Video Steganalysis Based on Transform Domain

Thu Thu Htet, Khin Than Mya

University of Computer Studies, Yangon

thuthuhtet@gmail.com,khinthanmya@gmail.com

Abstract

Steganalysis, the method to detect steganographically embedded hidden messages in digital data, has received an increasing interest in recent years. Although significant research efforts have been devoted to develop steganalysis techniques for still-images, video steganalysis remains largely an explored area. This paper proposes a video steganalysis method to detect the presence of hidden messages. Features are extracted from the histograms of the wavelet subbands and the statistical moment of the wavelet characteristic functions Co-Occurrences, water-filling features. K-Nearest Neighbor (K-NN) is utilized as the classifier.

Keywords- steganalysis, histogram characteristic function, statistical moment, exponential entropy, K-NN

1. Introduction

Steganography and steganalysis are like spear and shield. Steganography aims to hide the existence of the messages and steganalysis tries to defeat it. While many researchers make efforts to develop the secure steganographic algorithms, steganalysts try to detect the artifacts which are made by the embedding process. It is very difficult to determine whether there are the hidden messages or not without the cover media. In steganalysis one needs to determine whether the given object has data embedded in it or not.

There are two main categories in current steganalysis: specific or universal. Since the specific steganalysis aims to the targeted steganographic algorithm, its weaknesses can be extracted with the specific strategy and the high detection accuracy can be achieved [9], [11]. In the universal steganalysis, the feature should be modified and classified after embedding the message regardless of the steganographic algorithms [6], [10]. Its performance depends on the efficiency of the feature. The more different the feature is after message embedding, the better it is for the detection.

Even though the universal steganalysis does not perform better than the specific steganalysis from the viewpoint of the detection accuracy, the universal steganalysis scheme is the more practical solution because it can detect the altered images which are hidden by the new embedding algorithm.

Steganalysis can be active or passive [2]. In passive steganalysis the goal is to determine whether a given object is a steganogram or not. In active steganalysis the goal is to estimate the secret message itself. Estimating the hidden message means finding the length or location of the hidden message or the parameters used to create a steganogram (embedding method). Extensive research is done on image and audio steganalysis, but not much is done on video steganalysis.

Steganalysis also attempts to discover more information of the image and hidden message such as the type of embedding algorithm, the length of the message, the content of the message or the secret key used. A less theoretical and more practical categorization of Steganalysis is of the following

i) Targeted Steganalysis: In the case of a known algorithm, an attack that works for that specific algorithm is called Targeted Steganalysis.

ii) Blind Steganalysis: Steganalysis attacks that can be appropriate on all steganographic algorithms are called blind Steganalysis.

iii) Semi Blind Steganalysis: Steganalysis attacks that can apply on a selected set of steganographic algorithms are called semi-blind attacks.

The rest of this paper is organized as follows. In Section 2, some existing Steganalysis methods are explained. In Section 3, the proposed Steganalysis method is explained in detail. The final conclusions are drawn in Section 5.

2. Exiting Steganalysis Methods

In image steganalysis, as one of well-known detectors, Histogram Characteristic Function Center Of Mass (HCFCOM) was once successful in detecting noise-adding steganography [8]. Another well-known method is to construct the high-order moment statistical model in the multi-scale decomposition using wavelet-like transform and then apply learning classifier to the high order feature set [5]. Fridrich et al. presented a method to estimate the cover-image histogram from the stego-image [7]. Another new feature-based steg-analytic method for JPEG images was proposed where the features are calculated as an L1norm of the difference between a specific macroscopic functional calculated from the stego-image and the same functional obtained from a decompressed, cropped, and recompressed stego-image. Shi .et al. [1] proposed a Markov process based approach to effectively attacking JPEG steganography, which have remarkably better performance than general purpose feature sets. Based

on the Markov approach, Liu et al. [12] expanded the approach to the inter-bands of the DCT domains, combined the expanded Markov features and the polynomial fitting of the histogram of the DCT coefficients, and successfully improved the steganalysis performance in multiple JPEG images.

3. Proposed System

The proposed system is a Universal steganalysis method which uses a blind classifier.

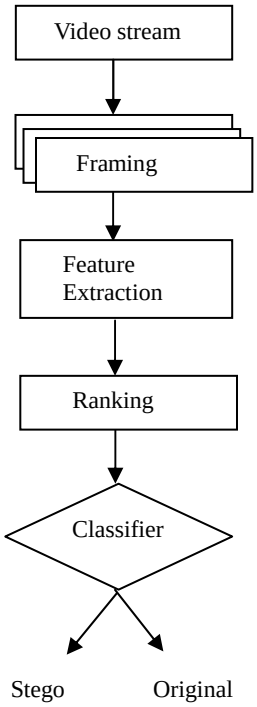

Fig.1. Propose architecture

The proposed system architecture is shown in Fig.1.Firstly visual feature is extracted from video file for feature calculation and each clip is divided into frames. Features are analyzed and extracted by using visual features histogram moments, wavelet moments, Co-Occurrences, water-filling features. In the second stage, features are eliminated by ranking algorithm. Its aims are to select a subset of the original features of a dataset which are rich in the most useful information. The benefits include improved data visualization, transparency, a reduction in training and utilization times. Finally, these features are classified for stego video or original video by k-NN classifier. Classifier accuracy can be increased as a result of feature selection, through the removal of misleading features.

The framework of blind steganalysis for video stream in this work has several modules. The pattern classifier should be able to discriminate between a stego and a cover video based on the input to the classifier, which is the estimate of the message in each frame. The first feature set is extracted by histogram change of the wavelet subbands after data embedding. Let us assume that the histogram of wavelet coefficients can be modeled by Laplacian distribution. The Laplacian distribution is defined as

$$f_K(x) = \frac{\alpha}{2}\exp(-\alpha|x|), \qquad (1)$$

where α is statistical parameter of the Laplacian distribution. Once, the probability distribution function (pdf) is determined, we can obtain the value x' which has area of K on the pdf as follows.

$$\int_0^{x'} \frac{\alpha}{2}\exp(-\alpha x)\,dx = K. \qquad (2)$$

Since the histogram of the cover image is different from the histogram of stego image, the value x' of the stego image and the cover image is also different. For this reason, the value can be a feature to detect the presence of hidden messages. To obtain x', we have

$$x' = -\frac{\log(1-2K)}{\alpha}. \qquad (3)$$

The α is presented in the form of variance σ^2

$$VAR[X] = \frac{2}{\alpha^2} = \sigma^2. \qquad (4)$$

Then, x' can be calculated as

$$x' = -\frac{\alpha\log(1-2K)}{\sqrt{2}}. \qquad (5)$$

In this paper, 3-level wavelet decomposition is performed using the Haar wavelet basis. We extract one feature from 9 high frequency subbands, respectively. Second feature set is the statistical moments of wavelet characteristic functions.

3.1. Moments of characteristic function

It is well-known that an image's histogram is essentially the probability mass function (pmf) of the image (only differing by a scalar). Multiplying each component of the pmf by a correspondingly shifted unit impulse results in the probability density function (pdf). Obviously, in the context of discrete Fourier transform (DFT), the unit impulses can be ignored, implying that we can treat pmf and pdf exchangeable. Thus, the pdf can be thought as the normalized version of a histogram. One interpretation of characteristic function (CF) is that the CF is simply the Fourier transform of the pdf (with a reversal in the sign of the exponent).

Owing to the decorrelation capability of discrete wavelet transform (DWT), the coefficients of different subbands at the same level are kind of independent to each other. Therefore, the features generated from different wavelet subbands at the same level are kind of independent to each other. This property is desirable for steganalysis.

The propose to use the statistical moments of the

CFs of both a test image and its wavelet subbands as features for steganalysis, which are defined as follows.

$$M_n = \sum_{j=1}^{(N/2)} f_j^n \left|H(f_j)\right| \Big/ \sum_{j=1}^{(N/2)} \left|H(f_j)\right| \qquad (6)$$

where H (f_i) is the CF component at frequency (f_i), N is the total number of points in the horizontal axis of the histogram. Note that we have purposely excluded the zero frequency component of the CF, i.e., H (f_0), from calculating the moments because it represents only the summation of all components in the discrete histogram. For an image, it is the total number of pixels. For a wavelet subband, it is the total number of the coefficients in the subband. In either case, it does not change during the data hiding process. As shown below, its exclusion can enhance moments' sensitivity to data hiding.

Denote histogram by $h(x)$,which is the inverse Fourier transform (in the above-mentioned sense) of the CF, *H (f)*.The following formula can be derived straightforwardly.

$$\left|\left(\frac{d^n}{dx^n}h(x)|_{x=0}\right)\right| = \left|(-j2\pi)^n \int_{-\alpha}^{\alpha} f^n H(f)df\right| \leq 2(2\pi)^n \int_0^{\alpha} f^n |H(f)|\,df \qquad (7)$$

This is to say that the magnitude of the n-th derivative of the histogram at x=0 is upper bounded by the n-th moments of the CF multiplied by a scalar quantity (simply stated below as "upper bounded by the n-th moments of the CF"). Using Fourier translation property, it can be shown that this upper bound is also valid for $x \neq 0$ .The moments defined in Equation (6) are non-increasing after data hiding.

3.3. The Water-Filling Algorithm

These propose an algorithm to extract features from the edge map directly without edge linking or shape representation. The idea is to look for measures for the edge length and edge structure and complexity by a very efficient graph traverse algorithm. As water fills the canals (edges), various information are extracted, which are stored as the feature primitives. Feature vectors can then be constructed based on these feature primitives. The time complexity of this algorithm is linear, proportional to the number of edge points in the image.

3.4. Feature ranking

We first rank the features according to their importance on clustering. Feature ranking is efficient since it requires only the computation of n scores and sorting the scores. Statistically, it is robust against overfitting because it introduces bias but it may have considerably less variance [4].Here, the designed ranking index belongs to exponential entropy.

Let $S_{p,q}$ be the similarity between two instances X_p and X_q, and let N be the number of samples on which the feature ranking index is computed. The feature ranking index is defined as:

$$H = \sum_{p=1}^{N} \sum_{q=1}^{N} \left[S_{p,q} \times e^{(1-S_{p,q})} + (1 - S_{p,q}) \times e^{S_{p,q}}\right] \qquad (8)$$

where $S_{p,q}$ takes value in [0.0-1.0] [1]. When $S_{p,q} \rightarrow 0(1)$, H decreases. However, $S_{p,q} \rightarrow 0.5$, H increases. In other words, the index decreases as the similarity (dissimilarity) between two patterns belonging to the same cluster (different cluster) in the feature space, increases. This is appropriate to character the clustering performance of the selected feature set.

For ranking of features we can use H in the following way. Each feature is removed in turn and H is calculated. If the removal of a feature results in the minimum H, the feature is the least relevant; and vice versa. The minimum H indicates the removed feature has the least effect on the distribution of sample in the data set, so it has least influence on the cluster. For the data set with large number of data points, we use a scalable method that is based on random sampling [3]. It should be noted that for H measure working well the cluster structure needs to be retained and should be largely independent of the number of data points. The ranking process is named as RANK.

3.5. K-NN classifier

The design of classifier is another key element in steganalysis. In this work, the *k*-nearest neighbors' algorithm (*k*-NN) is a method for classifying objects. More sophisticated approach, k-nearest neighbor (k-NN) classification, finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. There are three key elements of this approach: a set of labeled objects, e.g., a set of stored records, a distance or similarity metric to compute distance between objects, and the value of k, the number of nearest neighbors. To classify an unlabeled object, the distance of this object to the labeled objects is computed, its k-nearest neighbors are identified, and the class labels of these nearest neighbors are then used to determine the class label of the object.

Figure 2 provides a high-level summary of the nearest-neighbor classification method. Given a training set *D* and a test object $x = (x', y')$ the algorithm computes the distance (or similarity) between *z* and all the tranining object(x, y) $\in D$ to determine its nearest-neighbor list, D_z. (x is the data of a training object, while y is its class. Likewise, $\mathbf{x}'$ is the data of the test object and y' is its class.).Once the nearest-neighbor list is obtained, the test object is classified based on the majority class of its nearest-neighbors:

$$y' = argmax \sum_{(x_i,y_i)\in D_z} I(v = y_i) \qquad (9)$$

where v is a class label, y_i is the class label for the ith nearest-neighbors, and $\boldsymbol{I}(.)$ is an indicator function that returns the value 1 if its argument is true and 0 otherwise.

Input: D, the set of k training objects and test object $z = (x', y')$
Process:
Compute $d(x', x)$, the distance between z and every object, (x, y)$\in$ D.
Select $D_z \subseteq D$, the set of closest training objects to z
.**Output:**

$$y' = argmax \sum_{(x_i,y_i)\in D_z} I(v = y_i)$$

Fig.2. The k-nearest neighbor classification algorithm

4. Conclusion

This paper proposed a general blind steganalysis system for video sequences. The features for steganalysis are extracted from the histograms of the wavelet subbands and the statistical moment of the wavelet characteristic functions. After ranking the process will reduce the redundant features to useful feature. These features are classified by K-NN classifier. K-NN classification is an easy to understand and easy to implement classification technique.

Acknowledgements

This research was supported by SDRC (Software Development and Research Center) in University of Computer Studies, Yangon. SDRC is a Research Center designated by Myanmar Science and Engineering Foundation and Ministry of Science & Technology.

References

[1] Chen, C, W. Chen ,Y.Q. Shi, “A Markov process based approach to effective attacking JPEG steganography”, In Lecture Notes in Computer Sciences, 2007, vol.437, pp.249-264.

[2] Cox, J, J. Kilian, T. Leighton, T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673-1687, December 1997

[3] Dash, M. and H. Liu, “Feature selection for clustering”,Proc. Pacific Asia conf. KDD 2000, pp. 110-121.

[4] Elisseeff, A, I. Guyon and “An introduction to variableand feature selection”, JMLR, 3, 2003, pp. 1157-1182.

[5] Farid, H, S. “Lyu How Realistic is Photorealistic”, IEEE Trans. on Signal Processing, 2005, 53(2): 845-850.

[6] Farid, H, S. Lyu, S. Member, and Steganalysis using higher order image statistics,” IEEE Transactions on Information Forensics and Security, vol. 1, pp. 111–119, 2006.

[7] Fridrich, J, D. Hogeam, M. Goljan, “Steganalysis of JPEG Images:Breaking the F5 Algorithm”, Proc. of 5th Information Hiding Workshop, 2002, pp. 310-323.

[8] Harmsen ,J. J and W.A. Pearlman, “Steganalysis of Additive Noise Modelable Information Hiding”, Proc. of SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents V. 5020, 2003, pp.131-142.

[9] Huang, B. Li, J. and Y. Q. Shi,“ Steganalysis of yass,” Trans. Info. For. Sec ., vol . 4, no. 3, pp. 369–382, 2009.

[10] Lou, D.-C,C.-L. Lin, and C.-L. Liu, “Universal steganalysis scheme using support vector machines,” Optical Engineering ,vol. 46, no. 11, p. 117002, 2007.

[11] Mahdavi, M. S. Samavi, V. Sabeti,and S. Shirani, “*Steganalysis and payload estimation of embedding in pixel differences using neural networks*,” Pattern Recogn., vol. 43, no. 1, pp.405–415, 2010.

[12] Ribeiro, B, A. Sung, Q. Liu, R. Ferreira, “Steganalysis of Multi-class JPEG Images based on Expanded Markov Features and Polynomial Fitting”, 21st Proc. International Joint Conf. Neural Networks, in press