Research on Uncertainty of audio and Video Information Hiding Based on Semantic and Statistical Moment

Jiexin Zhang
Department of Computer Science, GuangXi Economic Management Cadre College, GuangXi Nanning 530007
Email: zjx7689@126.com

***Abstract*—nowadays, audio and video media data is already facilitates generation, transmission, storage and circulation on the global scale. Audio and video data is geometrically fast as the rate of growth, the video data processing and analysis have lagged behind the pace of development in the growth of data, resulting in large amounts of data is wasted. Therefore, it becomes an urgent need for efficient retrieval of video data content. Accordingly the SS hiding effectively, verify the presence of the secrete message in an important issue. In this paper we present two statistical analysis algorithms for SS hiding. Both the two methods are based on machine learning theory and discrete wavelet transform (DWT), which adopts the classification technology. In the algorithm I, we introduce Gaussian mixture model (GMM) and generalize Gaussian distribution (GGD) to character the probability distribution of wavelet sub-band. Then the absolute probability distribution function (PDF) moment is extracted as feature vectors. We use GMM to model the probability distribution of wavelet coefficient and calculate the absolute moment of statistical distribution as feature vector of each sub-band for statistical analysis. In the algorithm II, we propose distance metric between GMM and GGD of wavelet sub-band to distinguish cover. The experiment results of both two proposed classification algorithms may obtain better detecting performance. The probability distribution model takes GMM and GGD. We use de-noising method to get the estimation of cover audio, and then use four distances metric to measure the distortion.**

***Index Terms*—audio and video, information hiding, semantic, statistical moment**

I. Introduction

Digital watermarking and data hiding have become a vibrant research area. Various kinds of multimedia files can be downloaded freely from the Internet. Terrorists might have seen this as an opportunity to communicate secretly with each other. Thus, various methods have emerged as means to detect covert communication by terrorists [1]. Statistical is the scientific technology to decide whether a medium carries some hidden messages or not. In addition to preventing secret communication among terrorists, statistical analysis serves a method to judge the security performance of statistical techniques [2].

Spread spectrum technology has developed rapidly in the area of information hiding. As to audio statistical analysis, the hiding algorithm based on spread spectrum has accepted extensive attention owing to the advantages of good robustness and immunity to noise attack [3]. Consequently detecting the spread spectrum hiding effectively, verify the presence of the secrete message reliably and taking action accordingly is an important issue in information hiding field.

Video content retrieval can be divided into three levels from bottom to up: feature layer retrieval, object layer retrieval and semantic conceptual layer retrieval [4]. Feature layer concern is the video image in a particular visual features, spatial characteristics and composition. Object layer concerns the image of the target object and the temporal relationship. Semantic conceptual layer concerns the video images were abstract, the idea, and formation of conceptual content of thought. But people during the course of the video content retrieval, video content to determine the similarity of the image is not confined to the similarity of low-level visual features, often also concerned about the subjective thinking activities from the image of high-level semantic abstraction of concepts such as events, scene places, activities, objects and so on [5].

Usually, the statistical algorithms for image cover can achieve a satisfactory detection results, regardless of the underlying embedding algorithm. But these methods are not directly appropriate for audio cover, due to different characteristics of audio signals and images. The statistical regularities captured in image signals are inherent to the spatial composition of images that are simply not present in audio [6]. As to audio the statistical field, all these algorithms attempt to find good feature vectors from time domain or transform domain which are used to capture statistical changes caused by data embedding. But these features are not effective to spread-spectrum the statistical embedding [7]. In this paper, we aim to propose two effective algorithms of extracting feasible feature vectors for spread-spectrum hiding as to get higher detecting accuracy (lower false alarm rate and lower false negative rate), compared with the present the statistical algorithms for DSSS hiding. Both the two proposed algorithms are based on machine learning theory. The paper is organized as follows. Section 2 we introduce the spread spectrum coding. Sections 3 and Section 4 we describe our proposed algorithms and give detailed theoretic analysis. Section 5 we present the experiments results. Section 6 we conclude the paper.

doi:10.4304/jsw.6.6.1154-1162

II. Related Work

In 1951, Calvin Moores first proposed the concept of information retrieval [8]. He thinks information retrieval embraces the intellectual aspects of the description of information and its specification for search， and also whatever systems， techniques， or machines that are employed to carry out the operation. Video content retrieval is a vibrant branch of information retrieval. In 1992, US National Science Foundation presented a research task on visual information management system [9]. To this meeting as a dividing point in time, prior to 1991 focused on image and video content retrieval literature are rare, and then more and more research organizations, agencies conducted extensive research and exploration, a variety of results emerging out.

In 1995, IBM Almaden Research Center established QBIC system [10], which is the first image-based commercial video content retrieval of a typical system, implemented based on example, a user-drawn sketches and color, texture, lens, and target motion, etc. check the video features, the establishment of a high dimensional feature indexing, and implemented a text-based keyword search and content-based similarity search. In 1998, Columbia University proposed a Visual SEEK and Video Q system. Study of these two systems is the spatial relationship between image area query and visual features extracted from the compressed domain [11]. System uses a set of visual features in color and texture features based on wavelet transform, to speed up the retrieval process, using a search algorithm based on binary tree.

In 2000, Urbana-Champaign developed the MARS system [12]. Study of this system is not limited to find the single best description, but rather concerned with how to organize various visual features to retrieve the structure of a meaningful system, and hope that this system can be applied to different applications and different users. In 2006, Excalibu technology developed content-based image retrieval tool retrieval ware, its early focus on the neural network for image retrieval, the improved search engine can make use of texture, color, structure and aspect ratio as the search feature, and support for these features combination [13].

Fraid H [14] proposed detection algorithm based on higher-order statistics of wavelet coefficients and support vector machine. In this method, a statistical model (mean) based on the first and high-order magnitude statistics (variance, skewness and kurtosis) is used as features for image statistical analysis detection. For EzStego, OutGuess and Jsteg, experimental results show the effectiveness of the selected statistics. Harmsen [15] established a common model to additive noise steganography, providing (supplying)a theoretical basis for statistical analysis detection. According to the additive theory, specific variety algorithm can be designed for time-domain and transform domain, but using histogram characteristic function center of mass to determine whether the secret data, its accuracy is not satisfactory. Kalpana Seshadrinathan [16] concluded that there are significant difference between statistical distortion caused by spread spectrum and distortion caused by normal signal processing (compression, Gaussian blur, white Gaussian noise). So the author takes advantage of the natural scene (the statistical distribution of wavelet transform coefficients model) to effectively distinguish from the dependent performance of image noise. The shortcoming of this method is that it needs the original image as a reference. Kenneth Sulliva [17] has used the markov random chain to calibrate the correlation of pixels in images, and statistically analysis of the empirical matrix. However, this method doesn’t work on audio signal because of the empirical matrix can not perceive the difference caused by information embedding and the features are not sensitive to the embedding. Ying Wang [18] pointed out that Block-DCT Statistical analysis algorithm introduces non-stationary nature into cover image, making the distribution difference of adjacent pixels between inter block and intra block. Then the author uses Kolomogrov-smirnov tests to determine whether the image is hidden information. The disadvantage is that the default block distribution is 8*8, and pre-assuming carrier image is two-dimensional stationary process. This condition is very harsh in the real image. Shi et al [19] proposed a Markov-process based on approach to detect the information hiding behaviors in JPEG images. Markov process is applied to modeling difference JPEG 2-D arrays. The experimental works has outperformed the existing statistical analyzers in attacking OutGuess, F5 and MB1. Based on the Markov approach, Liu et al [20] expanded the Markov features to the inter-blocks of the discrete cosine transform (DCT) domains, and calculated the difference of the expanded Markov features between the testing image and the calibrated version and combined the expanded features and the polynomial fitting of the histogram of the DCT coefficients as detectors. The method successfully improved the statistical analysis performance in multiple JPEG images. Liu [21] proposed an effective statistical analytic technique based on statistical moments of differential characteristic function. The method calculates the first, second and third order partial differentiations, the first and second order total differentiations at pixel-locations in the test image. 18 statistical were computed from these five objects and intensity in image utilizing histogram and co-occurrence matrix. The presented method demonstrates higher detecting rates with lower false positives for Cox and Piva statistical. Other work on image statistical analysis based on high-order statistics have been done by Fridrich [22], Shi [23], Lyu and Farid [24]. The idea of using distortion measure to classify cover signal and statistical signal was introduced by Avcibas et al. [25]. Some image quality metrics have been identified based on the analysis of variance technique as feature sets to distinguish between cover-images and statistical-images. The classifier between cover and statistical-images is built using multivariate regression on the selected quality metrics and is trained based on an estimate of the original image.

Many research groups have investigated audio steganalysis. Ru et al. [26] presented a method based on negative resonance phenomenon. The proposed method can be very effectively used to detect hidden messages embedded by Hide4PGP, Stegowav and S-Tools4. Hamza Ozer [27] introduced audio statistical analysis algorithm based on audio quality distortion and classifier, but the accuracy of detection is directly related to the training results of audio quality indicators in audio database, resulting poor performance and high complexity. Avcibas [28] designed a feature set of content independent distortion measures for classifier design. By removing content dependency during the distortion measurement, this paper shows that the discriminatory power is enhanced. Oktay Altun [29] has introduced a method based on diminishing marginal distortion. The experiment results are good for the Gaussian random noise. Johnson [30] proposed statistical model through building a linear basis that captures certain statistical properties of audio signals. A low-dimensional statistical feature vector is extracted from this basis representation and used by a non-linear support vector machine for classification. This paper shows the efficiency on LSB embedding and Hide4PGP. Kraetzer and Dittmann [31] proposed a mel-cepstrum based analysis to perform detection of embedding hidden messages. Recently, Kraetzer [32] introduces the information fusion technology into audio statistical analysis. The results show for the test cases an increase of the classification accuracy for statistical analysis algorithms by math level fusions, no gain by decision level fusion and a considerably small impact of the key selection assumption on the statistical detectability. By expanding the Markov approach proposed by Shi et al. [33] for image statistical analysis, Liu et al.[34] designed expanded Markov features for audio statistical analysis. Yali Liu [35] proposed a novel distortion metric based on Hausdorff distance. The distortion measurement is obtained at various wavelet decomposition levels from which the high-order statistics are derived as features for a classifier to determine the presence of hidden information in an audio signal. Extensive experiment results for Least Significant Bit (LSB) substitution based statistical analysis tool shows that the proposed algorithm has a strong discriminatory ability. Additionally, to improve audio statistical analysis method [36], Liu [37] presented a method based on Fourier spectrum statistics and coefficients, derived from the second-order derivative of the audio signal. Experiment results show that the proposed method improves the detection accuracy on Invisible, Hide4PGP, LSB matching, and Steghide.

III. The Color Named Mapping Conversion Model

In the classical set theory, an element of a relationship and a set A there is only a part A and A does not belong to both identify a situation. Collection can be described by the characteristic function. Each collection A has a characteristic function.

This is the classical set theory a reflection on the law of excluded middle, but the support of fuzzy set theory is the case of violation of law of excluded middle, that in one or the other, there is also vital in the third case, in a way that elements of a belongs to set A. The degree of belonging to [0,1] to represent a number between, this value is called the membership degree, with a membership degree to describe the intermediate state, which introduced the concept of fuzzy sets, fuzzy set is defined as: Let U be a universe of discourse, the domain U to the real interval [0,1] random map.

Specific to the color name mapping, the color characteristic value can be adopted to deal with the membership function of a color value according to their degree of membership described, for example, a color value for the RGB (15,150,200), corresponding to different a subset of the semantic value of the possibility of color measure described as:

$$f(x) = RGB(15,150,200) \rightarrow \begin{array}{l} f(x, blue) = 0.72 \\ f(x, green) = 0.65 \\ f(x, red) = 0.05 \\ f(x, yellow) = 0.23 \end{array} \tag{1}$$

It indicates that the blue color value is the possibility of x 0.72, the formation of the fuzzy feature values can be used for color matching and queries.

A. Latent Semantic Retrieval Model Principle

It said the concept of using semantic vector space, the latent semantic indexing method can achieve the elimination of the correlation between words, the purpose of simplifying the document vector. The basic idea is that the starting point to retrieve document data in the vocabulary for a link between, that is some kind of implicit semantic structure, semantic structure of such statistical methods can be used to find and use by the latent semantic structure to represent words and documents. In the latent semantic indexing method, the index entry on the document by the singular value decomposition of matrix, dimension reduction can be generated that contains the concept of a number of orthogonal factor space; this space reserved for feature information and the original document matrix consistency index entries, but also reflects the semantic structure of the entire document collection, the document focused on the expression of lexical information mainly related model, eliminating one of the specific words for different forms of change brought word interference information.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} = \left(X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot m} \right) = \begin{pmatrix} X_{1\cdot} \\ X_{2\cdot} \\ \vdots \\ X_{n\cdot} \end{pmatrix} \tag{2}$$

B. Semantic Rules of Uncertainty Reasoning Based on Cloud Model

Extract video semantic concept, the semantics for the random concept, the concept of the scope of its boundary values may be vague, and therefore not suitable for a fixed threshold classifier, in order to solve the problem of classification and recognition of these uncertainties, you need to explore the use of cloud model approach to deal with uncertainties video semantic rule-based reasoning. Cloud model is a new theory, which is composed of cloud model, reasoning under uncertainty, and cloud transform. Cloud theory will combine fuzziness and randomness to solve the membership function of fuzzy set theory the concept of the inherent defects in the data analysis of quantitative and qualitative mapping approach laid the foundation.

C. The Definition of Cloud Model

Cloud model is a natural language that the qualitative value of the concept and its quantitative expression of uncertainty between the conversion models. Let U be an accurate numerical representation with the domain (one-dimensional, two-dimensional or multidimensional), U the corresponding qualitative concept, for an arbitrary element of the domain x, there exists a stable tendency of the random number called the degree of certainty of the concept x, x in the domain of the distribution is called the cloud model.

D. The Algorithm of Cloud Generator

Many cloud droplets from the clouds, each cloud droplet is a qualitative concept of a map to the number of domain space (one-dimensional, two-dimensional or multidimensional), a point of achieving a quantifiable; this implementation with uncertainty , is given the point cloud model can represent the qualitative concept of certainty. The universality of the normal distribution, so that the most basic normal cloud as the cloud, it is from qualitative to quantitative mapping of natural language expression of the value of the most basic language is very useful. When the number corresponding to the concept of one-dimensional domain, the normal cloud generator algorithm is as follows:

(1). input

The three qualitative features of represents the concept. That is Ex , En ,and He . The number of cloud droplets N.

(2). output

The quantitative value of cloud droplets N and representatives of each of the concept is cloud droplets to determine the degree.

(3). algorithm steps

The first step: generating a normal standard value En'. En is expectations, He is standard deviation.

The second step: generating a normal standard value *x*. En is expectations, En' absolute value is standard deviation.

The third step: *x* is the concept that a specific qualitative quantitative value, known as cloud droplets.

The fourth step: calculation $y = e^{\frac{-(x-Ex)^2}{2(En')^2}}$.

The fifth step: Y to x is that the concept of certainty qualitative.

The Sixth step: complete the current reflected by this time, the entire contents of the qualitative and quantitative transformation.

The algorithm uses two normal random number generations, which was the first random number random number generator when the second take, which is the normal cloud this algorithm to generate the key. Cloud generation algorithm is implemented using software, which you can also use the hardware in the form of implementation, the cloud generator, the positive generator as follows in Fig.1:

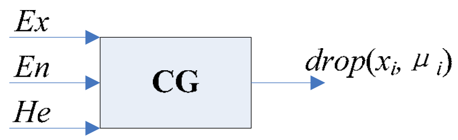

Figure 1. One-dimensional forward cloud generator

IV. STATISTICAL ALGORITHM 1

Statistical analysis is essential a detection problem. Given a carrier signal X , assuming that the attacker knows the probability density function P_0 and the statistical-signal probability density function P_1 , then the problem becomes into binary hypothesis testing, that is:

$$\begin{cases} H_0 : X \sim P_0 \\ H_1 : X \sim P_1 \end{cases} \qquad (3)$$

This detection problem can be Bayesian estimation, minimum and maximize detection, Neyman-Pearson test etc. If the statistical-signal has the existence of possible multiple distributions, then the test will be more complicated. Related entropy (Kullback-Leibler Divergence) is often used to discriminate the probability distribution. Defined as formula:

$$D(P_0 \parallel P_1) \triangleq \int_x P_0(x) \log \frac{P_0(x)}{P_1(x)} d_x \qquad (4)$$

It is associated with the probability of error detection, the value of $D(P_0 \parallel P_1)$ is the greater, the smaller the probability of error. Thus, usually using information theory to define the security of statistical analysis systems (detection capability) often takes the method. Only if:

$$D(P_0 \parallel P_1) = 0, i.e., P_0 \equiv P_1 \qquad (5)$$

Systems have an excellent safety performance. Under ideal conditions, following embedding of the message into the cover signal, the resulting statistical signal is required to have exactly the same probability distribution as the cover signal, and then no statistical test can reliably detect the presence of the hidden message. We refer to such statistical schemes as perfect secure. But in reality, statistical analysis is frequently faced with a dilemma that is unknown. Therefore, the statistical personnel may not be able to design a beautiful embedding function to get the perfect security features; accordingly statistical -analyst can not get a reliability detection performance.

In statistical analysis test of this hypothesis, there are usually two types of errors: false alarm rate and false negative rate. The corresponding error probability is expressed as α and β . Usually these two types of errors are difficult to be estimated. The threshold method was given as:

$$\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta} \le D(P_1 \parallel P_0) \qquad (6)$$

This type of errors include two types of threshold hypothesis testing related to the probability distribution of entropy: greater related entropy means the stronger detection capability. In order to make statistical analysis system security, statistical analysis needs to reduce the related entropy, or even make it to be zero to obtain a perfect secure statistical analysis system. On the contrary, in order to design good statistical analysis algorithm, we need to look for characteristics of the probability distribution of carrier signal and the statistical analysis. While on the signals (images, audio, etc.) modeling, recent studies have made great progress, but on a unified model of the signal has not been established. However, given the situation that contains two kinds of signals (the original carrier signal and the statistical analysis signal); this problem can be solved through supervised learning approach. Therefore, statistical analysis faces enormous challenges in order to avoid changing the statistical features of cover signal when secret message has been embedded; conversely, statistical analysis is to seek the statistical difference between the feature vectors caused by information hiding.

To sum up, as to how to distinguish between carrier signal and statistical-signal effectively, there are some common problems that is extracting feature vector and giving mathematical explanation. Next, this article will give the core proceeding including multi-scale decomposition, probability distribution models, feature vector extraction and optimization.

In this framework I, we aim to analyze the difference of statistical distribution in wavelet domain; the whole idea takes supervised learning. After extracting the feature vector, classifier is used to the distinguish carrier signal and the statistical-signal. The overall framework is as follows in Fig.2:

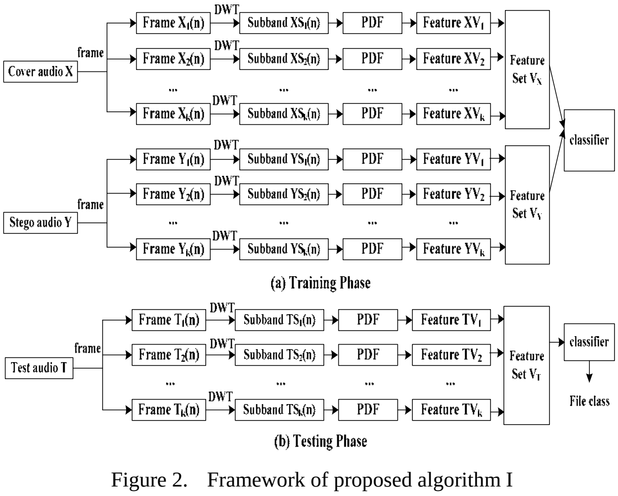

Figure 2. Framework of proposed algorithm I

The training and testing procedures for the statistical analysis is shown in Fig. 2. Let X and Y be the cover audio and statistical audio, respectively. Before the statistical analysis, the audio signal is divided into blocks with same length M . Assume that these three kinds of audio with same sample length L , then the total number k of segments is $\lfloor L/M \rfloor$.

Then in training phase, we can get frame signals set $\left(\mathrm{X}_1(n),\mathrm{X}_2(n),\ldots,\mathrm{X}_k(n)\right)$, $\left(\mathrm{Y}_1(n),\mathrm{Y}_2(n),\ldots,\mathrm{Y}_k(n)\right)$, where $1 \le n \le M$. After the DWT operation on each frame signal, the detail sub-bands are extracted. Let $\left(\mathrm{XS}_1(n),\mathrm{XS}_2(n),\ldots,\mathrm{XS}_k(n)\right)$, $\left(\mathrm{YS}_1(n),\mathrm{YS}_2(n),\ldots,\mathrm{YS}_k(n)\right)$, where $1 \le n \le M/2$, represent the sub-band signal of cover audio and statistical audio, respectively. Then we choose the appropriate probability density function to model the statistical distribution of each sub-band and extract effective statistical features. The feature set $\mathrm{V_X}$, $\mathrm{V_Y}$ for cover audio and statistical audio is $\left(\mathrm{XV}_1,\mathrm{XV}_2,\ldots,\mathrm{XV}_k\right)$ and $\left(\mathrm{YV}_1,\mathrm{YV}_2,\ldots,\mathrm{YV}_k\right)$. Finally both the feature sets are transported into classifier as to get training model. In the testing phase, the test audio file T will go through the same procedures of segmentation, DWT module, PDF module until the feature sets $\mathrm{V_T}$ is achieved, and then it is used to judge the test file type with training model. The core steps are PDF model and feature extraction which will be discussed in detail in the following section.

The wavelet transform is a technique for analyzing signals. It was developed as an alternative to the short time Fourier transform (STFT) to overcome problems related to its frequency and time resolution properties. More specifically, unlike the STFT that provides uniform time resolution for all frequencies the DWT provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies. Through the wavelet transform, the signal is decomposed into low-frequency components and high frequency components. The high-frequency components maintain the details of the audio signal, and it is sensitive to external noise and embedding information. In the spread-spectrum statistical analysis, the hidden information after being modulated by pseudo-random sequence extensions is embedded into the audio signal. The embedding operation interferes most of the band and changed the detail sub-bands of the signal in the wavelet domain, resulting the changes of coefficients in high-frequency components.

As a dollar evolution of Gaussian probability density function, Gaussian mixture model (GMM) can be approximation of any probability density distribution of arbitrary shape, which is widely used in speech recognition. This paper also uses it to model the wavelet coefficients. In general, the Gaussian mixture distribution model can be the following using limited form of distribution and said:

$$f_k(x) = \sum_{j=1}^{k} \pi_j \phi\left(x,\theta_j\right) \tag{7}$$

Here, $\phi\left(x,\theta_j\right)$ is the j^{th} component of GMM model, θ_j is the vector of the mixture parameters which consists of weight π_j , mean μ_j , variance σ_j^2 . The weight π_j must satisfy:

$$\pi_1 + \ldots + \pi_k = 1 \, , \, \pi_j \ge 0 \tag{8}$$

In this paper, after the details of DWT decomposition, we also use greedy EM algorithm to GMM modeling. Here we test sample "blues.wav" (music sample). In statistical field, in order to get a better estimated distortion, the length of the signal sub-frame should not be too long. Here, the sub-frame length is set to 1024 samples and the number of Gaussian components is set to 3. The largest number of iterations takes 20 times after GMM training, three Gaussian components of parameters in Fig.3.

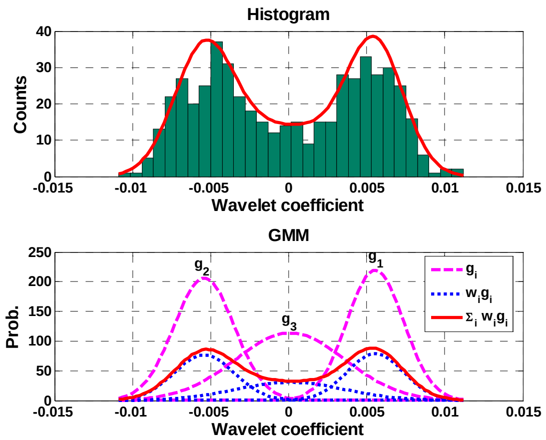

Figure 3. Histogram of wavelet coefficients and GMM model

Fig.3 shows the histogram and simulated GMM model of wavelet coefficient. g_1, g_2, g_3 denotes the probability of GMM component with the weight w_1, w_2, w_3, respectively. $\sum_i w_i g_i$ is the weighted sum of component's probability. We can see that Gaussian mixture model can simulate the distribution of wavelet coefficients (the red line display it). Here we calculate the probability density function of GMM model with various embedding strength α .

IV. Statistical Algorithm 2

As to a random variable X , its probability density function is denoted by $p(x)$. If it is Gaussian random variable with mean μ and variance σ^2 , then its probability density is denoted by $N(\mu,\sigma^2)$. The characteristic function defined as follows:

$$\Phi(t) = \int_{-\infty}^{\infty} p(x)e^{jtx}dx \tag{9}$$

Here, $j=\sqrt{-1}$. Corresponding probability density can also use the following type:

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi(t)e^{-jtx}dt \tag{10}$$

As to a sequence of independent and identically distributed terms whose probability density function is unknown, more natural statistical description is the probability density (hereinafter referred to as PDF) moments. The -order moment formula for calculating the probability density is as follows:

$$\hat{m}_n = \frac{1}{N}\sum_{i=1}^{N} x_i^n, n \ge 1 \tag{11}$$

Where, N is the length of random variable X .Equation (11) is an estimation of the n^{th} PDF moment as follows:

$$m_n = EX^n = \int_{-\infty}^{\infty} p(x)x^n dx \tag{12}$$

Based on the probability density function, four statistical moments (mean, variance, and kurtosis) are used for statistical analysis. The wavelet coefficients of original clean signal and statistical signal have the symmetry distribution around the zero value, so often the absolute moment is calculated:

$$\hat{m}_n^A = \frac{1}{N}\sum_{i=1}^{N}|x_i|^n, n \ge 1 \tag{13}$$

Here is the absolute PDF moment of order unbiased estimation:

$$m_n^A = E|X|^n = \int_{-\infty}^{\infty} p(x)|x|^n dx \tag{14}$$

Let X , Y , Z denote cover signal, statistical signal and embedded noise respectively. So the general additive statistical can be modeled as:

$$Y = X + Z \tag{15}$$

Here, Z and X are mutually independent. The secrete key is shared by embedding side and extraction side. In the independent and identically distributed (I.I.D) model, the mutual independence will lead to two edges PDF convolution:

$$p_Y(y) = \int_{x\in X} p_X(x)p_Z(y-x)dx \tag{16}$$

The above discussion of the model concerns the generic situation, but it did not tell me if it is sensitive to the statistical and analysis operation. Even if we do not know the exact statistical model and which embedded algorithm is used, but as described above, statistical analysis personnel can capture the signal is prior to knowledge of statistical models and general statistical characteristics of statistical analysis algorithms. The following section will apply these characteristics to the additive spread-spectrum statistical analysis model and analysis the statistical moments of cover audio and statistical audio modeled by GMM and GGD.

Where, $w_{x,i}$, $\sigma_{x,i}$, $\mu_{x,i}$ is the weight, variance, mean of i^{th} CMM component of cover audio respectively. n denotes the n^{th} order moment. $\Gamma(\cdot)$ is the Gamma function and $F(\cdot)$ is the confluent hyper-geometric function. After spread-spectrum hiding, the absolute PDF moment of the hidden signal Y is as follows:

$$\begin{aligned} m_{n,Y}^A &= \sum_{i=1}^{k} w_{y,i} E_{y,i}\left[|y|^n\right] \\ &= \sum_{i=1}^{k} w_{y,i}\sigma_{y,i}^n \cdot 2^{\frac{n}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}} F\left(-\frac{1}{2}n,\frac{1}{2},-\frac{1}{2}\left(\mu_{y,i}/\sigma_{y,i}\right)^2\right) \end{aligned} \tag{17}$$

In order to further examine difference between PDF moments caused by embedding, the ratio of the absolute PDF moment between cover audio X and statistical audio Y is calculated by the above formula (16) (17), the ratio r can be expressed as:

$$r = \frac{m_{n,Y}^A}{m_{n,X}^A} = \frac{\sum_{i=1}^{k} w_{y,i}\sigma_{y,i}^n F\left(-\frac{n}{2},\frac{1}{2},-\frac{1}{2}\left(\mu_{y,i}/\sigma_{y,i}\right)^2\right)}{\sum_{i=1}^{k} w_{x,i}\sigma_{x,i}^n F\left(-\frac{n}{2}n,\frac{1}{2},-\frac{1}{2}\left(\mu_{x,i}/\sigma_{x,i}\right)^2\right)} \tag{18}$$

From the equation (18), it can be seen that the intensity ratio changes with the moment order n as well as the quantity of Gaussian mixture model.

Here, in order to examine the impact of Gaussian mixture components, we set the fixed order $n = 3$, and then test four sets of experiments with various embedding strength α respectively. The results are shown in Fig.4

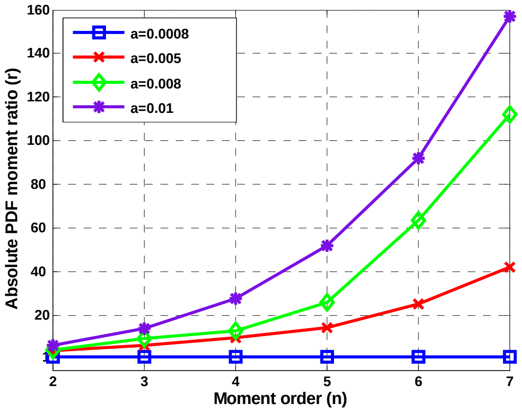

Figure 4. Absolute PDF moment and the GMM components number

Generalized Gaussian distribution has a wide range of applications in many fields. In the fields of image processing; it was used to describe DCT and wavelet coefficients. The GGD is defined as:

$$p_{\alpha,\beta}(s) \triangleq \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}\exp\left\{-\left(\frac{|s|}{\alpha}\right)^{\beta}\right\} = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}\exp\left\{-\left|\frac{x-\mu}{\alpha}\right|^{\beta}\right\}, \alpha>0, \beta>0, s\in\mathbb{R} \quad (19)$$

Where,

$$\alpha = \sqrt{\frac{\sigma^2\Gamma(1/\beta)}{\Gamma(3/\beta)}} = \sigma\sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}, \sigma > 0 \quad (20)$$

If the cover audio X has the PDF (equation (19)) then the standardized variable $V = (X-\mu)/\sigma$ will have the PDF:

$$f(v) = \frac{\beta\exp\left\{-|v|^{\beta}\right\}}{2\Gamma(1/\beta)} \quad (21)$$

It is easily seen that the k^{th} moment of variable V is given by

$$\mathrm{E}(V^k) = \frac{1+(-1)^k}{2\Gamma(1/\beta)}\Gamma\left(\frac{k+1}{\beta}\right) \quad (22)$$

Thus, the n^{th} moment of cover audio X can be obtained as:

$$\begin{aligned}\mathrm{E}(X^n) &= \mathrm{E}\left[(\mu_x+\sigma_x V)^n\right] \\ &= \sum_{k=0}^{n}\binom{n}{k}(\mu_x)^{n-k}(\sigma_x)^k\,\mathrm{E}(V^k) \\ &= \frac{(\mu_x)^n\sum_{k=0}^{n}\binom{n}{k}(\sigma_x/\mu_x)^k\left\{1+(-1)^k\right\}\Gamma\left((k+1)/\beta_x\right)}{2\Gamma(1/\beta_x)}\end{aligned} \quad (23)$$

Where, μ_x, σ_x is the mean, variance of cover audio respectively. β_x is shape parameters of GGD model. $\Gamma(\cdot)$ is the Gamma function

Then we use the GGD to model statistical audio Y, the n^{th} moment is:

$$\mathrm{E}(Y^n) = \frac{(\mu_y)^n\sum_{k=0}^{n}\binom{n}{k}(\sigma_y/\mu_y)^k\left\{1+(-1)^k\right\}\Gamma\left((k+1)/\beta_y\right)}{2\Gamma(1/\beta_y)} \quad (24)$$

Where, μ_y, σ_y is the mean, variance of statistical audio respectively. β_y is shape parameters of GGD model. $\Gamma(\cdot)$ is the Gamma function, so, when we use GGD to model wavelet sub-band, based on equation (23) and (24), the ratio r_{GGD} of PDF moment can be calculated as follows:

$$\begin{aligned}r_{GGD} &= \frac{\mathrm{E}(Y^n)}{\mathrm{E}(X^n)} \\ &= \frac{(\mu_y)^n\sum_{k=0}^{n}\binom{n}{k}(\sigma_y/\mu_y)^k\left\{1+(-1)^k\right\}\Gamma\left((k+1)/\beta\right)}{(\mu_x)^n\sum_{k=0}^{n}\binom{n}{k}(\sigma_x/\mu_x)^k\left\{1+(-1)^k\right\}\Gamma\left((k+1)/\beta\right)}\times\frac{\Gamma(1/\beta_x)}{\Gamma(1/\beta_y)}\end{aligned} \quad (25)$$

In the equation (25), the r_{GGD} may be negative due to the polarity of μ_x, μ_y and n. In order to make fair comparison, we calculate the n^{th} PDF moment of GMM model, not the absolute moment. The ratio r_{GMM} of GMM moment is calculated as follows:

$$r_{GMM} = \frac{\sum_{i=1}^{k} w_{y,i}\sigma_{y,i}^{n}\left(-i\sqrt{2}\,\mathrm{sgn}\,\mu_{y,i}\right)^n U\left(-\frac{n}{2},\frac{1}{2},-\frac{1}{2}\left(\frac{\mu_{y,i}}{\sigma_{y,i}}\right)^2\right)}{\sum_{i=1}^{k} w_{x,i}\sigma_{x,i}^{n}\left(-i\sqrt{2}\,\mathrm{sgn}\,\mu_{x,i}\right)^n U\left(-\frac{n}{2},\frac{1}{2},-\frac{1}{2}\left(\frac{\mu_{x,i}}{\sigma_{x,i}}\right)^2\right)} \quad (26)$$

V. SUMMARY

This study covers the basic semantics of video content search link in the lower part of perceived characteristics of treatment, the use of the semantic model of color naming, semantic digital color space and color space conversion; in the semantic concept generates link of video semantic concept of cloud association rule-based reasoning, the automatic extraction of semantic concepts; in the semantic retrieval links, by latent semantic analysis to eliminate the interference of semantic relevance. Spread-spectrum algorithm is common in the area of audio information hiding algorithm. In this paper, we propose two statistical analysis algorithms. Both the method is based on machine learning theory. In the proposed algorithm I, we use GMM to model the probability distribution of wavelet coefficient and calculate the absolute moment of statistical distribution as feature vector of each sub-band for statistical analysis. In the proposed algorithm II, the distortion metric based on probability distribution of wavelet sub-band is designed to detect modifications in audio signal. The probability distribution model takes GMM and GGD. We use de-noising method to get the estimation of cover audio, and then use four distances metric to measure the distortion. Results from simulation experiments show that the two proposed algorithms provide significantly higher detection rates than existing schemes. On the whole, proposed algorithm II has better performance than algorithm I. The experiment results on FHSS and Audio Wet systems show that our two algorithms have better extensibility to the other statistical analysis methods. In view of the block embedding, we randomly choose several frames from the whole audio piece to extract features; accordingly the algorithm performance is better to the full hiding situation. Future research may focus on feature dimension reduction methods, reducing the complexity and improving the detecting results with various embedding ratio. As to the feature optimization, there is some room for improvement.

REFERENCES

[1] Peter Wilkins, Alan F. Smeaton, Noel E. O'Connor et al., "K-Space Interactive Search," Proceedings of the 2008 ACM International Conference on Content-Based Image and Video Retrieval, Niagara Falls, Canada, 2008, pp. 555-556.

[2] [2] Ork de Rooij, Cees G.M. Snoek,Marcel Worring, "Mediamill: Fast and Effective Video Search Using the Forkbrowser," Proceedings of the 2008 international conference on Content-based image and video retrieval, Niagara Falls, Canada, 2008, pp. 561-562

[3] Shi-Yong Neo, Huanbo Luan, Yantao Zheng et al., "Visiongo: Bridging Users and Multimedia Video Retrieval," Proceedings of the 2008 international conference on Content-based image and video retrieval, Niagara Falls, Canada, 2008, pp. 559-560

[4] [4] Huang J.,S. R. Kumar, "Image Indexing Using Color Correlogram," IEEE Conference of ter Vision Pattern Recognition (CVPR 97), San.Juan, Puerto Rico, 1997, pp. 762-768

[5] Stricken M.,M. Orengo, Similarity of Color, SPIE Storage and Retrieval for Image and Viedeo Databases III, vol. 2185, pp. 381-392, 1995

[6] Ma, W.-Y,B. S. Manjunath, A Comparison of Wavelet Transform Features for Texture Image Annotation, Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 256-259, 1995

[7] Liu W. Y, Z. Su,et al., "A Performance Evaluation Protocol for Content-Based Image Retrieval Algorithms/Systems.," IEEE CVPR Workshop on Empirical Evaluation in Computer Vision, Kauai, Hawaii, USA, 2001

[8] M. Naphade, J. R. Smith, J. Tesic et al., Large-Scale Concept Ontology for Multimedia, Multimedia, IEEE, vol. 13, no. 3, pp. 86-91, 2006

[9] H. Farid (2002) Detecting hidden messages using higher-order statistical models. In Proceedings of IEEE International Conference on Image Processing, pp. 905-908

[10] Harmsen J J, Pearlman W A (2003) Steganalysis of additive noise modelable information hiding. SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents. vol.5022, 21-24

[11] Seshadrinathan K, Sheikh H R, Bovik A C (2005) Detecting spread spectrum watermarks using natural scene statistics. In Proceedings of IEEE International Conference on Image Processing, vol. 2, 1106-1109

[12] Kenneth Sullivan, Upamanyu Madow, Shivkumar Chandrasekaran et al (2005) Steganalysis of spread spectrum data hiding exploiting cover memory. In Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents V, pp.38-46

[13] Ying Wang, Moulin P (2003) Steganalysis of block-DCT image steganography. In IEEE workshop on Statistical Signal Process, pp.339-342

[14] Y. Shi, C. Chen, W. Chen (2007) A markov process based approach to effective attacking JPEG steganography. In Lecture Notes in Computer Science, vol. 437, pp. 249-264

[15] Q. Liu, A. Sung, B. Ribeiro, R. Ferreira (2008) Steganalysis of multiclass JPEG images based on expanded Markov features and polynomial fitting. In Proceeding of IEEE International Joint Conference Neural Networks, pp.3351-3356

[16] Zugen Liu, Xuezeng Pan, Lie Shi, Jimin Wang, Lingdi Ping (2006) Effective steganalysis based on statistical moments of differential characteristic Function. In International Conference on Computational Intelligence and Security, vol. 2, 1195-1198

[17] J. Fridrich (2004) Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In Lecture Notes in Computer Science, vol. 3200, pp.67-81

[18] Shi, Y.Q., Xuan, G., Yang, C., et al (2005) Effective steganalysis based on statistical moments of wavelet characteristic function. In: IEEE International Conference on Information Technology: Coding and Computing, pp.768-773

[19] S. Lyu, H. Farid (2006) Steganalysis using high-order image statistics. IEEE Transaction on Informaion Forensics Security 1(1):111-119

[20] Avcibas, I., Memon, N., Sankur, B (2003) Steganalysis using image quality metrics. IEEE Transactions on Image Processing 12(2):221-229

[21] X. Ru, Y. Zhang, F. Wu (2006) Audio steganalysis based on negative resonance phenomenon caused by steganographic tools. Journal of Zhejiang Univ. SCIENCE A 7(4):577-583

[22] Hamza Ozer, Ismail Avcibas, Bulent Sankur, et al (2003) Steganalysis of audio based on audio quality metrics. In Proceedings of SPIE on Security and Watermarking of Multimedia Contents, pp.55-66

[23] Avcibas (2006) Audio steganalysis with content-independent distortion measures. IEEE Signal Process Letter 13(2):92-95.

[24] Altun O, Sharma G, Celik M, Bocko M (2005) Morphological steganalysis of audio signals and the principle of diminishing marginal distortions. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.21-24

[25] M. Johnson, S. Lyu, H. Farid (2005) Steganalysis of recorded speech. In Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents VII, vol. 5681, 664-672

[26] C. Kraetzer, J. Dittmann (2007) Mel-cepstrum based steganalysis for voip-steganography. In Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents IX, vol. 6505, 650505.1-650505.12

[27] C. Kraetzer, J. Dittmann (2008) Pros and cons of mel-cepstrum based audio steganalysis using SVM classification. In Lecture Notes in Computer Science, vol. 4567, 359-377

[28] Krätzer, Christian; Dittmann, Jana (2009) The impact of information fusion in steganalysis on the example of audio steganalysis. Media Forensics and Security, vol. 7524, 725409-725409-12

[29] Q. Liu, A. Sung, M. Qiao (2008) Detecting information-hiding in WAV audios. In Proceeding of 19th International Conference on Pattern Recognition, pp 1-4
[30] Y.Liu, K.Chiang, C.Corbett, R.Archibald, et al (2008) A novel audio steganalysis based on high-order Statistics of a distortion measure with hausdorff distance. 11th Information Security Conference, Lecture Notes in Computer Science, vol.5222, 487-501
[31] Qingzhong Liu, Andrew H. Sung, Mengyu Qiao (2009) Temporal derivative-based spectrum and mel-cepstrum audio steganalysis. IEEE Transactions on Information Forensics and Security 4(3):359-368
[32] W Bender, D Gruh, N Morimoto,et al (1996) Techniques for data hiding. IBM System, vol.35, no.3&4, pp.313-336
[33] J.D.Gordy, L.T.Bruton (2000) Performance evaluation of digital audio watermarking algorithms. In Proceedings of the 43rd IEEE Midwest Symposium, Vol.1, pp.456-459
[34] Nedeljko Cvejic (2004) Algorithms for audio watermarking and steganography. Dissertation. University of OULU
[35] Ying Wang, Pierre Moulin (2008) Perfectly secure steganography: capacity, error exponents, and code constructions. IEEE Transactions on Information Theory, 54(6):2706-2722
[36] G. Xuan, Y. Q. Shi, J. Gao, et al (2005) Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In Proceeding of Information Hiding Workshop, pp.262-277
[37] M. Goljan, J. Fridrich, T. Holotyak (2006) New blind steganalysis and its implications. In Proceedings of the SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI, pp.1-13

Jiexin Zhang was born in August 1976. He received the B.S. degrees from Guangxi University of Technology in 2001, and the M.S. degrees from Guilin University of Technology in 2010. Now he is a lecturer in Department of Computer Science, GuangXi Economic Management Cadre College, China. His research interests include algorithm, optimization and data mining.