ANALYSIS OF WATERMARK DETECTION USING SPOMF

T. Kalker and A.J.E.M. Janssen

Philips Research Laboratories Eindhoven
Prof. Holstlaan 4
5656AA Eindhoven, The Netherlands
{kalker,janssena}@natlab.research.philips.com

ABSTRACT

An important aspect of almost every watermarking system is the performance of the method by which watermark patterns are correlated with (derived features of) the suspect content. In the JAWS video watermarking system, developed at Philips Research, correlation is realized by Symmetrical Phase-Only Matched Filtering. SPOMF detection has experimentally been shown to be an excellent correlation method. In this paper we address the problem of building a theoretical understanding of its performance. Being a non-linear correlation method, analysis of SPOMF detection performance is difficult. This paper is a first attempt to tackle the problem of deriving an analytical relationship between the energy of the embedded watermark and the reliability of detection. The theoretical results are compared with experimental results, and are found to be in good agreement.

1. INTRODUCTION

Watermarking is a technology that can be used to provide copyright protection for multimedia content [1] [2] [3]. A watermark is an imperceptible – or at least unobtrusive – label, which is holographically attached to the content. A watermark is designed to be robust with respect to any signal processing: as long as the content is of good /useful quality, the watermark should be detectable by dedicated hard- or software. Moreover, intentional attempts to remove the watermark should succeed only when content quality is severely reduced. Depending on the application, several other requirements may be imposed with respect to the complexity of embedding and detection, payload, rate of false positives and negatives, open or blind detection, and others.

At Philips Research, a new watermarking technology for copy protection and tracing of video has been developed.The technology is referred to as JAWS (Just Another Watermarking System). In this section we will give a short exposition of the basic features of the JAWS system. For more details we refer to [4].

Watermark embedding in JAWS basically treats each video field as a still image and the same watermark is embedded in each field. Watermark embedding consists of adding one or more normally distributed pseudo-random sequences W. Spatial masking is applied to control the visual impact of the watermark: image regions with little activity have little watermarking energy and regions with high activity – e.g. textured areas – have substantial watermark energy. There is no obvious choice for an image activity measure, but we have experimentally found that the response to a simple Laplacian filter L suffices,

$$L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

For reasons of complexity reduction, the sequence W is endowed with translational symmetry. In particular, the sequence W is obtained by *tiling* (with possible truncation) an $N \times N$ matrix W_0, where N is typically 128. The matrix W_0 is referred to as a primitive pattern. A watermark payload larger than one bit is obtained by embedding multiple watermarks. By tiling primitive patterns at different phases (i.e. with non-overlapping tiling grids) it is possible to re-use primitive patterns [5].

Correlation is used to detect watermarks. Due to the facts that (1) each video field is marked using the same noise pattern and (2) the noise patterns have translational symmetry, the complexity of correlation can be greatly reduced. A number of consecutive video fields is collected in an $N \times N$ *fold buffer* B using temporal accumulation and spatial folding. The content of the fold buffer B is subsequently correlated with a primitive pattern W_0 to obtain a *decision variable* d. If the value of d is large, the video content is considered to have an embedded watermark. Otherwise we assume that no watermark is present. In order to quantify the meaning of "large", it is important to estimate the standard deviation σ_d of d for unmarked content. It is not difficult to show that σ_d is proportional to the standard deviation σ_B of the fold buffer, and inversely proportional to the size N of a primitive pattern [6]. Typically a detection event is reliable if d is larger than $5\sigma_d$.

Detection can be boosted by pre-filtering the input video fields with an appropriate high pass and zero-phase filter $F(z)$. The filter $F(z)$ is designed such that (1) the standard deviation of the fold buffer B is considerably reduced while (2) a high correlation is retained between fold buffer B and primitive pattern W_0. A simple tensor product of the one-dimensional filter $A(z) = -\frac{1}{2}z + z - \frac{1}{2}z^{-1}$is in general sufficient for obtaining a boost in reliability of up to 20dB [6].

An important requirement of a video watermarking system is that it should be resistant to spatial shifts. An obvious method for obtaining shift invariance is to exhaustively search over all possible shifts. Shifts in the domain of video fields translate to cyclic shifts of the fold buffer B. Correlating a primitive pattern W_0 with all possible cyclic shifts of B can be formulated as the cyclic convolution $D = B \otimes \mathrm{conj}(W_0)$ of B and W_0, where "$\otimes$" and "conj" denote cyclic convolution and pointwise conjugation, respectively. The left-hand expression D denotes a matrix of correlation values. An efficient method for computing this cyclic convolution (and therefore for obtaining shift invariance) is by using the Fast Fourier Transform (FFT) $\mathcal{F}$,

$$D = \mathcal{F}^{-1}(\mathcal{F}(B) * \mathrm{conj}(\mathcal{F}(W_0))), \tag{1}$$

where "$*$" denotes point-wise multiplication.

There are a number of additional advantages in computing the correlation matrix D using an FFT. Firstly, the pre-filtering by $F(z)$ can be implemented in the Fourier domain. It is known that an optimum pre-filter results in a spectrally white signal. We can therefore take this pre-filtering to an extreme and retain only the phase $\phi(B)$,

$$[\phi(B)]_i = \frac{B_i}{|B_i|}, \tag{2}$$

of the fold buffer. By assumption, the pattern W_0 is already approximately spectrally white, but by forcing W_0 to be pure white, even better detection reliability is obtained,

$$D = \mathcal{F}^{-1}(\phi(\mathcal{F}(B)) * \phi(\mathrm{conj}(\mathcal{F}(W_0)))). \tag{3}$$

The non-linear correlation method of Eq. (3) is known as Symmetrical Phase-Only Matched Filtering (SPOMF). It is a well-known method in the field of pattern recognition [7].

Secondly, assuming that W_0 is embedded with a multiplicity μ ($\mu \ll N^2$), only a few of the values in D will be large. The other entries in D can be viewed as decision variables for unmarked content. By directly measuring the standard deviation of these non-extremal values, a yardstick for comparing the extremal value in D is obtained. As the number μ is assumed to be small in comparison with N^2, the extremal values in D can be included in the computation of the standard deviation. It is not difficult to derive (using Parseval) that the standard deviation of the matrix D computed in Eq. (3) is equal to $1/N$.

In view of the above observations, the presence of a watermark in a SPOMF buffer D can be asserted by the presence of entries larger than T/N, where T is a threshold parameter. Assuming a normal distribution for the buffer D, and setting $T = 5$, we obtain a false positive rate of $5.7 * 10^{-7}$.

In the context of SPOMF detection we now raise the following question: given a fold buffer B corresponding to unmarked content, how much watermark energy needs to be inserted to achieve a given reliability of detection T? The answer to this question is important for watermark embedding. Without a simple relationship between watermark energy and detection reliability, a watermark embedder will have to resort to an iterative procedure in order to achieve a given reliability T.

The goal of this paper is to investigate the relationship between watermark energy and reliability of SPOMF detection in an analytical manner. This paper is organised as follows. We start by stating a precise mathematical formulation of the problem. This formulation includes a model of the original unmarked content and the watermark pattern as auto-regressive processes. We subsequently give an analytical solution for the simple case that both content and watermark are spectrally white. These results have been confirmed experimentally. Finally we address the general case in which both content and watermark may have an arbitrary spectral colour.

2. MATHEMATICAL FORMULATION OF THE PROBLEM

For the remainder of this paper all sequences are assumed to be one-dimensional. Let $B = \{b_i\}$ and $W = \{w_i\}$, $i = 0, \ldots, N - 1$ be two uncorrelated, finite-length signals. We assume that B and W are instances of auto-regressive processes with correlation factors α_B and α_W, respectively. This can be expressed in the z-domain as

$$B(z) = \frac{\sqrt{1-\alpha_B^2}\, V_B(z)}{1-\alpha_B z^{-1}}, \tag{4}$$

$$W(z) = \frac{\sqrt{1-\alpha_W^2}\, V_W(z)}{1-\alpha_W z^{-1}}, \tag{5}$$

where $V_B(z)$ and $V_W(z)$ are the z-expressions for i.i.d. normally distributed noise. Without loss of generality we may assume that V_B and V_W are zero mean and have a standard deviation equal to 1. Then also the sequences $B(z)$ and $W(z)$ are zero-mean, have a standard deviation equal to 1 and have an autocorrelation factor α_B and α_W, respectively.

Let $S(B, W)$ denote the value of the DC entry of the SPOMF correlation given in Eq. (3), i.e.

$$\begin{aligned} S(B,W) &= \\ &= \left[\mathcal{F}^{-1}(\phi(\mathcal{F}(B)) * \phi(\mathrm{conj}(\mathcal{F}(W_0))))\right](0,0) \\ &= \frac{1}{N}\sum_{i,j}\left[\phi(\mathcal{F}(B)) * \phi(\mathrm{conj}(\mathcal{F}(W_0)))\right](i,j). \end{aligned} \tag{6}$$

In the context of this paper, watermark embedding corresponds to adding a scalar multiple of the pattern W_0 to the content of a fold buffer B. More mathematically, watermarked content corresponds to the expression $B + \lambda W_0$, where λ determines the energy of the embedding. Watermark detection corresponds to the computation of $S(B + \lambda W_0, W_0)$. As we are interested in the reliability of detection as a function of the embedding strength, we will study the function

$$f_{\alpha_B \alpha_W}(\lambda) = \mathrm{E}[S(B + \lambda W, W)], \tag{7}$$

where "E[-]" denotes the expectation operator. Whenever convenient, we will suppress the indices α_B and α_W and simply write $f(\lambda)$. The result below is an essential ingredient for the discussions in the following sections.

Lemma 1 *The coefficients of the Fourier transform of spectrally white or coloured sequences as in Eq. 4 and Eq. 5 are circular normal random variables. Moreover, coefficients with different indices are uncorrelated.*

Proof. Observe that it is sufficient to prove the lemma for spectrally white sequences. A spectrally coloured sequence is obtained by from a white sequence by complex multiplication in the Fourier domain with a fixed factor γ_k for each coefficient y_k. It is obvious that circularity, normality and correlation between indices are not affected by complex multiplication. For the remainder of this proof we assume spectrally white sequences.

We observe that the Fast Fourier transform $\mathcal{F}$ is a unitary operator on complex vector spaces V. It follows that $\mathcal{F}$ is an orthonogal operator on the underlying real vector space $V_{\mathbb{R}}$. The probability density function (*pdf*) for real-valued, i.i.d. and normally distributed sequences $x = (x_0, \ldots, x_{n-1}) \in V$ is of the form $p(x) = Ce^{-|x|^2/2\sigma^2}$, where C is an appropriate constant. The orthogonal operator $\mathcal{F}$ maps the $\mathbb{R}$-linear subspace of real-valued sequences to the $\mathbb{R}$-linear subspace of conjugate-symmetric sequences $y = (y_0, \ldots, y_{n-1})$, $y_k = \mathrm{conj}(y_k)$. Assuming n to be even, conjugate-symmetric sequences are parametrised by the vectors $z = (y_0 + i y_{n/2}, y_1, \ldots, y_{(n/2)-1})$. A simple calculation shows that the pdf $p(z)$ is of the form

$$p(z) = Ce^{-\frac{z_0^2+2|z_1|^2+\cdots+2|z_n|^2}{2\sigma^2}}, \tag{8}$$

which shows that the coefficients z_k are circular normal random variables. Exploiting the unitary property of $\mathcal{F}$ once more, one shows that $\mathrm{E}[z_k \operatorname{conj}(z_l)] = \alpha_k \delta_{kl} \sigma^2$, where $\alpha_k = 2$ for $k = 0$ and $\alpha_k = 1$ otherwise.

□

3. THE SIMPLE CASE

In this section we will assume that both the content B and the watermark W are spectrally white, i.e. we assume that $\alpha_B = \alpha_W = 0$. This implies that except for one (n is odd) or two (n is even) indices, the coefficients of $\mathcal{F}(B)$ and $\mathcal{F}(W)$ are all circular normal random variables with the same standard deviation. Inspecting Eq. (6) and Eq. (7) it follows that the computation of $f(\lambda)$ is equivalent to the computation of the expectation $\mathrm{E}[g_\lambda(z_1, z_2)]$, where z_1 and z_2 are circular normal random variables and where $g_\lambda(z_1, z_2)$ is defined by

$$g_\lambda(z_1, z_2) = \frac{z_1 z_2^* + \lambda z_2 z_2^*}{\left|z_1 z_2^* + \lambda z_2 z_2^*\right|}. \tag{9}$$

It is easy to see that for $\lambda \to \infty$, the expectation $\mathrm{E}[g_\lambda(z_1, z_2)]$ goes to 1, and that for $\lambda = 0$ the expectation value is equal to 0. For general λ, $0 \leq \lambda < \infty$, the value of $\mathrm{E}[g_\lambda(z_1, z_2)]$ is given by the following theorem.

Theorem 1 *With notations and assumptions as above, the following equality holds*

$$f(\lambda) = \mathrm{E}[g_\lambda(z_1, z_2)] = \int_0^{\pi/2} \frac{\sin^2(x)}{\sqrt{\lambda^{-2} + \sin^2(x)}} dx, \tag{10}$$

The right-hand side of Eq. (10) can also be written in terms of hypergeometric functions as

$$f(\lambda) = \frac{\pi}{4\sqrt{1+\lambda^{-2}}} F(1/2, 1/2; 2, 1/(1+\lambda^{-2})). \tag{11}$$

Proof. The result Eq. 11 follows immediately from Eq. (10) using [8, 331-92, p. 106]. It therefore remains to prove Eq. (10). This latter equation is proved in a number steps of which we only sketch the outline. In the first step we compute the conditional probability $p_\varphi(\varphi | z_2 = \underline{z}_2)$, where $\varphi = \arg(z_1 \operatorname{conj}(z_2) + \lambda |z_2|^2)$. Because of circularity we may assume that z_2 is real and that it is sufficient to compute $p_\varphi(\arg(\omega))$, where $\omega \sim \mathcal{N}(\lambda|\underline{z}_2|^2, |\underline{z}_2|)$. In the second step we resolve the dependence on $\underline{z}_2$ by evaluating the integral

$$\iint_{\mathbb{C}} p_\varphi(\varphi | z_2 = \underline{z}_2) p_{z_2}(\underline{z}_2) d\underline{z}_2, \tag{12}$$

which, after some manipulation, results in an integral of the form

$$\int_0^\infty e^{-ar^2} \operatorname{erfc}(br) dr \tag{13}$$

This latter integral can then be simplified using [9, 6.285, p. 649]. The resulting expression for $p_\varphi(\varphi)$ splits as a sum of a symmetric (i.e. invariant for $\varphi \to \varphi + \pi$) and an anti-symmetric term. The symmetric term does not contribute to the computation of $\mathrm{E}[e^{i\varphi}]$; the anti- symmetric term is of the form

$$-\frac{1}{2\pi} \int_0^{2\pi} g(\varphi) g''(\varphi) d\varphi, \tag{14}$$

where $g(\varphi) = \pi/2 + \arctan(\cos(\varphi)/\sqrt{\lambda^{-2} + \sin^2(\varphi)})$. The expected value of $e^{i\varphi}$ is therefore computed as

$$-\frac{1}{2\pi} \int_0^{2\pi} g''(\varphi) g(\varphi) \cos(\varphi) d\varphi. \tag{15}$$

In the final step, the expression in Eq. (15) is massaged in the required form of Eq. (10). □

The theoretical result of Theorem 1 is compared with experimental results in Figure 1. Referring to this figure, we see that the dashed curved, representing the theoretical formula of Theorem 1, completely overlays the the experimental results (the *crosses*-curve, obtained by software simulations). We therefore have complete agreement between theory and experiment.

We can also interpret the two (overlaying) curves as a measures of reliability of watermark detection by comparing it with the standard deviation of the noise. If we set N to be the number of samples, then the standard deviation of the noise level can be shown to be $1/\sqrt{N}$. The optimum reliability of SPOMF detection is therefore $1/\sqrt{N}$ and is reached for $\lambda \to \infty$. In the case of linear correlation, the reliability of watermark detection is given by the expression $\lambda/\sqrt{1+\lambda^2}\sqrt{N}$ (solid curve in Figure 1). This shows that in a spectrally white scenario, linear correlation is a slightly more reliable watermark detection method than SPOMF. However, in actual watermarking practice the energy in the watermark is considerably smaller than the energy of the signal to be marked. In particular the value of λ will usually be (substantially) smaller than 1, in which case SPOMF and linear correlation have the same performance.

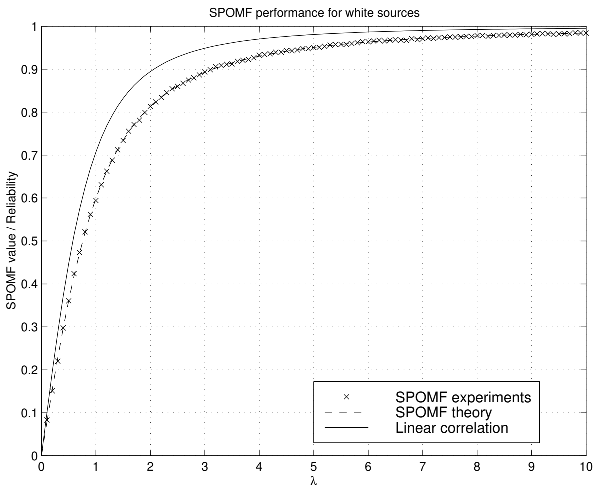

Figure 1: Theoretical and experimental performance of SPOMF.

4. THE GENERAL CASE

In the general case we impose no restrictions on the correlation factors α_B and α_W. The coefficients of the Fourier transforms of B and W are still independent circular normal random variables, but the standard deviation now depends on the index. The ratio $\lambda_k$between standard deviations for index k is given by $\lambda(2\pi k/N)$, where $\lambda(\varphi)$ is given by

$$\lambda(\varphi) = \lambda\sqrt{\frac{1+\alpha_B^2 - 2\alpha_B\cos(\varphi)}{1+\alpha_W^2 - 2\alpha_W\cos(\varphi)}}\sqrt{\frac{1-\alpha_W^2}{1-\alpha_B^2}}. \qquad (16)$$

Averaging over all indices k, and letting $N \to \infty$, we obtain the following result.

Theorem 2 *The function $f(\lambda)$ is given by the integral*

$$f(\lambda) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{E}[g_{\lambda(\varphi)}(z_1, z_2)]d\varphi. \qquad (17)$$

In Figure 2 experimental results of SPOMF detection with white watermarks on correlated sources are presented. With reference to this figure we see that SPOMF detection outperforms linear correlation for (highly) correlated sources.

Sofar we have failed to find an analytical expression for Eq. (17). We hope to resolve this unsatisfactory situation in a future publication, where we will also address the comparison with common (linear) matched filtering techniques.

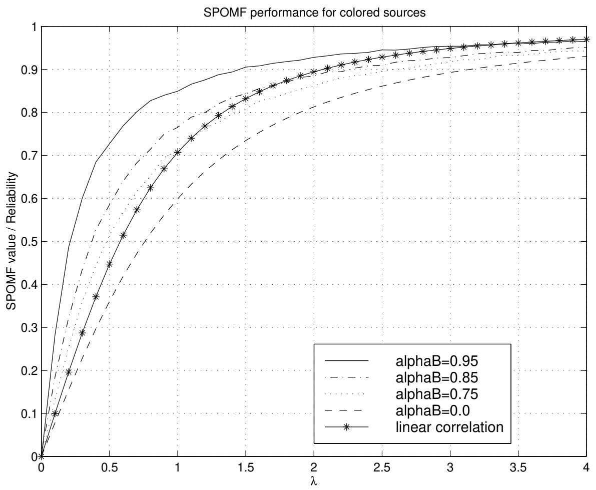

Figure 2: SPOMF performance for spectrally coloured sources.

5. CONCLUSIONS

In this paper we have addressed the problem of relating energy of embedded watermarks to the reliability of SPOMF detection. An analytical formula has been derived for spectrally white content and watermarks. For the more realistic case of auto-regressive content and watermarks no such an analytical formula has been found. However, experiments with correlated sources and white watermarks suggest that an analytical expression may exist. This will be addressed in future work.

6. REFERENCES

[1] Matt. L. Miller, Ingemar J. Cox, and Jeffrey A. Bloom, "Watermarking in the real world: An application to dvd," in *Proceedings of Workshop on Multimedia & Security, ACM Multimedia'98*, Bristol (UK), Sept. 1998.

[2] T. Kalker, "Digital video watermarking," in *Proceedings of the ICMCS 99*, Florence (Italy), 1999.

[3] Pascale Termont, Lieven De Strycker, Jan Vandewege, Jaap Haitsma, Ton Kalker, Maurice Maes, Geert Depovere, Andreas Langell, Claes Alm, and Per Norman, "Performance measurements of a real-time digital watermarking system for broadcast monitoring," in *Proceedings of the ICMCS 99*, Florence (Italy), 1999.

[4] Ton Kalker, Geert Depovere, Jaap Haitsma, and Maurice Maes, "A video watermarking system for broadcast monitoring," in *Proceedings of IS&T/SPIEE/EI25 1999*, June 1999.

[5] M.J.J. Maes, A. Kalker, J. Haitsma, and G. Depovere, "Exploiting shift invariance to obtain a high payload in digital watermarking," in *Proceedings of the ICMCS 99*, Florence (Italy), 1999, vol. I, pp. 7–12.

[6] Geert Depovere, Ton Kalker, and Jean-Paul Linnartz, "Improved watermark detection using filtering before correlation," in *Proceedings of the ICIP*, Chicago, Oct. 1998, vol. I, pp. 430 – 434.

[7] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, , no. 24, pp. 325–376, Dec. 1992.

[8] W. Gr obner and N. Hfreiter, *Integraltafel, 2. Teil, Bestimte Integrale*, Springer, Wien, 3. auflage edition, 1961.

[9] I.S. Gradshteyn and I.M. Rydzik, *Table of Integrals, Series and Products*, Academic Press, 1980.