A Novel Scheme for Hybrid Digital Video Watermarking: Approach, Evaluation and Experimentation

Pik Wah Chan, *Student Member, IEEE*, Michael R. Lyu, *Fellow, IEEE*, and Roland T. Chin

***Abstract*—We have seen an explosion of data exchange in the Internet and the extensive use of digital media. Consequently, digital data owners can quickly and massively transfer multimedia documents across the Internet. This leads to wide interest in multimedia security and multimedia copyright protection. We propose a novel hybrid digital video watermarking scheme based on the scene change analysis and error correction code. Our video watermarking algorithm is robust against the attacks of frame dropping, averaging and statistical analysis, which were not solved effectively in the past. We start with a complete survey of current watermarking technologies, and noticed that none of the existing schemes is capable of resisting all attacks. Accordingly, we propose the idea of embedding different parts of a single watermark into different scenes of a video. We then analyze the strengths of different watermarking schemes, and apply a hybrid approach to form a super watermarking scheme that can resist most of the attacks. To increase the robustness of the scheme, the watermark is refined by an error correcting code, while the correcting code is embedded as a watermark in the audio channel. It optimizes the quality of the watermarked video. The effectiveness of this scheme is verified through a series of experiments, in which a number of standard image processing attacks are conducted, and the robustness of our approach is demonstrated using the criteria of the latest StirMark test.**

***Index Terms*—Digital watermarking, discrete wavelet transform (DWT), hybrid, scene change, video.**

I. Introduction

With the rapid growth of the Internet and multimedia systems in distributed environments, it is easier for digital data owners to transfer multimedia documents across the Internet. Therefore, there is an increase in concern over copyright protection of digital contents [1]–[4]. Traditionally, encryption and control access techniques were employed to protect the ownership of media. These techniques, however, do not protect against unauthorized copying after the media have been successfully transmitted and decrypted. Recently, watermark techniques are utilized to maintain the copyright [4]–[7]. In this paper, we focus on engaging the digital watermarking techniques to protect digital multimedia intellectual copyright, and propose a new algorithm particularly for video watermarking purpose.

Manuscript received March 11, 2004; revised June 10, 2004, September 10, 2004, and February 24, 2005. This work was supported by RGC Project No. CUHK4182/03E and UGC Project No. AoE/E-01/99 of the Hong Kong Special Administrative Region, China

P. W. Chan and M. R. Lyu are with the Computer Science and Engineering Department, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: pwchan@cse.cuhk.edu.hk; lyu@cse.cuhk.edu.hk).

R. T. Chin is with the Computer Science Department, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (e-mail: roland@cs.ust.hk).

Digital Object Identifier 10.1109/TCSVT.2005.856932

We have performed a complete survey on the current watermarking technologies. It is noticed that none of the current watermarking schemes can resist all attacks. With this finding, we propose a hybrid watermarking scheme based on scene change analyze and error correction codes [8].

Video watermarking introduces a number of issues not present in image watermarking. Due to a large amount of data and inherent redundancies between frames, video signals are highly susceptible to piracy attacks, including frame averaging, frame dropping, frame swapping, statistical analysis, etc [4]. However, the currently proposed algorithms do not solve these problems effectively. We attack this problem by applying scene change detections and scrambled watermarks in a video. The scheme is robust against frame dropping, as the same part of the watermark is embedded into the frames of a scene. For different scenes, different parts of the watermark are used, making the scheme robust against frame averaging and statistical analysis [8]. To increase the robustness of the scheme, we propose several hybrid approaches. The first one is visual–audio hybrid watermarking scheme. As videos consist of both video and audio channels, the robustness of our scheme can be enhanced by including an audio watermark. Consequently, we embed error correcting codes of a video watermark as an audio watermark, which can refine the retrieved watermark during watermark detection. The second approach is another hybrid with different watermarking schemes. As no existing scheme is resistant against all attacks, we employ the hybrid scheme to embed different parts of a watermark into different scenes. There are different ways to embed the watermarks, and the details will be described in the following sections.

Our approach cultivates an innovative idea in embedding different parts of a watermark according to scene changes, in embedding its error correcting codes as an audio watermark, and in applying a hybrid approach to the proposed scheme. This approach is never explored in the literature, and its advantages are clear and significant. The effectiveness of this scheme is verified through a number of experiments.

This paper is organized into six sections. The next section surveys the related work of current watermarking technologies. Section III describes the details of the novel scene-based video watermark scheme. Section IV states the possible improvement for the proposed watermarking scheme, the hybrid approach. The experimental results are shown in Section V. Section VI presents a conclusion and the future work.

II. Related Work

As a method of intellectual property protection, digital watermarks have recently stimulated significant interest and become

TABLE I
COMPARISON BETWEEN DIFFERENT WATERMARKING SCHEMES. (A) LSB. (B) THRESHOLD-BASED CORRELATION. (C) M-SEQUENCE/M-FRAME. (D) SPREAD SPECTRUM. (E) MID-BAND DCT. (F) MID-BAND DWT. (G) DFT TEMPLATE MATCHING. (H) RADON TRANSFORM

Attack Class	LSB	Threshold - based Correlation	m-sequence / m-frame	Spread Spectrum
JPEG Lossy Compression	0	0.75	0.7	0.85
PSNR	0	0.82	0.89	0.9
Add Noise	0	0.7	0.75	0.89
Median Filter	0	0.4	0.4	0.35
Row / Column Removal	0.5	0.63	0.7	0.69
Cropping	0.62	0.65	0.75	0.78
Rescale	0	0.5	0.62	0.83
Rotation	0	0.52	0.61	0.85
Affine	0	0.46	0.56	0.76
Geometrical Distortions	0	0.42	0.5	0.62
Shearing	0	0.3	0.54	0.85

Attack Class	Mid-band DCT	Mid-band DWT	DFT template Matching	Radon Transform
JPEG Lossy Compression	1	0.75	0.74	0.83
PSNR	0.98	1	0.81	0.78
Add Noise	0.95	0.73	0.86	0.75
Median Filter	0.4	0.3	0.25	0.3
Row / Column Removal	0.65	0.5	1	0.75
Cropping	0.62	0.76	0.89	0.85
Rescale	0.53	0.75	0.78	1
Rotation	0.5	0.52	1	0.98
Affine	0.35	0.45	0.98	0.83
Geometrical Distortions	0.64	0.75	0	0.75
Shearing	0.35	0.42	1	0.6

a very active area of research. A variety of imperceptible watermarking schemes have been proposed over the past few years. In general, watermarking schemes can be roughly divided into two categories: spatial domain watermark, and transformed domain watermark. We have chosen some representative watermarking schemes in each category for implementation and performed experiments to compare their robustness. They are: least significant bit (LSB) based watermarking scheme [9]; threshold-based correlation watermarking scheme [10]; direct sequence watermark using m-frame [11]; discrete Fourier transform (DFT) with template matching [12]; discrete wavelet transform (DWT) based watermarking scheme [13]; discrete cosine transform (DCT) based watermarking scheme [14] and spread spectrum [15] watermarking scheme. To evaluate the algorithms, the StirMark 4.0 benchmark program [18], [19] and 30 different images are used. Each attack is considered by itself and it is applicable after watermarking. For each image, we assign a score of 1 if the watermark is correctly decoded in the case. A value of zero is assigned if the watermark is incorrect. The comparison is shown in Table I.

From the result, the frequency domain watermarking schemes are relatively more robust than the spatial domain watermarking schemes, particularly in lossy compression, noise addition, pixel removal, rescaling, rotation and shearing. DCT-based watermarking scheme is the most robust to lossy compression. Moreover, DWT-based watermarking scheme is the most robust to noise addition. DFT-based watermarking scheme with template matching can resist a number of attacks, including pixel removal, rotation and shearing. Radon transformation resists attacks by rescaling and geometric distortion. The weakness of the existing algorithms, however, includes the following. 1) The watermark is not robust to attacks which are specifically targeted at to videos, such as frame dropping, averaging and statistical analysis; 2) The bit rate of the watermark is low. Some algorithms embed only one bit information as the watermark. 3) Existing techniques are not aware of the usefulness of the audio channel in a video. 4) None of the existing watermarking schemes resists to all the attacks. 5) A frequency domain watermark is more robust than a spatial domain watermark. To tackle these problems, in this paper, we propose a novel watermarking scheme based on scene changes with a hybrid approach.

III. SCENE-BASED VIDEO WATERMARKING SCHEME

The new watermarking scheme we propose is based on scene changes in [8]. Fig. 1 shows an overview of our watermarking process. In our scheme, a video is taken as the input, and then a watermark is decomposed into different parts which are embedded in corresponding frames of different scenes in the original video.

As applying a fixed image watermark to each frame in the video leads to the problem of maintaining statistical and perceptual invisibility [20], our scheme employs independent watermarks for successive but different scenes. However, applying independent watermarks to each frame also presents a problem if regions in each video frame remain little or no motion frame after frame. These motionless regions may be statistically compared or averaged to remove the independent watermarks [16],

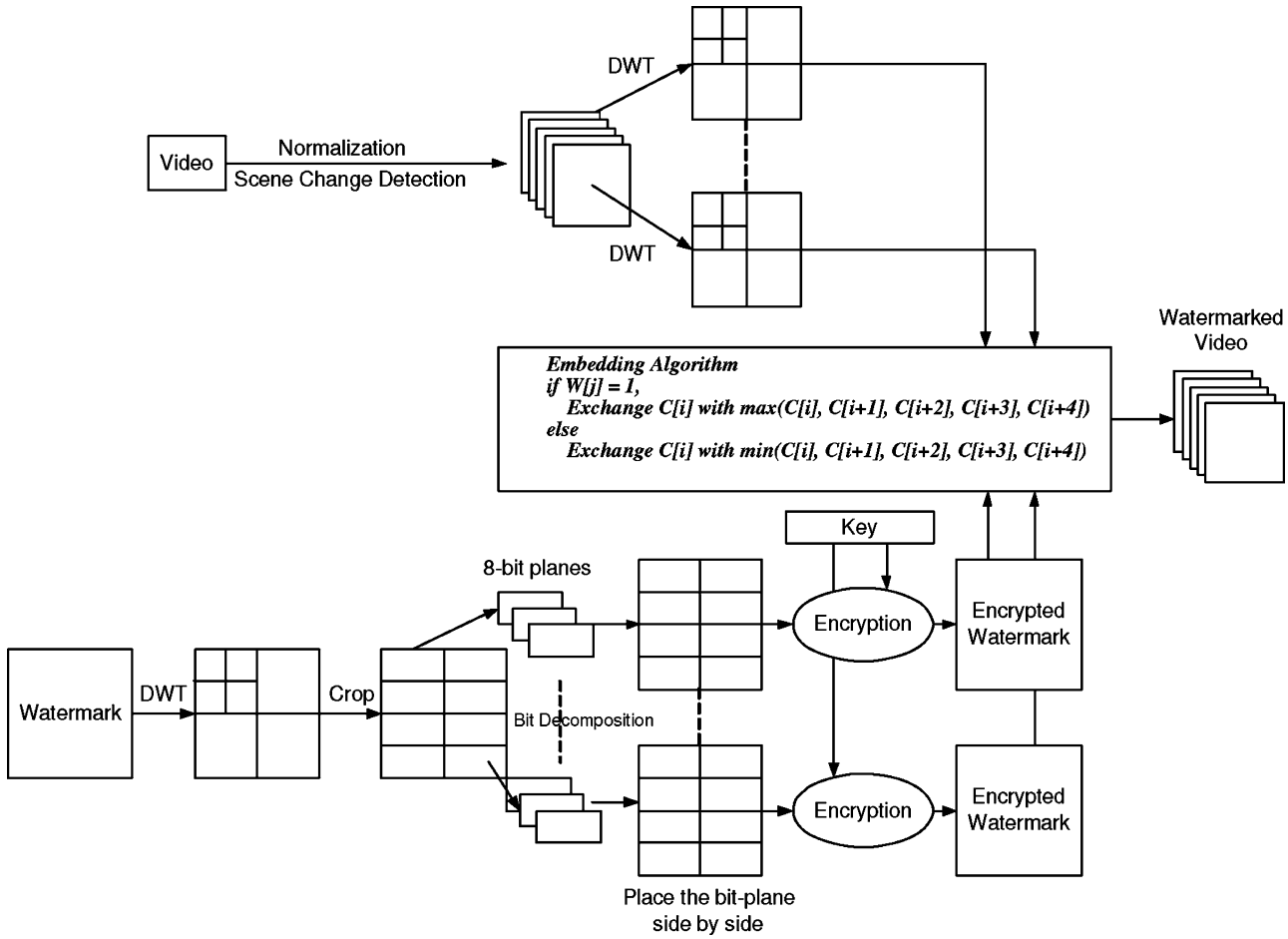


Fig. 1. Overview of the watermarking process.

[21]. Consequently, we use an identical watermark within each motionless scene. With these mechanisms, the proposed method is robust against the attacks of frame dropping, averaging, swapping, and statistical analysis. This newly proposed scheme consists of four parts, including: watermark preprocess, video preprocess, watermark embedding, and watermark detection. Details are described in the following sections.

A. Watermark Preprocess

A watermark is scrambled into small parts in a preprocess, and they are embedded into different scenes so that the scheme can resist a number of attacks toward to the video. A 256-grey-level image is used as the watermark, so 8 bits can represent each pixel. The watermark is first scaled to a particular size as follows:

$$2^n \leq m;\ n > 0 \tag{1}$$

$$p + q = n;\ p, q > 0 \tag{2}$$

where m is the number of scene changes and n, p, q are positive integers. The size of the watermark is represented as

$$64 \cdot 2^p \times 64 \cdot 2^q \tag{3}$$

Then the watermark is divided into 2^n small images with size 64×64. In the next step, each small image is decomposed into 8 bit-planes, and a large image m_n can be obtained by placing the bit-planes side by side only consisting of 0s and 1s. These processed images are used as watermarks, and totally 2^n independent watermarks are obtained. To make the scheme more robust, the processed watermarks m are transformed to the wavelet domain and encrypted [22]. Sample preprocessed watermarks are shown in Fig. 2, where (a) is the original watermark, (b)–(i) represent the scrambled watermarks in the spatial domain, and (j) shows the encrypted watermark of (b), i.e., m'_0.

B. Video Preprocess

Our watermark scheme is based on 4-level DWT. All frames in the video are transformed to the wavelet domain. The frames are decomposed in 4-level subband frames by separable two-dimensional (2-D) wavelet transform. It produces a low-frequency subband LL_4, and three series of high-frequency subbands LH_j, HL_j, HH_j, where $j < 4$. According to the energy distribution, LL_4 is the most important then LH_j, HL_j, and HH_j. For different levels, the higher the level, the more important the subbands. In our scheme, we only embed the watermark in the middle frequency subbands. Our scheme is based on 4-levels DWT, which is determined by experiments. If less than 4-levels is applied, the capacity of the scheme would be decreased; if larger than 4-levels is applied, the quality of the watermark video is affected.

Moreover, scene changes are detected from the video by applying the histogram difference method on the video stream. If

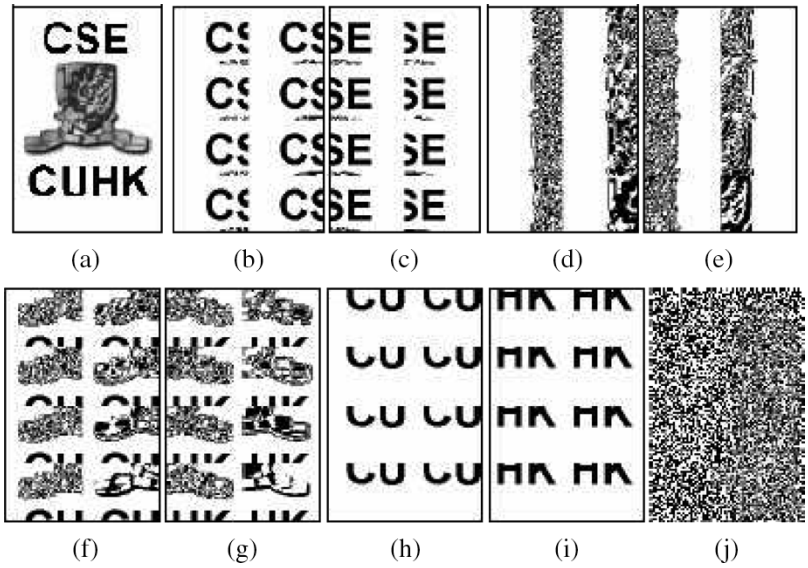


Fig. 2. (a) Original watermark. (b)-(i) Preprocessed watermark $m_0 - m_7$. (j) Encrypted watermark m_0'.

the difference of the two scenes is greater the threshold, we consider there is a scene change. The threshold is again determined by experiments.

Independent watermarks are embedded in frames of different scenes. Within a motionless scene, an identical watermark is used for each frame. Watermark m_1 is used for the first scene. When there is a scene change, another watermark m_3 is used for the next scene. The watermark for each scene can be chosen with a pseudo-random permutation such that only a legitimate watermark detector can reassemble the original watermark.

C. Watermark Embedding

The watermark is then embedded to the video frames by changing position of some DWT coefficients with the following condition:

if $W_j = 1$ **then**
  Exchange $\max(C_i, C_{i+1}, C_{i+2}, C_{i+3}, C_{i+4})$
else
  Exchange $\min(C_i, C_{i+1}, C_{i+2}, C_{i+3}, C_{i+4})$
end if

where C_i is the ith DWT coefficient of a video frame, and W_j is the jth pixel of a corresponding watermark image [23]. When the watermark $W_j = 1$, we perform an exchange of the C_i with the maximum value among C_i, C_{i+1}, C_{i+2}, C_{i+3}, C_{i+4}. When $W_j = 0$, we perform an exchange of the C_i with the minimum value among C_i, C_{i+1}, C_{i+2}, C_{i+3}, C_{i+4}. With this algorithm, the retrieval of the embedded watermark does not need the original image. The higher frequency coefficients of the watermark are embedded to higher frequency parts of the video frame, and only the middle frequency wavelet coefficient of the frame (middle frequency subband) is watermarked [16].

D. Watermark Detection

The video is processed to detect the video watermark. In this step, scene changes are detected from the tested video. Also, each video frame is transformed to the wavelet domain with 4-levels. Then the watermark is extracted with the following condition: where WC_i is the ith DWT coefficient of a watermarked video frame, and EW_j is the jth pixel of the extracted watermark [17]. When the watermark WC_j is greater than median value among WC_i, WC_{i+1}, WC_{i+2}, WC_{i+3}, WC_{i+4}, the extracted watermark is considered as one, i.e., $\mathrm{EW}_j = 1$; otherwise, it is considered as zero, i.e., $\mathrm{EW}_j = 0$. With this algorithm, the retrieval of the embedded watermark does not need the original image. This is an important property to video watermarking.

if
  $\mathrm{WC}_i > \mathtt{median}(\mathrm{WC}_i, \mathrm{WC}_{i+1}, \mathrm{WC}_{i+2}, \mathrm{WC}_{i+3}, \mathrm{WC}_{i+4})$
then
  $\mathrm{EW}_j = 1$
else
  $\mathrm{EW}_j = 0$
end if

As an identical watermark is used for all frames within a scene, multiple copies of each part of the watermark may be obtained. The watermark is recovered by averaging the watermarks extracted from different frames. This reduces the effect if the attack is carried out at some designated frames. Thus, we can combine the 8 bit-planes and recover the 64×64 size image, i.e., $1/2^n$ part of the original watermark.

If enough scenes are found and all parts of the watermark are collected, the original large watermark image can be reconstructed. This can be shown in Fig. 3, where the original frame, the watermarked frame, and the extracted watermark are depicted.

IV. Hybrid Watermarking Schemes

In the previous section, a novel scene-based watermarking scheme is proposed, which is resistant against the attacks of the video properties, including frame averaging, frame dropping,

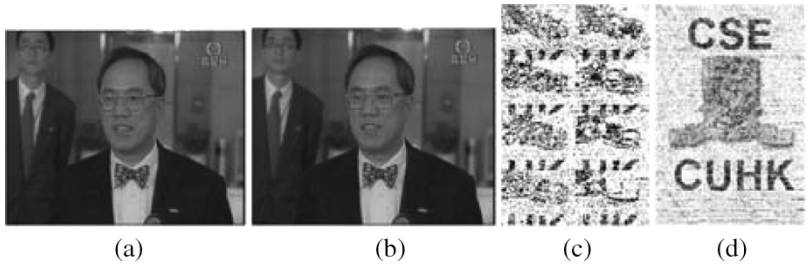


Fig. 3. (a) Original frame. (b) Watermarked frame. (c) Extracted watermark corresponding to Fig. 2(g). (d) Recovered watermark.

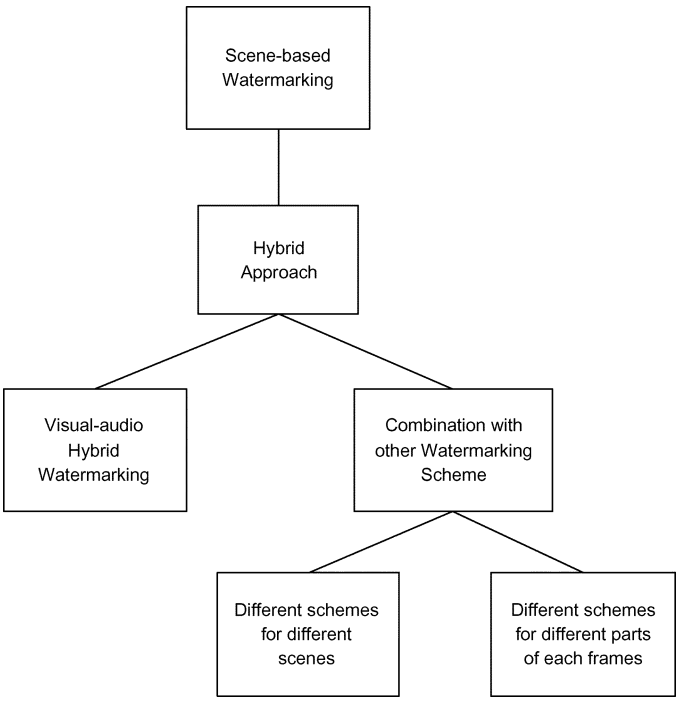


Fig. 4. Possible improvement for scene based watermarking scheme.

and statistical analysis. However, the scheme dose not improve the robustness against the attacks by image processing on the video frames. Therefore, we propose a hybrid approach to improve the performance and the robustness of the watermarking scheme based on the conclusion drawn from the survey and the properties of a video.

The scene-based watermarking scheme can be improved with two type of hybrid approaches; visual–audio hybrid watermarking and hybrid with different watermarking schemes. Fig. 4 shows the overall framework of the proposed scheme.

The visual–audio hybrid watermarking scheme applies both video and audio watermarks in a video. Error correcting codes are extracted from the video watermark and embedded as audio watermark in the audio stream. This approach takes the advantage of watermarking the audio channel, because it provides an independent means for embedding the error correcting codes, which carry extra information for watermark extraction. Therefore, the scheme is more robust than other schemes which only use video channel alone. The hybrid with different watermarking schemes can further be divided into to two classes: independent scheme and dependent scheme. From the survey, we find that no watermarking scheme can resist to all watermark attacks; hybrid with different watermarking schemes can be one of the solutions. It takes advantages of various watermarking schemes by combining them in different ways.

A. Visual–Audio Hybrid Watermarking

The visual audio watermarking scheme combines a video watermark and an audio watermark. We embed error correcting codes of a video watermark as an audio watermark and refine the retrieved video watermark during detection [8]. Fig. 5 shows an overview of our visual–audio watermarking process. In our scheme, an input video is split into audio and video streams, which undergo separate watermarking procedures. On the one hand, a video watermark is decomposed into various parts, embedded in corresponding frames of different scenes in the original video. On the other hand, error correcting codes are extracted from the watermarks and embedded as an audio watermark in the audio channel, which in turn makes it possible to correct and detect the changes from the extracted video watermarks. This additional protection mechanism enables our scheme to overcome the corruption of a video watermark, thus the robustness of the scheme is preserved under certain attacks.

1) Audio Watermark: The watermark embedded in the audio channel provides the error correction and detection capability for the video watermark. In the detection phase, it would be extracted and used for refining the video watermark. Disparate error correction coding techniques can be applied, such as Reed–Solomon coding techniques [21] and Turbo coding [24].

Error correcting codes play an important role in watermarking, especially when the watermark is damaged significantly. Error correcting codes overcome the corruption of a watermark, and make the watermark survive through serious attacks. Moreover, our scheme benefits from audio watermarking as it provides an independent channel for embedding the error correcting codes, which carry extra information for video watermark extraction.

The key to error correction is redundancy. The simplest error correcting code is repeating everything several times. However, in order to keep the audio watermark inaudible, we cannot embed too much information into an audio channel. In our scheme, we apply averaging to obtain the error correction code. Within a small region of an image, the pixels are similar. Hence, an average value of a small region can be fully utilized

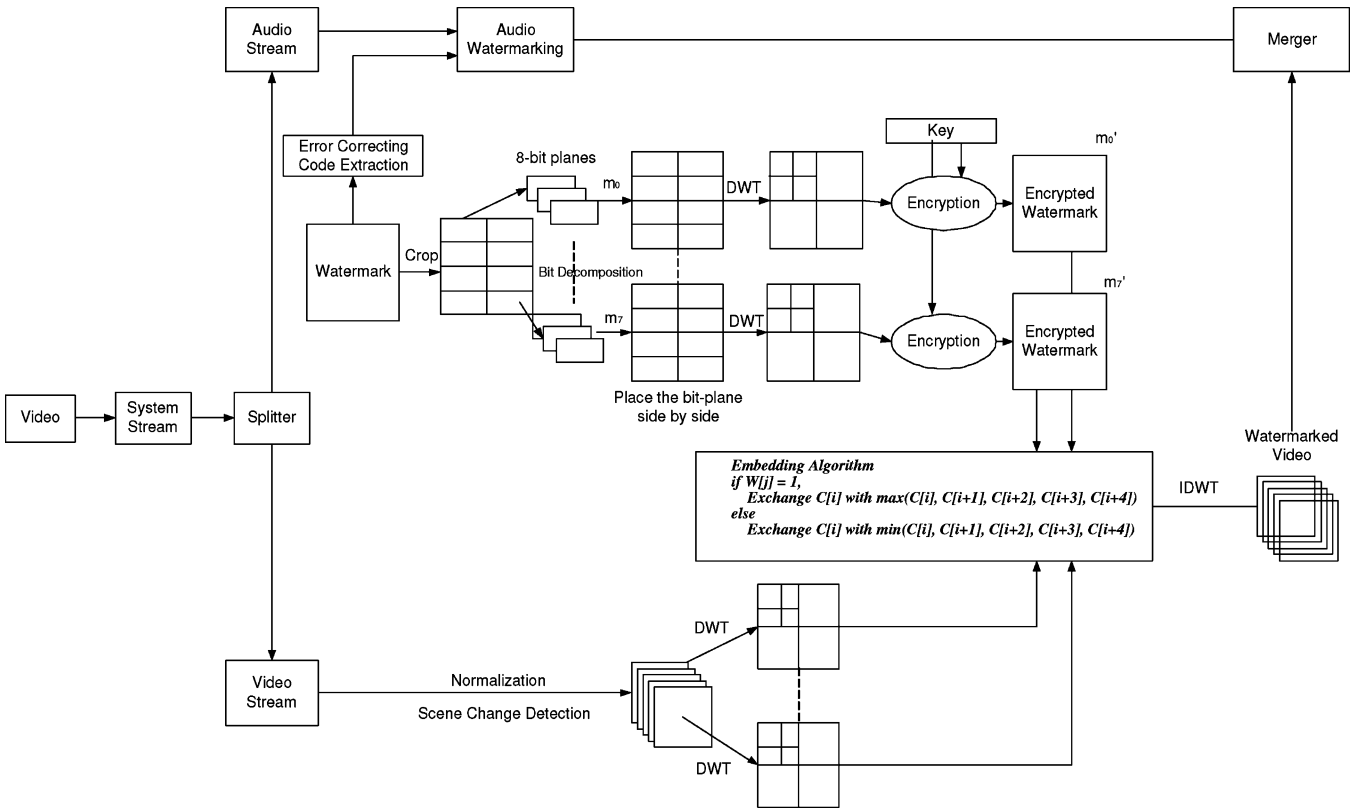
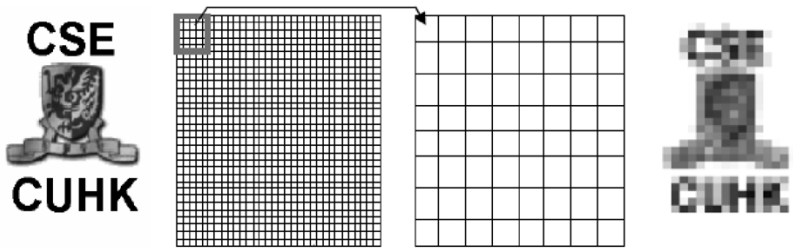


Fig. 5. Overview of visual–audio hybrid watermarking scheme.



(a) (b) (c)

Fig. 6. (a) Original video watermark. (b) Visualization of averaging. (c) Audio watermark (average of a).

to estimate the pixels within the particular region. The average value of the pixels in each region is calculated as (4)

$$\mathrm{Avg}_k = \frac{\sum_{i=0}^{x}\sum_{j=0}^{y} W(r\times x+i, s\times y+j)}{x\times y} \qquad (4)$$

where Avg_k is the average of the kth block, W is a pixel in the image(r, s) is coordinate of region $k, (i, j)$ is the coordinate of the pixel in region k, and $x \times y$ is the size of the block. A sample is shown in Fig. 6.

The audio watermarking is based on modulated complex lapped transform (MCLT) [25]. The MCLT is a $2\times$ oversampled DFT filter bank, used in conjunction with analysis and synthesis windows that provide perfect reconstruction. The MCLT is well suited for noise suppression and echo cancellation. For the typical 44.1 kHz sampling, we use a length-2048 MCLT. Only the MCLT coefficients within the 2–7 kHz subband are modified and considered in the detection process, to minimize carrier noise effects as well as sensitivity to downsampling and compression.

2) Synchronization Between the Video and Audio: Including synchronization in audio or video can provide security enhancement. Some attacks, such as cropping, cause change in the amount of data and misplacement of the data. In these cases, synchronization codes can help. To increase the robustness of our watermarking scheme against attacks such as cropping and rotation, a synchronization signal is embedded into both the video [28] and audio [29] channels with the watermark, thus, the chance of the watermark being recovered is increased.

a) Video Synchronization: The video watermark is synchronized based on profiled statistics [28]. The characteristic of the frames is extracted and sent as side information for synchronization. The video frames are transformed from 2-D model to a 3-D model. By summing the meshing function along the x- and y- directions of the 3-D mesh, we obtain the profiles of the

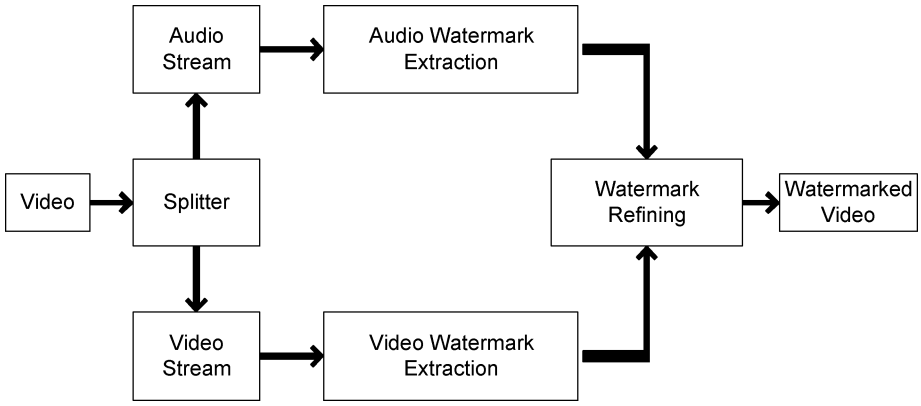


Fig. 7. Overview of detection of the watermark.

frames. The profiles of the image intensity $I(x,y)$ can be divided in two directions x and y as

$$HI(x) = \sum_{y=1}^{n} I(x,y) \qquad VI(y) = \sum_{x=1}^{m} I(x,y). \tag{5}$$

Then the characteristic vector of the orginal video frame C can be prsented as

$$C = \left[\mu_x \mu_y \sigma_x^2 \sigma_y^2\right] \tag{6}$$

where μ_x and μ_y are the means, σ_x^2 and σ_y^2 are the variances, along the x- and y- directions.

The parameter C_i represents the characteristics of frame i. The original video characteristics CV and the attacked video characteristics CV^a can be formed by the set of C_i and C_j^a, respectively. In the scene-based watermarking scheme, only sampled frames are selected to extract the characteristics as the synchronization code for each scene. As the same part of the watermark is embedded into a scene, the synchronization codes capable of identifying the correct positions of each scene are sufficient for extracting the watermark, since in our approach we only need to synchronize every scene, but not every frame. This improves the performance of the scheme and preserves the robustness of the synchronization code. By comparing the characteristics contained in CV and CV^a, the received frames can be resynchronized. The frame index resynchronization is determined by the nearest neighborhood rule in which the attacked frame should be fit to the most similar frame in the original video. The resynchronized index K^a for the video is

$$K^a = \left\{k_j | k_j = \max\nolimits_{(i\in 0,1,\ldots N-1)}^{-1} Z\left(C_i C_j^a\right), \forall j\right\} \tag{7}$$

where k_j is the resynchronized index for the jth frame and K^a is the set of the index sequence of k_j, N is the number of nearest neighborhood used, and $Z(.)$ is normalized correlation.

b) Audio synchronization: In addition to the error correcting code, synchronization code is embedded into the audio channel. Bark code with 12 bits is adopted as the synchronization code [29] embedded in time domain at the beginning of each block of the audio signal for performing MCLT. The auto-correction function of Bark code is defined as

$$C(t) = \sum_{j=1}^{L-1} a_j a_{j+1} \tag{8}$$

where L represents the code's length, a_j the jth element of the code, and $a_j \in \{+1, -1\}$. The Bark code "111 110 011 010" is chosen as the synchronization code to guaratee $|c(t)| \leq 1$ [29].

Then the synchronization code is embedded into the audio signal as follows:

$$A_i(j) = \begin{cases} A_i(j) - A_i(j) \bmod S + \frac{3S}{4}, & \text{if } B(j) = 1 \\ A_i(j) - A_i(j) \bmod S + \frac{S}{4}, & \text{if } B(j) = 0 \end{cases} \tag{9}$$

where $A_i(j)$ denotes the jth element in the block i of the audio signal, $B(j)$ denotes the synchronization code, S is the maximum value on the premise of inaudibility. $A_i(j) - A_i(j) \bmod S$ corresponds to the fact that the lowest bits of $\log_2 S$ are set to 0. After taking both robustness and constraint of inaudibility into our consideration, $\mathrm{S} = 8192$ is chosen to achieve the best performance in resisting the attacks by adding noises.

The detection of the synchronization code is based on the frame synchronization techniques in digital communication. We adopt the anti-decline frame synchronization scheme in our scheme. The input signal would be examined and searched for the synchronization code. The advantages of embedding the synchronization code in the time domain is the low cost for searching computation. To search synchronization code and establishing resynchronization, we first examine the entire audio signal and select those codes whose correlation functions in (8) are greater then a predefined threshold. Then, the code is selected and regarded as synchronization code if the distance of two adjacent codes meets the distance of division.

When extracting the watermark, the synchronization code is detected and the resynchronization is then established. According to the position of the synchronization code, the watermarked audio signal is split into a number of blocks with M samples for MCLT.

3) Watermark Detection: The watermark is detected through the process whose overview is shown in Fig. 7. A test

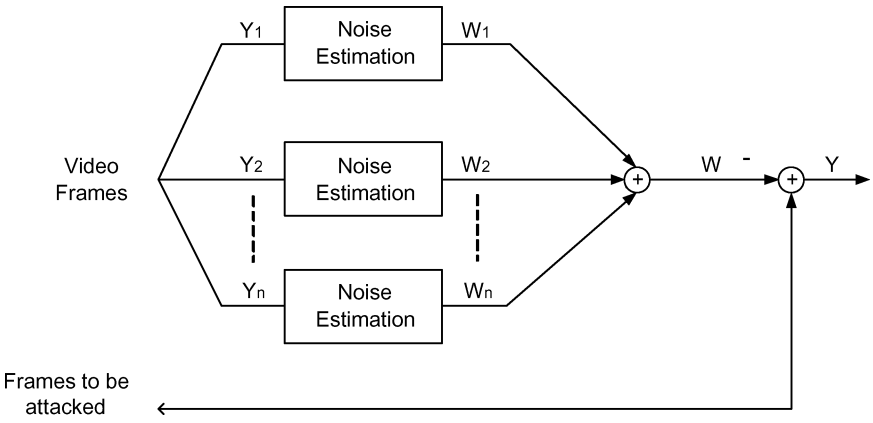


Fig. 8. Scenario of statistical averaging attack.

video is split into video stream and audio stream, and watermarks are extracted separately by audio watermark extraction and video watermark extraction. Then the extracted watermarks undergo a refining process.

The video stream is processed to get the video watermark. At the same time, error correcting codes are extracted from the audio stream and the video watermark extracted is refined by this information with the (10):

$$\mathrm{RW}_{ij} = \frac{\mathrm{EW}_{ij} \times f + \mathrm{Avg}_k \times g}{f + g} \tag{10}$$

where RW_{ij} is the refined watermark, EW_{ij} is the extracted video watermark from (7), Avg_k is the extracted audio watermark, k is the kth block of the average image, (i, j) is coordinate of the video watermark, and $f : g$ is a ratio of importance of the extracted video watermark to the audio watermark. In all the subsequent experiments, we assume $f = 0.5$ and $g = 0.5$, $f + g = 1$.

The important ratio can be adjusted by comparing the refined watermark with the original watermark. With different values of the important ratio, we can obtain different refined watermark. Then, we can calculate the normalized correlation (NC) values of the refined watermarks. The one with the highest NC values will be chosen to be the extracted watermark.

B. Hybrid Approach With Different Watermarking Schemes

No watermarking scheme is found in the current literature to be capable of resisting all watermark attacks. The hybrid approach can be a possible solution. As stated earlier, it can be classified into independent schemes and dependent schemes. Independent watermarking schemes include either different schemes for different scenes or different schemes for different parts of the frame. Dependent watermarking schemes embed a watermark in each frame with several different schemes.

We propose two approaches for the hybrid watermarking schemes. They combine alien schemes in disparate ways. Four watermarking schemes are chosen, each of which strives a different set of attacks. These four schemes are: DWT, DCT, DFT, and Radon Transform based watermarking (RADON). As they embed the watermark in various domains, their robustness properties are preserved. By combining the advantages of these watermarking schemes systematically, various kinds of attacks can be resisted altogether. In the paper, we propose two approaches to combine the employed watermarking schemes: Different schemes for different scenes, and different schemes for different parts of each frame.

a) Different schemes for different scenes: In this approach, a watermark is still decomposed into different parts which are embedded in the corresponding frames of different scenes in the original video. Each part of the watermark, however, is embedded with a different watermarking scheme. Within a scene, all the video frames are watermarked with the same part of a watermark by the same watermarking scheme.

When there is an attack on the watermarked video, different watermarking schemes are resistant against it. Consequently, some parts of the watermark still survive after the attack. This approach thus enhances the chance of survival under several attacks, and raises its robustness. The merit is that only one part of the watermark is damaged if the watermarked video is attacked, provided that at least one of the watermarking schemes is resistant against the attack. The disadvantage of this approach is that the accuracy of the extracted watermark is lower, compared with the other schemes specified to a particular attack.

b) Different schemes for different parts of each frame: This approach is similar to the previous approach. However, four different watermarking schemes are applied to each frame instead of different schemes for different scenes. Each video frame is divided into four parts, and the watermark for that frame is also divided into four parts. Then, each part of the watermark is embedded into the frame in different domains.

When a watermarked video is attacked, part of the watermark in each frame may still survive. Therefore, information for every part of the watermark can be retrieved, and the watermark can be approximately estimated. Although the accuracy of the extracted watermark is reduced, it is more resistant against attacks.

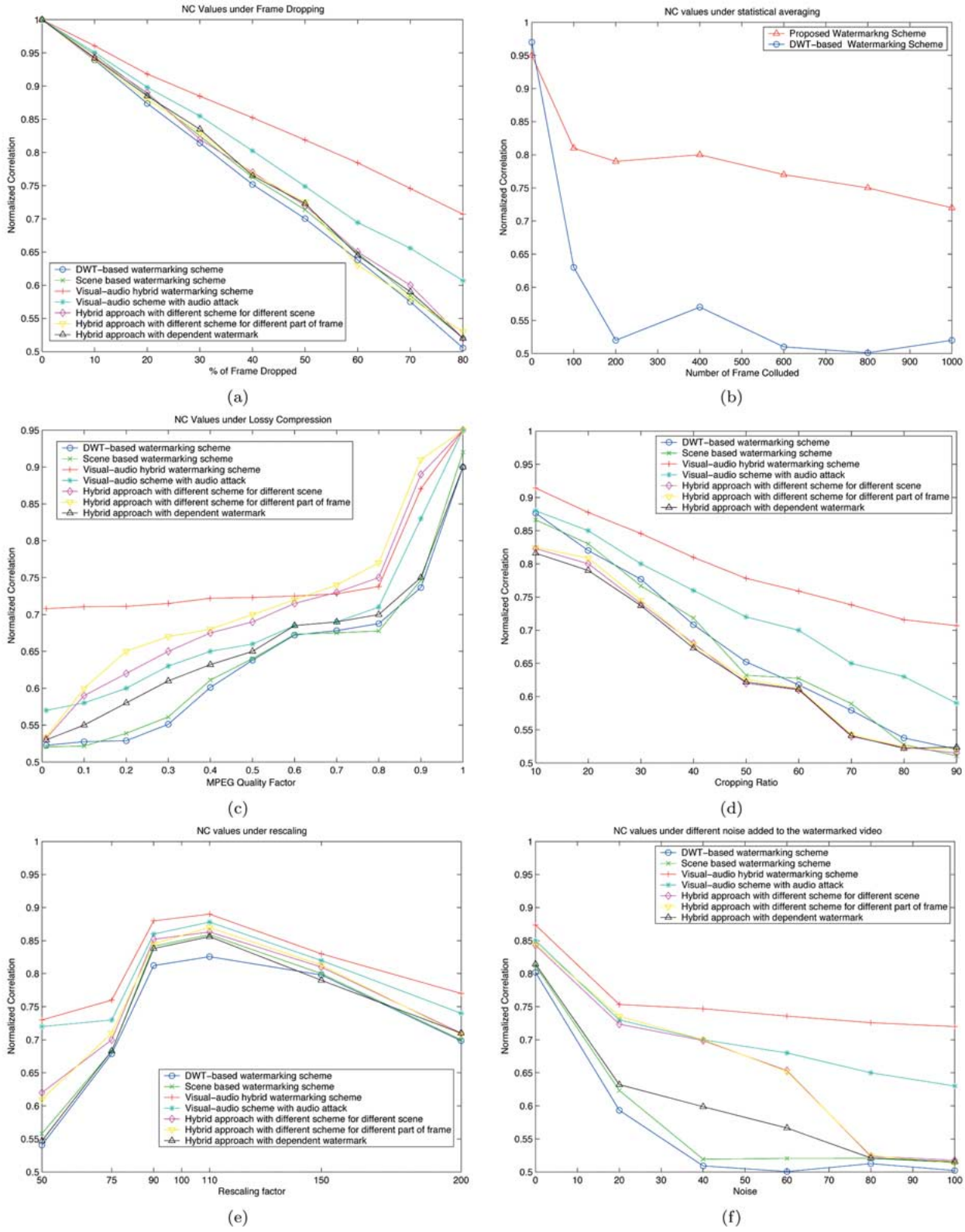


Fig. 9. Experiment results under different attacks.

V. Experimental Results

To implement the proposed watermarking scheme, the software VirtualDub[1] is employed. The performance of the new video watermarking scheme is evaluated through several experiments: the experiment with various dropping ratio, the experiment with various number of frame colluded, the experiment with various quality factor of MPEG, and the test of Robustness with StirMark 4.0. The audio channel is also attacked by adding some noises into it. Another DWT-based watermarking scheme

[1]VirtualDub is distributed under the GNU General Public License, written by Avery Lee. [Online] Available: http://www.virtualdub.org/

TABLE II
COMPARISON BETWEEN DIFFERENT WATERMARKING SCHEMES INCLUDING: (A) DWT-BASED; (B) SCENE-BASED; (C) VISUAL-AUDIO HYBRID; (D) VISUAL-AUDIO HYBRID WITH AUDIO ATTACK; (E) HYBRID APPROACH WITH DIFFERENT SCHEME FOR DIFFERENT SCENE; (F) HYBRID APPROACH WITH DIFFERENT SCHEME FOR DIFFERENT PART OF FRAME

Attack Class	DWT-based watermarking scheme	Scene-based watermarking scheme	Visual-audio hybrid watermarking scheme with audio attack
Lossy Compression	0.61	0.62	0.82
PSNR	0.80	0.76	0.86
Add Noise	0.63	0.60	0.76
Median Filter	0.54	0.54	0.74
Row / Column Removal	0.69	0.71	0.85
Cropping	0.68	0.66	0.78
Rescale	0.63	0.62	0.75
Rotation	0.60	0.61	0.73
Affine	0.55	0.55	0.78

Attack Class	Visual-audio hybrid watermarking scheme	Hybrid approach with different scheme for different scene	Hybrid approach with different scheme for different part of frame
Lossy Compression	0.69	0.71	0.72
PSNR	0.80	0.82	0.81
Add Noise	0.67	0.70	0.69
Median Filter	0.60	0.55	0.52
Row / Column Removal	0.75	0.77	0.78
Cropping	0.70	0.72	0.69
Rescale	0.69	0.71	0.68
Rotation	0.67	0.69	0.66
Affine	0.70	0.73	0.69

[31], which embeds an identical watermark in all frames, is implemented to compare with the proposed scheme. We use the cross-correlation normalized (NC) to measure the similarity of the extracted and the referenced watermarks to evaluate our scheme in the experiments.

A. Experiment With Frame Dropping

As a video contains a large amount of redundancies between frames, it may suffer attacks by frame dropping. This experiment is aimed at examining the robustness of the scheme under the frame dropping attack and the obtained results are shown in Fig. 9(a).

From the experiment, we find that the scheme achieves better performance than the DWT-based scheme without scene-based watermarks. It is because in each scene, all frames are embedded with the same watermark. It prevents the attackers from removing the watermark by frame dropping. If they try to remove one part of the watermark, they need to remove the whole trunk of frames (i.e., the whole scene), leading to a significant damage to the video. In addition, when the frames are dropped, the error is only introduced to a corresponding small part of the watermark. For the DWT-based scheme (i.e., nonscene-based), however, the error is introduced to the whole watermark, making the performance worse.

The performance of the scheme is significantly improved by combining with an audio watermark, the visual–audio watermarking scheme, as error correcting codes from the audio watermark provide information to correct the error and recover the corruption of the video watermark. Moreover, the error correcting codes are embedded in the audio channel. As frame dropping would not affect the audio channel much, our scheme benefits by allowing uninterrupted error correcting codes to refine the watermark.

When the error correcting codes in the audio channel are altered by the attack, the capability to recover the error in the video watermark is dropped. However, the result is still better than the scheme without an audio watermark, as the attacked audio watermark still contains some information to recover the watermark in the video channel.

B. Experiment With Frame Averaging and Statistical Analysis

Frame averaging and statistical analysis is another common attack to the video watermark. When attackers collect a number of watermarked frames, they can estimate the watermark by statistical averaging and remove it from the watermarked video [32], [33]. The scenario is shown in Fig. 8.

Firstly, noise estimation would be done on similar frames of the video. As watermark can be consider as noise in a frame. If the frames are similar, they can be compared and estimated the noise. After the noise is estimated, the watermark can be considered as the watermark. It will be compare with the frames to be attacked and remove the watermark in the video frames.

Experiments have been conducted to evaluate the proposed scheme under this attack, and the results are shown in Fig.(9b). It is found that the proposed scheme can resist to statistical averaging quite well. This is because our scheme crops a watermark into pieces and embeds them into different frames, making

the watermarks resistant to attacks by frame averaging for the watermark extraction. The identical watermark used within a scene can prevent attackers from taking the advantage of motionless regions in successive frames and removing the watermark by comparing and averaging the frames statistically [34]. On the other hand, independent watermarks used for successive, yet different scenes can prevent the attackers from colluding with frames from completely different scenes to extract the watermark.

C. Experiment With Lossy Compression

From the result Fig. 9(c), we note that the proposed scheme improves the robustness for watermark protection. The performance of the scheme is significantly improved by combining with audio watermark again, especially when the quality factor of MPEG is low. This is because when the quality factor of MPEG is low, the error of the extracted watermark is increased and the watermark is damaged significantly. As the error correcting codes are provided from the audio watermark, they are not affected by the lossy compression attack applied to the video channel. Consequently, the error correcting codes can overcome the corruption of the video watermark, achieving higher NC values.

The performance of the scheme is also improved by the hybrid approach with different watermarking schemes. From the survey, we find that the DCT-based watermarking scheme is the most resistant one against lossy compression. When compression is applied to the watermarked video, the watermark embedded in the video with DCT-based watermarking scheme survives. Therefore, at least one forth of the watermark can be retrieved from the video. This increases the robustness of the scheme.

D. Test of Robustness With Stirmark 4.0

We employ a benchmark StirMark 4.0 [18], [19] to test the robustness of the proposed schemes and the result are shown in Fig.9(d)–(f) and Table II.

Fig. 9(d) shows the result of the NC values of the watermark under different cropping ratios. The visual–audio watermarking scheme gives better performance. This shows that the audio watermark significantly improves the robustness of the watermarking scheme. However, the result shows that the performance of scheme is not improved by the hybrid approach with different watermarking schemes.

When the watermarked video is rescaled, the proposed scheme also portrays improvement. Fig. 9(e) depicts the NC values when the watermarked video is rescaled with various factors. The performance of the scheme is significantly improved again by visual–audio watermarking scheme, especially when the rescaling factor is large. Furthermore, the improvement becomes more evident with the increase of the rescaling factor.

With the hybrid approaches, the robustness of the scheme is increased. In wavelet domain, the coefficients vary when the size of frame or image is different. The coefficients in the Randon transformed domain, however, do not vary too much when rescaling. As shown in Fig. 9(e), the hybrid approaches with different schemes perform better than the scene-based watermarking scheme.

When additional noises are applied to the watermarked video, the proposed scheme also shows improvement. Fig. 9(f) depicts the NC values when different noises are added to the watermarked video. The performance of the scheme is significantly improved by combining with an audio watermark, especially when more noises are added.

There are several tests from the StirMark 4.0. The result is summarized in Table II. The proposed video watermarking scheme also shows improvement when the videos are under other attacks, including: row/filter removal, rotation, PSNR and affine.

From the above results, the effectiveness of the scene-based hybrid schemes are demonstrated. The scene-based watermarking scheme achieves higher NC values when attacks based on video properties are launched. This indicates that the watermarking scheme works well by applying scene change detection with scrambled watermarks. The performance of the scheme is further improved by combining with an audio watermark, especially when the video watermark is corrupted, such as the attack by lossy compression. When audio channel is also attacked, the error correction information is altered. The overall performance, however, still shows improvement. The robustness of the scheme is also raised by engaging other hybrid approaches.

VI. Conclusion and Future Work

This paper proposes an innovative scene-based hybrid video watermarking scheme. The process of this comprehensive video watermarking scheme, including watermark preprocessing, video preprocessing, watermark embedding, and watermark detection, is described in detail. Various improvement approaches are also presented. Experiments are conducted to demonstrate that our scheme is robust against attacks by frame dropping, frame averaging, and statistical analysis, and the robustness against the image processing attacks is tested with StirMark benchmark. Our scheme is verified to be resistant against attacks based on video characteristics and image processing techniques. It is particularly enhanced by combining with audio watermarks for error correction capabilities and the hybrid scheme for attack resisting. The effectiveness of this scheme is determined through a number of experiments. We conducted series of experiment to prove its effectiveness.

This proposed watermarking scheme can further be associated with different applications to achieve a sophisticated system and the the fidelity can be improved by applying genetic algorithm. This research can also be extended by applying the scheme to specific environments or applications and examine its effectiveness.

References

[1] A. Piva, F. Bartolini, and M. Barni, "Managing copyright in open networks," *IEEE Trans. Internet Computing*, vol. 6, no. 3, pp. 18–26, May–Jun. 2002.

[2] C. Lu, H. Yuan, and M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1579–1592, Oct. 2001.

[3] C. Lu, S. Huang, C. Sze, and H. Y. M. Liao, "Cocktail watermarking for digital image protection," *IEEE Trans. Multimedia*, vol. 2, no. 6, pp. 209–224, Dec. 2000.

[4] J. Lee and S. Jung, "A survey of watermarking techniques applied to multimedia," in *Proc. 2001 IEEE Int. Symp. Industrial Electronics (ISIE)*, vol. 1, 2001, pp. 272–277.

[5] M. Barni, F. Bartolini, R. Caldelli, A. De Rosa, and A. Piva, "A robust watermarking approach for raw video," presented at the 10th Int. Packet Video Workshop, Cagliari, Italy, May 1–2, 2000.

[6] F. Petitcolas, Ed., *Information Hiding Techniques for Steganography and Digital Watermarking Stefan Katzenbeisser*. Norwood, MA: Artech House, Dec. 1999.

[7] A. Eskicioglu and E. Delp, "An overview of multimedia content protection in consumer electronics devices," in *Proc. Signal Processing Image Communication 16 (2001)*, 2001, pp. 681–699.

[8] P. W. Chan and M. Lyu, "A DWT-based digital video watermarking scheme with error correcting code," in *Proc. 5th Int. Conf. Information and Communications Security (ICICS2003)*, vol. 2836. Huhehaote City, China, Oct. 10–13, 2003, pp. 202–213.

[9] N. Memon, "Analysis of LSB based image steganography techniques Chandramouli," in *Proc. 2001 Int. Conf. Image Processing*, vol. 3, Oct. 2001, pp. 1019–1022.

[10] G. Langelaar, I. Setyawan, and R. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Process. Mag.*, vol. 17, no. 9, pp. 20–43, Sep. 2000.

[11] B. Mobasseri, "Direct sequence watermarking of digital video using m-frames," in *Proc. 1998 Int. Conf. Image Processing*, vol. 2, Oct. 1998, pp. 399–403.

[12] S. Pereira and T. Pun, "Robust template matching for affine resistant image watermarks," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1123–1129, Jun. 2000.

[13] I. Hong, I. Kim, and S. Han, "A blind watermarking technique using wavelet transform," in *Proc. IEEE Int. Symp. Industrial Electronics*, vol. 3, 2001, pp. 1946–1950.

[14] F. Duan, I. King, L. Xu, and L. Chan, "Intra-block algorithm for digital watermarking," in *Proc. IEEE 14th Int. Conf. Pattern Recognition*, vol. 2, Aug. 1998, pp. 1589–1591.

[15] M. George, J. Chouinard, and N. Georganas, "Digital watermarking of images and video using direct sequence spread spectrum techniques," in *Proc. 1999 IEEE Canadian Conf. Electrical and Computer Engineering*, vol. 1, May 1999, pp. 116–121.

[16] M. Swanson, B. Zhu, and A. Tewfik, "Multiresolution video watermarking using perceptual models and scene segmentation," in *Proc. Int. Conf. Image Processing*, vol. 2, Washington, DC, Oct. 1997, pp. 558–561.

[17] J. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 612, pp. 1973–1987, Dec. 1997.

[18] F. Petitcolas and R. Anderson, "Evaluation of copyright marking systems," in *Proc. IEEE Multimedia Systems* , Florence, Italy, Jun. 1999, pp. 574–579.

[19] M. Kutter and F. Petitcolas, "A fair benchmark for image watermarking systems," in *Proc. Electronic Imaging '99, Security and Watermarking of Multimedia Contents*, vol. 3657, 1999, pp. 226–239.

[20] N. Checcacci, M. Barni, F. Bartolini, and S. Basagni, "Robust video watermarking for wireless multimedia communications," in *Proc. 2000 IEEE Wireless Communications and Networking Conf.*, vol. 3, 2000, pp. 1530–1535.

[21] L. Zhang, Z. Cao, and C. Gao, "Application of RS-coded MPSK modulation scenarios to compressed image communication in mobile fading channel," in *Proc. 2000 52nd IEEE Vehicular Technology Conf.*, vol. 3, 2000, pp. 1198–1203.

[22] P. Dang and P. Chau, "Image encryption for secure Internet multimedia applications," *IEEE Trans. Consum. Electron.*, vol. 46, no. 3, pp. 395–403, Aug. 2000.

[23] F. Duan, I. King, L. Xu, and L. Chan, "Intra-block max-min algorithm for embedding robust digital watermark into images," in *Proc. IAPR Int. Workshop Multimedia Information Analysis and Retrieval*, vol. 1464, H. H. S. Ip and A. W. M. Smeulders, Eds., Heidelberg, Germany, 1998, pp. 255–264.

[24] A. Ambroze, G. Wade, C. Serdean, M. Tomlinson, J. Stander, and M. Borda, "Turbo code protection of video watermark channel," *Proc. IEEE Vision, Image and Signal Processing*, vol. 148, no. 1, pp. 54–58, Feb. 2001.

[25] D. Kirovski and H. Malvar, "Robust covert communication over a public audio channel using spread spectrum," in *Proc. 1st Information Hiding Workshop*, Pittsburgh, PA, 2001, pp. 354–368.

[26] J. Princen, A. Johnson, and A. Bradley, "Subband/ transform coding using filter bank designs based on time domain aliasing cancellation," in *Proc. IEEE ICASSP*, Dallas, TX, Apr. 1987, pp. 2161–2164.

[27] H. Malvar, "A modulated complex lapped transform and its applications to audio processing," in *Proc. IEEE ICASSP*, Mar. 1999, pp. 1421–1424.

[28] S. Sun and P. Chang, "Video watermarking synchronization based on profile statistics," *IEEE Aerospace Electron. Syst. Mag.*, vol. 19, no. 5, pp. 21–25, May 2004.

[29] J. Huang, Y. Wang, and Y. Shi, "A blind audio watermarking algorithm with self-synchronization," in *IEEE Int. Symp. Circuits and Systems*, vol. 3, May 2002, pp. 627–630.

[30] C. Hzu and J. Wu, "Digital watermarking for video," in *Proc. 1997 13th Int. Conf. Digital Signal Processing, DSP 97*, vol. 1, Jul. 2–4, 1997, pp. 217–220.

[31] X. Niu and S. Sun, "A new wavelet-based digital watermarking for video," in *Proc. 9th IEEE Digital Signal Processing Workshop*, TX, Oct. 2000, p. 241.

[32] K. Su, D. Kundur, and D. Hatzinakos, "A novel approach to collusion-resistant video watermarking," in *Proc. Security and Watermarking of Multimedia Contents IV SPIE*, vol. 4675, E. J. Delp and P. W. Wong, Eds., San Jose, CA, Jan. 2002, p. 12.

[33] ——, "A content-dependent spatially localized video watermarked for resistance to collusion and interpolation attacks," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, pp. 818–821.

[34] Y. Wang, J. Doherty, and R. Dyck, "A wavelet-based watermarking algorithm for ownership verification of digital image," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 77–88, Feb. 2002.

[35] R. Wolfgang, C. Podilchuk, and E. Delp, "The effect of matching watermark and compression transforms in compressed color images," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 1, Chicago, IL, Oct. 1998, pp. 440–444.