

# FALSE POSITIVE ANALYSIS OF CORRELATION RATIO WATERMARK DETECTION MEASURE

*Jun Tian, Jeffrey A. Bloom*

Thomson Corporate Research  
2 Independence Way Suite 300  
Princeton, NJ 08540, USA

*Peter G. Baum*

Thomson Corporate Research  
Karl-Wiechert Allee 74  
D 30625 Hannover, Germany

## ABSTRACT

A critical issue for many watermarking applications is the probability that the watermark detector incorrectly identifies an unwatermarked work as watermarked. This false positive probability is well understood for detectors that use a normalized correlation detection measure. However, in an effort to increase watermark robustness, a number of researchers have proposed detection measures that combine multiple normalized correlations. One such measure is the ratio of the largest normalized correlation to the second largest when an extracted vector is compared to multiple reference vectors. In this paper we analyze the false positive probability of the correlation ratio detection measure under a reasonable assumption about the distribution of unwatermarked works. We derive an analytic formula for the false positive probability and validate this with empirical data.

## 1. INTRODUCTION

Digital watermarking is the process of embedding an invisible message into a digital work (such as an audio file, still image, or image sequence) for the purpose of communicating information about that work. Applications include copyright communication, content authentication, counterfeit deterrence or detection, forensic tracking, and broadcast monitoring. For a detailed review of digital watermarking, we refer to books by Arnold et al. and Cox et al. [1, 2].

Detection errors are inevitable in even the best-designed watermarking systems. A false positive error occurs when a watermark detector indicates the presence of a watermark in an unwatermarked work. Such an error can lead to the mistaken prevention of a legitimate operation or the accusing of an innocent customer. Thus the false positive probability has become a critical issue for many watermarking applications. This probability depends on the watermark detection algorithm, the manner in which the detector is used, and the distribution of unwatermarked works.

The problem of analyzing false detection behavior has received little attention in the watermark literature. Linnartz et al. [3] provide a model to predict false positive probability

in correlation-based watermarking methods and show that a non-white spectrum of a watermark causes the image content to interfere with watermark detection. Hernández and Pérez-González include false detection probability in their framework for discussing watermarking systems [4]. Miller and Bloom present a precise method of calculating the false positive probability when using a normalized correlation detector [5]. They provide an exact formula for the false positive probability under the assumption that the vectors extracted from unwatermarked works are drawn from a radially symmetric distribution. Lichtenauer et al. study the false positive probability in exhaustive geometric searches [6]. They show that image and key dependency in the watermark detector leads to different false positive probability for geometric searches. This current paper draws on all of these works.

For many applications, an extracted vector is compared to a number of different watermark reference vectors. Each reference vector is associated with a different message symbol. The message symbol associated with the reference vector that has the highest similarity to the extracted vector is reported as the detected symbol. The certainty of the detection is the degree of similarity. The most common similarity measure used is normalized correlation. By using the formula described in Miller and Bloom [5], the false positive probability requirement of the application can be used to set a threshold. When the detection value exceeds the threshold, the symbol is reported as present, otherwise the detector reports no symbol.

In an effort to improve the robustness of such techniques, some researchers have proposed certainty measures that combine the largest correlation value with the second largest correlation value (for example, [7, 8]). One such approach considers the difference of these two values as the certainty and another considers the ratio of the two values as the certainty. In both cases, false positive analysis are missing and without these, there is no way to set a threshold and perform a fair robustness comparison. In the remainder of this paper, we derive an analytic formula for the probability of false positive for the correlation ratio measure and present empirical data to support this analysis.

## 2. WATERMARK DETECTION

The probability of false watermark detection is determined by the design of the watermark detector and the distribution of unwatermarked content processed by it. The embedding algorithm is not relevant to computing this probability, because we assume that no watermark has been embedded.

Consider a specific, but typical normalized correlation watermark detector. The input work is first processed to extract an  $n$ -dimensional feature vector,  $V$ . Examples of watermark extraction include various combinations of frequency transforms, block averaging, spectral shaping, whitening, and sub-sampling. These processes are intended to increase robustness, to increase signal-to-noise ratio, to increase efficiency of detection, and/or to enforce the distribution assumed by the detection measure.

The extracted feature vector  $V$  is then compared to a set of  $n$ -dimensional reference vectors  $\{W_1, W_2, \dots, W_m\}$ ,  $m \leq n$ , to obtain a detection measure  $D_V$ . Each reference vector typically represents one of  $m$  message symbols and the symbol associated with the reference vector most similar to the extracted vector is the reported symbol. Only when the detection measure exceeds a detection threshold  $T$ , does the detector report a positive detection. Otherwise no watermark is detected.

The exact formula for computing the detection measure is critical to determine the false positive probability. Normalized correlation is one of the most common techniques that are employed in the detection measure. The *normalized correlation* between two  $n$ -dimensional vectors  $V$  and  $W_i$  is defined as

$$C_{V,W_i} = \frac{V \cdot W_i}{\sqrt{(V \cdot V)(W_i \cdot W_i)}},$$

where the symbol  $\cdot$  is the inner product. Two other published detection measures based on normalized correlation are the difference between the largest and second largest magnitude correlation and the ratio between the largest and second largest magnitude correlation.

Consider the correlation ratio detection measure. The set of  $|C_{V,W_i}|$  is calculated for all  $m$  reference vectors,  $W_i$ . These are then sorted from largest to smallest.

$$|C_{V,W_{i1}}| \geq |C_{V,W_{i2}}| \geq \dots \geq |C_{V,W_{im}}|$$

The correlation ratio detection measure is then defined as

$$D_V = \frac{|C_{V,W_{i1}}|}{|C_{V,W_{i2}}|}, \quad (1)$$

and the message symbol associated with  $W_{i1}$  is reported when  $D_V$  exceeds the detection threshold. Note that  $D_V$  is always greater than or equal to 1.

In the next section, we study the detection measure of Eqn. (1) and derive an analytic formula for the false positive probability, i.e., the probability that  $P(D_V > T)$  for an unwatermarked work, where  $T$  is the detection threshold.

## 3. FALSE POSITIVE PROBABILITY

### 3.1. Orthonormal Basis

Let's consider the set of reference vectors  $\{W_1, W_2, \dots, W_m\}$  where  $|W_i| = \sqrt{W_i \cdot W_i} = 1$  and  $W_i \cdot W_j = 0$  for  $i \neq j$ . In other words, the set  $\{W_1, W_2, \dots, W_m\}$  is an orthonormal basis for an  $m$ -dimensional vector space. When  $m < n$ , this set can be expanded to an orthonormal basis for the  $n$ -dimensional vector space, where  $V$  and  $\{W_i\}$  reside, by adding  $n - m$  unit length vectors to the set where each is orthogonal to all others and to those of  $\{W_i\}$ . We denote such an orthonormal basis in  $n$  dimensional as  $\{W_1, W_2, \dots, W_m, e_{m+1}, e_{m+2}, \dots, e_n\}$ .

Projection of two vectors from one orthonormal basis to another represents a rotation of the axes and this does not change the angle between the vectors. Thus, the normalized correlation between two vectors and the detection measure  $D_V$  are invariant to orthogonal transforms. Consider the projection of  $V$  onto this new basis  $\{W_1, W_2, \dots, W_m, e_{m+1}, e_{m+2}, \dots, e_n\}$  and let the coefficients in this space be denoted as  $\{v_1, v_2, \dots, v_n\}$ . Then the normalized correlation can be written

$$C_{V,W_i} = \frac{V \cdot W_i}{\sqrt{(V \cdot V)(W_i \cdot W_i)}} = \frac{v_i}{\sqrt{V \cdot V}}.$$

Since  $\frac{1}{\sqrt{V \cdot V}}$  is a common factor in all  $C_{V,W_i}$  for  $i = 1, 2, \dots, m$ , we can drop it off and sort  $|v_i|$  directly,

$$|v_{i1}| \geq |v_{i2}| \geq \dots \geq |v_{im}|.$$

The correlation ratio detection measure becomes

$$D_V = \frac{|v_{i1}|}{|v_{i2}|} \quad (2)$$

and a false positive event,  $D_V > T$ , is equivalent to

$$|v_{i1}| > T \cdot |v_{i2}|, \text{ for all } i \neq i1. \quad (3)$$

If we assume that each  $|v_i|$  has the same likelihood of being the largest (a reasonable assumption given that we are studying the case in which no watermark was embedded), then the probability of any normalized correlation being large enough to cause a false positive is  $m$  times greater than the probability of  $v_1$  causing a false positive. We denote this as follows:

$$P_{fp} = P(D_V > T) = m \cdot P(D_V > T \mid i1 = 1). \quad (4)$$

### 3.2. Uniform Distribution Assumption

From Eqn. (2), it is clear that a false positive event is scalar invariant. That is, if for some feature vector  $V$ , its detection measure  $D_V$  is greater than the threshold  $T$ , then for a scaled version  $\alpha \cdot V = \{\alpha \cdot v_1, \alpha \cdot v_2, \dots, \alpha \cdot v_n\}$ , where  $\alpha$  is a non-zero constant,  $D_{\alpha \cdot V}$  is also greater than  $T$ . Thus we can normalize

$V$  such that its  $l^1$  norm is equal to 1,  $\sum_{i=1}^n |v_i| = 1$ . This describes a point on the unit hyper-plane  $H_n = \{(x_1, x_2, \dots, x_n) \mid \sum_{i=1}^n |x_i| = 1\}$  (which is actually a union of  $2^n$  hyper-planes of dimension  $n - 1$ ). The distribution of  $V$  on the unit hyper-plane  $H_n$  depends on the definition of the feature vector. It plays an important role in the false positive study. In this paper, we adopt a uniform distribution.

**Assumption 1** *The normalized feature vector  $V$  from an unwatermarked work is uniformly distributed on the unit hyper-plane  $H_n$ .*

Assumption 1, combined with the fact that a false positive event is scalar invariant, suggests that the false positive probability  $P(D_V > T)$  is equal to the portion of normalized feature vectors on the unit hyper-plane  $H_n$  such that their detection measures are greater than the threshold  $T$ ,

$$P(D_V > T) = \frac{\text{Area}(V \in H_n, D_V > T)}{\text{Area}(H_n)}. \quad (5)$$

If we define a positive hyper-plane  $H_n^+ = \{(x_1, x_2, \dots, x_n) \mid x_i \geq 0, \text{ and } \sum_{i=1}^n x_i = 1\}$ , where  $H_n^+$  is one of the  $2^n$  hyper-planes in  $H_n$ , by symmetry, Eqn. (5) is equivalent to

$$P(D_V > T) = \frac{\text{Area}(V \in H_n^+, D_V > T)}{\text{Area}(H_n^+)}. \quad (6)$$

### 3.3. Independent of Dimension

Now we give a proof that the false positive probability in Eqn. (6) is independent of  $n$ , the dimension of the vector space where  $V$  and  $W_i$  reside. Let's put a subscript in  $P$ ,  $V$ , and  $H$  to denote the dimension. Then

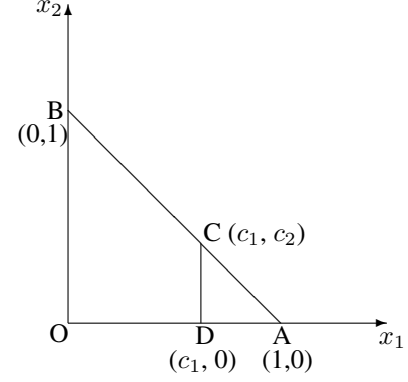
$$\begin{aligned} \text{Area}(V_{n+1} \in H_{n+1}^+, D_{V_{n+1}} > T) &= \\ \text{Area}\left(\sum_{i=1}^n v_i = 1 - v_{n+1}, v_{n+1} \geq 0, D_{V_n} > T\right). \end{aligned}$$

Since

$$\begin{aligned} \text{Area}\left(\sum_{i=1}^n v_i = 1 - v_{n+1}, v_{n+1} \geq 0, D_{V_n} > T\right) &= \\ \int_0^1 (1 - v_{n+1})^{n-1} \cdot \text{Area}(V_n \in H_n^+, D_{V_n} > T) dv_{n+1}, \end{aligned}$$

it follows that

$$\begin{aligned} P_{n+1}(D_{V_{n+1}} > T) &= \frac{\text{Area}(V_{n+1} \in H_{n+1}^+, D_{V_{n+1}} > T)}{\text{Area}(H_{n+1}^+)} \\ &= \frac{\text{Area}(\sum_{i=1}^n v_i = 1 - v_{n+1}, v_{n+1} \geq 0, D_{V_n} > T)}{\text{Area}(\sum_{i=1}^n v_i = 1 - v_{n+1}, v_{n+1} \geq 0)} \\ &= \frac{\int_0^1 (1 - v_{n+1})^{n-1} \cdot \text{Area}(V_n \in H_n^+, D_{V_n} > T) dv_{n+1}}{\int_0^1 (1 - v_{n+1})^{n-1} \cdot \text{Area}(H_n^+) dv_{n+1}} \\ &= \frac{\text{Area}(V_n \in H_n^+, D_{V_n} > T)}{\text{Area}(H_n^+)} \\ &= P_n(D_{V_n} > T). \end{aligned}$$



**Fig. 1.** Geometric Interpretation when  $m = 2$

Thus the false positive probability  $P(D_V > T)$  is independent of the dimension  $n$  of the feature vector. In particular, we can compute  $P(D_V > T)$  by setting  $n = m$ .

### 3.4. Geometric Solution

To derive an analytic formula, we begin with  $m = 2$  where a geometric interpretation can be clearly presented.

When  $m = 2$ ,  $H_2^+ = \{(x_1, x_2) \mid x_i \geq 0, \text{ and } x_1 + x_2 = 1\}$ . We can easily derive an analytic formula of  $P(D_V > T)$ . In Fig. 1,  $O = (0, 0)$ ,  $A = (1, 0)$ ,  $B = (0, 1)$ ,  $C = (c_1, c_2)$  is a point on the line segment  $AB$  such that  $c_1 = T \cdot c_2$ . Add a point  $D = (c_1, 0)$  on the line segment  $AO$ . Then

$$P(D_V > T | i1 = 1) = \frac{|AC|}{|AB|} = \frac{|AD|}{|AO|} = \frac{1}{T+1}.$$

So

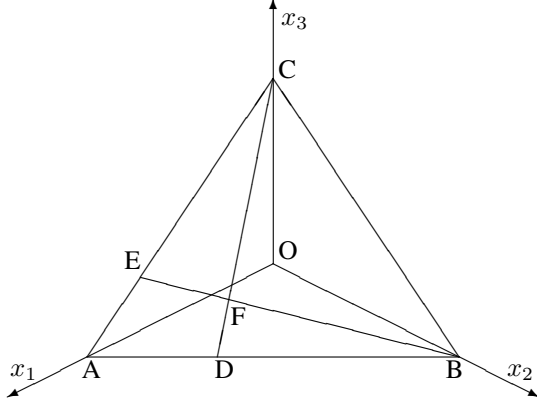
$$P(D_V > T) = 2 \cdot P(D_V > T | i1 = 1) = \frac{2}{T+1}.$$

When  $m = 3$ ,  $H_3^+ = \{(x_1, x_2, x_3) \mid x_i \geq 0, \text{ and } x_1 + x_2 + x_3 = 1\}$ . In Fig. 2,  $O = (0, 0, 0)$ ,  $A = (1, 0, 0)$ ,  $B = (0, 1, 0)$ ,  $C = (0, 0, 1)$ ,  $D = (d_1, d_2, 0)$  is a point on the line segment  $AB$  such that  $d_1 = T \cdot d_2$ ,  $E = (e_1, 0, e_3)$  is a point on the line segment  $AC$  such that  $e_1 = T \cdot e_3$ ,  $F$  is the intersection of  $CD$  and  $BE$ . Then

$$\begin{aligned} P(D_V > T | i1 = 1) &= \frac{\text{Area}(ADFE)}{\text{Area}(ABC)} \\ &= \frac{\text{Area}(ADC) - \text{Area}(CEF)}{\text{Area}(ABC)} \\ &= \frac{1}{T+1} - \frac{\text{Area}(CEF)}{\text{Area}(BCE)} \cdot \frac{\text{Area}(BCE)}{\text{Area}(ABC)} \\ &= \frac{1}{T+1} - \frac{1}{T+2} \cdot \frac{T}{T+1} \\ &= \frac{2}{(T+1)(T+2)}. \end{aligned}$$

So

$$P(D_V > T) = 3 \cdot P(D_V > T | i1 = 1) = \frac{6}{(T+1)(T+2)}.$$



**Fig. 2.** Geometric Interpretation when  $m = 3$

When  $m = 4$ ,  $H_4^+ = \{(x_1, x_2, x_3, x_4) \mid x_i \geq 0, \text{ and } x_1 + x_2 + x_3 + x_4 = 1\}$ . Then

$$\begin{aligned}
 P(D_V > T \mid i_1 = 1) &= \frac{\text{Area}(V \in H_4^+, v_1 > T \cdot v_2, v_1 > T \cdot v_3)}{\text{Area}(H_4^+)} - \\
 &\frac{\text{Area}(V \in H_4^+, v_1 > T \cdot v_2, v_1 > T \cdot v_3, v_1 \leq T \cdot v_4)}{\text{Area}(H_4^+)} \\
 &= \frac{2}{(T+1)(T+2)} - \frac{2 \cdot T}{(T+1)(T+2)(T+3)} \\
 &= \frac{6}{(T+1)(T+2)(T+3)}.
 \end{aligned}$$

So

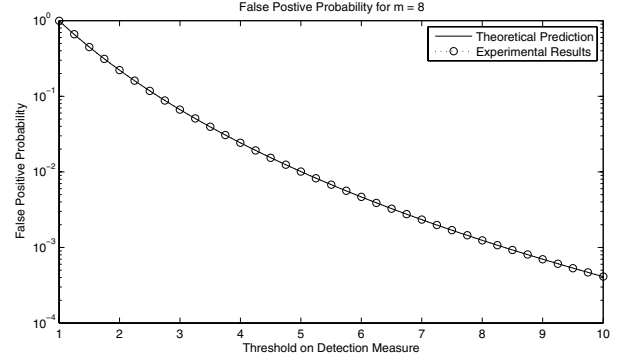
$$P(D_V > T) = \frac{24}{(T+1)(T+2)(T+3)}.$$

Realizing that  $P(D_V > T \mid i_1 = 1)$  for  $m$  and  $m - 1$  is different by a factor  $\frac{m-1}{T+m-1}$ , one can derive an analytic formula of the false positive probability  $P(D_V > T)$  for any  $m$ , which is

$$P_{fp} = P(D_V > T) = \frac{m!}{\prod_{i=1}^{m-1} (T+i)}. \quad (7)$$

#### 4. EXPERIMENTAL RESULTS

To verify that the formula in Eqn. (7) is correct, we compared its prediction against results obtained from 100,000 synthetic vectors drawn from a unit hyper-plane uniform distribution. The dimension of the feature vector is 1024 ( $n = 1024$ ). There are 8 orthogonal watermark vectors in the detector ( $m = 8$ ). Fig. 3 shows the results of our experiment compared against the predictions made by Eqn. (7). The analytic formula's predictions match very closely with the experimental results down to at least  $P_{fp} = 10^{-4}$ .



**Fig. 3.** False Positive Probability for  $m = 8$

#### 5. CONCLUSIONS

We have presented an analytic formula for the false positive probability of the correlation ratio watermark detection. This formula holds exactly if the distribution of feature vectors extracted from unwatermarked content is uniform on a unit hyper-plane. The relationship between this result and that of Miller and Bloom [5] will be studied in a forthcoming paper.

#### 6. REFERENCES

- [1] M. Arnold, S. D. Wolthusen, and M. Schmucker, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House Publishers, 2003.
- [2] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2001.
- [3] J-P. Linnartz, T. Kalker, and G. Depovere, "Modelling the false alarm and missed detection rate for electronic watermarks," in *Proc. 2nd International Workshop on Information Hiding*, 1998, pp. 329–343.
- [4] J. R. Hernández and F. Pérez-González, "Shedding more light on image watermarks," in *Information Hiding 1998*, D. Aucsmith, Ed. 1998, pp. 191–207, Springer-Verlag.
- [5] M. L. Miller and J. A. Bloom, "Computing the probability of false watermark detection," in *3rd International Workshop on Information Hiding*, 1999, pp. 146–158.
- [6] J. Lichtenauer, I. Setyawan, T. Kalker, and R. Lagendijk, "Exhaustive geometrical search and the false positive watermark detection probability," in *Security and Watermarking of Multimedia Contents V*, 2003, pp. 203–214.
- [7] G. B. Rhoads and R. K. Sharma, *Digital Watermark Screening and Detecting Strategies*, US Patent 6516079.
- [8] F. Ahmed and I. S. Moskowitz, "The binary phase only filter as an image watermark," NRL CHACS Tech Memo 5540-38TM, Naval Research Lab, 2004.