A Method for Watermark Casting on Digital Images

Ioannis Pitas

Abstract—Watermark casting on digital images is an important problem since it affects many aspects of the information market. We propose a method for casting digital watermarks on images, and we analyze its effectiveness. The satisfaction of some basic demands in this area is examined, and a method for producing digital watermarks is proposed. Moreover, issues like immunity to subsampling and image-dependent watermarks are examined, and simulation results are provided for the verification of the above-mentioned topics.

Index Terms—Copyright protecting, multimedia, watermark.

I. INTRODUCTION

The following analysis is a suggested approach in solving a quite interesting and demanding problem: casting digital watermarks on digital images. By the term “digital watermark,” we mean a signal which is superimposed on the digital image, in such a way that the following hold.

1) The visual perception of the image remains unaltered, and the watermark is unnoticed.
2) We are in a position to detect a certain digital watermark by examining the alterations caused by the superposition.
3) A great number of different digital watermarks, all distinguishable from each other, can be produced.
4) Distortion or removal of the digital watermark through general image operations and manipulations should be extremely difficult and, preferably, impossible.

The satisfaction of the above-mentioned demands provides a way to superimpose an “invisible” watermark on images. This signal completely characterizes the person who applied it and, as a result, proves the origin of the image. Thus, it can be used in copyright protection for digital media.

The proposed copyright protection strategy is the following. The copyright owner deposits the original digital image to a copyright authority. He also signs each distributed copy with his own watermark (one per copyright owner). When he finds a suspected illegal copy (e.g., in a www site) he uses watermark detection to check if the image contains his own watermark. If the test is positive, he searches whether he deposited the original in a copyright authority. If yes, he has a legal proof of his ownership.

The benefits of such a method are numerous. The copyright owners of digital images have a way of protecting their products against illegal copies by proving intellectual property in conjunction with the deposition of original in a copyright authority. An owner of an image database, for example, can use watermark casting on his images and, in the case of unauthorized replications of his products, prove his copyright ownership. Television channels are presently protected merely by their logo signs on a corner of the video signal. The digital watermark can ensure them against illegal recordings and retransmissions. Of course, this topic has strong legal aspects. However, we will limit ourselves to the technical discussion only.

Watermarks are applied either in the frequency or in the spatial domain [1], [2]. The approach we follow in this paper is based on statistical detection theory, and it is applied in the spatial domain.

II. DIGITAL WATERMARK DESCRIPTION

We consider the case where an $N \times M$ gray level image $I$ has to be transformed to an $N \times M$ image $I_s$ containing a digital watermark. Our proposal is that this watermark is applied on the spatial domain, and slightly affects the intensity of some pixels of $I$ by adding or subtracting a small integer value.

A digital watermark $S$ is actually a specific binary pattern of size $N \times M$ where the number of “ones” equals the number of “zeros”

$$S = \{s_{nm}, n \in \{0, 1, \cdots, N-1\}, m \in \{0, 1, \cdots, M-1\}\}$$

(1)

where $s_{nm} \in \{0, 1\}$.

By using $S$, we can split $I$ into two subsets of equal size and alter the intensity levels of the pixels of one subset. If we consider that the original image $I$ is represented as

$$I = \{x_{nm}, n \in \{0, 1, \cdots, N-1\}, m \in \{0, 1, \cdots, M-1\}\}$$

(2)

where $x_{nm} \in \{0, 1, \cdots, L-1\}$ is the intensity level of pixel $(n, m)$ and $L$ is the total number of intensity levels, we can split it into two subsets, by using $S$, as follows:

$$A = \{x_{nm} \in I, s_{nm} = 1\} \quad B = \{x_{nm} \in I, s_{nm} = 0\}$$

(3)

Each one of subsets $A$ and $B$ contains $N \times M/2$ pixels and $I = A \cup B$. The digital watermark is superimposed by changing the elements of the subset $A$ by the positive integer factor $k$:

$$C = \{x_{nm} + k, x_{nm} \in A\}$$

(4)

The watermarked image is given by

$$I_s = C \cup B$$

(5)
In the following, we will use the symbols $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ to denote the mean values of the subsets $A$, $B$, and $C$, respectively, and the symbols $s_{\alpha}$, $s_{\beta}$, and $s_{\gamma}$ will denote the sample variances.

III. SATISFACTION OF BASIC DEMANDS

We now show that the basic demands mentioned in Section I are satisfied by the proposed method.

First Demand: Although the visual perception of an image is a completely subjective matter, the alteration of the intensity of pixels by a small fraction works properly. The quantity $k$ that is added to the pixel $x_{nm} \in A$ to produce the set $C$ in (4) is actually sufficiently small, so that the ratio $k/x_{nm}$ remains small and its visual perception is negligible according to Weber's law. Especially, if the members of the subsets $C$ and $B$ do not form a recognizable pattern, then the picture does not seem distorted in any way.

Second Demand: The satisfaction of the second demand will be explained extensively since it is critical to watermark detection. The central key is the examination of the difference of the mean values of the two image subsets $A$ and $B$. Especially, if $C$ and $B$ are as much intermixed as possible, they are expected to give better results. First, we calculate the mean values $\bar{\alpha}$ and $\bar{\beta}$, and then apply the theory of hypothesis testing [5], [6] for the determination of the difference of the two mean values

$$\bar{\mu} = \bar{\alpha} - \bar{\beta}. \quad (6)$$

Our test statistic is [6]

$$q = \frac{\bar{\mu}}{\hat{\sigma}_{\mu}} \quad (7)$$

where $\hat{\sigma}_{\mu}$ depends upon the sample variances $s_\alpha$ and $s_\beta$, and it is given by the following relation:

$$\hat{\sigma}_{\mu}^2 = \frac{s_{\alpha}^2 + s_{\beta}^2}{P} \quad (8)$$

where $P = (N \times M)/2$. In our case, the null and the alternative hypotheses, respectively, are the following.

- $H_0$: There is no watermark in the image ($\bar{\mu} = 0$).
- $H_1$: There is a watermark in the image ($\bar{\mu} = k$).

Under the null hypothesis, the test statistic $q$ follows a student distribution with zero mean and $(2P - 2)$ degrees of freedom. However, a student distribution having a degree of freedom greater than 30 can be very well approximated by the normal distribution. This is exactly the case in our problem since, for an image as small as $16 \times 16$ pixels, the value of $P$ is 128 and $2P - 2 = 254 \gg 30$. When the alternative hypothesis holds, the test statistic $q$ is distributed according to the so-called noncentral student distribution with mean (known also as noncentrality parameter) equal to $(k/\sigma_{\mu})$. Therefore, knowledge of the variance $\sigma_{\mu}$ is required in order to determine the distribution of $q$. For a large number of samples, however, the distribution of $q$ can be approximated by a normal distribution having unit variance and mean equal to $(k/\sigma_{\mu})$. Furthermore, $\hat{\sigma}_{\mu}$ can be used instead of $\sigma_{\mu}$.

The possible detections are the following.

- Type I Error: Accept the existence of a watermark, although there is none.
- Type II Error: Reject the existence of a watermark, although there is one.

Graphically, the Type I error is the shaded region of Fig. 1(a), and the Type II error appears in Fig. 1(b). It is obvious that the $t$ percentile that minimizes both errors is given by

$$t_{1-\alpha} = \frac{k}{2\hat{\sigma}_{\mu}}. \quad (9)$$

Thus, we can obtain the relationship

$$k = \left[2\hat{\sigma}_{\mu}t_{1-\alpha}\right]. \quad (10)$$

As a result, during the watermark casting (or superposition) of the image, we can give as input the degree of certainty $(1-\alpha)$ which we want to have during the later phase of the detection of the watermark.
Concluding this part, we are now in a position to give the description of the two phases, namely, “superposition” and “detection” of the watermark. In both cases, we have the prior knowledge of \((1 - \alpha)\) (certainty level) and \(S\) (watermark domain).

Watermark Casting (Superposition): We calculate \(s_\epsilon^2\) and \(s_b^2\) and use them to calculate \(\hat{\sigma}_\text{w}^2\). We calculate \(k\) from (10). However, the quantization imposed by this equation changes the level of certainty to

\[
t_1-\alpha' = \frac{k}{2\hat{\sigma}_\text{w}} = \frac{[2\hat{\sigma}_\text{w}t_1-\alpha]}{2\hat{\sigma}_\text{w}},
\]

(11)

Therefore, the true level of certainty is \(1 - \alpha'\). Moreover, as will be seen in the description of the detection phase, the following assumption is made: \(s_c = s_\epsilon\). This is not exactly correct due to clipings in the case when the terms \(x_{nm} + k\) result in numbers outside the range \(\{0, \ldots, L - 1\}\). Finally, we create the watermarked image \(I_h\) by substituting the subset \(A\) of \(I\) with the subset \(C\).

Watermark Detection: We calculate \(\bar{c}, \bar{b}, \hat{u}\), and use them to calculate \(\hat{\sigma}_\text{w}\). We create \(s_c, s_b\), and use them to calculate \(\hat{\sigma}_\text{w}\).

We create the test statistic \(q\) from (7) and test it against \(t_1-\alpha\).

If \(q < t_1-\alpha\), we give the answer “there is no watermark”; else, “there is a watermark.” During the calculation of \(\hat{\sigma}_\text{w}\), we made the above-mentioned assumption. However, the type of error induced is not actually considerable.

Third Demand: We now move to the examination of the number of “different” watermarks provided by the above-mentioned scheme. Moreover, it is now time to suggest a method of creating watermark domains \(S\). The most general method is to employ random sets \(S\) for this purpose. Such domains can be easily created by pseudonumber generators.

For the moment, let us consider the case that any possible watermark pattern is acceptable. This means that the number of watermarks that can be applied on an image of size \(N \times M = 2P\) equals the number of ways we can select \(P\) items out of \(2P\) items. As a result, the number of possible watermarks \(N_s\) is given by

\[
N_s = \binom{2P}{P} = \frac{(2P)!}{(P)!^2} \approx \frac{2^2P}{\sqrt{\pi P}} (12)
\]

by using the Stirling formula.

For example, an image of size \(32 \times 32\) \((P = 512)\) can host as many as \(4.48 \cdot 10^{306}\) different watermarks applied on it. Of course, one might argue that two very similar watermark domains cannot be distinguishable under the previously mentioned detection algorithm due to domain overlapping. Let us consider the case when two watermarks have \(X\) out of \(P\) pixels in common (partial overlap). When we try to detect one of these watermarks, while this image was watermarked by using the other, we get the following:

\[
\bar{c}' = \bar{a} + \frac{X}{P}k
\]

(13)

\[
\bar{b}' = \bar{b} + \left(1 - \frac{X}{P}\right)k
\]

(14)

\[
\bar{w}' = \bar{c}' - \bar{b}' = (\bar{a} - \bar{b}) + \left(\frac{X}{P} - 1\right)k.
\]

(15)

In order to have a wrong answer, the following inequality must hold:

\[
\frac{(\bar{a} - \bar{b}) + \left(\frac{2X}{P} - 1\right)p}{\hat{\sigma}_w^2} > t_1-\alpha
\]

(16)

where \(\hat{\sigma}_w^2\) is given by

\[
\hat{\sigma}_w^2 = \frac{s_c^2 + s_b^2}{P}.
\]

(17)

Therefore, the probability of a wrong answer is given by

\[
\text{Prob}(q + h > t_1-\alpha)
\]

(18)

where \(q\) is the test statistic we would get if we examined the clear image and \(h\) is given by

\[
h = \left(\frac{2X}{P} - 1\right)k.
\]

(19)

Since \(q\) and \(h\) are independent random variables, the distribution function of their sum is given by the convolution of their distribution functions [7]. A very good approximation of the distribution of \(q\), as we already mentioned, is the normal distribution. Let us denote by \(N(x, \mu, \sigma^2)\) the c.d.f. of a normally distributed random variable \(x\) with mean value \(\mu\) and variance \(\sigma^2\). The distribution of \(q\) is given by \(N(q, 0, 1)\).

We will now try to find the distribution of \(h\). The first watermark actually divides our total set of \(2P\) items (pixels) in two equally sized subsets, namely, subset \(S_1\) with \(P\) items having \(s_{nm} = 1\) and subset \(S_2\) with \(P\) items, as well having \(s_{nm} = 0\). The probability that the second watermark will have exactly \(X\) items belonging in \(S_1\) and \(P - X\) items belonging in \(S_2\) is given by

\[
\frac{\binom{P}{X} \binom{P}{X}}{\binom{2P}{P}} = \frac{\binom{P}{X}^2}{\binom{2P}{P}}.
\]

(20)

\(X\) follows a hypergeometric distribution with mean value \(P/2\) and variance \(P^2/(8P - 4)\). A good approximation of the distribution of \(X\) can be achieved through a normal distribution \(N(X, P/2, P/8)\) having mean value \(P/2\) and variance \(P/8\). As a result, we derive that the distribution of \(h\) is \(N(h, 0, (k^2/2P\hat{\sigma}_w^2))\), according to (19).

However, the convolution of two normal distributions is also a normal distribution, having mean value the sum of means and variance the sum of variances [6]. Therefore, it follows that the distribution of \(z = q + h\) is \(N(z, 0, 1 + (k^2/2P\hat{\sigma}_w^2))\). The probability of a wrong answer given by the (18) becomes

\[
\alpha' = 1 - N\left(t_{1-\alpha}, 0, 1 + \frac{k^2}{2P\hat{\sigma}_w^2}\right)
\]

(21)

from which we derive

\[
t_{1-\alpha'} = \sqrt{\frac{t_{1-\alpha}}{1 + \frac{\hat{\sigma}_w^2}{\delta_{1-\alpha}^2} \frac{2k^2}{P}}}
\]

(22)
It is obvious that \( t_{1-\alpha'} < t_{1-\alpha} \), which means that our degree of certainty has indeed decreased. Moreover, the superposition of a watermark, as described so far, has the nature of additive noise. As a result, the overall variance of the image is increased, which means that \( \sigma_{\Pi'} < \sigma_{\Pi} \). We, therefore, obtain an upper and a lower limit of \( t_{1-\alpha'} \)

\[
\frac{t_{1-\alpha}}{\sqrt{1 + \frac{2\epsilon_{1-\alpha}}{P}}} < t_{1-\alpha'} < t_{1-\alpha}.
\]

(23)

We will try to estimate the uncertainty imposed due to the second watermark by using these upper and lower limits. This uncertainty is given by

\[
f(t_{1-\alpha}) = N(t_{1-\alpha}, 0, 1) - N\left(\frac{t_{1-\alpha}}{\sqrt{1 + \frac{2\epsilon_{1-\alpha}}{P}}}, 0, 1\right).
\]

(24)

The function \( f(x) \) has a maximum value, which can be approximated by the equation

\[
x = \sqrt{\frac{3P}{P-6}} \approx \sqrt{3}.
\]

(25)

This result can be verified by Fig. 2, where we show the function \( f(x; P) \) for such values of \( x \) and \( P \) that are used in the cases of percentiles and images, respectively.

In Fig. 3(a), we can see a graphical representation of the normalized ambiguity imposed by a similar watermark: \( f(t_{1-\alpha})/N(t_{1-\alpha}, 0, 1) \) for \( t_{1-\alpha} = \sqrt{3} \). Moreover, a value of \( t_{1-\alpha} \) given by (25) is highly improbable to be used in practice since it provides only 95.83% degree of certainty. Instead, a value \( t_{1-\alpha} = 4 \) is more typical if we want to be very sure about the existence of watermarks. Fig. 3(b) presents the normalized ambiguity for \( t_{1-\alpha} = 4 \). It is easy to understand that the problem of similar watermarks does not really increase our initial uncertainty seriously since, for
images as small as \(32 \times 32\) \((P = 512)\), this uncertainty increases by only a factor of \(2 \cdot 10^{-5}\) \((t_{1-\alpha} = 4)\), and for typical images of size \(256 \times 256\) \((P = 32768)\), this factor is \(2.6 \cdot 10^{-7}\).

IV. IMMUNITY TO SUBSAMPLING

Another issue about watermarks is their immunity to subsampling. Here, we consider the case of the mean value subsampling, where every four pixels are substituted by their mean value. In this case, if the original image \(I\) was of size \(N \times M\), the subsampled image \(I_{\text{sub}}\) is \((N/2) \times (M/2)\) large, and the intensity levels of its pixels are given by

\[
x'_{nm} = \frac{x_{2n+1,2n} + x_{2n+1,2n+1} + x_{2n,2n+1} + x_{2n+1,2n+1}}{4}
\]

where \(x'_{nm} \in I_{\text{sub}}\) and \(x_{2n+1,2n}, x_{2n+1,2n+1}, x_{2n,2n+1}, x_{2n+1,2n+1} \in I\), for \(n \in \{0, 1, \ldots, (N/2) - 1\}\) and \(m \in \{0, 1, \ldots, (M/2) - 1\}\).

In order to apply the detection algorithm on \(I_{\text{sub}}\), we first make a subsampled version of our \(N \times M\) watermark, using the following method.

Let \(s_1, s_2, s_3, s_4\) denote the four neighboring pixels to be subsampled, and let \(u = s_1 + s_2 + s_3 + s_4\) be their sum. The sample \(s\), which will substitute \(s_1, s_2, s_3, s_4\), is 0 for \(u = 2\), 1 for \(u \in \{3, 4\}\), and 0 or 1 (with equal probabilities) when \(u = 2\).

It is obvious that our subsampled watermark pattern again contains \(P' = N \times M/8\) pixels having level 1 and \(P'\) pixels having level 0. When we examine the subsampled image with this watermark, errors are introduced. In the watermarked and unwatermarked part of the image, eight different kinds of 2 \(\times\) 2 squares, for each case, can exist. When calculating \(\overline{m}\), we have

\[
\overline{c'} = \overline{c'} + \frac{1}{8}k + \frac{43k}{8} + \frac{3k}{8} - \overline{c'} + \frac{11}{16}k
\]

\[
\overline{u'} = \overline{u'} + \frac{4k}{8} + \frac{3k}{8} - \overline{u'} + \frac{5}{16}k
\]

\[
\overline{m'} = \overline{c'} - \overline{u'} = \overline{c'} - \overline{u'} + \frac{3}{8}k.
\]

V. IMAGE-DEPENDENT WATERMARKS

Using a single watermark on all images of equal size whose copyright is owned by the same individual is very convenient in terms of implementation simplicity and speed. However, such a practice is not a safe one. A simple way to recover and subsequently destroy a watermark embedded in a set of \(K\) images is to calculate an averaged “image” as follows:

\[
\overline{x}_{mn} = \frac{1}{K} \sum_{i=1}^{K} (x_{mn}^{(i)} - m_i)
\]

where \(x_{mn}^{(i)}\) denotes the intensity of the pixel \((m, n)\) in the image \(i\) and \(m_i\) is the mean intensity value of the image \(i\).

Then, for the watermarked pixels, \(\overline{x}_{mn}\) tends to \(k\); otherwise, it tends to zero. Thus, for a sufficient \(K\), the watermark can be recovered.

In order to robustify the watermarking technique against this type of attack, we have devised a variation of the basic method that generates image dependent watermarks, i.e., a technique that, for the same watermark key, leads to different watermark patterns when applied on different images. This is achieved using the following methodology: first, we construct the watermark pattern \(S\). Then we split the pixels in \(A\) into two subsets \(A_c, A_m\) containing \(hP\) and \((1 - h)P\) pixels, respectively. \((0 \leq h \leq 1)\). The pixels of the subset \(A_c\) form the fixed part of the watermark pattern. The position of these pixels does not change when the watermark is applied on different images. On the other hand, the spatial location of the pixels that form the subset \(A_m\) changes when the watermark is applied on different images. Each pixel in \(A_m\) is translated \((\Delta n, \Delta m)\) pixels away from its original position.

The whole procedure of moving a pixel that belongs to \(A_m\) into a new position should be insensitive to image distortions. Otherwise, the detection algorithm would not be able to calculate the positions of these pixels.

VI. SIMULATION RESULTS AND CONCLUSIONS

We tested the above-mentioned algorithm on a large number of images. We applied 3000 different watermarks on each of them, asking for the minimum certainty. Due to the quantization shown in (10), for \(k = 1\), the degrees of certainty \((1 - \alpha)\) obtained by

\[
t_{1-\alpha} = \frac{k}{2\delta_m}
\]

were 84.1 and 90.5%, respectively. It has also been proven by simulations that this method is also very resistant to attacks that just put additive noise on the watermarked image.

The simulation results were very close to these values, namely, 84.1 and 90.96%. We repeated the same simulation, but with unwatermarked images this time, and obtained certainties 83.96 and 91.8%, respectively.

One important issue that was tested by simulation was the watermark resistance to JPEG compression. It was found that the method, as presented here, is resistant to compression.
ratios up to $4:1$. This is already an interesting result if we take into account that the proposed watermark casting method essentially adds high-frequency noise to the image. Ongoing research is currently performed to modify this watermark-casting scheme so that it becomes more resistant to lossy compression. One such possibility is to cast watermarks in the DCT domain, rather than in the spatial domain [1]. Preliminary results are very encouraging.

In this paper, we propose a novel method for casting digital watermarks on images. This is basically done by adding a predetermined small luminance value to randomly selected image pixels. The luminance values are small enough to be undetected by the human eye. The seed of the random pixel generator is essentially the copyright holder watermark. We also propose a scheme for watermark detection that is based on statistical detection theory criteria. Although watermark domains may overlap, we have proven that the watermarks are easily distinguishable. We have also proven that the proposed watermark scheme is rather immune to subsampling. Unlike other watermark casting schemes proposed recently, our method is based on solid mathematic background given by statistical detection theory. The theoretical study has been verified by numerous simulation experiments.

The work reported in this paper is just a first approach to the problem at hand. Several problems are still under study, notably immunity to various image alterations (low-pass filtering, clipping, line or column removal). An especially important topic under study is the modification of the algorithm so that it is resistant to JPEG compression. Preliminary results toward this goal are very encouraging.

REFERENCES


[7] Ioannis Pitas received the Diploma of Electrical Engineering in 1980 and the Ph.D. degree in electrical engineering in 1985, both from the University of Thessaloniki, Greece.

Since 1994, he has been a Professor at the Department of Informatics, University of Thessaloniki. From 1980 to 1993, he served as Scientific Assistant, Lecturer, Assistant Professor, and Associate Professor in the Department of Electrical and Computer Engineering at the same university. He served as a Visiting Research Associate at the University of Toronto, Canada, University of Erlangen-Nuernberg, Germany, Tampere University of Technology, Finland, and as Visiting Assistant Professor at the University of Toronto. His current interests are in the areas of digital image processing, multidimensional signal processing, and computer vision. He has published over 250 papers and contributed to 8 books in his area of interest. He is the coauthor of the book Nonlinear Digital Filters: Principles and Applications (New York: Kluwer, 1990) and author of Digital Image Processing Algorithms (Englewood Cliffs, NJ: Prentice-Hall, 1993).

Dr. Pitas has been a member of the European Community ESPRIT Parallel Action Committee. He has also been an invited speaker and/or member of the program committee of several scientific conferences and workshops. He was Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS and coeditor of Multidimensional Systems and Signal Processing, and he is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS. He was chair of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing (NSIP95).