

From the Proceedings of the
Third International Workshop on Information Hiding, Dresden, Germany, 1999.

© 1999 Springer-Verlag.

This paper has been published by Springer-Verlag as part of the Lecture Notes in
Computer Science series. The LNCS series WWW site can be found at
<http://www.springer.de/comp/lncs/index.html>.

Computing the Probability of False Watermark Detection

Matt L. Miller and Jeffrey A Bloom

Signafy, Inc.* , 4 Independence Way, Princeton, NJ 08540

Abstract. Several methods of watermark detection involve computing a vector from some input media, computing the normalized correlation between that vector and a predefined watermark vector, and comparing the result against a threshold. We show that, if the probability density function of vectors that arise from random, unwatermarked media is a zero-mean, spherical Gaussian, then the probability that such a detector will give a false detection is given exactly by a simple ratio of two definite integrals. This expression depends only on the detection threshold and the dimensionality of the watermark vector.

1 Introduction

Digital data hiding, also referred to as digital watermarking, is the practice of making imperceptible changes in digitized media to hide messages. A basic watermarking system consists of a watermark embedder and a watermark detector (see figure 1). In general, the inputs to the embedder are a message to be hidden and some media in which to hide it, such as an audio stream, an image, or a video stream. The output of the embedder is media that is perceptually very similar to the input media, but contains the input message as a hidden watermark. The input to the detector is some media that may or may not contain the watermark. The detector's output is a judgment about whether or not a watermark is present in its input, along with the message that the watermark contains.

Several applications of such invisible watermarking systems have been suggested. These include identifying the content creator, identifying the content recipient in a transaction, monitoring broadcast and publication channels for tracking purposes, assuring that the media item is an authentic, unmodified copy of the original, and restricting the use of media to avoid illegal duplication. These applications of watermarking are discussed in more detail in [1].

Recently, watermarking systems have begun to be considered for widespread deployment as parts of copy-protection systems. Primarily for protection of video or audio media, watermarks would be used in addition to encryption techniques. Standardization movements that will likely involve watermarking include efforts

* The authors are now with NEC Research Institute, 4 Independence Way, Princeton, NJ, 08540.

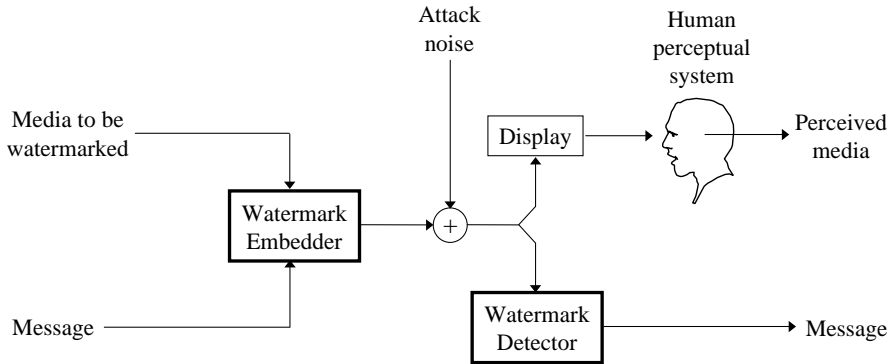


Fig. 1. Typical Watermarking System

by the Copy Protection Technical Working Group to protect DVD video [2], efforts by a group of 4 companies (IBM, Intel, Matsushita, and Toshiba) working through the DVD forum to protect DVD audio, and an effort by the Recording Industry Association of America (RIAA) through its Secure Digital Music Initiative (SDMI) to protect music distributed on the internet [3].

A critical issue for many applications is the probability that the watermark detector incorrectly identifies some unwatermarked media as containing a watermark. Such an error can lead to the mistaken prevention of a legitimate operation. For example, in the case of DVD copy protection, the presence of a watermark on a RAM disk will prevent that disk from being played on a DVD player. If the watermark is really there, this is the correct behavior, since copy protected video that appears on a home-recordable disk must have been copied illegally. But if the detection is false and the video does not really contain a watermark, then it is a legal copy and should have been playable. A false watermark detection could thus prevent people from watching their legitimate home recordings and would be viewed as an equipment malfunction. For this reason, the probability of false detection should be as low as the probability of malfunction in consumer electronics equipment.

The required probability of false detections can be extremely low in some applications. The Consumer Electronics Manufacturer's Association (CEMA) has set the requirement for DVD at roughly one false detection in 10^{10} seconds - roughly 300 years - of unwatermarked video. Such probabilities are too low to be verified by experimentation, so it is necessary to estimate the false detection behavior of a given detector analytically.

The problem of analyzing false detection behavior has received little attention in the literature. Linnartz and Kalker have made an in-depth analysis of the probability of false detection in an image watermarking method [4]. Hernández and Pérez-González include false detection probability in their framework for discussing watermarking systems [5]. Stone analyzed the behavior of watermark detectors that use the correlation coefficient as a detection metric [6] and his

analysis was applied to a specific video watermarking system by McKellips [7].

In the present paper, we analyze the behavior of watermark detectors that use correlation coefficient as their detection metric. In section 2 we give a basic framework for watermark detectors and discuss possible detection metrics that they can use. Section 3 reviews two earlier methods for estimating the probabilities of false detection when the detection metric is correlation coefficient. Section 4 gives our method, which is exact under a single, reasonable assumption. Finally, section 5 shows the results of experimental simulations, and section 6 offers conclusions and discusses future work.

2 Watermark detection

The probability of false watermark detection is determined by the design of the watermark detector and the distribution of media processed by it. The embedding algorithm is not relevant to computing this probability, because we assume that no watermark has been embedded (otherwise a detection would not be false). We therefore begin by describing a generic framework for watermark detection and will not discuss embedding.

Figure 2 shows a basic design for a watermark detector. The input media is first processed by a watermark extractor which generates an n -dimensional extracted vector, V . Examples of watermark extraction include various combinations of spatial registration [8], frequency transforms [9, 10, 11], block averaging [12, 13], spectral shaping [14, 15], whitening [16], and subtraction of the unwatermarked original media [9, 17]. Reasons for applying these processes include the following: increasing robustness to certain common types of attack, increasing signal-to-noise ratio, reducing the size of data to make detection cheaper, and generating vectors that are well distributed for standard detection measures.

The extracted vector is then compared against an n -dimensional watermark vector W , to obtain a detection measure C . In our analysis we assume that the detector uses a single watermark vector, which is a predefined constant. When used in a typical application, a watermark detector may check the extracted vector against multiple watermark vectors and decode the resulting detection measures into one of several possible messages. The effect of this decoding process on the probability of a false detection is usually a straightforward function of the probability of obtaining a false detection with only one watermark vector. Thus, when we limit our analysis to a single watermark vector, we obtain a result that is widely useful for computing false detection rates of real detectors with multiple watermark vectors.

Finally, the detection measure is compared against a threshold T , and the result of this comparison determines whether the detector reports that the watermark is present or not. Generally, larger C indicates greater probability that the watermark is present, so the detector reports a detection if $C > T$.

The exact formula used for computing the detection measure is critical to determining the probability of false detection. Three types of measure are correlation, correlation coefficient, and a slight variant of correlation coefficient that

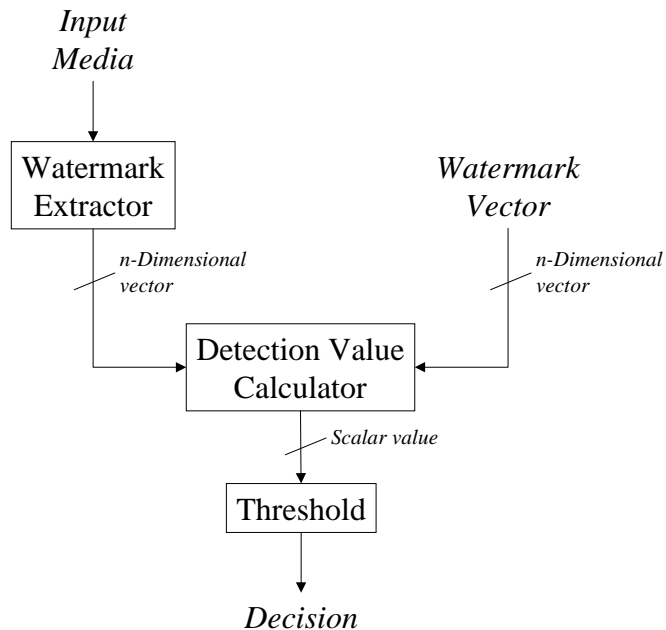


Fig. 2. Basic design for a watermark detector

we refer to as simply normalized correlation.

Correlation: The correlation measure is computed as

$$C_c = W \cdot V = \sum_i w_i v_i. \quad (1)$$

This type of detection measure is known to be optimal for certain types of communications channels, which suggests its use for watermark detection. However, it is argued in [5] and [18] that a watermark in digital media behaves very differently from those communications channels for which correlation is optimal. Furthermore, if we are to calculate the probability of false detection resulting from the use of correlation, we need to know the variance of the elements of vectors extracted from random media. This variance is difficult to obtain with enough precision to guarantee the kinds of extremely low probabilities required in many applications.

The probability of false detections resulting from an image watermark detector using correlation is analyzed in detail in [4]. This detection measure will not be considered further in the present paper.

Correlation Coefficient: To remove dependence on the variance of extracted vector elements, we can use the correlation coefficient, which is defined as

$$C_{cc} = \frac{\tilde{W} \cdot \tilde{V}}{\sqrt{(\tilde{W} \cdot \tilde{W})(\tilde{V} \cdot \tilde{V})}} \quad (2)$$

where

$$\tilde{W} = W - \bar{W}, \quad (3)$$

$$\tilde{V} = V - \bar{V}. \quad (4)$$

This differs from correlation by subtracting out the means of V and W , and dividing by the product of their standard deviations. The latter difference makes a qualitative change in the behavior of detection values which, as will be shown in section 4, causes the false detection probability to be independent of the variances of the extracted watermark elements.

Normalized Correlation: A detection measure that is simpler than correlation coefficient is obtained by not subtracting out the means of V and W . This yields

$$C_{nc} = \frac{W \cdot V}{\sqrt{(W \cdot W)(V \cdot V)}} \quad (5)$$

We refer to this measure as *normalized correlation*.

If we restrict W to have variance of exactly 1, then this detection metric is equivalent to the one used in [9], namely

$$C'_{nc} = \frac{W \cdot V}{\sqrt{V \cdot V}}. \quad (6)$$

Since $W \cdot W$ is a constant n , these two metrics differ only by a constant factor of \sqrt{n} and their behaviors are the same.

For the sake of simplicity, the analysis presented in section 4 concentrates only on normalized correlation. However, we can apply the same analysis to correlation coefficient by observing that subtracting the means from V and W amounts to projecting them into an $(n-1)$ -dimensional subspace. The correlation coefficient is then just the normalized correlation, computed in this subspace. Thus, the analysis of section 4 will apply equally well to both normalized correlation and correlation coefficient, with the proviso that the dimensionality, n , be replaced with $n - 1$ when computing the probability of false detections using correlation coefficient.

3 Approximate methods of computing probability of false detections

Two approximate methods have been used in the past to estimate the probability of false detections when using normalized correlation or correlation coefficient as a detection measure. Both of these rely on the assumption that the elements of watermarks extracted from unwatermarked media are drawn from identical, independent distributions.

3.1 Approximate Gaussian method

The simplest method, which we will refer to as the approximate Gaussian method, is to treat the distribution of normalized correlations as a Gaussian with standard deviation $1/\sqrt{n}$. If n is large, then the body of the distribution is very similar to a Gaussian. Thus, we can approximate the false detection probability as

$$P_{fd} \approx \operatorname{erfc}(T\sqrt{n}) = \int_{T\sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \quad (7)$$

For relatively low thresholds, and relatively high false detection probabilities, this approximation can be quite accurate. However, as the threshold increases out into the tails of the distribution, we begin to dramatically overestimate the probability of false detections. This is clear from the fact that the normalized correlation is bounded between plus and minus 1. Thus, if T is higher than 1, the true probability of a false detection (or, indeed, any detection at all) is zero. But the approximation based on assuming a Gaussian distribution yields a non-zero probability.

3.2 Fisher Z-statistic method

A more accurate approximation is described in [6], and employed in [7]. This relies on the Fisher Z-statistic [19], computed from the correlation coefficient as

$$Z = \frac{1}{2} \log \frac{1+C}{1-C}. \quad (8)$$

When elements of V are drawn from identical, independent, Gaussian distributions, the values of Z form an approximately Gaussian distribution with standard deviation¹ $1/\sqrt{n-2}$. If we assume the distribution of Z is truly Gaussian, then we can estimate the probability of false detection by

$$P_{fd} \approx \operatorname{erfc}(T_Z \sqrt{n-2}), \quad (9)$$

¹ In [6], the standard deviation was given as $1/\sqrt{n-3}$. The difference is because Stone was analyzing correlation coefficient, which has one fewer dimension than normalized correlation.

where

$$T_Z = \frac{1}{2} \log \frac{1+T}{1-T}. \quad (10)$$

This approximation is much more accurate than the approximate Gaussian method. However, as will be shown in section 5, it begins to underestimate the probability of false detections as the threshold approaches one.

4 Exact method of computing probability of false detections

In this section, we develop a new method for computing the probability of false detections when using normalized correlation or correlation coefficient. We begin with a single assumption about the distribution of vectors extracted from random, unwatermarked media:

Assumption: the probability density function (PDF) of vectors extracted from unwatermarked media is radially symmetric.

In other words, we assume that the relative likelihood of a given vector being extracted from unwatermarked media is dependent only on the vector's length, and independent of the vector's direction.

While the above assumption is true for many distributions, the most plausible such distribution in a watermark detector is the one that arises when the elements of extracted vectors are drawn from identical, zero-mean, independent, Gaussian distributions. This yields an n -spherical Gaussian. Thus, the above assumption can be seen as essentially the same as the assumption that underlies the Z-statistic method of approximating probability of false detections.

Next, we simplify the PDF by scaling the extracted vectors to unit length before computing the detection measure. The scaled, extracted vector will be denoted V' and is defined as

$$V' = \frac{V}{\sqrt{V \cdot V}}. \quad (11)$$

Observe that V' yields a detection if, and only if, V yields a detection, since

$$C_{nc} = \frac{W \cdot V'}{\sqrt{(W \cdot W)(V' \cdot V')}} = \frac{W \cdot V'}{\sqrt{W \cdot W}} = \frac{W \cdot V}{\sqrt{(W \cdot W)(V \cdot V)}}. \quad (12)$$

So the probability of false detections is unaffected by the introduction of this scaling step. Every V' lies on the unit n -sphere and has the same likelihood of occurring as any other V' because of the radial symmetry of the PDF for V .

Now, we turn our attention to a geometric interpretation of the detection region defined by a given watermark vector W , and a given threshold T . Since

$$\frac{W \cdot V'}{\sqrt{(W \cdot W)(V' \cdot V')}} = \cos \alpha, \quad (13)$$

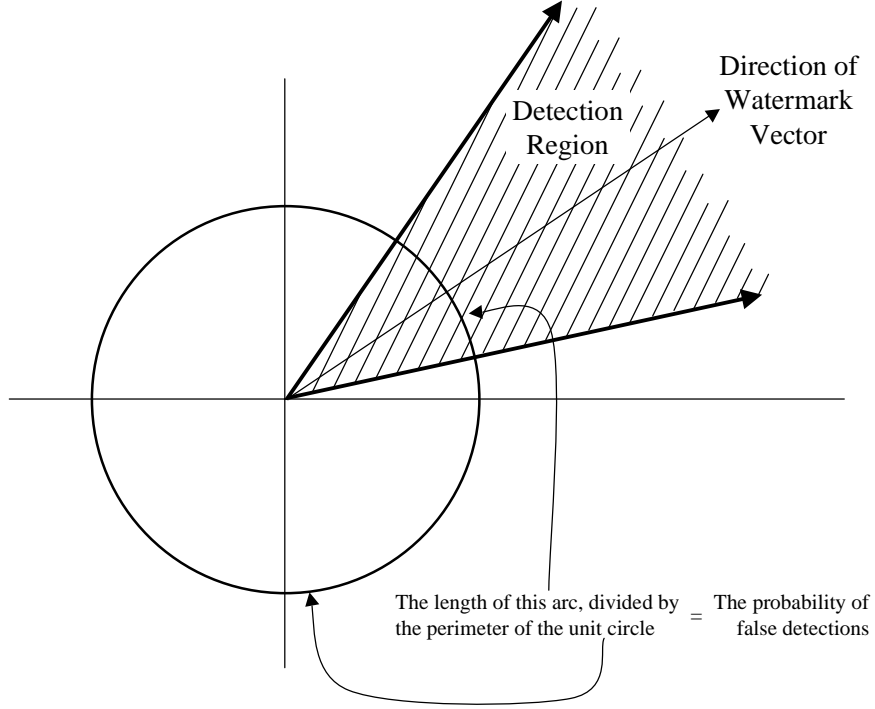


Fig. 3. Two dimensional detection region

where α is the angle between W and V' , we can replace the threshold on normalized correlation between W and V' with a threshold on the angle between them:

$$C > T \iff \alpha < T_\alpha, \quad (14)$$

where

$$T_\alpha = \cos^{-1} T. \quad (15)$$

This means that the detection region is an n -cone that subtends an angle of $2T_\alpha$. Figure 3 illustrates the 2-dimensional case. It shows the unit 2-sphere (circle), upon which V' is uniformly distributed, together with the detection region for a given watermark vector and T . Figure 4 is the same diagram for the 3-dimensional case.

The probability of false detections is given by dividing the $(n-1)$ -content of the unit n -sphere's surface into the $(n-1)$ -content of the intersection of the surface with the detection region. The following derivation of the resulting equation is obtained from [20].

The $(n-1)$ -content of the intersection between an n -cone and the surface of an n -sphere is given by

$$\text{Cap}(n, \theta) = S_{n-1} I_{n-2}(\theta) \quad (16)$$

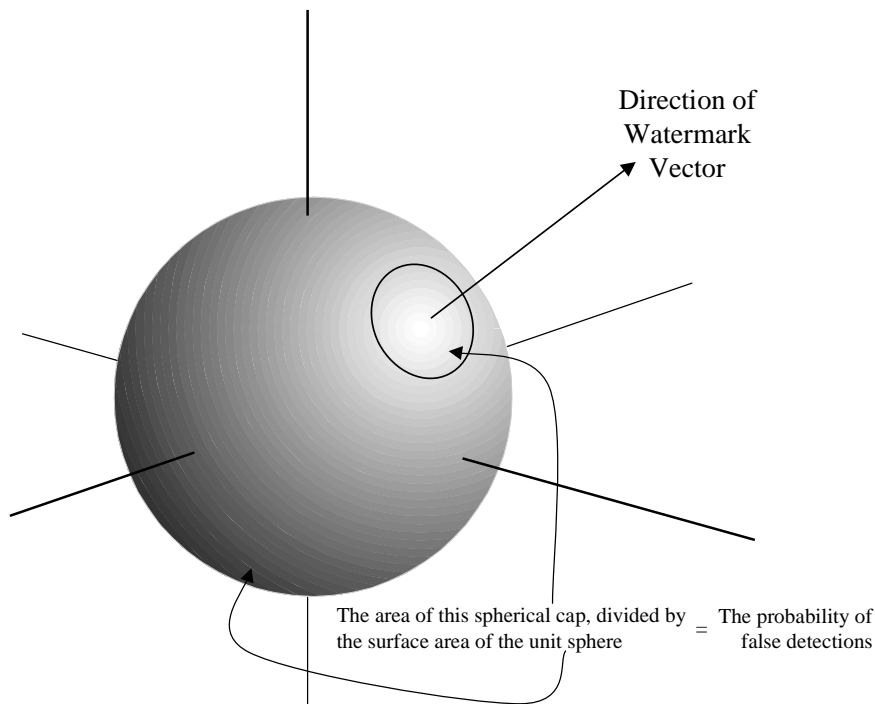


Fig. 4. Three dimensional detection region

where θ is half the angle subtended by the n -cone, and where

$$S_d = \frac{d\pi^{\lfloor d/2 \rfloor}}{\lfloor d/2 \rfloor!} \quad (17)$$

and

$$I_d(\theta) = \int_0^\theta \sin^d(u) du \quad (18)$$

for any d .

Note that $\text{Cap}(n, \frac{\pi}{2})$ is half the $(n-1)$ -content of the surface of the unit n -sphere. Thus, the ratio of the portion of the surface that is within the detection region to the whole surface and hence, the probability of a false detection, is given by

$$P_{fd} = \frac{\text{Cap}(n, T_\alpha)}{2\text{Cap}(n, \frac{\pi}{2})} = \frac{I_{n-2}(T_\alpha)}{2I_{n-2}(\frac{\pi}{2})} \quad (19)$$

where T_α is as defined in equation 15.

To compute this probability, we need to evaluate the function $I_d(\theta)$. Table 1 shows the closed-form solutions for $d = 1$ through $d = 5$. Values of $I_d(\theta)$ for $d > 5$ can be computed by means of the recursive formula at the bottom of this table.

d	$I_d(\theta)$
0	θ
1	$1 - \cos(\theta)$
2	$\frac{\theta - \sin(\theta) \cos(\theta)}{2}$
3	$\frac{\cos^3(\theta) - 3 \cos(\theta) + 2}{3}$
4	$\frac{3\theta - (3 \sin(\theta) + 2 \sin^3(\theta)) \cos(\theta)}{8}$
5	$\frac{4 \cos^3(\theta) - (3 \sin^4(\theta) + 12) \cos(\theta) + 8}{15}$
≥ 6	$\frac{d-1}{d} I_{d-2}(\theta) - \frac{\cos(\theta) \sin^{d-1}(\theta)}{d}$

Table 1. Closed-form solutions for $I_d(\theta)$

5 Results

To verify that the formula derived in section 4 is correct, we compared its predictions against results obtained from over a billion synthetic vectors drawn from a radially symmetric, pseudorandom distribution. Each vector had 10 elements ($n = 10$). This dimensionality gives high enough false detection rates that we can obtain reasonable statistics from a billion trials. It also shows the differences between predictions made using the new method and using the two approximate methods of section 3. Each element of a vector was generated using a pseudorandom number generator designed to produce normal distributions [21] with 0 mean and unit variance.

Figure 5 shows the results of our experiment compared against the predictions made by the new method and by the two approximate methods of section 3. The new method’s predictions match very closely with the experimental results, while the approximate methods quickly deviate from them.

In figures 6 and 7, we explore the relationship between the new method and the two approximate methods for vectors with more elements. It is clear that, at higher thresholds, the approximate Gaussian method tends to overestimate the false detection probabilities, while the Z-statistic method tends to underestimate false detection probabilities, though providing closer approximations to the exact result obtained with the new method. However, when the threshold is low, all three methods give very similar results, regardless of the dimensionality of the vectors. Thus, the approximate methods may be acceptable when a low threshold is predicted to yield the desired false detection rate, either because the desired rate is relatively high, or because the vectors are long. The new method is to be preferred when the detection threshold must be set at a higher level.

6 Conclusion and future work

In this paper, we have presented a formula (equation 19) for the probability that a watermark detector will report false detections, if the watermark detector uses normalized correlation as its detection measure. This formula holds exactly if the

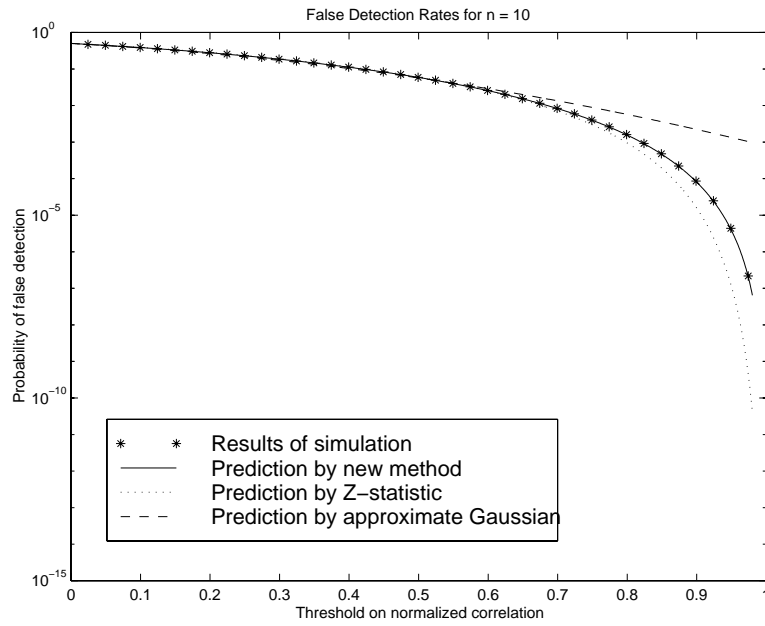


Fig. 5. False detection rates for various thresholds, watermark vector length 10

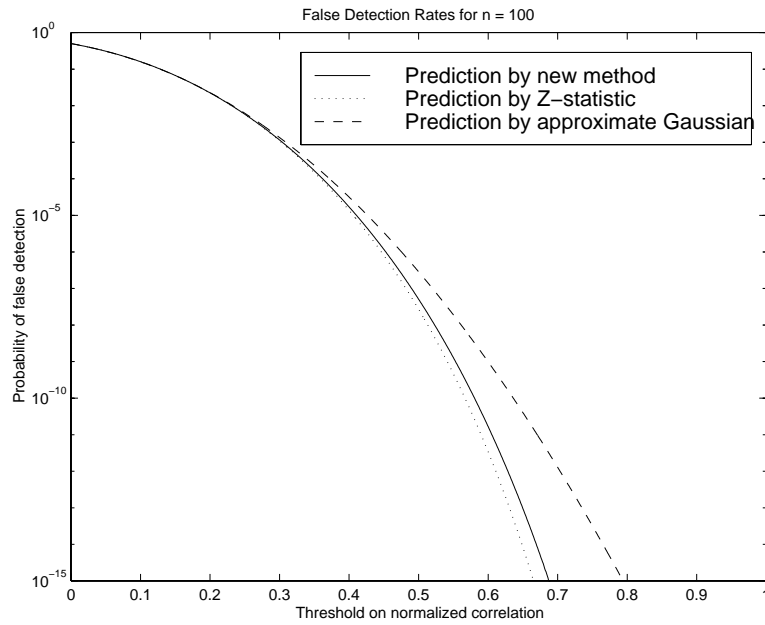


Fig. 6. False detection rates for various thresholds, watermark vector length 100

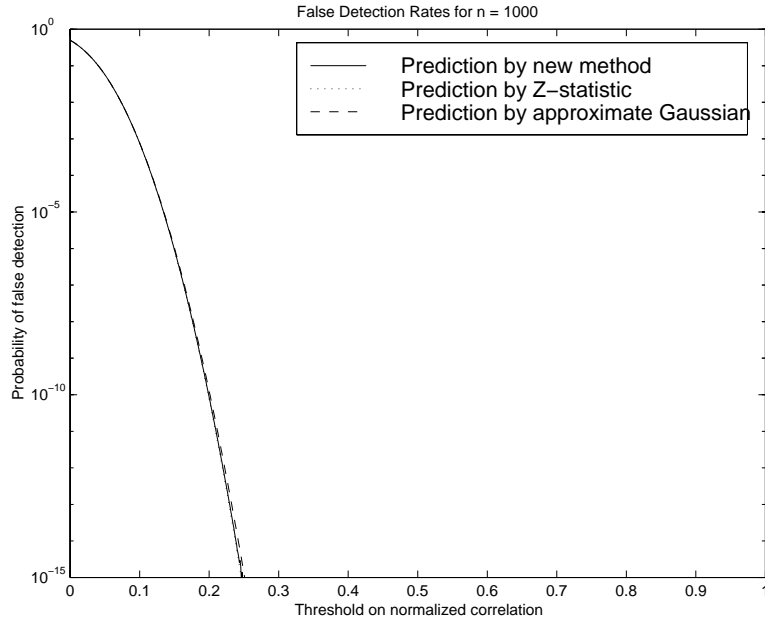


Fig. 7. False detection rates for various thresholds, watermark vector length 1000

distribution of vectors extracted from unwatermarked data is radially symmetric, which is the case when the elements of the vectors are drawn from independent, identical, zero-mean Gaussian distributions.

Of course, in an actual application, the distribution of extracted vectors is unlikely to be perfectly symmetric. Thus, the next problems to be studied are the effects of other types of distributions on the probability of false detections. In particular, we expect that real distributions can differ from symmetric distributions in the following ways:

- The elements of extracted vectors can be correlated.
- In a watermark detector that uses integer arithmetic, the elements of the vectors are quantized. This results in a joint probability distribution that is quantized to a rectilinear grid.
- In most systems, the possible values in extracted watermarks are bounded. This means that the probability distribution is bounded by an n -cube.

We plan to address these questions in the future.

Acknowledgements

The authors are grateful to Dr. Warren Smith and Dr. Harold Stone for their helpful discussions and insights.

References

1. I.J. Cox, M. Miller, J-P. Linnartz, and T. Kalker. A review of watermarking principles and practices. In K.K. Parhi and T. Nishitani, editors, *Digital Signal Processing for Multimedia Systems*, chapter 17. Marcel Dekker, Inc., 1999.
2. J. A. Bloom, I. J. Cox, T. Kalker, J-P. Linnartz, M. L. Miller, and B. Traw. Copy protection for DVD video. *Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information*, 87:1267–1276, 1999.
3. Secure digital music initiative. <http://www.riaa.com/tech/sdmiinfo.htm>.
4. J-P. Linnartz, T. Kalker, and G. Depovere. Modelling the false alarm and missed detection rate for electronic watermarks. *Proc. Second International Workshop on Information Hiding*, pages 329–343, 1998.
5. Juan Ramón Hernández and Fernando Pérez-González. Shedding more light on image watermarks. In David Aucsmith, editor, *Information Hiding 1998*, pages 191–207. Springer-Verlag, 1998.
6. Harold S. Stone. Analysis of attacks on image watermarks with randomized coefficients. *Tech. Rep. 96-045, NEC Research Institute, Princeton, NJ*, 1996.
7. Andrew McKellips. Watermark detection false positive analysis. *Signafy Technical Report, TR-118*, 1997.
8. Geoffrey B. Rhoads. Image steganography system featuring perceptually adaptive and globally scalable signal embedding. *U S Patent 5,748,763*, 1998.
9. I.J. Cox, J. Kilian, T. Leighton, and T. Shamoan. Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Proc.*, 6(12):1673–1687, 1997.
10. C. I. Podilchuk and W. Zeng. Image-adaptive watermarking using visual models. *IEEE Trans. on Selected Areas of Communications*, 16(4):525–539, 1998.
11. J. J. K. O’Ruanaidh, W. J. Dowling, and F. Boland. Phase watermarking of digital images. *Proc. ICIP 96*, pages 239–242, 1996.
12. Ton Kalker, Geert Depovere, Jaap Haitzma, and Maurice Maes. A video watermarking system for broadcast monitoring. *Proc. SPIE, 3657 Security and Watermarking of Multimedia Contents:103–112*, 1999.
13. I. J. Cox and K. Tanaka. *NEC data hiding proposal*. Technical report, NEC Copy Protection Technical Working Group. Response to call for proposal issued by the Data Hiding SubGroup. Available at <http://www.dvcc.com/dhsg>, July, 1997.
14. Ingemar J. Cox, Matt L. Miller, Kazuyoshi Tanaka, and Wakasu. Digital watermarking. *U S Patent filed November 5, 1996*, 1996.
15. R. D. Preuss, S. E. Roukos, A. W. F. Huggins, H. Gish, M. A. Bergamo, P. M. Peterson, and D. A. G. Embedded signalling. *U S Patent 5,319,735*, 1994.
16. Geert Depovere, Ton Kalker, and J-P Linnartz. Improved watermark detection reliability using filtering before correlation. *Proc. ICIP 98*, pages 430–434, 1998.
17. M. D. Swanson, B. Zhu, and A. H. Tewfik. Transparent robust image watermarking. *Proc. ICIP 96*, pages 211–214, 1996.
18. Ingemar J. Cox, Matt L. Miller, and Andrew L. McKellips. Watermarking as communications with side information. *Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information*, 87:1127–1141, 1999.
19. H. D. Brunk. *An introduction to mathematical statistics*. Ginn and Company, 1960.
20. Warren D. Smith. *Studies in computational geometry motivated by mesh generation*. Ph.D. Thesis, Applied Mathematics, Princeton University, 1988.
21. J.H. Ahrens and U. Dieter. Extensions of Forsythe’s method for random sampling from the normal distribution. *Math. Comput.*, 27:927–937, 1973.

