ELSEVIER

Signal Processing 66 (1998) 357–372

SIGNAL PROCESSING

A DCT-domain system for robust image watermarking

Mauro Barni, Franco Bartolini, Vito Cappellini, Alessandro Piva*

Dipartimento di Ingegneria Elettronica, Università di Firenze, via di S. Marta, 3, 50139 Firenze, Italy

Received 3 February 1997; received in revised form 21 November 1997

Abstract

Digital watermarking has been proposed as a solution to the problem of copyright protection of multimedia data in a networked environment. It makes possible to tightly associate to a digital document a code allowing the identification of the data creator, owner, authorized consumer, and so on. In this paper a new watermarking algorithm for digital images is presented: the method, which operates in the frequency domain, embeds a pseudo-random sequence of real numbers in a selected set of DCT coefficients. After embedding, the watermark is adapted to the image by exploiting the masking characteristics of the human visual system, thus ensuring the watermark invisibility. By exploiting the statistical properties of the embedded sequence, the mark can be reliably extracted without resorting to the original uncorrupted image. Experimental results demonstrate that the watermark is robust to several signal processing techniques, including JPEG compression, low pass and median filtering, histogram equalization and stretching, dithering, addition of Gaussian noise, resizing, and multiple watermarking.

Zusammenfassung

Digitale Wasserzeichen sind als eine Lösung für das Problem des Urheberrechtsschutzes von Multimediadaten in vernetzten Umgebungen vorgeschlagen worden. Sie ermöglichen, mit einem digitalen Dokument fest einen Code zu verbinden, der die Identifizierung des Urhebers, Eigentümers, autorisierten Benutzers der Daten, usw. gestattet. In dieser Arbeit wird ein neuer Wasserzeichen-Algorithmus für Digitalbilder vorgestellt: die Methode, die im Frequenzbereich arbeitet, bettet eine pseudozufällige Folge reeller Zahlen in eine ausgewählte Menge von DCT-Koeffizienten ein. Nach der Einbettung wird das Wasserzeichen an das Bild angepaßt, indem Verdeckungseigenschaften der menschlichen Sichtwahrnehmung ausgenützt werden und damit die Unsichtbarkeit des Wasserzeichens sichergestellt wird. Unter Ausnützung der statistischen Eigenschaften der eingebetteten Folge kann das Zeichen zuverlässig extrahiert werden, ohne auf das unverfälschte Originalbild zurückzugreifen. Experimentelle Ergebnisse zeigen, daß das Wasserzeichen gegenüber mehreren Signalverarbeitungsverfahren robust ist, worunter JPEG-Kompression, Tiefpaß- und Medianfilterung, Histogrammentzerrung und -dehnung, Zusetzen von Dither, Addition von gaußschem Rauschen, Größenveränderung und mehrfache Wasserzeichen fallen.

Résumé

Le watermarking numérique a été proposé comme solution au problème de la protection des droits d'auteur pour les données multimédia dans un environnement de réseau. Il rend possible l'association étroite d'un code permettant l'identification du créateur des données, propriétaire, consommateur autorisé, etc., à un document numérique. Un

* Corresponding author. Tel.: + 39-55-4796380; fax: + 39-55-494569; e-mail: piva@cosimo.die.unifi.it.

0165-1684/98/$19.00
PII S0165-1684(98)00015-2

algorithme nouveau de watermarking d'images numériques est présenté dans cet article: la méthode, qui opère dans le domaine fréquentiel, intègre une séquence de nombre réels pseudo-aléatoire dans un ensemble sélectionné de coefficients DCT. Après intégration, le filigrane (watermark) est adapté à l'image en exploitant les caractéristiques de masquage du système visuel humain, ce qui assure l'invisibilité du watermark. L'exploitation des propriétés statistiques de la séquence intégrée permet une extraction fiable de la marque sans avoir à utiliser l'image originale. Les résultats expérimentaux mettent en évidence que le watermark est robuste vis-à-vis de plusieurs techniques de traitement telles que la compression JPEG, les filtrages passe-bas et médian, l'égalisation d'histogramme et l'étirement, le dithering, l'addition de bruit gaussien, le changement d'échelle, et le watermarking multiple.

Keywords: Digital watermarking; Copyright protection; Security; Image authentication

1. Introduction

Networked multimedia systems have recently gained more and more popularity due to the ever increasing amount of information that is stored and transmitted digitally; the expansion will continue at an even more steep rate when advanced multimedia services such as electronic commerce, interactive TV, teleworking, etc., will be widely available. A limiting factor in the development of multimedia-networked services is that authors, publishers and providers of multimedia data are reluctant to allow the distribution of their documents in a networked environment because the ease of reproducing digital data in their exact original form is likely to encourage copyright violation. As a matter of fact, the future development of networked multimedia systems is conditioned by the development of efficient methods to protect data owners against unauthorized copying and redistribution of the material put on the network. Whereas encryption systems do not completely solve the problem, because once encryption is removed there is no more control on the dissemination of data, a possible solution envisages the digital watermarking of multimedia works to allow their dissemination to be tracked. In this way, the number of permitted copies is not limited, but the possibility exists to control the path of the original work has been disseminated through.

A digital watermark is a code carrying information about the copyright owner, the creator of the work, the authorized consumer and whatever is needed to handle the property rights associated to any given piece of information. The watermark is intended to be permanently embedded into the digital data so that authorized users can easily read it. At the same time, the watermark should not modify the content of the work but slightly (it should be unperceivable or almost unperceivable by human senses), and it should be virtually impossible for unauthorized users to remove it. By means of watermarking the work is still accessible, but permanently marked. To be really effective, a watermark should be [3,4,6,8]:

Unobtrusive: It should be statistically and perceptually invisible so that data quality is not degraded and attackers are prevented from finding and deleting it.

Readily extractable: The data owner or an independent control authority should easily extract it.

Robust: It must be difficult (hopefully impossible) to be removed by an attacker trying to counterfeit the copyright of the data; if only partial knowledge of the watermark is available, attempts to remove or destroy it should produce a remarkable degradation in data quality before the watermark is lost. In particular, the watermark should be resistant to the most common signal processing techniques, to collusion and forgery attacks by multiple persons each possessing a watermarked copy of the document.

Unambiguous: Its retrieval should unambiguously identify the data owner.

Innumerable: It should be possible to generate a great number of watermarks, distinguishable from each other.

This paper is focused on image watermarking algorithms; in this special case, the requirement of robustness calls for the watermark to be resistant to the most common image processing techniques such as digital-to-analog and analog-to-digital

conversions, resampling, dithering, compression, contrast or colour enhancement, and to common geometric distortions such as rotation, translation, cropping, scaling and line dropping.

Image watermarking techniques proposed so far can be divided into two main groups: those which embed the watermark directly in the spatial domain [7,8,11,14,18,19,21] and those operating in a transformed domain, e.g. the frequency domain [2–5,9,16,17,22]. Techniques can also be distinguished according to the way the watermark is extracted from the possibly distorted version of the marked image. In some cases the watermark is recovered by comparing the (distorted) marked image to the original non-marked one, in this way an extra degree of robustness is achieved which virtually makes impossible the removal of the watermark without a significant degradation of the original data. Examples of such an approach are reported in [3–5,9,16,17,21], where several methods are proposed which are resistant to a large variety of image processing techniques and possible attacks aiming at removing the watermark or at making it unreadable. Unfortunately, for these techniques to be applied, the possibility to access the original image, e.g. by means of a network connection to a database, must be granted. This raises a twofold problem, since on one side the set-up of a watermarking system becomes more complicated, and on the other side the owners of the original images are compelled to unsecurely share their works with anyone who wants to check the existence of the watermark. Of course, methods capable of revealing the mark presence without comparing the marked and original images would be preferable. In the sequel, techniques which recover the watermark without resorting to the comparison between the marked image and the non-marked one will be referred to as blind watermarking techniques.

In this paper, a DCT domain watermarking technique is presented which is suitable for the marking of grey-level images. The need to access the non-marked image in the detection phase is eliminated, thus achieving a major improvement with respect to methods relying on the comparison between the watermarked and the original images [3–5,7,9,16–18,21], though at the expense of a slight loss of robustness. The algorithm, however, is still robust enough and the embedded mark invisible as much as needed in most practical applications, so that our proposal may represent a good starting point towards the protection of image-like data to be disseminated through an open-network environment.

As in [4] the watermark consists of a pseudo-random sequence, which is superimposed to some of the coefficients of the full-frame DCT transform. Unlike the method in [4], however, the mark is always superimposed to the same set of coefficients, thus avoiding the need of the original image to determine where the pseudo-random sequence is hidden. In this way, the recovery of the mark is more difficult given that the original DCT values are unknown. To regain some robustness a new casting technique is introduced and longer, higher energy, random sequences used. This can raise some problems from the point of view of mark visibility, which are solved by properly choosing the set of DCT values the mark is superimposed to, and by perceptually hiding it in image areas characterized by high luminance variance.

The paper is organized as follows. In Section 2 some of the most robust watermarking algorithms operating in the frequency domain are reviewed. Particular attention is given to the method proposed in [4] since our algorithm relies on some of the ideas exposed there. In Section 3 the new watermarking algorithm is described; in particular, the casting and recovery steps are analysed, the rationale underlying them is discussed, and a careful theoretical analysis of the algorithm robustness is carried out. Experimental results are illustrated in Section 4, and finally some conclusions are drawn in Section 5.

2. Embedding the watermark in the frequency domain

To completely define a watermarking technique operating in a transformed domain, three main steps must be specified: image transformation, watermark casting and watermark recovery.

With regard to image transformation the DCT is used in virtually all the techniques proposed so far, with few exceptions, like in [9], where a watermark

is embedded in the phase of the DFT, and in [1], where either the DCT, Walsh transform or the Wavelet transform is used. According to the different approaches, however, the transformation can be applied to the image as a whole, as in [4], or to its subparts (blocks), as in [1,2,5,9,16,17,22]. To cast the watermark code in the image, some coefficients in the transformed domain are selected which will be modified according to a watermarking rule. The coefficients to be modified can concern the whole image or only some blocks may be marked. In the second case hybrid techniques are obtained, in which the watermark is added in the frequency domain, but spatial information is also exploited by marking only a subset of the image blocks. Usually, the set of coefficients the mark is superimposed to belongs to the medium range of the frequency spectrum, so that a tradeoff between perceptual invisibility and robustness to compression and other common image processing techniques is obtained; there are two techniques where, in direct contrast to this fact, the watermark is placed in perceptually significant spectral components of the signal: in [9], where a watermark is embedded in the phase of the DFT which is quite robust to tampering and possesses superior noise immunity when compared to the magnitude, and in [4], where the watermark is inserted in the 1000 largest DCT coefficients, excluding the DC term.

To recover the watermark, the original image is in some algorithms [1,3–5,16,17] compared to the possibly corrupted and watermarked image to provide extra robustness against the attacks, since the watermark is retrieved comparing the original coefficients to the watermarked ones; moreover, the use of the original image permits some preprocessing to be carried out before the watermark checking; rotation angles, translation and scale factors can be estimated, and missing parts of the image can be replaced by corresponding parts of the original one, like in [4].

In [4] the watermark consists of a sequence of 1000 randomly generated real numbers having a normal distribution with zero mean and unity variance: $X = \{x_1, x_2, \ldots, x_{1000}\}$; the DCT of the whole image is computed, and the 1000 largest DCT coefficients, excluding the DC term, are selected; the watermark is added by modifying the selected DCT coefficients $T = \{t_1, t_2, \ldots, t_{1000}\}$ according to the relationship

$$t'_i = t_i + \alpha t_i x_i, \tag{1}$$

where $i = 1, 2, \ldots, 1000$ and $\alpha = 0.1$.

Given the original image I and the possibly distorted image I^*, a possibly corrupted watermark X^* is extracted essentially by reversing the embedding procedure. The n DCT components with largest magnitude are selected in the original image, and the difference between the non-marked coefficients and those of the (corrupted) marked image is computed. In this way, an estimate X^* of the mark sequence is obtained. Then the similarity between X and X^* is measured by means of the formula

$$\mathrm{sim}(X, X^*) = \frac{X \cdot X^*}{\sqrt{X^* \cdot X^*}}, \tag{2}$$

where by $X \cdot X^*$ the scalar product between vectors X and X^* is meant. Experimental results reported in [4] are very interesting: the algorithm can extract a reliable copy of the watermark from images that have been significantly degraded through several common geometric distortions and signal processing techniques: scaling by 75% of image size, JPEG compression with quality factor 5%, dithering, clipping, and the sequence of printing, photocopying, rescanning and scaling. Robustness against geometric deformation is achieved by means of the use of the original image in the detection step.

Sometimes, as in [12,16,17], the characteristics of the human visual system (HVS) are taken into account to adapt the watermark to the data being signed in order to improve the watermark invisibility and to enhance its robustness (watermarks of larger energy content can be embedded).

3. The proposed watermarking system

Like in [4], the watermark $X = \{x_1, x_2, \ldots, x_M\}$ consists of a pseudo-random sequence of length M generated with a multiplicative congruential algorithm (see [13]); each value x_i is a random real number with a normal distribution having zero mean and unity variance. The choice of a normally distributed watermark is motivated by

the robustness to the attacks performed by trying to produce an unwatermarked document by averaging multiple differently watermarked copies of it (see [3]). For watermark detection it is important that the real numbers x_i constituting different watermarks are statistically independent; such characteristic is granted by the pseudo-random nature of the sequences. Furthermore, such sequences could be easily reproduced by providing to the generating algorithm the correct seed (key) [13].

3.1. Watermark casting

In this step the $N \times N$ DCT for an $N \times N$ gray-scale image I is computed and the DCT coefficients are reordered into a zig-zag scan, such as in the JPEG compression algorithm [20]. What changes here with respect to the Cox's system is that it is now impossible for the decoder to determine the position of the coefficients with the largest magnitude, since the non-marked image is no longer available. To get around the problem, the mark is always inserted in the same set of coefficients. In particular, the coefficients from the $(L+1)$th to the $(M+L)$th are taken according to the zig-zag ordering of the DCT spectrum, where the first L coefficients are skipped to achieve the perceptual invisibility of the mark, without a loss of robustness against signal processing techniques. With regard to the embedding of the watermark, a different rule is used to get rid of the necessity of comparing the marked and non-marked data. In particular, the vector $T' = \{t'_{L+1}, t'_{L+2}, \ldots, t'_{L+M}\}$ with the marked DCT coefficients is computed according to the following rule:

$$t'_{L+i} = t_{L+i} + \alpha |t_{L+i}| x_i, \tag{3}$$

where $i = 1,2,\ldots,M$. The reason for weighting the introduced watermark with the absolute value of the transform coefficient instead of its plain value (as in Eq. (1)) will be clear from the following theoretical analysis (Section 3.3). Finally, T' is reinserted in the zig-zag scan and the inverse DCT is performed, thus obtaining the watermarked image I'.

3.2. Watermark detection

Given a possibly corrupted image I^*, the $N \times N$ DCT transform is applied; the DCT coefficients of I^* are reordered into a zig-zag scan, and the coefficients from the $(L+1)$th to the $(L+M)$th are selected to generate a vector $T^* = \{t^*_{L+1}, t^*_{L+2}, \ldots, t^*_{L+M}\}$. Being it impossible to get an estimate of the mark by subtracting the non-marked DCT coefficients from T^*, the correlation between the marked and possibly corrupted coefficients T^*, and the mark itself is taken as a measure of the mark presence. More specifically, the correlation z between the DCT coefficients marked with a codemark X and a possibly different mark Y is defined as

$$z = \frac{Y \cdot T^*}{M} = \frac{1}{M} \sum_{i=1}^{M} y_i t^*_{L+i}. \tag{4}$$

According to the application at hand, the correlation z can be used to determine whether a given mark is present or not, or to distinguish between a set of known marks. In the first case, z is simply compared to a predefined threshold T_z, whereas in the second case z is computed for each of the marks and that with the largest correlation is assumed to be the one really present in the image.

3.3. Theoretical analysis

Let us denote with I, I' and I^*, the original, the watermarked and the watermarked and possibly corrupted images, respectively. The coder selects a vector T of M DCT coefficients in which the watermark is possibly embedded producing a watermarked vector T', according to the rule in Eq. (3). In watermark detection, a vector T^* is selected, and the correlation between T^* and a generic watermark Y is computed according to Eq. (4). If we suppose that the watermarked image has not been corrupted, we have (overlooking the index shift of L)

$$t^*_i = t'_i = t_i + \alpha |t_i| x_i. \tag{5}$$

Then

$$z = \frac{1}{M} \sum_{i=1}^{M} (t_i y_i + \alpha |t_i| x_i y_i). \tag{6}$$

If the testing watermark Y matches the watermark X embedded in the image, z becomes

$$z = \frac{1}{M}\sum_{i=1}^{M} (t_i x_i + \alpha |t_i| x_i^2). \tag{7}$$

The statistical characteristics of z have been studied under the following hypothesis: both t_i's and x_i's are zero mean, independent and equally distributed random variables. According to these assumptions, the mean and variance of z have been computed:

$$\mu_z = \begin{cases} \alpha\mu_{|t|} & \text{if } X = Y, \\ 0 & \text{if } X \neq Y, \\ 0 & \text{if no mark is present,} \end{cases} \tag{8}$$

$$\sigma_z^2 = \begin{cases} \dfrac{1+2\alpha^2}{M}\sigma_t^2 + \dfrac{\alpha^2}{M}\sigma_{|t|}^2 & \text{if } X = Y, \\ \dfrac{1+\alpha^2}{M}\sigma_t^2 & \text{if } X \neq Y, \\ \dfrac{1}{M}\sigma_t^2 & \text{if no mark is present,} \end{cases} \tag{9}$$

where $\mu_{|t|} = E[|t|]$, $\sigma_t^2 = \mathrm{var}[t]$ and $\sigma_{|t|}^2 = \mathrm{var}[|t|]$. By noting that $\sigma_{|t|}^2 < \sigma_t^2$ and by assuming $\alpha^2 \ll 1$, we can write

$$\sigma_z^2 \simeq \frac{\sigma_t^2}{M}, \tag{10}$$

either in the case $X = Y$ or in the case that the mark X is not present in the image ($X \neq Y$ or no mark is present). Such as depicted in Fig. 1 two Gaussian random variables z_1 (if the watermark the detector searches for does not match the embedded mark or no mark is present in the image) and z_2 (if the searched watermark matches the embedded one) approximately having the same variance σ_z^2 and means $\mu_1 = 0$, $\mu_2 = \alpha\mu_{|t|}$ are obtained. In order to get a low-error probability, the factor $k = \mu_z/\sigma_z$, i.e. the distance between the Gaussian curves, must be large enough. Eq. (8) shows that μ_z does not depend on the random sequence length M, and that it increases with α; in addition, since in the zig-zag scan the DCT coefficients decrease in absolute value, when the number of skipped coefficients L increases

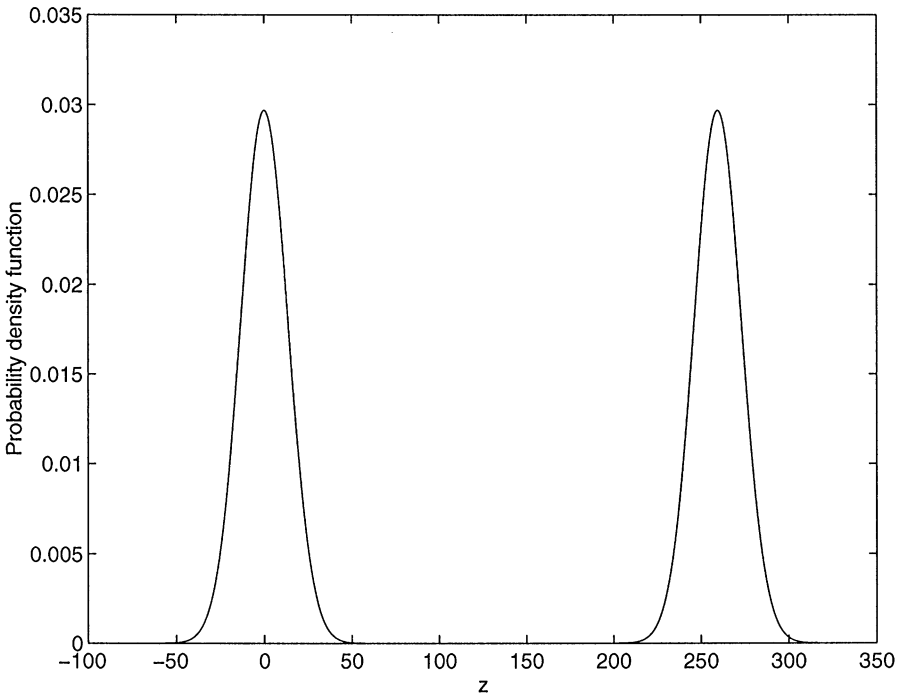

Fig. 1. Probability density functions of the random variable z, when the watermark detected does not match the embedded one (Left), and when the watermark matches the embedded one (Right).

Table 1
Evaluation of the factors $\mu_{|t|}$, σ_t and k for different values of L and M computed for the standard images 'Lenna' and 'Boat' ($\alpha = 0.2$)

Image	Sequence length	Coeff. skipped	$\mu_{\|t\|}$	σ_t	$k = \mu_z/\sigma_z$
Lenna	1000	1000	1.515	2.674	1.792
	8000	8000	0.414	1.008	3.674
	16 000	16 000	0.068	0.091	18.914
Boat	1000	1000	0.484	0.621	4.928
	8000	8000	0.137	0.177	13.791
	16 000	16 000	0.081	0.106	19.322

σ_t^2 and $\mu_{|t|}$ also decrease; nevertheless, the first factor decreases faster than the second. These considerations suggest to choose a random sequence length larger than that in [4], so that the factor k becomes sufficiently high (see Table 1). It is furthermore understood why signature (see Eq. (3)) is performed by weighting the watermark with the absolute value of the DCT coefficients instead of their plain value: in fact, the use of t would lead to $\mu_z = 0$ since $\mu_t = 0$; on the contrary, by using $|t|$ a non-zero μ_z is obtained, due to the fact that $\mu_{|t|}$ is always non-zero.

Once a threshold T_z is selected, an estimate of the error probability when no attack is present can also be given. In particular, by assuming $T_z = \mu_2/2$ and by letting $\sigma_{z_1} = \sigma_{z_2}$, we have

$$P_e = \frac{1}{\sqrt{2\pi\sigma_z^2}} \int_{T_z}^{\infty} e^{-x^2/2\sigma_z^2} dx = \frac{1}{2} \operatorname{erfc}\left(\frac{T_z}{\sqrt{2\sigma_z^2}}\right), \quad (11)$$

where $\operatorname{erfc}(x)$ is the complementary error function. To actually derive the error probability, σ_t^2 and $\mu_{|t|}$ must be estimated. This is a very difficult task, since the expected value of t_i over all possible images should be computed. Based on a test database composed by 170 grey-level images taken from a wide variety of application fields, we have found experimentally that when M and L range from 10 000 to 20 000 a good approximation is obtained by setting $\mu_{|t|} = 0.7$ and $\sigma_t^2 = 1$. By substituting these values in Eq. (11), and by assuming $\alpha = 0.1$ and $L = M = 16\,000$, an error probability approximately equal to 10^{-6} is obtained.

For the above analysis to be successfully applied to practical situations, two considerations are in order. On the basis of statistical analysis, we have assumed $\mu_{|t|} = 0.7$, which is quite a reasonable assumption; however, if an image has to be marked for which the mean absolute value of the DCT coefficients is significantly lower than 0.7, or, even worst, if some processing has been applied to the image such that the average value of $|t^*|$ is considerably reduced, an error is likely to occur when comparing z with $T_z = (\alpha/2)\mu_{|t|}$. In practical applications, then, it is better for the decoder to use a threshold T'_z which is evaluated directly on the marked image, i.e.,

$$T'_z = \frac{\alpha}{2M} \sum_{i=1}^{M} |t'_i|. \quad (12)$$

The second consideration concerns the choice of T'_z when the image has been corrupted by intentional or unintentional attacks. In such a case, the analysis carried out previously is no longer valid, since both the mean value and the variance of z may be altered because of attacks. Though the situation is not amenable to be discussed analytically, due to the large variety of possible attacks, by relying on experimental results it can be argued that when attacks are considered, σ_{z_1} remains approximately the same, whereas σ_{z_2} increases significantly. As to the average values of z_1 and z_2, we will assume that μ_{z_1} is null even in presence of attacks, and that μ_{z_2} can be reliably estimated by observing the marked, possibly corrupted, image. Therefore, by referring again to Fig. 1, we can say that because of attacks, two Gaussians are still present, but the one centred in μ_{z_2} has now a significantly larger variance. This suggests that T'_z should be set closer to zero, instead of midway between zero and $\mu_{|t|}$. Throughout the rest of the paper, we will assume

that

$$T'_z = \frac{\alpha}{3M}\sum_{i=1}^{M} |t_i^*|. \tag{13}$$

This choice of T'_z is also supported by experimental results, a part of which will be presented in Section 4.

3.4. Visual masking

As a matter of fact, the modulation law in Eq. (3) is designed to take into account the frequency masking characteristics of the HVS [15]. In fact, the perceptibility threshold of a sinusoidal grating depends on the amplitude of the iso-frequency signal to which it is superimposed. Indeed, when a DCT coefficient is modified as a consequence of watermark embedding, changes occur over the whole image, even in regions where a signal of that particular frequency is not actually present, thus, in such regions, the watermark fails to be masked. In order to enhance the invisibility of the watermark, the spatial masking characteristics of the HVS are also exploited to adapt the watermark to the image being signed: the original image I and the watermarked image I' are added pixel by pixel according to a local weighting factor $\beta_{i,j}$, thus getting a new watermarked image I'', i.e.,

$$y''_{i,j} = y_{i,j}(1-\beta_{i,j}) + \beta_{i,j}y'_{i,j} = y_{i,j} + \beta_{i,j}(y'_{i,j} - y_{i,j}). \tag{14}$$

The weighting factor $\beta_{i,j}$ takes into account the characteristics of the HVS: in regions characterized by low-noise sensitivity, where the embedding of watermarking data is easier (e.g. highly textured regions) $\beta_{i,j} \approx 1$ and $y''_{i,j} \approx y'_{i,j}$, i.e. the watermark is not diminished, whereas in regions more sensitive to changes, in which the insertion of the watermark is more disturbing, (e.g. uniform regions) $\beta_{i,j} \approx 0$ and $y''_{i,j} \approx y_{i,j}$, i.e. the watermark is embedded only to a minor extent. It is important to choose an appropriate visual characteristic of the image on the basis of which the factor $\beta_{i,j}$ changes. A simple way of choosing $\beta_{i,j}$ is here described: for each pixel $y_{i,j}$ a square block of fixed size $R \times R$ is considered (in our case $R = 9$) where the sample variance is computed; this variance is then normalized with respect to the maximum of all block variances. The factor $\beta_{i,j}$ is, thus, the normalized variance computed for pixel $y_{i,j}$. By means of Eq. (14) a twofold goal is aimed at: to increase the marking level α without compromising mark invisibility, and to make more difficult for an attacker to erase the mark, since, usually, non-uniform image regions cannot be significantly altered without degrading the image quality too much. By exploiting, in this way, visual masking, marks of higher energy can be embedded; the parameter α of Eq. (3) can be chosen in such a way that its mean value over the image, after weighting by factor $\beta_{i,j}$, is $\bar{\alpha} = 0.2$ without visible degradation of images. Threshold T'_z (Eq. (13)) has then to be estimated by using this $\bar{\alpha}$ value.

4. Experimental results

In order to test the new watermarking algorithm, 1000 watermarks were randomly generated. Some grey-scale standard images ('Boat', 'Lenna', 'Bridge', ...) were then labelled, and several common signal processing techniques and geometric distortions were applied to these images to evaluate if the detector can reveal the presence of the image owner's watermark, thus measuring the algorithm robustness to various kind of attacks. In this paper, experimental results obtained on the standard image 'Boat' in Fig. 2 (Left) are described, but similar results have been obtained with the other standard images. The original image was signed with $\bar{\alpha} = 0.2$, $M = L = 16\,000$, and block size $R = 9$ to obtain the watermarked copy shown in Fig. 2 (Right). The log of the magnitude of the response of the watermark detector to all the codemarks is shown in Fig. 3. Two possible interpretations can be given to the diagram depicted in the figure: according to the first, the response of a given mark is compared to T'_z to decide whether the mark is present or not; on the other hand, if one does not know which is the mark whose presence must be checked for, the responses to all the codemarks are compared and the largest one selected. In both the cases, there is no doubt as to the success of the decoder in making the right decision. In fact, the response to the correct mark is much stronger than the others,

Fig. 2. Original image 'Boat' (Left), and watermarked image 'Boat' with parameters $\bar{\alpha} = 0.2$, $M = L = 16\,000$, and block size $R = 9$ (Right).

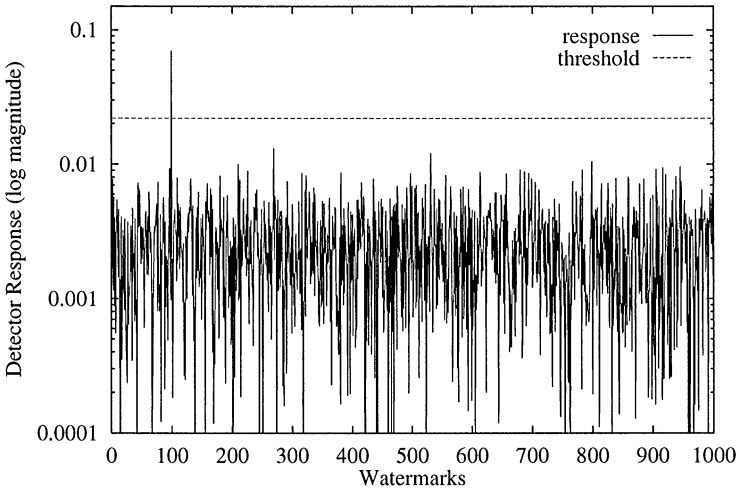

Fig. 3. The log of the magnitude of the detector response of the watermarked image in Fig. 2 (Right) to 1000 randomly generated watermarks. Only watermark number 100 matches that embedded.

thus suggesting the possibility of achieving very low false positive and false negative rates.

4.1. JPEG compression

The JPEG compression algorithm is one of the most important attacks the watermark should be resistant to. JPEG coding with 0% smoothing and decreasing quality was applied to the signed image. Obviously, when the JPEG compressed image quality decreases, the maximum detector response also decreases, however, the watermark is well above the threshold until quality is larger than 8%, corresponding to a compression ratio equal to 34 : 1 (Fig. 4 (Right)), although the image is visibly distorted (Fig. 4 (Left)). Besides, experimental results show that the response to the right mark keeps on

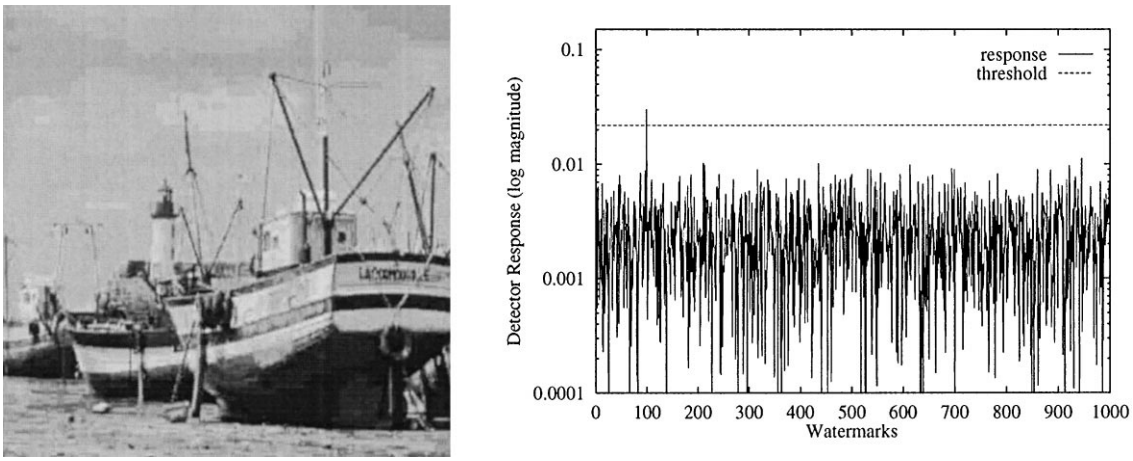

Fig. 4. JPEG compressed copy of the watermarked image 'Boat', with 4% quality and 0% smoothing (Left), and the corresponding log of the magnitude of the detector response (Right).

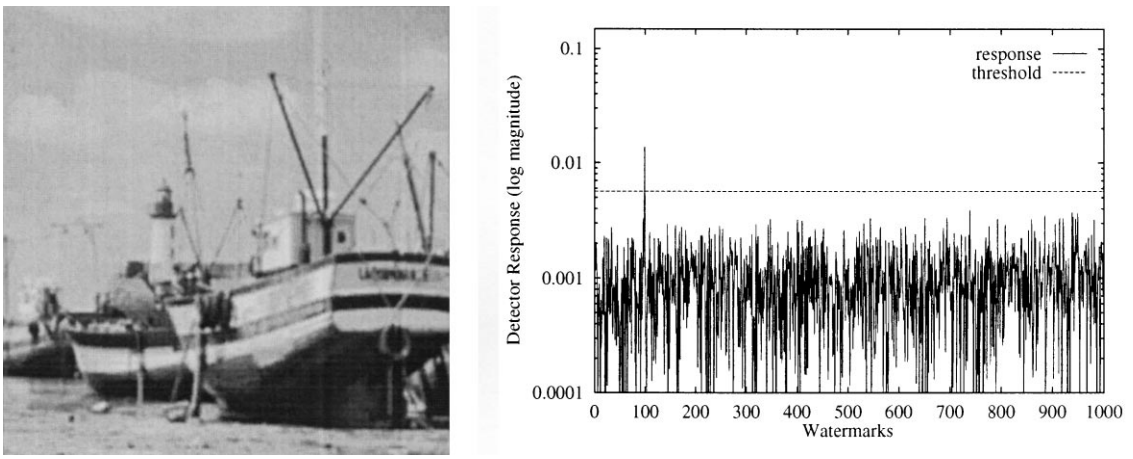

Fig. 5. Watermarked image 'Boat' low pass filtered 5×5 (Left), and the corresponding log of the magnitude of the detector response (Right).

being the largest one even if the quality parameter is set to 1%, corresponding to a compression ratio equal to 69 : 1.

4.2. Low pass filtering and median filtering

The watermarked image was filtered with a low pass filter and a median filter having increasing window size; the tests demonstrate that watermarking is robust to filters of window size 3×3 and 5×5: the responses are well above the threshold even if the images appear degraded (Figs. 5 and 6)

4.3. Histogram equalization and stretching

As shown in Figs. 7 and 8, operations on the image histogram do not degrade the watermark,

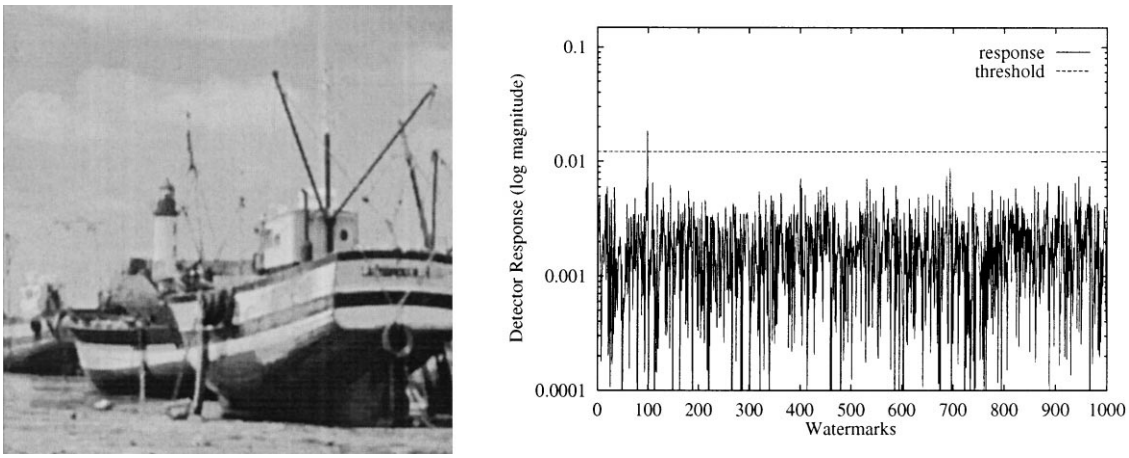

Fig. 6. Watermarked image 'Boat' median filtered 5×5 (Left), and the corresponding log of the magnitude of the detector response (Right).

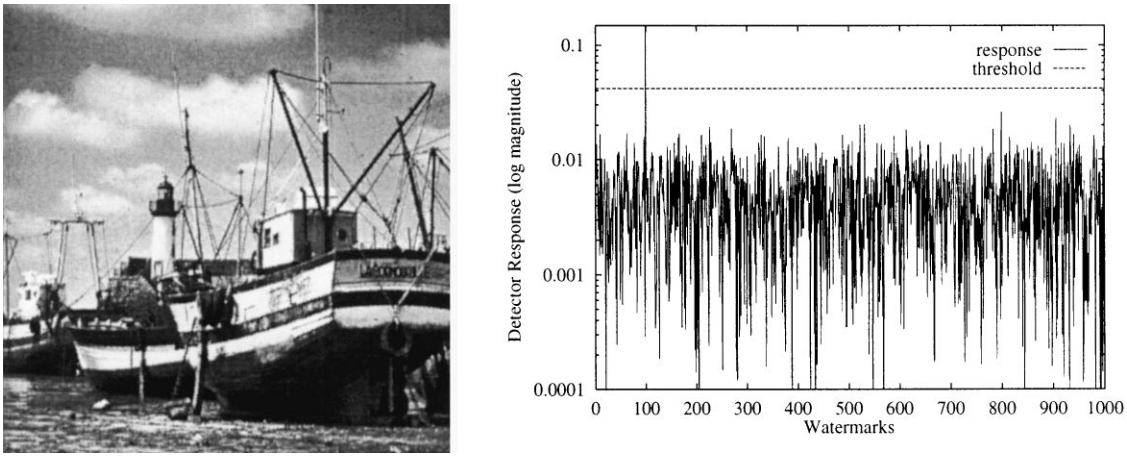

Fig. 7. Watermarked image 'Boat' after histogram equalization (Left), and the corresponding log of the magnitude of the detector response (Right).

indeed the detector response of the embedded watermark increases with respect to the response obtained on the unprocessed watermarked image. These results suggest that to enhance the algorithm performance, it is possible to preprocess the possibly corrupted image before the watermark detection by means of a histogram equalization or a histogram stretching.

4.4. Gaussian noise

As a further test, the *Boat* image was corrupted by the addition of Gaussian noise, thus obtaining the image reported in Fig. 9 (Left). A zero-mean Gaussian noise with variance $\sigma^2 = 4000$ was used. Though the image degradation is so heavy that it cannot be accepted in practical applications, the

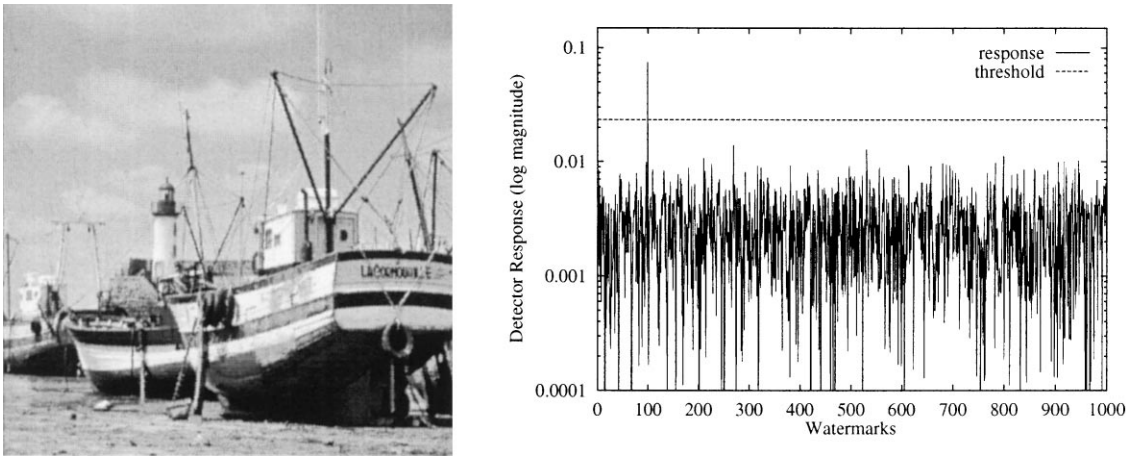

Fig. 8. Watermarked image 'Boat' after histogram stretching (Left), and the corresponding log of the magnitude of the detector response (Right).

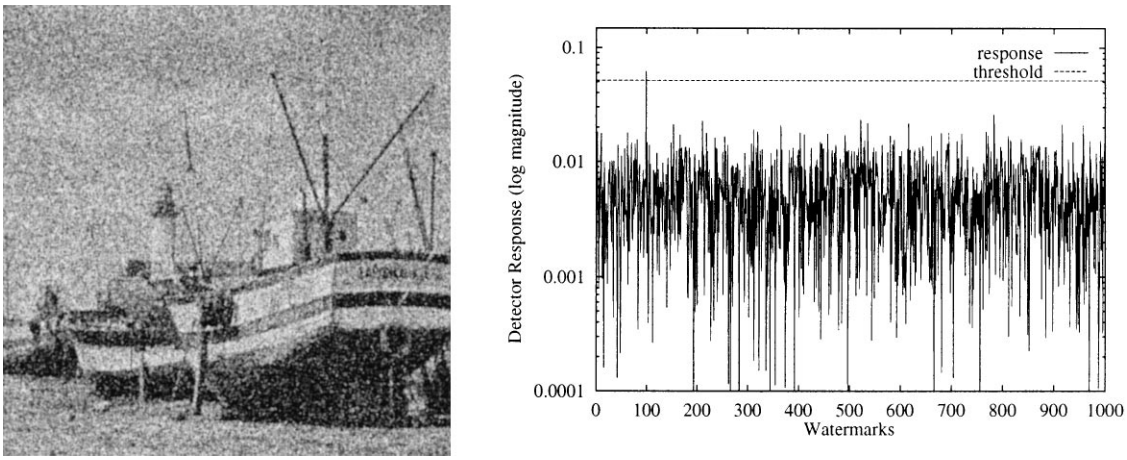

Fig. 9. Watermarked image 'Boat' with Gaussian noise having variance $\sigma^2 = 4000$ (Left), and the corresponding log of the magnitude of the detector response (Right).

mark is still easily recovered as shown in Fig. 9 (Right). Indeed, tests showed the decoder is able to recover the mark in presence of a noise with variance σ^2 up to 25 000.

4.5. Dithering

Fig. 10 (Left) shows a dithered version of the *Boat* image. Once again, the output of the decoder is satisfactory, since the detector response is well above the threshold (see Fig. 10 (Right)), thus permitting to unambiguously identify the mark present in the image. Note that the high resistance of the watermark to dithering suggests the system is also robust against all digital-to-analog conversions based on such techniques.

4.6. Geometric distortions: resizing

Virtually, all practical applications call for the watermark to be immune to geometric manipulations such as cropping and resizing. With regard to

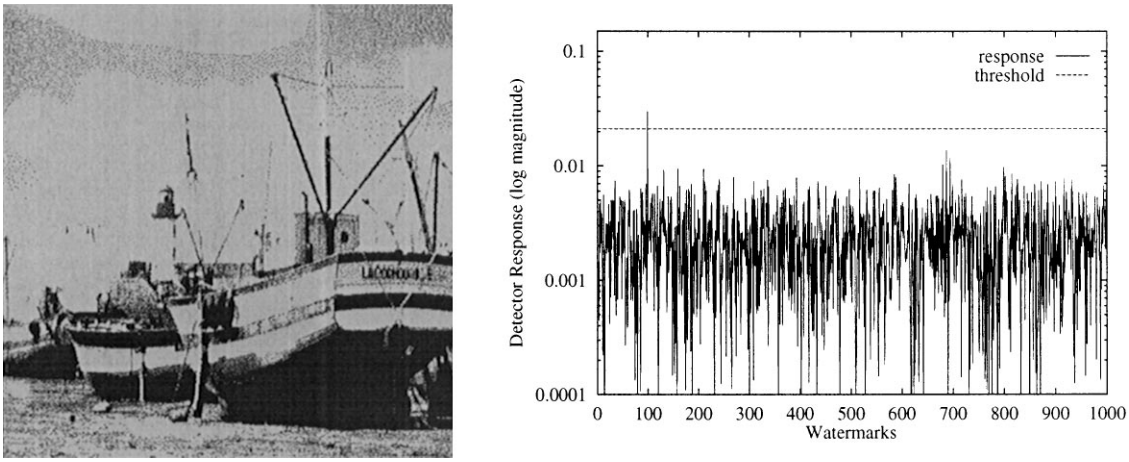

Fig. 10. Watermarked image 'Boat' after dithering (Left), and the corresponding log of the magnitude of the detector response (Right).

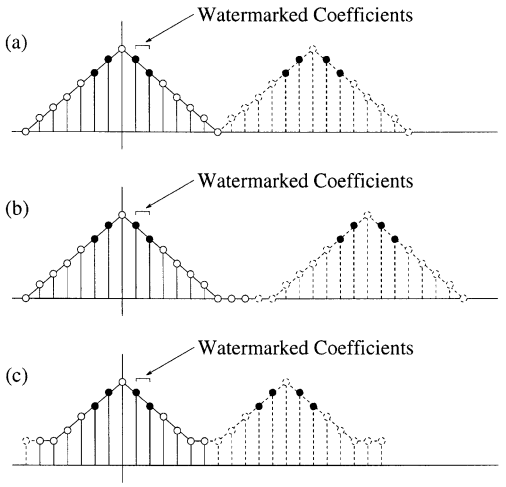

Fig. 11. Example of the effects of image resizing on DCT coefficients. The DCT spectrum of the uncorrupted watermarked image (a) is shown, as well as that of a magnified (b) and a shrunk (c) copies.

resizing the new algorithm described throughout the paper turns out to have an excellent behaviour. As a matter of fact, the response of the detector does not depend, or depends only slightly on the image size. To motivate the intrinsic robustness of the algorithm against resizing, let us consider this process in more detail [10]. The effect in the transformed domain of image resizing is exemplified in Fig. 11, where for sake of clarity the case of a one-dimensional signal is considered. In Fig. 11(a), the spectrum of the marked image is sketched with the marked coefficients highlighted. When the signal is magnified by means of an ideal interpolation process, the spectrum reported in Fig. 11(b) is obtained. As it can be seen the repetition period of spectrum replicas is enlarged, but, since the number of samples is increased by the same factor, the marked coefficients do not change. Conversely, when the signal is shrunk, replicas get closer thus causing some aliasing to occur. However, once again, if the shrinking factor is not too large, the portion of the spectrum the watermark is embedded in, does not change. Analogous considerations apply to the 2D case, even when a different scaling factor is applied in the horizontal and vertical directions, thus ensuring watermark robustness against both isotropic and anisotropic resizing. Very often, in practical applications, resizing is not achieved through an ideal interpolation process, however, due to its intrinsic robustness against this particular kind of geometric distorsion, the watermark turns out to be extremely resistant against all kinds of practical resizing algorithms (see Fig. 12).

4.7. Geometric distortions: cropping

In spite of the major role that resistance to cropping plaies in virtually all practical applications,

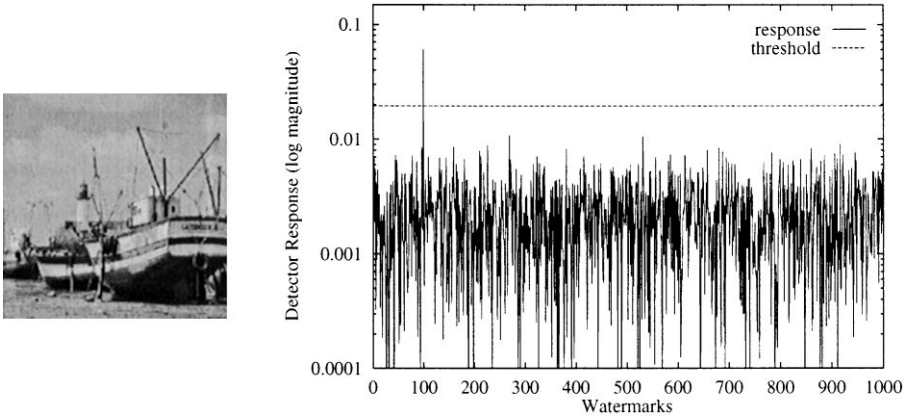

Fig. 12. Watermarked image 'Boat' after resizing from 512×512 to 256×256 (Left), and the corresponding log of the magnitude of the detector response (Right).

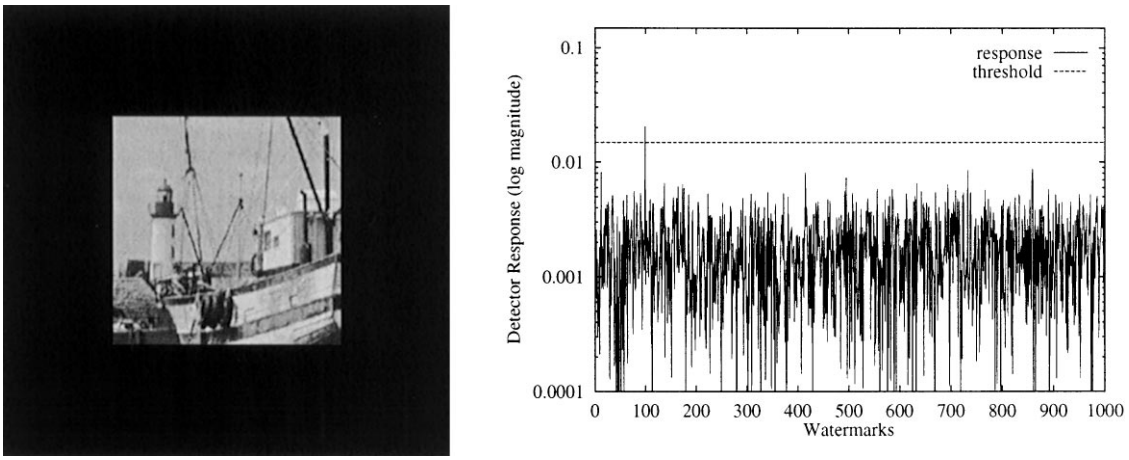

Fig. 13. Watermarked image 'Boat' after cropping (Left), and the corresponding log of the magnitude of the detector response (Right).

the proposed technique does not support the blind recovery of the watermark from a subpart of the original image. Briefly, this is mainly due to the change of the frequency sampling step which cropping results in, and to the high sensitivity of DCT transform to spatial translations. Nevertheless, experiments have been carried out proving that the information contained in a subimage is still sufficient to detect the presence of the watermark. In particular, supposed that the subimage can be replaced at exactly the same position it occupied in the original picture, the proposed system can detect the watermark if the cropped part is at least 40% of the original image (see Fig. 13).

4.8. Multiple marks and forgery attacks

Some applications require that more than one watermark is inserted in the image. For example, one could want two marks, one referring to the data creator and one indicating the authorized consumer, to be embedded in the image. Of course, all the marks embedded in the image should be

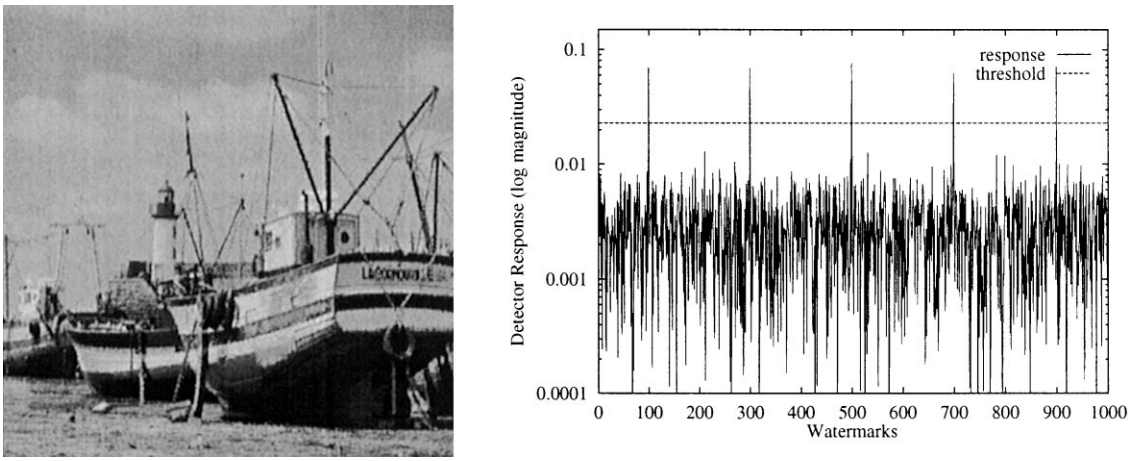

Fig. 14. Image 'Boat' with five different watermarks (Left), and the corresponding log of the magnitude of the detector response (Right).

detected by the decoder. Besides, several watermarks could be inserted aiming at making the original mark unreadable. To test our algorithm under this aspect, the original image was watermarked, then the watermarked copy was signed again with a different watermark, and so on until an image with five different watermarks has been obtained (see Fig. 14 (Left)). As shown in Fig. 14 (Right), the detector is able to retrieve all the watermarks embedded in the image.

5. Conclusions

In this paper a watermarking algorithm for digital images operating in the frequency domain is presented: a pseudo-random sequence of real numbers having normal distribution with zero mean and unity variance is embedded in a selected set of DCT coefficients. The set is produced by arranging the DCT coefficients in a zig-zag scan and by extracting the first $L + M$ coefficients; the lowest L coefficients are then skipped to preserve perceptual invisibility, and the watermark is embedded in the following M coefficients. After embedding, the watermark is adapted to the image being signed by exploiting the characteristics of noise masking of the HVS, to further ensure the watermark invisibility. Experimental results demonstrate that the watermark is robust to several signal processing techniques, including JPEG compression, low pass and median filtering, histogram equalization and stretching, dithering, Gaussian noise, resizing and multiple watermarking. Some questions arise about the maximum number of marks that can be generated satisfying the requirement of mutual independence between samples of either the same mark or different marks; however, this is not a real problem since, given that multiple watermarks can be embedded in the same image, composite marks can be used to code as much information as needed in most applications (note that even by inserting only three watermarks chosen among a set of 1000 possible marks, 10^9 different combinations are allowed).

Trying to outline the direction for future research, it seems that there is enough room for further improvement of the method. Future research will be devoted to investigate the use of DFT instead of DCT, in such a way to allow the watermarking system to resist to geometric translations. Research could also focus on colour image watermarking (currently colour images are marked by simply processing the luminance component of the image, thus ignoring the correlation between image bands), on the optimum selection of the mark length and its optimum positioning in the DCT spectrum. Also, the maximum number of marks that can be generated without compromising the algorithm robustness deserves deeper investigation.

Acknowledgements

The present work was developed with support of "Progetto Finalizzato Beni Culturali – C.N.R.". (Italian Finalized Project on Cultural Heritage – National Research Council).

References

[1] F.M. Boland, J.J.K. Ó Ruanaidh, C. Dautzenberg, Watermarking digital images for copyright protection, Proc. IEE Conf. Image Process. Appl. (July 1995) 326–331.

[2] A. Bors, I. Pitas, Image watermarking using DCT domain constraints, Proc. IEEE Internat. Conf. on Image Process. (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 231–234.

[3] I.J. Cox, J. Kilian, T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, NEC Research Institute Technical Report 95-10, 1995.

[4] I.J. Cox, J. Kilian, T. Leighton, T. Shamoon, Secure spread spectrum watermarking for images, audio and video, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 243–246.

[5] C.T. Hsu, J.L Wu, Hidden signatures in images, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 223–226.

[6] E. Koch, J. Rindfrey, J. Zhao, Copyright protection for multimedia data, Proc. Internat. Conf. on Digital Media and Electronic Publishing, Leeds, UK, 6–8 December 1994.

[7] G.C. Langelaar, J.C.A. van der Lubbe, J. Biemond, Copy protection for multimedia data based on labeling techniques, Proc. 17th Symp. Information Theory in The Benelux, Enschede, The Netherlands, May 1996.

[8] N. Nikolaidis, I. Pitas, Copyright protection of images using robust digital signatures, Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing (ICASSP-96), Vol. 4, May 1996, pp. 2168–2171.

[9] J.J.K. Ó Ruanaidh, F.M. Boland, W.J. Dowling, Phase watermarking of digital images, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 239–242.

[10] I. Pitas, Digital Image Processing Algorithms, Prentice-Hall, New York, 1993.

[11] I. Pitas, A method for signature casting on digital images, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 215–218.

[12] C. Podilchuk, W. Zeng, Perceptual watermarking of still images, Proc. The First IEEE Signal Processing Society Workshop on Multimedia Signal Processing, Princeton, NJ, June 1997.

[13] W.H. Press et al., Numerical Recipes In C: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1994.

[14] RACE Project M 1005, Access Control and COpyright Protection for Images (ACCOPI). Workpackage 8: Watermarking, Technical Report, June 1995.

[15] C.F. Stromeyer III, B. Julesz, Spatial frequency masking in vision: critical bands and spread of masking, J. Opt. Soc. Amer. 62 (10) (October 1972) 1221–1232.

[16] M.D. Swanson, B. Zhu, A.H. Tewfik, Transparent robust image watermarking, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 211–214.

[17] B. Tao, B. Dickinson, Adaptive watermarking in the DCT domain, Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing (ICASSP97), Munich, Germany, 21–24 April 1997

[18] R.G. van Schyndel, A.Z. Tirkel, C.F. Osborne, A digital watermark, Proc. IEEE Internat. Conf. on Image Processing (ICIP'94), Vol. 2, Austin, Texas, 13–16 November 1994, pp. 86–90.

[19] G. Voyatzis, I. Pitas, Applications of toral automorphisms in image watermarking, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. II, Lausanne, Switzerland, 16–19 September 1996, pp. 237–240.

[20] G.K. Wallace, The JPEG still picture compression standard, Commun. ACM 34 (4) (April 1991) 30–40.

[21] P. Wolfgang, E.J. Delp, A watermark for digital images, Proc. IEEE Internat. Conf. on Image Processing (ICIP'96), Vol. III, Lausanne, Switzerland, 16–19 September 1996, pp. 219–222.

[22] J. Zhao, E. Koch, Embedding robust labels into images for copyright protection, Proc. Internat. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, Austria, 21–25 August 1995, pp. 242–251.