Robust multimedia watermarking: Hidden Markov model approach for video sequences

Ersin ELBAŞI
The Scientific and Technological Research Council of Turkey
Tunus Caddesi No: 80 Kavaklıdere, Ankara-TURKEY
e-mail: ersin.elbasi@tubitak.gov.tr

Abstract

Because of the large amount of frames, similarity between frames and temporal attacks (frame dropping, frame averaging, frame swapping etc.), the process for video watermarking is more difficult than that for image watermarking. Current image watermarking methods are limited to solve all these difficulties. We propose a novel video watermarking system based on the Hidden Markov Model (HMM). This novel watermarking scheme splits the video sequences into a Group of Pictures (GOP) with HMM. Portions of the binary watermark are embedded into each GOP with a wavelet domain watermarking algorithm. The embedding process is the standard additive algorithm [1] in low pass (LL) and high pass (HH) bands in the wavelet domain. This novel system increases the robustness against geometric and temporal attacks, and increases the quality of the watermarked video.

Key Words: *Watermarking, Hidden Markov Model, Wavelet domain, Group of pictures.*

1. Introduction

Digital watermarking has received increasing attention in recent years. Distribution of movies, music, and images is now faster and easier via computer technology, especially on the Internet. Hence, content owners (e.g., movie studios and recording companies) are concerned about illegal copying of their content. Watermarking and cryptography are two standard multimedia security methods. However, cryptography is not an effective method because it does not provide permanent protection for the multimedia content after delivery to consumers [2].

The most important properties of a watermarking system are robustness, invisibility, data capacity, and security. An embedded watermark should not introduce a significant degree of distortion in the cover multimedia element. Robustness is the resistance of the watermark against normal Audio/Video (A/V) processes or intentional attacks. Data capacity refers to the amount of data that can be embedded without affecting perceptual transparency. The security of a watermark can be defined to be the ability to thwart hostile attacks such as unauthorized removal, unauthorized embedding, and unauthorized detection. There are basically two
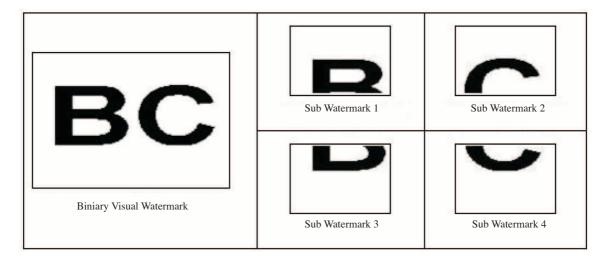
Figure 1. Watermark decomposition.

approaches to embed a watermark: spatial domain and transform domain (e.g., Discrete Cosine Transform, Discrete Fourier Transform, or Discrete Wavelet Transform) [3, 4]. In the spatial domain, the watermark is embedded by modifying the pixel values in the original cover image. Transform domain watermarking is similar to spatial domain watermarking; in this case, the transform coefficients are modified [2].

Most of the current video watermarking techniques are based on image watermarking methods [5], and directly applied to uncompressed or compressed video sequences [6]. However, these methods are not sufficient for copyright protection in video data. Video watermarking has a number of issues, and image based algorithms do not solve these problems. Embedding large amounts of data, redundancy between frames, and robustness against temporal attacks (e.g., frame averaging, frame dropping, and frame swapping) are some of the main problems in video applications [7]. Neither embedding the same watermark in each frame, nor embedding different watermarks in every frame of the video would be robust against all types of common attacks.

An important criterion for classifying watermark schemes is the type of information needed by the detector.

- Non-Blind Schemes [8]: Both the original image and the secret key(s) are needed in detection.

- Semi-Blind Schemes [9]: The secret key(s) and the watermark bit sequence are needed.

- Blind Schemes [10, 11]: Only the secret key(s).

To provide the necessary properties (robustness, invisibility, data capacity, and security), we propose a new watermarking method in video sequences. The proposed algorithm decomposes the binary visible watermark into m sub-images, and embeds each into GOPs. The best time durations for GOPs are calculated using Hidden Markov Models (HMM) [12]. This method provides robustness against common geometric and temporal attacks.

In this paper, we discuss how to obtain probabilities and select a group of pictures by Hidden Markov Models, and embedding and detection algorithms. Experimental results and conclusions are given at the end of the paper.

2. Preprocess

In this proposed system, video and watermark are two inputs, and they are both divided into portions. The scheme embeds different parts of a single watermark into different GOPs of the video in the wavelet domain. The best selection between watermark portion and GOPs and time duration for each GOP will be computed by the HMM algorithm. This algorithm helps us to make a decision of the beginning and ending positions of GOPs. The novel scheme has three main parts: preprocessing, GOP selection and embedding/extraction algorithms [2].

2.1. Watermark preprocess

In the system, we use a *NxM* binary watermark. Our purpose is to embed different watermarks into different GOPs. For this purpose, we divide the current binary watermark into m equal parts. This is a simple process and each portion of the watermark will be embedded into low and high frequencies of each frame in the video sequence.

2.2. Probability calculation

HMM-based watermarking works using probabilistic distribution in the video sequence. There are two methods to calculate probabilities: first, the Naive Bayes Classifier after feature extraction, second, the overall performance calculation for each frame, which indicates accuracy in watermarking based on the robustness and quality of the watermarked frame. The Naive Bayes Classifier predicts class membership probabilities, such as the probability that a given image belongs to class "*Watermark Present*" or "*Watermark Absent*". We use some statistical attributes, such as number of selected coefficients, mean, variance, range of the coefficients, and so forth as a feature set after transforming frames to DWT [13, 14].

In the second method, we calculate the probability of success for each frame, which shows the accuracy of the current embedding algorithm in that frame. There are several performance measurements in watermarking research. The requirements for invisibility, robustness and data capacity are as follows. Peak Signal-to-Noise Ratio (PSNR) is only for invisibility measurement, and Similarity Ratio (SR) should be used for robustness measurement. There is no measurement defined for all three requirements. In this work, we define a measurement for overall performance (OP).

$$OP = a_1 \times P_1 + ... + a_k \times P_k \tag{1}$$

The OP is the overall performance, a_k is the distribution probability of robustness and quality measurements (PSNR,M-SVD,SR etc.) and P_k is the probability calculated using these measurements. The a_k calculation is based on number of different measurements: half of the distribution is for robustness of embedded frame(PSNR and M-SVD), and the other half is for quality of extracted watermark quality (SR). The ratio between the calculated measurement value and the best expected value gives the probability P_k for each performance measurement. Each P indicates different measurements, such as success in invisibility, data capacity, cropping attack, rotation attack, and other attacks. Depending on the application, we can increase the number of the Ps.

$$\sum_{i=1}^{k} a_i = 1 \tag{2}$$

PSNR, M-SVD and SR help us to calculate the OP value.

The Peak-Signal-to-Noise Ratio (PSNR): The PSNR is most commonly used as a measure of the quality of reconstruction in image watermarking. It is the ratio between the maximum value of a signal and the magnitude of background noise.

Measure of Singular Value Decomposition (M-SVD): M-SVD is a newly developed measure, which expresses the quality of watermarked images. It is based on the Singular Value Decomposition (SVD). M-SVD is a bivariable measure that computes the distance between the singular values of the watermarked image block.

Similarity Ratio (SR): Defined by SR = S/(S+D), where S denotes the number of matching pixel values in compared images, and D denotes the number of different pixel values in compared images [8].

Overall Performance uses the average of these three measurements to produce probabilities in the video watermarking scheme.

Figure 2 shows the OP in a video sequence when we embed different sub-watermarks. Overall performance means probability of detection based on the different measurements mentioned above. For example, the overall performance is much better when we embed watermark 2 into frame 1 to 80 instead of embedding watermark 1 in the same frame sequence.

3. Group of picture selection

There are two optimization problems in video preprocessing:

1. Optimal matching between watermark portions and group of pictures.

2. Optimal time duration decision for each GOPs.

The Hidden Markov Model (HMM) is a finite set of states, each of which is associated with a probability distribution [12]. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state, an outcome or observation can be generated according to the associated probability distribution. HMMs are widely used, especially in speech recognition and other time-based systems.
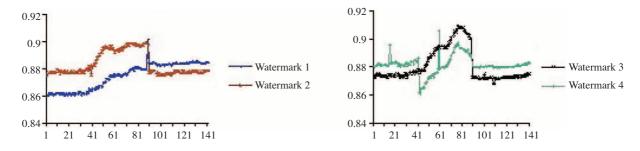

Figure 2. Overall performance for each frame.

A Hidden Markov Model is a statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. The extracted model parameters can then be used to perform further analysis. A HMM can be considered to be the simplest dynamic Bayesian network.

In a regular Markov model, the state is directly visible to the observer, and, therefore, the state transition probabilities are the only parameters. In a Hidden Markov Model, the state is not directly visible, but variables influenced by the state are visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states.

The basic process:

1. Environment is in some state.

2. As a function of the state, an observation is generated randomly.

3. As a function of the state, a new state is generated randomly.

Expected output:

1. Probability of a sequence.

2. Find HMM to maximize total probability of a set of sequences.

3. Find the most likely state sequence.

4. Probability of a part of the sequence.

The features extracted for each frame are the HMM input values in the proposed watermarking process: mean, variance, range, and number of selected coefficients.

A preprocessed set of watermarks is prepared in watermark preprocessing,

$$W = W_1, W_2, ..., W_N \tag{3}$$

where W(i) is the portion of the single visual binary watermark. The extracted feature vectors of each frame is another input for the HMM algorithm,

$$F = F_{(1,t)} = F_1, ..., F_t \tag{4}$$

where F_i is the feature vector for frame i, and t is the number of the frame in video sequences. The optimal criterion is to find the sequence of frames most likely to produce optimal watermarking with the watermark portion. Thus, we need to find the maximum value.

$P(W|F)$ is the probability of the complete automation state sequence of W, embedded with the most likely state transition timing, given the sequence of observations $F = F_{(1,t)}$.

$$P(W \mid F) = max_{t_1,t_2,..t_N} P(W_{1_{t_1,t_2-1}}, W_{2_{t_2,t_3-1}}, ...W_{N_{t_N,t}} \mid F) \tag{5}$$

$$P(W \mid F) = max_{\forall t_1,t_2,..t_N} \prod_{i=1}^{N} \frac{a_{i,i-1} P(W_{i_{t_i,t_{i+1}}} \mid F_{i_{t_i,t_{i+1}}})}{W_{i_{t_i,t_{i+1}}}} \tag{6}$$

Our main goal is to describe an algorithm to find the values of $t_1, t_2, ... t_N$ that maximize $P(W \mid F)$.

The direct computation of $P(W \mid F)$ at time T involves operations of $O(T^N)$ complexity, which is very expensive. An adaptation of the Viterbi algorithm will be used in the HMM to decrease the computational cost. R_i is the likelihood, where GOP_i at time t is the most likely transition timing between GOPs, given the sets of features $F_{(1,t)}$.

$$A_j = P(W \mid F) = max_{\forall t_1, t_2, .. t_N} \prod_{i=1}^{N} \frac{a_{i,i-1} P(W_{i_{t_i, t_{i+1}}} \mid F_{i_{t_i, t_{i+1}}})}{W_{i_{t_i, t_{i+1}}}} \tag{7}$$

$$R_i(t) = max_{t_{i-1} \le t_i \le t} A_i R_{i-1}(t_i - 1) \tag{8}$$

$$t_{i_{BEST}} = max_{t_{i-1} \le t_i \le t} \frac{a_{i,i-1} P(W_{i_{t_i, t}} \mid F_{i_{t_i, t}})}{W_{i_{t_i, t}}} \times R_{i-1}(t_i - 1) \tag{9}$$

The output is the sequence of GOPs occurring with optimum transition timing $t_{1_{BEST}}, t_{2_{BEST}}, ... t_{N_{BEST}}$. The output of the HMM algorithm is the beginning and ending time of each GOP, and optimal matching between GOPs and subwatermarks.

4. Embedding and extraction

For each frame in GOP, the following embedding procedure is applied. We use HMMs for two reasons in this work: 1) optimal matching between watermark portions and GOPs, and 2) optimal time duration for each GOP. After applying the HMM, we use an additive algorithm for sub-watermark embedding in GOPs. The binary watermark embedding procedure is given as follow [2, 8]:

Watermark Embedding Algorithm:

1. Convert the *NxM* RGB frame to YUV.

2. Compute the DWT of the luminance layer (Y) for each frame.

3. Modify the coefficients V_{ij} in the LL and HH bands in all frames.

4. $V_{w,ij} = V_{i,j}^k + a_k \times W_{i,j}$, where i = 1,...,n ; j = 1,...,m ; k=1,2.

5. Apply the inverse transformation to obtain the watermark cover frame I_w for each frame.

Watermark Detection Algorithm: There are some pieces of information stored during the embedding process, such as the duration of the group of frames. Based on this information, we apply the following procedure to extract the watermark [2, 8].

1. Split video sequence to group of frames.

2. Apply the following algorithm in frames.

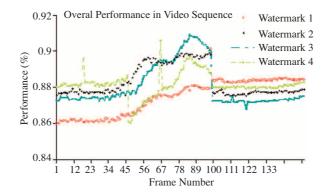   - Convert *NxM* RGB frames to YUV.

Figure 3. Probabilities in tennis video sequence.

- Compute the DWT of the luminance layer (Y) for each frame.

- Extract the binary visual watermark from the LL and HH bands.

- $W_{ij}^{*} = (V_{w,ij}^{*k} - V_{ij}^{k}) / a_k$, where i= 1,...,n; j = 1,...,m.

- If W_{ij}^{*} bigger than T, then $W_{ij}^{*} = 1$, else $W_{ij}^{*} = 0$, where T is the threshold between 0 and 1.

3. Combine all watermark portions, and output the watermark.

5. Experimental results

The DWT separates the image into a lower resolution image (LL), and horizontal (HL), vertical (LH) and diagonal (HH) detail components. High resolution sub-bands are used to locate edge and texture patterns in an image. The DWT is also computationally efficient and implemented using simple filter convolution. The magnitudes of DWT coefficients are larger in the lowest bands (LL) at each level of decomposition. Embedding the watermark in the higher level sub-bands increases the robustness of the watermark. However, the image visual fidelity may be lost, which can be measured by PSNR. With the DWT, the edges and texture can be easily identified in the high frequency bands like HH, LH, and HL. The large coefficients in these bands normally indicate edges in the image.

We validated the proposed algorithm using the tennis video sequence. The video sequence is about 20-40 seconds, and the frame size is 240x352.

Figure 3 shows the overall performance probability for each sub-watermark. In general, embedding different portions of the visual watermark gives better performance.

Figure 4 shows the extracted portions of the watermark. Experimental results show that SR values are higher in the LL and HH bands than the HL and LH bands. SR values show that the proposed algorithm increases performance. Figure 5 shows the composed sub-watermarks.

A watermark should be robust against attacks. We can classify attacks in several ways. Direct or indirect attacks:

1. Direct attacks attempt to remove, obscure, or render the watermarks undetectable in the content.
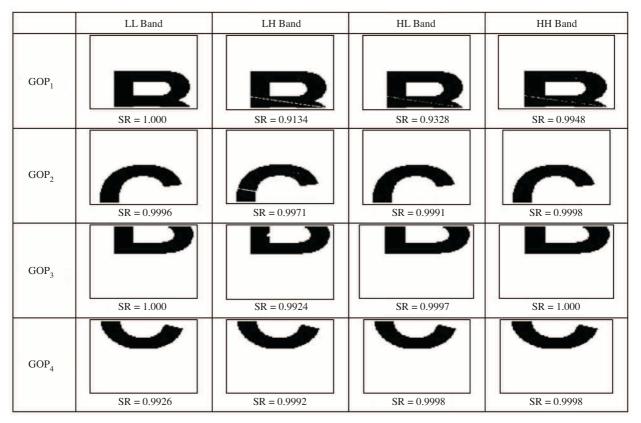
	LL Band	LH Band	HL Band	HH Band
GOP$_1$	SR = 1.000	SR = 0.9134	SR = 0.9328	SR = 0.9948
GOP$_2$	SR = 0.9996	SR = 0.9971	SR = 0.9991	SR = 0.9998
GOP$_3$	SR = 1.000	SR = 0.9924	SR = 0.9997	SR = 1.000
GOP$_4$	SR = 0.9926	SR = 0.9992	SR = 0.9998	SR = 0.9998

Figure 4. Extraction Results in GOP$_1$, GOP$_2$, GOP$_3$ and GOP$_4$.

SR = 0.9998

Figure 5. Composed watermark

2. Indirect attacks leave the watermark undamaged, but seek to undermine the validity of the scheme that uses the watermark as its basis.

Another attack classification scheme is based on the attack type. Geometric attacks and statistical attacks:

1. Common signal processing: The watermark should be detected after signal processing attacks, such as digital-to-analog, analog-to-digital conversion, resampling, Gaussian noise, histogram equalization, etc.

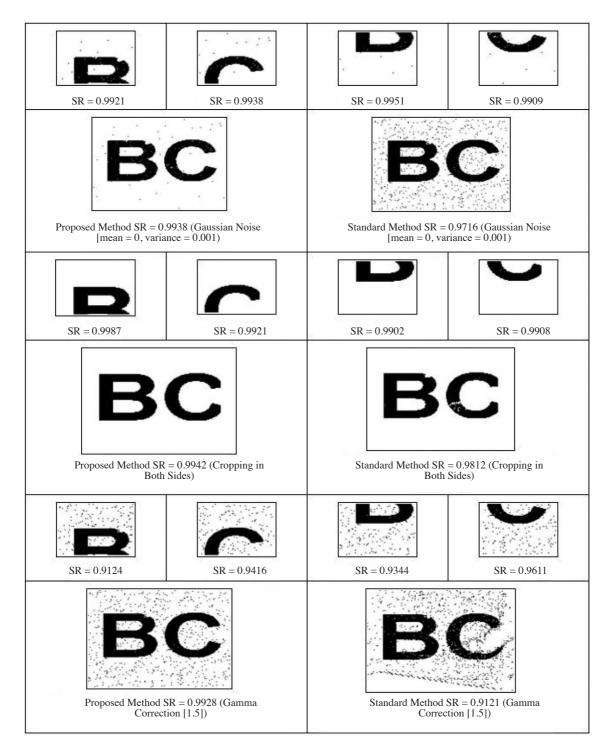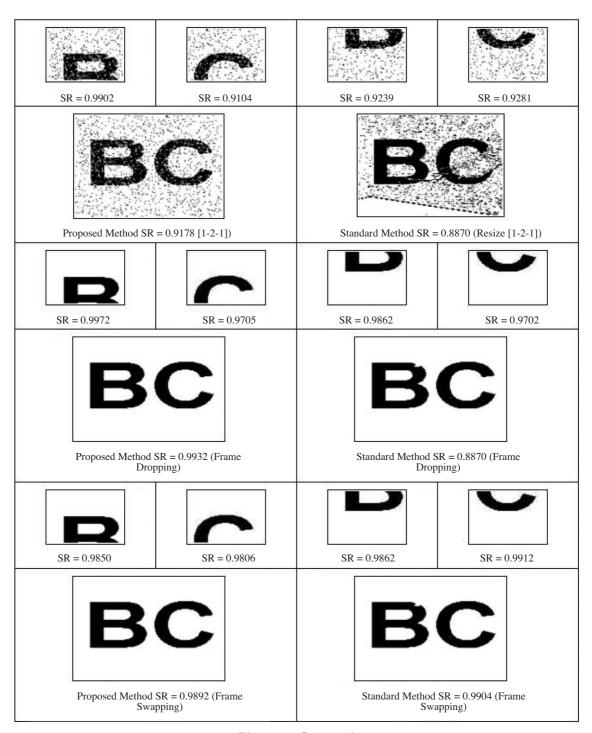2. Common geometric distortions: Rotation, resizing, cropping and scaling are the most common attacks in this class.

Figure 6. Extraction results after some common attacks.

Figure 6. Continued.

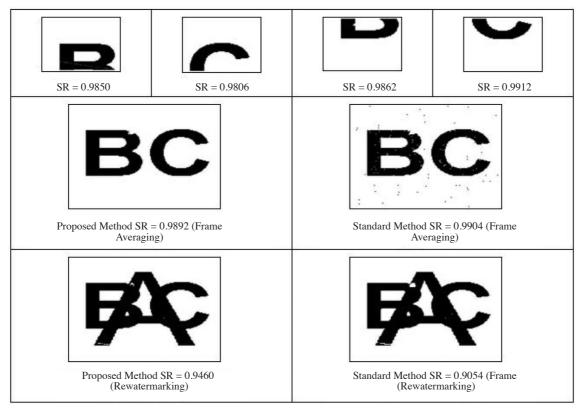|			
SR = 0.9850	SR = 0.9806	SR = 0.9862	SR = 0.9912
Proposed Method SR = 0.9892 (Frame Averaging)		Standard Method SR = 0.9904 (Frame Averaging)	
Proposed Method SR = 0.9460 (Rewatermarking)		Standard Method SR = 0.9054 (Frame Rewatermarking)	

Figure 6. Continued.

Figure 6 shows the extracted watermark portions and the composed watermark after common attacks. There are two frames for each attack in this figure; the first shows the extracted watermark based on the proposed algorithm, and the second shows the extracted watermark based on a standard additive algorithm [1]. SR values show that the proposed algorithm is more robust than the standard additive algorithm.

6. Conclusion

The most important properties of a video watermarking system are its robustness, invisibility, data capacity, and security. There is a significant amount of research in video base watermarking, but the problem is still an open research area because of a number of challenging problems: embedding large amount of data, redundancy between frames, and robustness against temporal attacks (e.g., frame averaging, frame dropping, and frame swapping). To solve these problems, we proposed a novel HMM based algorithm. Our scheme embeds different parts of a single binary watermark into different Group of Pictures of a video sequence in the wavelet transformation domain. The proposed Hidden Markov Model obtains optimal matching and time durations in video sequences. Basically, an iterative HMM-based Viterbi algorithm was applied to decrease complexity. In the detection process, some information from the embedding process is used, such as the duration of GOPs. Experimental results show that this new algorithm has higher visual video fidelity and robustness against common statical and temporal attacks.

Future direction is as follows:

- Extension of the proposed algorithm in MPEG and DVD multimedia elements.
- Developing new algorithms which are more robust against a large group of attacks.
- Extending the current applications to image and audio together in video sequences.
- Real-time embedding for video and audio.

References

[1] R. Schyndel, A. Tirkel, and C. Osborne. "A digital watermark." *IEEE Proceedings International Conference of Image Processing*, Vol. 2, pp. 86-90, 1994.

[2] E. Elbasi, A. M. Eskicioglu. "Robust Video Watermarking Scheme in Transform Domains." *Second Information Security and Cryptology Conference (ISC Turkey)*, Dec. 2007.

[3] R. Dugad, K. Ratakonda, N. Ahuja. "A New Wavelet-Based Scheme for Watermarking Images." *Proceedings of 1998 International Conference on Image Processing (ICIP '98)*, Vol. 2, pp. 419-423, Chicago, IL, Oct. 4-7, 1998.

[4] C. Hsu, J. Wu. "DCT-Based Watermarking for Video." *IEEE Transactions on Consumer Electronics*, Vol. 44, No. 1, pp. 206-216, Feb. 1998.

[5] G. Doerr, J. Dugelay. "A Guide Tour of Video Watermarking." *Signal Processing: Image Communication*, Vol. 18, No.4, pp. 263-282, Apr. 2003.

[6] F. Hartung, B. Girod. "Digital Watermarking of Raw and Compressed Video." *Proc. European EOS/SPIE Symposium on Advanced Imaging and Network Technologies*, Berlin, Germany, Oct. 1996.

[7] P. Chan, M. R. Lyu, R. T. Chin. "Copyright Protection on the Web: A Hybrid Digital Video Watermarking Scheme." *Proceedings of the 13th International World Wide Web Conference (WWW '04)*, pp.354-355, New York, May 17-22, 2004.

[8] P. Tao, A. M. Eskicioglu. "A Robust Multiple Watermarking Scheme in the DWT Domain." *Optics East 2004 Symposium, Internet Multimedia Management Systems V Conference*, pp. 133-144, Philadelphia, PA, Oct. 25-28, 2004.

[9] E. Elbasi, A. M. Eskicioglu. "MPEG-1 Video Semi-Blind Watermarking Algorithm in the DWT Domain." *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2006*, Las Vegas, NV, April 6-7, 2006.

[10] H. Wang, Z. Lu, J. Pan, S. Sun. "Robust Blind Video Watermarking with Adaptive Embedding Mechanism." *International Journal of Innovative Computing, Information and Control*, Vol. 1, No. 2, June 2005.

[11] P. Chan, M. R. Lyu. "Digital Video Watermarking with a Genetic Algorithm." *Proceedings International Conference on Digital Archives Technologies Technologies (ICDAT '05)*, pp. 139-153, Taipei, Taiwan, June 16-17, 2005.

[12] L. R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE*, Vol. 77, Iss. 2, pp. 257-286, Feb. 1989.

[13] L. Shaohui, Y. Hongxun, Wen G. "Neural Network Based Steganalysis in Still Images." *Multimedia and Expo, ICME'03*, Vol. 2, 2003.

[14] E. Elbasi, A. M. Eskicioglu. "Naive Bayes Classifier Based Watermark Detection in Wavelet Transform." *Int. Worksop on Content Representation, Classification and Security*, Istanbul, Sept, 11-13, 2006.