Video Steganalysis Using Histogram and Texture Features

Thu Thu Htet, Khin Than Mya
University of Computer Studies, Yangon
thuthuhtet@gmail.com,khinthanmya@gmail.com

Abstract

Information hiding in video streams with the development of network and multimedia technologies has played an important role in the steganographical field, and correspondingly video stegnanlysis techniques are catching attention of the security department of each government. This paper presents an improved steganalysis technique to detect the presence of hidden messages. In order to identify and classify, the two types of feature are used. The first type feature is texture feature that is computed by statistical, psychological and signal processing. The second type feature is the state of the art histogram based features. This system is classified by using a Support Vector Machine (SVM) .Classification is performed between actual video frames, steganograpyh frames from MSU (unique tool for hiding information in video) allows hiding text file in a video sequence. SVM is excellent classifier for two class problem that give higher detection accuracy rate for this system. Experimental results show that the proposed scheme can effectively detect whether a video has been processed by stego or not.

1. Introduction

The development of IT industry brings much convenience but many information security problems as well. Currently, steganalytic techniques mainly focus on still images [3, 6, 11]; researches on video steganalysis are developing slowly [9, 10]. On one hand, it's because video steganography and steganalysis require research backgrounds on video compression and its complex system; on the other, few well-developed video steganographic software appear in public. Because video resources will be the most important component of future Internet media, video steganography and steganalysis are gradually becoming the highlights in data hiding area. In the 1990s, data hiding techniques arose, which can protect messages by embedding them into innocent-looking cover objects and then transmitting them through open channels without suspects. With the release of some free steganography software, information security issues become more secure, and more attention is paid to steganalysis. However, researches mainly focus on attacking image steganography, and little work has been done on video steganalysis till now.

Current video steganography software fall into two categories, one of which exploits redundancy in file formats, including Steganography, Hider, Max File Encryption, etc; the other embeds messages into video contents, including MSU Stego Video tool, the only one found at present. In this paper proposes a scheme for detecting the information-hiding in videos. Two types of features are extracted from stego and plain sequence such as texture features and the grey level histogram. These features are classified by Support Vector Machine.

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 experiment results are reported in Section 4.The final conclusions are drawn in Section 5.

2. Related Work

Budia et. al [9] proposed a technique for video steganalysis by using the redundant information present in the temporal domain as a deterrent against secret messages embedded by spread spectrum steganography. Their study, based on linear collusion approaches, is successful in identifying hidden watermarks bearing low energy with good precision. The simulation results also prove the superiority of the temporal- based methods over purely spatial methods in detecting the secret message.

Pankajakshan a video steganalysis scheme [10] for the MPEG video coding standard in which a given frame is predicted from its neighboring reference frames using motion compensation. The MPEG coding scheme supports two types of predicted frames: the P frames and the B-frames. The prediction-error frames (PEFs) corresponding to the P and B- frames are then coded using transform coding techniques. The PEFs exhibit spatiotemporal correlation between the adjacent frames. The PEFs of a test video signal are decomposed using the 3-level DWT method and the first three moments of the characteristic functions in each of the sub-bands are computed. The resulting feature vectors are fed to train a pattern classifier to discriminate between the stego and non-stego videos.

One advantage of this scheme is that the PEFs can be obtained directly from a compressed sequence, thereby greatly reducing the computational requirements of the oracle. The drawback of these steganalysis techniques is that it is assumed that each pixel in the frames is modified by the watermark embedding process. However, this assumption may not valid since the steganographer may modify only a portion of the pixels to avoid detection by steganalysis techniques.

3. Proposed System

The general structure of the proposed steganalysis method consists of two main stages: feature extraction and classification. The overview system is shown in Figure 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 (Mean, Variance, Skewness, Kurtosis, Energy, and Entropy) and three very different approaches to computing texture features such as co-occurrence matrices, Tamura and Gabor wavelets. So that can extract 41 features from this video file.

The aim of this proposed system to improved the performance of video steganalysis by using 41 features.

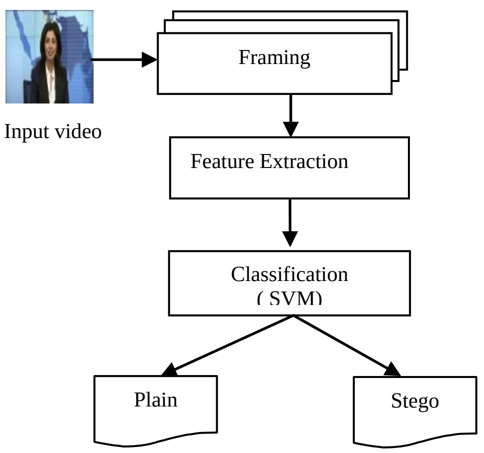

Figure 1. Overview of classification system flow

Finally, these overall 41features are classified for stego video or plain video by using SVM classifier. Classifier accuracy can be increased as a result of feature extraction. SVM is used to detect hidden data.

3.1. Feature Extraction methods

Texture is a key component of human visual perception. Everyone can recognize texture but, it is more difficult to define. Texture has qualities such as periodicity and scale; it can be described in terms of direction, coarseness, contrast and so on [7]. In this system chose three very different approaches to computing texture features: the first takes a statistical approach in the form of co-occurrence matrices, next the psychological view of Tamura's features and finally signal processing with Gabor wavelets.

3.1.1. Co-occurrence Matrices

Statistical features of grey levels were one of the earliest methods used to classify textures. Haralick [1] suggested the use of grey level co-occurrence matrices (GLCM) to extract second order statistics from an image. GLCMs have been used very successfully for texture classification in evaluations [5].

Table.1. Features calculated from the normalized co-occurrence matrix P (*i* , *j*)

Texture Feature	*Formula*
Energy	$\sum_i \sum_j P^2(i,j)$
Entropy	$\sum_i \sum_j P(i,j) log P(i,j)$
Contrast	$\sum_i \sum_j (i-j)^2 P(i,j)$
Homogeneity	$\sum_i \sum_j \frac{P(i,j)}{1+\lvert i-j \rvert}$

Where P =co-occurrence matrix
μ=mean of the co-occurrence matrix P
σ =standard variation of co-occurrence matrix P

3.1.2. Tamura

Tamura et al took the approach of devising texture features that correspond to human visual perception [7]. They defined six textural features (coarseness, contrast, directionality, line-likeness, regularity and roughness) and compared them with psychological measurements for human subjects. The first three attained very successful results and are used in our evaluation.

Coarseness has a direct relationship to scale and repetition rates and was seen by Tamura et al as the most fundamental texture feature. Computationally one first take averages at every point over neighborhoods the linear size of which are powers of 2. The average over the neighborhood of size $2^k \times 2^k$ at the point *(x, y)* is

$$A_k(x,y) = \sum_{i=x+2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y+2^{k-1}}^{y+2^{k-1}-1} \frac{f(i,j)}{2^{2k}} \qquad (1)$$

Then at each point one takes differences between pairs of averages corresponding to non-overlapping neighborhoods on opposite sides of the point in both horizontal and vertical orientations. In the horizontal case this is

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)| \qquad (2)$$

where k maximizes E in either direction. The coarseness measure is then the average of $S_{opt}(x, y) = 2^{kopt}$ over the picture.

Contrast aims to capture the dynamic range of grey levels in an image, together with the polarization of the distribution of black and white. The first is measured using the standard deviation of grey levels and the second the kurtosis α_4. The contrast measure is therefore defined as

$$F_{con} = \sigma/(\alpha_4)^n \quad where\ \alpha_4 = \frac{\mu_4}{\sigma^4}, \qquad (3)$$

μ_4 is the fourth moment about the mean and σ^2 is

the variance. Experimentally, Tamura found
n =1/4 to give the closest agreement to human measurements.

Directionality is a global property over a region. The feature described does not aim to differentiate between different orientations or patterns, but measures the total degree of directionality. Two simple masks are used to detect edges in the image. At each pixel the angle and magnitude are calculated. A histogram, H_d, of edge probabilities is then built up by counting all points with magnitude greater than a threshold and quantizing by the edge angle. The histogram will reflect the degree of directionality. To extract a measure from H_d the sharpness of the peaks are computed from their second moments.

3.1.3. Gabor

One of the most popular signal processing based approaches for texture feature extraction has been the use of Gabor filters. These enable filtering in the frequency and spatial domain. It has been proposed that Gabor filters can be used to model the responses of the human visual system. Turner [5] first implemented this by using a bank of Gabor filters to analyze texture.

This system implementation is based on that of Manjunath et al [8, 4]. The feature is computed by filtering the image with a bank of orientation and scale sensitive filters and computing the mean and standard deviation of the output in the frequency domain. Filtering an image $I(x, y)$ with Gabor filters g_{mn} designed according to results in its Gabor wavelet transform:

$$W_{mn}(x,y) = \int I(x,y) g^{*}_{mn}(x - x_1, y - y_1) dx_1 dy_1 \qquad (4)$$

The mean and standard deviation of the magnitude $|W_{mn}|$ are used to for the feature vector. The outputs of filters at different scales will be over differing ranges. For this reason each element of the feature vector is normalized using the standard deviation of that element across the entire database.

3.2. Gray level histogram

The aim of using the features is to obtain better feature vectors which can represent the characteristics of an image by considering statistical features of its pixels. The histogram tells us something about the distribution of the grey level. There are five aspects which will be analyzed for these features. Those aspects are mean, entropy, energy, variance, skewness, and kurtosis of the image histogram. The mean is computed from the image histogram by using this formula:

$$\mu = \sum_{h=0}^{L-1} f_h \, p(f_h) \qquad (5)$$

where f_h =the value of intensity level
$p(f_h)$=the first-order histogram probability

Variance can be computed by using this formula:

$$\sigma^2 = \sum_{h=0}^{L-1} (f_h - \mu)^2 p(f_h) \qquad (6)$$

Skewness from an image histogram can be obtained by using this formula:

$$skew = \frac{1}{\sigma^3} \sum_{h=0}^{L-1} (f_h - \mu)^3 p(f_h) \qquad (7)$$

Kurtosis can be computed by using this formula:

$$kurtosis = \frac{1}{\sigma^4} \sum_{h=0}^{L-1} (f_h - \mu)^4 p(f_h) - 3 \qquad (8)$$

Image histogram entropy measures the shape irregularity of the histogram curve. If the image is a simple image that has a good distribution on various intensity levels, the histogram of that image will have a small kurtosis value, and vice versa.

3.3. Standard SVMs classification of two-class

Supposing a set of training data,

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\} \qquad (9)$$

Where $x_i \in R, y_i \in \{-1,1\}$, and y_i is a class label.

The purpose of SVMs is to find decision function) f(x) by training, which can construct separating hyper-planes to classifier two different samples and maximize the margin. That is to say to solve the problem of quadratic programming [2], described as follows:

$$min\varphi(\omega, \xi) = \frac{1}{2}(\omega, \omega) + C \sum_{i=1}^{m} \xi_i$$
$$s.t. \qquad y_i[(\omega \bullet x_i) + b] \geq 1 - \xi_i$$
$$\xi_i \geq 0, i = 1,2, \ldots m \qquad (10)$$

It can be resolved into extreme value problem of quadratic function by seeking dual problem of the above equation:

$$max \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j \, K(x_i, x_j)$$
$$0 \leq \alpha_i \leq C, i = 1,2, \ldots, m$$
$$s.t. \qquad \sum_{i=1}^{m} \alpha_i y_i = 0$$
$$f(x) = sign\left[\sum_{i=1}^{m} \alpha_i y_i \square (x, x_i) + b\right] \qquad (11)$$

4. Proposed Classification and System Evaluation

The remaining point of the article will consequently be devoted to the detection of video steganography. The experimental study and results are reported in this section.

4.1. Experimental Study

In this system, 40 video samples taken from different movies each of duration 10 seconds. The secret text message is embedded using MSU stego video tool. MSU Stego Video (Unique tool for hiding information in video) allows hiding text file in a video sequence. When some hidden data is embedded in a cover video sequence, the encoding of the cover video sequence is analyzed and an algorithm is chosen which provides small data loss after video compression. MSU Stego Video supports multiple video compression formats. Data embed in video sequence using MSU Stego Video. It uses a specific data distribution mode to load secret data, so that the message can be extracted correctly even the stego video are compressed. In the experiment, the total frame number of these video sequences is 10000. The features are extracted from the plain and stego videos. After that, classifier train this features and classify the testing video sequences.

4.2. Experimental Result

Detailed experimental studies have been conducted to evaluate the performance of the proposed steganalysis scheme. In the first scheme texture and histogram features are extracted from each frame of the test sequence. This system captured different video sequences from different sources and the same video sequences from same sources, including the movies and on-line videos from CNN and YouTube. The classifier is trained with feature vectors extracted 60% of the frames in each sequence and the remaining frames are used for testing the classifiers. For cross-validation, the frames for training and testing the classifiers are randomly chosen and the reported results are the average of 100 experiments, each with a different training set. In this system, utilize a SVM classifier with *Gaussian* kernel to construct the classification model on the training and testing dataset. Figure 2 lists three frame samples.

(a) Different videos

(b) Same videos

Figure2. Three frame samples of Videos

The result is measured by classification accuracy defined as the number of correctly classified clips over total number of clips. Figure 3 shows testing results, consisting of True Negative (TN), False Positive (FP), True Positive (TP) and False Negative (FN). Training samples and testing samples are from different video chips. That is, if some frames from a certain chips are trained for setting up the classification model, other frames in the same chip are not permitted as testing samples.

If P and N denote the real number of positive and negative instances, and TP and FP denote the predicted number of true positives and false positives, respectively, then the true positive rate t_p is defined as:

$$t_p = \frac{TP}{P} \quad , \qquad (12)$$

and the false positive rate, f_p as:

$$f_p = \frac{FP}{N} \quad , \qquad (13)$$

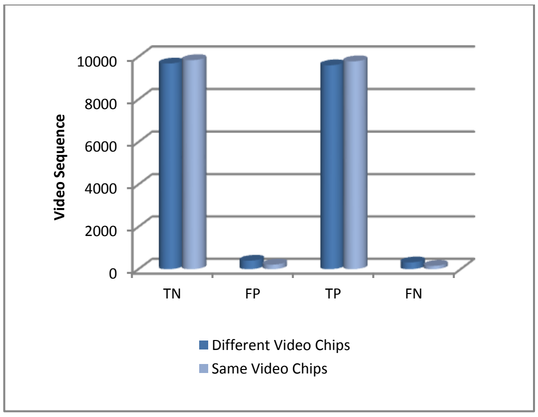

Figure 3. Testing Results for Different and Same Videos

Figure 4 describes the accuracy of testing videos. The different numbers of bits are embedded in these videos. There are a number of performance measures that are of interest in steganalysis. The most common measures are the false positive and false negative rates.

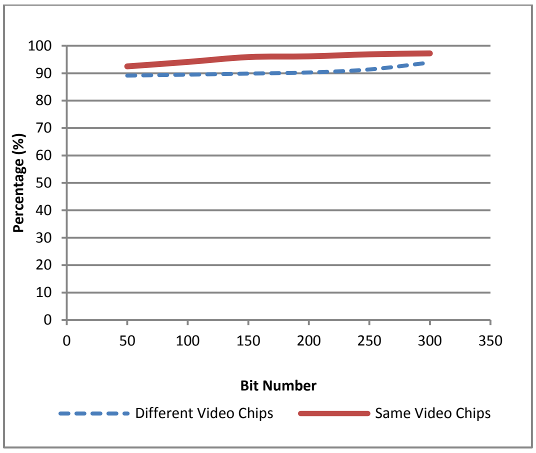

Figure 4. Accuracy Rate for Different and Same Videos

Receiver Operation Characteristics (ROC) curves to evaluate and compare classifiers. ROC is a graphical plot between the sensitivity and specificity. The ROC is used to represent the plotting of the fraction of true positives (TP) versus the fraction of false positives (FP).

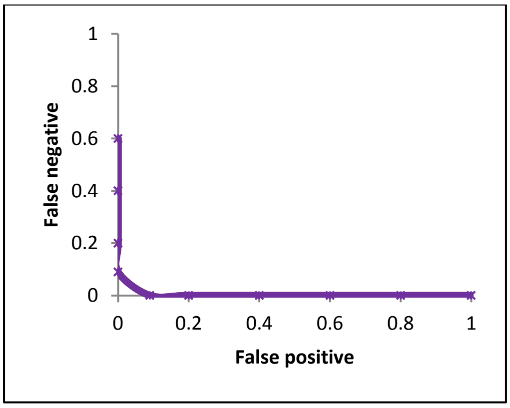

Figure 5.ROC curves for same video chips

Figure 5 shows ROC curves under the condition of training samples and testing samples from same video chips. The point (0, 1) is the perfect classifier, since it classifies all positive cases and negative cases correctly. Thus an ideal system will initiate by identifying all the positive examples and so the curve will rise to (0, 1) immediately, having a zero rate of false positives, and then continue along to (1, 1).

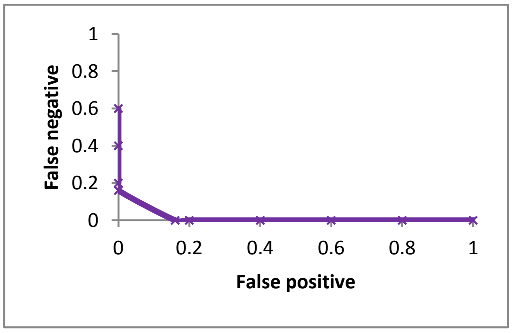

Figure 6.ROC curves for different video chips

Figure 6 shows the ROC curves under the condition of training samples and testing samples from different video chips. In each of these ROC plots, the x axis is the false alarm rate, calculated as the percentage of normal video frames considered as steganograms; the y-axis is the detection rate, calculated as the percentage of steganograms detected. The classification accuracy is defined as the ratio of the number of correctly classified clips to total number of clips in respective class and error rate is defined as the ratio of the number of incorrectly classified clips to total number of clips in respective class.

$$\text{Classification Accuracy} = \frac{\text{Correctly Classified Samples}}{\text{Classified Samples}} \quad (14)$$

$$\text{Classification Error Rate} = \frac{\text{Incorrectly Classified Samples}}{\text{Classified Samples}} \quad (15)$$

5. Conclusion

In this paper proposed a scheme to detect the information-hiding in MSU Stego-Video based on Texture feature and histogram features. The Features are analyzed and extracted from video sequences. These features may also be used to improve the classification performance of a classifier. Classification is performed between actual video frames, steganograpyh frames from MSU tool. Then these features are classified by SVM classifier. The average detection accuracy rate is above 96.61%. In the future, the characteristics of intra-frame and inter-frame in the MSU Stego Video to improve the detection result, and design more precise detection algorithm to find the hidden text file in the video resources.

References

[1] Haralick, R.: Statistical and structural approaches to texture. Proceedings of the IEEE 67 (1979) 786–804.
[2] J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2(1998)2, 121-167.
[3] J.Fridrich, and M.Goljan, "Practical steganalysis: state-of-the-art", Proceeding of SPIE Photonics West, Electronic Imaging 2002, San Jose, California, vol.4675, January 2002:1-13.
[4] Manjunath, B., Ma, W. Texture features for browsing and retrieval of image data.IEEE Trans on Pattern Analysis and Machine Intelligence 18 (1996) 837–842
[5] Ohanian, P., Dubes, R.: Performance evaluation for four classes of textural features. Pattern Recognition 25 (1992) 819–833
[6]. R.Chandramouli, K.P.Subbalakshmi, "Current trends in steganalysis: a critical survey", International Control, Automation, Robotics and Vision Conference 2004, Volume 2, December 2004:964 – 967.
[7]. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visualperception. IEEE Trans on Systems, Man and Cybernetics 8 (1978) 460–472
[8]. Turner, M.: Texture discrimination by Gabor functions. Biological Cybernetics 55 (1986) 71–82 10. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of image data.IEEE Trans on Pattern Analysis and Machine Intelligence 18 (1996) 837–842

[9]. U. Budhia, D. Kundur, and T. Zourntos, “Digital video steganalysis exploiting statistical visibility in the temporal domain”, IEEE Transactions on Information Forensics and Security, 2006, 1vol.1 pp.43-55.

[10]. V. Pankajakshan, A.T.S Ho, “Improving video steganalysis using temporal correlation”. Intelligent Information Hiding and Multimedia Signal Processing, 2007.Taiwan, Volume1, pp.287-290.

[11]. Wen-Nung Lie, Guo Shiang Lin, “A feature-based classification technique for blind image steganalysis”, IEEE Trans. on multimedia, 2005, 7(6):1007-1020.