INDEPENDENT COMPONENT ANALYSIS APPLIED TO DIGITAL IMAGE WATERMARKING

Francisco J. González-Serrano, Harold. Y. Molina-Bulla and Juan J. Murillo-Fuentes

Universidad Carlos III. DTC. ATSC.
Butarque, 15, 28911 Leganés, Madrid, Spain.
E-mail : murillo@tsc.uc3m.es http://alcaudon.tsc.uc3m.es/∼murillo/

ABSTRACT

The authors propose a new solution to the watermarking of digital images. This approach uses Independent Component Analysis (ICA) to project the image into a basis with its components as statistically independent as possible. The watermark is then introduced in this representation of the space. Thus, the change of basis is the key of the steganography problem. The algorithm applied to the fragile watermarking problem locates any change in the image since it also applies to the watermark. The problem of robust watermarking is also addressed from this new point of view. Some results are included to illustrate the method.

1. INTRODUCTION

Watermarking usually consists of the use of perceptually invisible authentification techniques. In this sense, "controlled" distortion is introduced in a multimedia element [1]. And the owner wants to "protect" this content. The goals are on the one hand the verification of the owner and the detection of forgeries of an original image. This is usually referred as Fragile Watermark (FW) and is mainly an authentification problem. A fragile watermark is readily destroyed if the watermarked data is only sligthly altered. On the other hand the Robust Watermarking (RW) focuses on the identification of illegal copies of the image and the prevention of unauthorized distribution. A typical scenario is that of an owner placing digital images on a network server. A robust watermark is resilient in the face of attacks that attempt to remove it, either intentionally or unintentionally.

In digital image watermarking the watermark is inserted in the image so that the watermarked image is public. The method to insert the watermark along with a private key characterizes the watermarking process. In RW, the illegal use of the watermarked image is usually associated to the image "attack". The goal of these attacks is to remove the watermark from the image. This must be as difficult as possible even if the adversary knows the method, as long as he does not have key. Many watermarking methods working either in spatial or transformed domain exists [2] [3] [4] [5]. Our approach may be classified in the second group as the watermarking is inserted in the image after it has been projected into a new space.

We propose to use the change of basis provided by ICA [6] or Blind Separation of Sources [7] methods. Consider a $k^2 \times 1$ vector $\boldsymbol{x}_t$ projected into a space of k^2 components $\boldsymbol{y}_t$ as statistically independent as possible. The ICA problem consists of finding this change of basis represented by a $k^2 \times k^2$ matrix B.

$$\boldsymbol{y}_t = B\boldsymbol{x}_t \quad t = 1, 2, \ldots \tag{1}$$

Let's matrix I be an intesity image of size $n \times m$. The matrix is divided into blocks of $k \times k$ $C_{p,q}$. The entry $C_{p,q}(i,j)$ with indexes

$$\begin{aligned} i, j &= 1, 2, ..., k \\ p &= 1, 2, ..., n/k \\ q &= 1, 2, ..., m/k \end{aligned} \tag{2}$$

is the sample $t = m(p-1)/k + q$ of the component $k(i-1)+j$ of vector $\boldsymbol{x}_t$. The analysis (and synthesis) expression yields

$$\begin{aligned} x(k(i-1)+j)_{(m(p-1)/k+q)} &= C_{p,q}(i,j) \\ &= I(k(p-1)+i, k(q-1)+j) \end{aligned} \tag{3}$$

By applying ICA to $\boldsymbol{x}_t$ we obtain the ICA components $\boldsymbol{y}_t$ of the image. Operating on these components leads to compression or codification algorithms [8]. Compression algorithms exploit this decomposition as they retain only the r ICA components with larger energy and it restores the image starting from these r components. The codification approach is based upon the idea that images with similar features may be restored from a common set of components. That is, it is possible to use ICA to define a set of basis functions to build a group of images. If the image is a RGB image, the process is applied to each color. We next

Thanks to Prof. Greg Heilemann at the University of New Mexico and to spanish CICYT for funding this project (TIC99-0219).

use the concepts above to propose a new solution to the watermarking problem.

2. WATERMARKING

In this section we apply ICA to the watermarking of an image. We first address the problem of inserting the watermark and then the watermak extraction.

2.1. Insertion

In the last section we described in (3) how to decompose an image I into components $\boldsymbol{x}_t^I$. By applying an ICA algorithm [9] [10] to the image to watermark, we get matrix B^I and ICA components $\boldsymbol{y}_t^I$. Besides, suppose the watermark is another image W of the same size than I, $n \times m$. We may obtain its ICA decomposition by dividing it in blocks of the same size as before, k. This pair of decompositions may be written as

$$\begin{aligned} \boldsymbol{y}_t^I &= B^I \boldsymbol{x}_t^I \\ \boldsymbol{y}_t^W &= B^W \boldsymbol{x}_t^W \quad t = 1, \ldots, mn/k^2 \end{aligned} \tag{4}$$

where the components, $y_t(i)$ $i = 1, ..., k^2$, have been arranged in descending order of energy.

ICA based compression methods remove the r less energy ICA components of an image. We propose to replace these r components from $\boldsymbol{y}_t^I$ with the first $k^2 - r$ ICA components of the watermark $\boldsymbol{y}_t^W$. Note that any other interchange of components may be valid. The components, y_t^V, of the watermarked image, V, yield

$$y_t^V(h) = y_t^I(h) \quad h = 1, \ldots, k^2 - r \tag{5}$$

$$y_t^V(k^2 - h + 1) = \boldsymbol{\alpha}(h) y_t^W(h) \quad h = 1, \ldots, r \tag{6}$$

where $\boldsymbol{\alpha}$ is a weight vector to control the perception of the watermark. The watermarking process outlined above may be summarized in the following algorithm

Algorithm 1: insertion.

1 *Image components*. Compute the components $\boldsymbol{x}_t^I$ and $\boldsymbol{x}_t^W$ of the image I and the watermark W by dividing it in blocks of $k \times k$ as in (3).

2 *ICA components*. Compute the ICA components $\boldsymbol{y}_t^I$ and $\boldsymbol{y}_t^W$ of the image and the watermark as in (4). Arrange the component in descending order of energy.

3 *ICA watermarked image components*. Compute the ICA components $\boldsymbol{y}_t^V$ by replacing the latest ICA components of image I by the ones of the watermark, W, as described in (5) and (6).

4 *Restoring watermarked image*. Restore the watermarked image V from components $\boldsymbol{x}_t^V = B_I^{-1} \boldsymbol{y}_t^V$ by using (3).

5 *Store the Key*. Store matrices B_I and B_W as the keys of the method.

Once we have the image with the watermark we need to define the procedure to extract it.

2.2. Extraction

The aim of this section is the extraction of the watermark W from the watermarked image V. We go back on the steps of algorithm 1. We first compute the components $\boldsymbol{x}_t^V$ as in (3). Next, the projected ones $\boldsymbol{y}_t^V$ by using B^I stored in step 5. Then we recover the ICA components of the watermark in (6)

$$y_t^W(h) = y_t^V(k^2 - h + 1) \quad h = 1, \ldots, r \tag{7}$$

and set the rest of them to zero. The image is restored by using matrix B^W in step 5. The watermark extraction yields

Algorithm 2: Extraction.

1 *Watermarked image components*. Compute the components $\boldsymbol{x}_t^V$ of the image V by dividing it in $k \times k$ blocks as in (3).

2 *Watermarked image ICA components*. Compute the components $\boldsymbol{y}_t^V$ of the image as $\boldsymbol{y}_t^V = B^I \boldsymbol{x}_t^V$.

3 *Watermark ICA components*. Compute the components $\boldsymbol{y}_t^W$ of the watermark as in (7).

4 *Restoring the watermark*. Compute the components $\boldsymbol{x}_t^W = B_W \boldsymbol{y}_t^W$ and use (3) to obtain the watermark W.

3. DISCUSSION

An important issue in watermarking extraction is the knowledge of the original image in the extraction process. Our aim is to propose an extraction method not based in this knowledge. Notice that if we had this information, the ICA components of the watermark could be added to the ones of the image instead replacing them as in (6) [11]. This way the method would be more robust to attacks.

In order to make our extraction algorithm independent of the image, the key B_I should be known at step 5. In [8] a common set of basis was proposed for natural images. In fact B in (1) is just a change of basis and it may be applied to any image. We applied this idea to our approach: Matrix B^I in step 2 of the insertion algorithm may be the ICA change of basis computed for any other image.

We use an image as the watermark. However, the watermark could be any signal such as a message. Notice that once the components of the watermark have been extracted, matrix B^W is used to restore the watermark. In this sense this second matrix is the key of a cryptographic problem.

As it will be described in the next section, the proposed method applied to FW allows to detect any change in the image by subtracting the original watermark from the extracted one. Any adversary must know the exact change of basis, the key, to manipulate the image. On the other hand, the method is flexible as the number r of components to replace is not fixed. Thus, in the case of RW this number may be increased. One of the main attacks to take into account is that of compression. In the case of JPEG compression, the watermark usually disappears from low frequency regions. A perceptual mask could be used to palliate this problem. Another point we do not face in this paper is that of the automatic detection of the watermark. We just focus on the watemark extraction and correlation with the original one. A simple threshold following this last operation would detect the watermark. Better detection techniques in the literature [12] could be adapted for our approach.

4. EXPERIMENTAL RESULTS

We first propose to apply FW to the image I in Fig.2.a. The image was divided in blocks of size 4×4. Then ICA was performed to compute matrix B_I. The first ICA component of the logo image of the University Carlos III de Madrid in Fig. 1 was introduced as the last ICA component of the watermarked image. We then applied some changes: we copied the beak of one arara on the other one and copied the eye area of this one to the other. These changes can be observed in Fig.2.b. Next we extracted the watermark and compared it to the original one to obtain image in Fig2.c. The watermark was detected except in the altered areas.

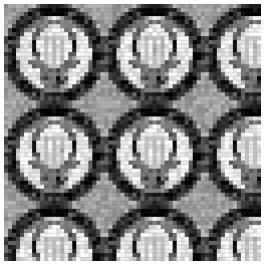

Fig. 1. First ICA component of the image used as watermark, the logo of the university Carlos III de Madrid.

We next include an example of RW applied to the intensity image of lena in Fig.3.a The watermark was, again, the logo in Fig.1. We computed the"ICA component" of the image as $\boldsymbol{y}_t^I = B_I \boldsymbol{x}_t^I$, where matrix B_I was the one obtained for image in Fig.2.a. The first ICA component of the watermark was introduced as the ICA components number 14, 15 and 16 multiplied by $\alpha = .5$. We performed the following attacks. In Fig.3.c the image was smoothed with a 3×3 mask with a ratio of 0.3 between any border entry of the mask and the central one. In Fig.3.e we cropped the 30% of the image and added white noise with variance σ^2, $\sigma/255 = 0.14$. And in Fig.3.g the image was converted to JPEG format with a compression factor of 80%. The correlation of the extracted watermarks with the watermark in Fig.1 are shown on the right hand column: Fig.3.d for the case of smoothing, Fig.3.f when cropping and adding white noise, and Fig.3.h. if the image is JPEG compressed. The watermark can be detected in these cases with a simple threshold. In the last case, only in the areas with higher frequency (face, hat and feathers) it can be observed the watermark.

(a)

(b)

(c)

Fig. 2. Illustration of Fragile Watermarking : (a) watermarked image, (b) altered watermarked image, (c) finding the altered regions.

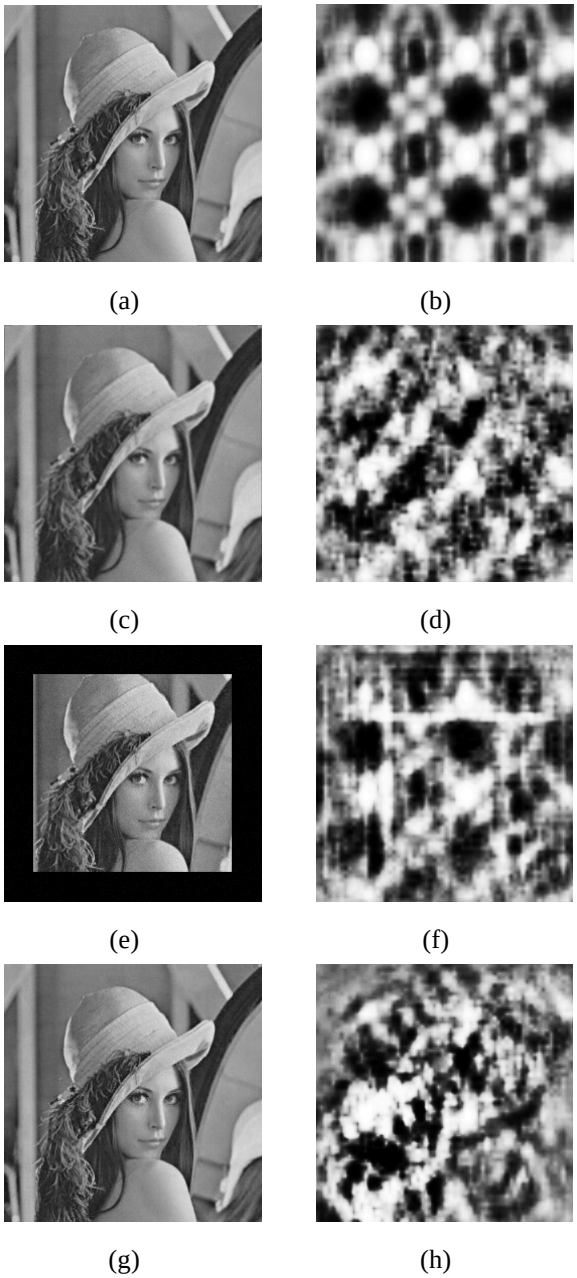

Fig. 3. Illustration of some attacks and their effect on the correlation of the extracted and the original watermark: (a) and (b) the watermarked image with no attacks, (c) and (d) smoothing, (e) and (f) cropping 30% and white noise of variance $\sigma^2 = 0.02$, and (g) and (h) JPEG compression of ratio 80%.

5. CONCLUSIONS

In this paper we present a new approach to image watermarking based on independent component analysis. The starting point is the ICA based image processing in [8]. We apply these concepts to write a new watermarking algorithm. The key of the steganographic method is the change of basis performed applying ICA to the image. This change of basis for the watermark is also a cryptographic key. The problems of robust and fragile watermarking are addressed. The experiments included show how this new method successes in extracting the watermark even when the image has been attacked. It also allows solving the authentification problem in fragile watermarking detecting any change in the image.

6. REFERENCES

[1] D. Kahn, "The history of steganography," *R. Anderson, Editor, Information Hiding, Springer Lectures Notes in Computer Science*, vol. 1174, pp. 183–206, 1996.

[2] V. Cappellini M. Barni, F. Bartolini and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, vol. 66, pp. 357–372, 1998.

[3] C.I. Podilchuk R.B. Wolfgang and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, pp. 1108–1126, 1999.

[4] J. R. Hernández and F. Pérez-González, "The impact of channel coding on the performance of spatial watermarking for copyright protection," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, May 1998, vol. V, pp. 2973–2976.

[5] N. Nikolaidis and I. Pitas, "Copyrigth protection of images using robust digital signatures," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, May 1996, http://poseidon.csd.auth.gr/signatures/, pp. 2168–2171.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[7] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceddings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[8] M. Lewicki T. Lee and T. Sejnowski, *Unsupervised classification with non-gaussian mixture models using ICA*, vol. 11, MIT Press, Cambridge MA., 1999.

[9] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, January 1999.

[10] J.J. Murillo-Fuentes and F. González-Serrano, "Independent component analysis with sinusoidal fouth-order contrast," in *International Conference on Audio, Speech and Signal Processing*, Submitted, May 2001.

[11] T. Leighton I. J. Cox, J. Kilian and T. Shamoon, "Secure spread spectrum watermarking for multimedia," Tech. Rep. 95-10, NEC Research Institute, www.neci.nj.nec.com/tr/index.html, May 1995.

[12] J. R. Hernández and F. Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE*, vol. 87, pp. 1142–1166, 1999.