A NEW BLIND IMAGE-ADAPTIVE WATERMARKING SCHEME: THEORETICAL AND EXPERIMENTAL RESULTS

Irene Karybali and Kostas Berberidis

Dept. of Computer Engineering and Informatics and RACTI/R&D
University of Patras, 26500 Rio-Patras, Greece
phone: +30 2610 960425, fax: +30 2610 991909, emails: {karybali, berberid}@ceid.upatras.gr

ABSTRACT

In this paper a new blind image-adaptive watermarking scheme is proposed. The scheme employs a new spatial mask whose construction is based on the HVS's (Human Visual System) properties. Specifically, the mask is a function of the local variance of the cover image prediction error sequence. An improved blind detection scheme has also been developed, based on a proper pre-whitening process. Due to the above modifications the proposed technique exhibits superior performance as compared to conventional HVS-based blind adaptive watermarking. The proposed detector's performance improvement has been justified theoretically for the cases of no attack, noise attack and linear filtering attack, and is also verified through extensive simulations. The theoretical analysis is independent of the proposed mask and the derived expressions can be used for any watermarking technique based on spatial masking. Moreover, proper expressions have been derived that enable the computation of detection thresholds adaptable to the attacking conditions.

1. INTRODUCTION

Copyright protection and authentication of digital data via watermarking is an issue of intensive research worldwide. Digital watermarking aims at embedding useful information in the data, in such a way that it is difficult to be removed. This information, the so-called watermark, is usually a key-generated pseudorandom pattern. 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 acceptable value of watermark's energy depends on the channel (original image) capacity. The image capacity is determined by the amount of information that can be inserted in an image without producing visible artifacts.

The employment of perceptual masks, which take into account the HVS's properties, turns out to be an effective way to improve the robustness of a watermark without affecting image quality [1]. The HVS is less sensitive to distortions around edges and in textured areas. In [2] a texture masking function based on local image properties is proposed. In [3] the watermark is embedded in the blue channel, exploiting the fact that human eye is less sensitive to this particular channel. In [4] the watermark is added to a number of low frequency DCT coefficients, adapted by the coefficients' strength. In [5], 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 [6] and [7].

The image-adaptive watermarking technique proposed in this paper operates in the spatial domain and has two novel characteristics. First, a new masking function is employed whose computation is 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. Second, a new similarity measure is proposed for the detection procedure (which is performed blindly). Specifically, this measure is the normalized correlation between the masked 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. The proposed mask enables the embedding of a higher energy watermark and consequently a more robust one. The detector's performance superiority has been proved theoretically, for the cases of no attack, noise attack and linear filtering attack. The theoretical analysis is independent of the proposed mask and thus the derived expressions can be used for any watermarking technique based on spatial masking. Moreover, additional expressions have been derived (mean values and variances of the correlation measures) for each attack case that enable the computation of detection thresholds adaptable to the attacking conditions. Extensive experimental tests have shown that the proposed watermarking scheme is robust to different types of attacks, such as additive white noise and linear filtering, non-linear filtering, JPEG and wavelet compression, dithering, thresholding etc.

In Section 2 the problem is formulated and the new perceptual mask is presented. The new detection scheme is discussed in Section 3 and theoretical results concerning its performance are derived. Experimental results are provided in Section 4 and finally, in Section 5, the work is concluded and further research directions are mentioned.

2. THE PROPOSED WATERMARKING SCHEME

2.1 Problem Formulation

Let x be a cover image and w_0 the watermark, which is a pseudorandom pattern with zero mean and variance $\sigma^2_{w_0}$. The watermark is of the same size (and uncorrelated) with the cover image. If spatial masking is used, then, denoting the involved mask by M, the watermarked image can be written as

$$y = x + M \odot w_0 \quad (1)$$

This work was supported by the Research Academic Computer Technology Institute of Patras and the Heraclitus program of the Greek Ministry of Education.

where $\odot$ stands for pointwise multiplication. The strength of the watermark (i.e., its standard deviation) is incorporated into w_0. Since mask M depends only on x, it can be readily shown that the masked watermark $u_0 \equiv M \odot w_0$ is also a zero mean white process and uncorrelated with the cover image x.

2.2 Perceptual Masking Based on Prediction Error

As already mentioned in the introduction, the construction of the proposed perceptual mask is based on the prediction properties of the cover image. This mask was first suggested in [8] and is further studied in this paper. Assuming stationarity, we first compute the prediction error filter. The desired prediction error sequence is derived as

$$e_x(i,j) = x(i,j) - \widetilde{\mathbf{a}}_x^T \widetilde{\mathbf{x}}(i,j) \tag{2}$$

where $\widetilde{\mathbf{a}}_x$ is a (p^2-1)-length vector containing the linear prediction coefficients taken row-wise and vector $\widetilde{\mathbf{x}}(i,j)$ 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)).

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

$$M(i,j) = 1 - \frac{1}{1+\sigma_{e_x}^2(i,j)} \tag{3}$$

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 [2]. 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 $\sigma_{e_x}^2(i,j) < \sigma_x^2(i,j)$ (which can be easily shown), we deduce that $M(i,j) < NVF(i,j)$. Therefore using the proposed prediction error-based (PE) mask enables the insertion of a higher energy watermark and consequently of a more robust watermark.

The performance of the new mask has been tested via extensive simulations as it is commonly done with all perceptual masks that have been previously proposed in literature. Some representative results which show the superiority of the proposed masking are summarized in Table 1. Indeed, as it can be seen there, for the same Peak Signal-to-Noise Ratio the new mask allows the insertion of a higher watermark strength.

Table 1: Comparison of the watermark strength for different watermark embedding methods.

Embedding method	***Boat: Watermark strength***		
Non-adaptive	5	10	15
NVF masking	5.6	11.2	17
PE masking	7.46	14.9	22.5
PSNR (dB)	34.1532	28.1326	24.5576

3. A 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 masked watermark is the prediction error sequence of the received image, which is the marked and possibly corrupted image. Using the perceptually masked watermark in detection is an issue that has not received much attention in the relevant literature. However, we have proved that if $P_M > \mu_M^2$ [8], where P_M is the mask's power and μ_M its mean value, it is always preferable to use the masked watermark in the detection procedure. Assuming that the received image retains (approximately) the predictability properties of the cover image, we can obtain a satisfactory estimate of the mask at the receiver's end. A 2-D non-causal linear prediction error filter is employed for the computation of the prediction error sequence. In fact, it turns out that the proposed scheme is an extension of well-established techniques in communications for detecting signals in colored noise. Note that in our case, the signal to be detected is the masked watermark while the colored noise is the attacked image. Of course, prewhitening techniques have already been presented in literature for the problem at hand (see [1] and the references therein). The proposed technique is an alternative approach and yields a lower residual error power as compared to the existing ones. This is achieved at the expense of more complexity, which however can be reduced considerably by using fast algorithms for the computation of the prediction error filter as well as for the convolution of this filter with the received image.

In the following, the performance of the proposed modified correlation measure is studied and compared with the conventional similarity measure. Our analysis has been conducted for the so-called given data case. Note that usually, ideal conditions are considered in the detection procedure. That is, each watermark is assumed to be a white process completely uncorrelated with the other watermarks and the cover image. However, in a practical situation (i.e., given data case) the above assumptions are only approximately true. In fact, all the involved auto- and cross-correlation quantities should be taken into account and sample averages should replace the expectation operators. Also, the correlation measure between the received image (or its prediction error) and another watermark (different than the embedded one) will no longer be zero.

The analysis has been conducted for the general case of linear filtering plus noise attack. From the derived general expressions the special cases of a) no attack, b) additive white noise attack, and c) linear filtering attack can be easily obtained. Our analysis was based on two alternative approaches, namely, i) a deterministic approach, in which the non-zero auto- and cross-correlation terms are assumed to have upper bounded magnitudes, and ii) a statistical approach, in which the correlation quantities are considered as gaussian random variables. In the latter approach, expressions for the mean values and variances of the involved random variables are derived enabling the computation of detection thresholds adaptable to the attacking conditions. That is, the threshold determination is not based solely on experiments as it is usually done in most of the existing methods.

Three detection scenaria have been investigated, i.e., a) detecting a non-masked watermark using w_0, b) detecting a masked watermark using w_0 and, c) detecting a masked watermark, using u_0. Due to the limited space, however, we present results only for the third scenario, which is actually the most general one.

The prediction error of the watermarked image i.e., $e_y(i,j) = y(i,j) - \widetilde{\mathbf{a}}_y^T \widetilde{\mathbf{y}}(i,j)$ is defined similarly to (2). The aim is to compare the derived correlation measures when ei-

ther the image itself or the prediction error of the image is used in the detection procedure.

3.1 Deterministic Approach

The normalized correlation measure computed for two $2D$ sequences x_1 and x_2 is defined as

$$C_{x_1,x_2} = \frac{\sum\sum_{i,j=1}^{N} x_1(i,j)x_2(i,j)}{\sqrt{\sum\sum_{i,j=1}^{N} x_1(i,j)^2}\sqrt{\sum\sum_{i,j=1}^{N} x_2(i,j)^2}} \quad (4)$$

where N is the number of pixels. The correlation measures between different watermarks, or between a watermark different from the embedded one and the image, are assumed to have a magnitude that is upper bounded by a small positive scalar ε.

In the case of linear filtering plus noise attack the image is given as $z = \mathbf{h}^T\mathbf{y}(i,j) + n$, where vector $\mathbf{h}$ contains the coefficients of a linear filter of size $l \times l$ taken row-wise. It is assumed that, in general, $l \geq p$. n is additive white gaussian noise, with zero mean and variance σ_n^2. The prediction error for the received image is given by $e_z(i,j) = z(i,j) - \widetilde{\mathbf{a}}_z^T\widetilde{\mathbf{z}}(i,j)$. After standard manipulations we have that

$$if \quad |h_0|P_M\sigma_{w_0}^2 L \geq \|\widetilde{\mathbf{a}}_z\|\,\|\widetilde{\mathbf{h}}\|P_M\sigma_{w_0}^2 + c \quad then \quad C_{e_z,u_0} \geq C_{z,u_0} \quad (5)$$

where $\|\cdot\|$ is the Euclidean norm, $L = 1 - \sqrt{\sigma_{e_z}^2/P_z}$, with P_z being the power of z, and $\widetilde{\mathbf{h}}$ is the truncated (p^2-1)-length central part of the linear filter vector $\mathbf{h}$ (excluding h_0). c is a very small scalar given by

$$c = \varepsilon\mu_M\Big[\Big(\sum_i |h_i| + 1\Big)K\|\widetilde{\mathbf{a}}_z\| + \Big(|h_0| + 1 + \|\widetilde{\mathbf{h}}\|K\Big)L\Big] + \varepsilon^2\|\widetilde{\mathbf{h}}\|LK + (A+B)\|\widetilde{\mathbf{a}}_z\| \quad (6)$$

where $K = \sqrt{p^2-1}$ and

$$A = \sqrt{\sum_{i\neq 0}\Big[\varepsilon^4\Big(\sum_{i\neq j}|h_j|\Big)^2\Big]},\; B = \sqrt{\sum_{i\neq 0}\Big[2\varepsilon^2 P_M\sigma_{w_0}^2|h_i|\Big(\sum_{i\neq j}|h_j|\Big)\Big]} \quad (7)$$

Thus, if the condition in (5) holds true then the proposed correlation measure performs better as compared to the conventional one. The expressions for the special cases of linear filtering attack, noise attack, and no attack can be easily derived by a proper selection of the involved parameters. In case of linear filtering attack it turns out that inequality (5) is valid if $|h_0| \geq \|\widetilde{\mathbf{a}}_z\|\,\|\widetilde{\mathbf{h}}\|$. Such filters are those with dominant central part, as the Laplacian, the Gaussian and the Unsharp filter. That is, for instance, the watermark is not detectable after a mean filter attack. Since the analysis is independent of the proposed mask's properties, the derived result is general and valid for any type of spatial masking.

3.2 Statistical Approach

Note that in the deterministic approach above we compared the detectors' performances assuming that the watermark involved in the correlation measure is the original (desired) one. However, in the given data case the performance of any correlation detector depends also on the correlation measure between the image (whether prewhitened or not) and a watermark different than the embedded one. In the following statistical approach both cases are tackled and corresponding expressions for the mean values and variances of the correlation quantities are derived.

The original masked watermark is denoted as u_0 whereas any other masked watermark (of the suitably designed bank of watermarks) is denoted as u_1. Due to space limitations, the derivations of the provided expressions are omitted. The following definitions are used in the derived expressions: $k_0 = \sqrt{P_z\sigma_{u_0}^2}$, $k_1 = \sqrt{P_z\sigma_{u_1}^2}$, $l_0 = \sqrt{\sigma_{e_z}^2\sigma_{u_0}^2}$, $l_1 = \sqrt{\sigma_{e_z}^2\sigma_{u_1}^2}$, $r_{M^2}(i) = E[M^2(n)M^2(n+i)]$ and $\mathbf{a}_z$ is a p^2-length vector containing the linear prediction coefficients taken row-wise and a_i is a prediction error filter coefficient.

The mean values and variances of the examined correlations are given below, enabling the computation of suitable thresholds.

$$\mu_{C_{z,u_1}},\; \mu_{C_{e_z,u_1}} = 0,\;\; \mu_{C_{z,u_0}} = \frac{h_0 P_M\sigma_{w_0}^2}{k_0},\;\; \mu_{C_{e_z,u_0}} = \frac{\mathbf{a}_z^T\mathbf{h}P_M\sigma_{w_0}^2}{l_0}$$

$$\sigma_{C_{z,u_1}}^2 = \frac{1}{Nk_1}\sigma_{w_1}^2\Big[P_M\mathbf{h}^T R_x\mathbf{h} + \Big(h_0^2 r_{M^2}(0) + \sum_{i\neq 0} h_i^2 r_{M^2}(i)\Big)\sigma_{w_0}^2 + P_M\sigma_n^2\Big] \quad (8)$$

$$\sigma_{C_{e_z,u_1}}^2 = \frac{1}{Nl_1}\sigma_{w_1}^2\Big[P_M\mathbf{g}^T R_x'\mathbf{g} + C\sigma_{w_0}^2 + P_M\sigma_n^2\|\mathbf{a_z}\|^2\Big]$$

$$\sigma_{C_{z,u_0}}^2 = \frac{1}{Nk_0}\sigma_{w_0}^2\Big[P_M\mathbf{h}^T R_x\mathbf{h} + \Big(3h_0^2 r_{M^2}(0) + \sum_{i\neq 0} h_i^2 r_{M^2}(i) + \frac{1}{N}h_0^2\sum_{i\neq 0} r_{M^2}(i) - Nh_0^2P_M^2\Big)\sigma_{w_0}^2 + P_M\sigma_n^2\Big]$$

$$\sigma_{C_{e_z,u_0}}^2 = \frac{1}{Nl_0}\sigma_{w_0}^2\Big[P_M\mathbf{g}^T R_x'\mathbf{g} + \Big(D + \frac{1}{N}(\mathbf{a}_z^T\mathbf{h})^2\sum_{i\neq 0} r_{M^2}(i) - NP_M^2(\mathbf{a}_z^T\mathbf{h})^2\Big)\sigma_{w_0}^2 + P_M\sigma_n^2\|\mathbf{a_z}\|^2\Big]$$

where R_x is the cross-correlation matrix of x, $\mathbf{g}$ is the convolution between the prediction error filter and the filter used for attack (with length $2p-1$) and R_x' is the cross-correlation matrix of the corresponding dimensions. The expressions for the subcases can be derived as in the deterministic case by omitting noise ($\sigma_n^2 = 0$), or/and setting h_0 to one and h_i (for $i \neq 0$) to zero. After noise attack for example, the corresponding expression for (8) is

$$\sigma_{C_{z,u_1}}^2 = \frac{1}{Nk_1}\sigma_{w_1}^2\Big(P_M P_x + r_{M^2}(0)\sigma_{w_0}^2 + P_M\sigma_n^2\Big)$$

The expressions for C and D are not given here due to limited space, but since they depend on the filter we give the corresponding expressions for $h_0 = 1$ and $h_i = 0$ (for $i \neq 0$):

$$C = r_{M^2}(0) + \sum_{i\neq 0} a_i^2 r_{M^2}(i), \quad D = 3r_{M^2}(0) + \sum_{i\neq 0} a_i^2 r_{M^2}(i)$$

It can be seen from the above expressions that when the proposed detection is used, the variances either in presence or in absence of the embedded watermark are smaller than those of the conventional detection. The mean values in absence of watermark are zero in both cases, while the mean value in presence of watermark is higher for the proposed scheme. Therefore the new detector is superior to the conventional one. Based on the above values proper thresholds can be easily computed.

4. EXPERIMENTAL RESULTS

Although extensive simulations have been conducted for several images of different types, here, due to limited space, we provide the results for the image "Boat" (470×500). The proposed perceptual mask was first derived for this image and after being multiplied with the corresponding watermark

(of strength 6.3) it was added to the original image. The strength was selected so as the watermark to be practically invisible (the Watson metric from Checkmark [9] was used). The steps of the watermark embedding procedure are shown in Figure 1, for the "Boat" image. As it was also shown in Table 1, the proposed masking enables the insertion of a higher energy (and consequently a more robust) watermark as compared to other techniques.

Subsequently, the watermarked image was attacked in different ways. The results shown in Table 2 have been obtained after applying on the received image the conventional direct detection (DD) (correlation between the attacked image and the masked watermark) and the prediction error based detection scheme (PED) (correlation between the prediction error of the attacked image and the masked watermark). A bank of 1000 different watermarks was used with the correct watermark having index equal to 500. As it can be easily deduced, the detection is much better for the PED and is feasible even in cases where the DD is unable to detect the watermark.

(a) Original image

(b) Prediction error

(c) Prediction error based mask

(d) Watermarked image

Figure 1: The steps of watermark embedding.

5. CONCLUSION AND FUTURE WORK

A new perceptual 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 attack and linear filtering attack. 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. A theoretical proof for our mask's effectiveness is also under investigation and moreover, a masking function based on adaptive prediction error filter will be tested. Finally, the robustness to geometrical attacks is also under consideration.

REFERENCES

[1] I.J. Cox, M.L. Miller and J.A. Bloom, "Digital Waterarking", Morgan Kaufman Publishers, 2002.

[2] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner and T. Pun, "A stochastic approach to content adaptive digital watermarking," in *Proc. 3rd International Workshop on Information Hiding*, Dresden, Germany, Sept. 1999, pp. 211–236.

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

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

[5] S. Pereira, S. Voloshynovskiy and T. Pun, "Optimal transform domain watermarking for multimedia," *Elsevier Signal Processing*, vol. 81, pp. 1251–1260, Jun. 2001.

[6] G. C. Langelaar, I. Setywan and R. L. Lagendijk, "Watermarking digital image and video data, a state-of-the-art overview," *IEEE Signal Processing Magazine*, vol. 17, no. 6, pp. 20–46, Sep. 2000.

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

[8] I. Karybali and C. Berberidis, "Blind image-adaptive watermarking," in *Proc. ICECS 2003*, Sharjah, UAE, Dec. 2003, pp. 894–897.

[9] S. Voloshynovskiy, S. Pereira, V. Iquise and T. Pun, "Attack modelling: Towards a second generation benchmark," *Elsevier Signal Processing*, vol. 81, pp. 1177–1214, Jun. 2001.

Table 2: Detectors' responses (using u_0) for different attacks. Dash means that detection is impossible.

AWGN	***30dB***	***20dB***	***10dB***	***0dB***
Boat (DD)	0.0259	0.0258	0.0247	0.0205
Boat (PED)	0.4033	0.2110	0.0785	0.0291
JPEG Compr.	***Q80***	***Q50***	***Q15***	***Q10***
Boat (DD)	0.0108	-	-	-
Boat (PED)	0.1504	0.0624	0.0437	0.0283
Colour Reduce	***Dithering***		***Thresholding***	
Boat (DD)	0.0382		0.0340	
Boat (PED)	0.0617		0.1553	
Sampledownup	***Case 1***		***Case 2***	
Boat (DD)	0.0146		-	
Boat (PED)	0.2405		0.0454	
Wiener Filtering	***3x3***		***5x5***	
Boat (DD)	-		-	
Boat (PED)	0.1696		0.1947	
Trimmedmean	***3x3***		***5x5***	
Boat (DD)	-		-	
Boat (PED)	0.0540		0.0740	
Median	***3x3***	***5x5***	***7x7***	
Boat (DD)	-	-	-	
Boat (PED)	0.0695	0.0234	0.0080	
Other Filters	***Laplacian***	***Gaussian***	***Unsharp***	
Boat (DD)	-0.3951	0.0158	0.1099	
Boat (PED)	-0.4882	0.5198	0.5130	