EURASIP Journal on Applied Signal Processing 2004:14, 2185–2200

A Secure and Robust Object-Based Video Authentication System

Dajun He
Institute for Infocomm Research (I^2R), 21 Heng Mui Keng Terrace, Singapore 119613
Email: djhe@i2r.a-star.edu.sg

Qibin Sun
Institute for Infocomm Research (I^2R), 21 Heng Mui Keng Terrace, Singapore 119613
Email: qibin@i2r.a-star.edu.sg

Qi Tian
Institute for Infocomm Research (I^2R), 21 Heng Mui Keng Terrace, Singapore 119613
Email: tian@i2r.a-star.edu.sg

Received 31 March 2003; Revised 6 December 2003

An object-based video authentication system, which combines watermarking, error correction coding (ECC), and digital signature techniques, is presented for protecting the authenticity between video objects and their associated backgrounds. In this system, a set of angular radial transformation (ART) coefficients is selected as the feature to represent the video object and the background, respectively. ECC and cryptographic hashing are applied to those selected coefficients to generate the robust authentication watermark. This content-based, semifragile watermark is then embedded into the objects frame by frame before MPEG4 coding. In watermark embedding and extraction, groups of discrete Fourier transform (DFT) coefficients are randomly selected, and their energy relationships are employed to hide and extract the watermark. The experimental results demonstrate that our system is robust to MPEG4 compression, object segmentation errors, and some common object-based video processing such as object translation, rotation, and scaling while securely preventing malicious object modifications. The proposed solution can be further incorporated into public key infrastructure (PKI).

Keywords and phrases: watermark, authentication, error correction coding, cryptographic hashing, digital signature.

1. INTRODUCTION

Nowadays, the object-based MPEG4 standard is becoming growingly attractive to various applications in areas such as the Internet, video editing, and wireless communication because of its object-based nature. For instance, in video editing, it is the object of interest, not the whole video, which needs to be processed; in video transmission, if the bandwidth of the channel is limited, only the objects, not the background, are transmitted in real time. Generally speaking, object-based video processing can simplify video editing, reduce bit rate in transmission, and make video search efficient. Such flexibilities, however, also pose new challenges to multimedia security (e.g., content authenticity protection) because the video object (VO) can be easily accessed, modified, or even replaced by another VO in object-based video application.

Consider a video surveillance system, shown in Figure 1. The captured video is sent to processing centers or end users via various channels. This video could even be further processed to serve as a legal testimony in the court in some surveillance applications such as automatic teller machine (ATM) monitoring system. In order to save the transmission and storage cost, only those video clips containing interesting objects are required to be sent and stored. Moreover, if the background changes very slowly, which is common in surveillance applications, a possible efficient solution is that only the objects are sent out frame by frame in real time while the background is sent once in a long time interval. In such application scenarios, it becomes very important to protect authenticity of the video, which involves two parts of the work: one is to protect the integrity of the object/background (i.e., any modifications on the video which result in the alteration of video meaning are not allowed), the other is to protect the identity of the transmission source (i.e., identify the video source). Although digital signature scheme is an ideal solution to protect the authenticity of the received data [1], it does not provide any robustness. The objective of this

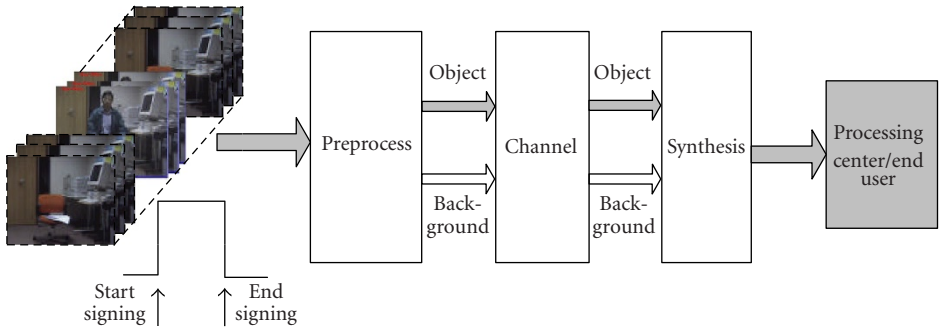

Figure 1: An example of video surveillance system. Only video clips containing interesting objects will be transmitted and authenticated; and the video frames are segmented into foreground (object) and background for individual transmission. So an object-based video authentication system is required.

paper is to propose a secure, robust, and object-based video authentication solution.

The requirements for an object-based video authentication system are different from those for a frame-based video authentication system; and these requirements are usually application dependent. As we know, it is very easy for the VO and background to be modified or even maliciously replaced by another object by the attacker during video transmission or storage. Hence, for security reasons, such kinds of attacks should be alerted and detected. (We define the distortions introduced by attacks as *intentional distortions*.) On the other hand, as shown in Figure 1, the video will usually be segmented into objects and background first, and the objects and background are then compressed individually before transmission; at the receiving site, the decompressed objects may even be scaled, translated, or rotated to interact with the end users. For robustness reasons, such kinds of video processing should be allowed. (We define the distortions introduced by above-mentioned object/video processing as *incidental distortions*.) Therefore, a practical object-based video authentication system should be robust to incidental distortions while being sensitive to intentional distortions for specific video applications.

Bartolini et al. [2] proposed a frame-based video authentication system for surveillance application. In their system, the camera ID, together with time and date, is used to generate the watermark. This watermark is then embedded into the raw video at the camera site. The watermarked video is sent to the central unit at the sending site. The central unit monitors and analyzes the video, and generates an alarm if necessary. The alarm, is transmitted with the video sequences causing this alarm to an intervention center via various channels. This watermarked video sequence can be used as a testimony in the court. They claimed that this system is robust to video compression with capability to detect tampering attack if the tampered region is not too small. However, their system is not suitable for object-based video authentication since the authentication level is at frame. Furthermore, the symmetric watermarking techniques used in this system may cause some security problems.

In an object-based video authentication system, every object in the video should be processed, signed, and authenticated separately. Several such algorithms have been proposed. MPEG4 Intellectual Property Rights by Adducing and Ordering (MIRADOR) project[1] was developed for copyright protection under the MPEG4 framework by integrating MPEG2 watermarking technologies into MPEG4. In this project, each object in the video is watermarked using a different value (identifier) and key in order to protect the copyright of the video sequence. If the illegal user exploits the object of interest in the video to create his own work, the content provider can detect such illegal usage. Piva et al. [3] also proposed a similar approach, where the raw video frame is segmented into foreground and background, denoted by VO0 (video object 0) and VO1, respectively; each VO is embedded with a different watermark in wavelet transform domain. These two works, however, only focus on the development of object-based watermarking algorithm.

Some other previous works on content-based watermarking/authentication are listed as follows. Lin and Chang [4] used the relationship between the DCT coefficients at the same position in different blocks of an image as the feature to generate the watermark. This feature is robust to multicycle JPEG compression, but it is not robust to image scaling or image rotation because the relationship between the DCT coefficients cannot be kept constant when image is scaled or rotated. Dittmann et al. [5] and Queluz [6] used the edge/corner of the image as the feature to generate a digital signature for authentication. Although this signature is robust to most video processing, its size is too large to create a content-based watermark; and the consistency of the edges/corners themselves under incidental distortions is also a problem.

To ensure that the authentication system is robust to incidental distortions, besides creating a robust content-based

[1] http://www.cordis.lu/infowin/acts/analysys/products/thematic/mpeg4/mirador/mirador.htm.

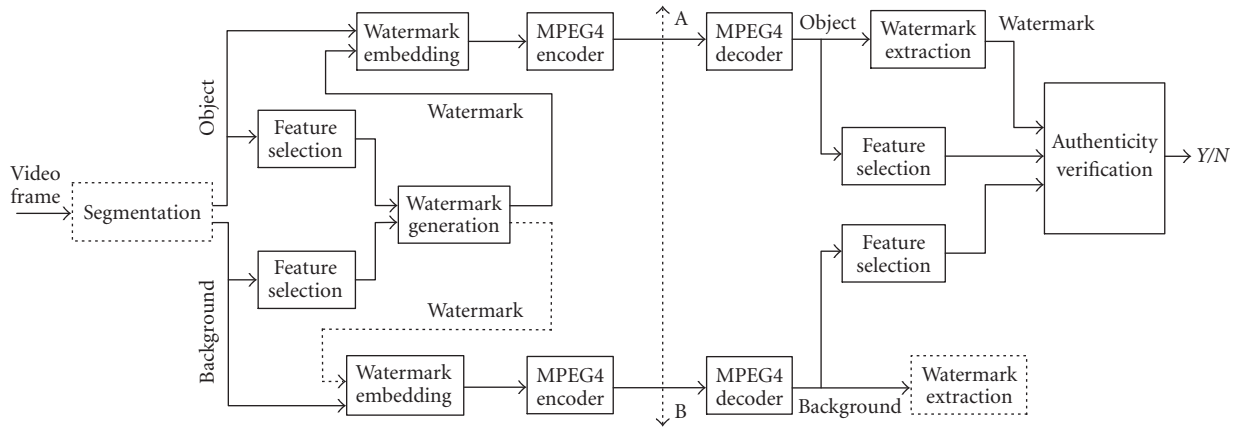

Figure 2: Block diagram of the proposed system. The block diagram is divided into two parts by the dashed line AB: on the left side is the procedure for signing, and on the right side is the procedure for video authenticity verification.

watermark, the watermarking algorithm should be able to protect the embedded watermark from such distortions. Many works have been done at object level. Boulgouris et al. [7], Lu and Liao [8], and Lie et al. [9] chose to embed the watermark in the DCT domain. But these algorithms are not robust to segmentation error since watermarked DCT blocks on the mask boundary are very sensitive to mask or segmentation error. To solve this problem, Bas and Macq [10] and Piva et al. [3] chose to embed a watermark in the spatial domain, and wavelet domain respectively.

In this paper, we propose a complete object-based video authentication system to protect the authenticity of video content, as shown in Figure 2. The procedures of watermark generation and video authentication are given in Figures 3 and 4. The feature of the object is error correction coding (ECC) encoded first; the resultant codeword, together with the feature of the background, is then hashed (e.g., MD5 or SHA-1 [11]) to get a short content-based message so that a secure link between the object and the background is created to protect the integrity of the video. The system's robustness to various video processing is guaranteed by ECC and watermarking; and the system's security is protected by cryptographic hashing which is a basic algorithm used to generate message authentication code (MAC [1]). We will explain these in Sections 2 and 3 in detail. It is worth mentioning here that this content-based message can also be signed by the system's private key to generate a digital signature (the part marked by the dotted lines in Figure 3) and therefore it can be easily incorporated into public key infrastructure (PKI).

The proposed system is the continuation of our previous works [12, 13]. The paper is organized as follows: the framework of the proposed system will be introduced in Section 2; the content-based watermark generation and verification are illustrated in Section 3; watermark embedding and extraction algorithm will be explained in Section 4; experimental results will be given in Section 5; conclusion and future works are presented in Section 6.

2. OVERVIEW OF THE PROPOSED SYSTEM

As we have described in the introduction section, the proposed system is designed to be robust to incidental distortions introduced by acceptable video processing. So we will define the acceptable video processing before briefing on the proposed system, because only after the acceptable video processing is defined, can we evaluate the robustness and security of the system.

2.1. Targeted acceptable video processing

Resizing (scaling)

The resources of end users may vary significantly. For example, monitors of some devices may not be able to display the video in its original resolution. So the received video has to be scaled to meet such limitation. As a result, the designed video authentication system must be robust to resizing (scaling) processing. In real application such as video transcoding, resolution is halved when the video is converted from CIF format to QCIF format. Since the information of the object is mostly preserved when the scaling factor exceeds 1, it is easier to design a robust video authentication system under this condition compared with the case when the scaling factor is less than 1. Thus, we will assume that the scaling factor is in a range from 0.5 to 1 in this system.

Rotation/translation

In video editing, the VO may be translated and/or rotated to meet end user's specific requirement. Because the video content remains unchanged after the object is rotated and/or translated, the proposed system should be robust to object rotation in any degree and object translation in any style.

Segmentation error

As we discussed in the introduction, the video is segmented into object and background for watermarking individually. Thus, segmentation may be required during video authentication. Again, we use ATM monitoring application as an example. Normally, the object and background are combined

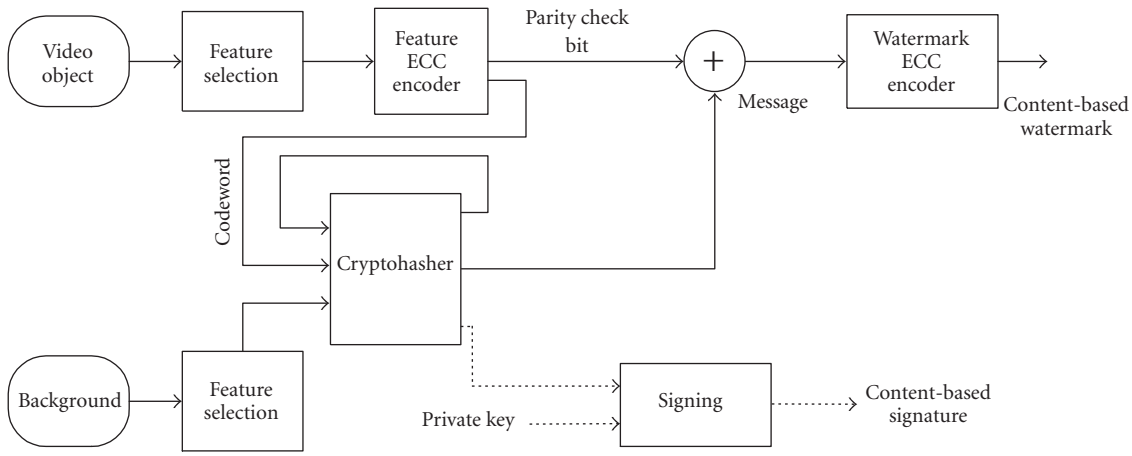

Figure 3: Procedure of watermark generation. The inputs are video object and background; and the output is a content-based watermark. A secure link between the object and background is created to protect the integrity of the video. The system can be incorporated into public key infrastructure.

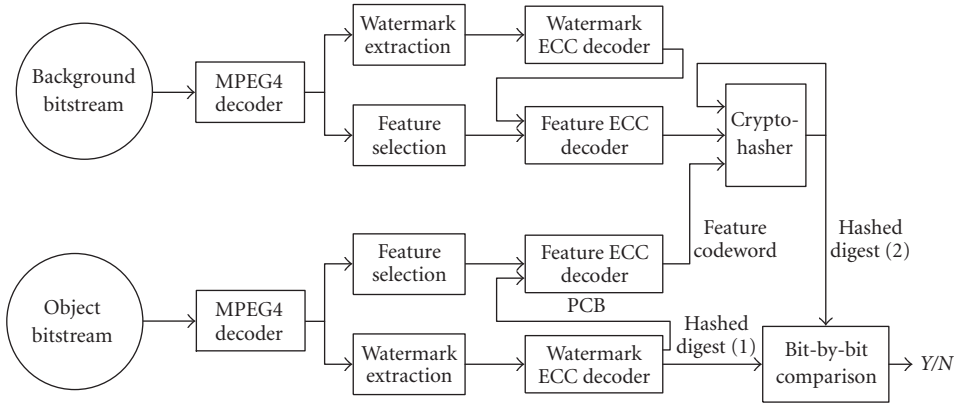

Figure 4: Procedure of video authentication. The inputs are video object and background; the output is the authenticity of video.

into a "raw video" for display and storage regardless of its authenticity at the receiving site. Only when necessary, such as the incidence of an illegal withdrawal of cash from an ATM machine, will the end users, bank in this case, check the video authenticity. So the video has to be segmented again before authentication can proceed. A segmentation error between the object at the sending site and the resegmented object at the receiving site may exist. Therefore, the proposed system should be robust to slight segmentation errors.

MPEG4 coding

The VO and background are usually compressed before they are transmitted. The proposed system should, hence, be robust to such kind of coding if the quality of video after compression is still satisfactory. In MPEG4 coding, two processes will affect the robustness of the video authentication system. One is the requantization process that is common in most coding schemes; the other is the repadding process that is unique for MPEG4 coding. In our system, we use bit rate to measure the requantization step.

Table 1: Acceptable video processing and their parameters for system evaluation.

Video processing	Parameters
Rotation	Any degree
Translation	Any direction and extent
Resizing	Scaling factor: 0.5–1.0
Requantization (MPEG4 coding)	Bit rate (512 Kbps)
Segmentation error	Edge error
Repadding (MPEG4 coding)	Zero padding

We summarize the typical acceptable video processing in Table 1. The parameters for evaluating the system are also listed here. (More details on experiments will be given in Section 5.)

2.2. System brief description

In the introduction, we have mentioned that only the video clips containing interesting objects are required to be transmitted or stored to save the transmission cost or storage

cost. So it is reasonable that we only protect the integrity of the video clips that contain interesting objects. How to define an interesting object is application dependent and usually very difficult. Fortunately, in surveillance applications, we have observed that the interesting objects, such as a person in an ATM monitoring system, are usually quite large. Thus, the size of the object is selected as a trigger to start signing and verification in the proposed system (refer to Figure 1). One video sequence may include many objects, and the objects themselves may overlap in some video frames. We will leave this case for future study. In this proposed system, we only focus on video sequence where every video frame contains one or more nonoverlap objects.

Figure 2 is the block diagram of the proposed system. It comprises two parts: signing and verification. The procedure for signing is on the left side, and the procedure for verification is on the right side.

In the signing procedure, the input could be in either raw video format (segmentation is needed in this case) or object/background MPEG4 compliant format while the outputs are signed MPEG4 bitstreams. Firstly, features of the object and its associated background are selected. Details of feature selection will be described later. Secondly, a content-based watermark is created based on the selected features, as shown in Figure 3. The selected feature of the object is ECC encoded first to ensure that the same feature can be obtained in the verification procedure in spite of the incidental distortion. For the reason of system security, the ECC codeword and the feature of the background are cryptographically hashed (e.g., MD5 or SHA-1) to get a content-based digest, which will be used to authenticate the video during verification. A content-based message is created based on this digest and parity check bit (PCB) data of the codeword. In practice, another ECC scheme, such as convolutional coding [14], is employed to increase the robustness of watermark to the incidental distortions. Following the similar procedure, the watermark for the background can also be created. For simplicity, we will only focus on the object processing (object feature selection, object watermark generation, and object watermark embedding and extraction) in this paper. The watermarks are embedded in the DFT domain of the object/background. Details will be given in Section 4. Finally, the watermarked object and background are compressed into MPEG4 bitstreams. Note that if we want to generate a digital signature for the video, what we need to do is to use the system's private key to sign on the hashed digest by the well-known digital signature scheme such as DSA or RSA [1].

To authenticate the received video, the MPEG4 bitstreams have to be decompressed to get the VO and background again. (In certain applications, if the inputs are already separated, object and background, this step can be skipped. If the input is a video sequence including object and background, segmentation should be employed first.) Following the same way as signing, features of the object and background can be obtained. Note that the features are computed from the received object and background. Meanwhile, watermark is extracted from the received object. Authenticity decision comprises two steps. In the first step, the PCB data contained in the watermark and the feature obtained from the received object are concatenated to form a feature codeword; the syndrome of the codeword is calculated to see whether it is correctable. If not, we claim that the video is unauthentic. If yes, we turn to the second step. The authenticity of the video is decided by bit-by-bit comparison between the two hashed digests. (One is the newly generated based on hashing ECC codeword and the other is extracted from the watermark.) Even if there is only one-bit difference between them, we claim that the authenticity of the video has been broken. More detailed description will also be given in Section 3.

From the above description, we can find that feature selection; watermark generation and authenticity verification; and watermark embedding and extraction are the three important parts of our proposed object-based video authentication system. In addition, a good segmentation tool is also very important for a successful authentication system. But it is beyond the scope of this paper.

3. WATERMARK GENERATION AND AUTHENTICITY VERIFICATION

Before discussing the watermark generation and authenticity verification, we will discuss feature selection.

3.1. Feature selection

The feature selected to create a watermark in our proposed video authentication system should have the following three properties: robustness, discriminability/sensitivity, and short length.

Robustness

The feature should have little or even no change if the video only undergoes acceptable video processing defined in Section 2.

Discriminability

The feature should be able to represent the specific VO. Features obtained from different objects should be different.

Short length

Because of the issue of watermark capacity, the feature size should be small.

Based on these criteria, angular radial transformation (ART) is selected as the feature of the VO in our solution. ART, which belongs to the broad class of shape analysis techniques based on moments, is one of the three visual shape descriptors, which are employed to represent the image object in MPEG7 [15, 16]. Studies show that ART has the following specific properties: firstly, it gives a compact and efficient way to describe the object; secondly, the descriptor is robust to segmentation errors and invariant to object rotation and shape distortions.

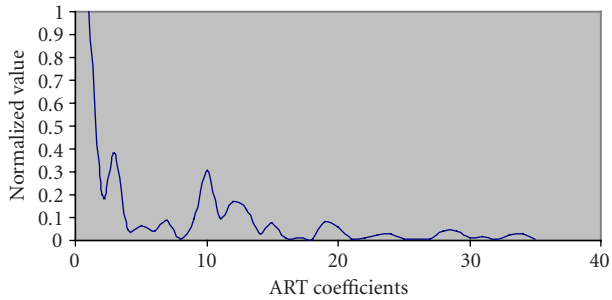

Figure 5: 36 normalized ART coefficients calculated from the first frame of video Akiyo. Usually high-order coefficients have smaller magnitudes.

The definition of ART is given in equation (1). F_{nm} is an ART coefficient of order n and m; $V_{nm}(\rho, \theta)$ is the ART basic function that is separable along the angular and radial direction; and $f(\rho, \theta)$ is a gray image in polar coordinates. In MPEG7, $f(\rho, \theta)$ represents mask function. Here, we modify this ART definition to make sure that the ART coefficients are sensitive to some intentional attacks, in which only the content of the object is modified while the shape of the object remains unchanged. This modification on ART definition, however, may bring an impact: the robustness of high-order ART coefficients may be affected by the interpolation error during rotation and scaling. Later, we will see how this impact could be overcome:

$$\begin{aligned} F_{nm} &= \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle \\ &= \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta), f(\rho, \theta) \rho \, d\rho \, d\theta, \end{aligned} \tag{1}$$

and

$$\begin{aligned} V_{nm}(\rho, \theta) &= A_m(\theta) R_n(\rho), \\ A_m(\theta) &= \frac{1}{2\pi} \exp(jm\theta), \\ R_n(\rho) &= \begin{cases} 1, & n = 0, \\ 2\cos(\pi n \rho), & n \neq 0. \end{cases} \end{aligned} \tag{2}$$

Figure 5 shows the 36 normalized ART coefficients of the first frame of video sequence "Akiyo," shown in Figure 6a. In MPEG7 specification, 35 ART coefficients (excluding the first coefficient) are recommended to be used, and each coefficient is quantized to 4 bits/coefficient. Hence, a region-based shape descriptor is a 140-bit data. But in our video authentication system, 140 bits is still too long to generate a watermark. Fortunately, from the following discussion, we will see that not all these 35 ART coefficients are necessary to be selected and not every coefficient is necessary to be quantized to 4 bits.

As we have mentioned above, the feature of the VO for generating a watermark should be robust and well discriminable. Among the ART coefficients, the low-order coefficients, which are also considered as low-order moments, represent the basic information of the VO while the high-order coefficients, which are also considered as high-order moments, represent the detailed information of the VO. So, the low-order coefficients should be part of the selected feature. The selection of high-order coefficients should consider both the robustness and the discriminability. For high discriminability, these coefficients, which represent the detailed information, must be kept. But for robustness, these coefficients, which may change due to loss of high-frequency information and interpolation error when object is scaled or rotated, should be discarded. It is a trade-off. The dissimilarity measure equation [15], which is used to measure the dissimilarity between two objects, is useful for us to choose the high-order coefficients:

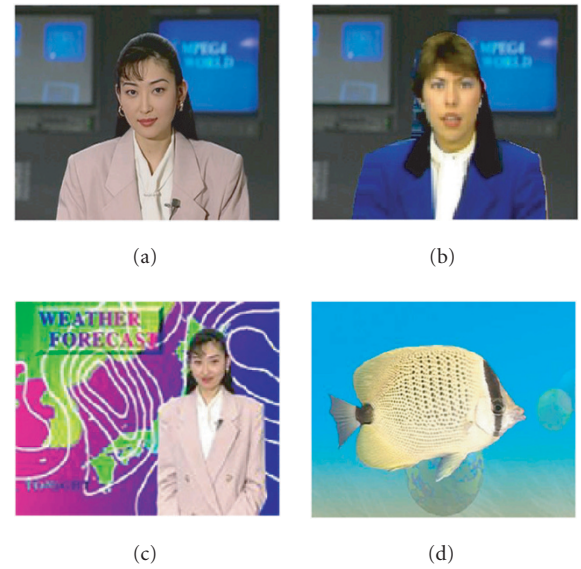

Figure 6: Video sequences for feature vector evaluation. (a) Akiyo, (b) Attacked Akiyo 1, and (d) Bream are CIF videos. (c) Weather is QCIF video. The shapes of (a) and (b) are identical.

$$\text{Dissimilarity} = \sum_i ||M_d[i] - M_q[i]||, \tag{3}$$

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

This equation clearly indicates that the coefficients with smaller magnitudes have less contribution to object recognition. So, discarding these coefficients will not affect the representation of the object significantly. Since high-order coefficients usually have smaller magnitudes, in the proposed system, we scan the 36 ART coefficients in zigzag format, and select the first 18 ART coefficients as the feature of the object. This information is known to both sending and receiving sites once the decision has been made. We will explain why we select the first 18 ART coefficients as the feature of the object later in this section.

In our proposed system, those coefficients with large magnitudes will be quantized into 5 levels while those with smaller magnitudes will be quantized into 3 levels to further shorten the length of feature. Moreover, the quantized value of each ART coefficient is converted into a quasigray binary code according to Tables 2 and 3. Finally, a feature vector

Table 2: Two-bit quantization.

Level	0	1	2
Code	00	01	11

Table 3: Four-bit quantization.

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

(FV) is created to represent a VO by concatenating the binary codes. The quasigray code conversion is to ensure that one-bit change of FV only represents one unit modification on the feature of video content so that the difference between two objects can be easily measured by just calculating the Hamming distance between their two corresponding FVs. Hence, equation (3) can be rewritten as equation (4)

$$\text{Dissimilarity} = ||\text{FV}_d - \text{FV}_q||, \tag{4}$$

where d and q represent two different VO.

Now, we start to explain why the first 18 ART coefficients are selected as the feature of VO based on feature selection criteria. The empirical way to select the ART coefficients will be given first. Experimental results will then be given to prove that such a selection is correct.

The empirical way to filter the ART coefficients is as follows.

(a) Scan the 36 ART coefficients in zigzag format.

(b) As we have mentioned previously, low-order ART coefficients should be part of our selected feature. We make a reasonable assumption that the first 9 ART coefficients are classified as low-order coefficients, and hence must be part of the selected feature. So we only consider the ART coefficients from 10 onwards. For a specified N_i ($10 \le N_i \le 35$), the first N_i ART coefficients are used to create an FV. Then, the maximum Hamming distance between the FV extracted from the original object and the FV extracted from the object processed by all kinds of acceptable video processing is computed. If this distance is small and robust for all training video sequences, N_i will be considered as the candidate of robustness group G_r, $G_r = \{N_i, \ldots, N_k\}$.

(c) For every member in the robustness group G_r, the Hamming distance between FVs extracted from two different VOs is calculated. If this distance is much larger than the maximum Hamming distance between the original VO and the processed VO (i.e., 3), we can say that this feature has good discriminability. All these kinds of N_i are defined as a group G_{rs}.

(d) The smallest number in the group G_{rs} is the number of ART coefficients that should be selected.

Four video sequences are used during the above-mentioned filtering. They are CIF videos "Akiyo," "Attacked Akiyo 1," "Bream," and QCIF video "Weather." Figure 6 shows the first frame of every video sequence. Among them, the shape information of video Attacked Akiyo 1 is identical to that of the video Akiyo. We give the experimental results when the first 18 ART coefficients are selected to create an FV.

Figure 7 shows the maximum Hamming distance between the FV extracted from the original object and that from the object having undergone various video processing. The video processing includes scaling (scaling factor is in a range from 0.5 to 0.9), rotation (rotation degree is in a range from 10° to 30°) and MPEG4 compression, and "Cubic" interpolation technique is employed during rotation and scaling processing. From the results, we can find that the maximum Hamming distance is not greater than 3. So we can claim that the selected feature is robust to object rotation, object resizing, and MPEG4 compression. Figure 8 shows the difference between the mask of original object and the mask of processed object. From Figure 8c, we can see that this difference is similar to segmentation error. The maximum error on the edge is about 4 pixels. Usually the maximum pixel error on the edge of segmented object is also around this number. We therefore simulate the segmentation error by using the mask difference. Because the selected FV is robust to such kind of mask difference, we claim that the selected feature is robust to slight segmentation error.

Figure 9a shows the Hamming distance between the FV extracted from Attacked Akiyo 1 and that from Akiyo. The distance between the FV extracted from Bream and that from Akiyo is given in Figure 9b. From both figures, we find that the Hamming distance exceeds 10 in most frames. This is much larger thanthe maximum Hamming distance between the originalobject and the processed object. Hence, we conclude that the selected feature has good discriminability, because we can easily deduce whether the two objects are distinct objects or just modified duplicates of each other from the FVs.

Above experimental results demonstrate that the FV created from the first 18 ART coefficients can meet the robustness and discriminability requirements. So we will select the first 18 ART coefficients as the feature of the object in the proposed video authentication system.

3.2. Watermark generation and authenticity verification

The procedure for a content-based watermark generation is shown in Figure 3. Firstly, a systematic ECC scheme [17] is utilized to encode the FV of VO to obtain a feature codeword. (We define this ECC scheme as feature ECC coding scheme.) A systematic ECC scheme means that after ECC encoding, its codeword can be separated into two parts: one is its original message and the other is its PCB data. Secondly, the feature codeword, together with FV of background, is hashed by a typical hash function such as MD5 or SHA-1 to get a hashed digest. By including the background FV into the object watermark, a secure link between the object and background has been created. So this object is not allowed for another background. Thirdly, the hashed digests from two consecutive frames are further processed (e.g., XORed) to create a secure link between these two frames. This can prevent the order

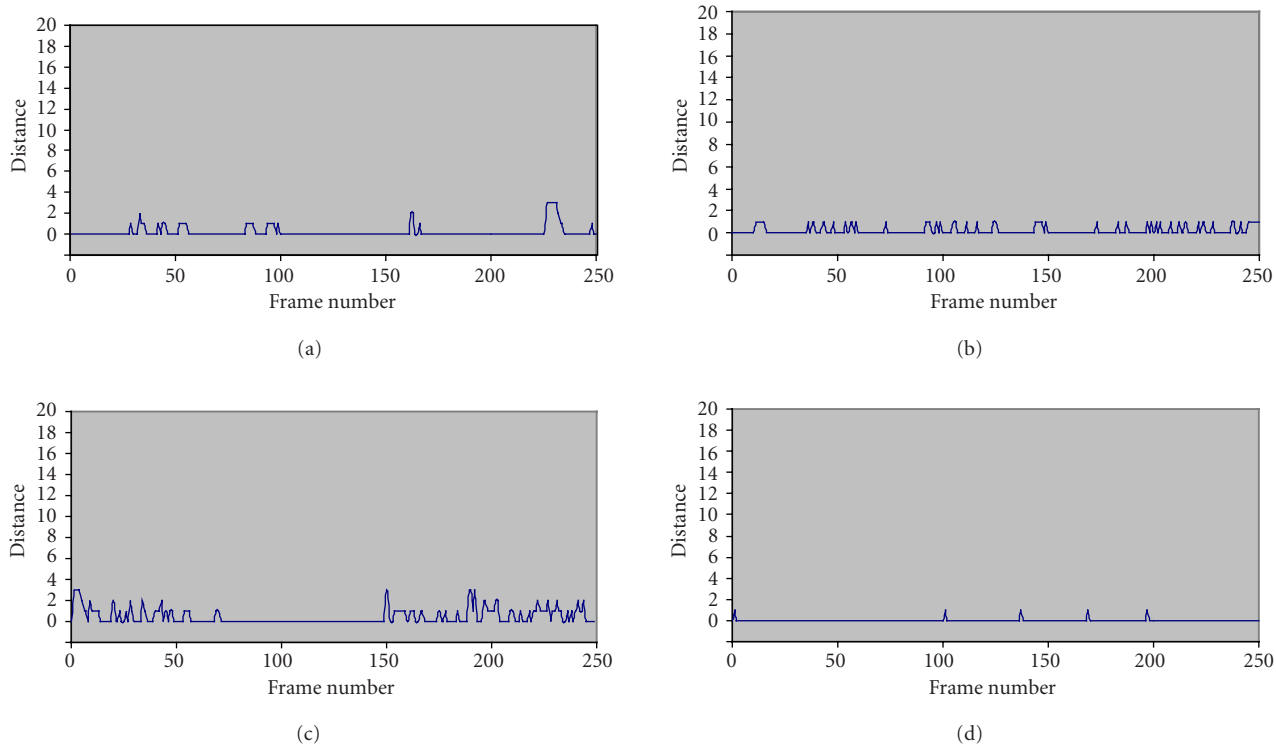

Figure 7: (a) Akiyo, (b) Attacked Akiyo 1, and (c) Weather show the maximum Hamming distance between the feature vector obtained from the original object and that from the object undergoing various video processing including rotation, resizing, and MPEG4 coding. (d) Bream (resizing only) shows the maximum Hamming distance between the original object and the scaled object. The horizontal axis represents the frame number, and the vertical axis represents the Hamming distance. From these figures, we can find that all maximum Hamming distances are not bigger than 3.

of video frames from being changed. Fourthly, the first n-bit data of the processed result and the PCB data of the feature codeword are concatenated to create a content-based message. Finally, this message is encoded again by another ECC scheme (we define it as watermark ECC coding scheme) to generate a content-based watermark. Figure 10 illustrates intermediate results during watermark generation. It is found that the watermark is a convolutional code by encoding the content-based message.

From the procedure of watermark generation, we can find that authenticity of video is protected securely. The security of the video is achieved by hashing the ECC codeword. Further processing of the hashed digests from two consecutive frames also improves the security of the video. In this system, only part of the hashed digest (30 bits) instead of all the 128 bits is used to create the watermark because of the limitation in watermarking capacity. Note that reducing length of hashed digest may cause some security risks; however, it is still acceptable in real applications, because a strong contextual property in video content also makes attacking content difficult. Clearly, our proposed object-based solution can be easily incorporated into PKI by generating a digital signature based on the hashed digest and a selected signature scheme such as RSA or DSA [1].

The procedure for authenticating received VO is shown in Figure 4. No original video is needed during authentication. Firstly, PCB data and the hashed digest are obtained by ECC decoding the extracted watermark. (It is the reverse process of watermark ECC coding in watermark generation.) Secondly, FV of the received object is calculated, which may be different from the FV of the original object since the received VO may have undergone various video processing; PCB data is used to correct this difference. Thirdly, a hashed digest for the received object can be calculated following similar steps in watermark generation. Finally, this hashed digest and the hashed digest obtained from the extracted watermark are compared bit by bit to decide the authenticity of the object: even if there is only one-bit difference, the video will be claimed as unauthentic.

We roughly explain the feature ECC coding scheme and the watermark ECC coding scheme employed in the proposed system. Please refer to [18] for more details on employing ECC scheme.

Feature ECC coding

As we have discussed in Section 3.1, the difference between FV obtained from the original object and that from the processed object may exist. This Hamming distance, however,

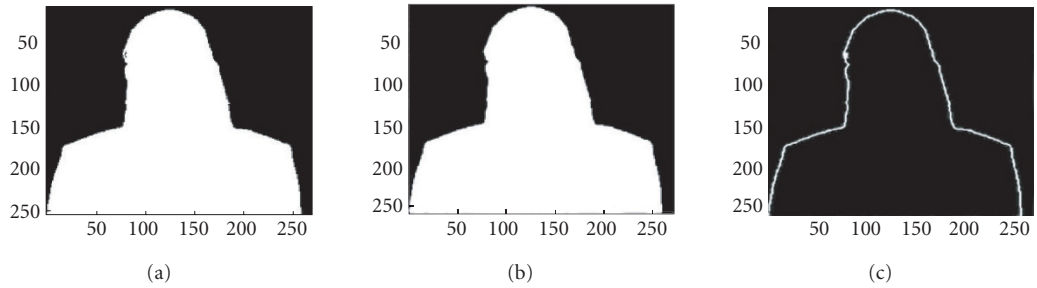

Figure 8: Mask difference (c) between the original object: mask (a) and the processed object mask (b). This difference is similar to the segmentation error.

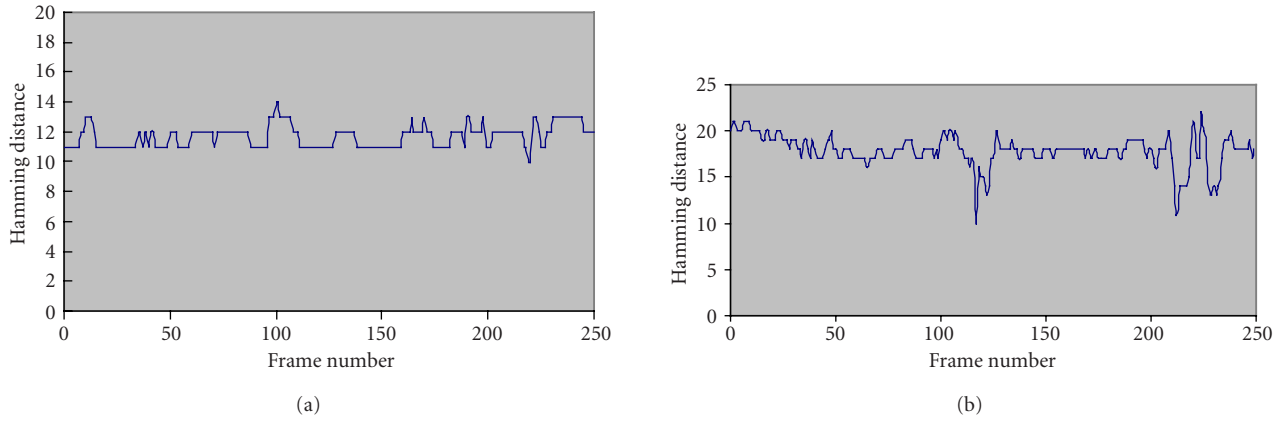

Figure 9: The distance between feature vectors obtained from different video objects (a) Akiyo and Attacked Akiyo and (b) Akiyo and Bream.

is relatively small compared with the Hamming distance between FVs obtained from two different VOs. So, the design of feature ECC coding scheme should follow such a rule: it should be able to correct the difference between FVs obtained from the original object and that from the object that has undergone acceptable video processing; on the other hand, it should not be possible for the difference between FVs obtained from different objects to be corrected. For instance, the maximum Hamming distance is 3 in Section 3.1, so the selected feature ECC coding scheme must be able to correct 3 errors in the feature codeword.

Watermark ECC coding

Since the received VO may undergo a series of acceptable video processing, the extracted watermark is usually not identical to the embedded watermark. But in our system, we have to get the original message (i.e., the PCB data of the feature codeword and hashed digest) from the extracted watermark. This problem is solved by introducing the watermark ECC coding scheme. The design of watermark ECC coding scheme should consider the error ratio between the extracted watermark and the original watermark in order to make sure that the original message can be recovered free of errors.

4. WATERMARK EMBEDDING AND EXTRACTION

To design an object-based video authentication system that is robust to various acceptable video processing, the watermarking algorithm should also be robust to these forms of video processing.

Fourier-Mellin transform is widely used in pattern recognition because of its invariance to image rotation, scaling, and translation. A watermarking algorithm based on Fourier-Mellin transform is first suggested by O'Ruanaidh and Pun [19]. Lin et al. [20] further proposed a rotation, scale, and translation resilient watermarking scheme closely related to the Fourier-Mellin transform. However, unbalanced sampling of DFT coefficients, introduced by log-polar mapping and inverse log-mapping in Fourier-Mellin transform, may cause a loss of image quality. The computation expense and watermark capacity of Fourier-Mellin transform-based watermarking algorithms still cannot meet the needs in designing our object-based video authentication system. This could be one of our future works.

In our system, we choose to embed the watermark in DFT domain because of good properties of DFT in object rotation and scaling. Before the detailed procedures of watermark embedding and extraction are given, we will discuss some specific solutions for keeping the watermark robust to acceptable video processing.

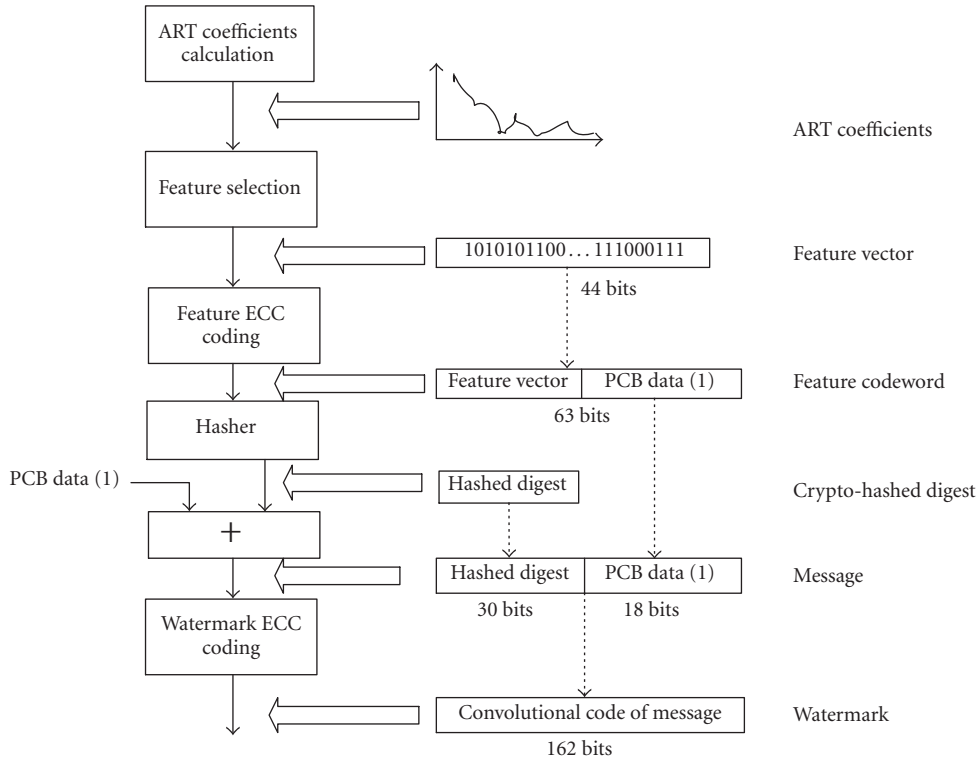

Figure 10: Illustration of watermark generation. The main procedure is on the left side, and the intermediate results are on the right side. PCB data (1) is the PCB data of feature codeword.

4.1. Challenges and solutions for object-based watermark embedding and extraction

Repadding

In order to perform DFT on VO, the VO has to be padded to form a rectangular area (image) first since the shape of the object is usually irregular. However, this padding will affect the robustness of watermarking because the watermark information will be not only in the object area but also in the padding area after watermark embedding and inverse DFT (IDFT). During MPEG4 encoding, most of this padding area will not be encoded and transmitted so that some watermark information may be lost. If the padding area is large, some watermark bits will not be correctly extracted. To solve this problem, the size of the padding area should be limited, and the relationship between two DFT coefficients instead of the individual DFT coefficient is used to embed the watermark.

Rotation

Lin et al. [20] also pointed out that the DFT of a rotated image is not the rotated DFT of the original image due to the rectilinear tiling, as shown in Figure 11. It indicates that if the watermarked object is rotated, it has to be rotated back for watermark extraction. However, such kind of rotation will introduce interpolation error that affects robustness of watermarking. To overcome this problem, the relationship between two groups of DFT coefficients, rather than the relationship between two DFT coefficients, is used to embed the watermark, as shown in Figure 12.

Rotation also causes the synchronization problem in watermark embedding and extraction. We adopt the method given in [8] to solve this problem. The eigenvectors of the object are calculated first, and then the VO is adjusted to align its maximum eigenvector with the x-axis in two-dimensional space before watermark embedding and extraction. Ideally, the alignment parameters (e.g., the rotation degree) used for watermark embedding should be identical to those used for watermark extraction. In practice, this assumption is not appropriate. Many factors such as mask error caused by the object manipulation will interfere in getting exactly the same alignment parameters in watermark extraction. By using the relationship between two groups of DFT coefficients to embed the watermark, such incurred distortion will be alleviated.

Resizing (scaling)

In this paper, we only focus on the cases when the scaling factor is less than 1. We will also assume this scaling factor is known during watermark extraction. This assumption is reasonable since we can get the resolution information about the original object by analyzing the MPEG4 bitstream. So the object can be scaled back to its original resolution before watermark extraction.

Although the received object can be scaled back to its original resolution before watermark extraction, the high-frequency information will be lost. An approximate relationship between the DFT coefficients before and after the image is scaled is given in equation (5). $F_p(u, v)$ represents DFT

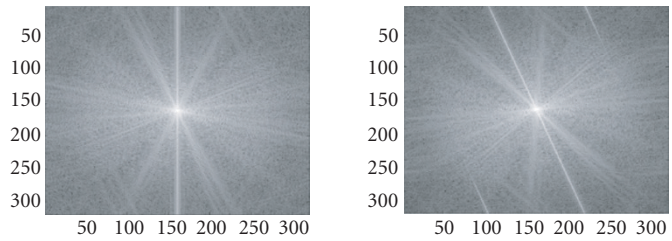

Figure 11: Comparison between (a) DFT of the original Akiyo image and (b) DFT of the 20° rotated Akiyo image. The DC of the DFT is located in the center of the image. These two figures show that the DFT of a rotated image is not the rotated DFT of its original image.

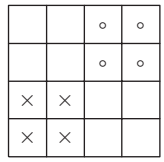

Figure 12: DFT coefficients grouping for watermarking. The bottom-left corner points to the origin of DFT domain. Eight DFT coefficients in 4×4 group are classified into two subgroups.

coefficients of processed image while $F_o(u, v)$ represents DFT coefficients of original image. α is scaling factor while D is half of the image resolution:

$$F_p(u, v) \approx F_o(u, v) \quad |u, v| < \alpha D. \tag{5}$$

Figure 13 shows this phenomenon; that only the low-frequency and middle-frequency DFT coefficients have no change or little change after the object is scaled. Equation (5) also reminds us that scaling factor α should be closely examined before we select a frequency area to embed watermark. For example, if the scaling factor is 0.5, watermark should be embedded in an area where all frequencies in this area are less than half of the maximum frequency. In addition, we have to bear in mind that the modification on low-frequency coefficients will significantly degrade the image quality. Considering these two factors, a low-middle frequency area (texture-filled part in Figure 14a) is selected to embed watermark in our algorithm. Note that Solachidis and Pitas [21] and Barni et al. [22] also have similar considerations in their proposed algorithms. In comparison, we also show the area selected by Barni et al. [22] to embed watermark in Figure 14b. Again, the interpolation error introduced by scaling object will also affect the robustness of the watermarking algorithm. Such incurred distortion, however, will be alleviated by using the relationship between two groups of DFT coefficients to embed watermark.

4.2. Watermark embedding

The embedding procedure comprises the following three steps: pre-embedding, embedding, and postembedding.

Pre-embedding

(a) Adjust the object to align its mask's maximum eigenvector with the x-axis in two-dimensional space.

(b) Expand the object into a rectangular image whose size is predefined. The selection of image size should consider the object size. Zero padding is used during expansion.

(c) Compute the DFT coefficients; shift the DFT coefficients in order that the DC coefficient locates in the center of DFT domain.

(d) Randomly select groups of DFT coefficients in the low-middle frequency area (shown in Figure 14a). The seed to generate the random number sequence is only known to watermark embedding site and watermark detector. Eight DFT coefficients in each 4×4 group are classified into two subgroups, defined as SG1 and SG2, respectively, as shown in Figure 12.

Embedding

(e) Compute the energies ($E_{1,2}$) for SG1 and SG2:

$$E_{1,2} = \sum_{\text{SG1,SG2}} ||F(u, v)||, \tag{6}$$

where $F(u, v)$ is the DFT coefficient at frequency (u, v).

(f) Calculate the adaptive threshold (TH) for each group:

$$\text{TH} = \alpha * \left(\frac{f}{M/2}\right)^{\beta}, \tag{7}$$

where $f = \sqrt{u^2 + v^2}$ stands for the frequency at point (u, v), M is the image size (e.g., 384 or 256), and α, β are two constants estimated from statistical distribution of all DFT coefficients of video sequence. For instance, we could derive $\alpha = 30\,000$, $\beta = 1.8$ for video Akiyo. Note that TH plays a very important role in maintaining the balance between the watermarked video quality and watermark robustness.

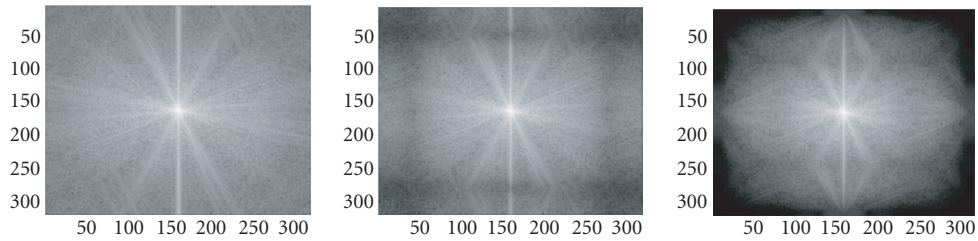

Figure 13: Comparison between (a) DFT of the original Akiyo image, (b) DFT of the scaled Akiyo image (scaling factor = 0.75), and (c) DFT of the scaled Akiyo image (scaling factor = 0.5). The DC of the DFT is located in the center of the image. Figures show that only the low-frequency coefficients and middle-frequency coefficients have little or no change after the image is scaled.

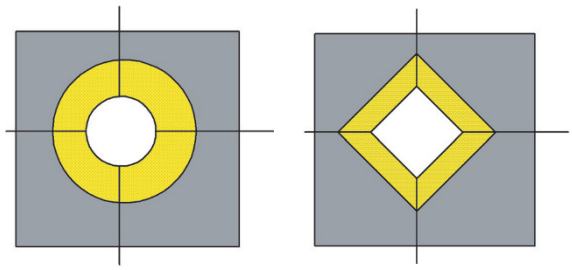

Figure 14: Area classification in DFT domain. The texture-filled area is selected to embed the watermark: (a) selected area in our proposed algorithm and (b) selected area in Barni's algorithm.

(g) Modify the threshold according to the following equation:

$$\mathrm{MTH} = \begin{cases} \dfrac{2}{3}\,\mathrm{TH} & (E_1 - E_2 > 0,\ \text{watermark bit} = 1) \\ & \text{or } (E_1 - E_2 < 0,\ \text{watermark bit} = 0), \\ \dfrac{4}{3}\,\mathrm{TH} & \text{others.} \end{cases} \quad (8)$$

Iteratively modify E_1 and E_2 until equation (9) is satisfied. Note that E_1 and E_2 should be kept as positive values during iterative calculation:

$$\begin{aligned} E_1 - E_2 &> \mathrm{MTH}, \quad \text{watermark bit} = 1, \\ E_2 - E_1 &> \mathrm{MTH}, \quad \text{watermark bit} = 0. \end{aligned} \quad (9)$$

(h) Based on the newly modified energies E_1 and E_2, adjust the magnitude of every DFT coefficient in SG1 and SG2 while keeping their phases unchanged. The modification of each DFT coefficient is with reference to its original magnitude.

(i) For every watermarked coefficient, its symmetric coefficient with reference to the DC component of DFT should also be modified to ensure that the watermarked video pixels have real values.

Figure 15: Surveillance video Dajun.

Postembedding

(j) The watermarked image is generated using IDFT. This image is rotated back to its original orientation.

(k) Finally, a watermarked VO is extracted again for MPEG4 encoding.

4.3. Watermark extraction

Watermark extraction is the reverse procedure of watermark embedding. The received VO is scaled back to its original resolution first. Then, similar to the embedding procedure, E_1 and E_2 for each subgroup are calculated. Finally, the watermark bit is extracted based on the following criterion:

$$\text{watermark bit} = \begin{cases} 1, & E_1 - E_2 > 0, \\ 0, & \text{else.} \end{cases} \quad (10)$$

4.4. Evaluation of the robustness of watermarking algorithm

To evaluate the robustness of watermarking algorithm, correct ratio of the extracted watermark is defined in the following equation:

$$\text{correct ratio} = \frac{\text{no. of watermark bits detected correctly}}{\text{length of watermark}}. \quad (11)$$

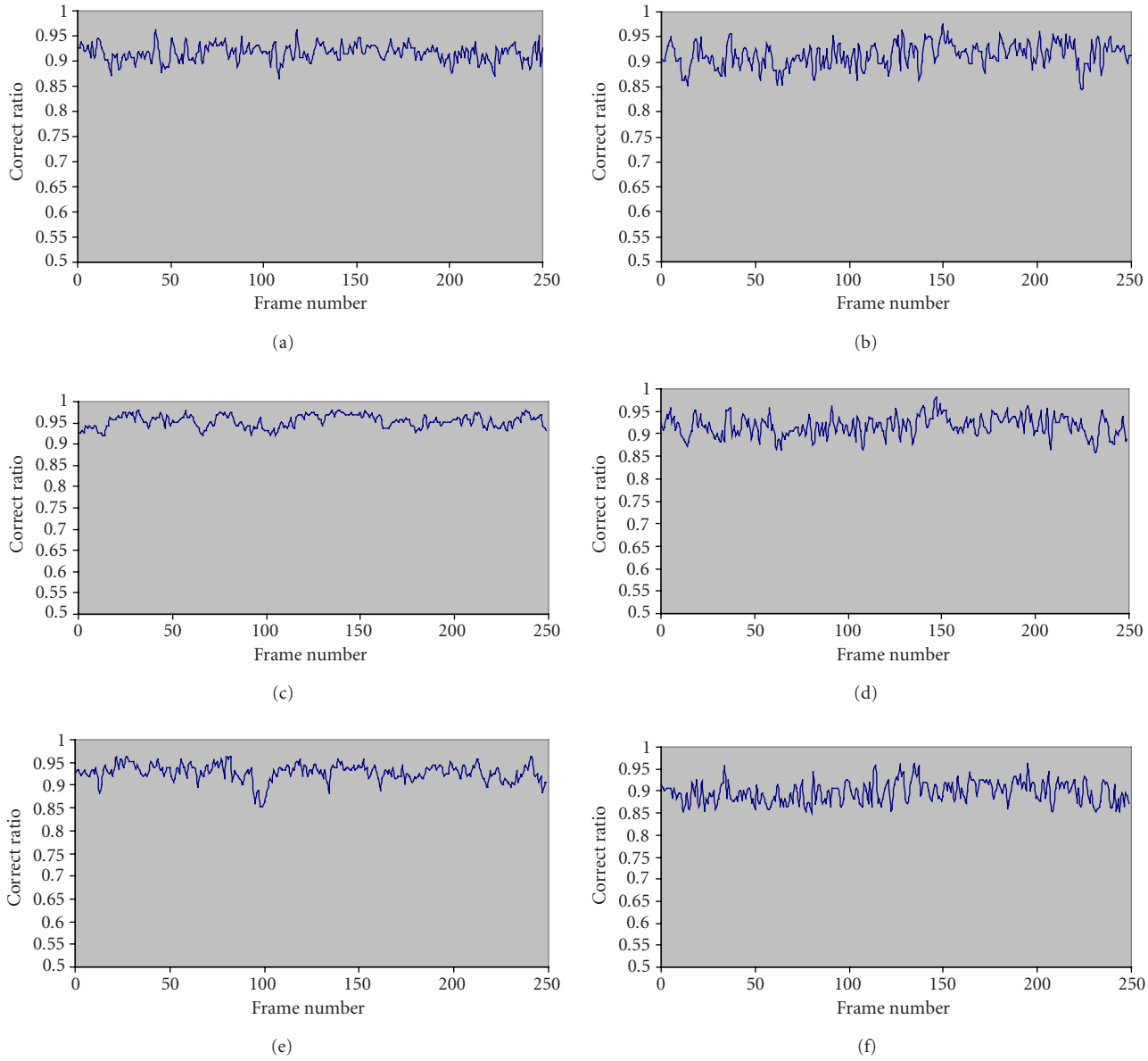

Figure 16: Correct ratio of extracted watermark under various video processing. All figures show that the correct ratio is larger than 0.85. (a) Akiyo (scaling factor = 0.5), (b) Bream (scaling factor = 0.5), (c) Akiyo (rotation degree = 30), (d) Bream (rotation degree = 20), (e) Akiyo (rotation/scaling/MPEG4 coding), (f) Bream (rotation/scaling/MPEG4 coding).

The videos used for evaluation are the same as what we have used in feature selection evaluation except that the QCIF video Weather is replaced by a surveillance video called "Dajun," shown in Figure 15. During the evaluation, a content-based 162-bit watermark is used. Some results of evaluation based on Akiyo and Bream are shown in Figure 16. From Figure 16, we find that the correct ratio is bigger than 0.85 if the object is rotated, scaled, or processed by a combination of rotation, scaling, and MPEG4 coding. Similar results can also be obtained from evaluation of other videos. All these results illustrate that the watermarking algorithm is robust enough to ensure that the original message for watermark generation can be correctly extracted if the watermark ECC coding scheme is properly designed.

5. EXPERIMENTAL RESULTS

The objective of this paper is to design an object-based video authentication system that can protect the integrity of video (object/background) while allowing various natural video processing. In our experiment, 300 frames from every video are used for evaluation. We mainly evaluate the system in following four aspects:

(a) whether this system is robust to acceptable video processing defined in Section 2;
(b) whether this system can detect object replacement;
(c) whether this system can detect background replacement;
(d) whether this system can detect object modification.

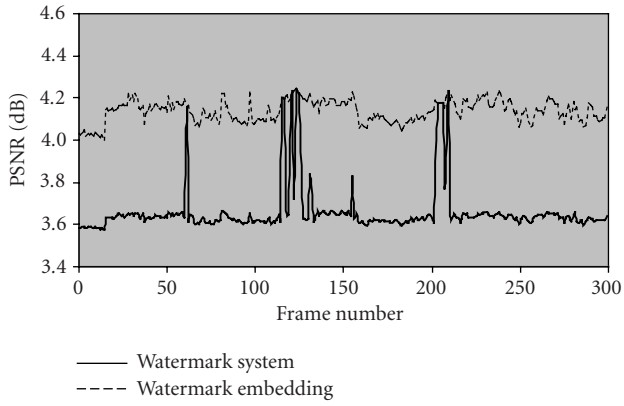

Figure 17: PSNR of watermarked video Akiyo. The dashed line represents the PSNR of object before and after watermarking with only watermarking distortion. The solid line represents the PSNR of watermarking system with not only watermarking distortion but also alignment distortion and interpolation error.

(a) (b)

Figure 18: Comparison between (a) the original frame and (b) the watermarked frame (of Akiyo).

We have developed this system using Visual C++ 6.0. In this system, the first 18 ART coefficients from total 36 ART coefficients in MPEG7 shape descriptor are selected as feature of the object. Among these 18 coefficients, 4 coefficients are quantized into 5 levels (4 bits), and the other 14 coefficients are quantized into 3 levels (2 bits). The resultant FV is a 44-bit data ($4 \times 4 + 2 \times 14 = 44$). We employ BCH (63,45,3) ECC coding scheme, which has 3 bits error correction capability (this capability is equal to the maximum Hamming distance between the FV obtained from the original object and the FV obtained from the object undergoing various video processing), as the feature ECC coding scheme. Eighteen-bit PCB data of the feature codeword and 30-bit hashed digest are concatenated to create a 48-bit message. Convolutional coding ($K = 7$, rate $= 1/3$) is used as the watermark ECC coding scheme. The start and end states are all set to 0. As a result, the watermark is a 162-bit binary data.

As we mentioned in Section 2, the size of object in video is chosen as a trigger to start signing and verification in this system. During evaluation, the system automatically starts signing or verification if ratio between object and its associated image is greater than 15% for a CIF video.

Recall that in watermark embedding procedure, we need to estimate the orientation of the object before aligning its mask's maximum eigenvector with the x-axis in two-dimensional space. After embedding watermark, the watermarked object needs to be rotated back to its original location. So two PSNR exist to measure the system's performance. One is the PSNR of object before and after watermarking with only watermarking distortion; the other is the PSNR between the original object and the watermarked object, including watermarking distortion, alignment distortion as well as interpolation error introduced by rotation. The dashed line in Figure 17 represents the former PSNR, and the solid line represents the latter PSNR. From Figure 17, we can find that two PSNR in some frames are equal. This is because no alignment is needed in these frames before and after watermarking since the mask's maximum eigenvector of the object is just parallel to the x-axis. This phenomenon also means that the quality is degraded about 5 dB after the VO is rotated.

Figure 18 shows the original object and the watermarked object. The number of frames that are correctly authenticated is shown in Table 4. From the table, we can claim that the proposed system is robust to acceptable video processing. However, the results obtained from the video sequences Bream and Dajun are not as good as the results acquired from the video sequences Akiyo and Attacked Akiyo 1, especially in scaling processing. This is because of relatively small size of the object in some frames in video sequences Bream and Dajun. To embed a 162-bit long watermark, even small object has to be expanded to a certain size using zero padding. As we have mentioned previously, this will reduce correct ratio of watermark detection. Moreover, small object will lose more information than large object when the object is scaled; this will also reduce correct ratio of watermark detection. Thus, compared with other types of video processing, watermark detection is much more sensitive to object size in scaling processing. Study of the relationship between the object size and watermarking will be our future work.

We use video Attacked Akiyo 1 to evaluate whether the system can detect object replacement. During signing, the watermark generated from video Akiyo is embedded into the object of video Attacked Akiyo 1. During verification, the system can detect that all the frames are unauthentic. So the proposed system can detect object replacement.

To evaluate the system's security (i.e., to detect whether the object and background belong to the same frame or integrity protection), we combined the object extracted from the embedded video Akiyo and the background taken from our laboratory to form a new video sequence called Attacked Akiyo 2, as shown in Figure 19. The verification results also show that the system performs well in protecting the integrity between the object and its associated background.

Finally, we modify the embedded video Akiyo to form a new video sequence "Attacked Akiyo 3," shown in Figure 20, to evaluate whether the system can detect object modification. For all the 300 frames in video sequence, 288 frames are detected as faked frames. But there exists a limitation

Table 4: System performance when the video object undergoes various video processing (the number of video frames that are correctly authenticated for a video containing 300 frames).

Video object	Resizing (scaling factor is 0.5–1.0)	Rotation (0°–30°)	MPEG4 coding	Repadding
Akiyo	300	300	300	300
Bream	261	297	297	300
Dajun	274	293	300	300

Figure 19: The background of the Akiyo has been replaced. This video is used to evaluate the system's security (integrity protection). (Attacked Akiyo 2.)

Figure 20: A flower is put on the embedded object. This is used to evaluate whether the system can detect object modification. (Attacked Akiyo 3.)

in this system. As we have mentioned before, any malicious modifications, which may cause the maximum Hamming distance between the original VO and modified VO to be less than the feature ECC error-correcting capability (3 bits in our system), will be regarded as acceptable video processing. Therefore, a careful selection of feature and ECC scheme is the key work in our system.

6. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed a new object-based video authentication system, where ART coefficients of the VO are used as the feature to represent the VO. A content-based watermark was then generated by using ECC schemes and cryptographic hashing to increase the robustness and security of the system. The watermark was embedded in a set of randomly selected DFT coefficient groups that locate in the low-middle frequency area. Experimental results further demonstrated that the proposed video authentication system is robust during MPEG4 compression and normal VO processing such as scaling, rotation, as well as segmentation and mask errors. The results also showed that this system can protect the integrity of video (object and background).

In this paper, we assume that we only protect those objects whose sizes are larger than a preset threshold. This assumption is practical in some applications such as surveillance. Although rigorous evaluation will involve extensive work in collecting testing videos and especially in generating attacked videos, we still plan to conduct it in two ways: one is to fine-tune the parameter setting and perform the evaluation by real applications; the other is to continue the theoretic study among watermarking, ECC, and cryptography techniques to derive an adaptation framework for robust and secure content-based authentication.

REFERENCES

[1] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, Wiley, New York, NY, USA, 2nd edition, 1996.

[2] F. Bartolini, A. Tefas, M. Barni, and I. Pitas, "Image authentication techniques for surveillance applications," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1403–1418, 2001.

[3] A. Piva, R. Caldelli, and A. De Rosa, "A DWT-based object watermarking system for MPEG-4 video streams," in *Proc. IEEE International Conference on Image Processing (ICIP '00)*, vol. 3, pp. 5–8, Vancouver, BC, Canada, September 2000.

[4] C.-Y. Lin and S.-F. Chang, "Robust image authentication method surviving JPEG lossy compression," in *Storage and Retrieval for Image and Video Databases VI*, I. K. Sethi and R. C. Jain, Eds., vol. 3312 of *Proc. SPIE*, pp. 296–307, San Jose, Calif, USA, December 1997.

[5] J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," in *Proc. IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pp. 209–213, Florence, Italy, July 1999.

[6] M. P. Queluz, "Towards robust, content based techniques for image authentication," in *Proc. IEEE 2nd Workshop on Multimedia Signal Processing*, pp. 297–302, Redondo Beach, Calif, USA, December 1998.

[7] N. V. Boulgouris, F. D. Koravos, and M. G. Strintzis, "Self-synchronizing watermark detection for MPEG-4 objects," in *Proc. IEEE 8th Conference on Electronics, Circuits and Systems (ICECS '01)*, vol. 3, pp. 1371–1374, Masida, Malta, September 2001.

[8] C.-S. Lu and H.-Y. M. Liao, "Video object-based watermarking: a rotation and flipping resilient scheme," in *Proc. IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 483–486, Thessaloniki, Greece, October 2001.

[9] W.-N. Lie, G.-S. Lin, and T.-C. Wang, "Digital watermarking for object-based compressed video," in *Proc. IEEE Int. Symp.*

Circuits and Systems (ISCAS '01), vol. 2, pp. 49–52, Sydney, NSW, Australia, May 2001.

[10] P. Bas and B. Macq, "A new video-object watermarking scheme robust to object manipulation," in *Proc. IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 526–529, Thessaloniki, Greece, October 2001.

[11] M. J. B. Robshaw, "MD2, MD4, MD5, SHA and other hash functions," Tech. Rep. TR-101, version 4.0, RSA Laboratories, Redwood City, Calif, USA, July 1995.

[12] D. He, Q. Sun, and Q. Tian, "An object based watermarking solution for MPEG4 video authentication," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '03)*, vol. 3, pp. III-537–III-540, Hong Kong, China, April 2003.

[13] D. He, Q. Sun, and Q. Tian, "A semi-fragile object based video authentication system," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '03)*, vol. 3, pp. III-814–III-817, Bangkok, Thailand, May 2003.

[14] C. Lee, *Convolutional Coding: Fundamentals and Applications*, Artech House, London, UK, 1997.

[15] W.-Y. Kim and Y.-S. Kim, "A new region-based shape descriptor," in *ISO/IEC MPEG99/M5472*, Maui, Hawaii, December 1999.

[16] M. Bober, "MPEG-7 visual shape descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716–719, 2001.

[17] W. W. Peterson and E. J. Weldon Jr., *Error-Correcting Codes*, MIT Press, Cambridge, Mass, USA, 1984.

[18] Q. Sun, S.-F. Chang, K. Maeno, and M. Suto, "A new semi-fragile image authentication framework combining ECC and PKI infrastructures," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '02)*, vol. 2, pp. II-440–II-443, Phoenix-Scottsdale, Ariz, USA, May 2002.

[19] J. J. K. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing*, vol. 66, no. 3, pp. 303–317, 1998.

[20] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui, "Rotation, scale, and translation resilient Public watermarking for images," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, 2001.

[21] V. Solachidis and I. Pitas, "Circularly symmetric watermark embedding in 2-D DFT domain," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, vol. 6, pp. 3469–3472, Phoenix, Ariz, USA, March 1999.

[22] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 755–766, 2001.

Dajun He received the B.S. degree from Tsinghua University, Beijing, China, in 1991 and the M.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1994, both in electrical engineering. From 1994 to 1995, he was a Lecturer in Shanghai Jiaotong University, where he developed the HDTV simulation system for the Chinese Broadcasting Ministry. From 1996 to 2001, he was a Senior Engineer in AIWA, Singapore, in charge of developing audio and visual consumer products. He joined Institute for Infocomm Research (I^2R), Singapore, in 2001, as an Associate Scientist. His main research interests include image/video processing, compression, and security.

Qibin Sun received his Ph.D. degree in electrical engineering from University of Science and Technology of China, Anhui, China, in 1997. Since 1996, he has been with the Institute for Infocomm Research, Singapore, where he is responsible for industrial as well as academic research projects in the area of face recognition, media security, and image and video analysis. He worked in Columbia University 2000–2001, as a Research Scientist.

Qi Tian is a Principal Scientist at the Media Division, Institute for Infocomm Research, Singapore. His main research interests are image/video/audio analysis, indexing, retrieval, media content identification and security, computer vision, and pattern recognition. He has B.S. and M.S. degrees from Tsinghua University, China, and a Ph.D. degree from the University of South Carolina, USA, all in electrical and computer engineering. He joined the Institute of System Science at the National University of Singapore, in 1992. Since then he has been working on robust character ID recognition and video indexing; a number of state-of-the-art technologies have been developed, including the first container number recognition system (CNRS) in the world, developed and deployed in the Port of Singapore Authority, one of the busiest ports in the world. He was the Program Director for the Media Engineering Program at the Kent Ridge Digital Labs, then Laboratories for Information Technology from 2001 to 2002. He is a Senior IEEE Member, and has served and serves on editorial boards of professional journals, and as Chair and member of technical committees of the IEEE Pacific-Rim Conference on Multimedia (PCM2001, PCM2002, and PCM2003), the IEEE International Conference on Multimedia and Expo (ICME'2002, 2003), and the International Multimedia Modeling Conference 2004.