Towards Digital Video Steganalysis using Asymptotic Memoryless Detection

Julien S. Jainsky
Texas A&M University
julienj@tamu.edu

Deepa Kundur
Texas A&M University
deepa@ece.tamu.edu

Don R. Halverson
Texas A&M University
halverson@ece.tamu.edu

ABSTRACT

This paper studies the potential for passive steganalysis in correlated image frames using non-classical detection theory. In particular, an algorithm for digital video steganalysis, named *MoViSteg* for Motion-based Video Steganalysis, is developed that exploits the temporal correlation among individual image frames in video signals to enhance steganalysis performance. The method differs from prior art in the use of motion interpolation and non-classical asymptotic memoryless detection that we believe is well-suited for video steganalysis. Results and discussion are provide in order to demonstrate the potential of our ideas for intrusion detection in a broad class of emerging multimedia applications.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: General—*Security and protection (e.g., firewalls)*

General Terms

Security

Keywords

Steganalysis, Detection theory, Motion estimation

1. INTRODUCTION AND MOTIVATION

Given the high degree of collaboration and cooperation in modern information systems such as emerging multimedia sensor networks, covert communications is a greater threat than ever. Network-level approaches to covert transmission have classically involved passing information innocuously via shared resources by having one communicating entity modulate network characteristics (such as transmission times or storage elements) such that a second party (who can monitor the resources) deduces the secret message. More modern approaches to covert transmission in networks have included the use of steganographic mechanisms in network protocol packets [8]; however, recent research [9] has suggested that communicating covertly at such a structured level is easily detectible making it more attractive for attackers to employ subversive communications at the multimedia content level.

Studies in enabling and preventing covert transmissions have repeatedly demonstrated a fundamental trade-off between the reduction of covert communications capacity and the performance of overt communications. As communications, computation and sensing converge to create advanced multimedia sensor networking, we assert that it will become practically impossible to design high performance networks that prevent covert communications. The high levels of redundancy of such networks provide a rich environment for data hiding without significantly affecting network performance. In addition, the scalable network design often requires collaboration on the part of network entities enabling subversive communications to a much greater extent. Furthermore, the acquisition of highly correlated multimedia provides fertile ground for advanced steganographic approaches.

Covert communications in multimedia networks is of special concern for several reasons. Given recent interest in employing multimedia sensor systems for tactical military and healthcare applications, such networks are natural targets for attack. A distinguishing assumption in threat models of these systems is the high likelihood of insider attack via corrupt network entities that facilitate subversive behavior. In addition, sensor network security strategies often entail intrusion detection mechanisms that only exploit deviations in overt communication statistics to assign a trust-level to each network entity encouraging stealthy behavior [10]. Moreover, covert communications among select network participants allows for strategic cooperation amongst corrupt nodes resulting in highly effective denial-of-service attacks [11].

Thus, we believe it is imperative to investigate methods to detect and discourage covert communications in multimedia networks that acquire highly correlated data. In this paper, we focus on the particular problem of passive digital video steganalysis; modern detection theory strategies suitable for video steganalysis and multimedia networking applications are proposed and evaluated as to their potential. Video information can be interpreted as a sequence of correlated image readings. Therefore, the straightforward problem of video steganalysis allows us to also determine more effective principles that we can apply to more general content networks where entities acquire correlated sensor readings. In

MM&Sec'07, September 20–21, 2007, Dallas, Texas, USA.

addition, our proposed strategies can be placed in context within the large body of steganalysis literature providing existing tools as well as an effective means of evaluating performance. Moreover, the formalism presented through the use of detection theory provides an effective methodology to develop steganalysis techniques suitable for a broad family of emerging multimedia networking applications.

The paper is organized as follows. In Section 2, we briefly overview related prior work in the field of steganalysis and detail the assumptions of our problem. Section 3 discusses and motivates our proposed method entitled *MoViSteg* (Motion -based Video Steganalysis) and Section 4 presents performance results. Final comments and future work conclude the paper.

2. PRIOR ART AND PRELIMINARIES

It is well-known that a steganalyst (classically referred to as the "warden") can be either *passive* or *active*. A passive steganalyst only detects the presence or absence of steganographic content and may infer other characteristics of the secret message such as its length. An active steganalyst often non-discriminately alters multimedia signals while in transit from the *stego-sender* to the *stego-receiver* in order to destroy or distort the presence of any secret message before it reaches its destination. In this paper, we focus on the problem of passive steganalysis because it has the potential to provide discriminating intrusion detection information that is vital to detect corrupt network entities; in this way, they can be subsequently prevented from participating in the network.

Most research on passive steganalysis has focused on detecting steganography in still images [1, 2, 3, 4, 5, 16, 17]. However, given the high levels of redundancy present in video, we believe that there is tremendous potential for creative data hiding and inspired steganalysis that extend to multimedia sensor networks. We therefore, take a structured and formal approach to this problem providing a solution amenable to adaptation to a family of related applications.

Early steganalysis methods were *reactive*, meaning that they targeted only a specific embedding algorithm. For instance, Westfeld and Pfitzmann [1] developed both visual and statistical attacks to counteract the EzStego method employing least significant bit (LSB) embedding. Later methods were applicable to a broader range of embedding methods. For example, Fridrich *et al.* developed RS-steganalysis [2], which can identify the application of a class of LSB embedding methods. Subsequently, in [3], Fridrich *et al.* used the characteristics of JPEG compressed images as "signatures" in order to detect data hiding.

More *proactive* steganalysis approaches that apply to a broad class of embedding methods have also been proposed. These methods are often called *blind* or *universal* steganalysis techniques since no (or very few) assumptions on the embedding process are made. Within this class, Farid *et al.* employed the higher order statistics of image features [4]. Here, the image is decomposed via a wavelet transform and the mean, variance, skewness and kurtosis of associated coefficients and their features are used to differentiate cover-images from stego-images. Since few assumptions on embedding are made, training to estimate thresholds and soft computing are employed. In [5], Avcibas *et al.* introduced the use of image quality metrics (IQMs) in order to discern between cover and stego-images by taking into account the more global characteristics of natural images. Their approach uses discriminative image statistics such as the Minkowsky metric, the spectral magnitude and the normalized mean square error as well as regression analysis to build a composite measure to identify the presence of hidden data.

More recently methods founded on detection theory have been presented with the hope of leading to a more universal and high-performance steganalysis solution. In [16, 17], Sullivan *et al.* developed a detection theoretic approach that employs a Markov chain (MC) model for spatial correlation in the cover-image order to identify the presence of hidden data. The authors argue that their approach provides a fundamental benchmark for evaluating the security of data hiding algorithms. Their detection algorithm is based on the observation that the divergence of the *transition matrix*, comprised of conditional probability values governing the MC source model, behaves somewhat predictably when data is embedded in an image; in particular, significant matrix values "spread out" from the main diagonal. To quantify this characteristic, features are extracted from an empirically generated transition matrix of the suspect-image and classification based on supervised learning strategies is employed to deduce whether the suspect is a cover-image or stego-image. Their approach illustrates the necessary interaction between detection theory and signal processing to develop a practical method governed in well-developed theory.

In this paper, we argue that this interaction is especially important for the problem of video steganalysis in which the temporal complexity should be taken into account and the subsequent detection complexity may be of great concern for emerging applications. To the best of the authors' knowledge, only one other video steganalysis method has been proposed to date by Budhia *et al.* [6, 7]. The proposed technique leverages the differences in correlation from frame-to-frame between the watermark and the cover-video in order to statistically separate the components. In particular, the method assumes that each frame of a stego-video contains a hidden spread spectrum watermark and employs the collusion attack on adjacent video frames in order to estimate the cover-video. Features of the difference between the suspect-video and cover-video are then classified using a kNN classifier that is trained a priori.

Given the success of video steganalysis in incorporating temporal correlations of the cover-video as well as the fundamental nature of detection theory in helping develop benchmarks and methods suitable for a variety of situations, we consider the development of a video steganalysis algorithm incorporating asymptotic relative efficiency (ARE)-based detection. Our contribution can be considered a performance-enhancing extension of [6, 7] and the extension of [16, 17] employing a non-classical detection-theoretic framework, new to the field of steganalysis. In the next section, we summarize our assumptions.

2.1 Assumptions

We make the following assumptions:

- A1 The suspect-media consists of a sequence of correlated image frames; in this work we consider them to comprise a video signal.
- A2 The watermark is independent (or has low dependence) from frame-to-frame; this is a reasonable assumption assuming high-capacity embedding.

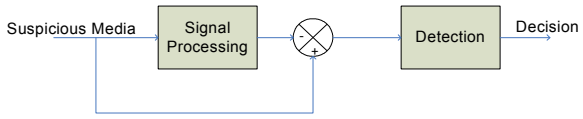

Figure 1: Steganalysis algorithm.

A3 The video signal obeys a Gauss-Markov correlation model temporally (i.e., from frame-to-frame).

A4 The watermark is zero-mean.

In this paper, no explicit assumptions on the probability density function (pdf) of the watermark are made other than the requirement for zero-mean. In addition, we do not explicitly consider spatial correlations as we focus on the steganalytic capabilities of the temporal domain of video. It is to be noted that although there are methods that do not obey A2, video steganalysis is a new field and we believe current video data hiding methods do not make use of informed embedding. We strive to design a technique that is low-cost for video data and leverages well-developed research fields of signal processing and detection theory that are suited for the problem at hand. Many steganalysis techniques inherently estimate the cover features of a media signal from the suspect and then evaluate features of the difference between the two in order to detect hidden information. In this work, we take a slightly different approach. We look at signal processing to enhance the hidden data temporally.

3. MOVISTEG DEVELOPMENT

A steganalytic algorithm is usually comprised of two parts as shown in Figure 1: a signal processing phase and a detection stage. The former step is often used to extract features of the multimedia signal that are affected by steganography; whereas the latter part provides the necessary decision-making often using user-determined parameters. Various signal processing techniques have been developed through the years, however, except for a handful of methods [16, 17], the detection approach has generally consisted of a form of comparison of the output of the signal processing phase to a simple threshold. We assert based on recent research showing the potential of classical detection theory for image steganalysis [16, 17] that the investigation of formal detection theory is a fruitful field with the potential to benefit steganalysis in a diverse set of applications.

The signal processing phase is often first executed by attempting to find an accurate estimate of the cover-media, or associated statistical metrics, from the suspect-media. Once this estimate is obtained, it is effectively subtracted from the suspect and the difference is input to the detection stage, which categorizes the suspect-media as being a cover-media or stego-media. In many techniques, a simple threshold is used for the detection step.

We use the same general process for *MoViSteg*. However, instead of estimating the cover-media, we attempt to create a new video signal, which when subtracted to the suspicious one will enhance the presence of steganography, if it exists. To do so, we use a motion estimation algorithm. The reason that we taken this alternative approach from existing methods is to make our approach more suited to applications in which only a subset of video frames are watermarked instead of all of them as in [6, 7].

The second stage of the steganalysis is an adapted detector. The use of a detector is crucial here as steganalysis basically boils down to a detection problem. Hence there is a need for a well-designed detector to take the best decision possible on the potential steganographic content of the suspicious sequence. We assert that the added complexity compared to simple detection schemes will prove to be worth it.

3.1 Motion Interpolation

As mentioned previously, the purpose of this stage is to maximize the difference between the suspicious and estimated sequences in order to make the detection of steganography easier.

3.1.1 Motivation

Motion estimation and interpolation techniques have often been used for the purpose of video restoration [12]. When a frame is poorly transmitted and hence is missing data, it is possible to retrieve most of the data by using the surrounding frames because of the high correlation existing between each frame.

The same idea is used here for the purpose of steganalysis. One frame, F_n, at a time is assumed to be "missing" from the video and needing interpolation. By using the immediately adjacent frames, denoted F_{n-1} and F_{n+1}, the estimated frame $F_n^\sharp$ is reconstructed. It is to be noticed that frame F_n does not contribute to the reconstruction of frame $F_n^\sharp$. This ensures that if F_n is corrupted by steganography it will not contaminate $F_n^\sharp$.

In presence of hidden data the associated disruption added to the frames will make the motion estimation more difficult and hence will introduce greater discrepancies between both the suspicious and estimated sequences. Ideally, the larger these deviations, the easier the steganalysis goal.

3.1.2 Algorithm

The first step in finding $F_n^\sharp$ is to compute the motion vectors between F_{n-1} and F_{n+1}. It should be noted that some video codecs provide these motion vectors directly making this stage unnecessary. However this is not true in general, which is why we propose a lightweight estimation step.

The motion estimation method we adopt is the sum of absolute differences (SAD) technique. It is attractive for our purposes because it is commonly employed, simple conceptually (which is convenient for analysis), yet still gives satisfying results for our purposes. It provides an effective and computationally simple estimation technique. More advanced motion estimation algorithms are available in the literature; however, in this case, the use of the SAD technique is not just for performance, but for lower complexity too, and straightforward assessment of this new approach. The algorithm needs to be lightweight in order not to waste considerable amounts of time and resources, especially if one must use video steganalysis continuously or in emerging resource constrained multimedia content networks.

To compute the motion vectors from F_{n-1} to F_{n+1}, the SAD technique takes each $N \times N$ block of pixels, $B_{n-1,k}$ (where k represents the block index) in F_{n-1} and determines the best matching block $B_{n+1,k}$ in F_{n+1} that minimizes the SAD:

$$B_{n+1,k} = min_i |B_{n-1,k} - B_{n+1,i}| \quad (1)$$

Because there are approximately 25 and 30 frames per sec-

ond in a sequence in PAL and NTSC standards respectively, we assume that F_{n-1} and F_{n+1} are close and therefore that the displacement between both frames is small. As a result, the search area to find $B_{n+1,k}$ is then confined to the coordinates adjacent to $B_{n-1,k}$.

For model simplicity the displacement from one frame to another is also considered linear. Therefore the interpolation step to find $F_n^\sharp$ reduces to taking half the length of the motion vectors previously obtained. The same algorithm is repeated for each frame of the suspect-video sequence to generate the newly estimated sequence, which is subsequently subtracted from the suspect-sequence.

We can classify the motion interpolation of a frame as being associated with one of several distinct cases summarized in Table 1. Case 1 considers the situation in which F_{n-1}, F_n and F_{n+1} all do not contain a watermark (i.e., they are "not corrupted" (NC) by the embedded data). When motion interpolation is applied, the resulting estimated frame $F_n^\sharp$ also is NC. Therefore, we expect that the difference between F_n and $F_n^\sharp$ to be minimal if interpolation is successful. In Case 2, only frame F_n is watermarked and thus "corrupted" (C) by the embedded data. Therefore, $F_n^\sharp$ generated from F_{n-1} and F_{n+1} will be NC and there will ideally be a detectable deviation between F_n and $F_n^\sharp$. Similarly in Cases 4 to 8 there is some deviation even when both F_n and $F_n^\sharp$ are corrupted by embedding because we assume that the embedding from frame-to-frame is independent (or nearly so). Thus, other than Case 1, there ideally should be a measurable deviation in the difference: $F_n - F_n^\sharp$. This difference-sequence is directly processed by the detector.

The reader should note that we do not extract features from $F_n - F_n^\sharp$ prior to the detection as in the case of many existing algorithms as we intend for the detector to work on all the information available to it.

3.2 Detector

3.2.1 Prior knowledge

Detection theory is generally used to detect the presence of a signal in some form of additive noise. It is necessary when one needs to make a choice between a finite set of hypotheses such as the following:

- $H_0 : Y_i = N_i$
- $H_1 : Y_i = N_i + \theta \cdot S_i$

where N_i and S_i represent the noise and the signal distributions of the $i-th$ frame, respectively, and θ denotes the strength of the hidden signal.

In *MoViSteg*, the video represents the classical notion of "noise" in detection theory, whereas the embedded message represents the "signal." In much of classical detection theory, the signal is considered to be a constant, which is not the case in our situation because the watermark must be necessarily random throughout the cover-media to have a non-trivial payload capacity. Therefore, we model both the "signal" and "noise" for detection as random variables necessitating and motivating the use of non-classical detection.

The optimal Neyman-Pearson detector is not used here because it requires the knowledge of the signal distribution. As we strive to be more proactive such an assumption would be too strong. Instead, we employ a detector based on the asymptotic relative efficiency (ARE) [15]. This particular

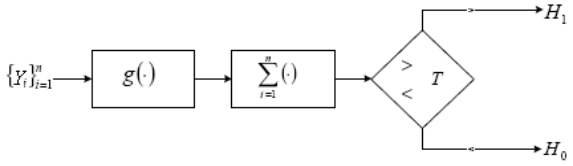

Figure 2: Proposed MoViSteg Detection Stage.

detector is efficient for large samples and weak signals, which fits our application model; to be transparent perceptually, the secret hidden data must be overpowered by the cover-video itself. The detector also needs to be tractable for practical algorithm design; for that reason we choose the ARE detector to be memoryless and dependent to account for the highly correlated suspect data to be analyzed. Figure 2 provides an overview of the detection scheme.

The input of the detector corresponds to the difference between the suspect-video and the estimated video from the motion interpolation phase with $\{Y_i\} = F_i - F_i^\sharp$. Without going into specifics, the use of the ARE detector requires the knowledge of a nonlinearity $g(x) = \sum_{i=0}^{M} a_i \cdot x^i$. This nonlinearity is a polynomial where coefficients a_i can be obtained once the efficacy η is maximized. The efficacy is a performance measure used in the ARE tests in which the larger the value of the efficacy the more efficient the tests. It is defined as follows [13]:

$$\eta(g) = \frac{\left[\frac{\partial^2}{\partial\theta^2} E_\theta\{g(Y_1)\}|_{\theta=0}\right]^2}{\sigma_0^2(g)} \quad (2)$$

with, by definition [13]:

$$\sigma_0^2(g) = E\{g(N_1)^2\} + 2 \cdot \sum_{j=1}^{\infty} [E\{g(N_1) \cdot g(N_{j+1})\}] > 0 \quad (3)$$

Once the efficacy is maximized, the chosen detector becomes optimal over the set of suboptimal dependent and memoryless detectors [14].

3.2.2 Calculations

In order to optimize the efficacy, we use the Lagrange multipliers technique which, while holding the denominator constant ($\sigma_0^2(g) = \alpha^2$), maximizes the numerator by finding the extrema of the function f defined as:

$$f = \left[\frac{\partial^2}{\partial\theta^2} E_\theta\{g(Y_1)\}|_{\theta=0}\right]^2 - \lambda(\sigma_0^2(g) - \alpha^2) \quad (4)$$

This infers that the coefficients a_i of the nonlinearity g can be determined by solving $\frac{\partial f}{\partial a_i} = 0$ for $i = 1...M$ which boils down to the system $W \cdot A = c \cdot Z$ with:

$$\frac{1}{2} \cdot W_{n\times n}(i,j) = E\{N_1^{i+j}\} + \sum_{k=1}^{\infty} \left[E\{N_1^i N_{k+1}^j\} + E\{N_1^j N_{k+1}^i\}\right] \quad (5)$$

$$Z_{n\times 1}(i) = i \cdot (i-1) \cdot E\{N_1^{i-2}\} \quad (6)$$

$$A_{n\times 1}(i) = a_i \text{ and } c \text{ is a constant} \quad (7)$$

Because of the $N_1^i N_{k+1}^j$ moments, further analyzes and computations are only possible if a distribution model is assigned

□	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8
F_{n-1}	NC	NC	C	C	NC	C	NC	C
F_n	NC	C	NC	NC	NC	C	C	C
F_{n+1}	NC	NC	C	NC	C	NC	C	C
$F_n^\sharp$	NC	NC	C	C	C	C	C	C

Table 1: Results of Motion Interpolation for various levels of frame-watermarking. Other than Case 1, there ideally exists a measurable difference between F_n and $F_n^\sharp$.
NC = Not Corrupted, C = Corrupted

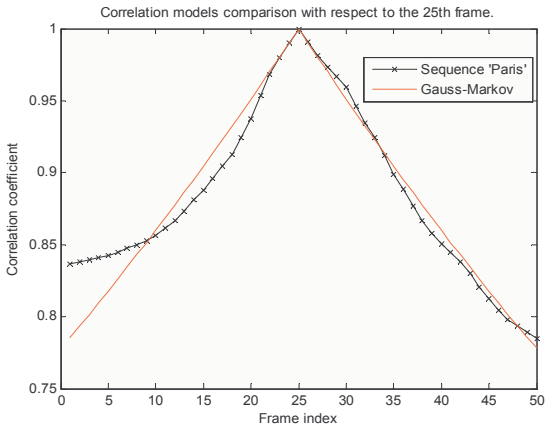

Figure 3: Correlation models for Real Video 'Paris' and Gauss-Markov distribution.

to the noise N (i.e., cover-video). This is not an assumption about the watermark signal statistics and therefore still provides some proactivity to our steganalysis approach. One model that has been used in order to model video-sequences is the Gauss-Markov model representing a noise process solution to the equation:

$$N_n = e^{-a} \cdot N_{n-1} + G_n \tag{8}$$

With $a \in R_+$ and $\{G_n\}_0^\infty \sim \aleph(0, 1 - e^{-2a})$ are *iid*. This model has potential because it provides a practical detector for steganalysis wile modeling reality to some extent. For example, in Figure 3 we show the temporal (frame-by-frame) correlation of the video 'Paris' (found at http: //trace.eas.asu.edu/ MDC/ index.html) and the Gauss - Markov distribution for $a = 0.01$. It is clear that the correlation models are reasonably close. The reader should note that the correlation is somewhat dependent on the video content (frames 1-10 vs. frames 11-50) eventually causing mismatching between the practical and theoretical models. However, the Gauss-Markov model can be triggered by modifying the value of the constant a in the previous equation. Here a is equal to 0.01 which is the most common value observed during the tests for matching real videos.

Employing this assumption leads to the following simplification:

$$N_1^i N_{k+1}^j = G_1^i \cdot \left(\sum_{m=1}^{k+1} e^{-(k+1-m)\cdot a} \cdot G_m \right)^j \tag{9}$$

From this simplification and the properties of the *iid* gaussian $\{G_n\}_0^\infty$, it becomes easier to solve the equation $\frac{\partial f}{\partial a_i} = 0$ for $i = 1...M$, as we can now define numerically the $N_1^i N_{k+1}^j$ moments. Finding the optimal coefficients a_i still involves heavy calculations because the system remains complex, however this is a one-time computation as the nonlinearity g remains the same whatever the suspicious video is. For the case where videos are tested five frames at a time, five coefficients are needed for g; these five coefficients are:

$$\{a_0, a_1, a_2, a_3, a_4\} = \{8.2643, 0, -51.729, 0, 23.27\} \tag{10}$$

3.2.3 *Threshold Selection*

In theory, the detector's threshold T should remain constant. Each suspicious video would be seen as a different realization of the same noise. However, the size of the videos varies as well as its content that is why an adaptative threshold is needed.

Examples of salient video characteristics that vary from one sequence to another include the size of the video as well as its standard deviation and its correlation coefficient; these therefore must be accounted for in our threshold selection. The number of frames to be analyzed at each passing into the detector also influences the decision and should appear into the threshold T assignment. Preliminary tests were conducted in order to find the best threshold as a function of these four parameters. Tests results demonstrate satisfying performance when the threshold is assigned to be:

$$T = \left[C \cdot StD(F_1) \cdot Corr(F_1, F_2) \cdot \frac{x \cdot y}{288 \cdot 352} \right]^2 \cdot \left[\frac{N_f}{5} \right] \tag{11}$$

where C is a constant, $StD(F_1)$ the standard deviation of the first frame of the video, F_1; $Corr(F_1, F_2)$ the correlation between the first and the second frames of the video, $\frac{x \cdot y}{288 \cdot 352}$ the size of each frame relative to the size of the videos used for the preliminary tests, $(x_{test}, y_{test}) = (288, 352)$, and $\frac{N_f}{5}$ is the number of frames analyzed at each passing into the detector relative to the number of frames analyzed during the preliminary tests ($N_{f,test} = 5$).

4. PERFORMANCE ANALYSIS

The detector has been derived to amplify the difference between cover-videos and stego-videos. For example, Figure 4 shows the output of the steganalyzer without and with detector for the sequence 'Paris'. In this case, the average ratio between the output for cover-videos and stego-videos is around 5 and 697 without and with detector respectively. Therefore the ratio becomes larger with the detector, hence making the differentiation between corrupted and non-corrupted sequences easier and proving the added performance of the detector.

Using the previously defined threshold, tests are executed. Several videos, with or without embedding, are imported into the steganalytic algorithm. The motion vectors are computed and the estimated sequence is examined by the

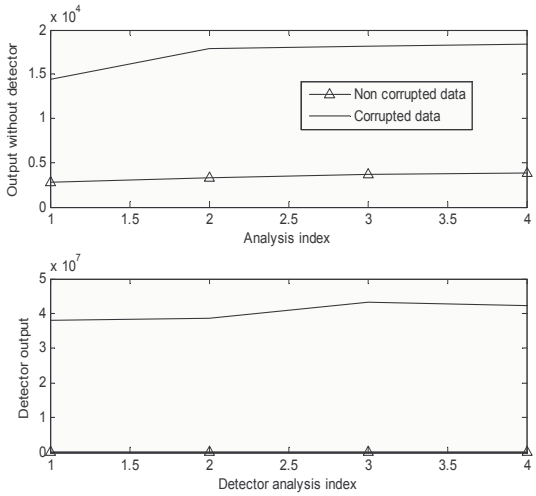

Figure 4: Output of Steganalyzer without and with Detector.

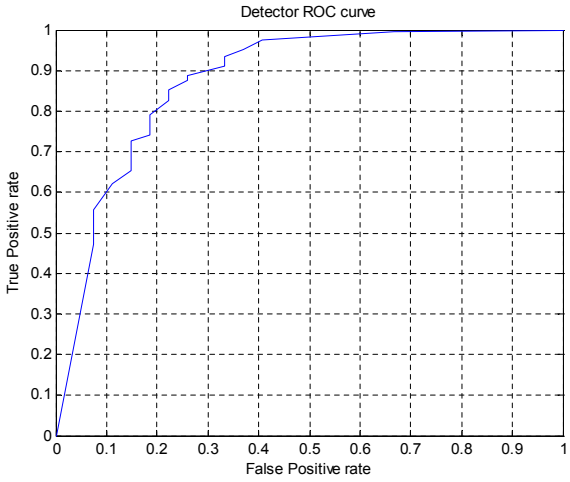

Figure 5: MoViSteg ROC curve.

detector which takes the final decision regarding the potential presence of steganography.

The tests consist in 28 different gray-scale sequences of 25 frames each and are conducted with a 5 frame-analysis during each passing into the detector. From each of these 28 uncorrupted sequences, 26 new sequences are derived, each possessing respectively from 0 to 25 corrupted frames. The corrupted sequences are generated by adding spread spectrum watermarking to the frames. Therefore a total of 728 sequences have been used for these tests. For this first series of tests, the corrupted frames are watermarked using a high peak signal-to-noise ratio (PSNR) of 75dB.

The detector receiver operating characteristic (ROC) curve is plotted in Figure 5. The varying parameter is the constant C from the threshold definition. One important point that differs from previous work on video steganalysis is the assumption about the distribution of hidden data. In [6, 7], the watermark signal is assumed to be zero-mean and present in every frame so that after doing a frame averaging the embedded data diminishes. However *MoViSteg* is developed to identify the presence of steganography even when a small number of frames are contaminated. It is also naturally observed that the more frames that contain the watermark the easier the steganalysis. In Figure 6 another ROC curve is presented. In this case 80% of the frames in the sequence are corrupted.

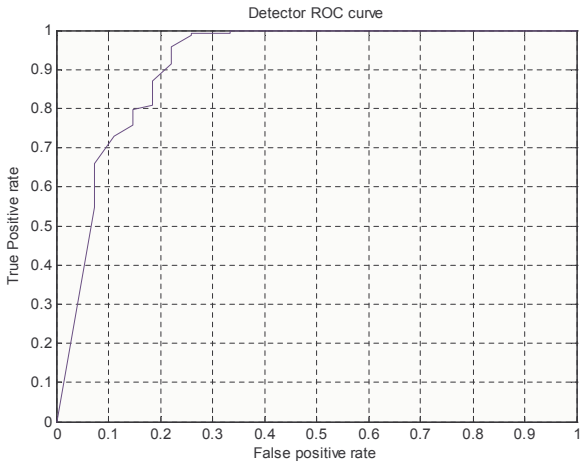

Figure 6: MoViSteg ROC curve for at least 20 frames corrupted out of 25.

In Figure 5, *MoViSteg* achieves a 60% true positive rate for a 10% false positive. Figure 6 shows a better performing steganalysis as a 70% true positive rate is achieved for a 10% false positive, therefore supporting that the steganalysis improves as the number of corrupted frames increases.

In order to test the performance of the detector against the strength of the watermark, we present the case where the embedding distortion, measured by the PSNR, is allowed. The tests are consist in a smaller sample of sequences than the previous ones as we only want to show the general behavior of the steganalytic algorithm for different PSNR. Figure 7 shows the ROC curves of *MoViSteg* for a PSNR ranging from 60dB to 75dB. The results confirm what was expected i.e., the stronger the watermark, the better the steganalysis.

In theory, it is argued that for a steganographic algorithm to be defeated, it is necessary for the steganography to be detected with more success than a random guess i.e., with a true positive rate greater than 50% and a false positive rate less than 50%. However it is commonly agreed that the real performance of an efficient steganalysis should be far better. In Figure 5, *MoViSteg* achieves a true positive rate of 60% for a false positive rate of 10% when the embedding distortion is about 75dB, therefore demonstrating its potential for practical steganalysis.

5. CONCLUSION

In the present paper, we proposed a steganalysis scheme, *MoViSteg*, consisting in two distinct stages. The first stage is the use of a signal processing algorithm that aims to emphasize the presence of hidden information in the sequence using a motion estimation scheme. The second stage is the formulation of a detection theory algorithm based on asymptotic relative efficiency. This algorithm uses a detection approach

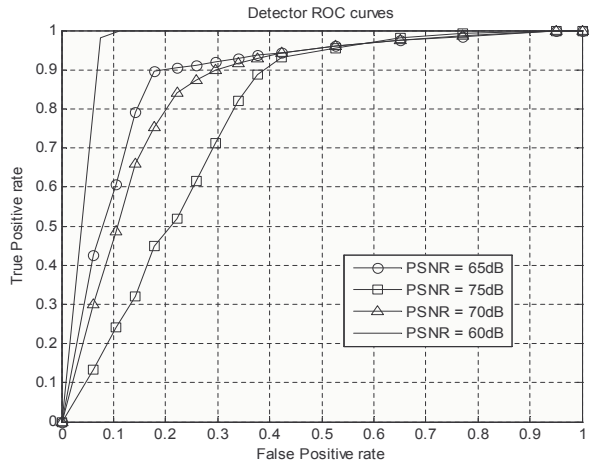

Figure 7: MoViSteg ROC curves for different PSNR.

in which both the cover-video ("noise") and watermark ("signal" to detect) are considered to be random variables. *MoViSteg* employs unrestrictive assumptions about the hidden data except that it is zero-mean making it more proactive in identifying a broader class of additive steganography. Future works include testing *MoViSteg* with other steganographic techniques than spread spectrum embedding and also adapting the correlation steganalysis scheme to visual sensor networks in order to detect and track node corruption.

6. ACKNOWLEDGMENTS

This research was supported by Sagem Morpho, Inc.

7. REFERENCES

[1] Westfeld and Pfitzmann,"Attacks on steganographic systems", IEEE transactions on communications, vol. 50, no. 11, November 1999.

[2] J. Fridrich and M. Goljan and R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images", Proc. of the ACM Workshop on Multimedia and Security., pp. 27-30, 2001, citeseer.ist.psu.edu/article/fridrich01reliable.html.

[3] J. Fridrich and M. Goljan and R. Du, "Steganalysis based on JPEG compatibility", Proc. SPIE Multimedia Systems and Applications IV, pp. 275-280, August 2001.

[4] S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines", Proc. 5th Int'l Workshop on Information Hiding, SpringerVerlag., 2002, citeseer.ist.psu.edu/article/lyu02detecting.html.

[5] N. Memon and I. Avcibas and B. Sankur, "Steganalysis Based on Image Quality Metrics", SPIE, San Jose, California, USA., vol. 4314, 2001, citeseer.ist.psu.edu/523259.html.

[6] U. Budhia and D. Kundur, "Video steganalysis using collusion sensitivity", Proc. SPIE, Sensors, Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense, Orlando, Florida, USA., vol. 5403, April 2004.

[7] U. Budhia and D. Kundur and T.Zourntos, "Digital video steganalysis exploiting statistical visibility in the temporal domain", IEEE Transactions on Information Forensics and Security., 2006.

[8] K. Ahsan and D. Kundur, "Practical Data Hiding in TCP/IP", Proc. Workshop on Multimedia Security at ACM Multimedia '02, French Riviera, December 2002.

[9] S. J. Murdoch and S. Lewis, "Embedding Covert Channels in TCP/IP", Proc. 7th International Workshop on Information Hiding, Barcelona, Spain, June 2005.

[10] S. Buchegger and J.-Y. L. Boudec, "Nodes Bearing Grudges: Towards Routing Security, Fairness, and Robustness in Mobile Ad Hod Networks", Proc. Euromicro Workshop on Parallel, Distributed and Network-based Processing, pp. 403–410, Canary Islands, Spain, 2002.

[11] C. Karlof and D. Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures", Proc. IEEE International Workshop on Sensor Network Protocols and Applications, pp. 113–127, May 2003.

[12] Y. Zheng and X. Li and C. Dai, "Video error concealment based on implicit motion models", Multimedia Systems and Applications VIII., Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, vol. 6015, pp. 140-149, October 2005.

[13] D.R. Halverson and G.L. Wise, "A detection scheme for dependent noise processes", J. Franklin Institute., vol. 309, pp. 287-300, May 1980.

[14] D.R. Halverson and G.L. Wise, "Asymptotic memoryless detection of random signals in dependent noise", J. Franklin Institute., vol. 312, pp. 13-29, July 1981.

[15] E.L. Lehmann and J.P. Romano, "Testing Statistical Hypotheses, 3rd. Edition.", Springer Texts in Statistics, 2005.

[16] , K. Sullivan and U. Madhow and S. Chandrasekaran and B.S. Manjunath, "Steganalysis of spread spectrum data hiding exploiting cover memory", Proc. IST/SPIE 17th Annu. Symp. Electronic Imaging Science Technology, San Jose, CA, USA., pp. 38-46, January 2005.

[17] K. Sullivan and U. Madhow and S. Chandrasekaran and B.S. Manjunath, "Steganalysis for Markov Cover Data With Apllications to Images", IEEE Transactions on Information Forensics and Security., vol. 1, pp. 275-287, June 2006.