# A SEMI-FRAGILE OBJECT BASED VIDEO AUTHENTICATION SYSTEM

*Dajun He, Qibin Sun and Qi Tian*

Media Engineering, Labs for Information Technology
21 Heng Mui Keng Terrace, Singapore, 119613
Email: {djhe, qibin, tian}@lit.a-star.edu.sg

## ABSTRACT

This paper presents a semi-fragile object-based authentication solution for MPEG4 video. To protect the integrity of the video objects / sequences, a content-based watermark is embedded into each frame in the Discrete Fourier Transform (DFT) domain before the MPEG4 encoding. A set of Angular Radial Transformation (ART) coefficients are selected as the robust features of the video objects. Error Correction Coding (ECC) is employed for watermark generation and embedding. Using this methodology, our semi-fragile authentication solution can robustly tolerate some natural object manipulations and errors (e.g. translation, scaling, rotation, lossy compression, segmentation errors, etc) while securely preventing other malicious modifications. Experimental results further demonstrate that the proposed method can achieve a good trade-off among system robustness, security and complexity.

## 1. INTRODUCTION

The increasing sophistication of computers has made digital manipulation of video very easy to perform, but difficult to detect. How to prove the authenticity of captured video becomes a serious issue. With the object based MPEG4 standard accepted as the international standard in Multimedia, the problem is becoming more prominent because the video object (VO) can be easily accessed, modified or copied from one video sequence to another in MPEG4 compliant applications. Thus an object based video authentication method is needed to protect the video integrity. Furthermore, during video editing, the video object may be scaled, translated or rotated to meet some specific requirements. The edited object may also undergo re-compression for the purposes of storage or transmission. All these conditions require the object / video authentication method to allow some natural object / video manipulations while preventing other malicious modification.

Figure 1 illustrates the content-based watermarking system for video authentication. Firstly, the raw video is segmented into foreground objects and background video. The watermark is generated using the features extracted from both the foreground and the background. The watermark is then embedded into foreground objects so that a secure link between objects and the background is created. At the receiver site, integrity between the object and background can be verified by comparing two sets of codes: one is the extracted watermark from the object and the other is re-generated from both the received object and the background. If these two sets of codes are the same, we then claim that the video content is authentic.

To ensure that the authentication method is robust against video processing such as translation, scaling, rotation and slight segmentation error, the feature extracted from the video object should tolerate variations during the above-mentioned video processing. In the meantime, the watermarking algorithm should be able to protect the embedded watermark from being distorted by such kinds of video processing. The detailed explanation of the watermarking algorithm is discussed in another paper we have submitted [1]: The video object is first transformed into the DFT domain. Then, a series of DFT coefficient groups in the low middle frequency area are selected, and the coefficients in every selected group are divided into two sub-groups based on pre-specified patterns. Finally, the characteristic relationship between these two sub-groups is used to hide the watermark bit sequence. The proposed watermarking algorithm is robust to translation, scale, rotation, as well as segmentation errors.

In this paper we will focus on the watermark generation and authentication algorithm. The paper is organized as follows. Section 2 is an analytical description of object based feature extraction. The detailed watermark generation and authentication algorithm is explained in Section 3. Experimental results and conclusions are presented in Section 4 and 5 respectively.

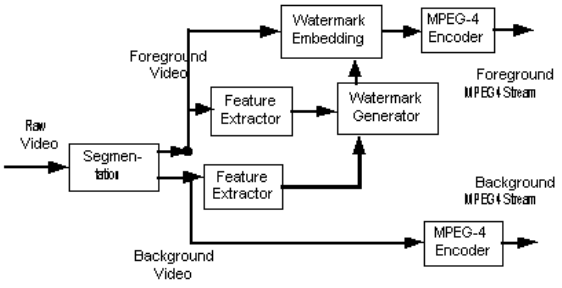



Figure 1 Proposed watermarking system for video authentication

## 2. OBJECT-BASED FEATURE EXTRACTION

In order to protect MPEG4 object / video integrity, the selected object-based features should meet the following requirements:

a. Robustness: Be invariant under normal video processing (translation, scaling and rotation). Other distortions from lossy compression and segmentation errors also need to be tolerated.

b. Security: Be sensitive to malicious modification. Any meaningful alteration to the objects should be rejected.

0-7803-7762-1/03/$17.00 ©2003 IEEE

Prior work used different features for different applications. In [2], block-based image intensity histogram was selected as the feature. Lin [3] used the relationship between the DCT coefficients at the same position in different blocks of an image as the features. However, these two kinds of features cannot meet our requirements as the object could be scaled, rotated or inaccurately segmented. Dittmann [4] and Queluz [5] used the edge / corner of the image as the feature to generate the digital signature. They claimed this feature is robust against high quality compression and scaling. But the disadvantage is that the signature generated based on the edge is too long, and the consistency of the edge itself is also a problem.

The main difference between the frame-based video application and the object-based video application lies in the utilization of shape information. The shape of the object plays an important role in the latter. The experiment has shown that human beings can recognize a characteristic object solely from its shape. That is why three kinds of visual shape descriptors are exploited to represent the image object in the MPEG7 [6][7]. In this paper, we employ Angular Radial Transformation (ART), a Region-Based Shape Descriptor, as the feature of the video object. The Region-Based Shape Descriptor has the following specific features that meet the requirement for our authentication application: Firstly, it gives a compact and efficient way to describe the video object. Secondly, the descriptor is also robust for segmentation errors. And finally, it is invariant to scaling, rotation, translation and various type of shape distortions.

Refer to [6][7], for a video shape, ART coefficients $F_{nm}$ can be extracted according to equation (1).

$$F_{nm} = \langle V_{nm}(\rho,\theta), f(\rho,\theta) \rangle = \int_0^{2\pi}\int_0^1 V_{nm}^*(\rho,\theta), f(\rho,\theta)\rho d\rho d\theta \quad (1)$$

Here, $F_{nm}$ is an ART coefficient of order $n$ and $m$, $V_{nm}(\rho,\theta)$ is the ART basis function, and $f(\rho,\theta)$ is a texture function of an arbitrary shape object instead of shape function in polar coordinates.

In the MPEG7 specification, 36 ART coefficients (n= 11, m= 2) are recommended for use. So, a region-based shape descriptor has 140 bits data. In our video authentication application, it is not necessary to exploit all these 36 coefficients to generate a watermark because of the following two reasons:

a. Among these 36 ART coefficients, the high-order coefficients, which are also considered as high order moments, represent the detailed information of the video object. These coefficients will change if the object is scaled or the object has slight segmentation error. To ensure that the watermark generated based on ART coefficients remains unchanged when the object is manipulated by acceptable manipulations, these high order ART coefficients should be cast away.

b. Equation (2) is the dissimilarity measure of the visual shape descriptor. From this equation, we can find that the coefficient with smaller magnitude has less contribution to image recognition. So, the ART coefficients with smaller magnitude can also be neglected:

$$\text{Dissimilarity} = \sum_i \left\| M_d[i] - M_q[i] \right\| \quad (2)$$

Where $d$ and $q$ represent different images and $M$ is the array of normalized magnitude of ART coefficients.

Based on the above considerations, we choose the ART coefficients in the following empirical way:

a. Given a training video sequence, calculate the all 36 ART coefficients of every frame.

b. For the same video sequence, calculate the coefficients of every frame under different scale factors that are from 0.5 to 0.9. Such scaling range is made for most transcoding applications where transmission bit-rate is a crucial issue. Figure 2 shows ART coefficients of first frame of "Akiyo" video sequence under different scaling factors.

c. Select the coefficients with large magnitudes but small magnitude fluctuation under different scaling factors.

We have tested several video sequences, and selected 15 ART coefficients from a total of 36 coefficients.

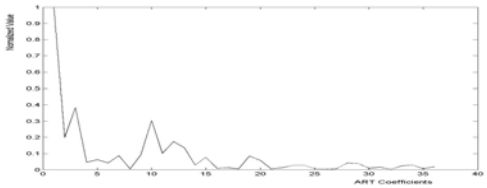



Figure 2. ART Coefficients of First Frame of "Akiyo"

## 3. AUTHENTICATION ALGORITHM

The procedure for generating a content-based watermark is shown in Figure 3. A series of ART coefficients of the video object are calculated first. Then, some are selected and quantized into a binary feature code called the feature vector. We will explain how to quantize the ART coefficients in Section 3.1. Next, an Error Correction Coding (ECC) scheme is utilized to encode the feature vector (Please refer to [8] for more details on exploiting ECC scheme). The ECC scheme selection should consider the difference between the feature vector extracted from the original object and the feature vector extracted from the manipulated video object. In our algorithm, the Parity Check Bit (PCB) of the ECC codeword should be capable of correcting this difference if this manipulation is acceptable. The ECC codeword is then hashed by a typical cryptographic hashing function such as MD5 to get a hashed value. The input of the hashing function also includes a user key and the background feature code, in order to create a secure link between foreground object and background video. By embedding the background feature into the object watermark, the integrity of video sequence can be protected. It means this object is not allowed for another background.

The hashed values of two consecutive frames are XORed. This can further improve the security of the watermark. The first $n$ bits of the XORed value, plus the PCB data, are used to generate

the message for watermarking. The reason that we only embed part of the XORed value is that the output of the MD5 hashing function is 128 bits, too long for an object based watermark. Finally, this watermark message is encoded again by another ECC scheme called message ECC Encoding. This step is necessary for the proposed authentication algorithm. It ensures that the watermark message can be extracted correctly from the received video object under acceptable manipulation.

The procedure for authenticating a received video object is shown in Figure 4. Actually, it is an inverse procedure of watermark generation. Firstly, the feature code is extracted from the received object. In the meantime, PCB data and the hashed result are also acquired by ECC decoding of the extracted watermark. Then, the PCB data is used to correct the extracted feature code. If the video object has been manipulated during video editing and transcoding, resulting in a difference between the feature code extracted from original object and that extracted from the manipulated object, this difference can be corrected if such a manipulation is acceptable. Following the same steps as the watermark generation, a hashed value based on the received object can be calculated. This hashed value and the hashed result obtained from the extracted watermark are compared bit by bit to decide whether the object is authentic.

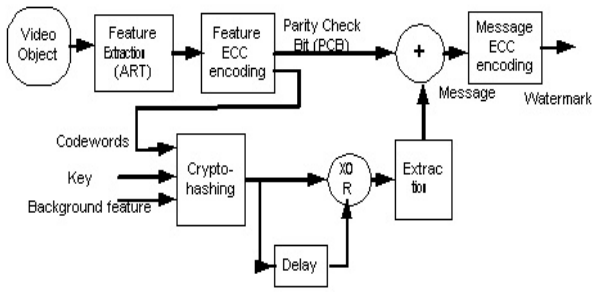



Figure 3. Content-based Watermark Generation

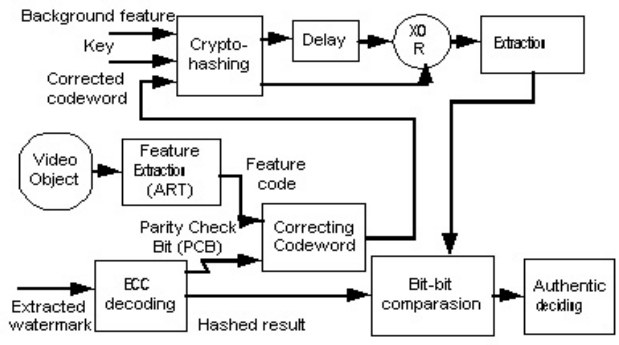



Figure 4. Video Object Authentication

### 3.1 ART Coefficients Quantization

Since the ART coefficient with large magnitude contributes greatly to video object recognition, we will assign more bits to represent this ART coefficient. In our algorithm, those coefficients with large magnitude will be quantized into 5 levels while those with less magnitude will be quantized into 3 levels. The quantization step size is pre-defined. We use Table 1 and Table 2 to convert the quantized value into quasi-gray binary code (feature vector). We can see that the Hamming distance between two adjacent codes is only 1. This is very important because every bit in the codeword can only represent one unit modification of the video content. Through this conversion, we can measure the difference of two objects by just calculating the distance between two corresponding feature codes.

| Level | Code |
|---|---|
| 0 | 00 |
| 1 | 01 |
| 2 | 11 |

Table 1. 2 bits Quantization

| Level | Code |
|---|---|
| 0 | 0000 |
| 1 | 0001 |
| 2 | 0011 |
| 3 | 0111 |
| 4 | 1111 |

Table 2. 4 bits Quantization

## 4. EXPERIMENTAL RESULTS

We have tested our algorithm on one QCIF format object-based video sequence "Weather" and three CIF format object-based video sequences: one is original "Akiyo" and another two are "Attacked Akiyo". "Attacked Akiyo" is created for the evaluation purpose of our authentication system performance. The shape information of "Akiyo" and "Attacked Akiyo" is the same. We took 250 frames from every video sequence for testing. Figure 5 (a) shows the first frame of original "Akiyo". Figure 5 (b) and (c) show two attacked videos: one is by replacing foreground object and the other is by replacing background.

Figure 5. (a) Original Aki (b) Attacked Aki 1 (c) Attacked Aki 2

Among the 15 selected ART coefficients, 4 coefficients are quantized into 5 levels while the other 11 coefficients are quantized into 3 levels. So the feature vector is 38 bits long. The dissimilarity of two video objects is measured according to Equation (3), which actually represents the Hamming distance between two feature vectors.

$$\text{Dissimilarity} = \left\| FV_i - FV_j \right\| \qquad (3)$$

Where $i, j$ represent two different video object

The robustness evaluation was performed on scaling, rotation and MPEG4 compression. Figure 6(a), 6(b) and 6(c) show the maximum Hamming distance between the original object and manipulated object in three video sequences under different scaling factors (From 0.5 to 0.9). Figure 6(d) shows the distance between the original object and the compressed object in the "Akiyo" sequence. From these Figures, we can find that the maximum Hamming distance is not greater than 3. During the experiment, we also found that the maximum distance between

the original object and the rotated object in a video sequence is not more than 1. So, if we adopt an ECC scheme that can correct 3 bits error in the codeword, the generated watermark will be robust for acceptable manipulation such as scaling, rotation and MPEG4 compression.

To prove that the 38 bits long feature vector is enough to represent the video content in our proposed authentication algorithm, we show the Hamming distance between the feature vector extracted from "Attacked Akiyo" and the feature vector extracted from "Akiyo" in Figure 7(a). The distance between the feature vector extracted from "Weather" and the feature vector extracted from "Akiyo" is given in Figure 7(b). From these two figures, we found that the distance in most frames exceeds 10. This is much larger than the maximum distance between the original video object and manipulated video object (i.e., 3). In Figure 7(b), we also found that distance suddenly decrease during the period from frame 200 to frame 230. By checking this period of video content, we found that these two video objects have some visual similarities. However, the minimum distance is still 7, which is lager than maximum distance between the original video object and manipulated video object (i.e., 3).

Since the maximum distance between the original video object and manipulated video object is 3, a BCH (63,45,3) which can correct 3 bits error in a block is exploited to generate the PCB data in this proposed algorithm. The length of PCB is 18. In [1], we have pointed out that more than 80% of the watermark bits can be correctly detected in our watermarking algorithm. For a total of 195 bits long watermark, the length of the message can be 48 if ECC for watermark encoding is another BCH (63,16,11). Therefore, including the 18 bits PCB data, 30 bits hashed value of the feature codeword from a total of 128 bits can be used to generate the watermark.

From the above results, we conclude the following: Using the selected feature vector, we can deduce whether the two objects are distinct objects or just modified duplicates of each other. So, this feature vector can be used to create a content-based watermark for our authentication applications.

## 5. CONCLUSION

In this paper, we have proposed a new object-based video authentication method. Region based shape descriptor, which is first adopted in MPEG7, is used as the feature of the video object. The feature vector is encoded by an ECC scheme. The Parity Check Bits (PCB) and the hashed value of the codeword are concatenated to generate a content-based watermark with a secret key. Experimental results further demonstrated that the proposed solution is robust during MPEG4 compression and normal MPEG4 object manipulations such as scaling, rotation and translation.

Our future work includes more tests and fine-tuning system parameters.

## 6. REFERENCES


[1] Dajun He, Qibin Sun and Qi Tian, "An Object Based Watermarking Solution for MPEG4 Video authentication", submitted to *ICASSP 2003*

[2] Schneider, M.; Shih-Fu Chang, "A Robust Content Based Digital Signature For Image Authentication", *Image Processing*, 1996. Proceedings, International Conference on , Volume: 3 , 1996 Page(s): 227 -230 vol.3

[3] C.-Y Lin and S.-F. Chang, " A Robust Image Authentication Method Surviving JPEG Lossy Compression", *SPIE Storage and Retrieval of Image/Video Database*, San Jose, January 1998

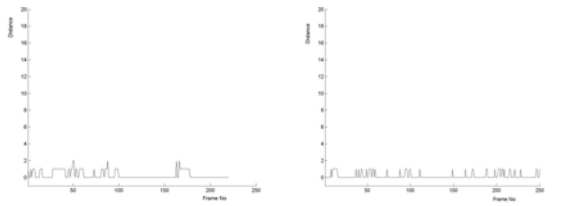
[4] Ditmann, J.; Steinmetz, A.; Steinmetz, R., "Content-based digital signature for motion pictures authentication and content-fragile watermarking", *Multimedia Computing and Systems*, 1999. IEEE International Conference on , Volume: 2 , 1999, Page(s): 209 -213 vol.2

[5] Queluz, M.P., "Towards robust, content based techniques for image authentication", *Multimedia Signal Processing*, 1998 IEEE Second Workshop on , 1998,Page(s): 297 –302

[6] W.-Y. Kim and Y.-S. Kim, " A New Region-Based Shape Descriptor ," *ISO/IEC MPEG99/M5472*, Maui, Hawaii, Dec. 1999

[7] Bober, M. "MPEG-7 Visual Shape descriptors", IEEE Transactions on *Circuits and Systems for Video Technology*, Volume: 11, June 2002, Page(s) 716-719.

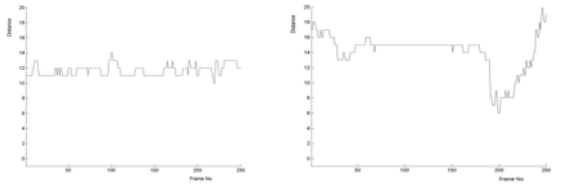
[8] Qibin Sun, Shih-Fu Change, Maeno, K and Suto,M , " A New Semi-fragile Image Authentication Framework Combining ECC and PKI Infrastructures", IEEE International Symposium on Circuits and Systems ,2002, ISCAS 2002, Volume: 2, Page(s): 440-443.

(a) "Akiyo" (Scaling) (b) "Attacked Akiyo"

(c) "Weather" (d)"Akiyo" (Compression)

Figure 6. Robustness under different manipulation



(a) "Attacked Aki" Vs "Aki" (b)"Weather" Vs "Akiyo"

Figure 7. Distances between different video object

Horizontal: Frame number 1-250 with the interval 50
Vertical: Distance –1 to 20 with the interval 2.