Video Object Trajectory Perturbation Based Data Hiding Satisfying Statistical and Perceptual Invisibility

Abdullah Cay
Electrical and Computer Engineering
Old Dominion University
Norfolk, VA USA 23529

Zia-ur Rahman
Electrical and Computer Engineering
Old Dominion University
Norfolk, VA USA 23529

Dimitrie C Popescu
Electrical and Computer Engineering
Old Dominion University
Norfolk, VA USA 23529

Abstract—A new semantic steganographic algorithm for hiding data in the motion trajectories of Video Objects (VOs) is presented. First, a set of primitives (objects bounding-box coordinates and its centroid) are used to define a VO. The centroid of VO is tracked in each frame and its coordinate is stored in a trajectory vector. The proposed algorithm embeds data in the VO object trajectory by presenting a relationship between the hidden data and the trajectory data in the form of a motion drift or perturbation. The perturbation introduced by the embedding algorithm is implemented by motion compensating the pixels in the VO bounding box. Assessment of the steganographic algorithm is done by both subjective judgments based on the perceived quality of the perturbed video sequence and by some commonly used statistical steganalysis metrics. Both subjective assessments and statistical measures indicate visual and statistical invisibility.

I. INTRODUCTION

Data hiding is the art of hiding a message signal in a host signal, without any serious perceptual distortion in the host signal. In [1], [2], [3], [4] data hiding is formulated as communications with side information and relevant information theoretic aspects are studied for different applications. The data hiding process includes an embedder, decoder and an attack channel that are depicted in the model in Fig.1. The embedding function $E$ can be defined as

$$s = E(c, m, K)$$

where $c$ is the cover signal (audio, video, text etc., or any 1-D, 2-D feature vector obtained from them), $m$ is a sequence of bits to be embedded and $K$ is the key conveying information about $c$ to the decoder and available at both sides. The embedding mechanism introduces perturbations to cover data features through the embedding function. These perturbations degrade the perceptual quality of the cover data and hence slight perturbations are allowed to have stego data which is perceptually similar to the cover object. In steganographic applications, the very initial goal is to hide the existence of the embedded information against passive or active attackers who have a battery of both statistical and perceptual quality metric tools to analyze the stego data. Based on this constraint the focus of the designer is to keep both perceptual and statistical visibility of the embedded information low to prevent the detection of hidden information within a cover data. The paper is organized as follows: We discuss previous work in Section II and present proposed method in Section III. Experimental results and discussion is given in Section IV. And finally the achievement and summary is concluded in Section V.

II. PREVIOUS WORK

State of the art video data hiding methods [5], [6], [7], [8], [9] focus on perturbing either the transform domain features (DCT, WT, DFT etc.) in a linear and/or non-linear fashion both in raw and compressed video domain or pixel domain features (LSB, Pixel locations etc.).

Although transform and spatial domains have been utilized as the primary operating domains for information embedding there are some methods devised for information embedding in temporal domain of the video [10], [11], [12], [13], [14], [15]. Those methods are basically aimed at modifying selected subset of motion vectors to convey the data to the receiver side. The selection of motion vectors is based on the magnitude and direction angle of the motion vectors and in some cases the position of the motion vector with respect to a reference grid [12]. For instance in [11] they proposed a video watermarking method that embedded information into
selected motion vectors based on magnitude of motion vectors. Large vector magnitude indicates fast moving objects in which case, human eyes cannot perceive motion vector perturbations compared to the smaller magnitude motion vectors. In [13] the motion vectors are altered according to the texture. The motion vectors in high-texture area are altered slightly, while in low-texture area can be altered more.

Those methods reported high PSNR values indicating perceptual invisibility, robustness against noise, compression, blurring and rotation. On the other hand they exhibit lower data rates compared to the classical transform or pixel domain embedding methods and hence can be considered as best for low capacity data hiding applications. Although they have limited amount of features (motion magnitude, x-y components, and motion direction etc.) to be perturbed, those methods could well be employed to provide extra capacity with the existing methods.

III. PROPOSED METHOD

We propose an oblivious quantization based semantic data hiding method providing both statistical and perceptual invisibility. Specifically we utilize Video Object (VO) trajectory to convey the secret message to the decoder side via perturbations of the trajectory centroid coordinates. When modifying the trajectory the semantic meaning of the VOs motion trajectory is preserved as human vision is less sensitive to the form and appearance of objects compared to their velocity and position.

We introduce a data hiding method in spatio-temporal domain. The block diagram of the proposed method is illustrated in Fig.2. The idea behind the method is to perturb the motion trajectory of a video object (VO) in a fashion which preserves the smoothness of the trajectory while modifying the direction and the magnitude at each feature point.

At data embedding stage the user selects the VO bounding box coordinates in the first frame of the sequence. Tracking algorithm tracks the centroid coordinate of the bounding box as feature point in the following frames. Binary message is conveyed to the receiver through perturbations introduced into to the trajectory centroid coordinates. During the perturbation, the direction and the magnitude of the motion is quantized using two different set of quantizers. Finally the new centroid coordinate is used to motion compensate the VO. At the decoder side the decoder takes the incoming video sequence and generates the trajectory of the VO. Tracking module returns the centroid coordinates of the bounding box of the VO. The message is decoded by using the same set of quantizers for which the parameters (range and step sizes) are assumed to be known by both the embedder and the decoder.

The proposed data hiding block diagram modules are discussed in detail next.

A. VO Tracking

In the framework of MPEG-4 a VO basically represents a semantic entity which has one or more regions which are homogeneous with respect to predefined criteria such as color, texture, motion etc. and do not have any semantic meaning at all.

The first step in tracking is to extract or segment VO. In our scheme a supervised segmentation approach is taken where the user defines a bounding box around the VO of interest in the first frame. Fig.3 illustrates the user defined bounding box for a VO. This particular case Ball is the VO. This approach of employing bounding box to define VO overcomes the problems inherent in tracking non-rigid moving objects when precise semantic VO boundary segmentation is employed. In our tracking scheme the centroid coordinates of the bounding box are tracked.

The proposed method is designed for data hiding in sports videos in which the ball undergoes different motion. So for instance the Ball in Table tennis sequence is identified as a VO whose trajectory is utilized data hiding. To track the ball VO we use salient features of the ball i.e. Color, Size and Shape and employ the video sequences in which the ball has smooth motion between frames. In the tracking phase we utilize a
search region in the next frame, based on the previous frame motion magnitude, and search for the VO based on the salient features. The outcome of the search is the VO bounding box centroid coordinates \((c_x, c_y)\). Fig. 4 shows sample trajectory obtained with this model. Once the centroid coordinates at each frame are obtained the trajectory of the VO can be represented as

\[
T\{(c_x, c_y)\} = \{(c_{x1}, c_{y1}), \ldots, (c_{xn}, c_{yn})\}
\]  

(2)

B. Data Embedding Through Trajectory Perturbation

Proposed data hiding method is quantization based with oblivious decoding at the receiver side. The quantization based data embedding and decoding scheme is discussed next. Let the displacement of the VO be defined in terms of centroid coordinates as \(\Delta c_x = (c_{x_i}, c_{x_j})\) and \(\Delta c_y = (c_{y_i}, c_{y_j})\) where \(i, j\) define the frame number and \(i > j\). The motion magnitude and angle can be represented in polar coordinates as

\[
\theta = \arctan \frac{\Delta c_y}{\Delta c_x} \quad r = \sqrt{(\Delta c_x)^2 + (\Delta c_y)^2}
\]  

(3)

At the embedding stage the features used for quantization are the displacement angle and its magnitude. Basically the message is embedded in motion magnitude and its direction via quantization to convey the message to the receiver side. The message \(m[i]\) is assumed to be a binary bit stream with \(i \in \{0, 1\}\). At each frame, L-bit binary code words \(\{b_1, b_2, \ldots, b_L\}\) where \(b_1\) is embedded through angle quantization and the remaining L-1 bits are embedded in magnitude through magnitude quantization as discussed below. First bit of the L-bit binary code word is embedded in motion angle \(\theta\) by using following quantization scheme

\[
\theta_Q = Q(\theta, \Delta_\theta, m[i]) = \Delta_{\text{ground}}(\frac{\theta + m[i] \frac{\Delta_\theta}{2}}{\Delta_\theta}) + m[i] \frac{\Delta_\theta}{2}
\]  

(4)

To embed remaining L-1 bits of the binary code word, motion magnitudes are quantized with a uniform polar quantizer as follows. We first define a range \([r_{\text{min}}, r_{\text{max}}]\) that represents the motion for the VO in the entire video sequence. This value can be made a priori or can be determined by histogram analysis of the VO displacements as defined in for consecutive frames. The dynamic range obtained by this setting is divided into disjoint adjacent rings in the polar quantizer as illustrated in Fig. 5. The magnitude values that fall into corresponding disjoint ring are quantized with uniform polar quantizer \(Q = \{q_1, q_2, \ldots, q_M\}\) with step size \(\Delta\). The rationale behind the partitioning the range into disjoint quantizers is to decrease the quantization introduced distortion. As an example when \(r_1\) first quantizer with range \([0, r_1]\) is used, when \(r_1 < r_2\) the quantizer with the range defined by the ring \((r_1, r_2]\) is used and so on.

This quantization strategy allows us to build up look-up tables at embedder and decoder side. Only side information needs to be communicated to the decoder side are the binary code word length L and disjoint rings. Moreover if we use uniform partitioning of R we only need to send L, R and \(\theta_Q\) back to Cartesian coordinates as

\[
c_{x_{\text{new}}} = r_Q \cos(\theta_Q)
\]

\[
c_{y_{\text{new}}} = r_Q \sin(\theta_Q)
\]  

(5)

C. Statistical Invisibility

In [16] Cachin introduced a requirement for perfect undetectibility (statistical invisibility) based on the Kullback-Leibler divergence

\[
D(P_c||P_s) = \sum_{i=0}^{n} P_c(i) \log \frac{P_c(i)}{P_s(i)}
\]  

(6)
between two pmfs where $P_c$ and $P_s$ represent the stego and cover data pmfs respectively. A steganographic method considered to be $\varepsilon$-secure against passive steganalysis if $D(P_c||P_s) \leq \varepsilon$ and if $\varepsilon = 0$ the steganographic method is called perfectly secure. Therefore basic requirement for statistical invisibility is to have $P_c \approx P_s$ as much as possible.

D. Experimental Results and Discussion

When testing proposed method we consider a VO having moderate to large motion in consecutive frames with smooth trajectories. Some examples of this type of motion pattern can easily be found in sports (soccer, football, basketball etc.) sequences.

The proposed embedding scheme was implemented by using 54 frames of Table Tennis video sequence in which the Ball VO trajectory is used for perturbation. At each frame the ball, which itself is a VO having a semantic meaning and a global motion direction, is tracked and the data is embedded through perturbations introduced in the ball trajectory. Although it is difficult to track, we intentionally choose the ball because the ball has smooth motion and its area is small as compared to those of the players to ensure imperceptibility in perceptual and statistical sense.

We set binary message codeword length $L=3$ bits where 1 is embedded using angle quantization and 2 bits for the magnitude quantization. Based on the histogram analysis of the displacement magnitude, we empirically defined $R=[0,30]$ and the magnitude quantization rings were set to $r_1=[0,8), r_2=[8,16), r_3=[16,24)$ and $r_4=[24,32)$. The embedder selects the quantizer based on the original motion magnitude which falls in the appropriate ring. For angle quantization we used $\Delta\theta$ as defined in (5).

Perceptual invisibility assessment was done by examining perceptual distortions and semantic meaning of VOs trajectory (sudden unacceptable changes in the magnitude direction and magnitude). From the assessments it was observed that the perceptual quality of the data embedded video sequence and semantic meaning were almost identical to the original video sequence without obvious perceptual distortion. Also to assess the quality of perturbed frames PSNR values, Mean Squared Error and Mean Absolute Error values were computed and the results are shown in Fig.6 and Fig.7 respectively. High per frame PSNR values also support the high perceptual quality of the perturbed frames. Note that the gaps in per frame PSNR values indicate infinite values as we have almost zero MSE for those frames.

In addition to the subjective judgments, to verify whether the data embedding algorithm has altered any statistical property of the original video frame we employed commonly used histogram based metrics given in the Appendix. Fig. 8 shows the result of the histogram based measures. The results clearly indicate low statistical evidence of the perturbation to the original frames. Per frame spectral difference between the original frames and the perturbed frames is illustrated in Fig. 9 also prove that the degree of distortion in DFT domain is almost negligible.

IV. CONCLUSION

Preliminary results of a new data hiding method via motion trajectory perturbation are presented. The proposed method aimed at hiding data in sports videos in which the ball itself is a VO. The subjective assessment of data embedded video sequence showed that the perceptual qualities of the frames are identical to the original ones with almost no visual artifacts. We also studied statistical visibility by employing some commonly used statistical measures on the data embedded frames and found that our algorithm did not leave any statistical signature after the embedding process. Hence, our algorithm can achieve the goals of both statistical invisibility and imperceptibility. Finally, to increase the data hiding capacity, proposed algorithm can be applied to hide data into trajectory segments (partial trajectories) of multiple VOs in a sequence simultaneously or in a random fashion. The selection of trajectory segments can be made key dependent providing another level of security as only the embedder and decoder will
know which VOs and their corresponding trajectory segments are used for perturbation based data hiding. Further research will focus on this direction.

APPENDIX

In [17] Avcibas et al. introduce a set of quality measures commonly used by steganalysis detectors. We use following set of statistical, perceptual and spectral measures from [17] and additional histogram measures to test the performance.

A. Video Perceptual Distortion Measures

These measures calculate the total distortion between cover and stego frames based on pixelwise differences between the two or certain moments of the difference frame.

Let \( C_k(i, j) \) represents the pixel value at \((i, j)\) in band \( k\) of the frame. Note when RGB color space is used \( k = 1, 2, 3 \) and the pixel values can be in the range \( \{0, \ldots, 255\} \) at each band. Let \( S_k(i, j) \) represent the \( k \) band stego frame, \( M_i, i = 1, 2, \ldots, 10 \) represent the features used in the steganalyzer and finally \( \epsilon_k \) denotes the error over all pixels. Specifically

\[
\epsilon_k = C_k(i, j) - S_k(i, j) = \sum_{k=1}^{K} |C_k(i, j) - S_k(i, j)|^2
\]

will denote the sum of errors in each band at pixel \((i, j)\).

1) Minkowsky Measures:

\[
M_\gamma = \frac{1}{K} \sum_{k=1}^{K} \left\{ \frac{1}{N^2} \sum_{i,j=1}^{N} [C_k(i, j) - S_k(i, j)]^{\gamma} \right\}^{\frac{1}{\gamma}}
\]

\( \gamma = 1 \) corresponds to mean absolute error \( M_1 \) and \( \gamma = 2 \) is the mean square error (MSE), \( M_2 \) respectively.

2) Peak Signal-to-Noise Ratio (PSNR):

\[
M_3 = 10 \log_{10} \left( \frac{\max(C(i, j))^2}{MSE} \right)
\]

where \( \max(C(i, j)) \) is the maximum pixel value and MSE is the mean squared error that can be found using metric \( M_2 \).

B. Correlation Based Measures:

1) Image Fidelity:

\[
M_4 = 1 - \left[ \frac{1}{K} \sum_{k=1}^{K} \frac{\sum_{i,j=0}^{N-1} [C_k(i, j) - S_k(i, j)]^2}{\sum_{i,j=0}^{N-1} [C_k(i, j)]^2} \right]
\]

2) Normalized Cross-Correlation:

\[
M_5 = \frac{1}{K} \sum_{i,j=0}^{N-1} \frac{C_k(i, j)S_k(i, j)}{\sum_{i,j=0}^{N-1} C_k(i, j)^2}
\]

C. Spectral Measures

1) Spectral Magnitude Distortion:

\[
M_6 = \frac{1}{N^2} \sum_{m,n=0}^{N-1} \left| F(u, v) - \hat{F}(u, v) \right|^2
\]

where Discrete Fourier Transforms (DFT) of the original and data embedded (stego) frames be denoted as \( F(u, v) \) and \( \hat{F}(u, v) \) respectively.

D. Histogram Measures

1) K-L Divergence:

\[
M_7 = \sum_i P_i \log \frac{P_i}{H_o(i)} = \sum_i \frac{H_o(i)}{N_{total}} \log \frac{H_o(i)}{H_o(i)}
\]

where \( H_m \) and \( H_o \) represents the modified (stego) and original frame histogram respectively.

2) \( \chi^2 \) (Chi-Square) Metric:

\[
M_8 = 0 \leq \chi^2(H_m, H_o) = \sum_{j=1}^{B} \frac{[H_m(j) - H_o(j)]^2}{H_m(j) + H_o(j)} \leq 1
\]

where \( H_m, H_o \) and \( B \) represent the modified (stego), original frame histograms and number of bins respectively.
REFERENCES


