Proceedings of 2009 APSIPA Annual Summit and Conference, Sapporo, Japan, October 4-7, 2009

Automatic Attack Identification and Frame Recovery Using Versatile Video-based Watermarking

Yi-Chong Zeng
Institute of Information Science, Academia Sinica, Taipei, Taiwan
E-mail: yichongzeng@gmail.com   Tel:+886-2-27883799-2419

Abstract— **Aim at attack identification and frame recovery, we propose a versatile video-based watermarking scheme. This scheme integrates halftoning and watermarking techniques, and the watermark is particularly composed of two types of data, namely analogue data and index data. To examine the extracted watermark is not only to authenticate the video content, but it is capable of attack identification and corrupted frame recovery. Furthermore, a big breakthrough of the proposed scheme is that original data is unavailable, we merely use the extracted watermark compared with the covered frame. The experiment results demonstrate that our scheme actually diagnoses the frame/video statement, detects the tampered areas and then recovers them.**

I. Introduction

Recently, many powerful softwares are available, the public can use them to edit video frames easily and recompress video with various compression rates. On the other hand, pirates also can easily copy, tamper and edit video, thus, the studies in video authentication get more attention. The researchers investigated various watermarking techniques in order to protect the authorized multimedia. In [1], Mobasseri proposed a spatial-based watermarking method, which adds watermark to raw video directly. Su et al. solves problems of collusion and interpolation attacks by developing a content dependent spatially localized watermarking method [2]. Zhang et al.'s algorithm embeds copyright information in motion vectors [3].

Survey most of watermarking approaches, researchers used to authenticate multimedia content by original data, such as, original watermark, image, video, etc. For instance, the watermark detector measures difference between extracted watermark and existing watermarks in Cox et al.'s method [4], the response of the watermark detector indicates whether multimedia is authorized or not. In [5], Lu et al. used similar detector to indicate the existence of specified-authorized watermark. In [6], Yong et al. calculate the correlation coefficient between the original watermark and extracted watermark to check the existence of watermarking. However, once detector lacks original data, the authentication process will be failure.

Furthermore, researchers interest in the multipurpose watermarking methods. In [7], Lu and Liao proposed a scheme to simultaneously embed robust and fragile watermarks into host image for copyright protection and content authentication. Lin and Chang's method hides two kinds of bits, namely authentication bit and recovery bit, based on DCT domain [8]. Their method can reject image modification once covered image is compressed with pre-determined quality factor. In [9], Lee and Lin's method utilizes extracted watermark to detect tampered areas in image and to recover the tampered areas.

In this study, we propose a versatile watermarking scheme for video protection, it is capable of achieving multiple purposes, including, watermark verification, attack identification, tamper detection, and frame recovery. Our scheme employs the complex technique, which is the integration of watermarking and halftoning techniques. The previous works has demonstrated that jointing the halftoning and watermarking technique capability can localize the tampered areas and then recover them for still image [10], and the watermarking technique can diagnose video automatically [11]. Furthermore, a big breakthrough of the proposed scheme is that the original data is unavailable. Only the extracted watermark compares with the covered frame for video content authentication.

This paper is organized as follows. Sections II and III introduce the halftoning and watermarking techniques. Section IV describes the details of attack identification, tamper detection and frame recovery. Section V shows the experiment results, and conclusion is drawn in Section VI.

II. Halftoning Technique

Most of studies use either a sequence of random numbers or a meaningful logo as watermark [1-11]. The proposed method is different from previous works in watermark generation. In our study, the watermark should be versatile, it means that the watermark is not only used for checking the existence of watermarking, but it can identify the attacks imposed on video and recover the tampered areas. Moreover, the proposed method must work well once original data is unavailable.

The halftoning technique is employed to generate the halftone of every intra-frame as watermark, because the halftone is considered as monochromatic representation of frame. The restored frame derived from the halftone is similar to the original frame. Several well-known halftoning approaches have been presented, i.e., error-diffusion [12, 13], ordered dither [13], look-up-table [14], etc. Similarly, there are many inverse halftoning approaches proposed to restore the gray-level image, i.e., average filtering, Gaussian lowpass filtering, look-up-table [15], etc. However, the existing halftone approaches are not suitable in this study, and there are two issues not considered: the halftone dimension and the quality of restored frame. The halftone dimension is limited due to watermark capacity, and the quality of restored frame depends on inverse halftoning approach. Most of halftoning and inverse-halftoning techniques are irrelevant to each other except for look-up-table approach. Because the original data is unavailable in this study, the look-up-table approach is inapplicable as well. The conventional approaches, such as, error-diffusion and ordered dither, generate the halftone, but those approaches cannot ensure the better quality in restored frame. For this reason, we propose a new halftoning scheme using mix-integer quadratic programming (MIQP), the halftoning scheme is related to the inverse halftoning scheme, which is a 3×3 average filter, and our method can obtain high-quality restored frame.

As mentioned above, the halftone dimension is limited with respect to watermark capacity, and watermark capacity is related to watermarking technique as well. For this reason, we down-scaled the original frame to generate watermark. A halftone is generated from an $h_s \times w_s$ down-scaled intra-frame (I_s), the dimensions are adjusted to the power of 4. Assume that A is a 4×4 non-overlapped pixel block divided from the down-scaled frame, and a_{ij} is the (i,j)-th element of A. The halftoning process converts pixel block A into binary block B, which consists of 16 binary values b_{ij}. In the inverse halftoning process, the average filter is performed on B to yield the restored block A'. The proposed method estimate the optimal value b_{ij} so that A' is similar to A, the relationships of a_{ij}, a'_{ij} and b_{ij} are defined as follows,

$$\begin{aligned}
&\text{Minimize}: \sum_k \varepsilon_k^2 \\
&\varepsilon_k = a_{ij} - a'_{ij} \\
&a'_{ij} = 255 \times \frac{1}{N_k} \sum_{i-1}^{i+1} \sum_{j-1}^{j+1} b_{ij}, \\
&b_{ij} \in \{0,1\}, \\
&-\infty \le \varepsilon_k \le \infty \\
&\text{where } i, j \in \{0,\ 1,\ 2,\ 3\}, \text{and } k = i*4+j.
\end{aligned} \tag{1}$$

where ε_k is the error between a_{ij} and a'_{ij}, and a'_{ij} is the (i, j)-th element of the restored block A'. N_k is the total number of elements which include b_{ij} and its eight-connected neighbors. There are 2^{16} probabilities for binary matrix B, however, it is a heavy work to find the best solution. In fact, to estimate b_{ij} is exactly to solve the problem of quadratic programming. The general form of mix-integer quadratic programming is formulated as,

$$\begin{aligned}
\underset{\mathrm{x}}{\text{minimum}} \quad & f(\mathrm{x}) = \tfrac{1}{2}\mathrm{x}^{\mathrm{T}}\mathrm{F}\mathrm{x} \\
\text{subject to} \quad & \mathrm{S}\mathrm{x} = \mathrm{b} \\
& \mathrm{x}_{\mathrm{L}} \le \mathrm{x} \le \mathrm{x}_{\mathrm{U}}
\end{aligned} \tag{2}$$

where S is a constraint matrix, x_L and x_U are the lower-bound and upper-bound vectors of x, respectively. To rewrite (1) to the formulae of mix-integer quadratic programming in (2), the vector x is defined as x=[b_{00}, b_{01}, b_{02}, b_{03}, b_{10}, …, b_{33}, ε_0, ε_1, …, ε_{15}]^T, which consists of 32 elements. The 32×32 matrix F is symmetric given by,

$$\mathrm{F}(i,j) = \begin{cases} 1, & \text{if } i = j \text{ and } i, j > 15 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The 16×1 column vector b is defined as b = [a_{00}, a_{01}, …, a_{ij}, a_{ij+1}, …, a_{32}, a_{33}]^T. Two 32×1 column vectors x_L and x_U are set to,

$$\mathrm{x}_{\mathrm{L}}(i) = \begin{cases} 0, & \text{if } 0 \le i \le 15 \\ -\infty, & \text{otherwise} \end{cases} \tag{4}$$

, and

$$\mathrm{x}_{\mathrm{U}}(i) = \begin{cases} 1, & \text{if } 0 \le i \le 15 \\ \infty, & \text{otherwise} \end{cases}$$

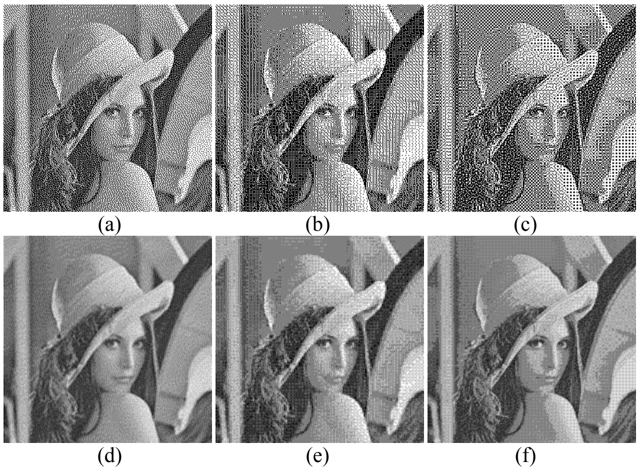
Eventually, the 16×32 constraint matrix S is defined as,

$$s_{ij} = \begin{cases} \phi_{ij}, & \text{if } 0 \le i, j \le 15 \\ -1, & \text{if } 16 \le j \le 32 \text{ and } j = i + 16 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$\Phi = \frac{255}{36}\begin{bmatrix}
9 & 9 & 0 & 0 & 9 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6 & 6 & 6 & 0 & 6 & 6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 6 & 6 & 6 & 0 & 6 & 6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 9 & 9 & 0 & 0 & 9 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6 & 6 & 0 & 0 & 6 & 6 & 0 & 0 & 6 & 6 & 0 & 0 & 6 & 0 & 0 & 0 \\
4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 0 & 0 & 0 \\
0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 6 & 6 & 0 & 4 & 6 & 6 & 0 & 0 & 6 & 6 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 6 & 6 & 0 & 0 & 6 & 6 & 0 & 0 \\
0 & 0 & 0 & 0 & 4 & 6 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 \\
0 & 0 & 0 & 0 & 0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 & 0 & 4 & 4 & 4 \\
0 & 0 & 0 & 0 & 0 & 4 & 6 & 6 & 0 & 0 & 6 & 6 & 0 & 0 & 6 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 9 & 0 & 0 & 9 & 9 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 6 & 6 & 0 & 6 & 6 & 6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 6 & 6 & 0 & 6 & 6 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 9 & 0 & 0 & 9 & 9
\end{bmatrix} \tag{6}$$

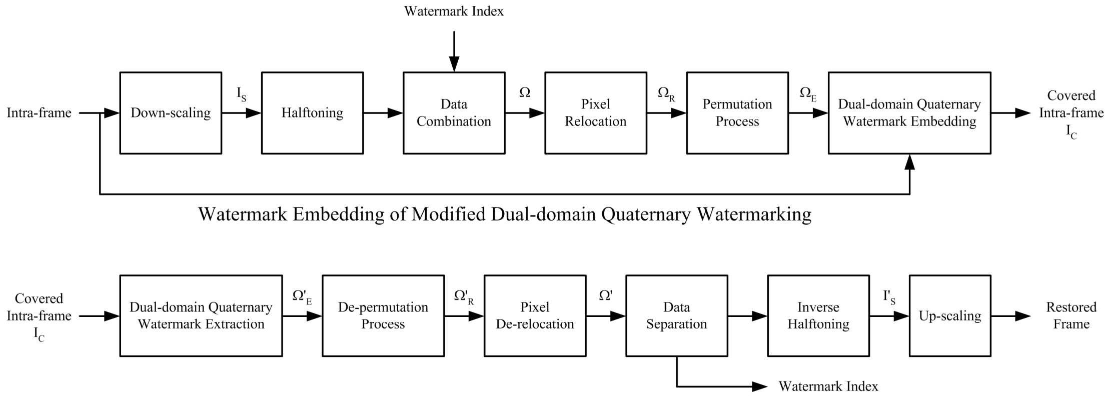
where ϕ_{ij} is the (i, j)-th element of the weighting matrix Φ.

An example is shown in Fig.1. Once the error-diffusion is implemented to the entire image, the halftone and restored image are shown in Figs.1(a) and (d), respectively. The peak signal-to-noise ratio (PSNR) of Fig.1(d) is 21.71 dB. The


(a) (b) (c)

(d) *PSNR*=21.71dB (e) *PSNR*=20.60dB (f) *PSNR*=24.39dB

Fig.1. Halftone Generation: (a) Error-diffusion method is implemented to the entire image to generate halftone, (b) error-diffusion method is implemented to the divided blocks to generate halftone, (c) halftone is generated by the proposed method, and (d), (e), (f) are the restored images of (a), (b), (c), respectively.



Watermark Embedding of Modified Dual-domain Quaternary Watermarking

Watermark Extraction of Modified Dual-domain Quaternary Watermarking

Fig.2. Block Diagram of Modified Dual-domain Quaternary Watermarking Method

error diffusion is implemented to all divided blocks of size 4×4, and the halftone and restored image with PSNR=20.60dB are shown in Figs.1(b) and (e), respectively. In our method, the halftone and restored image are shown in Figs.1(c) and (f), respectively. The PSNR of Fig.1(f) is 24.39dB, it demonstrates that our method obtain the high-quality restored image among three methods.

III. Watermarking Technique

Dual-domain quaternary watermarking is employed to embed significant information into intra-frames. In the previous work, DQW is developed for DCT-based image and video protection, and this approach can be used to diagnose video automatically [11]. Fig.2 illustrates the block diagram of the modified dual-domain quaternary watermarking (MDQW), the particular features of MDQW are encryption mechanism and watermark content. The watermark encryption includes two phases: pixel relocation and permutation process. In order to improve watermark imperceptibility and accuracy of tamper detection, the watermark pixels need to be encrypted. The details of MDQW are described as follows.

A. Watermark Specification

The embedding procedure of dual-domain quaternary watermarking is performed on discrete-cosine-transform (DCT) blocks, however, the quaternary watermark can be extracted from both of DCT blocks and spatial sub-blocks. In this study, the quaternary watermark is composed of two types of data, namely analogue data and index data. The analogue data is the halftone of down-scaled intra-frame as mentioned in Section II, and the index data represents the 8-bit watermark index, which ranges from 0 to 255 and are embedded repeatedly.

Assume that an $H\times W$ intra-frame is divided into several $h\times w$ DCT blocks. The total size of the quaternary watermark is $h\times w\times 8$ bits, because every DCT block is embedded four 2-bit watermark symbol. Moreover, due to the block halftoning is implemented on 4×4 pixel block, the dimension of halftone is adjusted to $h_f=\lfloor h\times 2^{-0.5}\rfloor\times 2^2$ and $w_f=\lfloor w\times 2^{-0.5}\rfloor\times 2^2$, which are power of 4, where h_f and w_f represent the height and width of the halftone, respectively. Moreover, the size of index data is s_{idx} bits, where $s_{idx}= (h\times w\times 8) - (h_f\times w_f)$. The index data is divided into 8 groups, and each group represents one bit of 8-bit watermark index. The voting system determines the group represents either 0 or 1 value. For example, if the intra-frame dimension is 288×352 and the size of quaternary watermark is 36×44×8 bits, hence, the dimension of analogue data is 100×124, and there is the 272-bit index data (which is calculated by 272=36×44×8−100×124).

B. Pixel Relocation

Once we generate the analogue data (halftone) and index data, those data are reshaped the two one-dimensional vectors, and we combine those two vectors and reshape them a two-dimensional matrix. For instance, the 100×124 analogue data and the 272-bit index data are combined and reshaped as the (72×2)×88 initial watermark (Ω).

Subsequently, the pixels of initial watermark are relocated to resist tampering attack, because tampering attack destroys one or more regions in frame, and it results in the corrupted watermark pixels gathering. Since watermark extraction implements, the pixel de-relocation can distribute the corrupted watermark pixels. The pixel relocation is formulated to,

$$\Omega_R(i,j) = \left\{ \Omega(u,v) \left| \begin{array}{l} u = d\times(i \bmod \dfrac{w'}{d}) + \left\lfloor i\times\dfrac{d}{w'}\right\rfloor, \\ v = d\times(j \bmod \dfrac{h'}{d}) + \left\lfloor j\times\dfrac{d}{h'}\right\rfloor. \end{array} \right. \right\} \tag{7}$$

where Ω_R denotes the relocated watermark, and d is the spreading number. The variables w' and h' represent the width and height of watermark, respectively.

C. Permutation Process

The purpose of permutation process is to achieve the watermark imperceptibility. We utilize one-to-one mapping sequence [9], which is defined as,

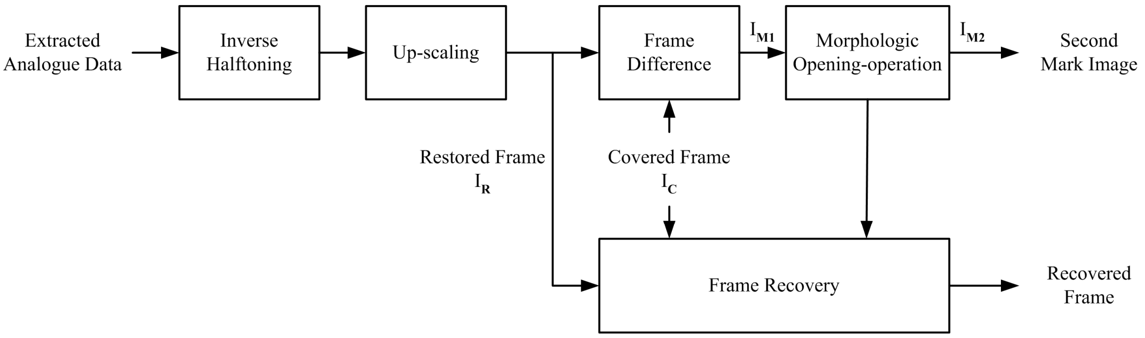


Fig.3. Block Diagram of Tampered Area Detection and Frame Recovery

$$X'_{\mathrm{P}} = \left\{ \Psi(X') \left| \begin{array}{l} \Psi(X') = \left((k \times X') \bmod N\right) \\ \gcd(k,\ N) = 1 \\ X', X'_{\mathrm{P}} \in \{0, 1, \cdots, N-1\} \end{array} \right. \right\} \quad (8)$$

where X' and X'_{P} denote one-dimensional coordinates of original watermark and permuted watermark, respectively, Ψ represents the permutation function, N is the total numbers of watermark bits, and the integers k and N are relatively prime. The initial watermark is processed with the pixel relocation and permutation process in order, the resultant is called the encrypted watermark (Ω_{E}).

Since the watermark embedding implements, the encrypted watermark is separated to two $2h\times2w$ sub-watermarks, namely first and second sub-watermarks (Ω_1 and Ω_2). Those two watermarks are combined as the quaternary watermark Ω_{Q}, which is defined as $\Omega_{\mathrm{Q}}=2\Omega_1+\Omega_2$. Inversely, the proposed method extracts the quaternary watermark from the covered intra-frame, and the quaternary watermark is separated into two $2h\times2w$ sub-watermarks and reshaped to the encrypted watermark. Subsequently, the extracted-encrypted watermark is de-permuted and de-relocated to obtain the analogue data and index data.

D. *Voting System*

While the index data is derived from the extracted-encrypted watermark, the voting system estimates the watermark index, which is defined as follows,

$$V(g_i) = \begin{cases} 1, & \text{if } \sum_{j=0}^{\|g_i\|} g_i(j) \geq \left\lceil \|g_i\| \times \frac{1}{2} \right\rceil \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

$$IDX = \sum_{i=0}^{7} V(g_i) \times 2^i,$$

where $g_i(j)$ and $\|g_i\|$ represent the j-th binary value and the total numbers of bits in the i-th group, respectively. The variable IDX denotes the 8-bit watermark index, which ranges from 0 to 255.

IV. Attack Identification, Tamper Detection and Frame Recovery

The previous work in video diagnosis was addressed in [11], the difference between the extracted and original quaternary watermarks was compared and analyzed in order to identify the attack categories, including, frame attack, temporal attack and composite attack. The previous method works well under the original watermarks are available. Even the recent studies need either the original watermark or original video/image content for checking the existence of watermarking [1-8]. However, the proposed method can use the extracted analogue data and index data for watermark verification and attack identification without the originals. Furthermore, the analogue data is capable of frame recovery while the frame is corrupted by tampering attack. The block diagram of tampered area detection and frame recovery is shown in Fig.3, and the details are described as follows.

A. *Attack Identification and Tamper Detection*

After extracting the encrypted watermark, a 3×3 average filter is performed on the extracted analogue data to obtain the restored frame, and the watermark index is estimated from index data by (9). If the frame is authorized with embedding the encrypted watermark, the restored frame derived from extracted analogue data should be similar to the covered frame. Otherwise, the extracted analogue data is meaningless. Therefore, we calculate the normalized correction (NC) between the covered frame ($\mathrm{I_C}$) and up-scaled restored frame ($\mathrm{I_R}$), which is formulated to,

$$NC = \frac{\sum_i \sum_j \mathrm{I_C}(i,j)\mathrm{I_R}(i,j)}{\sqrt{\sum_i \sum_j \mathrm{I_C}(i,j)^2}\sqrt{\sum_i \sum_j \mathrm{I_R}(i,j)^2}} \quad (10)$$

If the NC value is larger than threshold τ_1, the frame is authorized; otherwise, the frame will be called non-watermarked frame (NW).

Because pixel relocation and permutation process take effect, the corrupted watermark pixels are distributed over the entire watermark. The first mark image ($\mathrm{I_{M1}}$) records position of the corrupted pixel, which is defined according to the absolute difference between the two pixel intensities of $\mathrm{I_C}$ and $\mathrm{I_R}$ is larger than threshold τ_2,

$$\mathrm{I_{M1}}(i,j) = \begin{cases} 1, & \text{if } \left|\mathrm{I_C}(i,j) - \mathrm{I_R}(i,j)\right| > \tau_2 \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Noise removal process is implemented to the first mark image using opening-operator with a 5×5 structure element. The resultant is caller second mark image ($\mathrm{I_{M2}}$), which is utilized to identify the frame attack. For example, the tampered frame, up-scaled restored frame, first mark image and second mark

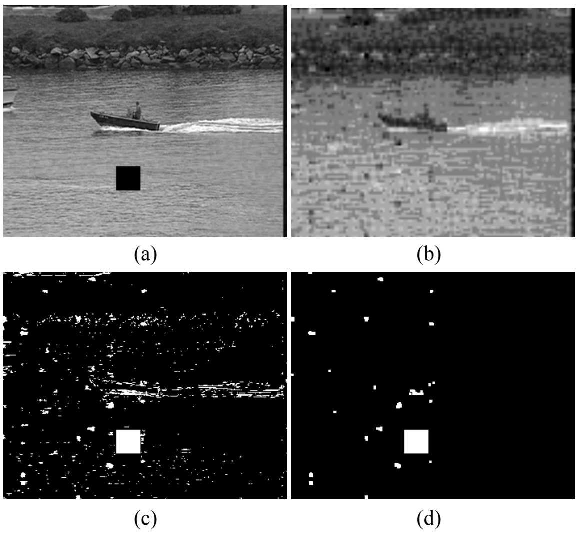

(a) (b) (c) (d)

Fig.4. (a) Tampered frame, (b) up-scaled restored frame, (c) first mark image, and (d) second mark image

image are shown in Figs.4(a)-(d), respectively.

The global-based frame attacks, i.e., lowpass filtering, median filtering, and recompression, cause the corrupted watermark pixels distributed over I_{M2}. On the contrary, the local-based frame attacks, i.e., region modification, salt & pepper noise, make the corrupted watermark pixels gathered. Therefore, the entropy of I_{M2} represents the attribution of attack. The second mark image is divided into 16×16 non-overlapped blocks, and the entropy is calculated by

$$E = \sum_{i=0}^{255} -p(i)\log_2 p(i), \tag{12}$$

$$p(i) = \frac{\sum_j I_{M2,i}(j)}{\sum_i \sum_j I_{M2,i}(j)}.$$

where E and $p(i)$ represent the entropy and the probability of i-th block in I_{M2}. The $I_{M2,i}(j)$ denotes the j-th pixel value of i-th block in I_{M2}, where $I_{M2,i}(j) \in \{0, 1\}$. Three kinds of frame attack statement (FAS) are identified according to,

$$\mathrm{FAS}(i) = \begin{cases} \mathrm{NF}, & \text{if } E \le \alpha_1 \\ \mathrm{LA}, & \text{if } \alpha_1 < E \le \alpha_2, \\ \mathrm{GA}, & \text{otherwise} \end{cases} \tag{13}$$

where FAS(i) represents the frame attack statement of i-th frame. The frame attack statements include non-frame attack (NF), local-based frame attack (LA) and global-based frame attack (GA). α_1 and α_2 are two thresholds.

We examine the estimated watermark index, and the temporal attack statement (TAS) is identified according to,

$$\mathrm{TAS}(i) = \begin{cases} \mathrm{NT}, & \text{if } idx_i = idx_{i-1}+1 \text{ and } idx_i = idx_{i+1}-1 \\ \mathrm{NT}, & \text{if } idx_i = idx_{i-1}+1 \text{ and } idx_i \ne idx_{i+1}-1 \text{ and } \mathrm{FAS}(i+1) = \mathrm{GA} \\ \mathrm{NT}, & \text{if } idx_i = idx_{i-1}+1 \text{ and } idx_i \ne idx_{i+1}-1 \text{ and } I_{i+1} \text{ is NW} \\ \mathrm{NT}, & \text{if } idx_i \ne idx_{i-1}+1 \text{ and } idx_i = idx_{i+1}-1 \text{ and } \mathrm{FAS}(i-1) = \mathrm{GA} \\ \mathrm{NT}, & \text{if } idx_i \ne idx_{i-1}+1 \text{ and } idx_i = idx_{i+1}-1 \text{ and } I_{i-1} \text{ is NW} \\ \mathrm{NT}, & \text{if } idx_{i-1} = idx_{i+1}-2 \text{ and } \mathrm{FAS}(i) = \mathrm{GA} \\ \mathrm{TA}, & \text{otherwise} \end{cases} \tag{14}$$

where idx_i and TAS(i) denote the watermark index and temporal attack statement of i-th frame (I_i), respectively. The temporal attack statements include non-temporal attack (NT) and temporal attack (TA). If both frame attack and temporal attack are detected in the same frame, we consider this frame is corrupted by the composite attack (CA).

Table I. Four types of video statement are conducted by summarizing the frame attack statement (FAS) and temporal attack statement (TAS).

FAS TAS	Non-frame Attack (NF)	Local-based Frame Attack (LA)	Global-based Frame Attack (GA)
Non-temporal Attack (NT)	Non-attacked Video (NV)	Frame-attacked Video (FV)	
Temporal Attack (TA)	Temporal-attacked Video (TV)	Composite-attacked Video (CV)	

The video statement is concluded by summarizing the frame attack statement and temporal attack statement, and four types of video statement are determined as shown in Table I, including, non-attacked video (NV), frame-attacked video (FV), temporal-attacked video (TV) and composite-attacked video (CV).

B. Frame Recovery

The restored frame is an $h_f \times w_f$ gray-level frame, whose size is the same as the analogue data. Using linear interpolation, the restored frame is up-scaled to an $H \times W$ restored frame (I_R). The frame recovery is defined by,

$$\hat{I}_k(i,j) = \begin{cases} I_R(i,j), & \text{if } I_{M2}(i,j) = 1 \text{ and } \mathrm{FAS}(k) = \mathrm{LA} \\ I_R(i,j), & \text{if } I_{M2}(i,j) = 1 \text{ and } \mathrm{FAS}(k) = \mathrm{CA} \\ I_C(i,j), & \text{otherwise.} \end{cases} \tag{15}$$

where $\hat{I}_k$ denotes the k-th recovered frame. From above equation, it shows that the frame recovery process is only implemented to the tampered frame, which only encounters local-based frame attack.

C. Performance Evaluation

Besides the peak signal-to-noise ratio (PSNR) is employed to evaluate the quality of recovered frame, we define two measurements to evaluate the proposed method in tamper detection, those two measurements are the detection rate (DR) and false alarm rate (FA) as formulated to,

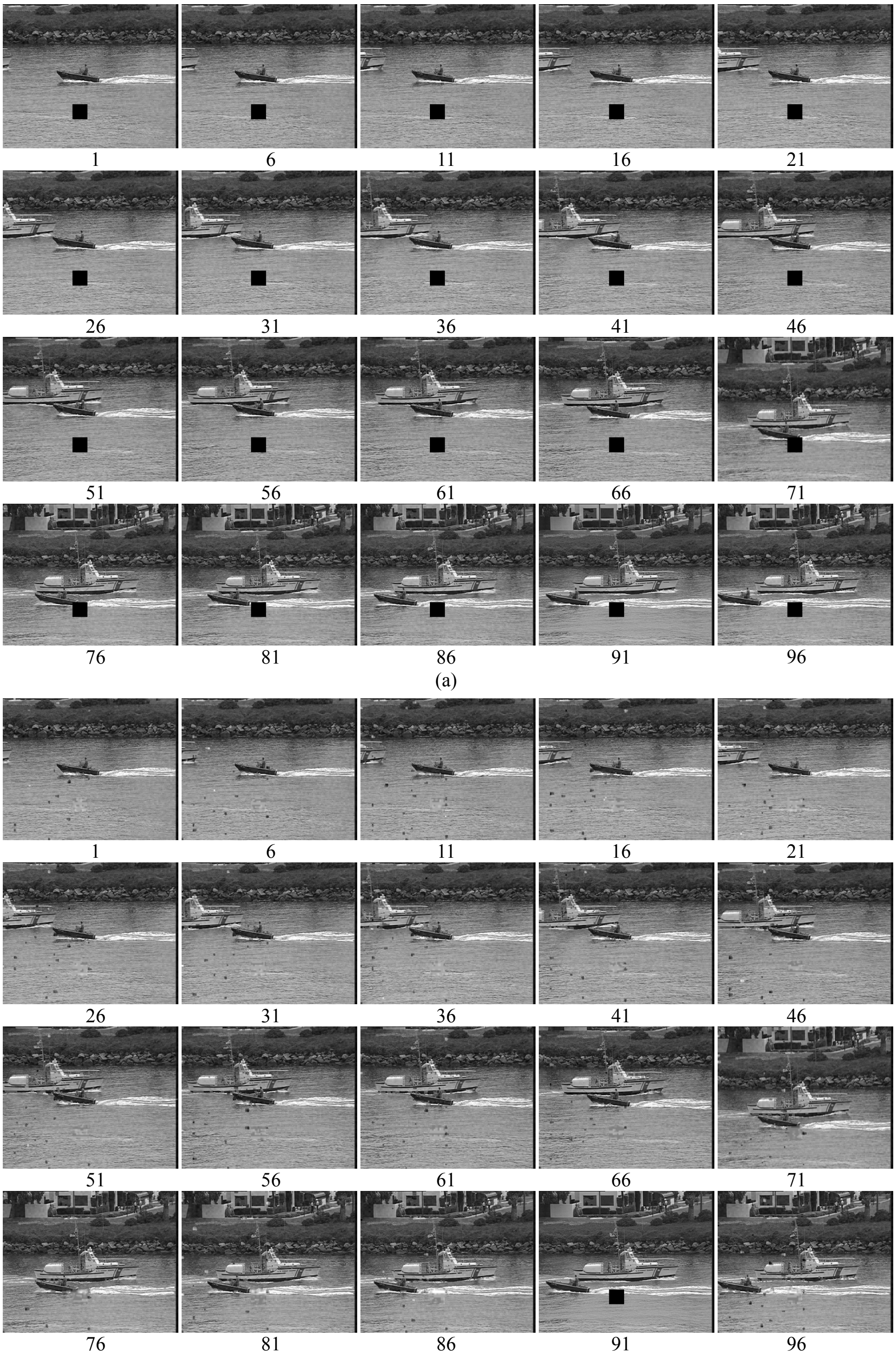
$$DR = \frac{\sum_i \sum_j I_{M2}(i,j)\,\mathrm{GT}(i,j)}{\sum_i \sum_j \mathrm{GT}(i,j)} \tag{16}$$

$$FA = \frac{\sum_i \sum_j I_{M2}(i,j)\left(1-\mathrm{GT}(i,j)\right)}{\sum_i \sum_j \left(1-\mathrm{GT}(i,j)\right)} \tag{17}$$

where GT represents the ground-truth of realistic tampered areas. The ideal performances are DR=1 and FA=0.

V. Experiment Results

We determine the proper parameters from the 1000 tested frames in an experiment, and the other experiments demonstrate the proposed method is efficient in attack identification, tamper detection and frame recovery. The



(a)

(b)

Fig.5. There are (a) twenty tampered Intra-frames, and (b) the corresponding recovered Intra-frames. The frame index is captioned below the frame.

dimension of all test intra-frames is 288×352, and those frames are embedded the (72×2)×88 encrypted watermark, which is composed of analogue data and index data. The dimension of analogue data is 100×124, and the size of index data is 272 bits. In addition, the threshold is set τ_2=45 in (11), the spreading number is d=2, and the parameters in permutation process is set k=13 and N=12672. Figs.5(a) and (b) show the tampered frames and the corresponding recovered frames, respectively.

A. Parameters Setting

Fig.6 shows three *NC* curves of non-attacked video, attacked video and non-watermarked video. The *NC* value represents the similarity between the analogue data and cover

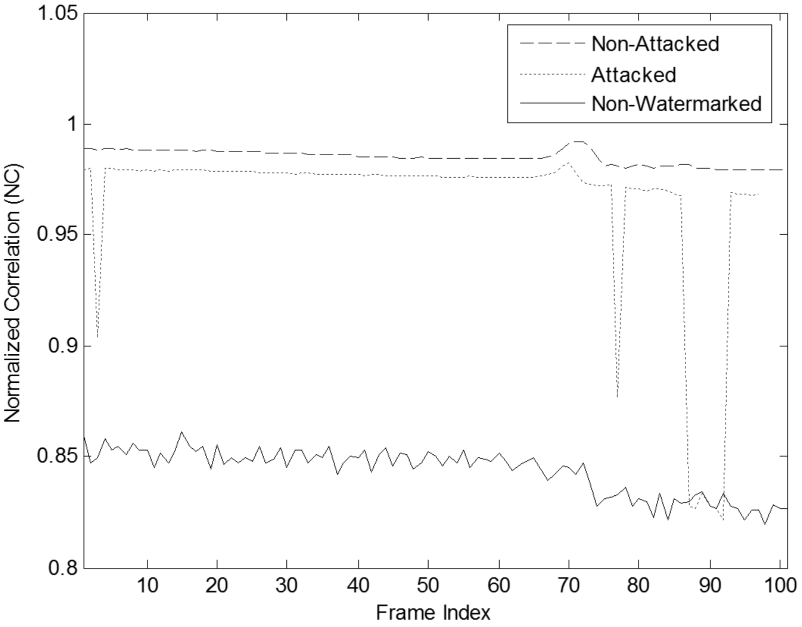


Fig.6. Three *NC* Curves of Non-attacked Video, Attacked Video and Non-watermarked Video

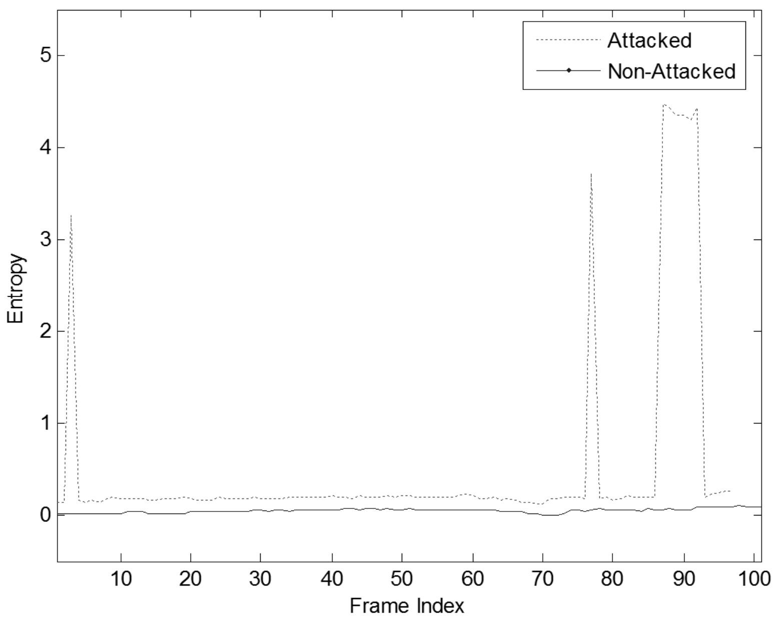


Fig.7. Two Entropy Curves of Non-attacked Video and Attacked Video

frame. Once *NC* value is large, it means that the analogue data is like the cover frame. If *NC* value is lower than the threshold τ_1, the frame is unauthorized. According to a large numbers of experiments, we find the proper threshold τ_1=0.87.

In (13), the frame attack statement is determined by two thresholds, α_1 and α_2, and the thresholds are estimated with respect to the entropies of tampered areas in the corrupted frames. Fig.7 shows two entropy curves of non-attacked video and attacked video plotted with dotted line and solid line, respectively. If the video is not corrupted by any attacker, the entropy of tampered area is close to 0. Furthermore, the entropy of the tampered area in local-based frame attack is smaller than that in global-based frame attack. Hence, 1000 test frames are examined in the experiment, and the proper parameters are set α_1=0.061 and α_2=2.

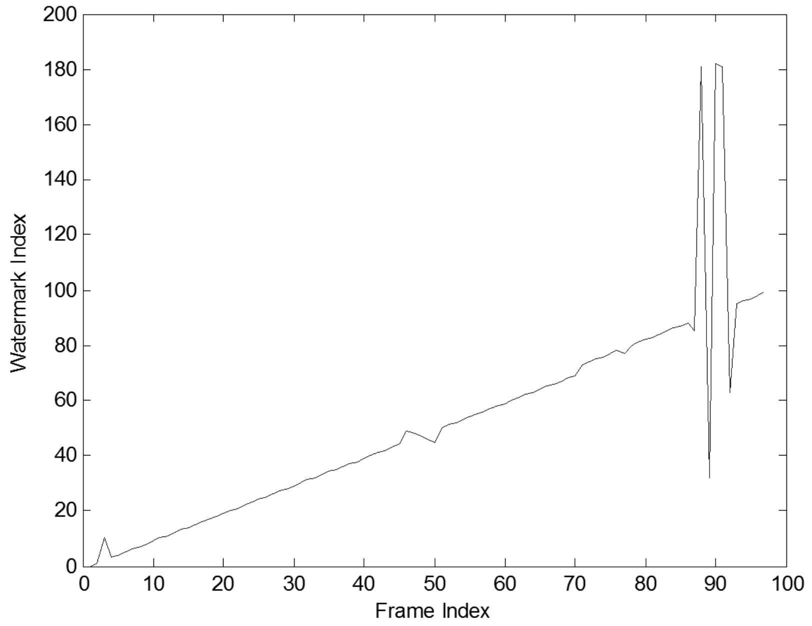


Fig.8. Watermark Indices of Test Video

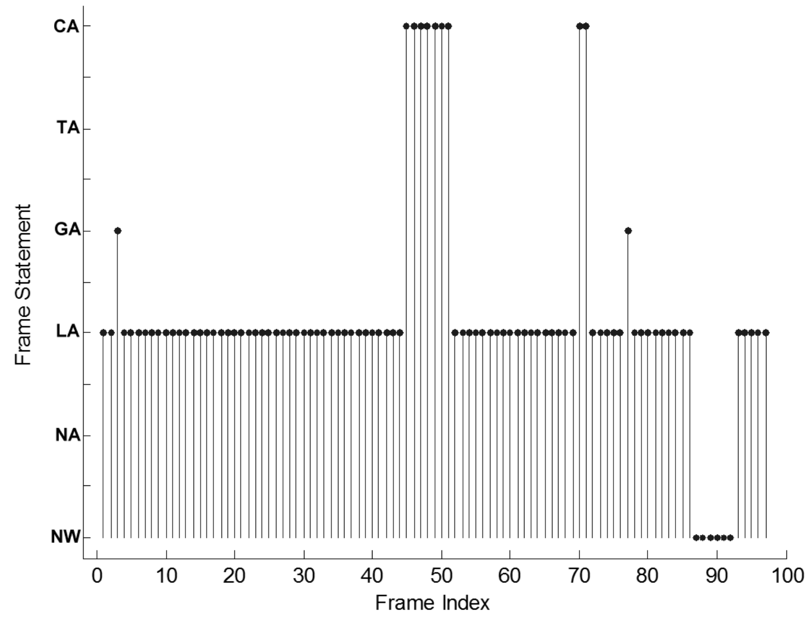


Fig.9. Video Diagnosis: The figure illustrates the resultant of video diagnosis, and the six frame statements are NW (non-watermarked frame), NA (non-attacked frame), LA (local-based frame attack), GA (global-based frame attack), TA (temporal attack) and CA (composite attack).

B. Tamper Detection and Video Diagnosis

The next experiment is to simulate that the varied attacks destroy the test video, those attacks include the regional modification with 30×30 region imposed on all intra-frames, median filtering at 3rd frame, lowpass filtering at 80th frame, sequence reversing from 46th to 50th frames, the frame deletion from 71st to 73rd frames, and the unauthorized frames insertion from 90th to 95th frames. The region modification is local-based frame attack, and media filtering and lowpass filtering are global-based frame attacks. There are two kinds of temporal attacks, i.e., sequence reversing and frame deletion.

The proposed method measures the *NC* values, entropies and watermark indices of the corrupted frames. In Fig.6, the *NC* values of the 87th to 92nd frames are lower than threshold

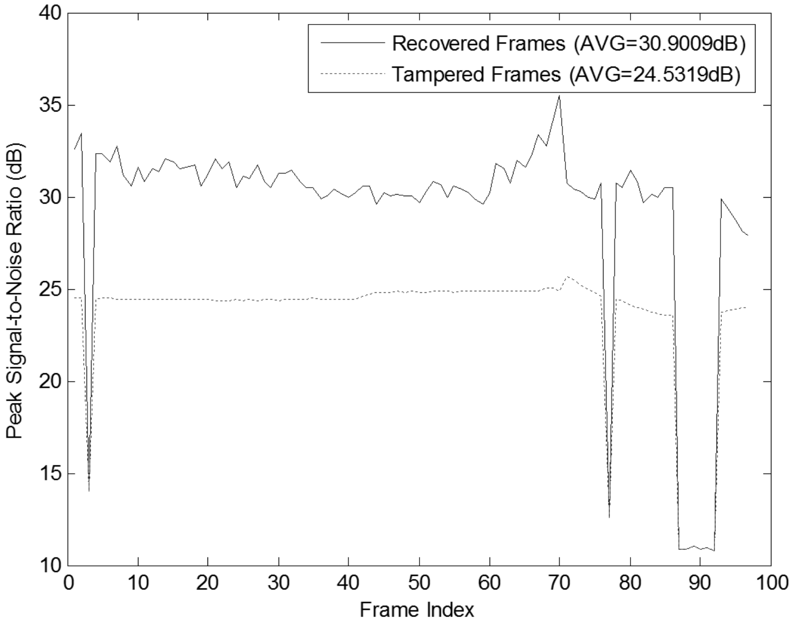


Fig.10. Two PSNR Curves of Tampered Frames and Recovered Frames

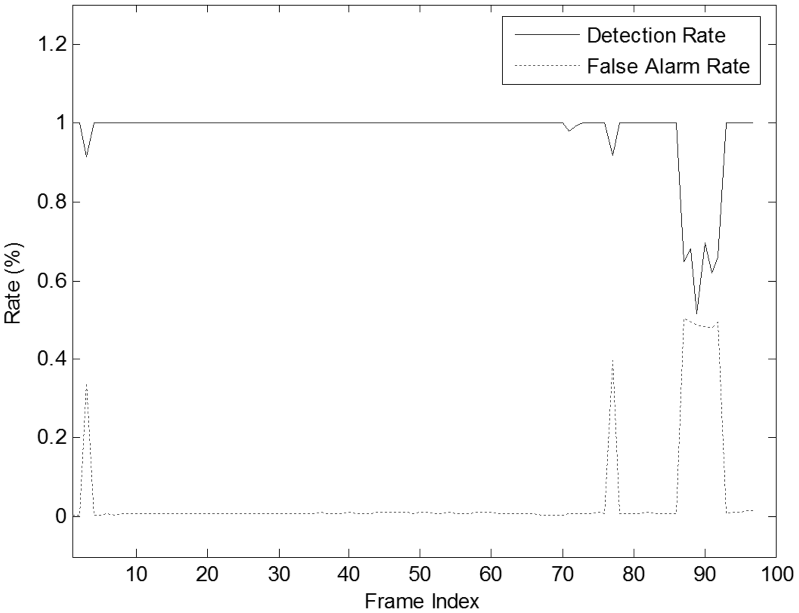


Fig.11. Detection Rate and False Alarm Rate of Tampered Areas Detection

τ_1, those frames are considered as the unauthorized frames and not embedded the specified watermark, and those frames are identified as non-watermarked frames (NW). In addition, the global-based frame attack (GA) has been detected at 3rd and 77th frames, and the local-based frame attack (LA) has been detected in the rest of intra-frames by analyzing entropies of Fig.7 via (13). To examine watermark indices of Fig.8 via (14), our method detects the frames corrupted with temporal attacks, where appear at 45th to 51st frames and 70th to 71st frames. Those frames have been detected the local-based frame attack, hence, we consider those frames are destroyed by composite attack. Consequently, the resultant of video diagnosis is shown in Fig.9.

C. *Frame Recovery*

One of contributions in our method is to recover the corrupted frame. The frame recovery process is only performed on the tampered frame corrupted by local-based frame attack. Fig.10 shows two PSNR curves of tampered frames and recovered frames, the average PSNRs of tampered frame and recovered frame are 24.53dB and 30.90dB, respectively. The resultant demonstrates the proposed method actually improves the frame quality, which increases 6.37dB averagely.

To analyze Fig.10, due to global-based frame attack seriously destroys encrypted watermark, and it results in the corruption of analogue data. Hence, the recovery process is quit at 3rd and 77th frames. In addition, the recovery process is also quit at 87th to 92nd unauthorized frames, which are not embedded encrypted watermark.

D. Evaluation

In order to evaluate the performance of tamper detection, we define two measurements, namely detection rate in (16) and false alarm rate in (17). Fig.11 depicts that once the frame is only damaged by local-based frame attack or temporal attack, the detection rate and false alarm rate are ideally close to 1 and 0, respectively. While global-based frame attack superimposes on frame, the detection rate is decreased and false alarm rate is increased as a result of global-based attack seriously destroys the hidden watermark. Consequently, it is quite obvious that the false alarm rates of 3rd and 77th frames are high, because those two frames are corrupted by media filtering and lowpass filtering.

VI. Conclusions

A versatile video-based watermarking scheme is proposed for attack identification and tamper detection, automatically. The hidden watermark not only can verify the existence of watermarking, but it is capable of frame recovery. A big breakthrough of our method is that the original video and watermark are unavailable, we only analyze the difference between the covered frame and the restored frame, which is derived from the extracted analogue data. The experiment results demonstrate three achievements of the proposed scheme: first, our method can identify whether the frame is embedded the encrypted watermark or not, and it can detect the frame attack, temporal attack and composite attack. Second, the high detection rate and low false alarm rate signify that the proposed scheme is efficient in tamper detection. Thirdly, the tampered areas are recovered with the extracted analogous data, and the frame quality is increased 6.37dB averagely.

References

[1] B. G. Mobasseri, “A Spatial Digital Video Watermark that Survives MPEG,” *ICIT*, pp.68-73, 2000.

[2] K. Su, D. Kundur, and D. Hatzinakos, “A content dependent spatially localized video watermark for resistance to collusion and interpolation attacks,” *ICIP*, vol.2, pp.818-821, 2001.

[3] J. Zhang, J. Li, and L. Zhang, “Video watermark technique in motion vector,” *Brazilian Symposium on Computer Graphics and Image Processing*, pp.179-182, Oct. 2001.

[4] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Source spread spectrum watermarking for multimedia,” *IEEE Trans. on Impage Processing*, vol.6, no.12, pp.1673-1687, Dec. 1997.

[5] C.-S. Lu, S.-K. Huang, C.-J. Sze, and H.-Y. M. Liao, “Cocktail watermarking for digital image protection,” *IEEE Trans. on Multimedia*, vol.2, no.4, pp.209-224, Dec. 2000.

[6] Y. Ma, Y. Tian, and Y. Qu, “Adaptive video watermarking algorithm based on MPEG-4 streams,” *ICARCV*, pp.1084-1088, 17-20 Dec. 2008.

[7] C.-S. Lu and H.-Y. M. Liao, “Multipurpose watermarking for image authentication and protection,” *IEEE Trans. on Image Processing*, vol.10, no.10, pp.1579-1592, Oct. 2001.

[8] C. Y. Lin and S. F. Chang, “Semi-fragile watermarking for authenticating JPEG visual content,” *SPIE Security and Watermarking of Multimedia Content II*, vol.3971, no.13, EI’00, San Jose, USA, Jan. 2000.

[9] T.-Y. Lee and S. D. Lin, “Dual watermark for image tamper detection and recovery”, *Pattern Recognition*, vol.41, pp.3497-3506, 2008.

[10] S.-C. Pei and Y.-C. Zeng, “Joint bi-watermarking and halftoning technique capability for both tampered areas localization and recovery of still image,” *ICIP*, vol.3, pp.281-284, Sept. 2007.

[11] Y.-C. Zeng and S.-C. Pei, “Automatic video diagnosing method using embedded crypto-watermarks,” *ISCAS*, pp.3017-3020, May 2008.

[12] R. Floyd and L. Steinberg, “An adaptive algorithm for spatial grey scale,” *Proceedings of the Society of Information Display*, vol.17, no.2, pp.75-66, 1976.

[13] R. Ulichney, Digital Halftoning, Massachusetts Institute of Technology, 1987.

[14] M. Mese, P. P. Vaidyanathan, “Look up table (LUT) method for image halftoning,” *ICIP*, vol.3, Vancouver, Canada, pp.993-996, 2000.

[15] M. Mese, P. P. Vaidyanathan, “Tree-structured method for LUT inverse halftoning and for image halftoning,” *IEEE Trans. on Image Processing*, vol.11, no.6, pp.644-655, Jun. 2002.