y}_t^V$, of the image as $\boldsymbol{y}_t^V = \boldsymbol{B}\boldsymbol{x}_t^V$.
3 *Extraction.* Extract the watermark from $\boldsymbol{y}_t^V$.
4 *Detection.* Estimate the message and the probability of watermark detection.

Since we use a SS watermark, detection can be easily achieved by simple correlation, i.e., by using a matched filter. We first compute the IC of the watermarked image $\boldsymbol{y}_t^V = \boldsymbol{B}\boldsymbol{x}_t^V$. Then we average all components $h : \alpha_h \neq 0$, improving the signal (watermark) to noise (image+attacks) ratio,

$$\hat{\boldsymbol{y}}_t^W = \sum_{h:\alpha_h \neq 0} \boldsymbol{y}_t^V(h) \tag{6}$$

and reshape the resulting vector into matrix $\hat{\boldsymbol{W}}$. Finally, we estimate the copyright message by computing matrix $\boldsymbol{Q}$ in (4) as

$$\hat{\boldsymbol{Q}} = \boldsymbol{P} \otimes \hat{\boldsymbol{W}}^{\mathrm{S}} \tag{7}$$

Table 1. Averaged bit error rate for different attacks and images performed on ICA and DCT watermarked images, PSNR=41 dB

Attack	**BER ICA**
AWGN	0.0243
Quantization 2^2 levels	0
Median (5×5)	0.0104
JPEG 20 %	0.0747
Cropping 10%	0.0278

where $^{\mathrm{S}}$ denotes symmetry ($\boldsymbol{W}^{\mathrm{S}}(i,j) = \boldsymbol{W}(j,i)$). The values of the peaks of this convolution are the bits of the message. The probability of detection can be easily estimated by comparing these peaks to the rest of pixels modeled as zero mean Gaussian noise.

4 Experimental Results

We next include an example of robust watermarking applied to nine 512×512 intensity images. We first computed $\boldsymbol{x}_t^I$ with $k = 3$ and then the IC of the image as $\boldsymbol{y}_t^I = \boldsymbol{B}\boldsymbol{x}_t^I$, where matrix $\boldsymbol{B}$ was the one obtained for another image. The watermark was generated as the spread version of a 2-dimensional message of 8×8 bits. The watermark was added to the IC of the image number $h = 2, 3, 4, 5, 6$. The final peak signal-to-noise ratio (PSNR) was 41 dB. At this point we must emphasize that for 41 dB the visual perception of the watermark in the DCT approach was much more significant than in the ICA method. Besides, it as been further improved by means of a perceptual mask.

We performed the following attacks and obtain the averaged bit error rate (BER) included in Tab. 1. We first added white Gaussian noise with standard deviation $\sigma = 0.15$, we requantized the image to 2^2 levels, then we applied a 3×3 median filter, the image was also JPEG compressed to 20% of its original size and finally we cropped the 90% of the image.

5 Conclusions

In this paper we propose a new blind robust image watermarking algorithm, where we embed the watermark into the independent components of the image. The orthogonality between spread spectrum signals and IC is exploited in this novel blind design. This method has a little perceptual impact compared to the DCT approach. Besides, we propose a double-key algorithm, where the ICA projection and the spreading codes are needed in the retrieval of the watermark. We show that by introducing some useful and simple features in the embedding stage we greatly improve the performance of the method. These are the use of a perceptual mask and the holographic approach. In the experiments included we illustrate the good performance of this method.

References

1. Cox, I., Kilian, J., Leighton, T., Shamoon, T.: Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing **6** (1997) 1673–1687
2. Noel, S.E., Szu, H.H.: Multimedia authenticity with ICA watermarks. In: Proc. SPIE Vol. 4056, p. 175-184, Wavelet Applications VII, Harold H. Szu; Martin Vetterli; William J. Campbell; James R. Buss; Eds. (2000) 175–184
3. Yu, D., Sattar, F.: A new blind watermarking technique based on independent component analysis. In: IWDW. (2002) 51–63
4. Liu, J., Zhang, X., Sun, J., Lagunas, M.A.: A digital watermarking scheme based on ICA detection. In: Proc. ICA2003, Nara, Japan (2003) 215–220
5. Murillo-Fuentes, J., Molina-Bulla, H., González-Serrano, F.: Independent component analysis applied to digital image watermarking. In: Proc. ICASSP'01. Volume III., Salt Lake City, USA (2001) 1997–2000
6. Bounkong, S., Toch, B., Saad, D., Lowe, D.: Ica for watermarking digital images. J. Mach. Learn. Res. **4** (2003) 1471–1498
7. Murillo-Fuentes, J.J.: Independent component analysis in the watermarking of digital images. In: Proc. ICA2004, Granada, Spain (2003)
8. Hyvrinen, A., Karhunen, J., Oja, E.: Independent component analysis. John Willey and Sons (2001)
9. Murillo-Fuentes, J.J., Gonzlez-Serrano, F.J.: A sinusoidal contrast function for the blind separation of statistically independent sources. IEEE Trans. on Signal Processing. **52** (2004) 3459–3463
10. Bell, A.J., Sejnowski, T.J.: Edges are the independent components of natural scenes. In Mozer, M.C., Jordan, M.I., Petsche, T., eds.: Advances in Neural Information Processing Systems. Volume 9., The MIT Press (1997) 831
11. Lee, T., Lewicki, M., Sejnowski, T.: Unsupervised classification with non-gaussian mixture models using ICA. In: Advances in Neural Information Processing Systems. Volume 11., Cambridge, MA, The MIT Press (1999) 58–64
12. S Hornillo-Mellado, R Martn-Clemente, J.I.A., Puntonet, C.G.: Application of independent component analysis to edge detection and watermarking. In: Proc. IWANN'03, Mahn, Spain (2003) 270–280
13. Bugallo, M.F., Dapena, A., Castedo, L.: Image compression via independent component analysis. In: Learning, Leganés (2000)
14. Kerckhopffs, A.: La cryptographie militaire. Journal des Sciences Militaires, IX:5-38 (Jan), 161-191 (Feb) (1883)
15. Mora-Jimenez, I., Navia-Vazquez, A.: A new spread spectrum watermarking method with self-synchronization capabilities. In: Proc. ICIP2000, Vancouver, BC, Canada (2000)
16. Wolfgang, R., Podilchuk, C., Delp, E.J.: Perceptual watermarks for digital images and video. Proceedings of the IEEE **87** (1999) 1108–1126