Vulnerability of DM watermarking of non-iid host signals to attacks utilising the statistics of independent components

Patrick Bas, Jarmo Hurri

Vulnerability of DM watermarking of non-iid host signals to attacks utilising the statistics of independent components

Patrick Bas (1,2)
Jarmo Hurri (3)

(1) Laboratoire des Images et des Signaux de Grenoble
961 rue de la Houille Blanche Domaine universitaire
B.P. 46 38402 Saint Martin d'Hères cedex FRANCE

(2) Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 5400 FI-02015 HUT FINLAND

(3) Helsinki Institute for Information Technology
Basic Research Unit, P.O. Box 68
FIN-00014 University of Helsinki FINLAND

Abstract

Security is one of the crucial requirements of a watermarking scheme, because hidden messages such as copyright information are likely to face hostile attacks. In this paper, we question the security of an important class of watermarking schemes based on Dither Modulation (DM). DM embedding schemes rely on the quantisation of a secret component according to an embedded message, and the strategies used to improve the security of these schemes are the use of a dither vector and the use of a secret carrier. In this paper we show that contrary to related works that deal with the security of spread spectrum and quantisation schemes, for non-iid host signals such as images, Principal Component Analysis is an inappropriate technique to estimate the secret carrier. We propose the use of a blind source separation technique called Independent Component Analysis (ICA) to estimate and remove the watermark. In the case of DM embedding, the watermark signal corresponds to a quantisation noise independent of the host signal. An attacking methodology using ICA is presented for digital images; this attack consists firstly in estimating the secret carrier by an examination of the high-order statistics of the independent components, and secondly in removing the embedded message by erasing the component related to the watermark. The ICA-based attack scheme is compared with a classical attack that has been proposed for attacking DM schemes. The results reported in this paper demonstrate how changes in natural image statistics can be used to detect watermarks and devise attacks. Different implementations of DM watermarking schemes such as pixel, DCT and ST-DM embedding can be attacked successfully. Our attack provides an accurate estimate of the secret key and an average improvement of 2dB in comparison with optimal additive attacks. Such natural image statistics -based attacks may pose a serious threat against watermarking schemes which are based on quantisation techniques.

1 Introduction

After more than ten years of active development by the watermarking scientific community, many of the proposed watermarking techniques are considered to be mature because they are robust while preserving the quality of the host data. However, if robustness and fidelity are mandatory requirements for an usable watermarking scheme, security is also a very important issue that is unfortunately rarely addressed. While *robustness* denotes the ability to decode the watermark after various operations (compression, filtering, noise addition or geometric transforms), and *fidelity* denotes the property that the watermark is imperceptible, *security* is a more complex notion.

1.1 Security of watermarking schemes

According to [1][2], a watermarking scheme is considered secure if it is not possible to access the hidden message channel. For example if an attacker is able to replace a hidden message by another one, then the watermarking scheme does not satisfy the security constraint. Many watermarking schemes claim to be secure because they use a secret key during the embedding and detection process. However, this hypothesis is often too weak in real application scenarios and several security attacks have already been proposed. They can be based on a full access to the detection process [3], the use of a symmetric detection scheme [4], or information leakage when a database of hosts is watermarked using the same secret key [2][5][6].

This paper focuses on the security of an important class of watermarking schemes based on quantisation called Dither-Modulation (DM). The contribution of this paper is to show how changes in the statistics of the host signal can be used to detect watermarks and devise attacks against watermarking schemes. In general, watermarking may change the statistical properties of the host signal, and this can be used to devise detection and attack schemes. Here we show how a method used in blind source separation (BSS) called Independent Component Analysis (ICA) can be used to separate the watermark component from the host component in the case of host signals that are not independent or not identically distributed (non-iid). As demonstrated by the results of our paper, such host statistics -based attacks may pose serious threats against watermarking schemes.

The use of BSS techniques for watermarking has already been proposed in different contexts. BSS techniques can be used either to decompose the host signal into a transform domain that will be watermarked, to extract the watermark, or to remove the watermark.

1.1.1 BSS and decomposing the host signal

In [7], an ICA decomposition of digital images was used as a transform domain to embed the watermark. The components were ordered according to their order of energy and the least significant ones were substituted by the watermark. The detection was done by extracting again the images and the components.
In [8], an ICA decomposition of images was used for the same purpose. The authors proposed to watermark several independent sources by using quantisation techniques. They also presented a MAP decoding strategy that relies on the distributions of the independent sources.

1.1.2 BSS and extracting the watermark

In [9] a method to hide images in images was proposed. The embedder provides several mixtures containing the host image and the image to be hidden, the decoder uses ICA to identify the parameters of the mixture and extract the hidden image.

1.1.3 BSS and designing removing attacks

In [10], BSS techniques were used for spread spectrum steganography [11] to extract a carrier that has been embedded with different magnitudes but on same host signals. The paper gave conditions for secure spread spectrum steganography: the statistic of the watermark and the host signal that should have both gaussian distributions to prevent separation by ICA techniques. In [12], authors proposed a removing attack for spread spectrum watermarking scheme that uses the sparse code shrinkage algorithm [13]. This algorithm uses a decomposition of images into independent components. It analyses and alters the sparseness of each component to denoise the image. Authors also proposed a watermarking scheme that includes the denoising process at the embedding stage to cope with this attack.

As explained in detail below in this paper, our method addresses the identification and estimation of the watermark and does not only consider BSS techniques as a tool to remove noise from the host signal as in [12]. Additionally, contrary to [10] the proposed attack does not require that the watermarked contents are different watermarked versions of the same original content. We use two important properties of the watermark to design the attack: the fact that it is independent of the host signal and the fact that it has a singular distribution. Independency enables us to decompose the watermarked signal on a basis of independent vectors, one of which is the watermark component. The distribution of the watermark enables to identify the vector that is related to the watermark.

1.2 Contents of the paper

The rest of the paper is divided into six sections. First, principles of DM and Spread Transform DM (ST-DM) watermarking schemes are presented in section 2. The security of these schemes is also discussed and we outline the different strategies that have been proposed to achieve security. In section 3, we describe a previously proposed attack against ST-DM watermarking, and outline its limitations in the case of non-iid host signals. Section 4 proposes a technique to estimate the secret carrier after DM and ST-DM embedding. Our methodology is applied on digital images which are a good example of non-iid signals. We motivate the use of Independent Component Analysis as a tool to separate the watermark component from the features of the image. The decomposition of natural images into independent basis vectors is also presented. Section 5 presents the details of the two main ideas of this paper: the estimation and identification of the secret carrier and the attack that is used to remove the watermark. These two methods both use the decomposition of image blocks into a basis of independent vectors. In Section 6 we report the performance of the estimation and denoising scheme for four different implementations of DM: pixel DM, frequency-space DM on 2 different DCT coefficients, and ST-DM. We compare our attack with another attack and show that our attack provides a better ability to remove the watermark with very small distortion. Section 6 also provides a study of the performance of the attack as a function of the frequency of the secret carrier. Finally section 7 draws conclusions and gives directions for future work.

2 Dither Modulation schemes

The aim of this section is to present a well-known class of watermarking schemes called Dither Modulation (DM). The extension of DM using a spread transform is also described. A general discussion of the different ways to achieve secure DM schemes concludes this section.

2.1 Principles of DM watermarking

The class of DM watermarking schemes was first presented by Chen and Wornell [14]. The principle of the basic DM technique is to embed a binary message $b(m)$ in an host sample x by applying a quantiser on a modified component

to obtain the watermarked component x_w:

$$x_w(x;b) = q(x + d(b)) - d(b)$$

where $q(x)$ is a quantisation function with a quantisation step equal to Δ, and $d(b)$ is called the dither vector that is function of the transmitted bit b:

$$d(b=1) = \begin{cases} d(b=0) + \Delta/2 & \text{if} \quad d(b=0) < 0 \\ d(b=0) - \Delta/2 & \text{if} \quad d(b=0) \geq 0 \end{cases}$$

One basic solution is to choose $d(b=0) = 0$ which is equivalent to having a set of two disjoint quantisers where each quantiser has a quantisation step equal to Δ and each quantisation cell at a distance of $\Delta/2$ from the closest one. Such embedding quantisation grid is illustrated in Fig.1.

The main problem of this basic and very simple embedding scheme is that it is public, consequently anybody can

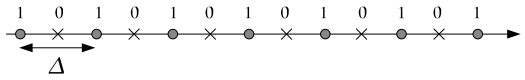


Figure 1: Quantisation grid obtained after applying DM and a null dither component.

access the watermarked components and decode the watermark once the quantisation step Δ is known. Computing the pdf of the watermarked components, or in the practical case, a histogram allows one to estimate the parameter Δ using maximum-likelihood as for example in [15][16]. In order to cope with this problem an implementation of DM consists of choosing a dither vector $d()$ that is a pseudo-random noise that prevents the estimation of the quantisation cells (this issue is presented in section 2.3.1).

In addition to the basic DM, a variant with the name of Distortion Compensated DM, also called Scalar Costa's Scheme (SCS), has been presented by [14, 17] as a way to control the distortion introduced during the DM embedding. In SCS, the quantisation error signal is multiplied by a distortion parameter $\alpha \in [0;1]$. For sake of simplicity, our study focuses on classical implementation of DM schemes which is equivalent to the case $\alpha = 1$. Note that this choice is not restrictive, since choosing $\alpha \neq 1$ will only reduce the variance of the quantisation noise and not the shape of its pdf.

2.2 Spread Transform Dither Modulation schemes

Spread Transform Dither Modulation (ST-DM)[18] can be seen as an extension of the DM scheme where the quantised value is not a host coefficient, but the projection of the host signal vector $\mathbf{x}$ on a spread vector $\mathbf{v}$:

$$\mathbf{x_w} = \mathbf{x} - x\mathbf{v} + x_w\mathbf{v}$$

where $x = \mathbf{x}^T\mathbf{v}$ and $x_w = q(x + d(b)) - d(b)$ and $\|\mathbf{v}\| = 1$. The term $-x\mathbf{v}$ cancels the interference between the spread vector and the host signal, and the term $x_w\mathbf{v}$ is the spread vector with a quantised magnitude. ST-DM has two main advantages. Firstly it provides increased robustness against additive noise. Secondly, provided that the spread vector is a white pseudo-random signal, the distortion is spread over space and frequency.

2.3 Security measures for DM watermarking

In this section we point out the different measures that are used in DM watermarking in order to achieve security, e.g. to prevent an attentional modification of the embedded message. Note that the robustness of a DM watermarking

scheme is compromised if the security measures fail. For example one easy way to alter the embedded message is to perform a "moving attack", as presented in [19], which consists in moving the quantised value in such a way that the attacker can be sure that the modified coefficient will after encode the opposite bit. The "move" can be done by adding $\pm(\Delta/4+\epsilon)$ as illustrated in Fig.2.
It is important to point out that the moving attack cannot be performed if the attacker does not know which coefficients in the host signal are watermarked. If the watermarked coefficient is not known, this attack has to be applied on all coefficients of the host signal and the resulting distortion will be larger. This suggests that secrecy of the watermarked coefficients is a key issue in achieving security for DM watermarking schemes. The following solutions have been proposed to prevent a blind estimation of the watermarked coefficient.

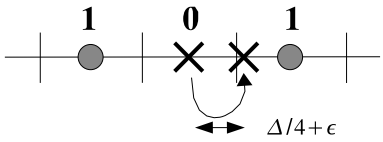


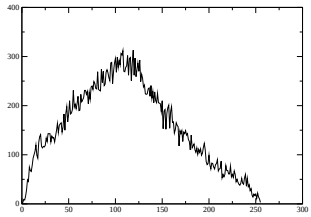
Figure 2: Principle of the moving attack.

2.3.1 Secrecy by use of a dither vector

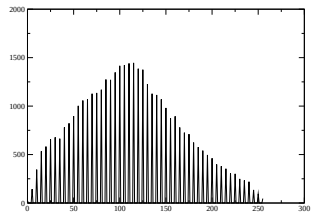
One can try to to achieve security by designing the embedding scheme in such a way that the distribution of the component that is watermarked is similar before and after the embedding. This condition has been proposed by Cachin and is based on the distance between the pdf of the original and watermarked components [20]. In [14] the authors proposed to use a pseudo-random dither component $d(b=0)$ that depends on a secret key (the seed of the random number generator for example) to fulfil this condition. The dither sequence may for example represent a uniform distribution. Under the hypothesis that the host signal is locally uniform around the quantisation cell, the watermarked signal will stay locally uniform after the embedding. If we consider samples such as image pixels, the quasi-invariance of the pdf after DM embedding can also be presumed for smooth distributions and small quantisation steps as illustrated in Fig.3.
One might argue that security has been achieved because the quantisation cells are no longer disclosed and it is not possible to access the watermarked components without knowing the dither sequence. From the point of view of security, dither modulation can be seen as a method to hide the real distribution of the watermarked component. After embedding, the marginal distribution of the component will stay very close to the original marginal distribution. This means that as far as we only consider the marginal distributions of a uniform host signal, the Cachin criterion of security is fulfilled.
However, in the case of DM watermarking, due to the quantisation process, the embedding can be modelled as the addition of uniform noise between $[-\Delta/2;\Delta/2]$ (this property is true even if the dither component is null) and this noise is independent of the host signal. This last property, independence between the watermark signal and the host image, is utilised in this paper in the design of an efficient attack.

2.3.2 Applying DM on a secret subspace

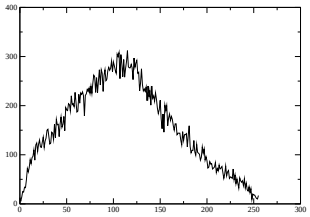
Another simple way that is used to achieve secrecy in a number of watermarking schemes, including the DM scheme, is to watermark only a set of selected coefficients of the host image, or a projection of the host signal on a secret vector as is done in ST-DM. This can be seen as the extraction of a secret subspace from the original host space. In DM, the selection of the coefficient can be done using a secret key that is used to generate the position of the embedded coefficients. In ST-DM, the pseudo-random projection vector $\mathbf{v}$ that depends on a secret key is used to generate the



Original blocks.



DM with a null dither vector.



DM with a uniform dither vector.

Figure 3: Histograms of 40,000 pixels randomly chosen from different images. With a dither vector the embedding cannot be observed just by looking at marginal distribution.

subspace. Without this knowledge, the attacker will have to process all of the possible coefficients or possible spread vectors to perform a successful attack. This security measure can be of course combined with dither modulation.

3 Previous work on attacks utilising carrier estimation

In this section we present earlier work on the different ways to perform blind carrier estimation, for different classes of watermarking schemes. We focus especially on the difficulty of estimating the secret carrier in the case of non-iid signals.

3.1 Previous work utilising ICA

A consequential work on watermarking security was done by Cayre *et. al.*[2][21]. The authors analysed both the theoretical and practical security of blind spread spectrum watermarking schemes and quantisation schemes. In these schemes, the watermark is embedded using a set of pseudo-random and orthogonal carriers (one carrier for each transmitted bit) that are modulated and added to the host signal. One part of the work was devoted to the estimation of the carriers using content data that had been watermarked using the same carriers but different messages. The authors expressed such a situation by the following equation:

$$\mathbf{Y} = \mathbf{X} + \alpha \mathbf{U}\mathbf{A}$$

where α is a scale proportional to the power of the watermark and the meaning of the other matrices $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{U}$ and $\mathbf{A}$ that compose this equation is explained in Fig.4. Each content is considered as a realization of an independent and identically distributed (iid) gaussian process.

The authors have proposed to estimate the matrices $\mathbf{U}$ (the carriers) and $\mathbf{A}$ (the embedded messages) using only the

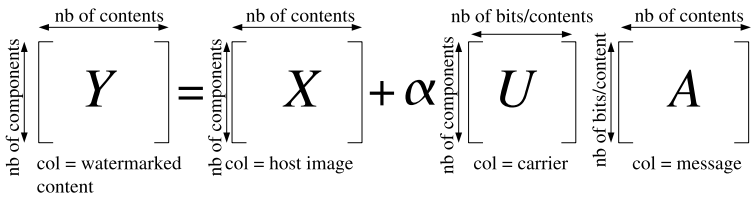


Figure 4: Details of the model equation proposed by Cayre *et. al.* to model BSS watermarking for a database of contents.

matrix $\mathbf{Y}$ (the set of watermarked images) by using Independent Component Analysis [22]. In this case, the independent sources are the embedded messages, since they are independent of each other. Thus ICA is an appropriate tool to estimate both $\mathbf{U}$ and $\mathbf{A}$. A limitation of this technique is the fact that both the decoded message and the carriers

are estimated up to sign (or bit flipping), this is due to the ICA technique itself. While it is not possible to decode the watermark, it is however possible to remove it or to alter it, and the authors have successfully used this technique to remove watermarks embedded in images and have shown that spread spectrum watermarking techniques are not secure under these hypotheses.

3.2 PCA and DM carrier estimation

In [21] the authors have also proposed to deal with the case of one unique carrier and to apply their methodology on quantisation schemes. Because independence can not be exploited anymore in this particular case (there is only one carrier), the authors propose to use Principal Component Analysis (PCA) to estimate the carrier (a similar approach has been proposed in [23] for spread-spectrum watermarking techniques and watermark subspace estimation). PCA is a method that estimates *principal components*, i.e., uncorrelated components along which the data has variance optima (including the components with the largest and smallest variances) [22]. Considering the host as an iid process, the principal component of the host should be negligible in comparison with the component given by the carrier.
However, in the case of non-iid host signals, it is not possible to use PCA to estimate the secret carrier that is used by DM watermarking. Consider that the watermarked observations $\mathbf{Y}$ are given by:

$$\mathbf{Y} = \mathbf{X} + \mathbf{u}\mathbf{a}^T$$

where $\mathbf{Y}$ and $\mathbf{X}$ are defined as in 3.1, $\mathbf{u}$ is a vector that represents the carrier to be estimated, and $\mathbf{a} = \{a_1, ..., a_i, ..., a_m\}$ is a vector whose ith element represents the magnitude of the embedding of the carrier $\mathbf{u}$ for the block i. The principle of PCA is to compute the covariance matrix of $\mathbf{Y}$ and its eigenvectors (the principal components). If the host signal $\mathbf{X}$ is iid, with variance σ_X^2, the covariance matrix of $\mathbf{Y}$ becomes:

$$\mathbf{C_Y} = (\mathbf{u}\mathbf{a})^T\mathbf{u}\mathbf{a} + \sigma_X^2\mathbf{I}$$

If we assume that $\mathbf{a}$ is iid, with a variance equal to σ_a^2 (with the relation $\sigma_a^2 = (\Delta/2)^2/12$ for DM with a uniform host), the last equation becomes:

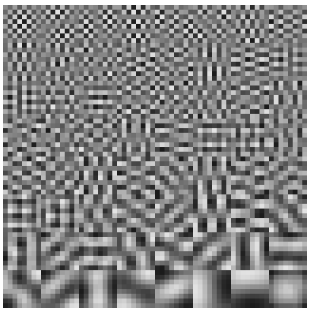
$$\mathbf{C_Y} = \sigma_a^2\mathbf{u}^T\mathbf{u} + \sigma_X^2\mathbf{I}$$

and examining the eigenvalues of $\mathbf{C_Y}$ and selecting the eigenvector associated with the largest eigenvalue $(\sigma_X^2 + \sigma_a^2)$ should provide a good estimate of vector $\mathbf{u}$. However, this estimation is not possible if the host vector is not iid because then the eigenvalues and eigenvectors of $\mathbf{C_Y}$ depend mainly on $\mathbf{C_X}$ and even in this ideal setup the addition of coloured noise will mask the watermark.

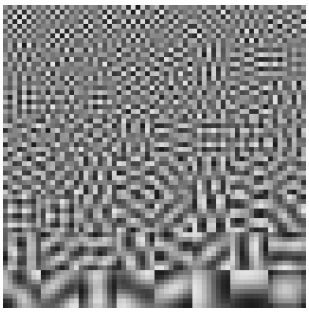
For example the assumption of an iid host is not realistic in the case of natural images when the carrier has the size of a small block (8x8, 16x16, 32x32). For small patches of natural images, the host can not be modelled by an iid process and PCA results in low frequency components with large associated variances (eigenvalues of the covariance matrix) and high frequency component with low variances (cf Fig.5). Consequently the carrier can not be estimated using PCA on small image blocks.

4 Security of DM schemes applied on digital images

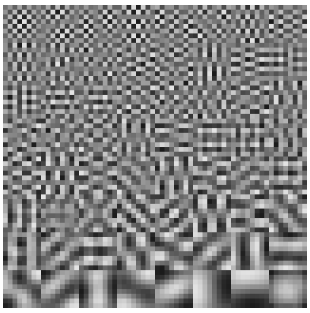
The rest of the paper presents a method developed to attack DM-watermarked digital images; these images represent one class of non-iid signals. We show that blind source separation techniques enable us to extract the carrier as an independent component of the host signal. The methodology that we use is the decomposition of watermarked images



Principal components of blocks. Original image.



Principal components of watermarked blocks, spatial embedding.



Principal components of watermarked blocks, DCT embedding.

Figure 5: PCA basis vectors for 20,000 original and watermarked image blocks of size 8×8. Blocks are ranked from the top-left corner to the bottom-right corner according to their variance in increasing order. For spatial embedding, the embedding is done on one pixel of each block and the resulting PSNR is equal to 51.1dB. For DCT embedding, the embedding is done on the coefficient (4,4) and the resulting PSNR is equal to 43.8dB. No qualitative difference is noticeable between the three sets. To counter the inherent sign-indeterminacy problem in PCA, several principal components have been flipped by multiplying them with -1 to better illustrate the qualitative match of the components.

into independent components. Note that similar attacks can be designed for other signals such as audio or video signals with a similar approach.
This section first defines the different elements used by the attacker: the nature of his observations, the possible DM algorithm and his knowledge on the watermarked contents. The principle of image decomposition into an independent basis is introduced thereafter, and the decomposition is applied to DM-watermarked images.

4.1 Attacker's material

4.1.1 Studied DM schemes

This study focuses on the security of three different embedding DM schemes for digital images (but the method can also be applied to other coefficient embeddings):

- Embedding in the pixel domain: one pixel in each $N \times N$ block embeds one bit using DM, the other pixels are left unchanged. The goal is to find the location of the watermarked pixel and then perform an attack which will alter the message while minimising the distortion.
- Embedding in the DCT domain: this situation is similar to embedding in the pixel domain, but here one DCT coefficient carries the message.
- Embedding in one projected domain: using the ST-DM method, each $N \times N$ block is watermarked using a secret carrier $\mathbf{v}$. Our goal is to first estimate $\mathbf{v}$ and then perform an attack which will alter the message while minimising the distortion.

Note that these three embedding schemes are all cases of ST-DM. For pixel embedding the spread vector $\mathbf{v}$ is equal to δ_{k_1,k_2} where δ is the Kronecker delta function and (k_1, k_2) is the location of the quantised pixel.
For DCT embedding the equivalent spreading vector is given by:

$$\mathbf{v}_{k,l} = \{v_{i+jN} = \frac{2}{N}\cos(\frac{\pi k(i+0.5)}{N})\cos(\frac{\pi l(j+0.5)}{N}), 0 \leq i < N, 0 \leq j < N\}$$

where (k, l) denotes the frequencies of the DCT coefficient.

4.1.2 Knowledge of the attacker

It is also important to point out what sort of information the attacker has at his disposition in order to perform an attack. Our work assumes that the attacker has access to content data that are watermarked with the same key. There is no restriction on the embedded messages; these are assumed to be unknown. This attack belongs to the class of WOA (for Watermarked contents Only Attack) defined in [2].

4.1.3 Practical hypotheses

In the rest of this paper we assume that the watermark is embedded in a block, and only one "component" of the block is watermarked using the DM algorithm. By the term "component" we mean either a pixel of the block (which is often the case when the location is secret), a DCT coefficient of the DCT transform of the block (which can also be used for robust DM watermarking) or a projection of the block on a secret carrier $\mathbf{v}$ in ST-DM.
The attacker processes a set of blocks of size $N \times N$, with the goal to estimate the secret carrier that has been used and subsequently to remove the watermark. The carrier can represent the pixel location, the DCT coefficient location, or the spread sequence. This set-up may correspond to two different watermarking scenarios:

1. The "database and location" scenario: here the secrecy relies on the fact that a secret subset of blocks is watermarked for each image. The same key is used for each image in the database, meaning that the set of secret watermarked locations is the same for each image. The attacker has to analyse each possible location in order to find the locations of the watermarked blocks.

2. The "one coefficient" per block scenario: in this case we assume that the attacker has only one watermarked host image that is partitioned into blocks of size $N \times N$ and in each vector only one component is modified using DM. This scenario is for example used by watermarking schemes embedding a watermark at a capacity of one bit per block.

4.2 Independent basis vectors of natural and watermarked images

4.2.1 Carrier estimation and independent component analysis

Estimating and removing a carrier can be seen as a blind source separation problem where one has a decomposition of the watermarked content into two subspaces: one representing the watermark and the other representing the features of the original content. This paper proposes to estimate the secret carrier by exploiting the fact that the carrier is statistically independent of the components of host content. This property enables us firstly to use the decomposition in independent vectors to estimate the secret carrier, and secondly to design an attack by processing the independent signal related to the watermark. Next we introduce the model that is used to generate a basis of independent vectors from blocks of images.

4.2.2 Estimation of independent basis vectors

Independent component analysis can be used as a generative model of image data [22]. In this model each image block $\mathbf{x}$ can be expressed as a linear combination of independent components referred to as the source signals $\{s_i\}$:

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_i \mathbf{a}_i s_i$$

where $\mathbf{A}$ is a constant matrix called the mixing matrix and the vector $\mathbf{a}_i$ denotes the i^{th} column of $\mathbf{A}$. These vectors are called the features or basis vectors. The decomposition of images patches (blocks) into independent basis vectors has been widely used in both image processing and computational modelling of the visual system. The independent basis vectors are similar to edge and lines detectors; neurons performing similar feature detection have been found in the mammalian visual system [24][25]. Moreover, because in natural image data the distribution of basis vectors is often sparse, the decomposition of images into independent components has also been used for image denoising [13] and image coding [26]. ICA techniques can be used to estimate the mixing matrix $\mathbf{A}$ from a set of N image blocks $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ (which are unfolded into vectors) using the matrix formulation:

$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

To illustrate, each term of this equation is explained in Fig.6. The estimation of matrices $\mathbf{S}$ and $\mathbf{A}$ can be performed

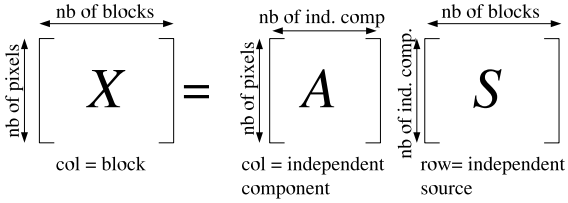


Figure 6: Details of the equation used for computing independent basis vectors based on images patches.

using different ICA algorithms; we have decided to use FastICA because this algorithm achieves good performance both in computational cost and reliability of the extracted basis vectors [22]. An example of a set of ICA basis vectors obtained from 8×8 blocks of natural images is given in Fig.7. Verbally, these basis vectors are typically described as localised, oriented and band-pass filters.

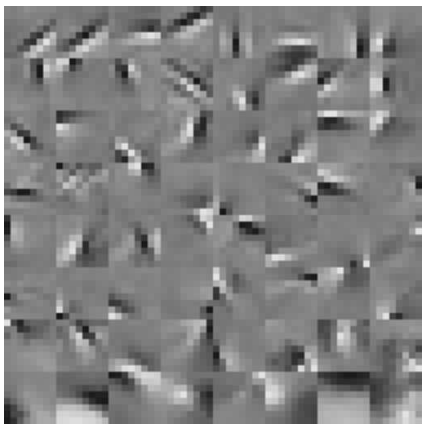


Figure 7: A complete set of 64 ICA basis vectors, computed from 20,000 8×8 pixel samples taken from natural images.

4.3 Natural image statistics and the singularity of the watermark distribution

The possibility to identify the carrier with ICA is based on the following principles:

1. the watermark is statistically independent of the image data
2. the distribution of the watermark is singular in the set of independent components, that is, the distribution of the watermark is qualitatively different from the distribution of the independent components observed in natural images.

The details of the estimation of the carrier using these principles is covered below; here we provide an intuitive overview.

A fundamental property of Independent Component Analysis is that it finds source estimates s_i that are "maximally nongaussian", and an information-theoretic measure called *negentropy* can be used as a measure of this nongaussianity. While "maximal nongaussianity" may at first seem logically separate from the principle of statistical independence, the two become the same when we observe that

1. by the definition of the source signals s_i, in ICA we are dealing with latent random variables that have zero mean and fixed unit variance (in practice this is achieved by centring and whitening the observed mixtures $\mathbf{x}$); the gaussian random variable plays a special role here since in the family of random variables with zero mean and unit variance it is the one that has the *largest* differential entropy [27]
2. information-theoretic analysis shows that the fundamental measure of statistical dependence, the mutual information, of zero-mean unit-variance random variables is minimised when the differential entropy of the individual sources $H(s_i)$ is minimised [22].

Therefore, doing ICA is equal to minimising the differential entropy of the sources, and since in the ICA family of random variables a gaussian random variable has maximal differential entropy, doing ICA is equal to finding the maxima of negentropy, which is defined as

$$J(s) = H(s_{\text{gauss}}) - H(s), \tag{1}$$

where s_{gauss} denotes a gaussian random variable with the same mean and variance as s.

The principles of statistical independence of the watermark, nongaussianity of the sources and singularity of the distribution of the watermark are illustrated in Figure 8. This figure shows the independent component basis vectors (columns of matrix $\mathbf{A}$) estimated from watermarked natural image data, and the histograms of four different estimated independent components, s_1, s_2, s_{32}, and s_{63}. The basis vectors shown in the image have been ordered – from left to right and top to bottom – according to the negentropy $J(s)$ of the corresponding source: source s_1 (top left in the figure) has the lowest negentropy and source s_{63} (bottom right) the highest. In this case watermarking has been done in the pixel space, using a carrier that corresponds to a single pixel. The following observations can be made from these results:

- First, regarding the statistical independence of the watermark, since the watermark is independent of the image data, it has been estimated by the ICA algorithm as one of the sources, and the corresponding carrier becomes one of the basis vectors (top left).
- Second, regarding nongaussianity of the sources, note that the majority of the components have high negentropy values and are highly nongaussian. This is a general property of the independent components of natural image data [22].

- Third, regarding the singularity of the watermark, note that when compared with the other sources, the watermark s_1 has a very low negentropy. This low negentropy value can be used to identify the watermark from the other sources.

The low negentropy value of the watermark is a result of DM embedding process. For uniform quantisers, the quantisation noise has a uniform pdf [28] and is independent (at least using a linear formulation of independency) of the original signal. In the case of DM embedding the quantisation noise can be considered as uniform between $-\Delta/2$ and $\Delta/2$. If we consider that independent components of a natural image have very sparse distributions [13], then the negentropy of the component related to the quantisation noise is smaller than the negentropy of the other components. Note that other criterions have been used in ICA to estimate independent components such as Kurtosis, Infomax or joint approximate diagonalisation of eigenmatrices [22]. However, the use of negentropy provides a fast and reasonably reliable way to estimate independent components. The reliability of estimating the carrier depends on the introduced distortion, the number of analysed blocks and the selection criterion. In the case of negentropy, research conducted to assess the reliability of the estimated components can be found for example in [29].

Fig.9 illustrates the capability of ICA to estimate the secret carrier in the case of both pixel and DCT DM. In the pixel domain embedding, we have watermarked the central pixel of 20,000 blocks taken from natural images; the resulting PSNR after the embedding is 51.1dB. These results – obtained by using the FastICA algorithm with the tanh nonlinearity [30] – show that the secret carrier has been detected in the basis vector that is located on the bottom right in the set of patches. Similar results for DCT embedding are shown in the same figure; here we chose to watermark (4,4) coefficient. The resulting PSNR is equal to 43.8 dB. As illustrated in Fig. 9, the spatial representation of this DCT component is also clearly identified on the bottom right corner of the set of basis vectors. The normalised correlation between the original carrier and its estimate is 0.82 for pixel embedding and 0.94 for DCT embedding.

5 Estimating and removing the watermark

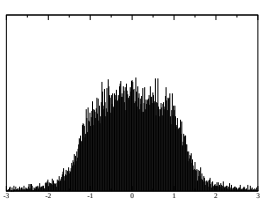
The goal of this section is to present a simple scheme to automatically identify the secret carrier and to remove the associated watermark. The estimation of the carrier relies on the decomposition in a basis of independent vectors, as described in the previous section. The removal of the watermark is performed by processing the independent component related to the watermark and applying the mixing operation to generate the attacked image.

5.1 Carrier estimation

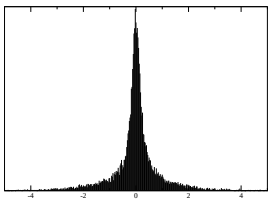
Estimation of a secret carrier is a necessary step to perform a successful attack. The carrier will be estimated as a vector that is included in the set of basis vectors obtained using an ICA algorithm. Note that the proposed method estimates the carrier without any a-priori information about the nature of the watermarked component.

5.1.1 Detection criteria

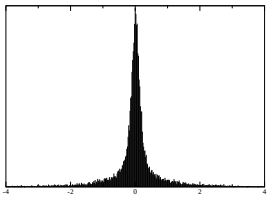
We first need to select the carrier from the set of basis vectors obtained using the ICA algorithm, e.g. the columns of the mixing matrix $\mathbf{A}$. The deflationary ICA algorithm we used works by estimating the independent vectors one by one, and the carrier is extracted as one of the last estimated vectors. This property is due to the fact that the watermark is the 'least nongaussian' of the independent components, as measured by the objective function maximised by the ICA algorithm. Consequently we have decided to choose the basis vector related to the carrier by taking into account the negentropy of its related component. The negentropy of this subgaussian component is lower than the negentropy of the components of natural images which are sparse (see section 4.3). Consequently, we have decided to



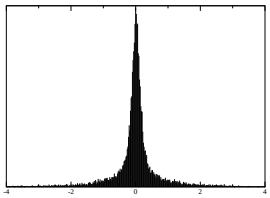
Distribution of component s_1
$J(s_1) = 0.01$



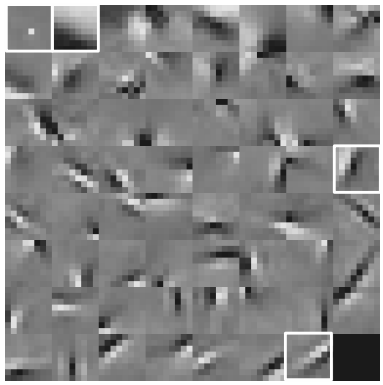
Distribution of component s_2
$J(s_2) = 0.37$



Distribution of component s_{32}
$J(s_{32}) = 0.72$

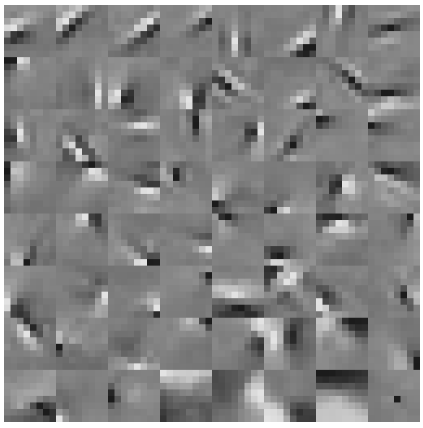


Distribution of component s_{63}
$J(s_{63}) = 0.88$

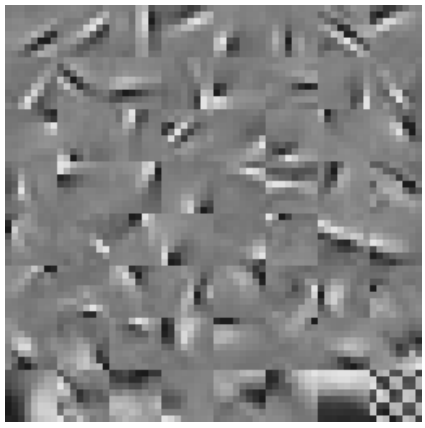


Extracted basis vectors.
From left to right and top to bottom: vectors associated with the selected components s_1, s_2, s_{32}, s_{63}

Figure 8: Top: distributions of four independent components extracted after DM had been applied to one center pixel of each block ($PSNR = 49.1dB$, 20,000 blocks). Bottom: ICA basis vectors and selected vectors. The component on the bottom right is the DC vector, which in this case was fixed as one component before running the algorithm; discarding the DC component in this or some other way is a standard technique in applying ICA to image data, since the mean grayscale value is not considered as a relevant local feature.



ICA basis vectors for
watermarked blocks in the pixel domain.



ICA basis vectors for
watermarked blocks in the DCT domain.

Figure 9: ICA outputs for 8×8 watermarked blocks. ICA is able to estimate the secret carrier, located at the bottom right block in each set of vectors.

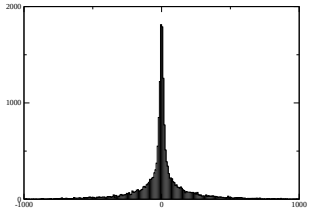
select the vector with the lowest negentropy as the estimate of the carrier.

5.1.2 Improvements using pre-filtering of the DCT components

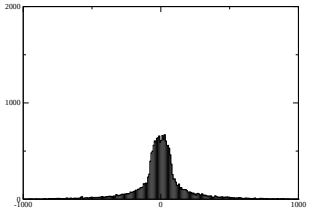
Using the criterion of the lowest negentropy, we have noticed that, in the case of the DCT embedding, this estimation of the carrier gives a low estimation error for high frequency components, but gives a poor estimate of low frequency components (see Table 1). This problem is due to the fact that the difference between the negentropy between the original and the watermarked component is much larger for high frequency components than for low frequency components (see Fig.10). For low frequency components, the energy of the watermark is small in comparison with the energy of the host component, and the watermarking process has therefore a smaller impact on negentropy values.

Coeff	(1,1)	(2,2)	(3,3)	(4,4)
MSE	0.0209	0.0216	0.0047	0.0037

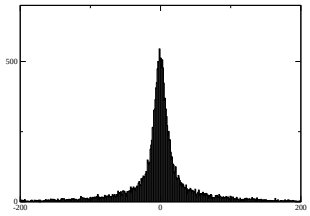
Table 1: Mean Square Error between the estimated and true carrier as a function of the watermarked DCT coefficient ($\Delta = 50$).



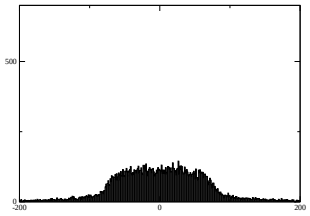
Histogram of the (1;1)
DCT component, original image.
$J = 0.42$



Histogram of the (1;1)
DCT component, watermarked image.
$J = 0.26$



Histogram of the (4,4)
DCT component, original image.
$J = 0.63$



Histogram of the (4,4)
DCT component, watermarked image.
$J = 0.0012$

Figure 10: Histograms of original DCT coefficients and watermarked coefficients ($\Delta = 50$) with their associated negentropies. Subgaussianity is more pronounced for the high frequency component (4,4) than for the low frequency one (1,1).

To improve the estimation of the carrier when the embedding is done using a carrier which has low-frequency component, we can consider a-priori information on the natural pdf of low frequency DCT coefficients. We have reduced the impact of high-amplitude DCT coefficient values by only considering coefficient values that are under a given threshold (for example Δ). This operation consists of computing the DCT coefficients of each blocks, zeroing the coefficient values that are above the threshold, and finally computing the inverse DCT transform of the block before

applying the ICA algorithm to the whole set of blocks. Consequently ICA is performed considering only the centre of the distributions of the DCT coefficients, where difference between the original and watermarked component is the largest. This preprocessing technique can be related to BSS techniques that use innovation information [31], which is the error between a given model and the observed data. In our case we assume that the tails of the distribution of the component are part of the model for digital images and the centre of the distribution is the innovation. Such a filtering method is one example of an important class of BSS schemes called Denoising Source Separation (DSS), where denoising is used to increase the performance of the estimation of the sources [32]. In practice such a pre-filtering technique enables us to achieve lower estimations errors (cf. Table 2).

Coeff	(1,1)	(2,2)	(3,3)	(4,4)
MSE	0.0009	0.0015	0.0015	0.0014

Table 2: Mean Square Error between the estimated and true carrier when the pre-filtering of DCT component is used. The DM algorithm was applied to different DCT components (as denoted by column heading in the table).

5.2 Removing the watermark

Once we have an estimate of the carrier that has been used to convey the watermark, we have access to the communication channel. It is then possible to design a specific attack to reduce the quantisation noise produced in DM as much as possible and destroy the watermark. If we denote with $\mathbf{s}_k$ the k^{th} row of matrix $\mathbf{S}$, $\mathbf{s}_k$ contains all the information that is related to the watermark vector. One straightforward way to remove the watermark is to simply reset $\mathbf{s}_k$. However, if we consider that the component $\mathbf{s}_k$ may contain image information especially for heavily textured patches, we may apply a softer function by keeping high values. This leads to the application of a shrinkage function to each sample of the vector $\mathbf{s}_k$. The used function is depicted in Fig.11; similar functions have been used for image denoising applications [33] and blind watermarking removal [34].

Our watermark identification and removal algorithm can be summarised as follows:

1. Build the matrix $\mathbf{X}$ from the set of watermarked image blocks (each block in a column of $\mathbf{X}$).
2. Using FastICA, compute $\mathbf{A}$ and $\mathbf{S}$ such as $\mathbf{X} = \mathbf{AS}$.
3. Estimate the watermark carrier $\mathbf{a}_k$ and the related source $\mathbf{s}_k$.
4. Modify $\mathbf{s}_k$ using a shrinkage function to obtain $\hat{\mathbf{s}}_k$.
5. Substitute $\mathbf{s}_k$ by $\hat{\mathbf{s}}_k$ to obtain $\hat{\mathbf{S}}$.
6. Compute the matrix $\mathbf{X}_a = \mathbf{A}\hat{\mathbf{S}}$ that represents the attacked set of blocks.

6 Results

The goal of this section is to describe the performance of the presented attack both in the estimation of the carrier and the removal of the watermark. We used four different sets of image blocks containing 900, 3,600, 10,000 and 40,000 blocks of size 8×8. These blocks were picked randomly from a panel of 16 pictures containing standard images (such as lena, mandrill, barbara, boat, peppers and also 2 synthetic images). The embedded messages were randomly generated which corresponds to the WOA setup. Four embedding schemes based on DM were attacked: DM of one pixel, DM of the (1,1) DCT coefficients (the DCT being located in (0,0)), DM of (2,2) and ST-DM with a pseudo-random gaussian

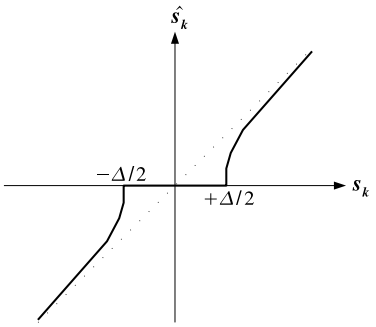


Figure 11: Shape of the used shrinkage function.

carrier.

For purposes of comparison we performed two different embeddings in the case of pixel embedding: one with the same distortion as in the case of DCT and ST-DM embeddings (PSNR=43.0dB, $\Delta = 50$), one with a smaller and unnoticeable distortion ((PSNR=52.8dB, $\Delta = 16$). For DCT and STDM embeddings (PSNR=43.0dB, $\Delta = 50$), the distortion was also unnoticeable.

The estimation of the secret carrier was done using the FastICA algorithm in the deflation mode (estimation of the independent components is done sequentially) with the tanh nonlinearity. For the estimation of the DCT coefficient, the pre-processing step presented in sec. 5.1.2 was used. Furthermore, to improve the estimation of the carrier, the DC component was forced as an independent component (this assumption is realistic and improves the estimation when the number of analysed blocks is small).

To improve the reliability of the results, the performance measures were computed from 10 different trials with different initialisations of the ICA algorithm. The reported values are means over these 10 trials.

6.1 Carrier estimation

Since knowing the carrier enables one to have a full access to the watermark subspace and to perform a successful attack, the carrier estimation step is crucial in the attacking process. The accuracy of carrier estimation is measured by calculating the normalised correlation between the estimated carrier $\hat{\mathbf{v}}$ and the original carrier $\mathbf{v}$: $c = \frac{\mathbf{v}\hat{\mathbf{v}}}{\|\mathbf{v}\|\|\hat{\mathbf{v}}\|}$.

Figure 12 presents the estimation accuracy as a function of the number of analysed blocks. For those schemes which have the same distortion, the normalised coefficient c was above 0.9 when the analysis was based on 40,000 blocks. For the different embeddings we notice that the gain in term of estimation accuracy is significant between 900 and 10,000 blocks but rather small between 10,000 and 40,000 blocks. The estimation of the carrier for ST-DM and pixel embedding is easier than estimating a carrier for DCT embedding for a same distortion. The estimation of the carrier for DM embedding in the pixel domain seems more difficult when the distortion is smaller (43dB). The estimation of a low-frequency component (DCT coefficient $(1, 1)$) is also more difficult. This issue is addressed in 6.3.

For PSNR=43dB, the first estimated carrier corresponds to the original one in every trial. For PSNR=52.8dB and pixel embedding, the estimated carrier corresponds to the original one in 80% of the trials when the number of blocks is below 3600, and in every trial for 10,000 or 40,000 blocks.

6.2 Watermark removal

The goal of this section is to evaluate the performance of the presented denoising attack as a function of the number of observed watermarked blocks. This attack also illustrates the ability of the scheme to separate the quantisation noise due to DM embedding from the original image and to reconstruct an estimate of the original image.

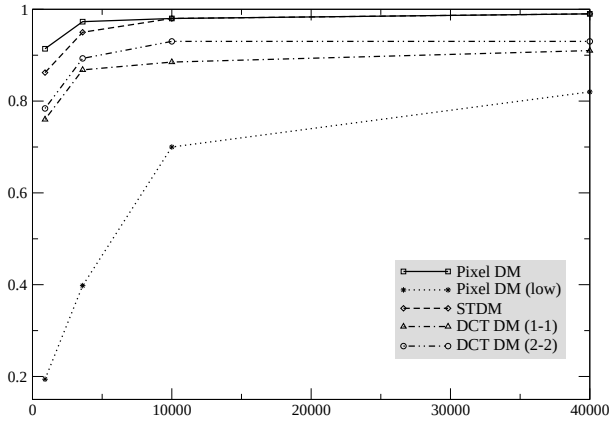


Figure 12: The normalised correlation c between the estimated and the original (secret) carrier as a function of the number of blocks. For DCT and ST-DM embedding, the PSNR is 43.0dB. For pixel embedding the PSNR is 52.8dB (low) or 43.0dB.

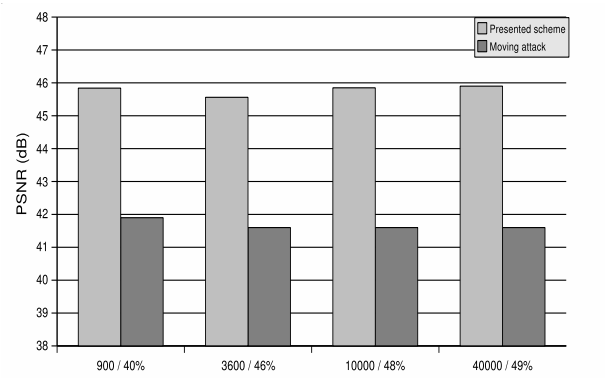
To test the denoising performance we computed the PSNR between the original and watermarked image and the PSNR between the original and attacked image. We also computed the resulting Bit Error Rate (BER) related to the attack. For purposes of comparison we calculated the PSNR after the moving attack (presented in section 2.3) that does not take into account the statistics of the host signal. In the case of the moving attack, we assume that the estimation of the secret carrier has been already done and is perfect.
The results are shown in Figure 13. Three embeddings were tested: pixel-domain DM, DCT domain DM and ST-DM. For the DCT embedding, the coefficients (2,2) and (1,1) were watermarked ((0,0) is the DC coefficient). In each embedding, the denoising attack was able to increase the PSNR between the attacked and original image in comparison with the PSNR between the watermarked and original image. This means that the power of the watermark has been decreased, and corresponds to reliable estimation of the watermark component. These results highlight the fact that the estimation error of the secret carrier $\mathbf{a}_k$, and also the independent source $\mathbf{s}_k$ (which depends on the estimation of all the basis vectors), depends on the number of observations (the number of columns of $\mathbf{X}$). In contrast, the distortion produced by the moving attack does not depend on the number of processed blocks but only on the targeted error rate.
Based on these results several conclusions can be drawn:

- The performance of the denoising attack improves as the number of observations increases; a gain of between 0.5dB (pixel embedding and same BER) and 2.5dB (ST-DM embedding, improved BER) can be achieved by using 40,000 blocks instead of 900. Note that if the PSNR does not increase with the number of observations, the BER does. For example, for DCT embedding on coefficient (1,1), the PSNR for 900 and 40,000 blocks is the same, but the BER increases 18%.

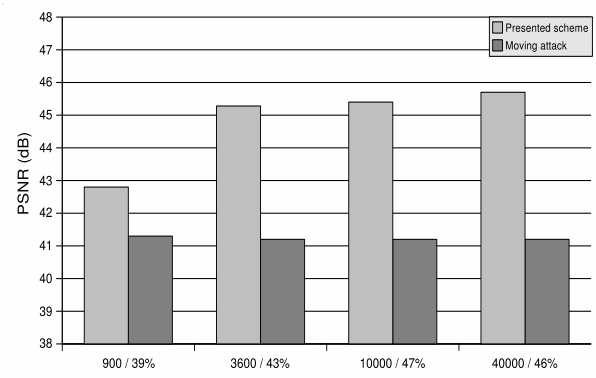
- The comparison with the moving attack highlights the superiority of our BSS-based attack for high bit error rates. For 10,000 blocks, the difference in term of introduced distortion is between 1.5dB and 4dB. Such a difference is due to the fact that the presented attacked does not only erase the watermarking signal but also estimates the components that are part of the original image.

6.3 Remarks on the security of DCT quantization schemes

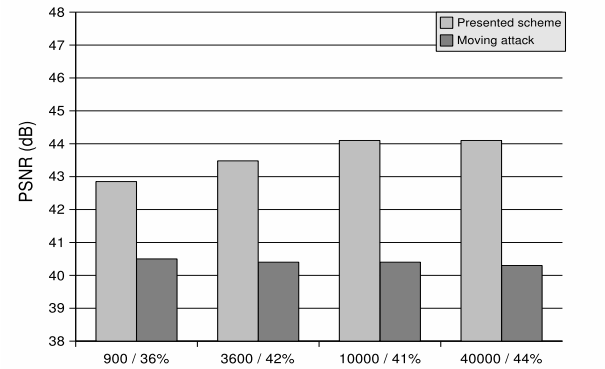
In this part of our study, the ability to erase the watermark was studied as a function of the position of the watermarked DCT coefficient. The measured dependency between the location of the DCT coefficient and the distortion between



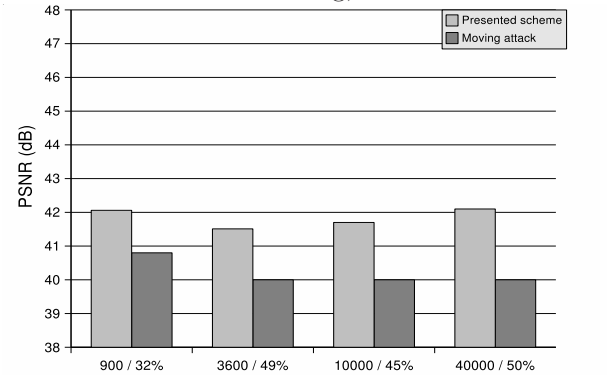
Pixel embedding, PSNR=43 dB



ST-DM embedding, PSNR=43 dB



DCT embedding coeff (2,2), PSNR=43 dB



DCT embedding coeff (1,1), PSNR=43 dB

Figure 13: Comparison of the presented ICA-based attack with the moving attack for four different DM implementations. The terms x/y under each bar denote the number of analysed blocks and the achieved bit error rate, respectively.

the attacked and original image is illustrated in Table 3. As a general rule, the performance of the proposed attack is lower for low-frequency coefficients than for high-frequency ones. For example, if the coefficient (1,0) is watermarked, the PSNR is only 34.1 dB vs. 46.5dB for an embedding in the coefficient (3,3). Such a result is due to the fact that low-frequency coefficients convey more information in natural images than high-frequency ones. Consequently it is more difficult to separate the information related to the image from that related to the watermark when the embedding is performed on low frequency features. This observation strengthens the motivation given in [35] which suggests to watermark low-frequency components for security purposes. However, it is also important to point out that a compromise between the security and the visibility has to be considered by the embedder: the perturbation of low-frequency DCT coefficients produces a larger impact in the perceived image. Moreover, note that restricting to low-frequency components greatly reduces the number of possible secret keys.

Coeff	(0,1)	(1,0)	(0,2)	(1,1)	(2,0)	(1,2)	(2,1)	(0,3)	(2,2)	(1,3)	(3,0)	(3,1)	(2,3)	(3,2)	(3,3)
PSNR	34.1	35.9	38.9	39.1	40.3	41.5	42.6	42.7	43.7	43.9	44.3	44.8	45.2	45.6	46.5

Table 3: Attack performance for different DCT frequencies (4800 blocks, PSNR=45.29 dB, $\Delta = 42$).

7 Concluding remarks

This paper addresses the security of quantisation-based schemes for non-iid host signals which are often supposed to be secure since they use a dithering vector and a secret carrier. We have show that, under a set of assumptions (natural images where one component per block has been watermarked, small blocks) it is possible to estimate the secret carrier used by the DM scheme. Because Principal Component Analysis is not a suitable tool for non-iid host signals, we have proposed to use Independent Component Analysis (ICA) to estimate the secret carrier. This is due to the fact that the DM watermarking process can be seen as the addition of a uniform noise that is independent of the host signal and can therefore be extracted using ICA. Our results suggest that the ability to separate the watermark from the original image improves as a function of the spatial frequency of the component and as a function of the number of processed blocks.

In the future, we will study whether a similar approach can be used for other popular substitutive schemes working in the DCT domain, such as the scheme proposed by [36]. In general, we expect that attacks utilising the statistics of natural images will play an important role in the security of image watermarking schemes.

8 Acknowledgements

The authors would like to thank Dr Ricardo Vigário of CIS/HUT, Finland, for his fruitful comments on Independent Component Analysis of digital images.

The work described in this paper has been supported (in part) by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT, the National French project Fabriano and by the Academy of Finland (project # 205742).

References

[1] Kalker, T.: Considerations on watermarking security. In: Proc. of MMSP, Cannes, France (2001) 201–206

[2] Cayre, F., Fontaine, C., Furon, T.: Watermarking security part I: Theory. In: Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII. Volume 5681., San Jose, USA (2005)

[3] Kalker, T.: A security risk for publicly available watermark detectors. In: Proc. Benelux Inform. Theory Symp., Veldhoven, The Netherlands (1998)

[4] Furon, T., Duhamel, P.: An asymmetric watermarking method. IEEE Trans. on Signal Processing **51** (2003) 981–995 Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[5] Comesaña, P., Pérez-Freire, L., Pérez-González, F.: Fundamentals of data hiding security and their application to spread-spectrum analysis. In: 7th Information Hiding Workshop, IH05. Lecture Notes in Computer Science, Barcelona, Spain, Springer Verlag (2005)

[6] Pérez-Freire, L., Comesaña, P., Pérez-González, F.: Information-theoretic analysis of security in side-informed data hiding. In: 7th Information Hiding Workshop, IH05. Lecture Notes in Computer Science, Barcelona, Spain, Springer Verlag (2005)

[7] Francisco J. Gonzalez-Serrano, H.Y.M.B., Murillo-Fuentes, J.J.: Independent component analysis applied to digital image watermarking. In: Proceedings of ICASSP, Salt-Lake-City, USA (2001)

[8] Bounkong, S., Toch, B., Saad, D., Lowe, D.: Ica for watermarking digital images. Journal of Machine Learning Research **4** (2004) 1471–1496

[9] Shen, M., Zhang, X., Sun, L., Beadle, P.J., Chan, F.H.Y.: A method for digital image watermarking using ica. In: Fourth International Symposium on Independent Component Analysis and Blind Signal Separation. Lecture notes in computer science, Springer Verlag, Berlin, Germany (2003)

[10] Chandramouli, R.: Mathematical approach to steganalysis. In: Proc. SPIE Vol. 4675, p. 14-25, Security and Watermarking of Multimedia Contents IV, Edward J. Delp; Ping W. Wong; Eds. (2002) 14–25

[11] Marvel, L.M., Boncelet, CG, Jr, Retter, C.T.: Spread spectrum image steganography. IEEE Tr. Im. Proc. **8** (1999) 1075–1083

[12] Lu, C.S., Liao, H.Y., Kutter, M.: Denoising and copy attacks resilient watermarking by exploiting prior knowledge at detector. IEEE Tr. Im. Proc. **11** (2002) 280–292

[13] Hyvärinen, A.: Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. Neural Computation **11** (1999) 1739–1768

[14] Chen, B., Wornell, G.W.: Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory **47** (2001) 1423–1443

[15] Shterev, I., Lagendijk, R., Heusdens, R.: Statistical amplitude scale estimation for quantization-based watermarking. In: Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VI, San Jose, USA (2004)

[16] Balado, F., Whelan, K., Silvestre, G., Hurley, N.: Joint iterative decoding and estimation for side-informed data hiding. IEEE Trans. on Signal Processing **53** (2005) 4006–4019

[17] Eggers, J.J., Buml, R., Tzschoppe, R., Girod, B.: Scalar costa scheme for information embedding. IEEE Trans. on Signal Processing **51** (2003) 1003–1019

[18] Chen, B., Wornell, G.W.: Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. In: IEEE Transaction on information theory, Vol. 47, N. 4. (2001) 1423–1443

[19] Vila-Forcén, J.E., Voloshynovskiy, S., Koval, O.J., Pérez-González, F., Pun, T.: Practical data-hiding: Additive attacks performance analysis. In: IWDW. (2005) 244–259

[20] Cachin, C.: An information-theoretic model for steganography. Information Computation **192** (2004) 41–56

[21] Cayre, F., Fontaine, C., Furon, T.: Watermarking security part II: Practice. In: Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII. Volume 5681., San Jose, USA (2005)

[22] Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)

[23] Doërr, G.J., Dugelay, J.L.: Danger of low-dimensional watermarking subspaces. In: ICASSP 2004, 29th IEEE International Conference on Acoustics, Speech, and Signal Processing, May 17-21, 2004, Montreal, Canada. (2004)

[24] van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society of London B **265** (1998) 359–366

[25] van Hateren, J.H., Ruderman, D.L.: Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. Proc. Royal Society **B** (1998) 2315–2320

[26] J.Ferreira, A., Figueiredo, M.: Class-adapted image compresion using independent component analysis. In: Proc. ICIP, Barcelona (2003)

[27] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons (1991)

[28] Proakis, J., Salehi, M.: Communication Systems engeneering. Prentice Hall International Edition (1994)

[29] Himberg, J., Hyvärinen, A.: Icasso: software for investigating the reliability of ica estimates by clustering and visualization. In: Proc. 2003 IEEE Workshop on Neural Networks for Signal Processing (NNSP2003), Toulouse, France (2003) 259–268

[30] Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks **10** (1999) 626–634

[31] Hyvärinen, A.: Independent component analysis for time-dependent stochastic processes. In: nt. Conf. on Artificial Neural Networks (ICANN'98), Skövde, Sweden (1998) 541–546

[32] Särelä, J., Valpola, H.: Denoising source separation. Journal of Machine Learning Research **6** (2005) 233–272

[33] Hyvärinen, A., Hoyer, P.O., Oja, E.: Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation. In: Proc. of Advances in Neural Information Processing Systems 11 (NIPS*98). (1999) 473–479

[34] Voloshynovskiy, S., Pereira, S., Herrigel, A., Baumgärtner, N., Pun, T.: Generalized watermark attack based on watermark estimation and perceptual remodulation. In Wah Wong, P., Delp, E.J., eds.: EI'2000: Security and Watermarking of Multimedia Content II. Volume 3971 of SPIE Proceedings., San Jose, California USA (2000) 358–370

[35] Cox, I., Killian, J., Leighton, T., Shamoon, T.: Secure spread sprectrum watermarking for images, audio and video. In: Int. Conf. on Image Processing (ICIP). Volume 3., IEEE (1996) 243–246

[36] Zhao, J., Koch, E.: Embedding robust labels intro images for copyright protection. In: Int. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienne, Autriche (1995)