OPTIMAL DETECTOR STRUCTURE FOR DCT AND SUBBAND DOMAIN WATERMARKING

Athanasios Nikolaidis *Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE, {nikola, pitas}@zeus.csd.auth.gr

ABSTRACT

Most of the watermarking schemes that have been proposed until now employ a correlator in the detection stage. The current paper proposes a new detector scheme that can be applied in the case of additive watermarking in the DCT or DWT domain. Certain properties of the probability density function of the coefficients in these domains are exploited in order to construct an asymptotically optimal detector based on well known results of the detection theory. Detection is performed without the use of the original image, as in methods employing different detectors. Experimental results prove the superiority of the proposed detector over the correlator.

1. INTRODUCTION

Image watermarking as a tool for copyright protection has attracted a lot of attention in the last few years [1, 2]. The watermark is superimposed on the image in either an additive or a multiplicative way, either in the spatial or in a transform domain. In either case, the detection stage involves, for the majority of the reported techniques, a simple correlator. This is done under the general implicit assumption that the image pixels or transform coefficients on which the watermark is embedded follow a normal (Gaussian) distribution. If this is indeed the case, the correlator is the optimal detector in the sense that, under a constant probability of false alarm, the probability of false rejection is minimized [3]. However, the host data distribution cannot, in most cases, be adequately approximated by a Gaussian pdf.

Luckily enough, the pdf of coefficients of certain transform domains can be sufficiently approximated by a generalized Gaussian distribution. Such an assumption has been adopted in [4], where the embedding domain is that of a band of DCT coefficients except for the DC term. A generalized Gaussian distribution is considered for all coefficients of this band, and an additive embedding rule is adopted. The ML criterion is employed in order to derive a detector structure that is optimal under the assumption that the watermark power is known [4].

The present paper proposes a detector structure that displays superior performance compared to other detectors, including the correlator, under certain assumptions. Watermarking in the DCT and DWT domains is considered in our work. Although the pdf assumption for the coefficients in these domains is the same as in [4], we derive a different detector structure that is optimal under the assumption of unknown watermark power, by employing a Rao test that is equivalent with a GLRT (generalized likelihood ratio test). The resulting detector is asymptotically optimal, meaning that it is optimal under the assumption of a large data record. The derivation of the optimal detector given the additive embedding model and assuming ignorance about the watermark power is presented in Section 2. In Section 3, a description of the algorithms involved in estimating the generalized Gaussian distribution parameters is presented. Section 4 presents the specifics for embedding in the DCT and DWT domains. In Section 5, experimental results confirm the superiority of the proposed detector compared to the correlator, even in the presence of compression attacks. Robustness against geometric attacks is not examined, since the aim of the paper is only to propose an optimization of the detection stage. Thus, synchronization compensating for such attacks is assumed to be present. Finally, certain conclusions about the performance of the proposed detector are drawn in Section 6.

This work has been partially supported by the European Project IST-1999-10987 CERTIMARK.

2. OPTIMAL DETECTOR STRUCTURE

Since the embedding model is that of an additive watermark inserted in a transform domain (e.g. DCT, DWT) and modulated by a factor corresponding to the watermark power, the hypothesis test under consideration throughout this paper will be of the form:

$$\begin{aligned} \mathcal{H}_1 : Y[\boldsymbol{k}] &= X[\boldsymbol{k}] + \alpha W[\boldsymbol{k}] \\ \mathcal{H}_0 : Y[\boldsymbol{k}] &= X[\boldsymbol{k}] \end{aligned} \tag{1}$$

where X is the original image transform coefficient that follows a certain pdf model, whose parameters can be estimated from the watermarked image transform coefficient. W is the watermark that is embedded in the original image transform coefficient, and α is an amplitude parameter that corresponds to a generally unknown watermark power.

Most of the methods that have been proposed until now employ a correlator-detector. This detector is the easiest to implement and is based on evaluating the correlation of the possibly watermarked (or even attacked) image transform coefficients and the watermark under investigation:

$$D(Y) = \sum_{\boldsymbol{k}} Y[\boldsymbol{k}] W[\boldsymbol{k}] \tag{2}$$

where $\boldsymbol{k}$ are the indices of the transform coefficients where the watermark is supposed to be embedded, Y is the watermarked (and possibly attacked) transformed image, and W is the two-dimensional watermark. The probability distribution of the correlator detector is approximately normal under both hypotheses. According to detection theory, this detector is expected to be optimal, that is, to present the lowest possible error probabilities, in the case that the noise (in our case, the original image transform coefficient) is white Gaussian (WGN), meaning that it has a white spectrum and follows the normal distribution.

However, in practice most image transform coefficients cannot be considered as white Gaussian noise. This leads to the conclusion that watermarking algorithms that are based on a correlator-detector perform only suboptimally. Thus, a more sophisticated detector structure in conjunction with a better, image-dependent approximation of the pdf of the embedding domain coefficients is necessary in order to achieve superior detector performance.

In this paper we assume, as in [4], that all the DCT or DWT transform coefficients, in which the watermark is embedded, follow a generalized Gaussian distribution with zero mean value and the same parameters β and σ^2:

$$p(X[\boldsymbol{k}]) = \frac{c_1(\beta)}{\sqrt{\sigma^2}} \exp\left(-c_2(\beta)\left|\frac{X[\boldsymbol{k}]}{\sqrt{\sigma^2}}\right|^{\frac{2}{1+\beta}}\right) \tag{3}$$

where:

$$c_1(\beta) = \frac{\Gamma^{\frac{1}{2}}\left(\frac{3}{2}(1+\beta)\right)}{(1+\beta)\Gamma^{\frac{3}{2}}\left(\frac{1}{2}(1+\beta)\right)} \tag{4}$$

$$c_2(\beta) = \left[\frac{\Gamma\left(\frac{3}{2}(1+\beta)\right)}{\Gamma\left(\frac{1}{2}(1+\beta)\right)}\right]^{\frac{1}{1+\beta}} \tag{5}$$

β is called the *shape* parameter ($\beta > -1$), σ^2 is the variance and $\Gamma(x)$ is the Gamma function. This assumption has been proven to hold both for DCT and DWT domain coefficients, and has already been employed in coding applications [5]. It is proven in [6] that a generalized Gaussian model is more suitable for all DCT coefficients but the DC, because it provides the lowest MSE (mean-squared error) during quantization.

Kay [7] has proved that, for such a problem, a Rao test has asymptotically optimal performance (i.e. it is indeed optimal for large images) that is equivalent to that of a generalized likelihood ratio test (GLRT). This test can be written as:

$$D_R(Y) = \frac{\left[\sum W[\boldsymbol{k}]\frac{p'(Y[\boldsymbol{k}])}{p(Y[\boldsymbol{k}])}\right]^2}{\frac{1}{N}\sum W^2[\boldsymbol{k}]\sum\left[\frac{p'(Y[\boldsymbol{k}])}{p(Y[\boldsymbol{k}])}\right]^2} \tag{6}$$

where N is the number of samples that have been watermarked, p is the pdf and p' is the derivative of the pdf of the image transform coefficients under hypothesis $\mathcal{H}_0$.

Let us now consider the case of a known binary watermark ($W[\boldsymbol{k}] \in \{-1, 1\}$). Let us also assume that the possibly watermarked transform coefficients approximately follows a generalized Gaussian distribution with the same parameters as the original image transform coefficient. This can easily be proved to hold for small values of watermark power. Thus, parameter estimation can be performed using the watermarked image transform coefficients instead of the original, which are not available in the detection stage.

Formula (6) can be rewritten as:

$$D_R(Y) = \frac{\left(\sum sgn(Y[\boldsymbol{k}])W[\boldsymbol{k}]|Y[\boldsymbol{k}]|^{\frac{1-\beta}{1+\beta}}\right)^2}{\sum |Y[\boldsymbol{k}]|^{\frac{2(1-\beta)}{1+\beta}}} \tag{7}$$

where $sgn(x)$ is the signum function and β is the shape parameter of the generalized Gaussian pdf. It has been shown that, for large data records, this detector follows a chi-squared distribution under both detection hypotheses. More specifically, the distribution of the detector under $\mathcal{H}_0$ is χ_1^2 that is, a chi-squared distribution with one degree of freedom, whereas under $\mathcal{H}_1$ it is $\chi'^2(1,\lambda)$ that is, a non-central chi-squared distribution with one degree of freedom and non-centrality parameter λ. The performance of this detector is the same as that of the generalized log-likelihood ratio test ([7]). Since a random variable that follows a $\chi'^2(1,\lambda)$ distribution is equivalent to the square of a normal random variable with mean value $\sqrt{\lambda}$ and variance equal to 1, it follows that the probabilities of false alarm and false rejection (P_{FA} and P_{FR}) which characterize the detector performance are given by:

$$\begin{aligned} P_{FA} &= 2Q(\sqrt{\gamma'}) \\ P_{FR} &= 1 - Q(\sqrt{\gamma'} - \sqrt{\lambda}) - Q(\sqrt{\gamma'} + \sqrt{\lambda}) \end{aligned} \tag{8}$$

where γ' is the detection threshold. The watermark detection performance depends on the so-called *non-centrality parameter* λ. For the case of the generalized Gaussian pdf for a certain range of values of the parameter β ($-1 < \beta < 3$), parameter λ can be defined as:

$$\lambda = N\alpha^2 \frac{4\Gamma(\frac{3-\beta}{2})\Gamma(\frac{3}{2}(1+\beta))}{(1+\beta)^2\sigma^2\Gamma^2(\frac{1}{2}(1+\beta))} \tag{9}$$

It is interesting to see that P_{FR} decreases monotonically with λ. This means that, for increased watermark power α or for larger sample number N, the probability of false alarm decreases, as expected.

3. GENERALIZED GAUSSIAN PARAMETER ESTIMATION

As derived in Equation (7), the detector formula involves the value of the parameter β of the generalized Gaussian approximation of the transform coefficients. In order to assess the performance of the new detector using 9, we should also have an estimate of σ so that the parameter λ can be calculated.

The variance of the generalized Gaussian distribution that fits best to the data can be estimated using the sample variance:

$$\hat{\sigma}_Y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu}_Y)^2 \tag{10}$$

where $\hat{\mu}_Y$ is the sample mean, N is the number of image data samples and Y_i are the data samples of the watermarked transformed image under test. The estimates are quite satisfactory since the number of data samples are of the order of 10^4 for images considered in our experiments.

As far as the shape parameter β_Y is concerned, we wish to estimate it in a computationally efficient way. For this reason, the technique proposed in [8] is adopted. This is based on exploiting the relation between the variance σ_Y^2, the mean of the absolute values $E\left[|Y|\right]$ and the shape parameter β_Y:

$$r(\beta_Y) = \frac{\sigma_Y^2}{E^2\left[|Y|\right]} = \frac{\Gamma\left(\frac{1}{2}(1+\beta_Y)\right)\cdot\Gamma\left(\frac{3}{2}(1+\beta_Y)\right)}{\Gamma^2(1+\beta_Y)} \tag{11}$$

also called the *generalized Gaussian ratio function*. After computing the estimate of the variance $\hat{\sigma}_Y^2$ by using Equation (10) and the estimate of the mean of the absolute values using $\hat{E}\left[|Y|\right] =$

$(1/N)\sum_{i=1}^{N}|Y_i - \hat{\mu}_Y|$ (the mean value is subtracted since we want an estimate of the parameters that is as close to the original data as possible), we calculate the ratio $\rho = \hat{\sigma}_Y^2/\hat{E}^2[|Y|]$ and finally solve the equation 11 for β_Y using e.g. a lookup table.

4. EMBEDDING IN DCT AND SUBBAND DOMAINS

4.1. DCT domain

An investigation on the possible exploitation of the (non-DC) DCT coefficients distribution has been previously done in [4]. Our approach assumes that the watermark power is unknown to the detector, which is realistic. The resulting detector structure is optimal for the asymptotic case, which is fulfilled in practice.

The dataset on which we embed the watermark is the set of the non-DC coefficients resulting from a 8×8-block DCT transformation. We estimate the parameters of the generalized Gaussian distribution based on all coefficient samples. An example is given in Figure 1 where the watermarked DCT coefficients of the Baboon image are approximated by a generalized Gaussian pdf, with shape parameter $\beta = 1.76$ and variance $\sigma^2 = 0.0171$.

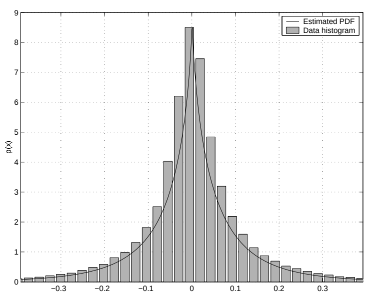


Fig. 1. Watermarked data samples histogram of DCT domain of Baboon and its generalized Gaussian pdf model fitting.

The DCT coefficients data set is traversed from top left to bottom right in a zig-zag order (the DC term is located in the upper left corner of the block). We can choose to use a diagonal band of low-to-middle frequency coefficients for watermark embedding in order to achieve robustness to lowpass filtering and compression. For example, we can choose those coefficients $C(u, \nu)$ that satisfy $3 \leq u + \nu \leq 6$ for a 8×8 DCT block.

4.2. Subband domain

As pointed out in [6] as well as in the classic paper of Mallat on wavelets [9], subband image data can be fairly well represented by generalized Gaussian distributions. This holds for all frequency bands but the lowest one. An example is given in Figure 2, where the watermarked coefficients of the HL1 subband of the Baboon image are approximated by a generalized Gaussian pdf with shape parameter $\beta = 1.79$ and variance $\sigma^2 = 0.0137$.

The 2-D DWT decomposes an image into space-frequency subbands by applying lowpass and corresponding highpass filters to the original image at each dimension and subsequently downsampling the result by a factor of 2. In this way, the so-called *detail* images are produced, as well as a smoothed image. This can be repeatedly performed up to the desired resolution level.

5. EXPERIMENTAL RESULTS

In order to confirm the superiority of the new detector compared to the classic correlator, and to the correlator proposed in [4], we

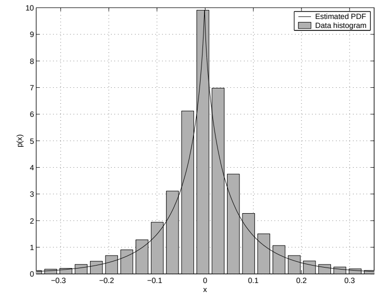


Fig. 2. Watermarked data samples histogram of HL1 subband of Baboon and its generalized Gaussian pdf model fitting.

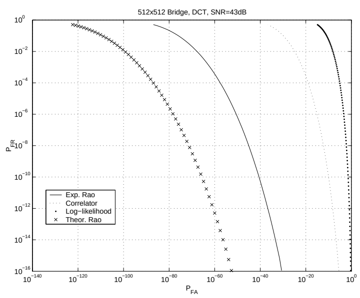
conducted a number of experiments on several real images of size 512×512 and for several SNR values (equivalently, watermark power values). Experiments were performed by embedding different binary pseudo-random watermarks in the coefficients of either the DCT or the DWT domain, and afterwards detecting them by the proposed detector, the correlator and the detector proposed in [4], having always in mind that the watermark power is not known at the detection stage. For this reason, a constant value of 1 is implied for watermark power in the case of the detector in [4]. The performance of all detectors was evaluated on undistorted images as well as JPEG compressed images (in the case of DCT embedding) and SPIHT compressed images (in the case of DWT embedding).

An example of a result for DCT domain embedding is shown in Figure 3. The original and watermarked versions of the Bridge image of size 512×512 are shown in Figures 3(a) and 3(b) respectively, for SNR$\simeq$43dB. Because of the high SNR value, no visual artifacts can be noticed. Figure 3(c) shows the ROC curves (plots of P_{FA} versus P_{FR} for several detection thresholds) for all considered detectors, when no attack (i.e. intentional or unintentional image processing operation) is imposed. The superiority of the proposed detector is obvious, whereas the other two methods exhibit similar performance. A curve that is derived theoretically based on the estimated values of the pdf parameters, is delineated in the same figure with $\mathbf{x}$. The respective experimental curve demonstrates lower performance, due to the bias of the experimental data from the estimated pdf. After a moderate JPEG compression attack of quality factor 95%, the performance of all detectors degrades but the new detector still demonstrates a better performance, as it can be seen in Figure 3(d).

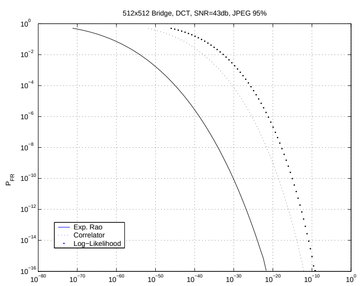
In the case of DWT domain embedding, the watermark is embedded in all coefficients that correspond to subbands of a 2-level transformation except for the smoothed image that corresponds to the LL2 subband. Figure 4(a) shows the ROC (receiver operating characteristics) curves resulting from the watermark detection in subband HH1 of the Boat image for SNR$\simeq$37dB, for additive embedding in the subbands of the 2-level DWT transformation (except LL2). The new detector displays, again, better performance, especially when compared to the log-likelihood-based detector. It is also superior to the correlator detector. The curve that is derived theoretically after pdf parameter estimation is delineated with $\mathbf{x}$. We can notice that the experimental performance is, again, worse than the theoretical but it is substantially superior to that of the other detectors. The same observations also hold after a moderate SPIHT compression (3bpp), as is shown in Figure 4(b).

(a)

(b)



(c)

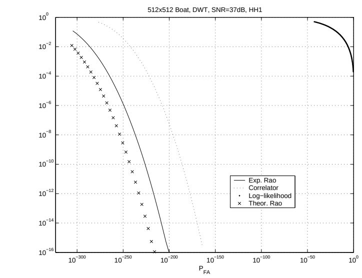


(d)

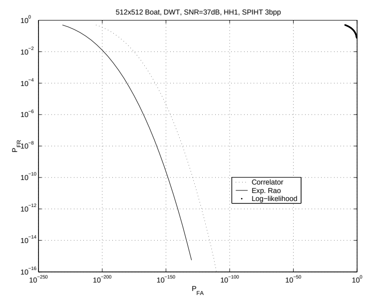
Fig. 3. (a) Original Bridge image. (b) Bridge watermarked in the DCT domain (SNR≃43dB). (c) ROCs without attack. (d) ROCs after JPEG compression (95%).

6. CONCLUSIONS

A new detector structure for watermarking schemes has been proposed in this paper. This is derived from well grounded theoretical results of statistical detection theory. A detector that is equivalent to a GLRT is introduced, that is asymptotically optimal in GWGN pdf models. Optimality is ensured in the case of watermarking since the datasets are adequately large. A generalized Gaussian distribution is assumed both in the DCT and DWT watermark embedding domains. The pdf parameters can be estimated sufficiently from the watermarked transform coefficients instead of the original transform coefficients without violating the test requirements. The detector is proven to perform optimally in practice when no attacks are imposed, but also under moderate compression levels, considering the watermark power employed in our experiments. Results are compared to those of the correlator-detector and the detector derived based on the log-likelihood ratio function, when no knowledge of the watermark power is available.



(a)



(b)

Fig. 4. (a) ROCs without attack for watermarked Boat image (SNR≃37dB). (b) ROCs after SPIHT compression (3bpp).

7. REFERENCES

[1] M.D. Swanson, M. Kobayashi and A.H. Tewfik, "Multimedia Data Embedding and Watermarking Technologies", in*Proceedings of the IEEE*, **86**(6), pp. 1064-1087.

[2] F. Hartung and M. Kutter, "Multimedia Watermarking Techniques", in*Proceedings of the IEEE*, **87**(7), pp. 1079-1107.

[3] S. M . Kay, "Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory", Prentice Hall, New Jersey, 1998.

[4] J. R. Hernández, M. Amado and F. Pérez-González, "DCT-Domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure", in *IEEE Trans. on Image Processing*, **9**(1), pp. 55-68.

[5] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images", in *IEEE Trans. on Communications*, **31**(6), pp. 835-839, 1983.

[6] K. A. Birney and T. R. Fischer, "On the Modeling of DCT and Subband Image Data for Compression", in *IEEE Trans. on Image Processing*, **4**(2), pp. 186-193.

[7] S. M. Kay, "Asymptotically Optimal Detection in Incompletely Characterized Non-Gaussian Noise", in *IEEE Trans. on Acoustics, Speech and Signal Processing*, **37**(5), pp. 627-633, 1989.

[8] K. Sharifi and A. Leon-Garcia, "Estimation of Shape Parameter for Generalized Gaussian Distributions in Subband Decompositions of Video", in *IEEE Trans. on Circuits and Systems for Video Technology*, **5**(1), pp. 52-56.

[9] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", in IEEE Trans. on Pattern Analysis and Machine Intelligence, **11**(7), pp. 674-693.