BLIND IMAGE-ADAPTIVE WATERMARKING

Irene Karybali, *Kostas Berberidis*

Dept. of Computer Engineering and Informatics
and Research Academic Computer Technology Institute / R&D
University of Patras, 26500 Rio-Patras, Greece
e-mails: {karybali, berberid}@ceid.upatras.gr

ABSTRACT

In this paper a new blind image-adaptive watermarking technique is proposed. The main contributions in this work are the following. First, a new spatial mask taking into account the Human Visual System (HVS) properties, is proposed. The mask is constructed based on the local variance of the cover image prediction error sequence. Second, an improved detection scheme has been developed, which is blind, in the sense that no knowledge concerning the cover image is required. The similarity measure used in the detector is the normalized correlation between the reproduced watermark and the prediction error of the watermarked and possibly attacked image (instead of the image itself). Due to the above modifications the proposed technique exhibits superior performance as compared to the conventional HVS-based blind adaptive watermarking. This performance improvement has been justified theoretically and verified through extensive simulations. In particular, the proposed technique is robust to additive white noise, JPEG and Wavelet compression, filtering etc.

1. INTRODUCTION

The spreading of digital data (image, audio, video) via network necessitates measures for copyright protection and authentication. Digital watermarking is a possible solution to this problem. It consists in embedding useful information in the data, in such a way that it is difficult to be removed. This information, the so-called watermark, provides the identity of the data owner. A watermark is usually a key-generated pseudorandom pattern. The key should have a secure length so that cryptographic attacks can be avoided. The embedded watermark should not affect the image quality in a visible manner, but at the same time it has to be robust to attacks. Obviously, a high energy watermark is more robust than a low energy one. The watermark's energy depends on the channel (original image) capacity. The image capacity in turn is determined by the amount of information that can be inserted in an image without producing visible artifacts.

An effective way to improve the robustness of a watermark without affecting image quality, is to increase its energy in a spatially adaptive manner, using a visual mask that exploits the HVS's properties. The HVS is less sensitive to distortions around edges and in textured areas. In [1] a texture masking function based on local image properties is suggested. In [2] the watermark is embedded in the blue channel, exploiting the fact that human eye is less sensitive to this particular channel. In [3] the watermark is added to a number of low frequency DCT coefficients, adapted by the coefficients' strength. In [4], an alternative transform watermarking has been proposed taking into account spatial domain constraints. More references regarding masking techniques in spatial and transform domain can be found in [5] and [6].

In the new image-adaptive watermarking technique proposed in this paper, the involved masking function is constructed based on the prediction error variance of the cover image. The prediction error sequence matches quite well the HVS characteristics since the errors are expected to be smaller in smooth areas than in edges and textured areas. Now, concerning the detection procedure, this is performed blindly, requiring only the watermark key. Moreover, we propose as similarity measure the normalized correlation between the watermark and the prediction error sequence of the received image. Recall, that commonly the detection is done by computing the correlation between the watermark and the image available at the detector.

The above mentioned modifications result in considerably improved performance as compared to conventional masking and detection. This performance superiority, that is mainly due to the new detection scheme, has been proved theoretically, for the cases of no attack, noise attack and linear filtering attack. Extensive simulations and checks have shown that the proposed watermarking scheme is robust to many attacks, as additive white noise, JPEG and wavelet compression, filtering, dithering, thresholding etc. The basic steps of the proposed watermarking scheme and the theoretical analysis for the detection are presented in Section 2. Experimental results are provided in Section 3 and finally, in Section 4, the work is concluded and further research directions are mentioned.

2. THE PROPOSED WATERMARKING TECHNIQUE

2.1. Problem Formulation

Let x be a cover image and w the watermark, which is a pseudorandom pattern with zero mean and variance σ_w^2. Also, the watermark is of the same size (and uncorrelated)

This work was supported by the Research Academic Computer Technology Institute of Patras.

with the cover image . The watermarked image is given by

$$y = x + w \tag{1}$$

If spatial masking is used, then, denoting the involved mask by M, the watermarked image can be written as

$$y = x + M \odot w \tag{2}$$

where $\odot$ stands for pointwise multiplication. The strength of the watermark is incorporated into w. Since mask M depends only on x, it can be readily shown that the masked watermark $u \equiv M \odot w$ is also a zero mean white process and uncorrelated with the cover image x.

2.2. Visual Masking Based on Prediction Error

As already mentioned in the introduction, the proposed visual mask is computed indirectly based on the prediction properties of the cover image. Having assumed stationarity, we first compute the prediction error filter by minimizing the cost function

$$J_{LP} = E\{|x(i,j) - \widehat{x}(i,j)|^2\} \tag{3}$$

where $\widehat{x}(i,j)$ is the predicted value of $x(i,j)$ given by

$$\widehat{x}(i,j) = \boldsymbol{a}_x^T \widetilde{\boldsymbol{x}}(i,j) \tag{4}$$

where $\boldsymbol{a}_x$ is a (p^2-1)-length vector containing the linear prediction coefficients taken row-wise and vector $\widetilde{\boldsymbol{x}}(i,j)$ contains row-wise the corresponding pixels of the $p \times p$ non-causal neighborhood of $x(i,j)$ (except for the central one at (i,j)). Minimization of the above cost function with respect to $\boldsymbol{a}_x$ leads to the following system of equations

$$R_x \boldsymbol{a}_x = \boldsymbol{r}_x \tag{5}$$

where R_x is the autocorrelation matrix of x and $\boldsymbol{r}_x$ the corresponding cross-correlation vector. The solution of the above system yields the linear predictor $\boldsymbol{a}_x$. The desired prediction error sequence is derived as

$$e_x(i,j) = x(i,j) - \boldsymbol{a}_x^T \widetilde{\boldsymbol{x}}(i,j) \tag{6}$$

The prediction error sequence of the original image varies spatially in a manner which is well suited for the HVS. It has lower values for the smooth areas of the image (that are more predictable) than for the edges and the textured areas (that are less predictable). The proposed masking function is defined as

$$M(i,j) = 1 - \frac{1}{1 + \sigma_{e_x}^2(i,j)} \tag{7}$$

where $\sigma_{e_x}^2(i,j)$ denotes the local variance of the prediction error in the neighborhood of pixel (i,j). Note that the above definition is similar to that of the so-called Noise Visibility Function (NVF) suggested in [1]. The difference is that in masking function $NVF(i,j)$ the local variance of the pixel values, i.e. $\sigma_x^2(i,j)$, is used instead of $\sigma_{e_x}^2(i,j)$.

Since, in general, $\sigma_{e_x}^2(i,j) < \sigma_x^2(i,j)$, we deduce that $M(i,j) < NVF(i,j)$. This means that the watermark strength (i.e. the multiplicative factor) can be higher for the proposed mask without smoothing the local differences. It should be noted that the image edges and textured areas are better represented by the prediction error, since the “smooth” component of the image, i.e. the predicted part of the image, has been taken away.

2.3. New Blind Detection Scheme

Commonly, the blind watermark detection procedure employs a similarity measure based on the correlation between the watermark and the received image. In the proposed detection scheme, the sequence correlated with the watermark is the prediction error sequence of the received image, which is the original image after watermarking and possible attack. It turns out that, in fact, the proposed scheme is an extension of well-established techniques in communications for detecting stochastic signals in non-white noise. In the case under consideration, the signal to be detected is the watermark while the non-white noise is the attacked image. Next, we show that the modified correlation measure, yields better results as compared to the conventional one. The analysis has been performed for the cases of a) no attack, b) attack with additive white noise, and, c) linear filtering attack.

2.3.1. Detection after no attack

Let us first consider the case of detecting a non-masked watermark, i.e., the received image is $y = x + w$. The prediction error of the watermarked image, i.e., $e_y(i,j) = y(i,j) - \boldsymbol{a}_y^T \widetilde{\boldsymbol{y}}(i,j)$ is defined similarly to (6). The aim is to compare the correlation between y and w, with the correlation between e_y and w. These correlation measures are defined as

$$C_{y,w} = \frac{E[y(i,j)w(i,j)]}{\sqrt{E[y(i,j)y(i,j)]}\sqrt{E[w(i,j)w(i,j)]}} \tag{8}$$

$$C_{e_y,w} = \frac{E[e_y(i,j)w(i,j)]}{\sqrt{E[e_y(i,j)e_y(i,j)]}\sqrt{E[w(i,j)w(i,j)]}} \tag{9}$$

Starting from (8) and (9) it can be shown, after standard manipulations, that

$$C_{e_y,w} \geq C_{y,w} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + \sigma_w^2 \|\boldsymbol{a}_x\|^2 \tag{10}$$

where P_x is the power of the cover image and $\|\cdot\|$ is the Euclidean norm . That is, the output of the proposed detector is larger than the conventional one as long as the right inequality is valid, which is always the case.

Let us now consider the case $y = x + u$, where $u = M \odot w$. Then, starting again from (8) and (9) we get

$$C_{e_y,w} \geq C_{y,w} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + P_M \sigma_w^2 \|\boldsymbol{a}_x\|^2 \tag{11}$$

where P_M is the power of the mask. The condition, as expressed by the right inequality holds true in any practical case.

If in the detection procedure we use u instead of w the correlations change to

$$C_{y,u} = \frac{E[y(i,j)u(i,j)]}{\sqrt{E[y(i,j)y(i,j)]}\sqrt{E[u(i,j)u(i,j)]}} \tag{12}$$

$$C_{e_y,u} = \frac{E[e_y(i,j)u(i,j)]}{\sqrt{E[e_y(i,j)e_y(i,j)]}\sqrt{E[u(i,j)u(i,j)]}} \quad (13)$$

The resulting inequality is the same as in (11), i.e.,

$$C_{e_y,u} \geq C_{y,u} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + P_M \sigma_w^2 \|\boldsymbol{a}_x\|^2 \quad (14)$$

However, comparing (12) with (8) and (13) with (9) we obtain

$$C_{y,u} \geq C_{y,w} \quad \text{and} \quad C_{e_y,u} \geq C_{e_y,w} \Leftrightarrow P_M \geq \mu_M^2 \quad (15)$$

where μ_M is the mean value of the mask. That is, we deduce that, it is always preferable to use u in the detection procedure. To compute u we need the mask which is unknown but can be adequately approximated based on y (and not on x which is not available).

2.3.2. Detection after noise attack

Let $y = x + w + n$, where n is additive white gaussian noise with zero mean and variance σ_n^2. The resulting inequalities , corresponding to (10), (11) and (14), are as follows,

$$C_{e_y,w} \geq C_{y,w} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + (\sigma_w^2 + \sigma_n^2)\|\boldsymbol{a}_x\|^2 \quad (16)$$

$$C_{e_y,w} \geq C_{y,w} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + (P_M\sigma_w^2 + \sigma_n^2)\|\boldsymbol{a}_x\|^2 \quad (17)$$

$$C_{e_y,u} \geq C_{y,u} \Leftrightarrow P_x \geq \sigma_{e_x}^2 + (P_M\sigma_w^2 + \sigma_n^2)\|\boldsymbol{a}_x\|^2 \quad (18)$$

The comments made above (see (15)), concerning use of u instead of w in the detection, are valid here, too.

2.3.3. Detection after linear filtering attack

In this case the image is given as $z = \boldsymbol{h}^T \boldsymbol{y}(i,j)$, where $\boldsymbol{h}$ contains the coefficients of a linear filter of size $l \times l$ taken row-wise. It is assumed that, in general, $l \geq p$. The prediction error for the received image is given by $e_z(i,j) = z(i,j) - \boldsymbol{a}_z^T \widetilde{\boldsymbol{z}}(i,j)$. The correlation measures under comparison here are given by,

$$C_{z,w} = \frac{E[z(i,j)w(i,j)]}{\sqrt{E[z(i,j)z(i,j)]}\sqrt{E[w(i,j)w(i,j)]}} \quad (19)$$

$$C_{e_z,w} = \frac{E[e_z(i,j)w(i,j)]}{\sqrt{E[e_z(i,j)e_z(i,j)]}\sqrt{E[w(i,j)w(i,j)]}} \quad (20)$$

which can also be written as

$$C_{z,w} = \frac{h_0 \sigma_w^2}{\sqrt{P_z}\sqrt{\sigma_w^2}} \quad (21)$$

$$C_{e_z,w} = \frac{(h_0 - \boldsymbol{a}_z^T \widetilde{\boldsymbol{h}})\sigma_w^2}{\sqrt{\sigma_{e_z}^2}\sqrt{\sigma_w^2}} \quad (22)$$

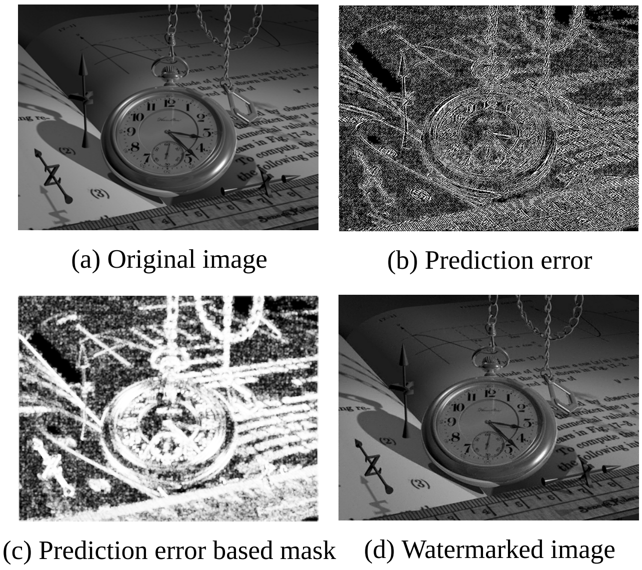
where P_z is the power of z, and $\widetilde{\boldsymbol{h}}$ is the truncated (p^2-1)-length central part of the linear filter vector $\boldsymbol{h}$ (excluding h_0). The proposed detection scheme would be better than the conventional one if $C_{e_z,w}$ were greater or equal to $C_{z,w}$, under reasonable conditions. This can be shown for a wide range of filters but the proof is skipped due to the limited space.

Given Data Case

The analysis in the above subsections is based on ideal conditions. That is, it has been assumed that each watermark is a white process, perfectly orthogonal to the other watermarks as well as completely uncorrelated with the cover image. Of course, in the given data case the above situation can only be approximately true. In such a case, all non-zero auto- and cross-correlations should be taken into account and sample averages should replace the expectation operators. The correlation coefficient between the received image (or its prediction error) and another watermark (different than the embedded one) will no longer be zero. Thus, in the given data case it is important to study the ratio between the peak value of the correlation coefficient and the maximum secondary value corresponding to another watermark. Also, thresholds should be derived so as to systematize the detection procedure. Preliminary results, in the case of noise attack, have already been derived confirming the results presented in the ideal case above. This analysis is skipped here due to limited space.

3. EXPERIMENTAL RESULTS

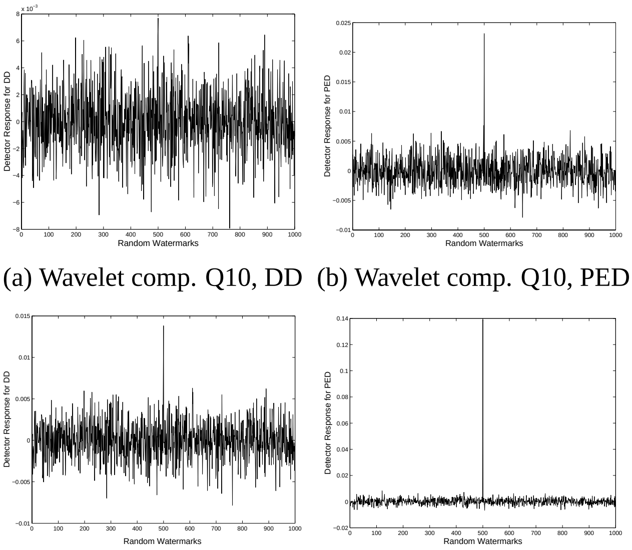
Although extensive simulations have been conducted for several images of different types, here we provide only representative results for two images, the “Clock” (375×500) and “Lenna” (256×256). The proposed visual mask was first derived for each image and after being multiplied with the corresponding watermark (of strength 5) it was added to the original image. The steps of the watermark embedding are shown in Figure 1, for the “Clock” image. Subsequently, the watermarked image was attacked. Most of the attacks were derived using Checkmark [7]. The results shown in Table 2 have been obtained after applying on the received image the conventional direct detection (DD) and the prediction error based detection scheme (PED). A bank of 1000 different watermarks was used with the correct watermark having index equal to 500. As


(a) Original image (b) Prediction error

(c) Prediction error based mask (d) Watermarked image

Figure 1. The steps of watermark embedding.

it can be easily deduced, the detection is much better for the PED and is feasible even in cases where the DD cannot find the watermark. Two characteristic examples of the detectors' responses are shown in Figure 2.



(a) Wavelet comp. Q10, DD (b) Wavelet comp. Q10, PED

(c) Wiener, DD (d) Wiener, PED

Figure 2. Direct and proposed detection.

4. CONCLUSION AND FUTURE WORK

A new visual masking function as well as a new detection scheme have been proposed. Their performance merits have been justified theoretically for the cases of no attack, noise attacks and linear filtering attacks. Extensive experiments have shown that the proposed technique performs equally well to several other type of attacks. The theoretical justification for these other attacks is an issue under investigation. Moreover, masking function based on adaptive prediction error filter will be tested and compared with other existing masks. Finally, the robustness to geometrical attacks is under consideration.

5. REFERENCES

[1] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. Third International Workshop on Information Hiding*, Dresden, Germany, Sept.29 – Oct.1, 1999, pp. 211–236.

[2] M. Kutter, F. Jordan, and F. Bossen, "Digital signature of color images using amplitude modulation," in *Proc. SPIE Electronic Imaging '97, Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 1997, pp. 518–526.

[3] I. Cox, J. Killian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[4] S. Pereira, S. Voloshynovskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *IEEE Signal Processing Magazine*, vol. 81, issue 6, pp. 1251–1260, June 2001.

[5] G.C. Langelaar, I. Setyawan, and R.L. Lagendijk, "Watermarking digital image and video data, a state-of-the-art overview," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, Sept. 2000.

[6] R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp, "Perceptual watermarks for digital images and video," in *Proc. of the IEEE*, vol. 87, no. 7, pp. 1108–1126, July 1999.

[7] S. Voloshynovskiy, S. Pereira, V. Iquise and T. Pun, "Attack modelling: Towards a second generation benchmark", *IEEE Signal Processing Magazine*, vol. 81, issue 6, pp. 1177–1214, June 2001.

Table 1. Detectors' responses (using u) for different attacks. Dash means that detection is impossible.

AWGN	***30dB***	***20dB***	***10dB***	***0dB***
Clock (DD)	0.0389	0.0389	0.0376	0.0255
Clock (PED)	0.3879	0.2486	0.1037	0.0312
Lenna (DD)	0.0344	0.0349	0.0348	0.0290
Lenna (PED)	0.4126	0.2438	0.1001	0.0387
JPEG Compr.	***Q80***	***Q50***	***Q15***	***Q10***
Clock (DD)	0.0209	0.0140	0.0089	-
Clock (PED)	0.1188	0.0495	0.0190	0.0131
Lenna (DD)	0.0214	0.0137	-	-
Lenna (PED)	0.1893	0.0819	0.0210	-
Wavelet Compr.	***Q80***	***Q50***	***Q20***	***Q10***
Clock (DD)	0.0366	0.0282	0.0132	-
Clock (PED)	0.3832	0.2933	0.0756	0.0232
Lenna (DD)	0.0326	0.0243	0.0131	-
Lenna (PED)	0.4370	0.3070	0.0646	0.0150
Colour Reduce	***Dithering***		***Thresholding***	
Clock (DD)	0.0452		0.0464	
Clock (PED)	0.0543		0.1224	
Lenna (DD)	0.0467		0.0441	
Lenna (PED)	0.0615		0.1514	
Sampledownup	***Case 1***		***Case 2***	
Clock (DD)	0.0234		0.0109	
Clock (PED)	0.1734		0.0280	
Lenna (DD)	0.0204		0.0129	
Lenna (PED)	0.2297		0.0427	
Wiener Filtering	***3x3***		***5x5***	
Clock (DD)	0.0138		0.0125	
Clock (PED)	0.1390		0.1666	
Lenna (DD)	-		-	
Lenna (PED)	0.1283		0.1449	
Trimmedmean	***3x3***		***5x5***	
Clock (DD)	-		-	
Clock (PED)	0.0294		0.0516	
Lenna (DD)	0.0130		0.0128	
Lenna (PED)	0.0546		0.0640	
Median	***2x2***	***3x3***	***4x4***	
Clock (DD)	0.0127	0.0095	-	
Clock (PED)	0.0710	0.0524	-	
Lenna (DD)	-	-	-	
Lenna (PED)	0.0684	0.0694	-	
Other Filters	***Laplacian***	***Gaussian***	***Unsharp***	
Clock (DD)	0.0273	0.0255	0.1323	
Clock (PED)	0.3602	0.3940	0.4521	
Lenna (DD)	0.0259	0.0226	0.1254	
Lenna (PED)	0.4293	0.5010	0.5127	