Independent Component Analysis for Domain Independent Watermarking

Stéphane Bounkong, David Saad, and David Lowe

Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK,
bounkons@aston.ac.uk, saadd@aston.ac.uk, lowed@aston.ac.uk,
WWW home page: http://www.ncrg.aston.ac.uk/

Abstract. A new principled domain independent watermarking framework is presented. The new approach is based on embedding the message in statistically independent sources of the covertext to mimimise covertext distortion, maximise the information embedding rate and improve the method's robustness against various attacks. Experiments comparing the performance of the new approach, on several standard attacks show the current proposed approach to be competitive with other state of the art domain-specific methods.

1 Introduction

Interest in watermarking techniques has grown significantly in the past decade, mainly due to the need to protect intellectual property rights (IPR). Research has mainly focused on digital images, audio or video data, where economic interests are more apparent, with a plethora of techniques. In spite of their common root, the techniques developed are domain specific and cannot easily be transferred across domains, making it difficult to provide a principled comprehensive theoretical approach to watermarking. The latter is a prerequisite to a methodological optimization of watermarking methods. The present paper describes a domain independent watermarking framework which aims at maximising the information embedding rate and the robustness against various attacks while mimimising the information degradation.

2 Domain Independent Watermarking

In the past few years, significant attention has been drawn to blind source separation by Independent Component Analysis (ICA) [1]. The recent discovery of efficient algorithms and the increase in computational abilities, have made it easier to extract statistically independent sources from given data.

ICA is a general purpose statistical technique which, given a set of observed data, extracts a linear transformation such that the resulting variables are as statistically independent as possible. Such separation may be applied to audio signals or digitized images [1], assuming that they constitute a sufficiently uniform class so that a statistical model can be constructed on the basis of observations. Experiments conducted on a set of digitized images that we examined, show that this hypothesis holds, giving us a general domain independent framework [1].

The suggested framework can be based on various generative methods. In this paper we will focus on a particular method for identifying statistically independent sources - ICA. We now describe the ICA generative model and a simple watermarking scheme based on it. Technical details have been omitted for brevity.

2.1 ICA Generative Model

ICA describes a set of latent variables, also termed Independent Components (IC), which can be observed only through their linear combination. By definition, these variables are random and statistically mutually independent.

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{il}s_l, \text{ for all } i = 1, \ldots, n \tag{1}$$

where $a_{i,j}$ are real coefficients, s_i are the latent independent variables and the x_i are observed measurements. Using a matrix notation, the previous equation can be written as $\boldsymbol{x} = A\boldsymbol{s}$; and the inverse (de-mixing) process can be described by $\boldsymbol{s} = W\boldsymbol{x}$, where W is the de-mixing matrix and inverse (or pseudo-inverse if $n \neq l$) of A.

2.2 Basic Watermarking Scheme

Basic watermarking schemes can be described in three steps. Firstly, a given message m, also termed a watermark, is embedded into the covertext X (e.g. a digitized image, audio or a transformed version) providing a watermarked covertext $\hat{X}$. Then, the watermarked text may be attacked either maliciously or non-maliciously, resulting in the attacked covertext Y. Finally, a decoder tries to extract m from Y given or not side information. This is summarised in figure 1.

2.3 Domain Independent Watermarking (DIW) Scheme

In the framework studied in this paper, X may be derived from any media, such as audio signals or digitized images. The de-mixing matrix W obtained by the ICA algorithm for the different domains are different but the principle remains the same: representing the covertext through a set of IC.

Given a covertext, a set of relevant IC are chosen and modified such that they carry m. Various efficient approaches have been suggested for hiding/embedding information. We used the distortion-compensated Quantization Index Modulation (QIM) method [5], that has been shown to be close to optimal in the case

[1] In the case of multiple, significantly different, covertext groups, one may construct a different model for each group.

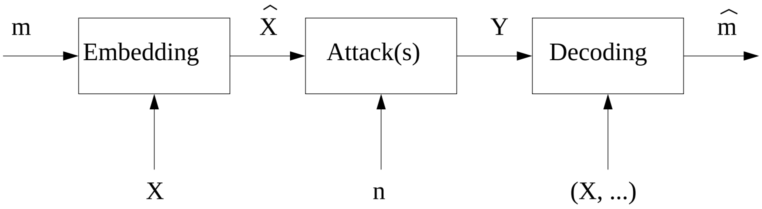


Fig. 1. A general watermarking scheme where m is the embedded message, X is the covertext, $\hat{X}$ the watermarked covertext, Y the attacked covertext and $\hat{m}$ an estimate of m.

of additive Gaussian attacks and is easy to use. It is based on quantizing the covertext real-valued IC to some central value, followed by a quantized addition/subtraction representing the binary message bit. This may also be modified by a prescribed noise template making it difficult to identify the QIM embedding process and its parameters.

The watermarked covertext $\hat{X}$ is then mixed back to the original covertext space, generating the watermarked covertext, as illustrated in Fig.2.

The decoding process proceeds in a similar way. The description of the attacked text is computed from the attacked covertext by employing the de-mixing matrix W giving us the corrupted source Y. $\hat{m}$ is computed from Y in conjunction with other available information (e.g. attack characteristics, original covertext, cryptographic key, $\ldots$; see Figure 2).

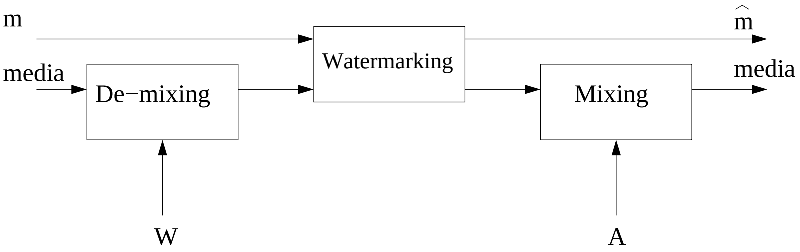


Fig. 2. This figure represents a domain independent watermarking scheme where m is the embedded message and $\hat{m}$ is an estimate of m. A and W are, respectively, the mixing and de-mixing matrices used to get the independent components.

3 Experimental Results

We carried out a few experiments, comparing the performance of our approach to other watermarking methods. The covertext used in our experiments was arbitrarily chosen to be digitized images. For the DIW approach, the latter are divided in contiguous patches. Each patch is marked independently following the method described above, see 2.3.

For comparison purposes, two other watermarking schemes have been tested under the same attacks and using the same embedding and decoding methods. Both methods operate in the discrete cosine transform (DCT) domain.

Comp1 This scheme is based on the DCT of the whole image, X, selecting a random coefficient set for the message m to be embedded in using QIM.

Comp2 In the second scheme, the image is divided into contiguous patches. The DCT of each patch is used as covertext X. A set of coefficients is selected and then quantized for embedding m.

In both schemes, $\hat{X}$ undergoes an inverse DCT, to provide the watermarked image. Notice that *local* methods such as Comp 2 and DIW are much more computationally efficient than *global* methods like Comp 1. Furthermore, watermarking parameters have been optimized in all methods, and separately for each specific attack.

3.1 Experiments

We carried out four experiments where watermarked pictures are attacked either by: a) white noise (WN) of mean zero and of various standard deviation values; b) JPEG lossy compression with different quality levels; c) resizing with various factors; d) a combination of attacks: resizing with a factor of 0.5, followed by JPEG compression with a quality factor of 70, followed by WN of zero mean and of standard deviation 15.

These attacks are, arguably, the most commonly used attacks as a benchmark in this field. The set of images used comprises eleven gray-scale pictures representing *natural*, as opposed to computer generated, scenes. The experiments are carried out ten times for each set of parameters for each picture, providing both mean performance and error bars on the measurements.

Each algorithm embeds, using a quantization method characterized by a quantization step δ, a message m of length 1024 bits with a maximum distortion of 38 dB as suggested in [3,4]. The distortion induced by the watermarking systems is measured by the peak signal to noise ratio (PSNR). A simple decoding scheme based on nearest decoding is also used for all systems. Table 3.1 summarises the parameters used in the experiments.

Table 1. Summary of the watermarking schemes parameters

Attack			Noise		JPEG		Resizing	
Scheme	Transform	Patch size	Coef. Rg.	δ	Coef. Rg.	δ	Coef. Rg.	δ
DIW	ICA	16 by 16	38-50	155	6-10	36	6-10	36
Comp1	DCT	-	101-1124	70	2081-20624	70	2-1985	70
Comp2	DCT	16 by 16	6-23	80	2-19	80	4-18	80

3.2 Results

Figure 3a, shows that all schemes are quite robust considering that the 38 dB attack distortion threshold is reached for a standard deviation of about 3. It also shows that DIW is the most robust method of those examined for a WN attack. In the case of DIW and the decoding method used, it is easy to see a direct relation between δ and the robustness of the process, since the noise in the feature space is also Gaussian. This may not be the case if other decoding methods, such as the Bayesian approach will be used. Moreover it also shows that one potential weakness of the DIW scheme, the ICA restriction of extracting only non-Gaussian sources, is not highly significant, even in the case of a Gaussian noise attack.

Figure 3b shows that all systems are quite robust against JPEG compression. However, for very low quality levels, under 15, performances decrease significantly, and are less stable as shown by the error bars. Furthermore the threshold of 38 dB distortion is reached at a quality level of about 90. DIW achieves here the best results on average.

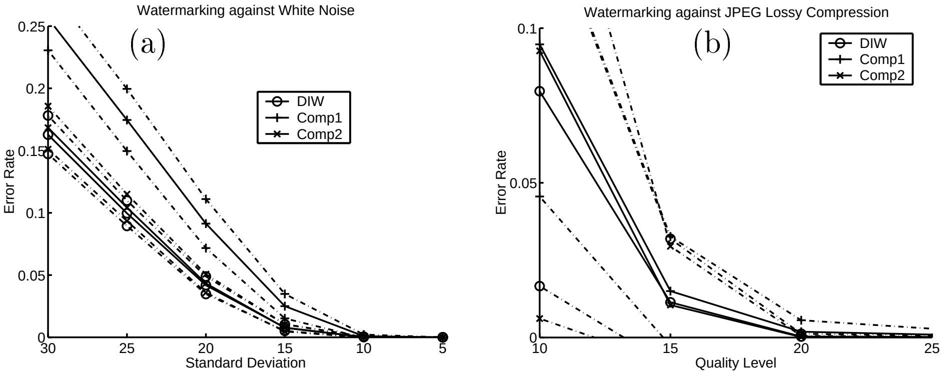
Figure 3c shows excellent performances for Comp1 under resizing attacks. DIW and Comp2 achieve excellent results for resizing factor greater than 0.5; their performances decrease significantly for stronger attacks. Intiutively this can be explained by their patches' localised nature. Low resizing factors affect severly the capacity of these schemes and the picture quality. For a 0.25 resizing factor, the picture size is reduced by more than 93% in storage.

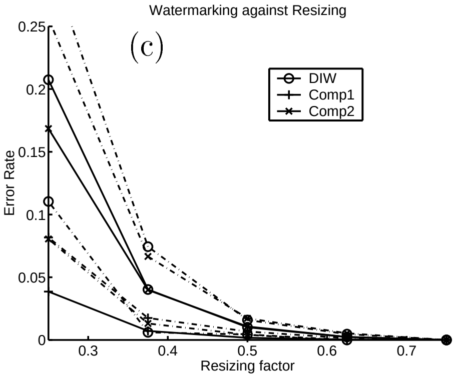
Figure 3d shows the results of the schemes against a combination of attacks based on a possible scenario. It appears that Comp2 performs better than DIW (which performs better than Comp1), presumably due to the resizing component.

4 Conclusions

A new principled domain independent watermarking framework is presented and examined. Experiments show highly promising performance in comparison with other state of the art methods on a limited set of attacks. The attacks include four of the most common attacks: white noise attack, JPEG lossy compression, resizing and a combination of attacks.

The main advance is that since the watermarking combines an information-theoretic embedding across a space of statistically independent sources, the same technique works across different media. Being based on local information and a linear transform, our method is economical in the computational costs required (unlike global methods relying on non-linear transforms like Comp1) and offers additional security in the use of *specific* mixing/de-mixing matrices that are not easy to obtain (in contrast to methods based on a simple transformation like Comp1 and Comp2). Further research will focus on theoretical aspects of this scheme, optimizing the decoding process and other improvements of its robustness against specific attacks.





(d)

			Error Rate	
Scheme	Coef. Rg.	δ	Mean	Std. Dev.
DIW	6-10	36	0.23	0.016
Comp1	2-1985	70	0.27	0.019
Comp2	4-18	80	0.18	0.015

Fig. 3. The performance of the three watermarking tested: DIW, Comp1 and Comp2, against various attacks; solid lines and symbols represent the mean values; dashed lines denote error bars. (a) White noise of different standard deviation values. (b) JPEG lossy compression for different quality levels. (c) Resizing for different factor. (d) Combination of attacks: resizing 0.5, followed by JPEG 70, followed by WN 15, the attack distortion has a PSNR of about 23 dB.

References

1. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis.* Wiley-Interscience, NY (2001).
2. Cox, I., Miller, M. L., Bloom, J. A.: *Digital Watermarking.* Principles and Practice, Morgan Kaufmann, SF (2001).
3. Petitcolas, F. A. P., Anderson, R. J.: *Evaluation of Copyright Marking Systems.* IEEE Inter. Conf. on Multimedia Computing and Systems. **1**, 574–579 (1999).
4. Petitcolas, F. A. P., Anderson, R. J., Kuhn, M. G.: *Information Hiding - a Survey.* Proceeding of IEEE Multimedia Systems 99. **87-7**, 1062–1078 (1999).
5. B. Chen and G.W. Wornell, *Quantization Index Modulation : A Class of Provably Good Methods for Digital Watermarking and Information Embedding*, IEEE Trans. Inform. Theory. **47-4**, 1423–1443 (2001).