DETECTING ELECTRONIC WATERMARKS IN DIGITAL VIDEO

Jean-Paul Linnartz, Ton Kalker and Jaap Haitsma

Philips Research
Nat.Lab. WY 8, Holstlaan 4,
5656 AA Eindhoven, The Netherlands

ABSTRACT

Electronic watermarking is an active area of research with many applications being foreseen. Watermarks may become an essential tool for copy management in future Consumer Electronic or PC devices. With simple circuits, detection of watermarks after noise addition, MPEG compression, D/A conversion, pixel shifts appears feasible, but detection after transformations, such as cropping and stretching, remains a challenge. We propose a model to evaluate the effect of scaling on the detector reliability and verify it with experiments.

1. INTRODUCTION

Digital multimedia technology opens many opportunities for new applications and services, but content owners are afraid to loose revenues as copies of digital content can be generated rapidly, perfectly, at large scale and without limitations to the number of generations of copies. Copy management came on the "critical time path" of the market introduction of several digital products, including the Digital Video Broadcasting (DVB), Digital Versatile Disc (DVD), HD television, the IEEE 1394 digital interface and improved digital audio carriers (e.g. Super-Audio CD). Tools for copy protection in the digital world are sought in two directions: cryptography and embedded signalling. The latter method, also called "watermarking" inserts copy-data into the copy-restricted video sequence itself. It is intended to prevent illegal copying by telling a compliant device not to copy it. Hence, the watermark should survive MPEG-2 compression and digital-to-analog-to-digital conversions. If the video fidelity remains high, then the watermark should remain detectable. It can also reduce the value of illegal copies by preventing them from being *played* on compliant devices. This means that consumers will have a choice between *a*) compliant devices, which can play legal, commercially released discs that were encrypted, but cannot play pirated discs, and *b*) non-compliant devices, which can play pirated material, but cannot play encrypted discs.

The Copy Protection Technical Working Group (CPTWG) of DVD initiated the Data Hiding SubGroup (DHSG), which released a call for proposals in May 1997. All possible (video)-content should fall into one of four categories: Free Copy, Copy Never, One Copy Allowed and Copy No More. Since watermark detectors must be built into millions of low-cost, consumer devices, a critical requirement is that the detector must be extremely simple and cheap, say implementable in "a few kilogates". These requirements are challenging design specifications.

In the process of determining a watermark detection standard for enhanced DVD copy protection, the "copy-once" and "conditional playback" became important requirements. This has lead to the development of methods to signal and dynamically modify the copy state of watermarked content. We reviewed generation management in [1].

This paper addresses some issues in the detection of watermarks. Section 2 formulates a mathematical model. Section 3 and 4 summarise earlier results on detector error rates. Section 5 addresses new results on detector behaviour for stretched video. Section 6 concludes this paper.

2. FORMULATION OF A MODEL

We consider two stochastical processes: W which generates watermarks and P which generates images. The image and watermark are random vectors of size N_1 by N_2 and M_1 by M_2, respectively. The intensity level of the pixel with coordinates $\boldsymbol{n} = (n_1; n_2)$, ($n_1 \in \{1, 2, .. N_1\}$ $n_2 \in \{1, 2, .. N_2\}$) is denoted as $p(\boldsymbol{n})$. We restrict our discussion to gray scale images, in which $p(\boldsymbol{n})$ takes on real or integer values in a certain interval. In previous publications [3, 4, 5], we found that a simplified theoretical model agrees well with experiments on real-world images, such as "Lenna". The model assumptions included:

- The stochastic processes W and P are mutually independent.
- W and P are wide-sense stationary. So the autocorrelation function $R_{p;p}(\boldsymbol{n};\boldsymbol{m})=\mathrm{E}[p(\boldsymbol{n})p(\boldsymbol{m})]$ becomes $R_{p;p}(\boldsymbol{n}-\boldsymbol{m})$. We denote the spatial offset $\boldsymbol{f} = \boldsymbol{n} - \boldsymbol{m}$.
- W and P are ergodic. So we approximate the statistical autocorrelation function $R_{p;p}(\boldsymbol{f})$ by the spatial

0-7803-5041-3/99 $10.00 © 1999 IEEE

autocorrelation function $\Gamma_{p;p}(f) = (1/N_1N_2)\,\Sigma_n p(n)p(n-f)$ where we assume $n+f$ to wrap around when it formally falls outside the image.

- Images are modeled by assuming a first order separable autocorrelation function $R_{p;p}(f)=m_1^2 + \sigma^2\alpha^{|f|}$ where the i-th moment is $m_i = Ep^i(n)$. The standard deviation σ is found from $\sigma^2 = m_2 - m_1^2$. We define distances as the sum of horizontal and vertical components.
- In the watermark detector all signals have been processed by subtracting the DC-component ($m_1 = 0$).

This model seems a crude approximation of the typical properties of images. However, from experiments, it appeared that error rates based on this crude model can be reasonably accurate for the purpose of this evaluation.

2.1 Watermark Model

A watermark $w(n)$ is modeled as a sample drawn from the stochastical process W. The energy E_w in a watermark w equals $E_w = \Sigma_n w^2(n)$. To avoid complicating side-effects deteriorating the detection reliability [3], the watermark is assumed to be DC-free, i.e., $\Sigma_n w(n) = 0$. We found that it is *not* sufficient to require only that $Ew(n) = 0$ for all n. A "white" watermark has statistically independent luminance values in any two (non-equal) locations n. The autocorrelation is a discrete dirac function. Low-pass watermarks, which we will use in section 4–6, are generated by spatially filtering a white watermark with a first-order two-dimensional spatial smoothing IIR filter.

An often used embedding method is to create a watermarked image $q(n)$ with $q(n) = p(n) + a(n)\,w((n_1 \bmod M_1+1;\, n_2 \bmod M_2+1))$. Here, the embedding depth $a(n)$ is a function of the image properties p, usually in a small area surrounding n.

2.2 Detection

Correlation detectors are a mathematical generalization of the basic method covered in several early papers (e.g. [2]). Another reason to address correlators (in particular "matched filters") is that these are known to be the *optimum* detector for a commonly encountered situation in radio communications, namely the Linear Time-Invariant (LTI), frequency non-dispersive, Additive Gaussian Noise (AWGN) channel, when the receiver has full knowledge about the alphabet of waveforms used to transmit messages [5]. Less ideal situations often are addressed with appropriate modifications to the matched filter, e.g. by adding a whitening filter, as in Figure 1.

In the correlator detector, a set of decision variables $d(\varepsilon)$ are extracted from the received image $r(n)$ by correlating with a spatially shifted version of the locally stored copy of the watermark $w(n)$, i.e., $d(\varepsilon) = \Sigma_n\, r(n)\, w(n+\varepsilon)$, where ε reflects the shift searched by the detector. If whitening is applied prior to correlation, both $r(n)$ and $w(n)$ are spatially (high-pass) filtered with the same filter [4], such that the image component in $r(n)$ becomes spectrally white.

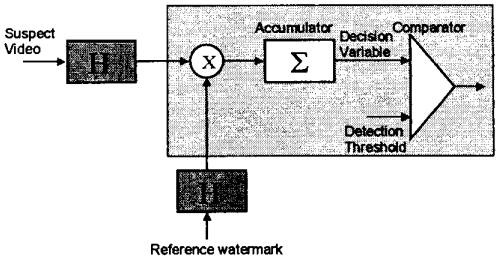

Figure 1: Whitened matched filter for watermark detector.

3. False Negative

For a properly designed watermarking tool, the probability that a detector erroneously does not see a watermark is negligibly small if the marked video has not been modified. More interesting is the question of how the detection is affected by processing of the marked video. Section 5 will focus on a critical operation, namely scaling.

4. False Positives

We derived in [3, 4] that for a detection threshold at 50% of the correlation value of a pure watermark, the false positive probability can be expressed as

$$P_{fp} = \frac{1}{2}\,erfc\left(\sqrt{\frac{E_w}{8\sigma^2}}\,\frac{1-\alpha\beta}{1+\alpha\beta}\right)$$

where σ is the standard deviation of the luminance of a pixel in the image. Interestingly, the size of the image (N_1, N_2) does not appear in the above expression, although it determines the energy E_w that can be accommodated in the image. It appears a low-pass nature if image and watermark ($\alpha > 0$, $\beta > 0$) reduces reliability.

Using a first order 2-dimensional filter to whiten the image, the false positive rate is [4]

$$P_{fp} = \frac{1}{2}\,erfc\left(\sqrt{\frac{E_w}{8\sigma^2}}\,\frac{1-2\alpha\beta+\alpha^2}{1-\alpha^2}\right)$$

Experiencing a typical α of 0.9 .. 0.95, we have verified experimentally [4] that whitening can give a 10 to 20 dB improvement in reliability of the detection.

5. Scaling

We express the value of the correlation peak seen by the detector if the video is stretched and cropped. We model the decision variable (which is typically calculated as a *sum* over pixels) as an *integration* over the image. That is, we address a continuous spatial representation, as for instance on a scan line of analog NTSC television. This assumption conflicts with the discrete pixel locations implemented in the digital practice of today, but it avoids that we must consider any details of how the stretching and scaling of the video has occurred before it arrived at the detector. Such scaling may involve nearest neighbor selection of pixels, linear or bicubical interpolation, or band pass filtering; in fact for time being its precise processing method is less of interest to us.

For simplicity of analysis we now restrict ourselves to the one-dimension case. Assume that the original image takes on luminance values at (real-valued) locations between 0 and N_1. This image is stretched to size 0 and $N_1 + \gamma$, and cropped to recover the original size 0 .. N_1. Such stretching can be interpreted as a linearly decreasing pixel shift. Assume that due to scaling, the "shift" f of the watermark linearly decreases from $f(0) = f_1 = \gamma_1 > 0$ to $f(N_1) = f_2 = -\gamma_2 < 0$. Here γ_1 is the number of pixels removed at left hand size of the image and γ_2 the number of pixels removed from the right hand side, and $\gamma_1 + \gamma_2 = \gamma$. In the following analysis it is not critical whether the shift increases from left to right (shrinking) or in decreases (stretching).

A correlator detector determines

$$d(\varepsilon)=\frac{1}{N_1}\int_o^{N_1}\left[p\left((1-\frac{\gamma}{N_1})s+\gamma_1\right)+\alpha(.)w\left((1-\frac{\gamma}{N_1})s+\gamma_1\right)\right]w(s+\varepsilon)ds$$

Assuming unity average embeding depth (Eε = 1), the expected value of the watermark component in the decision variable is

$$h(\varepsilon)=\frac{1}{N_1}\mathrm{E}\int_o^{N_1}w\left((1-\frac{\gamma}{N_1})s+\gamma_1\right)w(s+\varepsilon)ds=$$

$$=\frac{1}{N_1}\int_o^{N_1}R\left(\varepsilon-\gamma_1+\frac{\gamma}{N_1}s\right)ds$$

Replacing $f = (s\gamma)/N_1 - \gamma/2$, we see that the correlation value becomes

$$h(\varepsilon)=\frac{1}{\gamma}\int_{f_1}^{f_2}R(f-\varepsilon)df$$

for $f_2 \neq f_1$ and $R(0)$ for $f_2 = f_1$. We denote $\gamma = \Delta_f = f_2 - f_1$. Intuitively one expects the largest correlation peak to occur for $\varepsilon_0 = (f_2 + f_1)/2$. This can be verified for an autocorrelation function with negative derivative $R'(f) \leq 0$ for $f > 0$. Let's assume that a correlation peak occurs at ε_x ($\varepsilon_x \neq \varepsilon_0$). It is easily seen that $h(\varepsilon_x) < h(\varepsilon_0)$ if ε_x is outside the interval (f_1, f_2). For $f_1 < \varepsilon_x < f_2$, we write

$$h(\varepsilon_x)=\frac{1}{\gamma}\left[\int_{f_1}^{\varepsilon_x}R(f-\varepsilon_x)df+\int_{\varepsilon_x}^{f_2}R(f-\varepsilon_x)df\right]$$

We obtain the maximum by requiring that the derivative equals zero for f_x. We get

$$\frac{d}{d\varepsilon_x}\int_0^{\varepsilon_x-f_1}R(f)df+\frac{d}{d\varepsilon_x}\int_0^{f_2-\varepsilon_x}R(f)df=0$$

or $R(\varepsilon_x - f_1) = R(f_2 - \varepsilon_x)$ thus $\varepsilon_x = (f_2 + f_1)/2$. The largest correlation value occurs for optimum shift $\varepsilon = \gamma/2$, for which

$$h(\frac{\gamma}{2})=\frac{2}{\gamma}\int_o^{\frac{\gamma}{2}}R(f)df$$

We now apply a watermark generated using a random number generator, first-order low-pass filtered to having an autocorrelation function $R(s) = \beta^{-|s|} = \exp(-c|s|)$, where $1/c$ is the correlation distance of the watermark expressed in pixels. The correlation between two adjacent watermark pixels is e^{-c} which equals 0.37 in our examples using $c = 1$.

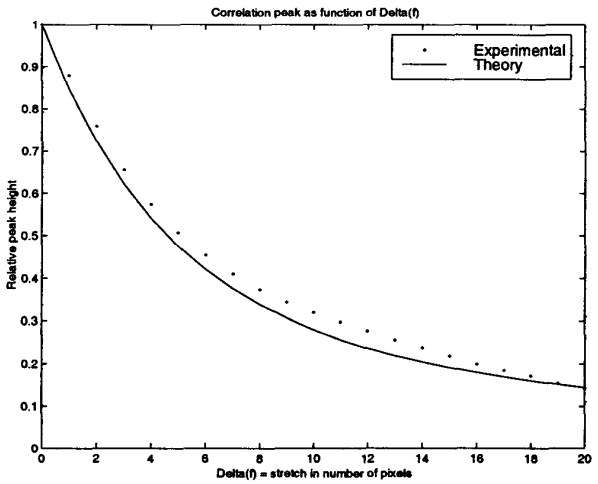

Figure: Detected correlation versus shift offset for a low-pass watermark (c = 1)

We now evaluate $h(\varepsilon_0)$ for a watermark that has a first-order low-pass spectrum, i.e., with $R(f) = \exp(-c|f|)$. The largest correlation peak has amplitude

$$h(\varepsilon_0) = \frac{2}{c\gamma}\left[1-\exp\left(-\frac{c\gamma}{2}\right)\right]$$

For $\gamma \rightarrow 0$, $h(\varepsilon_0)$ goes to one. For large γ, $h(\varepsilon_0)$ reduces to zero. A first order expansion for small $\gamma > 0$, gives $h(\varepsilon_0) = 1 - c\gamma/2$.

In the above expressions, $c\gamma$ appears as a product, but we do not encounter one of these parameters appearing separately. Interestingly several other parameters, including the image size N_1 and N_2 do *not* appear in the expression. We conclude that the correlation peak does not depend on the size of the image or watermark pattern, but only on the amount of stretching, relative to the autocorrelation in the watermark. Figure 1 compares the theoretical results with an experiment.

To find the shape of the correlation peak after stretching, we evaluate $h(\varepsilon)$ for arbitrary ε. We find

$$h(\varepsilon) = \frac{1}{c\gamma}\left[\begin{array}{l}\frac{1}{2}(\mathrm{sgn}(\varepsilon - f_2) - 1)(\mathrm{sgn}(f_1 - \varepsilon) - 1) + \\ \mathrm{sgn}(f_1 - \varepsilon)\exp(-c|f_1 - \varepsilon|) + \mathrm{sgn}(\varepsilon - f_2)\exp(-c|f_2 - \varepsilon|)\end{array}\right]$$

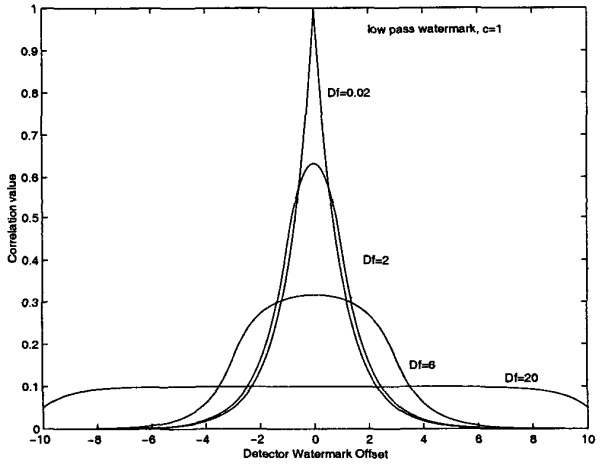

Figure 3: Detected correlation $h(\varepsilon)$ versus shift offset ε for a low-pass watermark ($c = 1$). $E_w = 1$. Df denotes the stretching γ in pixels.

Figure 3 illustrates $h(\varepsilon)$ versus $\varepsilon - f_0$, for $c = 1$ and various amounts of horizontal stretching $\gamma = \Delta_f$. For nearly white watermarks ($\beta << 1$, $c >> 1$), the effect smoothing due to of pixel interpolation should also be considered. Nearest neighbor interpolation gives a rectangular impulse spatial response, and a corresponding triangular autocorrelation function $R(f) = 1 + f$ for $-1 < f < 0$, $R(f) = 1 - f$ for $0 < f < 1$ and $R(f) = 0$ otherwise. Then $h(f_0) = 1 - \gamma/4$ for $0 < \gamma < 2$ and $1/\gamma$ for $\gamma > 2$. Presumably the smoothing associated with pixel interpolation explains why our experimental results in Figure 2 are slightly larger than predicted from the considered autocorrelation.

6. CONCLUSIONS

This paper reviewed and summarized some theoretical results derived at Philips Nat.Lab. for the reliability of watermark detection. Motivation for the formulation of a model has been the need to quantify false positive rates (false alarms) of a watermark detector used in CE and PC equipment to detect and prevent playback of illegally copied video. The rate of occurrence of such false positives need to be kept below (or not substantially larger than) typical product failures. Experimental verification of rates less that say 10^{-12} would not be practical. Theoretical models are needed. We showed that it is not correct to a neglect correlation of image pixels. Even though such assumptions are often made, these tend to substantially underestimate false positive rates, particularly if the watermark also has a low-pass nature.

New results on scaling have also been described. Scaling results in a smearing of the correlation peak. A model has been proposed and verified with experimental results. It appeared that the reduction of the height of the correlation depends primarily on the absolute amount of scaling ("pixels" not "percents") and the spectrum of the watermark, but it is insensitive to the image size. Results can be used for instance to determine an appropriate step size for searching the factor to which video has been stretched.

7. ACKNOWLEDGEMENTS

The authors thank Joop Talstra for fruitful discussions.

8. REFERENCES

[1] J.P.M.G. Linnartz, "The ticket concept for copy control based on embedded signalling", ESORICS '98, Louvain-La-Neuve, September 1998, Lecture Notes in Computer Science, 1485, Springer, pp. 257-274.

[2] W. Bender, D. Gruhl, N. Morimoto, "Techniques for Data Hiding", Proceedings of the SPIE, 2420:40, San Jose CA, USA, February 1995

[3] J.P.M.G. Linnartz, A.C.C. Kalker, and G.F. Depovere, "Modelling the false-alarm and missed detection rate for electronic watermarks". Workshop on Information Hiding, Portland, OR, 15-17 April, 1998. Springer Lecture Notes on Computer Science, No. 1525, pp. 258-272, pp. 329-343.

[4] G.F.G. Depovere, A.C.C Kalker, and J.P.M.G. Linnartz, "Improved watermark detection reliability using filtering before correlation", Int. Conf. on Image Processing, ICIP, October 1998, Chicago IL.

[5] www.eecs.berkeley.edu/~linnartz