Digital Video Steganalysis Exploiting Statistical Visibility in the Temporal Domain

Udit Budhia, Deepa Kundur, *Senior Member, IEEE*, and Takis Zourntos, *Member, IEEE*

***Abstract*—In this paper, we present effective steganalysis techniques for digital video sequences based on interframe collusion that exploits the temporal statistical visibility of a hidden message. Steganalysis is the process of detecting, with high probability, the presence of covert data in multimedia. Present image steganalysis algorithms when applied directly to video sequences on a frame-by-frame basis are suboptimal; we present methods that overcome this limitation by using redundant information present in the temporal domain to detect covert messages embedded via spread-spectrum steganography. Our performance gains are achieved by exploiting the collusion attack that has recently been studied in the field of digital video watermarking and pattern recognition tools. Through analysis and simulations, we evaluate the effectiveness of the video steganalysis based on linear collusion approaches. The proposed steganalysis methods are successful in detecting hidden watermarks bearing low energy with high accuracy. The simulation results also show the improved performance of the proposed temporal-based methods over purely spatial methods.**

***Index Terms*—Collusion attack, information forensics, pattern recognition, video steganalysis, video steganography.**

I. Introduction

THE purpose of steganalysis is to detect the presence of covert data within innocuous-looking media, called cover media, such as digital images or video. Steganalysis is an art of covert signal detection in which the signal in question has been "embedded" within another, often a more prominent, signal using steganography. A steganalyst may be passive (in which only the presence of a hidden message or the use of a particular steganographic algorithm is to be detected) or active (in which, additionally, some characteristic of hidden data, such as embedding location or length of the message is to be estimated) [1]. Steganalysis has gained attention in the fields of computer forensics [2] and homeland security [3], [4] in which threats of covert communications hold serious consequences. In addition, automated steganalysis techniques are effective in civilian applications to monitor Trojan horse programs, viruses, spywares, adwares, and other malicious data that may be hidden in digital media to adversely affect computer use [5].

Many practical steganalysis methods to date are designed to be passive. We briefly survey several techniques in this class. For instance, Fridrich *et al.* [6] propose a successful method to detect least-significant bit (LSB) embedding in 24-b color images by observing that the number of unique colored pairs decreases after LSB embedding. For JPEG images, it has been shown by Fridrich *et al.* [7] that steganalysis is possible by exploiting the unique fingerprints left by the JPEG quantization matrix. Methods based on first-order statistical analysis involving the Chi-square test on pairs of values by Westfeld and Pfitzmann [8] and the center of mass of the histogram classification function by Harmsen and Pearlman [9] have also been proposed. Provos [10] has subsequently demonstrated a way to design a steganography algorithm in which the associated first-order statistics are preserved making it necessary to employ higher-order statistics (HOS) tests for steganalysis.

For instance, Farid and his colleagues [11]–[14] have designed blind steganalysis methods that employ mean, variance, and HOS, such as skew and kurtosis to measure the disruption of statistical regularity due to steganography in wavelet coefficients of digital images. Linear and nonlinear classification methods, such as the Fischer linear discrimination and the support vector machine are applied. The low and HOS are believed to be rich enough to detect a broad class of steganography in digital images.

Image-quality metrics and multivariate regression analysis has also been proposed for steganalysis by Avcibas *et al.* [15], [16]. It is observed that the distance between image features carrying covert data and those of a filtered version is greater than for a purely cover image (with no hidden data). The most effective image quality metrics to measure this change in distance for a number of embedding methods are determined [17]. Part of the work proposed in this paper may be interpreted as being analogous to [15] in that temporal filters are applied to aid in steganalysis.

The steganalysis methods in [18] and [19] by Liu *et al.* focus on detecting wavelet-based steganography, popular due to the use of compression standards, such as JPEG2000. The parameters for a generalized Gaussian distribution to model the subband coefficients in a three-level wavelet decomposition of an image are calculated and then input into a trained neural network [19]. In another method by the authors [18], the energy of the wavelet coefficients is computed using the discrete Fourier transform (DFT), and the corresponding strength of the energy curve spikes is compared to a threshold.

A. Video Steganalysis, Watermarking Attacks, and Statistical Visibility

The data rate of covert data transmission using steganography is low in order to keep the covert data imperceptible within the

Manuscript received January 24, 2005; revised May 26, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hany Farid.

U. Budhia is with Mobilygen Corporation, Santa Clara, CA 95054 USA (e-mail: ubudhia@hotmail.com).

D. Kundur and T. Zourntos are with the Zachry Engineering Center, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA (e-mail: deepa@ece.tamu.edu; takis@ece.tamu.edu).

Digital Object Identifier 10.1109/TIFS.2006.885020

cover medium. This data rate is somewhat proportional to the volume of the cover medium and, for this reason, digital video is a convenient choice for moderate rate steganography. The temporal domain provides fertile ground for embedding higher volumes of covert data by exploiting the temporal masking characteristics of the human visual system (HVS). However, at the same time, the redundancy in this domain allows greater opportunity for steganalysis.

To the best of the authors' knowledge, the steganalysis methods proposed to date (with the exception of the authors' own work in [20]) do not apply to digital video. Although raw video streams are essentially sequences of images, the application of image steganalysis to video on a frame-by-frame basis can result in suboptimal solutions. Given the need for automated tools to monitor widespread steganography [21], we address the problem of passive steganalysis of digital video. This research generalizes, extends, and provides thorough analytic justification and simulations to [20].

Steganalysis tools have been developed over the years, forming a library of tests that can be used to flag suspicious communication. Many of these methods are reactive and, thus, are designed to address steganalysis for a specifically developed steganographic algorithm. In this work, we develop a methodology that leverages toolsets from research fields of watermarking attacks and pattern recognition. In this way, there is a natural means to more proactively develop steganalysis tools by leveraging the rapidly evolving field of watermarking attacks. In addition, good watermark attack operators are attractive features for steganalysis because they demonstrate the following characteristics: 1) they successfully remove hidden data if present in test media and 2) leave the test media essentially intact if hidden data do not exist. Thus, one can argue that some difference measure between an original test media and an attacked version is different in cases when steganography has and has not taken place providing a detection feature for steganalysis.

For video steganalysis, we make use of "statistical visibility" in the temporal domain as studied in [22] and [23] in order to assess the usefulness of temporal correlations for steganalysis. Linear collusion has been proven to be an effective and efficient attack operator for video-embedding algorithms exhibiting temporal statistical visibility. Thus, we focus on developing an approach that makes use of collusion for video steganalysis. There is a tradeoff between the detection accuracy and the applicability of steganalysis to a broad class of embedding algorithms. The inspiration for this work is drawn from a number of currently available steganalysis techniques aimed at detecting hidden messages from a variety of embedding schemes [9], [11], [12], [16].

B. Contributions of this Paper

In this paper, the following occur.

1) We propose a general framework for developing steganalysis methods that exploit advancements in watermark attack research and pattern recognition. The goal is to provide a general methodology to produce a library of timely steganalysis algorithms that are suited for a class of applications.
2) Using this framework, we design efficient steganalysis techniques for video sequences that take advantage of temporal redundancy. We develop composite methods that can be used to detect messages hidden using spread-spectrum steganography in the spatial as well as the frequency domain.
3) We highlight the limitations of data hiding in video. We assert that the chances of the detection of hidden messages greatly improve due to the presence of temporal redundancy in video. We show, with analytic arguments and simulations, that it is infeasible to hide data in those parts of video that are nonmoving or have translational motion.
4) We study the tradeoff between "statistical invisibility" and robust embedding of hidden messages in a video sequence. Through analysis and simulations, we show the lower bounds on the embedding strengths of the hidden message that leads to the failure of the proposed steganalysis method.

In the next section, we discuss the nomenclature that we employ in this paper and formally define the video steganalysis problem. In Sections III and V, we discuss the collusion-based steganalysis approach and its enhancements. Comparisons and simulations are provided in Section VI followed by conclusions and insights in Section VII.

II. Preliminaries

A. Nomenclature

A steganographic system involves two parties: the sender who embeds the secret message in the cover media to produce the stego media and the receiver who extracts it. Security comes, in part, from the presence of a symmetric secret key K in the system that details how the secret message is embedded and extracted. We assume that K is securely exchanged between the sender and receiver prior to covert communication; this key is particular to the steganography algorithm and may impose specifics such as how strongly and where in the cover object the secret information is embedded, and/or seed information for pseudorandom number generation.

Our video steganographic system scenario is summarized in Fig. 1. The sender takes the "host" video sequence called the cover video and embeds a secret binary message vector using K to produce a stego video sequence that is perceptually identical to the cover video. The stego video is then communicated along a public channel to the receiver. At the receiver, the stego-object and secret key K are used to extract the secret binary message. The public channel may be monitored by an active or a passive steganalyst whose goal is to detect the presence of covert communication.

The cover video is denoted by $U_k(m,n)$ where $1 \leq k \leq N$ is the frame number and m,n are the row and column indices of the pixels, respectively. The binary secret message is embedded into the host by modulating it into a signal we call the watermark [24] denoted by $W_k(m,n)$. Since the influence of the secret message is carried on to the watermark, we will use the terms hidden message and watermark interchangeably throughout this paper. Detection of the watermark will imply the presence of hidden information in the medium. For notational compatibility,

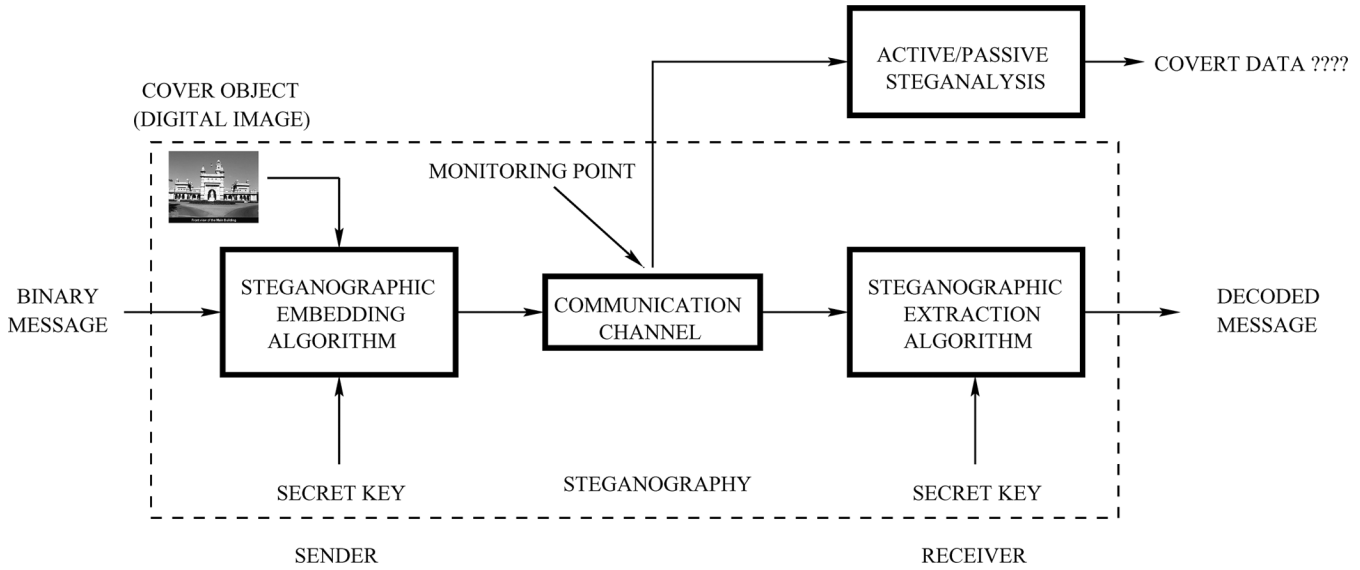


Fig. 1. Steganography and steganalysis.

the watermark $W_k(m,n)$ is defined over the same domain as the host $U_k(m,n)$; the reader should note that this holds even if the watermark is inserted in a nonspatial domain such as the discrete cosine transform (DCT). The stego-video signal is represented by the commonly used equation [22]

$$X_k(m,n) = U_k(m,n) + \alpha_k(m,n) \cdot W_k(m,n), \quad k = 1,2,3 \ldots N \quad (1)$$

where $\alpha_k(m,n)$ is a scaling factor used to manipulate the strength of the hidden message to tradeoff perceptibility and robustness. For simplicity of analysis, α is considered to be constant over all of the pixels and frames to give

$$X_k(m,n) = U_k(m,n) + \alpha \cdot W_k(m,n), \quad k = 1,2,3 \ldots N. \quad (2)$$

The scaled watermark $\alpha \cdot W_k(m,n)$ is a function of the binary hidden message, secret key K, and the host $U_k(m,n)$. The relation between these parameters is decided by the embedding algorithm.

B. Problem Formulation and Assumptions

One goal of this paper is to evaluate the importance of exploiting temporal correlations for video steganalysis. Thus, we first focus on video processing in the temporal domain; image methods that work in the orthogonal spatial domain can then be easily incorporated to enhance performance. Another goal is to develop a proactive steganalysis framework that applies to a larger genre of steganography algorithms or cover media and which exploits the actively evolving field of digital watermarking attacks. Thus, we focus on the steganalysis of Gaussian spread-spectrum-based steganographic methods [24], [25] due to its influence in the research literature.

Fig. 2 summarizes the basic framework. Two essential blocks are present: a watermarking attack stage used to estimate the host media from the possibly watermarked media, and a pattern recognition stage used for the detection of steganographic

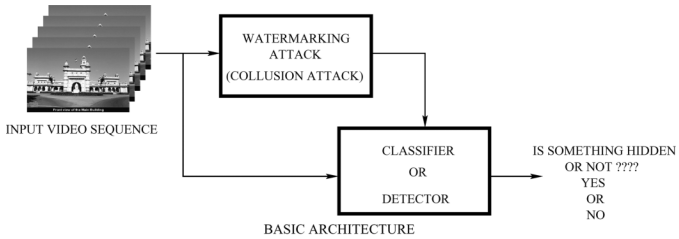


Fig. 2. Video steganalysis framework.

activity. Different algorithms can be substituted for each block to produce steganalysis techniques for a variety of applications. Instead of borrowing from libraries of image processing and statistical functions to identify potential primitives for steganalysis, the block-based structure also allows one to borrow from recent advancements in the related field of watermark attacks and pattern recognition for a timely steganalysis algorithm.

We make the following necessary assumptions.

1) The host frames U_k[1] are assumed to be from a distribution having mean μ and variance σ_u^2.
2) The correlation among the host frames follows the first-order Markov model where the correlation coefficient between frame U_i and U_j is given by $\rho^{|i-j|}$, and ρ is the correlation coefficient between any two adjacent frames.
3) The watermark frames W_k are assumed to be independent of U_k and of each other, and derived from a Gaussian distribution having mean 0 and variance σ_w^2. Since the watermark is embedded with an embedding strength of α, the effective variance of the watermark is $\alpha^2\sigma_w^2$.

For slow-moving sequences, we can assume that the frames have approximately the same mean and variance as stated in Assumption 1). If the mean of the video frames is not constant, it can be estimated on a frame-by-frame basis and subtracted out to produce a zero-mean video sequence for processing. In

[1]Please note that we have removed the subscripts m, n from our notations for clarity. For the rest of this paper, we will assume that all operations are done on the entire frame unless stated otherwise.

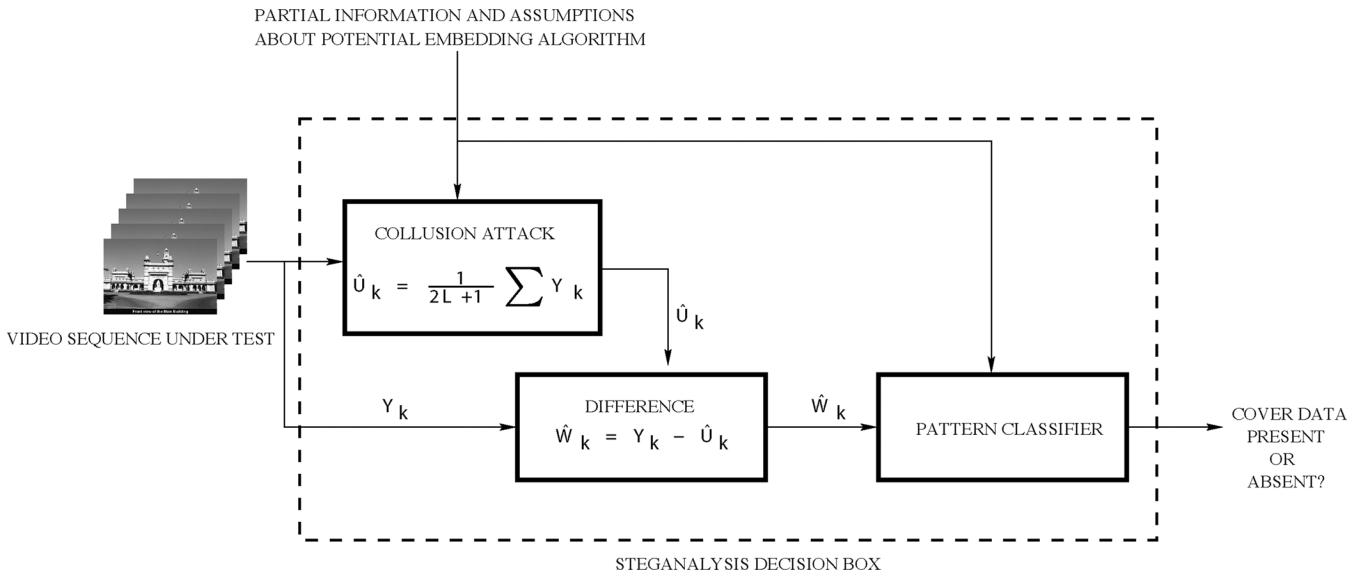


Fig. 3. Proposed framework for steganalysis.

Assumption 2), we employ the first-order Markov correlation model commonly found in the literature [26]. Assumption 3) directly comes from the structure of higher capacity spread-spectrum steganography methods where the watermarks embedded in each frame are from a zero mean Gaussian distribution and are independent from each other and the host frames [27]. The interframe independence may suggest higher capacity embedding since the same payload is not restricted to be repeated throughout the video sequence and can change from frame to frame. This assumption also applies to zero mean Gaussian watermarks embedded in other linear domains such as the DCT.

We also assume that the sender embeds a watermark into each pixel/coefficient of every frame of the video sequence; this assumption is reasonable because to maximize the steganographic capacity, a sender must make use of as much of the host signal as possible for information embedding. There is, however, a tradeoff between steganographic security and transmission capacity as we discuss later.

The figures of merit used to assess success of the algorithm are the probability of false positive detection and the probability of false negative detection defined as follows. The probability of false positive detection is the likelihood of detecting that hidden information is present in a given video sequence when nothing has been embedded (i.e., $\alpha = 0$); that is, a given video signal is flagged stego video when it is not. The probability of false negative detection is the likelihood of detecting that hidden information is not present when, in fact, it has been embedded (i.e., $\alpha \neq 0$); that is, a given video signal is declared cover video when it is not. A good steganalysis technique should strive to minimize both error probabilities. Some research prioritizes false negative detection rates using the philosophy that a false positive can be subsequently corrected using further offline video processing techniques. However, as pointed out by two reviewers of this paper, due to the low numbers of video that carry covert information, it is impractical to sacrifice false positive probabilities and a good balance between these error rates must be achieved for practical application.

III. Collusion-Based Steganalysis

The spirit of most steganalysis methods is to devise a function that differentiates between the general characteristics of a signal with and without embedding. This function is normally compared implicitly or explicitly to a threshold in order to decide whether a given signal Y_k contains hidden information. Much research on image steganalysis has focused on identifying image features that change when steganography algorithms are applied. Researchers have traditionally employed image processing and statistical toolsets that in some form attempt to estimate a potential host $\hat{U}_k = \mathfrak{H}[Y_k]$ signal from Y_k. This host estimate $\hat{U}_k$ is then compared in some way to Y_k in order to detect if something is hidden. The basic hypothesis is that the deviation of specific characteristics of Y_k and $\hat{U}_k$ will differ if something is embedded in Y_k (i.e., $Y_k = X_k = U_k + \alpha \cdot W_k$) in comparison to when nothing is embedded in Y_k (i.e., $Y_k = U_k$). Pattern classification is often employed to characterize this deviation effectively.

In this section, we specify the design of the blocks in the basic architecture proposed for video steganalysis in Section II-B. Fig. 3 presents our framework. The video sequence under consideration Y_k is passed through a digital watermarking attack block (in this case, a temporal collusion attack) that attempts to estimate the host signal to produce $\hat{U}_k$. This block may assume knowledge of the embedding algorithm (if any is used) to be effective. The estimate of the watermark $\hat{W}_k$, calculated by taking the difference between Y_k and $\hat{U}_k$, is passed through an appropriate pattern classifier. If Y_k is a stego video, then the input to the pattern classifier is a watermark signal corrupted by some interference related to temporal filtering from the watermark attack. On the contrary, if Y_k is a pure video signal without hidden data, the estimate $\hat{W}_k$ will only consist of the noise due to filtering. In an ideal case, if the filter is able to perfectly reconstruct the host, the estimate $\hat{W}_k$ will consist of the embedded watermark in case of a stego video or will be zero otherwise. By employing some *a priori* information about the embedding algorithm, the distinction between these two cases can be made to detect the presence of covert communication.

We conjecture that the linear collusion attack, used to remove the presence of independent digital watermarks in a sequence of images or video frames [22], is ideal to address our goals. First, the attack focuses on temporal correlations between video frames to estimate a host video sequence that can be easily incorporated into our framework. Second, much analytic and simulation-based work [22], [23], [28], [29], [27] focuses on this area, providing a strong foundation upon which to build a steganalysis method. Finally, the attack is computationally simple making our steganalysis approach practically feasible for real-time applications.

A. Collusion

Collusion for digital watermarking and steganography refers to the use of multiple image frames (that may or may not form a video sequence) to remove the presence of a watermark in one or more of the frames. In general, the collusion may be linear or nonlinear exploiting the differences and similarities among frames to judiciously reduce the energy of the watermark in relation to that of the host information. We represent the collusion of a sequence of video frames as

$$\hat{U}_k = \mathfrak{C}_P(Y_k) = C[Y_1, Y_2, \ldots, Y_n] \tag{3}$$

where $\hat{U}_k$ is called the colluded result and in this paper represents the estimate of the kth host frame U_k. $\mathfrak{C}_P$ is the collusion operator with parameters P (which, in this paper, represents the collusion window length) that exploits the similarities and differences among all or a select subset of possibly watermarked image frames $Y_1, Y_2, \ldots, Y_N$ to produce $\hat{U}_k$. As we discuss, the colluded result $\hat{U}_k$ will ideally contain a significantly less contribution from W_k compared to Y_k. Common forms of the collusion operator $\mathfrak{C}$ include taking the pixel-by-pixel maximum, minimum, mean, or median over a range of image frames [22], [28], [29].

Linear collusion is a special case in which $\mathfrak{C}_P$ represents a weighted average operation of select video features and frames. Intuitively, linear collusion on a sequence of video frames amplifies parts of the frames that are similar and attenuates components that are different. In the next subsection, we consider the case where the collusion weights applied to each frame are equal. For the remainder of this paper, we refer to this as the simple linear collusion scheme.

B. Simple Linear Collusion

Let us assume that we use a sliding window to denote the temporal neighborhood used for frame averaging; this window is assumed to contain visually similar frames. Specifically, we take a window size of $2L + 1$ frames centered at frame k to average the video sequence. The estimate of the kth host frame is given by

$$\hat{U}_k = \mathfrak{C}_L(Y_k) = \begin{cases} \frac{1}{2L+1}\sum_{i=1}^{2L+1} Y_i, & 1 \le k \le L \\ \frac{1}{2L+1}\sum_{i=k-L}^{k+L} Y_i, & L < k \le N-L \\ \frac{1}{2L+1}\sum_{i=N-2L}^{N} Y_i, & N-L < k \le N \end{cases} \tag{4}$$

where k is the frame under consideration to produce $\hat{U}_k$, an estimate of U_k. The first and the third cases of (4) account for edge effects of the window moving out of the range $1 \le k \le N$.

The effectiveness of $\hat{U}_k$ as an approximation of U_k depends on the value of L in relation to the rate of motion in the video sequence. Through analysis, we show that an optimum value of L will lead to the cancelation of the Gaussian watermarks and ensure the assumption that $\hat{U}_k \approx U_k$ holds true.

If collusion is applied to a given video sequence Y_k that may or may not contain a watermark, we believe that in both cases for slowly varying video and an appropriately selected value of L, the result will be an approximation of U_k. Thus, if a watermark is embedded in the video, subtracting $\hat{U}_k$ from Y_k gives $Y_k - \hat{U}_k \approx Y_k - U_k = \alpha W_k$, an estimate of the scaled zero mean Gaussian watermark. If no watermark is present in Y_k, then the result will be independent of any characteristics such as Gaussianity that we assume for the watermark. This difference is exploited by a pattern classifier discussed in the next section for steganalysis.

C. Classification

Our objective is to build a classifier that discriminates between an estimate of the scaled watermark and no watermark. The two main components of a typical classifier are feature extraction and the discriminator [30]. Feature extraction derives characteristics from the signal under consideration to provide relevant information to the discriminator for classification. Since we assume that steganography occurs through the addition of Gaussian watermarks, we employ features that can measure the level of Gaussianity in a signal. These include kurtosis, entropy, and the 25th percentile.

Kurtosis [31] is a value that partially measures the "shape" of a distribution. Kurtosis for a Gaussian distribution is 3 and varies for most of the other distributions. The kurtosis estimate for the sample set $\{x_i\}$ is defined as

$$\text{Kurtosis} = \frac{1}{\sigma^2 N}\sum_{\forall i}(x_i - \mu)^4 \tag{5}$$

where σ and μ represent the variance and mean of the distribution.

Entropy [31] helps to determine the degree of "randomness" in a given distribution. For a fixed variance, the Gaussian distribution has maximum entropy. Thus, the estimates obtained from a watermarked video sequence should have a higher entropy than those obtained from nonwatermarked sequences. The entropy estimate is given by

$$\text{Entropy} = -\sum_{\forall i}(p_X(i)\log(p_X(i))) \tag{6}$$

where $p_X(i)$ is an estimate of the distribution of $\hat{W}_k$.

The last feature that we consider is the 25th percentile of a given distribution defined as the value above which 25% of the points in the histogram reside.

Once the features are extracted, we build a kNN classifier [30], [32]. More sophisticated classifiers using support vector machines and neural networks [32] could have been employed

for discrimination, but are higher in complexity without providing significantly improved performance from our preliminary tests. The kNN classifier must be trained to be able to operate for steganalysis. Cross validation [30], [32] is employed to determine the video set which yields the lowest probability of false positive and false negative.

IV. Design and Development

This section provides analysis and intuitive explanations to justify the choice of steganalysis building blocks and parameters.

A. Effectiveness of Simple Linear Collusion

Linear collusion has recently received analytic and experimental attention in the digital video watermarking community [22], [23], [27]. It has been shown analytically that if the linear correlation among host video frames U_i for some i differs from that of the watermark frames W_i over the same range of i, then linear collusion will be successful in either attenuating or amplifying the presence of the watermark in the resultant frame $\hat{U}_k$ [22].

We focus on the application of higher capacity spread-spectrum steganography on video sequences that implies W_i is independent for each frame (as discussed in Section II-B). We assume that the motion in the video sequence is "slow" which implies that adjacent video frames are similar (as presented by our correlation model of Section II-B). Because of this visual correlation, it is expected that over a neighborhood of i centered at k, the watermarked video frames can be averaged in order to attenuate the presence of the watermark in the kth frame.

Substituting $X_i = U_i + \alpha \cdot W_i$ for all i from (2) into (4), we obtain

$$\hat{U}_k = \frac{1}{2L+1}\sum_i U_i + \frac{\alpha}{2L+1}\sum_i W_i \tag{7}$$

where the summations are over the appropriate domains for the various ranges of k shown in (4). Since the watermarks W_i are independent and zero mean, the second term of the left-hand side of (7) approaches zero as L increases. Furthermore, because we assume $U_i \approx U_k$ for all i in the neighborhood of the sliding window centered at k, the first term will dominate resulting in the approximation $\hat{U}_k = U_k$. We note that choosing an appropriate window size, the colluded frame is a good estimate of the host frame.

The success of the steganalysis method, which leverages the watermark characteristics, depends on the success in estimating the watermark in each frame. The estimate of the scaled watermark is given by

$$\begin{aligned}\hat{W}_k &= Y_k - \hat{U}_k = Y_k - \mathfrak{C}_L(Y_k)\\ &= U_k + \alpha W_k - \mathfrak{C}_L(U_k + \alpha W_k).\end{aligned} \tag{8}$$

In case of nonwatermarked video, $Y_k = X_k = U_k + \alpha W_k$, where $\alpha = 0$. Therefore

$$\hat{W}_k = U_k - \mathfrak{C}_L(U_k) \tag{9}$$

since $\alpha = 0$ for nonwatermarked sequences. For simplicity of notation, we let $n_k = U_k - \mathfrak{C}_L(U_k)$.

This residual "noise" from simple linear collusion, represented by n_k, is a measure of the invariance of the collusion operator on legitimate nonwatermarked data. Ideally, we would like $\mathfrak{C}_L(U_k) \approx U_k$.

In case of watermarked sequences, $Y_k = X_k = U_k + \alpha W_k$. Since $\mathfrak{C}_L(a+b) = \mathfrak{C}_L(a) + \mathfrak{C}_L(b)$, the estimate of the scaled watermark is given by

$$\begin{aligned}\hat{W}_k &= U_k - \mathfrak{C}_L(U_k) + \alpha W_k - \alpha \mathfrak{C}_L(W_k)\\ &= n_k + W'_k\end{aligned} \tag{10}$$

where we let $W'_k = \alpha(W_k - \mathfrak{C}_L(W_k))$.

In the case of the watermarked sequences, the estimate of the watermark is the sum of the noise due to collusion and a Gaussian signal which bears a high correlation with embedded watermark W_k. In the situation when all of the host frames are the same, n_k will be zero and the estimate of the watermark $\hat{W}_k$ will be the embedded watermark W_k.

We study the expected mean squared error (MSE) between the estimate of the watermark $\hat{W}_k$ and the embedded watermark αW_k to find conditions for which simple linear collusion will be successful in extracting the watermark from the original frames. The expected MSE is given by

$$\begin{aligned}\mathbf{E}[(\hat{W}_k - \alpha W_k)^2] &= \mathbf{E}[(Y_k - \hat{U}_k - \alpha W_k)^2]\\ &= \mathbf{E}[(U_k + \alpha W_k - \hat{U}_k - \alpha W_k)^2]\\ &= \mathbf{E}[(U_k - \hat{U}_k)^2].\end{aligned} \tag{11}$$

Proposition 1: Given a sequence of watermarked video frames $X_k, k = 1, 2, \ldots, N$, as defined by (2). Under Assumptions (1), (2), and (3), the expected MSE between the original watermark and the estimated watermark obtained from simple collusion is given by

$$\begin{aligned}\mathbf{E}[(\hat{W}_k - \alpha W_k)^2] = \sigma_u^2 \Bigg[&\frac{z-1}{z} - \frac{2\rho}{z(1-\rho)}\\ &+ \frac{4\rho^{\frac{z+1}{2}}}{z(1-\rho)} - \frac{2\rho(1-\rho^z)}{z^2(1-\rho)^2}\Bigg] + \frac{\alpha^2\sigma_w^2}{z}\end{aligned} \tag{12}$$

where $z = 2L + 1$.

Proof: See Appendix, Sec. A.

In the next proposition, we evaluate the MSE when $L = 0$, that is, no temporal collusion is applied.

Proposition 2: Under assumptions 1), 2), and 3), the expected MSE between the original watermark and the estimated watermark when there is no temporal collusion is given by

$$\mathbf{E}[(\hat{W}_k - \alpha W_k)^2] = \alpha^2\sigma_w^2. \tag{13}$$

Proof: See Appendix, Sec. B. Since the estimate of the watermark when $L = 0$ is always zero, the expected MSE between the watermarks is always equal to the variance of the effective watermark embedded (i.e., $\alpha^2\sigma_w^2$).

The next proposition analyzes when collusion is successful for steganalysis. We study the ratio of the variance of the host

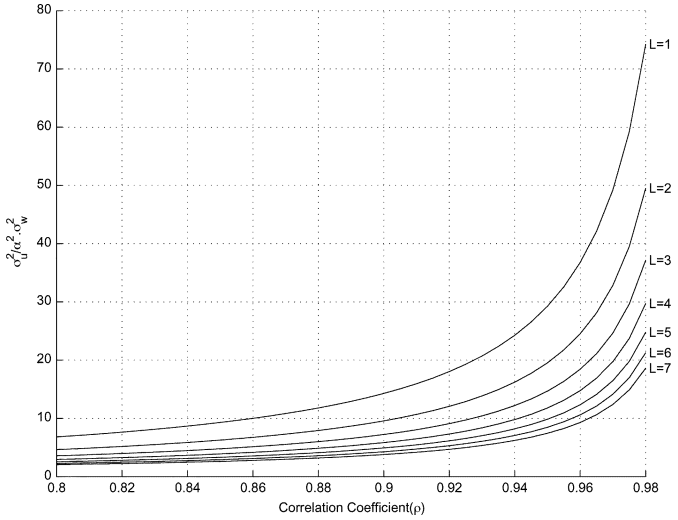


Fig. 4. Upperbound on $(\sigma_u^2)/(\alpha^2\sigma_w^2)$ as a function of ρ, L.

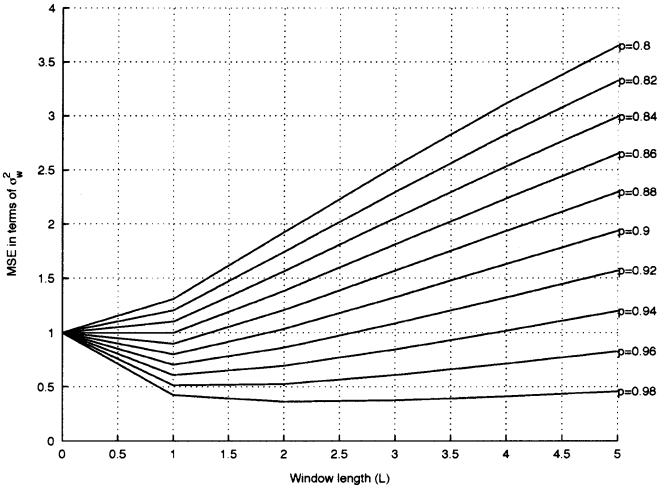


Fig. 5. MSE as a function of L, ρ.

to the variance of the embedded watermark, an inverse measure of watermark-signal-to-noise ratio (WSNR) often denoted "host-to-watermark ratio."

Proposition 3: Under assumptions 1), 2), and 3), the collusion attack is successful if the following condition is satistfied:

$$\frac{\sigma_u^2}{\alpha^2\sigma_w^2} < \frac{1}{1-\frac{2\rho}{(z-1)(1-\rho)}+\frac{4\rho^{\frac{z+1}{2}}}{(z-1)(1-\rho)}-\frac{2\rho(1-\rho^z)}{z(z-1)(1-\rho)^2}} \tag{14}$$

where $z = 2L + 1$.

Proof: See Appendix, Sec. C.

We arrive at the bounds on the above ratio by laying a constraint that the expected MSE of the watermark estimate in case of simple linear collusion is smaller than the expected MSE encountered when there is no collusion at all. The accuracy of steganalysis is related to two issues: 1) the ability of collusion to reduce the watermark strength (which increases reliability) and 2) the degree of residual noise present from temporal collusion (which decreases reliability). This constraint ensures issue 1) outweighs 2) so that the overall effect of collusion is successful in reducing MSE (and, hence, aids in steganalysis). The associated inequality of (14) provides insights on the conditions under which the proposed collusion-based method is successful.

Fig. 4 graphs the upperbound of $(\sigma_u^2)/(\alpha^2\sigma_w^2)$ as a function of ρ for various L, which relates a range of window lengths $z = 2L + 1$ for successful collusion to the similarity between video frames and the strength of the watermark, if present. For example, we see that for a correlation coefficient of $\rho = 0.94$ between adjacent frames and $L = 4$, the maximum $(\sigma_u^2)/(\alpha^2\sigma_w^2)$ can be 10. This means that if the variance of the host is greater than 10 times the effective variance of the watermark, a collusion length L of 4 (or less) will provide benefits for steganalysis in estimating the watermark. Although Fig. 4 provides us with an idea of when the collusion approach to steganalysis holds promise, it does not, however, provide information about the optimal value of L to produce the best estimate of the watermark.

The reader should note that practical simulations discussed in Section VI suggest that the bound of Fig. 4 is conservative. Inverse WSNR values can often be in the order of thousands implying that the value of ρ needs to be close to 1 for collusion to be effective for steganalysis. However, our simulations demonstrate that our collusion-based steganalysis works for the majority of video test sequences (even those with high inverse WSNRs and a lower value of ρ). We believe that this discrepancy is due, in part, to the use of the Markov model for video frame correlation. This well-known model is used to provide a tractable series of analytic design insights. However, after studying the model and comparing it to practical data, we believe the model provides a conservative assessment for steganalysis performance because the correlation between frames does not drop as quickly as the model suggests in many test sequences. We may, at times, be able to interpret this as the existence of a high value of ρ (above 0.98), which is variable. In addition, the use of the kNN classifier stage (which is not modeled in the formulation of Fig. 4) provides an additional level of robustness to the technique when applied to real data.

Fig. 5 displays the expected watermark estimate MSE in terms of $\alpha^2\sigma_w^2$ (i.e., $\mathbf{E}[(\hat{W}_k - \alpha W_k)^2]/\alpha^2\sigma_w^2$) computed by (12) as a function of L and ρ assuming $(\sigma_u^2)/(\alpha^2\sigma_w^2) = 10$. We see that for a given $\rho = 0.96$, $L = 2$ (i.e., window length of $z = 2L + 1 = 5$) provides an optimal watermark estimate in terms of MSE. Although the MSE is an intermediate signal of the steganalysis architecture, it is believed that a more accurate watermark estimate into the pattern classifier will usually result in a more successful overall steganalysis.

The reader should note that in the case of fast-moving video sequences (smaller values of ρ in Fig. 5), the simple linear collusion scheme will not result in a reasonable approximation for U_k; the minimum MSE occurs for $L = 0$. This motivates our work in Section V where we provide a practical alternative to improve simple linear collusion performance for steganalysis that involves using block-based collusion.

B. Justification of Feature Selection

Estimation of the watermark is one phase of our steganalysis framework. The nonlinear pattern classification stage is dis-

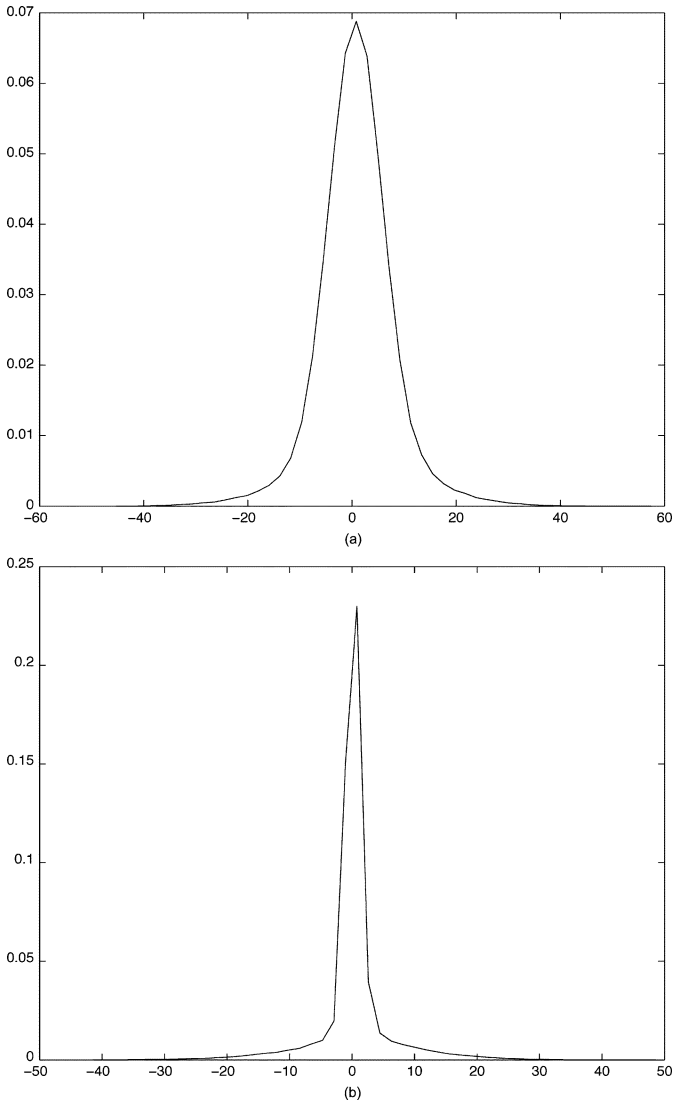


Fig. 6. Distribution of the watermark estimates for a video sequence (a) with and (b) without steganographic data embedded. (a) Watermarked sequence. (b) Nonwatermarked sequence.

cussed in this section. We justify the choice of kurtosis, entropy, and 25th percentile of the watermark estimate as our feature vectors to aid the pattern classifier stage. Fig. 6 gives a representative example of the histogram distribution of the estimated watermark $\hat{W}_k$ for a frame from a watermarked and a nonwatermarked video sequence. It is clear that there exists a difference between the two cases that can be quantified through statistical features; the case in which no watermark is present results in a distribution that is not Gaussian.

1) Kurtosis: Kurtosis [31] is a measure of the degree of flatness or peakness of a distribution compared to the Gaussian case. A higher value implies a distribution with a higher peak than the Gaussian distribution. We expect $\hat{W}_k$ from a watermarked sequence to have a kurtosis close to 3. The estimate from a nonwatermarked sequence should yield a higher kurtosis value owing to its peakiness as shown in Fig. 6.

2) Entropy: Entropy measures the degree of randomness in a data set. We argue that the estimate of the watermark from a watermarked video sequence will have more entropy compared to the estimate from a nonwatermarked sequence. This supports [33] where Anderson and Petitcolas define a good steganographic algorithm as one that can minimize the increase in entropy due to embedding. Let us represent the entropy of $\hat{W}_k$ obtained for nonwatermarked sequences as E_0 and the entropy of the estimate of the watermark from a watermarked sequence as E_1.

Proposition 4: In the case of simple linear collusion, the entropy of $\hat{W}_k$ obtained from a watermarked sequence (E_1) is greater to the entropy of $\hat{W}_k$ obtained from a nonwatermarked sequence (E_0) (i.e., $E_1 > E_0$).

Proof: See Appendix, D.

3) Distribution Percentile: The last feature that we consider is the 25th percentile of a given distribution defined as the value above which 25% of the points in the histogram reside. From Fig. 6, it is clear that the distribution when a watermark is present is more "spread" than when no watermark is present resulting in a difference in this percentile value.

Fig. 7 represents a scatter plot of specific statistical features of $\hat{W}_k$ for different video sequences that do and do not contain steganographic information. The features are estimates of the kurtosis, entropy, and 25th percentile of the distribution of $\hat{W}_k$ to form a 3-D feature vector that is plotted for different video frames in two different test video sequences (shown as parts (a) and (b) in the figure). The colored vector points represent the results of different video containing hidden information and the clear points are the results for no hidden information. The separate clustering for the two cases is clear which makes classification possible.

V. Enhancement: Block-Based Collusion

In case of fast-moving sequences or sequences having nontranslational motion, simple collusion may be suboptimal in producing a watermark-free frame. However, if we consider collusion at the block-unit (e.g., 8×8 pixels) instead of at the frame level, it may be possible to compensate for motion by effectively matching blocks of distinct frames via techniques similar to those found in MPEG/H.263x coding schemes.

Block-based collusion for five frames is demonstrated in Fig. 8. The frame corresponding to the center of the window Y_k is called the reference frame. For each block in this reference, the best block match is found in all of the other frames $(Y_{k-2}, Y_{k-1}, Y_{k+1}, Y_{k+2})$ via an MSE measure. A new set of "reconstructed" frames $(Y'_{k-2}, Y'_{k-1}, Y'_{k+1}, Y'_{k+2})$ is then formed by repositioning the matched blocks, so that they are at the position corresponding to the associated block in the reference frame. Once the reconstructed frames are formed, steganalysis is applied via collusion and pattern classification.

An insight that is drawn is that the effective embedding data rate that can be achieved in a video sequence can be significantly reduced if block-based collusion is employed instead of frame-based collusion attack. This is because the effective correlation between the blocks will be higher for nonmoving parts and will help in detecting messages embedded in those areas. Thus, from an embedder's point of view, he or she can judiciously hide the messages only in frame areas for which a good match cannot be found in the other frames of the collusion window.

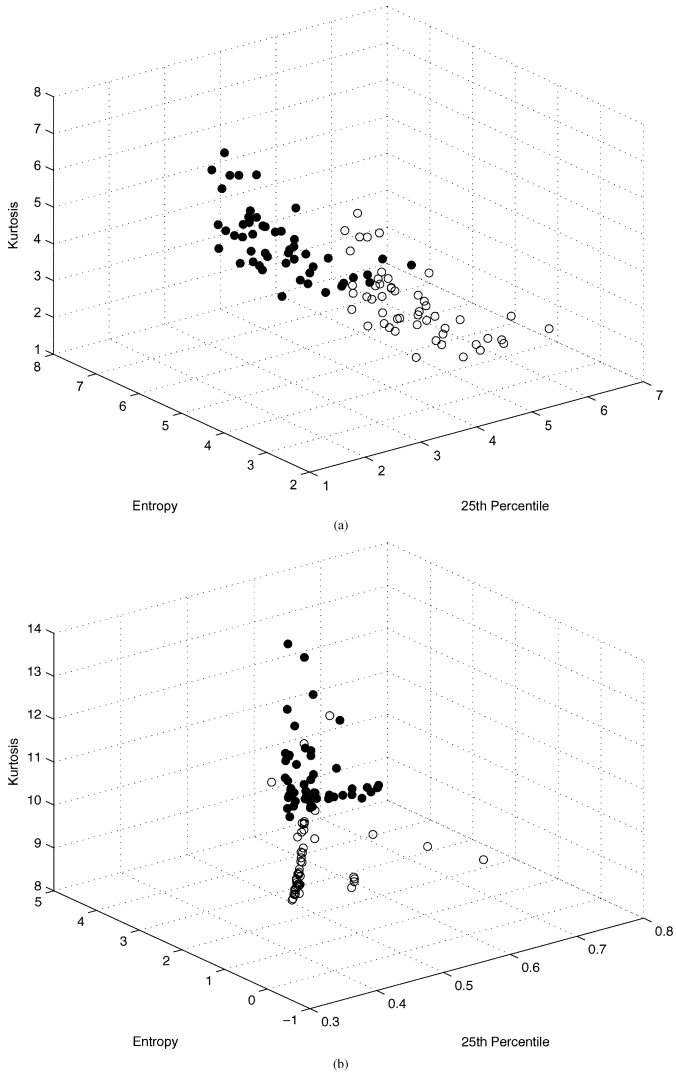


Fig. 7. Scatter plots of kurtosis, entropy, and 25th percentile feature vectors extracted from each frame. Colored and clear points represent the cases with and without a watermark present in the video, respectively. (a) Scatter plot for "Backyard" video sequence. (b) Scatter plot for "Hotel" video sequence.

VI. Results

The sequences[2] that were chosen for simulation and testing consist of grayscale video sequences of different resolutions in raw format. As discussed in Sections II-A and II-B, the messages are embedded in the spatial domain of each video frame to test the performance of our technique. However, the reader should note that our approach to steganalysis will still work if the embedding is done in another linear transform domain such as the DCT; tests for the embedded DCT domain were conducted, resulting in similar conclusions and are thus not included for reasons of space. The embedding was done by adding watermarks W_k from a zero-mean unit variance (i.e., $\sigma_w^2 = 1$) Gaussian distribution as presented in (2) into every pixel of each frame. The watermark strength parameter α is varied to test the effects on secrecy. The values used in our simulations are $\alpha = 1, 3, 5$ resulting in an embedded watermark variances of values $\alpha^2\sigma_w^2 = 1, 9, 25$, respectively. The smaller the value of α, the less perceptible the mark is both visually and through steganalysis, but the lower the capacity or robustness of the covert data embedding is, making it more vulnerable to active wardens.

Because of the variation of σ_u^2 for each test video sequence, the inverse WSNR values have been found to be in the range 38–3500, 4–390, and 1.5–140 for $\alpha = 1, 3$, and 5, respectively, providing a diverse basis for testing.

[2]The sequences were downloaded from http://ise.stanford.edu/video.html and http://www.articom.info/1489.html.

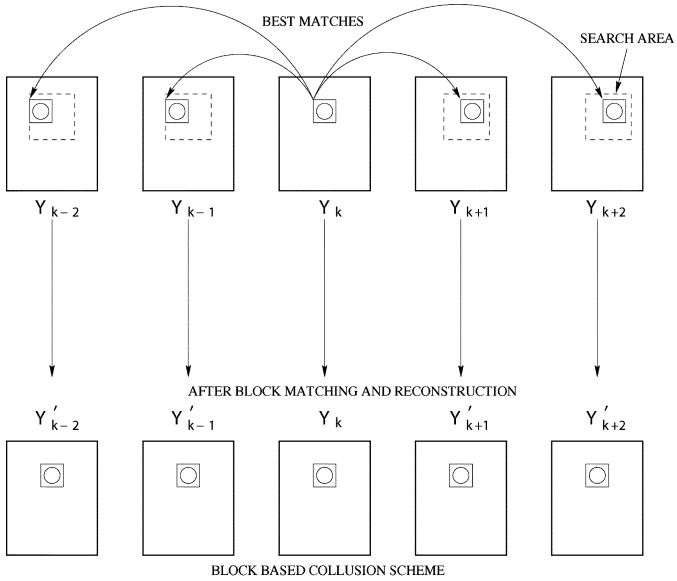


Fig. 8. Block-based collusion attack.

A. Window Length for Collusion

As mentioned in Section III-B we use a sliding window to perform the collusion attack. Different window lengths were employed for simple linear collusion on test video sequences containing watermarks X_k to produce $\hat{U}_k$. The difference $Y_k - \hat{U}_k$ was then obtained to provide an estimate of αW_k. To determine the success of the window length for steganalysis, the pairwise correlation coefficient $\rho(\alpha W_k, \hat{W}_k)$ was computed, where

$$\rho(A, B) = \frac{\mathrm{cov}(A, B)}{\sqrt{\mathrm{var}(A) \cdot \mathrm{var}(B)}} \tag{15}$$

where $\mathrm{cov}(\cdot, \cdot)$ denotes the covariance and $\mathrm{var}(\cdot)$ denotes the variance of the argument random variable(s).

Fig. 9(a)–(c) shows the average correlation between the embedded watermarks and the estimated watermarks over 40 frames using simple linear collusion for different values of embedding strength and window lengths for various sequences. We see that in Fig. 9(a), the correlation is highest for a majority of the sequences for a window length of 3. The optimum collusion window length increases for higher embedding strengths as can be seen from Fig. 9(b) and (c). This is in accordance with our earlier assertion that for a fixed value of ρ and σ_u^2, an increase in WSNR will lead to a increase in the value of the optimum L as demonstrated in Fig. 4. From Fig. 9(a)–(c), we see that the optimum collusion length of sequence "alex" is 3, 9, and 13 for embedding strengths of 1, 3, and 5, respectively. However, we see that for the sequences "carphone," "mobile,"

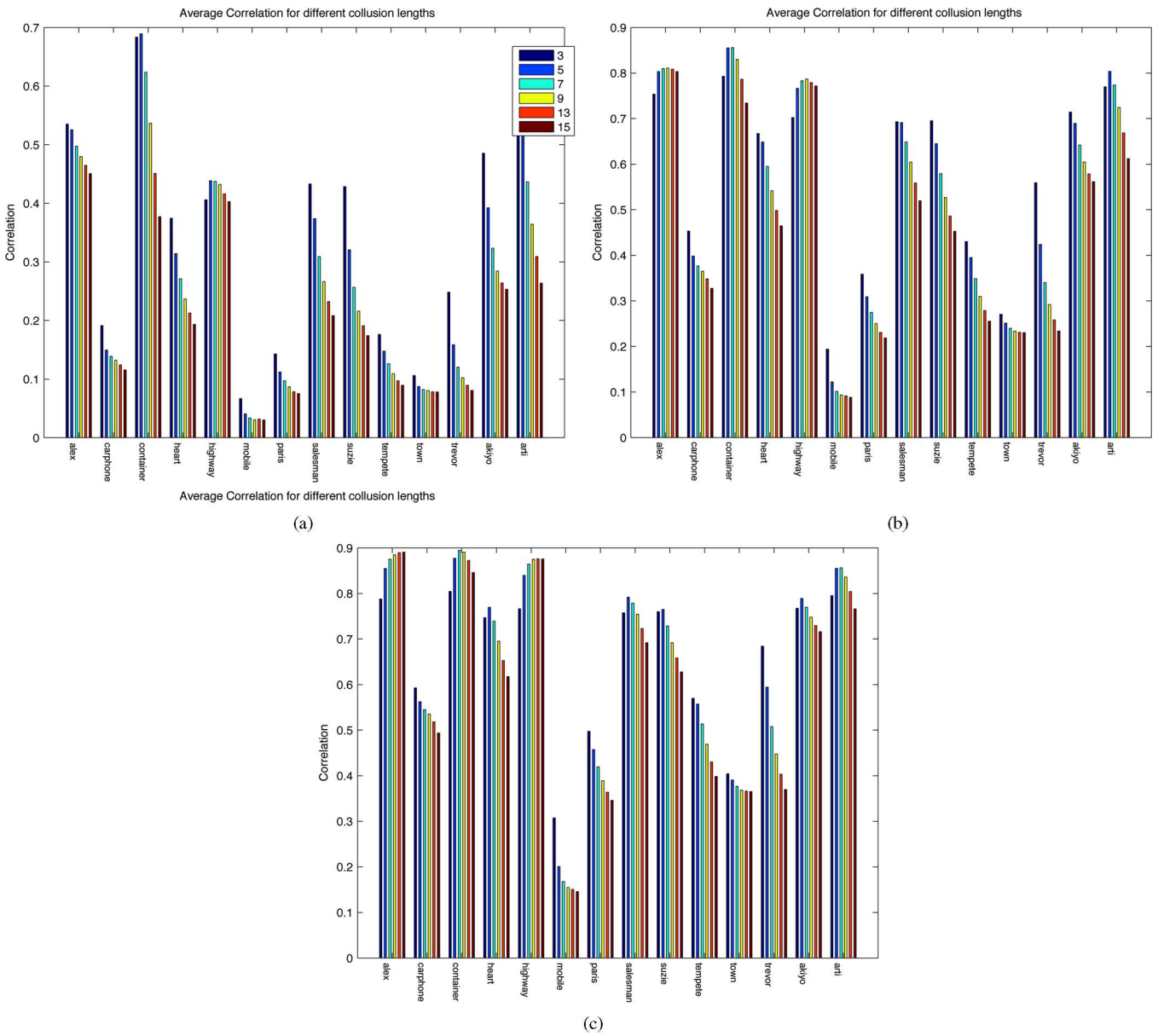


Fig. 9. Average correlation between W_k and $\hat{W}_k$ for different sequences: (a) $\alpha = 1$, (b) $\alpha = 3$, and (c) $\alpha = 5$. (Color version available online at: http://ieeexplore.ieee.org.)

"paris," "tempete," and "trevor," the optimum collusion length does not change with an increase in the embedding strength. This is because the increase in the embedding strength does not decrease the inverse WSNR (which was found to be very high for most of these sequences) sufficiently to cause an increase in the collusion length as predicted in Fig. 4.

We assume that the embedder uses a low embedding strength for watermark insertion since a higher embedding strength would leave significant statistical imprints. With this assumption, we chose the optimum collusion length to be $2L + 1 = 5$ for future tests of the proposed steganalysis method. However, the reader should note that the optimum technique would choose the collusion window length based on characteristics of the video sequences, such as correlation coefficient, degree of global motion, and variance of frames, and is a potential area of future work.

B. Training and Parameter Selection for the kNN Classifier

Other issues that require optimization are the training and parameter selection of the kNN classifier. The number of video sequences required for training for effective classification is application dependent. In our training, we employed 14 video sequences consisting of 40 frames each. Each video sequence was watermarked using the particular spread-spectrum embedding approach to represent the class of watermarked videos. The other class comprised of the same sequences without any embedded data. Features representing the watermarked and unwatermarked classes were extracted from $\hat{W}_k$ in each frame for every video sequence using collusion. We minimized the probability of false negative detection given a maximum false positive rate of 25% in simulations. The parameter k in the kNN classifier [30], [32] that determines the number of "nearest neighbors"

TABLE I
FALSE ALARM RATES (%) FOR SPATIAL-DOMAIN STEGANOGRAPHY USING $\alpha = 1$

$\alpha = 1$						
Method	Wiener		Averaging		Block-based	
Sequence	P_{FN}	P_{FP}	P_{FN}	P_{FP}	P_{FN}	P_{FP}
Seq No:15	0	100	40	60	0	97.5
Seq No:16	2.5	0	0	0	40	17.5
Seq No:17	30	97.5	0	10	5	100
Seq No:18	37.5	100	20	75	10	52.5
Seq No:19	100	0	40	32.5	12.5	0
Seq No:20	25	47.5	35	62.5	20	67.5
Seq No:21	27.5	40	27.5	72.5	12.5	52.5
Seq No:22	0	92.5	10	0	0	2.5
Seq No:23	0	100	5	95	12.5	82.5
Seq No:24	87.5	0	5	0	97.5	0
Seq No:25	82.5	0	5	15	2.5	5
Seq No:26	65	27.5	35	52.5	5	2.5
Seq No:27	37.5	100	40	2.5	12.5	0

searched to reach a classification decision was set to $k = 1$ to give a low probability of false negative and positive with low complexity; higher values of k did not improve performance.

C. Performance Results

The probabilities of false negative P_{FN} and false positive P_{FP} for frame-by-frame detection were computed for a given test video sequence by counting the number of misdetections over each frame in the sequence (40 frames per sequence were employed); thus, if one video frame out of the 40 results in a false detection, the error probability is 2.5%. We estimated P_{FN} by embedding a Gaussian watermark into a given cover-video sequence and then applying collusion to estimate the watermark present. The fraction of failed detections was counted to estimate P_{FN}. Similarly, the same approach was applied to unmarked video sequences to estimate P_{FP}.

Our aim is to detect the presence of covert data in a video sequence on the whole rather than estimating the presence of watermarks in individual frames. So if P_{FN} and P_{FP} are less than 0.5, we still have a successful steganalysis attack. Let us assume that the P_{FN} for a watermarked sequence is 0.3. This means that 30% of the total number of frames from a watermarked sequence are classified as nonwatermarked and the remainder as watermarked. By adopting a majority-takes-all strategy, this flags the overall video sequence as containing hidden data. The steganalysis method was tested on three different variations of the spread-spectrum steganography as discussed in [24] providing similar results. We provide the results for the embedding approach described in Section II-A.

Tables I–III show the probability of false negative P_{FN} and the probability of false positive P_{FP} for the proposed steganalysis method for embedding in spatial domain via (2). The tables show the error encountered in detection of watermarked and nonwatermarked sequences for different values of α using different steganalysis methods. The comparisons between the spatial-based steganalysis method based on Wiener filtering [34] to estimate the hidden watermark and the temporal methods using simple linear collusion and the block-based collusion have been provided. As we can see from Table I, for an embedding strength of $\alpha = 1$, the P_{FN} is reasonably low for most test video sequences.

TABLE II
FALSE ALARM RATES (%) FOR SPATIAL-DOMAIN STEGANOGRAPHY USING $\alpha = 3$

$\alpha = 3$						
Method	Wiener		Averaging		Block-based	
Sequence	P_{FN}	P_{FP}	P_{FN}	P_{FP}	P_{FN}	P_{FP}
Seq No:15	0	55	35	65	0	100
Seq No:16	10	0	0	0	0	0
Seq No:17	0	0	0	0	0	97.5
Seq No:18	0	0	2.5	27.5	2.5	0
Seq No:19	0	0	0	0	15	0
Seq No:20	0	0	0	0	2.5	0
Seq No:21	0	0	0	2.5	0	12.5
Seq No:22	0	0	0	0	0	0
Seq No:23	0	100	0	100	0	100
Seq No:24	0	0	0	0	72.5	0
Seq No:25	0	0	0	0	0	0
Seq No:26	20	0	0	0	0	0
Seq No:27	10	100	0	0	0	0

TABLE III
FALSE ALARM RATES (%) FOR SPATIAL-DOMAIN STEGANOGRAPHY USING $\alpha = 5$

$\alpha = 5$						
Method	Wiener		Averaging		Block-based	
Sequence	P_{FN}	P_{FP}	P_{FN}	P_{FP}	P_{FN}	P_{FP}
Seq No:15	0	100	20	72.5	0	100
Seq No:16	0	0	0	0	0	0
Seq No:17	0	0	0	0	0	0
Seq No:18	0	0	0	25	0	0
Seq No:19	0	0	0	0	0	0
Seq No:20	0	0	0	0	0	0
Seq No:21	0	0	7.5	5	0	0
Seq No:22	0	0	0	0	0	0
Seq No:23	0	100	0	100	0	50
Seq No:24	0	0	0	0	0	0
Seq No:25	0	0	0	0	0	0
Seq No:26	0	0	0	0	0	0
Seq No:27	0	0	0	0	0	0

Tables II and III also show how the performance of the steganalysis technique improves as the magnitude of the embedding strength α increases. It follows that a steganalysis technique that works well for a lower value of α will work at least as well for higher values. Thus, our analysis of small values of α provides a minimum performance limit on the algorithm.

The proposed steganalysis techniques produced high error rates for Sequence 15 and Sequence 23. This is due to the high degree of camera movement or global change in the scenes. The nontranslational nature of the motion in these sequences causes the block-based approach to fail as well. Due to the lack of sufficient temporal correlation, the energy of the residual noise due to collusion n_k is high, thus confusing the pattern classifier into believing that an additional component from watermarking exists in $\hat{W}_k$. Again, this observation verifies the trend of Fig. 4 that collusion is less effective for steganalysis when the correlation between frames is lower.

Overall, we observe comparable or improved performance of the purely temporal-based techniques over the purely spatial approach based on Wiener filtering demonstrating the usefulness

of temporal processing for video steganalysis. The performance of the block-based scheme is naturally better than the simple linear collusion scheme.

We would like to point out that the proposed steganalysis method has to undergo no change apart from the generation of the training set in order to successfully detect DCT-based spread-spectrum steganography [24]. The DCT transform is linear and, hence, any Gaussian watermark added in the DCT domain remains Gaussian in the spatial domain. The feature vectors are thus robust enough to address watermarks using varied embedding schemes. The results, found in [34], are comparable to the spatial domain and are not included due to space reasons.

VII. Conclusion

The work presented in this paper demonstrates the potential of our framework and the use of temporal processing for effective steganalysis. To the best of our knowledge, we developed the first video steganalysis algorithm that takes advantage of the temporal redundancy present in the video. We see improved performance in our method over the spatial methods that work on a frame-by-frame basis.

An advantage of simple linear collusion is that it has low complexity and is suitable for real-time applications. For every frame that is under a steganalysis test, there is a latency of 2–4 future frames before one can perform windowed collusion. At a display rate of 30 frames per second, this corresponds to a time lag of $1/10$ of a second. The processing time is low and does not add any non-negligible processing delay.

We demonstrate how statistical redundancy in the cover video can aid a steganalyst in detecting hidden watermarks. Increased interframe correlation improves performance of collusion. Furthermore, the block-based scheme demonstrates how slow-moving video sequences are not an ideal choice for steganography as supported by [1].

Through this paper, we studied the tradeoff between robust embedding of messages and detection capability of our steganalysis method. We see that the steganalytic detection rate increases with an increase in the watermark embedding strength suggesting that robustness increases the chances of detection. The analyses suggest using a range of 1 to 3 for α to foil collusion-based steganalysis. We note from simulations that for an embedding strength of 1, P_{FN} and P_{FP} are relatively high compared to higher embedding strengths of $\alpha \geq 3$. However, employing such low embedding strengths makes steganography susceptible to active wardens that can easily remove the watermark and, thus, prevent covert communications without employing steganalysis.

A. Limitations and Future Directions

Our steganalysis scheme presumes that the sender alters every pixel of each frame to embed a watermark. In order to maximize the capacity of the hidden data, this assumption is reasonable. However, future investigations must consider how the effects of interleaving the watermark in selected pixels, frames, or video features affect the steganalysis detection accuracy. Such interleaving will provide the sender with greater secrecy at the expense of capacity or robustness. We expect that there is a threshold for interleaving below in which steganalysis detection will become inaccurate. Thus, this value determines the effective covert communication capacity that cannot be detected.

To develop a strategy that works for a broader class of embedding schemes, one must robustly incorporate information about the statistics of the video [11], [13], [15], [16] rather than solely consider the statistics of a possibly hidden message. One area of possible future research involves incorporating cover-medium characteristics into the proposed framework.

Another modification that shows promise involves detecting the presence of a watermark at the block level rather than at a frame level. A collective decision, such as majority wins, can be made on each frame using the individual detection results on the blocks. The detection results for each frame can then be used to detect the presence and absence of a message in the entire video sequence. It is shown in [35] that such a distributed framework can help lower the probability of false negative and false positive suggesting the promise of this approach.

For situations in which the watermark possesses interframe correlation, one may consider applying the collusion attack strategies of [22] for more reliable steganalysis. Adaptation of the analysis by the second author [22] to the present framework provides fertile ground for theoretical research that can apply to algorithmic development.

We are currently exploring the advantages of employing weighted linear collusion instead of simple collusion and have derived values for optimal weights. However, the optimal weights require knowledge of elements of the cover video and the spread-spectrum embedding parameters, which may be difficult to estimate from Y_k, subsequently resulting in weak performance. We are as-of-yet uncertain if the additional complexity is worth the possible performance increase, which at this stage seems slight, at best.

Finally, we intend to consider employing an adaptive value of L for a given video sequence for more effective collusion in addition to more complex nonlinear collusion models.

This first study on temporal domain steganalysis demonstrates that using the time domain provides comparable or better performance than exclusively spatial-domain approaches. Since both domains are orthogonal and provide valuable information for the overall objectives of steganalysis, our long-term goal is to develop a framework for video steganalysis that makes efficient use of both domains jointly.

Appendix
Proofs

A. Proof of Proposition 1

The expected MSE between the estimated and the orginal watermark as defined in (11) is given by

$$\begin{aligned}&\mathsf{E}[(\hat{W}_k - \alpha W_k)^2] \\ &\quad = \mathsf{E}[(U_k - \hat{U}_k)^2] \\ &\quad = \mathsf{E}\left[(U_k - \frac{1}{2L+1}\sum_{i=k-L}^{k+L} Y_i)^2\right]\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}\left[U_k^2 + \frac{1}{(2L+1)^2}\left(\sum_{i=k-L}^{k+L}(U_i+\alpha\cdot W_i)\right)^2\right.\\
&\quad\left. - \frac{2}{2L+1}U_k\left(\sum_{i=k-L}^{k+L}(U_i+\alpha\cdot W_i)\right)\right]\\
&= \mathsf{E}U_k^2 + \frac{1}{(2L+1)^2}\left(\mathsf{E}\left[\left(\sum_{i=k-L}^{k+L}U_i\right)^2\right]\right.\\
&\quad + \mathsf{E}\left[\alpha^2\left(\sum_{i=k-L}^{k+L}W_i\right)^2\right]\\
&\quad + 2\alpha\mathsf{E}\left[\sum_{i=k-L}^{k+L}U_i\cdot\sum_{i=k-L}^{k+L}W_i\right]\\
&\quad\left. - \frac{2}{2L+1}\left(\sum_{i=k-L}^{k+L}\mathsf{E}[U_k\cdot U_i+\alpha\cdot U_k\cdot W_i]\right)\right]\\
&= \sigma_u^2 + \frac{\alpha^2\sigma_w^2}{2L+1} + \frac{1}{(2L+1)^2}\mathsf{E}\left[\left(\sum_{i=k-L}^{k+L}U_i\right)^2\right]\\
&\quad - \frac{2}{2L+1}\sum_{i=k-L}^{k+L}\mathsf{E}[U_i\cdot U_k]\\
&\quad (\text{Using Assumption 2}, \mathsf{E}W_k = 0\\
&\quad \mathsf{E}W_i\cdot W_j = 0 \text{ for } i\neq j \text{ and}\\
&= \sigma_u^2 + \frac{\alpha^2\sigma_w^2}{2L+1} + A - B \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
A &= \frac{1}{(2L+1)^2}\mathsf{E}\left[\left(\sum_{i=k-L}^{k+L}U_i\right)^2\right]\\
B &= \frac{2}{2L+1}\sum_{i=k-L}^{k+L}\mathsf{E}[U_i\cdot U_k]
\end{aligned}
$$

Now

$$
\begin{aligned}
B &= \frac{2}{2L+1}\sum_{i=k-L}^{k+L}\mathsf{E}[U_i\cdot U_k]\\
&= \frac{2}{2L+1}\mathsf{E}[U_k\cdot U_{k-L} + U_k\cdot U_{k-L+1} + \cdots\\
&\quad + U_k\cdot U_{k-1} + U_k\cdot U_k + U_k\cdot U_{k+1} + \cdots\\
&\quad + U_k\cdot U_{k+L-1} + U_k\cdot U_{k+L}]\\
&= \frac{2\sigma_u^2}{2L+1}(\rho^L + \rho^{L-1} + \cdots\\
&\quad + \rho + 1 + \rho + \cdots + \rho^{L-1} + \rho^L)\\
&\quad (\text{Using Markov Model defined in Assumption 2})\\
&= \frac{2\sigma_u^2}{2L+1}\left(1 + \frac{2\rho(1-\rho^L)}{1-\rho}\right)\\
&\quad (\text{Assuming } |\rho|<1, \text{fails for } \rho = 1)
\end{aligned}
$$

Now

$$
\begin{aligned}
A &= \frac{1}{(2L+1)^2}\mathsf{E}\left[\left(\sum_{i=k-L}^{k+L}U_i\right)^2\right]\\
&= \frac{1}{(2L+1)^2}\mathsf{E}\left[\sum_{i=k-L}^{k+L}U_i\sum_{j=k-L}^{k+L}U_j\right)\Bigg]\\
&= \frac{1}{(2L+1)^2}\mathsf{E}\left[\sum_{j=k-L}^{k+L}U_{k-L}\cdot U_j\right.\\
&\quad + \sum_{j=k-L}^{k+L}U_{k-L+1}\cdot U_j + \cdots\\
&\quad\left. + \sum_{j=k-L}^{k+L}U_{k+L-1}\cdot U_j + \sum_{j=k-L}^{k+L}U_{k+L}\cdot U_j\right]
\end{aligned}
$$

The 1^{st} term is

$$= \frac{\sigma_u^2}{(2L+1)^2}(1+\rho+\rho^2+\cdots+\rho^{2L-1}+\rho^{2L})$$

2^{nd} term is

$$= \frac{\sigma_u^2}{(2L+1)^2}(\rho+1+\rho+\cdots+\rho^{2L-2}+\rho^{2L-1})$$

$$\vdots$$

$2L+1^{th}$ term is

$$= \frac{\sigma_u^2}{(2L+1)^2}(\rho^{2L}+\rho^{2L-1}+\rho^{2L-2}+\cdots+\rho+1).$$

The terms can be put together as rows of a Toeplitz matrix and the sum of all terms is given by the sum of all elements in the matrix. Adding the terms and assuming $z = 2L+1$, we have

$$
\begin{aligned}
A &= \frac{\sigma_u^2}{z^2}[z + 2(z-1)\rho + 2(z-2)\rho^2 + \cdots\\
&\quad + 2(z-(z-1))\rho^{z-1}]\\
&= \frac{\sigma_u^2}{z^2}\left[z + 2z\rho\frac{(1-\rho^{z-1})}{1-\rho} - 2\rho\sum_{j=1}^{z-1}j\rho^{j-1}\right]\\
&= \frac{\sigma_u^2}{z^2}\left[z + 2z\rho\frac{(1-\rho^{z-1})}{1-\rho} - 2\rho\frac{d}{d\rho}\sum_{j=1}^{z-1}\rho^j\right]\\
&= \frac{\sigma_u^2}{z^2}\left[z + 2z\rho\frac{(1-\rho^{z-1})}{1-\rho}\right.\\
&\quad\left. - 2\rho\left(\frac{(1-\rho^z - z\rho^{z-1}(1-\rho))}{(1-\rho)^2}\right)\right].
\end{aligned}
$$

Substituting the values of A, B and $z = 2L+1$ in (16) and simplifying, we have

$$
\begin{aligned}
&\mathsf{E}[(\hat{W}_k - W_k)^2]\\
&\quad = \sigma_u^2\left[\frac{z-1}{z} - \frac{2\rho}{z(1-\rho)} + \frac{4\rho^{\frac{z+1}{2}}}{z(1-\rho)}\right.\\
&\quad\quad\left. - \frac{2\rho(1-\rho^z)}{z^2(1-\rho)^2}\right] + \frac{\alpha^2\sigma_w^2}{z}.
\end{aligned}
$$

The mean μ of the host frames in our proof has been ignored and is assumed to be zero. However, the final term will be independent of the mean even if we take it into account.

B. Proof of Proposition 2

The expected MSE between the watermarks in the case there is no collusion (i.e., $L = 0$ or $z = 1$ in (12)) reduces to

$$\begin{aligned}&\mathsf{E}[(\hat{W}_k - \alpha W_k)^2] \\ &\quad = \sigma_u^2 \left[-\frac{2\rho}{(1-\rho)} + \frac{4\rho}{(1-\rho)} - \frac{2\rho}{(1-\rho)} \right] + \alpha^2 \sigma_w^2 \\ &\quad = \alpha^2 \sigma_w^2.\end{aligned}$$

C. Proof of Proposition 3

To determine the conditions for which simple collusion reduces the MSE of the watermark estimate, we consider the case in which the estimated MSE obtained from use of collusion (i.e., the right-hand side (RHS) of (12) is smaller than the estimated MSE obtained for no collusion (i.e., the RHS of (13). By multiplying the RHS of the first inequality by $0 \leq (z-1)/z \leq 1$, it is straightforward to determine that

$$\begin{aligned}&\sigma_u^2 \left[\frac{z-1}{z} - \frac{2\rho}{z(1-\rho)} + \frac{4\rho^{\frac{z+1}{2}}}{z(1-\rho)} - \frac{2\rho(1-\rho^z)}{z^2(1-\rho)^2} \right] \\ &\qquad + \frac{\alpha^2 \sigma_w^2}{z} < \alpha^2 \sigma_w^2 \\ &\Rightarrow \frac{\sigma_u^2}{\alpha^2 \sigma_w^2} < \frac{1}{1 - \frac{2\rho}{(z-1)(1-\rho)} + \frac{4\rho^{\frac{z+1}{2}}}{(z-1)(1-\rho)} - \frac{2\rho(1-\rho^z)}{z(z-1)(1-\rho)^2}}.\end{aligned}$$

D. Proof of Proposition 4

In the case of nonwatermarked sequences, $\hat{W}_k = n_k$ and in the case of watermarked sequences, $\hat{W}_k = n_k + W'_k$, where $n_k = U_k - \mathfrak{C}_L(U_k)$ and $W'_k = \alpha(W_k - \mathfrak{C}_L(W_k)$. Due to assumption 3), n_k and W'_k are independent. Therefore, the entropy of the watermark estimate from watermarked frames (denoted as E_1) in relation to the entropy of the watermark estimate from nonwateramrked frames (denoted as E_0) is given by

$$\begin{aligned}E_1 &= H(n_k + W'_k) > H(n_k + W'_k \,|\, W'_k) \\ &= H(n_k \,|\, W'_k) = H(n_k) = E_0\end{aligned}$$

where we assume that $\sigma_w^2 > 0$ and $\alpha \neq 0$ for the case of watermarked sequences to ensure strict inequality.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor, Prof. H. Farid, for handling the review of this paper as well as the three anonymous reviewers for their useful comments.

REFERENCES

[1] R. Chandramouli, "A mathematical framework for active steganalysis," *ACM Multimedia Syst. J., Special Issue Multimedia Watermarking*, vol. 9, no. 3, pp. 303–311, 2003.

[2] G. Kessler, Steganography: Implications for the prosecutor and computer forensics examiner Apr. 2004 [Online]. Available: http://www.garykessler.net/library/ndaa.stego.html.

[3] J. Kelly, Terror Groups Hide Behind Web Encryption May 2001 [Online]. Available: http://www.usatoday.com/tech/news/2001-02-05-binladen.html.

[4] D. Lewis, Terrorists and Steganography Sep. 2001 [Online]. Available: http://www.linuxsecurity.com/content/view/110558/151/.

[5] B. Krebs, Danger of Image-Borne Viruses Looms Sep. 2004 [Online]. Available: http://www.washingtonpost.com/wp-dyn/articles/A45126-2004Sep23.html.

[6] J. Fridrich, R. Du, and L. Meng, "Steganalysis of LSB encoding in color images," presented at the IEEE Conf. Multimedia Expo, New York, Jul./Aug. 2000.

[7] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatibility," presented at the SPIE Multimedia Syst. Appl. IV, Denver, CO, Aug. 2001.

[8] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," presented at the 3rd Information Hiding Workshop, Dresden, Germany, Sep./Oct. 1999.

[9] J. Harmsen and W. Pearlman, "Steganalysis of additive noise modelable information hiding," in *Proc. SPIE Security and Watermarking of Multimedia Contents V*, Santa Clara, CA, Jan. 2003, vol. 5022.

[10] N. Provos, "Defending against statistical steganalysis," presented at the 10th USENIX Security Symp., Dresden, Germany, Aug. 2001.

[11] H. Farid, "Detecting hidden messages using higher-order statistical models," presented at the IEEE Int. Conf. Image Processing, Rochester, NY, Sep. 2002.

[12] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," presented at the 5th Int. Workshop Information Hiding, Noordwijkerhout, The Netherlands, 2002.

[13] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," presented at the IEEE Workshop on Statistical Analysis in Computer Vision, Madison, WI, 2003.

[14] S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines," presented at the SPIE Symp. Electronic Imaging, San Jose, CA, 2004.

[15] I. Avcibas, B. Sankur, and N. Memon, "Steganalysis based on image quality metrics—Differentiating between techniques," presented at the IEEE Workshop on Multimedia, Cannes, France, Oct. 2001.

[16] I. Avcibas, B. Sankur, and K. Sayood, "Image steganalysis with binary similarity measures," in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, Jun. 2002, vol. 3, pp. 645–648.

[17] ——, "Statistical evaluation of image quality measures," *J. Electron. Imaging*, vol. 11, no. 2, pp. 206–223, Apr. 2002.

[18] S. Liu, H. Yao, and W. Gao, "Steganalysis of data hiding techniques in wavelet domain," in *Proc. Int. Conf. Information Technology: Coding Computing*, Las Vegas, NV, Apr. 2004, vol. 1, pp. 751–754.

[19] S. Liu, H. Yao, and W. Gao, "Steganalysis based on wavelet texture analysis and neural network," presented at the 9th China Conf. Machine Learning, Shanghai, China, Oct. 2004.

[20] U. Budhia and D. Kundur, "Video steganalysis using collusion sensitivity," in *Proc. SPIE: Sensors, Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense*, E. M. Carapezza, Ed., Orlando, FL, Apr. 2004, vol. 5403, pp. 210–221.

[21] G. Mohay, A. Anderson, B. Collie, O. de Vel, and R. McKemmish, *Computer and Intrusion Forensics*. Norwood, MA: Artech House, 2003.

[22] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 43–51, Feb. 2005.

[23] ——, "Spatially localized image-dependent watermarking for statistical invisibility and collusion resistance," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 52–66, Feb. 2005.

[24] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[25] L. Marvel, C. B. Jr., and C. Retter, "Spread spectrum image steganography," *IEEE Trans. Image Process.*, vol. 8, no. 8, pp. 1075–1083, Aug. 1999.

[26] A. Bovik, *Handbook of Image and Video Processing*. New York: Academic, 2000.

[27] J. Kilian, F. Leighton, L. Matheson, T. Shamoon, R. Tarjan, and F. Zane, Resistance of digital watermarks to collusive attacks Comput. Sci. Dept., Princeton Univ., Princeton, NJ, Tech. Rep. TR-585-98, Jul. 1998.

[28] W. Trappe, M. Wu, Z. Wang, and K. J. R. Liu, "Anti-collusion fingerprinting for multimedia," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1069–1087, Apr. 2003.

[29] Z. Wang, M. Wu, H. Zhao, W. Trappe, and K. J. R. Liu, "Collusion resistance of multimedia fingerprinting using orthogonal modulation," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 804–821, Jun. 2005.

[30] R. Duda, P. Hart, and D. Stork, *Pattern Classification and Scene Analysis*, 2nd ed. New York: Wiley, 2001.
[31] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
[32] R. Gutierrez-Osuna, Cpsc 689-604: Special Topics in Pattern Analysis Transl.:Lect. Notes Sep. 2003 [Online]. Available: http://research.cs.tamu.edu/prism/lectures.html.
[33] R. Anderson and F. Petitcolas, "On the limits of steganography," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 474–481, May 1998, (Special Issue on Copyright and Privacy Protection).
[34] U. Budhia, "Steganalysis of video sequences using collusion sensitivity," Ph.D. dissertation, Texas A&M Univ., College Station, 2005.
[35] R. Chandramouli and N. Memon, "A distributed detection framework for steganalysis," in *Proc. ACM Workshop on Multimedia Security*, Los Angeles, CA, Nov. 2000, pp. 123–126.

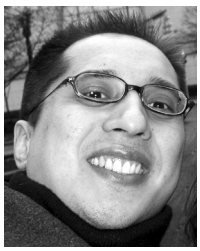
Udit Budhia was born in Ranchi, India. He received the B.E. degree in electrical engineering from Birla Institute of Technology, Ranchi, in 1998 and the M.S. degree in electrical engineering from Texas A&M University, College Station, in 2005.

Currently, he is a DSP Engineer with Mobilygen Corporation, Santa Clara, CA. His research interests include image processing, multimedia signal processing, and data hiding.

Deepa Kundur (S'93–M'99–SM'03) was born in Toronto, ON, Canada. She received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical and computer engineering in 1993, 1995, and 1999, respectively, from the University of Toronto, Toronto, ON, Canada.

In 2003, she joined the Department of Electrical and Computer Engineering at Texas A&M University, College Station, where she is an Assistant Professor and leads the Sensor Media Algorithms and Networking for Trusted Intelligent Computing (SMANTIC) research group of the Wireless Communications Laboratory. Her research interests include security and privacy for scalar and broadband sensor networks, multimedia security, digital rights management, steganalysis for computer forensics, and dynamical systems theory.

Dr. Kundur has given tutorials in the area of information security at ICME-2003 and Globecom-2003, and was Guest Editor of the June 2004 Proceedings of the IEEE Special Issue on Enabling Security Technologies for Digital Rights Management. She is currently Associate Editor of the IEEE COMMUNICATIONS LETTERS, on the editorial board of the *EURASIP Journal on Information Security,* and Vice-Chair of the Security Interest Group of the IEEE Communications Technical Committee.



Takis Zourntos (M'00) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 1993, 1995, and 2003, respectively,

Currently, he is an Assistant Professor with the Department of Electrical and Computer Engineering at Texas A&M University, College Station. His research interests are in lightweight algorithm design, nonlinear control and system theory, analog computation for robotics, and optimization and integrated-circuit implementation.