

On Sequential Watermark Detection

R. Chandramouli and Nasir D. Memon*

Abstract

This paper introduces a new watermark detection paradigm based on sequential detection. It is shown that this framework leads to significantly faster watermark detection compared to more commonly used fixed sample size detectors. Fundamental issues such as joint optimization of watermark encoding and detection are addressed within this framework. Theoretical results are validated through experiments. It is observed that the proposed framework can detect watermarks 70% faster than fixed sample size based detectors. We believe that the proposed method has a good potential to be successful in many practical applications.

1 Introduction

Digital watermarks have been recently proposed for a variety of applications including authentication and copy protection of multimedia content. In a general sense, digital watermarking can be described as the process of embedding, by means of a secret key, an imperceptible digital signal (the watermark) into multimedia content. Knowledge of this secret key is essential to subsequent extraction or detection of the watermark. In this paper, we assume the secret key is known to the detector and hence for the sake of brevity, we choose to ignore it. We do this as our results are not affected by the choice of any specific key but have to do with the nature of the detection process itself, once the key is known. Also for ease of exposition, we assume in the rest of this paper that the content being watermarked is a still image, although our results, in principle, are equally applicable to audio and video data.

Adopting the definition in [1], we choose to represent the watermark embedding step as follows:

$$Y = X + W \quad (1)$$

where $X = \{x_1, x_2, \dots, x_n\}$ is the original image, expressed either in terms of pixels or transform coefficients or any other feature set values. $W = \{w_1, w_2, \dots, w_n\}$ is the watermark information being embedded, and $Y = \{y_1, y_2, \dots, y_n\}$ represents the watermarked image.

Note that the above characterization is general enough to capture any watermarking technique that preserves signal length. This is due to the fact that the watermark itself can be just viewed as the difference between the original image and the watermarked image, no matter how it actually gets embedded. However, as noted in [1], such a characterization is not useful in many instances. For example, with a watermarking technique that embeds by performing geometric deformations, it may not make sense to characterize the watermark or the watermarking process as given in Eq. (1). Nevertheless, the additive framework represented in Eq. (1) is true and also meaningful for a large number of watermarking techniques in the literature. Such techniques embed by adding a suitably shaped watermark signal, to the original image, either in the spatial domain or in the transform

*To appear in Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery, IEEE Trans. on Signal Processing, 2003.

domain. In this paper we restrict our attention to additive watermarking techniques as represented in Eq. (1).

Depending on the way the watermark is inserted, and depending on the nature of the watermarking algorithm, the watermark detection method can take on very distinct approaches. In some watermarking schemes, a watermark can be extracted in its exact form. We refer to this procedure as *watermark extraction*. In other cases, we can only detect whether a specific given watermarking signal is present in an image. We refer to this as *watermark detection*. Both of these approaches have meaningful applications. For example in an image authentication application [4], a robust hash function of the image is embedded as a watermark. Later, this embedded hash value is extracted and compared against the computed hash value to assert whether an image is authentic or has been significantly modified. In contrast, for a broadcast monitoring application [4], a monitoring station is placed by a news agency in the broadcast region of each local station. This monitoring station continually examines local broadcast content for clips that contain the specific watermark of the news agency. Presence of a watermark indicates that its content was aired by the local station and for which charges would be due. Here, the monitoring station need only detect the presence of a watermark. There are no information bits that need to be extracted from the watermarked content.

In this paper we restrict our attention to watermarking techniques and applications where the watermark detection process just requires detection of the presence of a specific watermark. In general the detection process varies from technique to technique. But broadly speaking, one can view the detection process as being one of statistically differentiating between a watermarked image and an unwatermarked image, given a specific watermark. This in turn could be done by a simple correlation operation or by hypothesis testing [4] or by any other suitable procedure.

Irrespective of the specific procedure used for detection, most popular watermark detection techniques process a fixed number of observations. A test statistic is computed by the detector based on this fixed number of observations and compared against a threshold to determine if a specific watermark is detected or not. Such an approach, in our opinion, has some drawbacks as we list below:

- Fixed sample size watermark detection is not efficient in the sense of minimizing the required number of signal observations for detection.
- Fixed sample size hypothesis tests introduce a trade-off between the false alarm and miss probability. They do not allow simultaneous control of these probabilities during watermark detection.
- Fixed sample size detection could require a prohibitively large number of signal samples in order to achieve very small error probabilities. This is not suitable for applications such as detecting watermarks in large image databases or video watermarking.

In this paper we explore an alternate paradigm for watermark detection, namely *sequential watermark detection*. Sequential watermark detection techniques allow both false alarm and miss probabilities to be controlled simultaneously. That is, we can fix these probabilities and then allow the number of observed features during detection to vary such that they are achieved. Naturally, the question arises whether there is something to be gained by fixing error probabilities and allowing the number of observations used by the detector to vary. We show in this paper that there is indeed much to be gained by this alternate approach. Specifically, a sequential detector allows one to minimize the average number of observations that need to be processed for detection. It results in, what has been called in the literature as the quickest watermark detector [5].

In many practical scenarios, quickest watermark detection can be very beneficial. Specifically, consider the following example scenarios:

- *MediaBridge*TM is a commercial watermarking product from Digimarc [2]. In *MediaBridge*, watermark detection is done by first capturing a sequence of still images of a printed (potentially watermarked) image by using a simple PC camera. Since these individual frames vary in their quality, the *MediaBridge* reader sequentially searches each frame for Digimarc's watermark [3]. If a frame is not found to be of good quality it is discarded and the next frame in the sequence is searched and so on. In the above process it is clear that a watermark detector that can minimize the time taken to search each frame for the presence of a watermark will be highly desirable. This argument can be extended to video watermark detection in general where efficient detection can play a crucial role.
- Suppose one wishes to search a large database for images for a particular watermark. Then it is very useful to have a watermark detection algorithm that can detect the presence or absence of the watermark as quickly as possible, *i.e.*, the algorithm must be able to use as little as information as possible (such as pixel or feature values) to make a decision. This will increase the speed of watermark detection.
- In progressive image transmission, *most significant* information is transmitted first followed by enhancements. If the watermark decoder can make on-line decisions about the watermark content as and when it receives the information, then it will be immensely helpful in bandwidth limited applications where progressive transmission is used.
- Low computation also translates quite proportionally to low power consumption. For connected hand held devices that need to detect the presence of a watermark in downloaded content, quick watermark detection means low power consumption which is increasingly becoming an important factor in mobile computing.

We note that the proposed sequential watermark detection approach has other benefits apart from resulting in quicker watermark detection. Most watermark detection techniques in the literature treat watermark design and detection as two basically separate issues. However, there is an inherent relationship between these two operations. A good detection scheme can compensate for a poor watermark design. Therefore, it may be prudent to choose a watermarking scheme or its parameters based on the watermark detector. It is well known that watermarking involves trade-offs between the amount of modification made to the data on one hand and the degree of immunity to host signal attacks [6] and visual quality degradation on the other. A large amount of signal modification can often lead to significant degradation in the host signal. Given this fact, it is essential that we choose the *optimal* number of pixels (or coefficients, or feature points) to modify so that the detector is able to detect the watermark within the desired error probabilities. Traditionally, the choice of the watermark length has been somewhat *ad hoc*—Cox. et. al. [10] choose 1000 discrete cosine transform AC coefficients to embed a watermark, Pitas et. al. [7] choose half the number of the pixels to watermark, etc. There seems to be no specific scientific reason for these choices. So, what is a good choice for the embedded watermark length from a joint embedding-detection perspective? This can be answered using the proposed sequential detection paradigm. We show, for a specific watermarking technique, how to compute the watermark length that is enough to facilitate its detection within desired probability of detection errors. We observe that the proposed framework is quite general and it is easy to incorporate many popular watermarking algorithms within this analysis.

The paper is organized as follows. Mathematical analysis of the proposed sequential detection framework is presented in Section 2. Section 3 deals with a spatial watermarking example showing all the details of the application of the analysis presented in Section 2. In Section 4 we show how to compute the optimal watermark length in a mathematical manner. Performance analysis and experimental results are given Section 5. Concluding remarks can be found in Section 6.

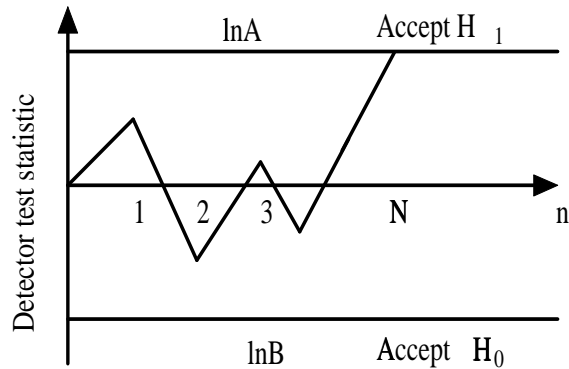


Figure 1: Example sample run of a sequential detector.

2 Sequential Watermark Detection

In this section we describe the analysis involved in designing a sequential watermark detector as a binary hypothesis test and explain how some of the popular watermarking techniques can be detected by this method.

Sequential hypothesis testing was pioneered by Wald [8]. The main feature of a sequential hypothesis test that differentiates it from a fixed sample size (FSS) hypothesis test is the number of observations (or samples) required by the sequential detector. Unlike a FSS detector, this number is not pre-determined, but a random variable. It varies depending on the uncertainty or information content of successive observations. If the uncertainty, for example, measured in terms of the sample variance, is high, then we can expect the average number of samples consumed by the sequential detector to achieve desired error probabilities to be high also. A sequential detector accumulates the information provided by each observation and accepts a hypothesis when just enough information has been collected. A FSS detector on the other hand operates on a block of data and therefore does not possess the capability to identify samples that do not convey additional information about the hypotheses.

Broadly, the steps involved in a sequential test for a hypothesis H are as follows. An experiment is performed, the outcome is observed and a test statistic is computed. Based on the test statistic the sequential detector makes one of the following decisions: (a) accept the hypothesis H , (b) reject H , or (c) continue the experiment and make an additional observation. This procedure continues until the test terminates by either accepting or rejecting the hypothesis. Thus, we see that the number of observations required by the detector before it terminates is a random variable. Fig. 1 shows an example sample run of a sequential detector to test between $H = H_1$ or $H = H_0$. The X-axis in the figure denotes the number of the observation and Y-axis is the value of the test statistic computed after observing that sample. We see that the test statistic of the detector oscillates

between the two decision boundaries, $\ln A$ and $\ln B$, and finally accepts H_1 by terminating in the upper boundary.

We know that detectors based on the the Neyman-Pearson criterion [9] (Ch. 3) utilize a fixed number of samples to test between two hypotheses for a given false alarm ($\alpha = \text{P}(\text{decide } H_1 | H_0 \text{ true})$) and miss probability ($\beta = \text{P}(\text{decide } H_0 | H_1 \text{ true})$). A test statistic is computed as a function of the observations and then compared against a decision threshold. If it exceeds the threshold H_1 is decided, if not H_0 is decided. It is known that in these types of FSS detectors only the desired false alarm can be fixed while the detector achieves a certain miss probability. In general, in FSS detection, both these error probabilities are not controllable. One is traded-off of for the other as given by a receiver operating characteristic [9] (Ch. 3).

We now introduce sequential watermark detection for discrete-time observations. Suppose the observations are represented by a sequences of real valued random variables $\{Y_k; k = 1, 2, \dots\}$ that are independent and identically distributed according to,

$$\begin{aligned} H_0 : Y_k &\sim P_0, \quad k = 1, 2, \dots \\ H_1 : Y_k &\sim P_1, \quad k = 1, 2, \dots \end{aligned} \tag{2}$$

where P_0 and P_1 are two possibilities for the probability distribution of the observations. Here, H_0 corresponds to the case when a watermark is absent and H_1 stands for the presence of the watermark. As an example, if the spread-spectrum watermark [10] is employed then the samples $\{Y_k\}$ represent the discrete cosine transform (DCT) coefficients. In this case, the variance of Y_k will differ depending on whether H_0 or H_1 is the true hypothesis. This is because, in spread-spectrum watermarking, the watermark is a zero-mean, Gaussian distributed random vector with finite variance. Therefore, if H_1 is true then the variance of Y_k will be equal to the sum of the variances of the corresponding DCT coefficient and the watermark coefficient's variance. The hypothesis test in this case attempts to identify this difference in the variance. Another example is the spatial domain watermark proposed by Pitas et. al. [7]. Here the watermark perturbs the spatial domain mean value of the host image and the hypothesis test will test for a change in this value.

A sequential watermark detector can be represented by a *sequential decision rule* which is a pair: (*stopping rule*, *terminal decision rule*). The stopping rule determines a *stopping time* (N) which corresponds to the time (or sample number) at which the detector stops considering further observations. Clearly, N is a random variable because it depends on the observed random variables. Note that $P(N < \infty) = 1$ means that the sequential test terminates (*almost surely*) after only a finite number of observations. Once the value of N is computed the sequential decision rule determines which one of the two hypothesis is indeed accepted by the detector. In the formulation discussed so far we have assumed that the hypotheses are completely known. This means, the parametric models and their parameters are known for the observed samples. If this assumption holds, then we can derive an optimal sequential watermark detector. Suppose this assumption does not hold, say, for example, the estimate of the original image in blind watermark detector is coarse (noisy). It is then known that the optimal sequential detector is sensitive to such uncertainties [9] (Ch. 3). In order to alleviate this problem, we also discuss a sub-optimal but robust watermark detector based on quantized observations. The next two subsection are devoted to the optimal and sub-optimal sequential watermark detectors.

2.1 Optimal Sequential Watermark Detection

Let the probability density functions of $\{Y_k\}$ conditioned on H_1 and H_0 be denoted by $f_1(y_1, y_2, \dots, y_n)$ and $f_0(y_1, y_2, \dots, y_n)$, $n \geq 1$, respectively. Select two decision thresholds A and B such that

$-\infty < B < A < \infty$. Then the optimal detector based on the sequential probability ratio test (SPRT) is given by [8] (Ch. 3),

$$\mathcal{L}(y_1, y_2, \dots, y_n) = \ln \left[\frac{f_1(y_1, y_2, \dots, y_n)}{f_0(y_1, y_2, \dots, y_n)} \right] \quad \begin{cases} \geq & \ln A \text{ decide } H_1 \\ \leq & \ln B \text{ decide } H_0 \\ \text{else} & n = n + 1 \end{cases} \quad (3)$$

where $\mathcal{L}(y_1, y_2, \dots, y_n)$ denotes the logarithm of the likelihood ratio of the first n observations. If the log-likelihood ratio does not exceed one of two decision thresholds, one more observation is taken and the new log-likelihood ratio is compared against the two thresholds. This process continues until one hypothesis is accepted and the test stops. The stopping rule (φ_n) and the decision rule (χ_N) for the SPRT are given by,

Stopping rule :

$$\begin{aligned} N &= \inf\{n : \mathcal{L}(y_1, y_2, \dots, y_n) \geq \ln A \text{ or } \mathcal{L}(y_1, y_2, \dots, y_n) \leq \ln B\} \\ \varphi_n &= \begin{cases} 1 & \text{if } n = N \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (4)$$

Decision rule:

$$\chi_N(y_1, y_2, \dots, y_N) = \begin{cases} H_1 & \text{if } \mathcal{L}(y_1, y_2, \dots, y_N) \geq \ln A \\ H_0 & \text{if } \mathcal{L}(y_1, y_2, \dots, y_N) \leq \ln B. \end{cases} \quad (5)$$

The constants B and A are computed using the false alarm and miss probability constraints and are given by the following lemma [8] (Ch. 3).

Lemma 1 *For given false alarm and miss probabilities the following inequalities hold,*

$$\begin{aligned} A &\leq \left(\frac{1 - \beta}{\alpha} \right) \\ B &\geq \left(\frac{\beta}{1 - \alpha} \right). \end{aligned} \quad (6)$$

With negligible loss in the error probabilities, the above inequalities can be taken to be equalities, thus giving the values for the decision thresholds—called the Wald’s approximation [8] (Ch. 3). Let $E(N|H_i)$, $i = 0, 1$ denote the average number of samples required by the sequential test to terminate given that hypothesis H_i is true. Then according to the optimality theorem of the SPRT [8] (Appendix A.7) the sequential watermark detector based on SPRT is optimum in the sense of requiring the minimum average number of observations to detect the presence or absence of a watermark among all sequential and fixed sample size detectors. Note here that we have made no assumptions about the specific watermark encoding algorithm. Therefore, this detector can be used to detect watermarks that produce a statistical difference between the original and the watermarked image. It is the above optimality property that makes the sequential watermark detection procedure very attractive for fast or quickest watermark detection. A natural question to ask now is: what is the guarantee that the sequential test will ever terminate? It can be shown that under certain general regularity conditions the sequential test terminates with probability 1 [8] (Appendix A.1). We do note that, for most practical watermark detection problems the conditions for theoretical convergence of the sequential detector hold. We refer to [8] (Appendix A. 1) for the statement of these conditions and a formal proof of convergence of the sequential detector.

2.2 Sub-optimal Sequential Detection

It is known that the parametric, optimal sequential detector is sensitive to uncertainties in the parametric model [9] (Ch. 3). For example, in blind image watermark detection, the original image can be estimated from the watermarked one using neighborhood averaging. If the prediction error is modelled as a Gaussian probability distribution when it is (say) Laplacian, then the performance of the optimal sequential detector could suffer. It is in such circumstances we may need a robust detector even if it is sub-optimal. We discuss one such robust, sub-optimal detector that operates on two-level quantized observations called the *sequential sign detector*.

A sign detector uses only the sign of an observation for detection thus quantizing an observation to two levels, namely, ± 1 . Since the magnitude information is lost before detection, this detector is sub-optimal. Therefore, the only knowledge required by the detector is the probability of a positive or negative observation. The sign detector is well known for its robustness [9] (Ch. 3) and for requiring less information about the probability distribution of the prediction error, attacks etc. Following quantization of the observation a SPRT for the resultant binomial random process is performed. Since the steps follow on the lines of the optimal sequential detector, we do not repeat them again. We will show later that even with the quantized information, this robust, sub-optimal sequential sign detector can outperform a FSS detector in terms of the average number of samples required for detection.

3 An Example: Sequential Spatial Watermark Detection

In the previous section we described the general framework for sequential watermark detection and some of its properties. Now, we consider a specific watermarking scheme and explain all the steps that are involved in sequential detection. We investigate the operating characteristic of the detector and the required average sample number for detection. The average sample number is a measure of the *speed* of the sequential detector, *i.e.*, the smaller this number is faster the detection. Therefore from a watermark detectors point of view this number is significant.

To illustrate the working of the proposed watermark detector we choose a version of a spatial domain watermark proposed by Pitas et. al. [7]. The reasons for this are the following:

- Many popular watermarks are additive and usually have zero mean and finite variance. Further most transform-domain based zero-mean, finite variance, additive watermarks can be equivalently represented as zero-mean, additive watermarks in the spatial domain when the transform is linear.
- This watermark encoding scheme is simple to explain and a wide variety of spatial domain watermarking algorithms can be formulated in a similar fashion.

We briefly describe the watermark encoding algorithm here. Consider an image, $I = \{x_{ij}, (i, j) \in \Omega\}$ where Ω is an index set. The set Ω is partitioned into two subsets Λ and Λ^c . The cardinality of the set, Λ , denoted by $|\Lambda|$ is fixed *a priori* to be equal to $\frac{1}{2}|\Omega|$. Then, the watermarking algorithm produces a watermarked image, $Y = \{y_{ij}, (i, j) \in \Omega\}$ in the following way:

$$y_{ij} = \begin{cases} x_{ij} + K, & \text{if } (i, j) \in \Lambda \\ x_{ij}, & \text{if } (i, j) \in \Lambda^c \end{cases} \quad (7)$$

where K is called the embedding factor.

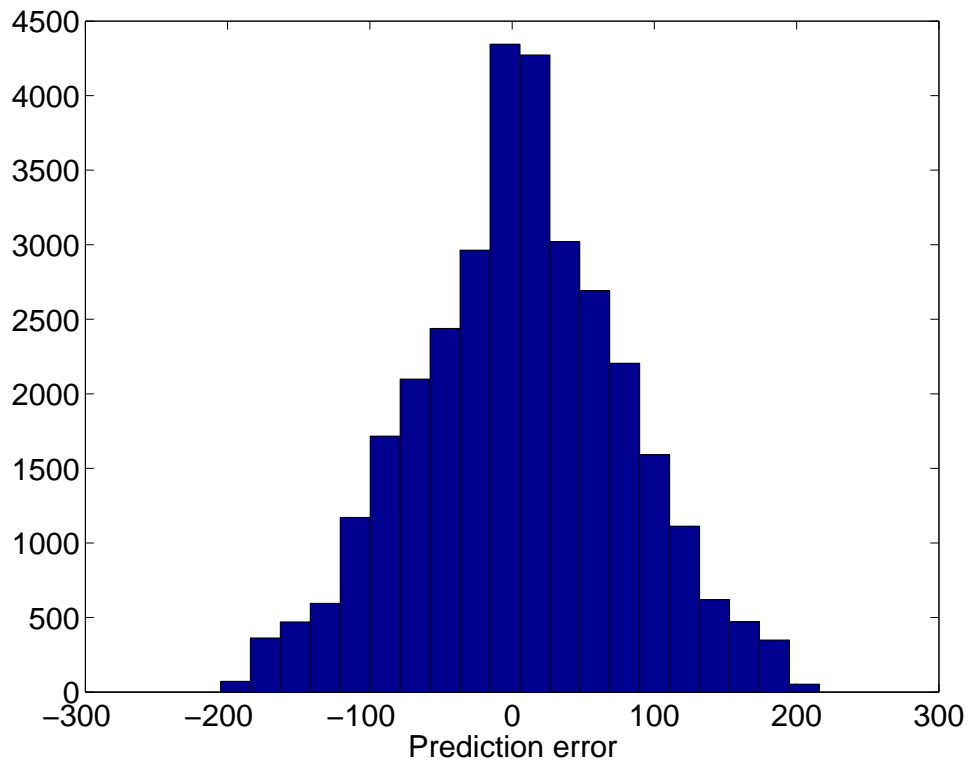


Figure 2: Histogram of prediction error. Host image is predicted using the nearest neighbouring pixel from the set Λ^c with the watermarking key information.

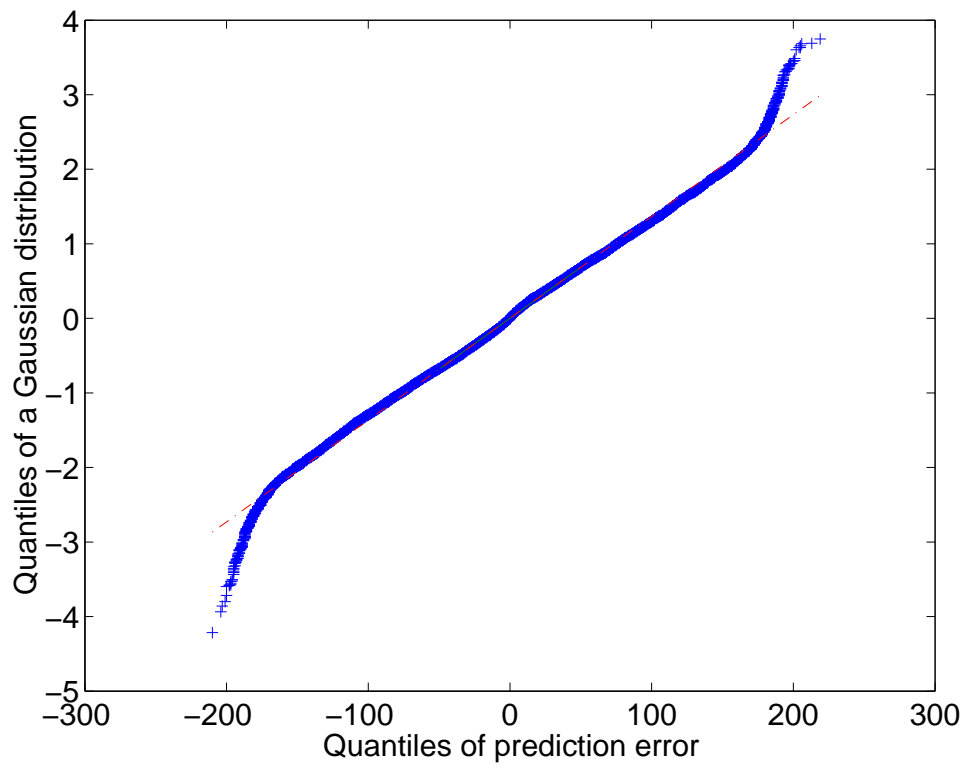


Figure 3: Quantile-quantile plot of prediction error versus standard Gaussian distribution.

3.1 Optimal Sequential Watermark Detector

The watermark detector conducts a binary hypothesis test between H_0 : no watermark and H_1 : watermark present. The two kinds of detection error probabilities α and β are assumed to be given. We also assume that the original image is not available at the detector side. For simplicity, the original image is estimated at the detector side as the nearest neighboring unmarked pixel. This is possible because the detector knows the watermark key. The uncertainty created due to the estimation error is modelled by a zero-mean, white Gaussian distribution with finite variance. To justify this model we performed several experiments and only the result for Lenna image is reproduced here for the sake of brevity. The histogram of the prediction error shown in Fig. 2 gives us valuable clues to choose an appropriate probabilistic model. We see from this figure that the histogram visually looks like more or less like a Gaussian distribution. In order to validate this model more rigorously a quantile-quantile test was performed as shown in Fig. 3. On the X-axis are the quantiles of the prediction error distribution and the Y-axis represents the quantiles of a standard Gaussian distribution. We observe that this plot is almost a straight line meaning a Gaussian model is a good fit for the prediction errors. Some theoretical reasons to justify a Gaussian model for the observations at the detector under certain conditions are as follows. Many watermark attacks such as compression, smoothing etc. can be modelled reasonably well as additive noise under suitable assumptions. Further, the Gaussian distribution has the maximum entropy among all probability distributions with a given variance. Also, if multiple random attacks are performed that can be modelled as additive, then, by the central limit theorem, the combination of these attacks can be represented by a Gaussian distribution. Finally, it is not uncommon in parametric estimation theory [9] (Ch. 2) to approximate the estimation errors using a Gaussian distribution. As a note, an optimal sequential detector can be derived for other probabilistic models also using the general framework presented in Section 2.

Now the sequential watermark detection problem is formulated as the following binary hypothesis test,

$$\begin{aligned} H_1 & : y_{ij} = x_{ij} + K \\ H_0 & : y_{ij} = x_{ij}, \quad i, j \in \Lambda. \end{aligned} \quad (8)$$

Let the estimate of I be denoted by $\hat{I} = \{\hat{x}_{ij}\}$ and let $l_{ij} = y_{ij} - \hat{x}_{ij}$. By replacing the double index $\{ij\}$ by a single index k for notational convenience we obtain the following equivalent problem:

$$\begin{aligned} H_1 & : l_k = K + n_k \\ H_0 & : l_k = n_k, \quad k = 1, 2, \dots, |\Lambda| \end{aligned} \quad (9)$$

where $\{n_k\}$ stands for a zero mean, white Gaussian noise with variance σ^2 . It represents the estimation errors, attacks, etc. as discussed previously. Then the SPRT based sequential watermark detector (Eq. (3)) becomes,

$$S(n) = \mathcal{L}(l_1, l_2, \dots, l_n) = \frac{K}{\sigma^2} \sum_{k=1}^n (l_k - K/2) \begin{cases} \geq & \ln A \text{ decide } H_1 \\ \leq & \ln B \text{ decide } H_0 \\ \text{else} & n = n + 1 \end{cases} \quad (10)$$

where the values of A and B are given by Eq. (6). Therefore, Eq. (10) completely describes the optimal sequential watermark detector for the current case. As an aside, note that $\{S(n)\}$ is a random walk process.

3.1.1 Operating Characteristic Function and Average Sample Number

We have assumed so far that the observations presented to the detector either have mean zero or K . But, in practice, the sequential detector could operate under conditions when the observations do not come from a probability distribution that has a mean neither equal to zero (corresponding to H_0) or K (corresponding to H_1). Three practical scenarios where this could happen are the following: (a) the watermark embedding factor being tested is incorrect, (b) when an estimate of the original image is subtracted from the watermarked image before blind detection estimation errors corrupt the watermark itself, and/or (c) attacks on the watermark change the watermark embedding factor. Under these and other circumstances the operating characteristic function is a valuable tool to infer the performance of the detector.

The operating characteristic (OC) function of the sequential detector is a function $L(k_0)$ which is then defined as the probability that the sequential detector will terminate by accepting H_0 when k_0 is the true value of the parameter (here, embedding factor) under test. Therefore $k_0 = 0$ corresponds to hypothesis H_0 (no watermark) and $k_0 = K$ corresponds to H_1 while the other values of k_0 represent all the other cases. Intuitively, the closer the value of k_0 is to zero, the higher must be the probability of the detector favoring H_0 . Similarly, when k_0 is closer to K , H_0 must be rejected with a high probability. Clearly, $k_0 = K/2$ represents the *worst case condition* for the detector. The derivation of the OC function for the optimal sequential detector for testing between the two hypotheses follows along the similar lines presented by Wald [8] (Ch. 3) and can be shown to be,

$$L(k_0) \approx \begin{cases} \frac{\ln((1-\beta)/\alpha)}{\ln((1-\beta)/\alpha) - \ln(\beta/(1-\alpha))}, & k_0 = K/2 \\ \frac{((1-\beta)/\alpha)^{h(k_0)} - 1}{((1-\beta)/\alpha)^{h(k_0)} - (\beta/(1-\alpha))^{h(k_0)}}, & k_0 \neq K/2 \end{cases} \quad (11)$$

where $h(k_0) = (K - 2k_0)/K$. The average sample number (ASN) required by the detector to terminate by accepting one of the hypothesis when the true value of the parameter under test is k_0 can be shown to be [8] (Ch. 3),

$$E_{k_0}(N) \approx \begin{cases} \frac{-\ln((1-\beta)/\alpha)\ln(\beta/(1-\alpha))}{(K/\sigma)^2}, & k_0 = K/2 \\ \frac{L(k_0)\ln(\beta/(1-\alpha)) + (1-L(k_0))\ln((1-\beta)/\alpha)}{(K/2\sigma^2)(2k_0 - K)}, & k_0 \neq K/2. \end{cases} \quad (12)$$

Numerical evaluation of these functions will be presented in Section 5.

3.2 Sub-optimal Sequential Watermark Detection

As discussed before, the optimal SPRT is sensitive to parametric uncertainties. In order to *robustify* the SPRT we start by quantizing the observations before detection. Therefore, instead of using l_{ij} directly in the detection procedure we use a two-level quantized version of l_{ij} . This can be shown to be robust against outliers [5] and allows the theory of random walks to be applied to study the sequential detection problem with quantization. The two-level quantizer is chosen to be $q_{ij} = \text{sgn}(l_{ij} - K/2)$, where,

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases} \quad (13)$$

Therefore the problem now to be solved is the following:

$$\begin{aligned} H_1 & : P(q_{ij} = 1) = p, P(q_{ij} = -1) = q \\ H_0 & : P(q_{ij} = 1) = q, P(q_{ij} = -1) = p, (i, j) \in \Lambda \end{aligned} \quad (14)$$

where $p = P(q_{ij} \geq 0|H_1) = P(q_{ij} < 0|H_0)$ due to the symmetry of the Gaussian distribution and $q = 1 - p$. It is then quite straightforward to show that,

$$\begin{aligned} p &= \int_{K/2}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(l-K)^2/2\sigma^2} dl \\ &= \frac{1}{2} \operatorname{erfc} \left(\frac{-K}{2\sqrt{2\sigma^2}} \right) \end{aligned} \quad (15)$$

where $\operatorname{erfc}(t) = 2/\sqrt{\pi} \int_0^t e^{-x^2} dx$ stands for the complement of the error function. We now rename and re-index the random variables $\{q_{ij}; (i, j) \in \Lambda\}$ as $\{\tilde{q}_k; k = 1, 2, \dots, |\Lambda|\}$ for notational convenience. Then, the SPRT for these binomially distributed random variables $\{\tilde{q}_k\}$ after the n th observation is given by,

$$(p/q)^k (q/p)^{n-k} \begin{cases} \geq T_1 & \text{decide } H_1 \\ \leq T_2 & \text{decide } H_0 \\ \text{else } n = n + 1 \end{cases} \quad (16)$$

where k denotes the number of \tilde{q}_k 's that are non-negative. Constants T_1 and T_2 are chosen based on the error probability constraint. By taking the natural logarithm on both sides of the above equation, renaming the corresponding new decision thresholds and after algebraic simplification we obtain,

$$S(n) = \sum_{i=k}^n \tilde{q}_k \begin{cases} \geq A_1 & \text{decide } H_1 \\ \leq -B_1 & \text{decide } H_0 \\ \text{else } n = n + 1 \end{cases} \quad (17)$$

Since the test statistic is an integer we choose the decision thresholds $A_1 > 0$ and $B_1 > 0$ to be integers. We observe that $\{S(n)\}$ is a random walk sequence and the sequential detection procedure stops when this random walk is absorbed in one of the two boundaries, A_1 or $-B_1$. It can be shown that the random walk hits one of the boundaries with probability 1 under certain conditions on p [8] (Ch. 5) which is valid for watermark detection problems. Clearly, the stopping time random variable $N = \inf\{S(n) \notin (-B_1, A_1)\}$ is now a function of $\{\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_n\}$.

3.2.1 Average Sample Number and Operating Characteristic Function

The watermark detection problem has been now reduced to that of studying random walk sequence within two absorbing boundaries. Noting that $\{S(n)\}$ is a discrete-time Markov chain with state space given by \mathcal{Z} the set of integers, the probability of the random walk terminating at the boundaries can then shown to be [5],

$$P(\text{terminate at } A_1|H_1) = p^{A_1} \frac{p^{B_1} - q^{B_1}}{p^{A_1+B_1} - q^{A_1+B_1}} \quad (18)$$

$$P(\text{terminate at } A_1|H_0) = q^{A_1} \frac{p^{B_1} - q^{B_1}}{p^{A_1+B_1} - q^{A_1+B_1}} \quad (19)$$

$$(20)$$

Since the random walk is absorbed in one of the boundaries w.p.1 $P(\text{terminate at } -B_1|H_i) = 1 - P(\text{terminate at } A_1|H_i)$, $i = 0, 1$ and the finite moments of N exist. Further, the expected

values of the sample size required by this sequential detector conditioned on hypothesis H_i , $i = 0, 1$ can be shown to satisfy [5],

$$\begin{aligned} E(N|H_1) &= \frac{B_1}{q-p} - \frac{A_1 + B_1}{q-p} \frac{1 - (q/p)^{B_1}}{1 - (q/p)^{A_1+B_1}} \\ E(N|H_0) &= \frac{B_1}{p-q} - \frac{A_1 + B_1}{p-q} \frac{1 - (p/q)^{B_1}}{1 - (p/q)^{A_1+B_1}}. \end{aligned} \quad (21)$$

Now, we investigate the choice of the decision thresholds and the OC function. We know that the optimal thresholds must be chosen so that $P(\text{terminate at } -B_1|H_1) \leq \beta$ and $P(\text{terminate at } A_1|H_0) \leq \alpha$ which from Eq. (18) in turn implies,

$$1 - \frac{r^{A_1+B_1} - r^{A_1}}{r^{A_1+B_1} - 1} \leq \beta \quad (22)$$

and

$$\frac{r^{B_1} - 1}{r^{A_1+B_1} - 1} \leq \alpha \quad (23)$$

where $r = p/q > 1$. The smallest integer pair A_1 and B_1 that satisfy the above inequalities are the desired decision thresholds. It is not difficult to see that [11], these are given by $A_1 = \lceil \ln((1-\beta)/\alpha)/\ln r \rceil$ and B_1 is the smallest integer such that, if $\gamma(1-\beta) < 1$,

$$\frac{1 - \alpha/\gamma}{\beta} \leq r^{B_1} \leq \frac{1 - \alpha}{1 - \gamma(1-\beta)} \quad (24)$$

and, if $\gamma(1-\beta) \geq 1$,

$$\frac{1 - \alpha/\gamma}{\beta} \leq r^{B_1} \quad (25)$$

where $\gamma = r_1^A(\alpha/(1-\beta)) \geq 1$. The operating characteristic function of this detector based on Binomial observations can be derived using the result found in [8] (Ch. 5) to get,

$$L(u) \approx \begin{cases} \frac{\ln((1-\beta)/\alpha)}{\ln((1-\beta)/\alpha) - \ln(\beta/(1-\alpha))}, & u = 0.5 \\ \frac{((1-\beta)/\alpha)^{h(k_0)} - 1}{((1-\beta)/\alpha)^{h(k_0)} - (\beta/(1-\alpha))^{h(k_0)}}, & u \neq 0.5 \end{cases} \quad (26)$$

where u is the probability of a positive observation when the true watermark embedding value is k_0 . It can be shown that ([8], Ch. 5) $h(k_0)$ satisfies the following relation,

$$k_0 = \frac{1 - ((1-p)/p)^h}{(p/(1-p))^h - ((1-p)/p)^h}. \quad (27)$$

Following the analysis in [8] (Ch. 5) the average sample number as a function of the true value of u can be shown to be,

$$E_u(N) \approx \begin{cases} \frac{(1-L(u))A_1 - L(u)B_1}{2u-1} & \text{if } u \neq 0.5 \\ A_1B_1 & \text{if } u = 0.5 \end{cases} \quad (28)$$

With all these derivations in place we are now ready to embark on the attempt to estimate the length of a watermark from the embedder's perspective. It is to be noted that this estimate is not unique but depends on the type of watermark detector employed. The next section is devoted to estimating the embedded watermark length as a function of the detector and its parameters.

4 Detector-dependent Watermark Length Computation

As noted in the previous sections there seems to be a lack of scientific reasoning in choosing a watermark length in current literature. We understand that the longer the watermark, the higher the gain in terms of received signal to noise ratio at the detector. But, as discussed before, this gain could cause additional perceptual distortion to the host image due to redundant watermarking. In some other instances these choices for the watermark length can even be insufficient to guarantee a target detection error probability. Redundant watermarking—when watermark length is much more than required (for desired detection error rates) can also lead to wasted additional computational effort from the detectors point of view.

Since the embedded watermark length has an impact on the probability of detection, choice of a detector dependent watermark length makes much sense. For instance, if a watermark embedding algorithm exploits the knowledge that a sequential watermark detector is being used to achieve certain detection error rates then it can accordingly choose the total number of features to watermark. Since the optimal sequential detector requires the minimum average number of observations for detection, theoretically, at least choosing the watermark length to be equal to $\max_i E(N|H_i)$ number of pixels (or features) for a desired α and β may be a reasonable choice, *i.e.*, choose $|\Lambda| = \max_i E(N|H_i)$ if the proposed sequential watermark detector is used. But, note that such a choice of $|\Lambda|$ will only represent an average value for the watermark length. This means, there is a finite probability that this choice may be violated during detection. Therefore it may be necessary to choose more conservative estimates for the watermark length to minimize the probability of N exceeding the chosen watermark length. Such possible choices for the length are given by $E(N|H_j) + t\sigma_{N|H_j}$, $t = 1, 2, \dots, t_{max}$ where $j = \arg \max_i E(N|H_i)$ and $\sigma_{N|H_j}^2$ is the conditional variance of N . The value of t can be chosen based on how conservative we want the length estimate to be—computed as a confidence level using the probability distribution of N . If we choose such an estimate for the watermark length then it is not difficult to see that the $P(N > E(N|H_j) + t\sigma_{N|H_j})$ becomes sufficiently small even for reasonably small values of $t > 0$. This can be seen from the fact that this quantity represents a tail probability and by using Chebyshev inequality an upper bound on this probability can be obtained and shown to be vanishing rapidly with increasing t .

We notice that in order to compute the conservative estimate of the watermark length we have to compute the conditional variance of N . Of course this conditional variance will be different for the optimal and sub-optimal sequential watermark detectors. But, the sub-optimal detector requires a higher average sample number and hence we compute the watermark length estimate for this detector. This estimate will also be an upper bound for the watermark length computed for the optimal sequential detector. Since the test statistic for the sub-optimal detector is a discrete-time random walk with two absorbing boundaries the second moments of N can be shown to be [5],

$$\begin{aligned} E(N^2|H_1) &= V(p, q) \\ E(N^2|H_0) &= V(q, p) \end{aligned} \tag{29}$$

where

$$\begin{aligned} V(p, q) &= \frac{2^{B_1+1}}{A_1 + B_1} q^{B_1} \sum_{i=1}^{\lfloor (A_1+B_1)/2 \rfloor} \left(\frac{B_1^2}{1-\theta_i} \frac{4B_1\theta_i}{(1-\theta_i)^2} \frac{4\theta_i(1+\theta_i)}{(1-\theta_i)^3} \right) \cos^{B_1-1} \left(\frac{\pi i}{A_1 + B_1} \right) \sin \left(\frac{\pi i}{A_1 + B_1} \right) \times \\ &\sin \left(\frac{\pi B_1 i}{A_1 + B_1} \right) + \frac{2^{A_1+1}}{A_1 + B_1} p^{A_1} \sum_{i=1}^{\lfloor (A_1+B_1)/2 \rfloor} \left(\frac{A_1^2}{1-\theta_i} + \frac{4A_1\theta_i}{(1-\theta_i)^2} + \frac{4\theta_i(1+\theta_i)}{(1-\theta_i)^3} \right) \times \\ &\cos^{A_1-1} \left(\frac{\pi i}{A_1 + B_1} \right) \sin \left(\frac{\pi i}{A_1 + B_1} \right) \sin \left(\frac{\pi A_1 i}{A_1 + B_1} \right) \end{aligned} \tag{30}$$

and $\theta_i = 4pq\cos^2(\pi i/(A_1 + B_1))$. Clearly, the conditional variance of N can now be computed from the first and second moments as $\sigma_{N|H_j} = E(N^2|H_j) - [E(N|H_j)]^2$. In the next section we provide numerical evaluation of these analytical results and discuss their impact from a practical perspective.

5 Performance and Experimental Analysis

In this section we discuss the performance of the optimal and sub-optimal sequential detectors and the FSS detector to draw conclusions. For the sake of completeness we also briefly analyze a truncated sequential detector which is a compromise between a sequential detector and a FSS detector.

It is well known that when the desired decision error probabilities are very small $E(N)$ (and $|\Lambda_{FSS}|$) can be large. To avoid decision delays in such cases it is possible to truncate the decision making process after a certain fixed number of observations (say, N_1) and a terminal decision can be made. The trade-off in such a truncated sequential watermark detection process is a loss in the target error probabilities, *i.e.*, the constraint on the decision error rates can be violated as discussed later in detail. The truncation of the detection process is achieved by shaping the decision boundaries. One possibility for a truncated sequential detector is as follows:

$$\mathcal{L}(y_1, y_2, \dots, y_n) \begin{cases} \geq & \ln A \left(1 - \frac{n}{N_1}\right)^r & \text{decide } H_1 \\ \leq & \ln B \left(1 - \frac{n}{N_1}\right)^s & \text{decide } H_0 \\ \text{else} & n = n + 1 \end{cases} \quad (31)$$

Note here that the two decision boundaries converge when $n = N_1$ and the rate of convergence of these boundaries is controlled by the parameters $0 \leq r \leq 1$ and $0 \leq s \leq 1$ in Eq. (31). Larger values of these parameters means that the sequential watermark detection process will be truncated faster but with a higher loss in the desired decision error probabilities. If a final decision is not made before $n = N_1$ then the detector truncates the procedure and chooses the maximum likelihood hypothesis based on the N_1 observations. If $E_T(N|H_i)$, $i = 0, 1$ denotes the conditional average sample number for the truncated sequential watermark detector then by following the discussion in Section 2 we can see that,

$$\begin{aligned} E_T(N|H_1) &\approx \frac{E(N|H_1)}{1 + \frac{r}{N_1}E(N|H_1)} \\ E_T(N|H_0) &\approx \frac{E(N|H_0)}{1 + \frac{s}{N_1}E(N|H_0)} \end{aligned} \quad (32)$$

and the resultant error probabilities are given by,

$$\begin{aligned} \hat{\alpha} &= \alpha \left[1 + \frac{r \ln(A) E(N|H_1)}{N_1 + r E(N|H_1)} \right] \\ \hat{\beta} &= \beta \left[1 + \frac{s \ln(B) E(N|H_0)}{N_1 + s E(N|H_0)} \right] \end{aligned} \quad (33)$$

Now, for $r > 0$ and $s > 0$ the denominators in Eq. (32) are greater than 1 showing that $E_T(N|H_i) < E(N|H_i)$, $i = 0, 1$. Similarly, from Eq. (33) we see that $\hat{\alpha} > \alpha$ and $\hat{\beta} > \beta$. We do not present experimental results for the truncated watermark detector because it is along the lines of the optimal sequential watermark detector.

A FSS detector operates on a fixed sample size to compute a test statistic. This test statistic is then compared against a threshold to determine which of the hypotheses is true. We know that the Neyman-Pearson FSS detector [9] (Ch. 3) is optimum to detect a constant signal in Gaussian noise for a fixed false alarm probability. The fixed sample size (FSS) watermark detector using the model in Eq. (9) is given by,

$$\mathcal{L}(l_1, l_2, \dots, l_{|\Lambda_{FSS}|}) = \ln \left[\frac{f_1(l_1, l_2, \dots, l_{|\Lambda_{FSS}|})}{f_0(l_1, l_2, \dots, l_{|\Lambda_{FSS}|})} \right] \quad \begin{cases} \geq \tilde{T} & \text{decide } H_1 \\ < \tilde{T} & \text{decide } H_0 \end{cases}$$

which after algebraic manipulations gives,

$$\sum_{k=1}^{|\Lambda_{FSS}|} l_k \quad \begin{cases} \geq T & \text{decide } H_1 \\ < T & \text{decide } H_0 \end{cases}$$

for some optimally chosen threshold T . Then the sample size $|\Lambda_{FSS}|$ and T must satisfy,

$$\beta = \int_{-\infty}^T f_G(l - |\Lambda_{FSS}|K, |\Lambda_{FSS}|\sigma^2) dl \quad (34)$$

$$\alpha = \int_T^{\infty} f_G(l, |\Lambda_{FSS}|\sigma^2) dl \quad (35)$$

where $f_G(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$, $-\infty < x < \infty$. Eliminating T from Eq. (34) and Eq. (35) we get the following estimate for the watermark length when a fixed sample size detector is used:

$$|\Lambda_{FSS}| = \frac{2\sigma^2}{K^2} [erf^{-1}(1 - 2\alpha) - erf^{-1}(2\beta - 1)]^2 \quad (36)$$

where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

For experiments many different images were used. All these images produced results that agreed with the theoretical analysis. Here we present experimental results for a 256 gray valued Lenna (256×256) image watermarked using the embedding technique presented in Section 3 with embedding factor $K = 5$. Half the number of pixels were watermarked. Since we assume that the original Lenna image is not available an estimate was obtained from the watermarked image. Since the detector knows the watermark key, it approximates the original pixel value for a marked pixel by the nearest unmarked neighboring pixel. This estimate is then subtracted from the marked pixel and the difference forms an input to the detectors. From our experiments we found that the standard deviation of the estimation error (σ^2) was equal to 72 on an average for different watermark keys. This value can be further reduced by using more sophisticated prediction techniques. The computed detector test statistic $S(n)$ and the final decision are shown in Fig. 4. In another experiment the image was not watermarked and given as input to the detector. Then, as shown in Fig. 5 the detector test statistic hits the lower decision boundary and detects the absence of the watermark. The number of samples used by the optimal sequential detector in these two cases were nearly 3000 and 1800 respectively which is much less than 32768, the watermark length used by the Pitas [7] embedding algorithm. Therefore the watermark redundancy is 29768 and 30968 pixels for this experimental run.

The same set of experiments were conducted to test the sub-optimal sequential watermark detector and the corresponding detector test statistics are shown in Figs. 6 and 7. We again see that the detector makes the correct decision in both cases. One important fact to notice from these two figures is that the number of samples consumed by the sub-optimal sequential detector in these

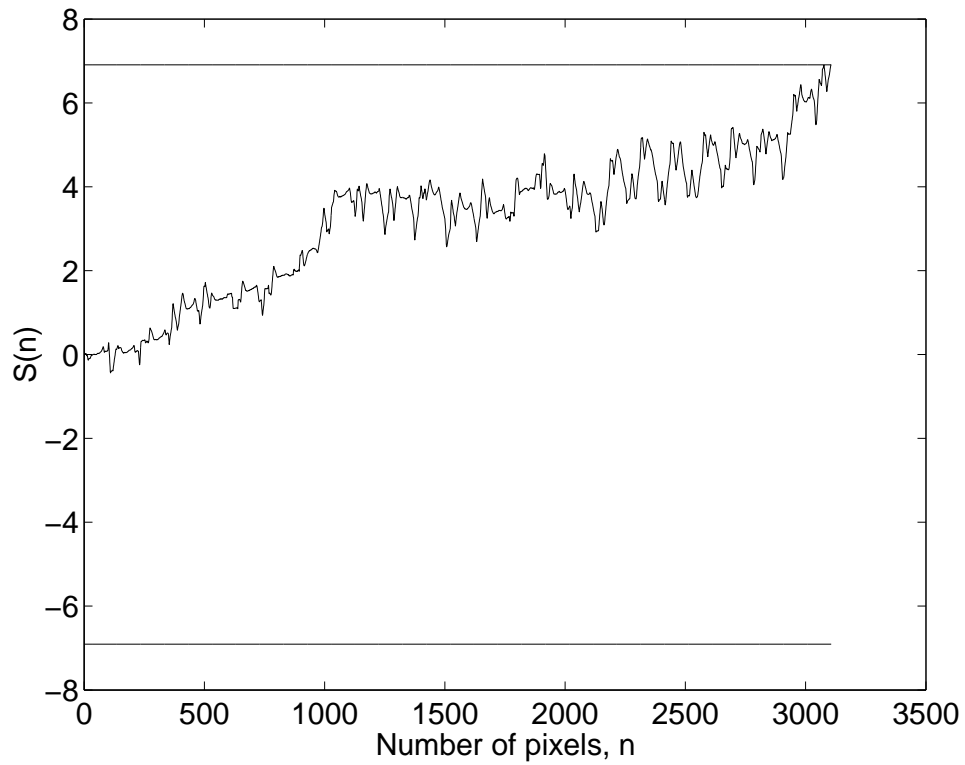


Figure 4: Detector test statistic corresponding to the optimal sequential detector. Watermark is detected correctly by the sequential detector with the test statistic reaching the upper decision boundary when $\alpha = \beta = 10^{-3}$.

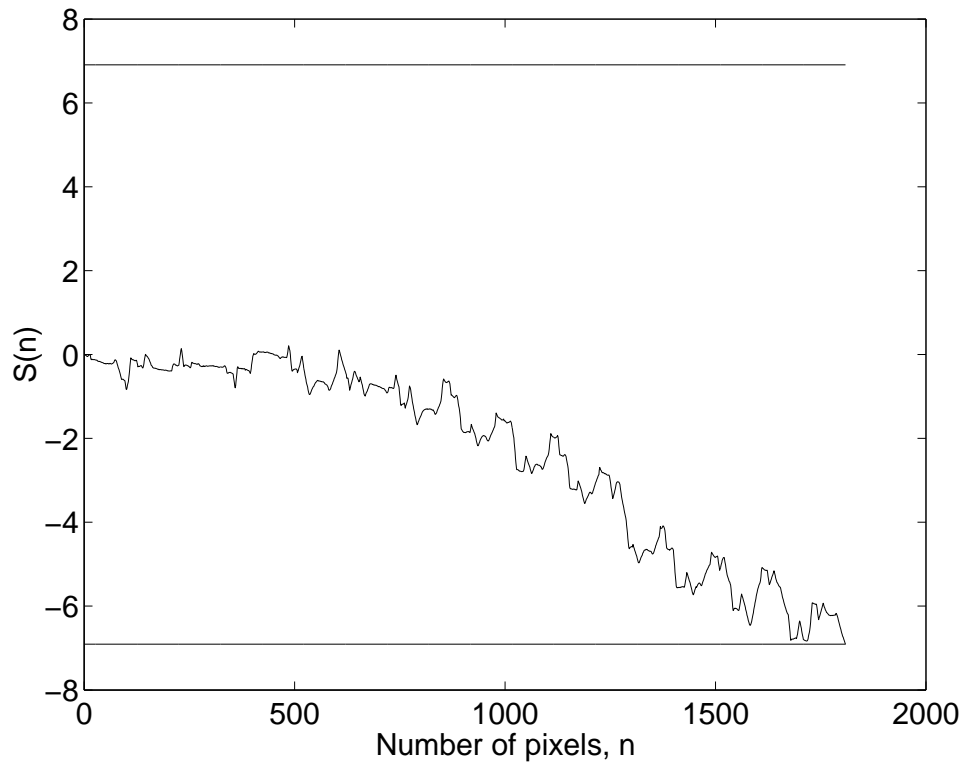


Figure 5: Detector test statistic corresponding to the optimal sequential detector. Absence of watermark is detected correctly by the sequential detector with the test statistic reaching the lower decision boundary when $\alpha = \beta = 10^{-3}$.

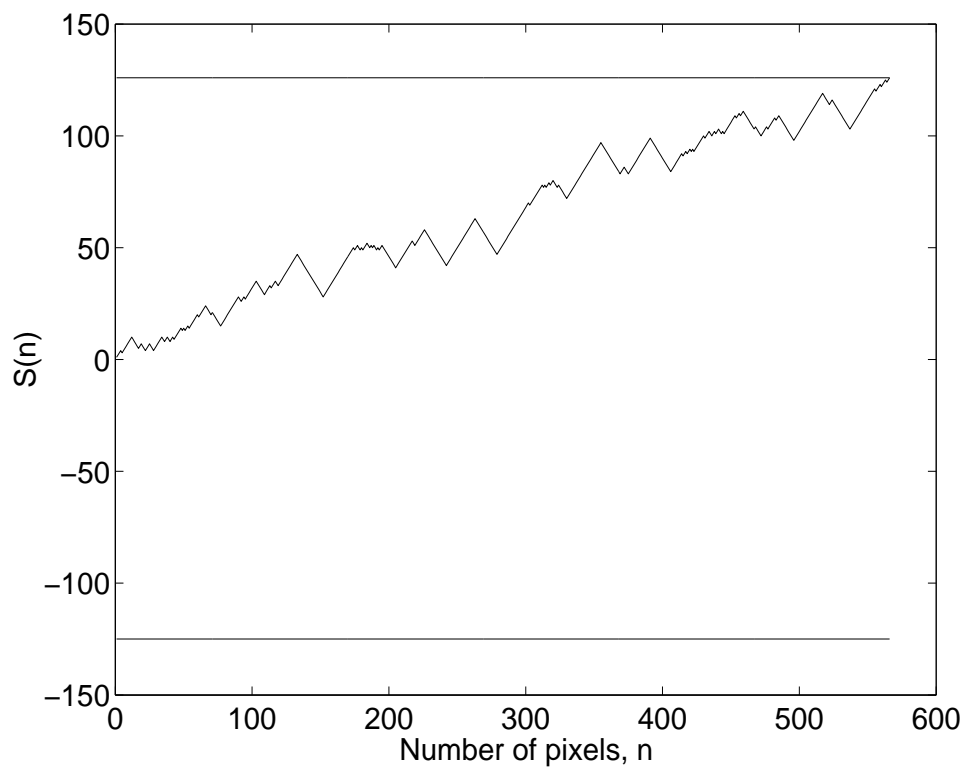


Figure 6: Detector test statistic corresponding to the sub-optimal sequential detector. Watermark is detected correctly by the sequential detector with the test statistic reaching the upper decision boundary when $\alpha = \beta = 10^{-3}$.

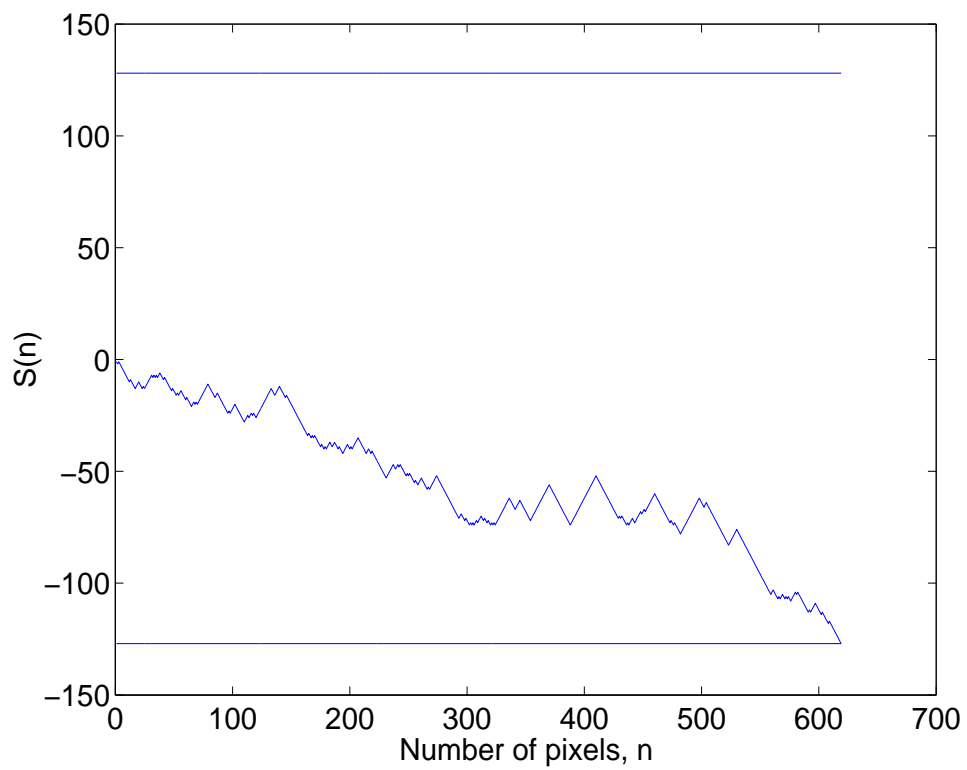


Figure 7: Detector test statistic corresponding to the sub-optimal sequential detector. Absence of watermark is detected correctly by the sequential detector with the test statistic reaching the lower decision boundary when $\alpha = \beta = 10^{-3}$.

Table 1: Comparison of average sample number for the proposed sequential watermark detectors and the fixed sample size detector.

$\alpha = \beta$	$E_{opt}(N H_1) = E_{opt}(N H_0)$	$E_{sub-opt}(N H_1) = E_{sub-opt}(N H_0)$	$ \Lambda_{FSS} $
10^{-3}	2858	4504	7920
10^{-4}	3818	6028	11472
10^{-5}	4774	7509	15087
10^{-6}	5729	9026	18741

Table 2: Estimate of watermark length for sub-optimal sequential detector.

$\alpha = \beta$	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Watermark length	6909	8827	10636	12454

two specific experimental runs is less than the corresponding numbers for the optimal detector. This is not surprising because the number of samples tested by a sequential detector is a random number. This is why average sample numbers serve as a good measure to consistently compare two sequential watermark detectors.

Table 1 gives a comparison of the sample numbers needed by the three detectors for different detection error rates. We see that there is a significant advantage in using the optimal sequential detector over the popular FSS detector. Even the sub-optimal detector results in almost 50% savings in the average number of samples required for detection compared to the FSS watermark detector. The optimal sequential detector seems to reduce the average sample number up to 70%. We found in our experiments that the theoretical and empirical estimates of the average number of pixels required by the detectors match reasonably well. Finally, observe that the watermark length computed according to [7] is 32768.

An estimate of the watermark length when the sub-optimal sequential watermark detector is employed is tabulated in Table 2. The estimate is given by $E(N|H_1) + \sigma_{N|H_1}$ since $E(N|H_1) = E(H_0)$ for equal α and β . We note from this table that this conservative estimate for the watermark length for the sub-optimal sequential detector is still less than that of the FSS watermark detector. We can expect that the gain in watermark length for the optimal sequential detector will be much more pronounced. Also, observe that $P(N > E(N|H_1) + \sigma_{N|H_1})$ is sufficiently small.

Figure 8 shows the average number of samples consumed by the optimal sequential detector when the true value of the watermark at the detector is k_0 and $\alpha = \beta = 10^{-3}$ and 10^{-4} . We explain the figure as follows. When the image is not watermarked and k_0 is close to zero, the detector quickly finds that there is no watermark resulting in a smaller average sample number. In the same way, when k_0 is nearly equal to 5, the sequential detector requires a smaller average number of samples for detection. However, when the received watermark value is $k_0 = 2.5$, we see that this value is mid-way between no watermark ($K = 0$) and the watermark's presence ($K = 5$). Therefore, this value of k_0 is the worst case input to the detector and hence it requires a large number of samples on an average before deciding on the outcome. This is seen as the peak in Figure 8. When the detection error probability constraint is stringent more samples are needed on an average by the detector. Similar observations and explanations hold for the sub-optimal

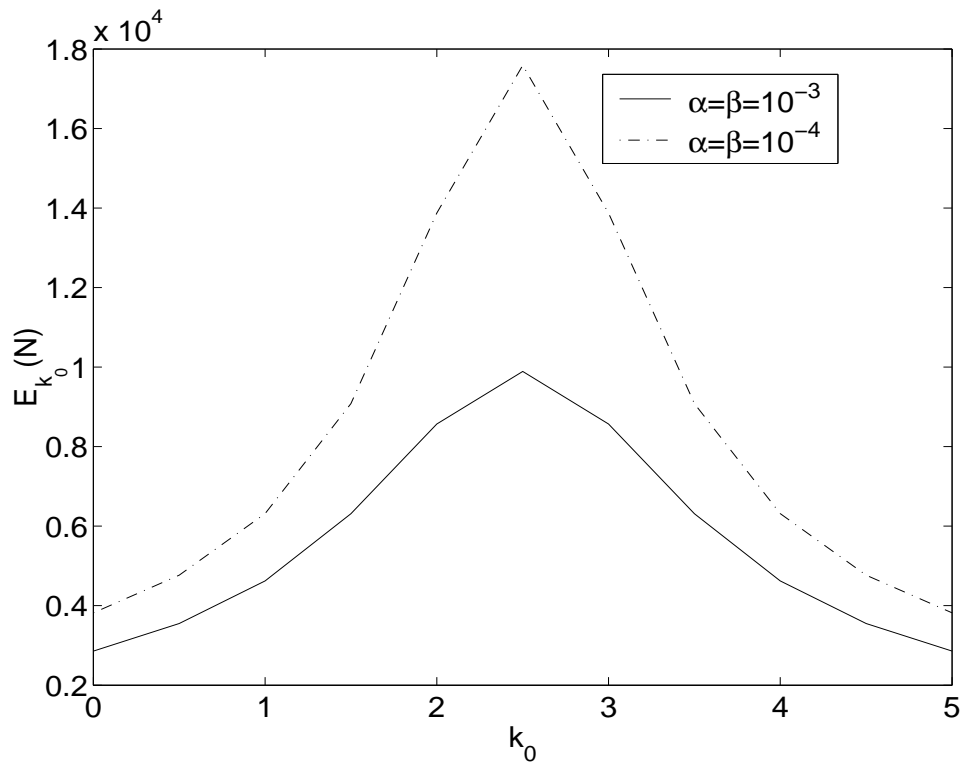


Figure 8: Average sample number for the optimal sequential watermark detector to test for non-design watermark embedding factor when $\alpha = \beta = 10^{-3}$ and 10^{-4} .

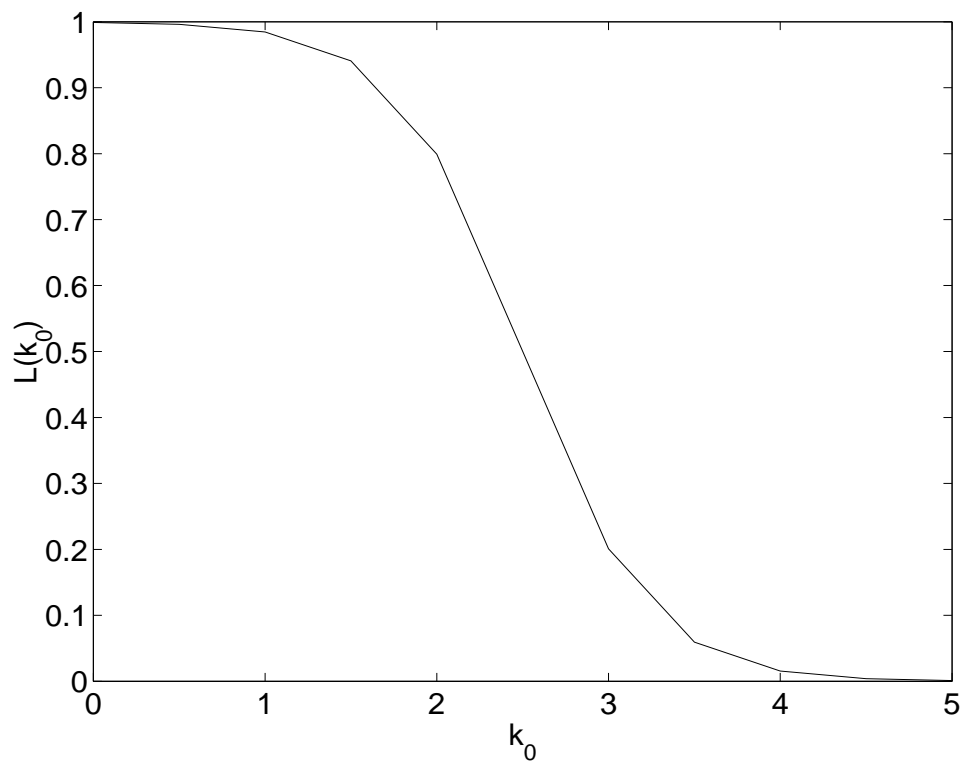


Figure 9: Operating characteristic curve for optimal sequential watermark detector.

sequential detector and therefore is not presented here.

Recall that points on the OC curve give the probability of accepting H_0 when k_0 is the true value of the watermark embedding factor. We see from Fig. 9 that the optimal sequential watermark detector accepts H_0 with high probability for a wide range of value of $0 \leq k_0 < 2.5$. When $k_0 = 2.5$ both H_0 and H_1 are accepted by the detector with equal probability (as expected). For $k_0 > 2.5$ the probability of accepting H_0 is significantly small as desired. Similar observations have been made for the sub-optimal sequential detector also.

6 Conclusions

A new paradigm for watermark detection called *sequential watermark detection* is proposed. Some fundamental questions are raised within this framework such as:

- Is it beneficial to jointly choose watermark embedding and detection parameters in some optimal sense?
- Is there a mathematical method that can lead to estimating the optimal watermark length?

These questions are answered using both theoretical and experimental approaches. It is shown that there is much to be gained by jointly optimizing a watermark embedder and the detector. We believe this paper could be the first attempt in rigorously formulating and solving this problem. It is shown that the proposed sequential watermark detectors outperform the traditional fixed sample detector in terms of the average number of samples required to detect the presence or absence of a watermark—a 70% reduction in average sample number can be achieved. Estimates for the embedding watermark length that depends on type of the detector and detection error probability constraints are given. It is observed that even a conservative estimate of the watermark length for the proposed sequential detector framework is much less than the corresponding number for the commonly used FSS based watermark detectors. On the whole, it appears that the proposed sequential watermark detection technique may be promising for many real-life applications where detection error probabilities need to be fully controllable and quick watermark detection may be necessary.

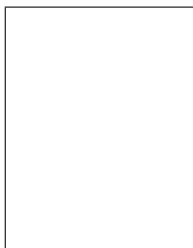
Acknowledgements

We thank the reviewers for their many useful comments that helped in improving the presentation of this paper. This material is based on research sponsored by Air Force Research Laboratory under agreement number F306020-02-2-0193. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

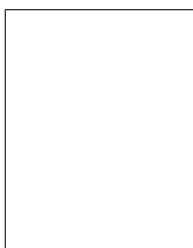
References

- [1] J.J. Eggers and B. Girod, “Quantization effects on digital watermarks,” *Signal Processing*, Vol. 81, no. 2, pp. 239-263, 2001.
- [2] URL: <http://www.digimarc.com>

- [3] Personal communication with Steve Decker, Digimarc.
- [4] R. Chandramouli, N.D. Memon, and M. Rabbani, *Digital Watermarking*, To appear in Encyclopedia of Imaging Science and Technology, 2002.
- [5] R. Chandramouli and N.D. Memon, "How many pixels to watermark?," Special session on Multimedia Content Protection, *Proc. IEEE International Conference on Information Technology: Coding and Computing*, Las Vegas, pp. 11-15, March 2000.
- [6] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3 & 4, 1996.
- [7] N. Nikolaidis and I. Pitas, "Robust image watermarking in the spatial domain," *Signal Processing*, vol. 66, pp. 385-403, 1998.
- [8] A. Wald, *Sequential Analysis*, NY, Wiley, 1947.
- [9] H.V. Poor, *An introduction to signal detection and estimation*, Springer Verlag, NY, 1994.
- [10] J. Killian, T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for images, audio and video," *Proceedings of the IEEE Int. Conf. on Image Processing*, pages 243-246, Lausanne, Switzerland, 1996.
- [11] S. Tantaratana and J.B. Thomas, "On sequential sign detection of a constant signal," *IEEE Trans. on Info. Theory*, May 1977, pp. 304-315.



R. Chandramouli. R. Chandramouli is an Assistant Professor in the Department of Electrical and Computer Engineering (ECE) at Stevens Institute of Technology. He co-directs the *MSyNC*: Multimedia Systems, Networking, and Communications research laboratory at Stevens. He received a Ph.D. in Computer Science and Engineering (1999) and M.A. in Mathematics (1999) from the University of South Florida, M.E. in Electrical Communications Engineering (1994) from the Indian Institute of Science, and B.Sc. in Mathematics (1990) from Loyola College, Madras. Prior to joining Stevens, he worked as an Assistant Professor in the Department of ECE at Iowa State University, Junior Research Fellow in the Department of Computing Science at University of New England, Australia, and the digital signal processing group in Motorola, India. His research interests include wireless networking, information security, low power VLSI, and applied probability. Since 1999 he has been serving as an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology. He is a recipient of the NSF CAREER award and the IEEE Richard E. Merwin award.



Nasir Memon. Nasir Memon is an Associate Professor in the Computer Science Department at Polytechnic University, New York. He received his B.E. in Chemical Engineering and M.Sc. in Mathematics from the Birla Institute of Technology, Pilani, India and received his M.S. and Ph.D. degrees from the University of Nebraska, in Computer Science. He has held an assistant professor position at Arkansas State University and at Northern Illinois University. He was also a visiting faculty at Hewlett Packard Laboratories, Palo Alto, California during 1997-98. His research interests include Data Compression, Data Encryption, Image Processing, Multimedia Data Security and Multimedia Communication and Computing. He has published more than a 100 articles in journals and conference proceedings. He was an associate editor for IEEE

Transactions on Image Processing from 1999 to 2002 and is currently an associate editor for the the ACM Multimedia Systems Journal.