DOI: 10.1007/s11432-006-0313-6

Geometrically robust video watermarking based on wavelet transform

ZHAO Yao

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
(email: yzhao@center.njtu.edu.cn)

Received October 26, 2005; accepted January 10, 2006

Abstract Geometrical attacks can destroy most watermarking systems at present. So how to efficiently resist such kind of attacks remains a challenging direction in watermarking research. In this paper, a novel sequence watermarking scheme, which exploits a geometrical invariant, i.e. average AC energy (AAE) to combat arbitrary geometrical attacks, is presented. The scheme also uses some other measures, such as synchronization and optimal whitening filter to resist other attacks and improve detection performance. The experimental results show that the scheme can efficiently improve the visual quality of the watermarked video and achieve good robustness against random geometrical attacks. The scheme also has good robustness against other attacks, such as low-pass filtering along time axis and frame removal.

Keywords: watermarking, geometrical transform attacks, optimal whitening filter, wavelet transform.

1 Introduction

Compared with relevant analog data, digital media (digital audio, image, video, etc.) have some significant properties: high quality, easy manipulation, perfect copying and easy transmission. Because of these good properties, the development and application of digital media are growing explosively. However, these properties also pose a threat of unauthorized possession and illegal usage of digital media. An additional threat is the illegal tampering and modification of digital media. The need for a method of protecting the copyright of digital images thus arises.

The idea of using a robust digital watermark to detect and trace copyright violations has stimulated significant interests among artists and publishers. A digital watermark is an invisible mark embedded in a digital medium, which may be used for a number of purposes including image captioning and copyright protection[1]. Since the first paper was published by Schyndel *et al.* [2], the research has inspired many researchers and a variety of approaches have been proposed.

Most approaches can resist attacks such as compression, filtering, enhancing and other

signal processing operations. However, recently it has become clear that even very small geometric distortions can prevent the detection of a watermark. This problem is most pronounced when the original image is unavailable to the detector[3]. So how to efficiently resist such kind of attacks remains a challenging direction in watermarking research and some schemes have been proposed.

The current approaches to resist geometrical attacks may be classified into three categories: 1) Exhaustive search. The typical geometrical transformations commonly used in the image edition are applied on the whole image, and in many ways can be easily represented by a mathematical operation. One basic idea to identify the transformation is to perform an exhaustive detection considering all possible geometrical transformations of the marked image. In this case, the computing cost will dramatically increase[4]. 2) Geometrically reverse transformation. We know that after geometrical transformation, the watermark signal is still there. The detector cannot detect it since the generated random sequence and the embedded random sequence are not synchronized. So if we know the geometrical transform applied on the marked image, then we can reverse the transformation and then can extract the watermark bit by correlation calculation. This category can be further divided into two classes: semi-blind correlation and blind correlation. In the semi-blind correlation, the detector needs the original image to identify the geometrical transform by the matching pairs of the feature points of the original image and the attacked marked image[4], or to compensate the distortion caused by the attacks[5,6]. Actually, we cannot solve the problem even when we make use of the original image, especially when the image is rotated and scaled or it is printed and scanned[7,8]. Since it needs not the original image in the detection process, the blind correlation becomes the main direction and some approaches have been proposed. Some researchers proposed resynchronization based on the template technology[9,10]. In the embedding process, the scheme embeds meaningful watermark, and meanwhile embeds a special template, after malicious geometrical attacks, the detector can predict the geometrical transform according to the shape change of the special template. Kutter[11] proposed a self-reference method. The scheme repeatedly embeds the same watermark in four different positions of the cover image and in the detection. It calculates autocorrelation function (ACF) of the predicted watermark, resulting 9 peaks and thus determine the geometrical transform. 3) Using geometrical invariants. Researchers try to exploit the moments or features invariant to geometrical transforms, and modulate the invariants with the watermark signals. Some researchers introduced a watermarking scheme exploiting some properties of the Fourier transform[12]. In the scheme, mark embedding is performed using an RST (rotation, scaling, and translation) invariant Fourier-Mellin domain.

Since the mean luminance of video is least sensitive to spatial geometrical operations on video frames, Haitsma *et al.*[13] proposed a video watermarking scheme that hides watermark signals in the luminance mean series. Zhao *et al.*[14] proposed an improved scheme. However, since the watermark signals directly modify the average luminance of every frame, the watermark strength is quite limited and very strong watermark signals often make the video visually flick.

In order to overcome the drawbacks of the scheme in ref. [14], we exploit another

geometrical invariant, i.e. AAE and embed watermark signals in low frequency AC energy while maintaining the luminance means unchanged.

The rest of the paper is organized as follows: Section 2 introduces AAE and its properties, section 3 describes the embedding process, section 4 describes the extraction procedure and section 5 presents the principle of optimal whitening filter, in section 6 some experimental results are presented, at last, section 7 concludes the paper.

2 Average AC energy (AAE)

Obviously, if a system embeds watermark signals in some invariants to arbitrary geometrical transform, then the system can resist geometrical transforms.

For the tth frame $f_t(x, y)$ of a sequence, after geometrical transforms, such as rotation, shifting, the coordinate (x, y) changes to (x', y'), however, the pixel value remains the same. That is

$$f_t(x, y) = f_t(x', y'). \tag{1}$$

So

$$\sum_{x,y} f_t(x, y) = \sum_{x',y'} f_t(x', y'), \tag{2}$$

$$\sum_{x,y} f_t^2(x, y) = \sum_{x',y'} f_t^2(x', y'). \tag{3}$$

Definition. Since $\sum_{x,y} f_t^2(x, y)$ is the total energy of a frame, $\frac{1}{KL}\left(\sum_{x,y} f_t(x, y)\right)^2$ is the DC energy. So define eq. (4) as the average AC energy (AAE) of frame $f_t(x, y)$.

$$v_t = \frac{1}{KL}\left(\sum_{x,y} f_t^2(x, y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x, y)\right)^2\right), \tag{4}$$

where K, L are the dimensions of frame $f_t(x, y)$.

AAE has the following properties:

1) After the geometric attack, if the dimensions of the frame remain unchanged, then v_t keeps unchanged.

From eqs. (2) and (3), we can easily get the conclusion.

$$v_t' = \frac{1}{K'L'}\left(\sum_{x',y'} f_t^2(x', y') - \frac{1}{K'L'}\left(\sum_{x',y'} f_t(x', y')\right)^2\right)$$

$$= \frac{1}{KL}\left(\sum_{x,y} f_t^2(x, y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x, y)\right)^2\right) = v_t. \tag{5}$$

2) If the frame spatially scales with a factor s, then v_t still keeps unchanged.

$$v_t' = \frac{1}{s^2KL}\left(s^2\sum_{x,y} f_t^2(x, y) - \frac{1}{s^2KL}\left(s^2\sum_{x,y} f_t(x, y)\right)^2\right)$$

$$= \frac{1}{KL}\left(\sum_{x,y} f_t^2(x,y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x,y)\right)^2\right) = v_t . \tag{6}$$

3) If the frame is rotated or shifted, and part of the new image is padded with black, then v_t may change.

$$\begin{aligned} v_t' &= \frac{1}{K'L'}\left(\sum_{x',y'} f_t^2(x',y') - \frac{1}{K'L'}\left(\sum_{x',y'} f_t^2(x',y')\right)^2\right) \\ &= \frac{1}{K'L'}\left(\sum_{\text{Image part}} f_t^2(x',y') + \sum_{\text{Padding part}} f_t^2(x',y')\right. \\ &\quad \left. - \frac{1}{K'L'}\left(\sum_{\text{Image part}} f_t^2(x',y') + \sum_{\text{Padding part}} f_t^2(x',y')\right)^2\right) \\ &= \frac{1}{K'L'}\left(\sum_{x,y} f_t^2(x,y) + \sum_{\text{Padding part}} f_t^2(x',y')\right. \\ &\quad \left. - \frac{1}{K'L'}\left(\sum_{x,y} f_t^2(x,y) + \sum_{\text{Padding part}} f_t^2(x',y')\right)^2\right) \\ &= \frac{1}{K'L'}\left(\sum_{x,y} f_t^2(x,y) - \frac{1}{K'L'}\left(\sum_{x,y} f_t^2(x,y)\right)^2\right) \approx \frac{KL}{K'L'} v_t . \end{aligned} \tag{7}$$

Even though v_t may change, but v_t and v_t' have nearly a linear relationship. In the following section, we will observe that the watermark detection may be affected a little, but because of the nearly linear relationship, we can still successfully detect watermark.

From the above analysis, we know that embedding watermark signals in v_t series can achieve good geometrical robustness.

3 Embedding process

The watermark embedding system is shown in Fig. 1.

In order to access the watermark randomly from any frame and resist frame removal attack, synchronization bits are embedded alternately with the meaningful watermark. The embedding frame structure is shown in Fig. 2.

Any meaningful watermark and synchronization can be interpreted as 1 and −1 sequence, denoted as $b_1 b_2 \cdots b_i \cdots$ We use spread spectrum to hide the bit sequence.

A pseudo random sequence (PRS) generator is used to generate two sets of PRSs, namely one for watermark and one for synchronization according to two different keys. Also the two sets may be of different lengths.

In order to improve the robustness against attacks in the temporal axis, we use the low-pass watermark instead of white Gaussian noise. Since HVS tells us that the human eyes' sensitivity at 1 Hz is relatively smaller than that at 10 Hz or so[16], we design a low-pass

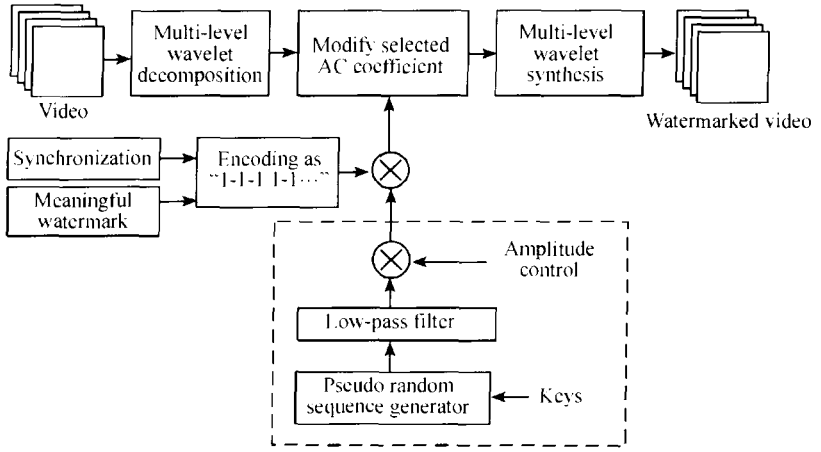

Fig. 1. The watermark embedding diagram.

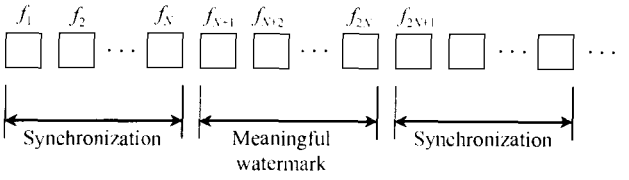

Fig. 2. The embedding frame structure.

filter whose cut-off frequency is approximately 1 Hz. The filter transfer function is

$$H(Z)=\frac{1}{1-0.4505Z^{-1}-0.988Z^{-2}+0.0429Z^{-3}+0.5112Z^{-4}}. \tag{8}$$

Since the construction, embedding and extraction processes of PRSs for synchronization and watermark are similar, we use w_t to denote a low-pass filtered PRS that can represent either a synchronization bit or watermark bit.

The construction of w_t holds zero mean, i.e. $E[w_t] = 0$, and we control the maximum amplitude Am to adjust embedding strength.

The watermark embedding process is to embed v_t into AAE series of the video, i.e. to modify the AC component of each frame. So for each frame of size 128×128, decompose it into 7 levels with bi-orthogonal wavelet. The coefficients of the wavelet filters are listed in Table 1.

Table 1 The bi-orthogonal wavelet filer

n	Decomposition filter		Synthesis filter	
	low-pass filter $h(n)$	high-pass filter $g(n)$	low-pass filter $s(n)$	high-pass filter $r(n)$
0	0.8527	0.7885	0.7885	0.8527
+1, −1	0.3774	−0.4181	0.4181	−0.3774
+2, −1	−0.1106	−0.0407	−0.0407	−0.1106
+3, −3	−0.0238	0.0645	−0.0645	0.0238
+4, −4	0.0378			0.0378

According to Fig. 1, the embedding process is to modify v_t with $b_i \cdot w_t$, that is

$$v_{tw} = v_t + b_i \cdot w_t, \tag{9}$$

where $b_i = 1$ or -1.

From eq. (4), we have

$$\begin{aligned} v_t + b_i \cdot w_t &= \frac{1}{KL}\left(\sum_{x,y} f_t^2(x,y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x,y)\right)^2\right) + b_i \cdot w_t \\ &= \frac{1}{KL}\left(\sum_{x,y} f_t^2(x,y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x,y)\right)^2 + KL \cdot b_i \cdot w_t\right) \\ &= \frac{1}{KL}(v_{tAC} + KL \cdot b_i \cdot w_t), \end{aligned} \tag{10}$$

where $v_{tAC} = \sum_{x,y} f_t^2(x,y) - \frac{1}{KL}\left(\sum_{x,y} f_t(x,y)\right)^2$ denotes the AC energy of the tth frame.

Eq. (10) means that the modification is to increase the AC energy by $KL \cdot b_i \cdot w_t$.

After wavelet decomposition, we will select some AC coefficients for modification. When selecting AC coefficients, we will consider the robustness and the video quality, and thus we only select the AC coefficients of level 2 to level 7, as shown in Fig. 3.

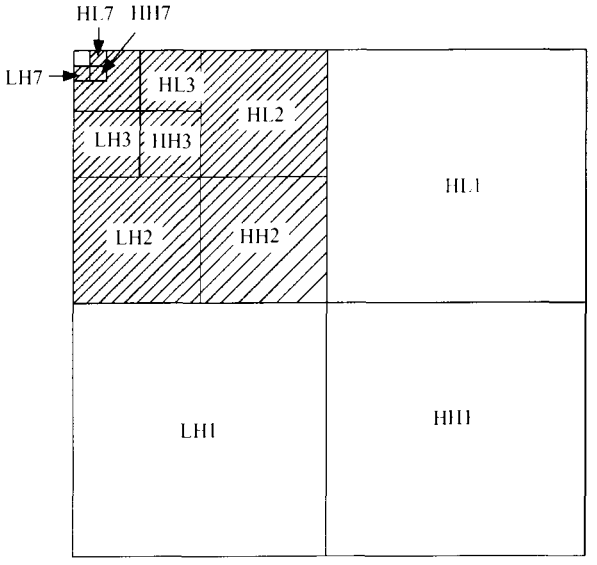

Fig. 3. The coefficients selected for modification.

Let Ω denote the set of all the selected frequency points. Then the AC energy of the selected coefficients can be denoted as $\sum_{(u,v)\in\Omega} F_t^2(u,v)$. According to eq. (10), the modified energy is

$$\sum_{(u,v)\in\Omega} F_{tw}^2(u,v) = \sum_{(u,v)\in\Omega} F_t^2(u,v) + KL \cdot b_i \cdot w_t. \tag{11}$$

There are a lot of modification methods for $F_t(u, v)$ to satisfy eq. (11). In our scheme, we modify every selected AC coefficient in the same proportion, let α denote the proportion. That is

$$F_{tw}(u, v) = \alpha \cdot F_t(u, v). \tag{12}$$

From eqs. (11) and (12), we can determine α as

$$\alpha = \sqrt{\frac{\sum_{(u,v)\in\Omega} F_t^2(u,v) + b_i \cdot KL \cdot w_t}{\sum_{(u,v)\in\Omega} F_t^2(u,v)}}. \tag{13}$$

After the modification, inverse transform $F_{tw}(u, v)$ and get the watermarked frame $f_{tw}(x, y)$. In the scheme, watermark is only embedded in the luminance component of a color video.

4 Extraction process

Fig. 4 illustrates the watermark extraction process. The AAE v_{tw} is calculated according to eq. (4). The detector first detects the synchronization along temporal direction. If no synchronization detected, that means the sequence does not contain watermark; otherwise, we calculate if the total frame number between two successive synchronizations is right; if yes, we can detect the meaningful watermarks from the sub-sequence. Otherwise, it means that some frames between the two synchronizations may be removed or inserted, the watermark cannot correctly extracted from this sub-sequence, and the extractor will process next two synchronizations. With synchronization, we can access the watermark from any frame and can resist frame removal attacks.

As shown in Fig. 4, the cross-correlation function between v_{tw} and watermark is used to detect the presence of watermark.

$$\begin{aligned} R_{v_w,w}(0) &= E(v_{tw}w_t] = E[(v_t + b_i \cdot w_t)w_t] \\ &= E[v_t w_t] + b_i E[w_t^2] = R_{v,w}(0) + b_i E[w_t^2]. \end{aligned} \tag{14}$$

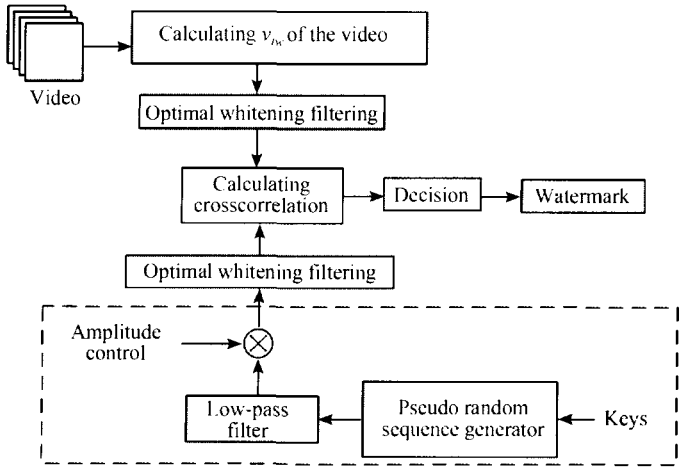

Fig. 4. Watermark extraction diagram.

Since the v_t and w_t are independent, so ideally

$$R_{v,w}(0) = E[v_t w_t] = E[v_t] \cdot E[w_t] = 0 . \tag{15}$$

So if the sequence is embedded with a watermark, then eq. (14) equals to $b_i E[w_t^2]$, and we can further determine 1 or -1 according to its sign; if not, then eq. (14) equals to 0. Thus we can easily extract the watermark bit.

However, since the length of v_t and w_t is finite, so the practical formula of estimating eq. (15) becomes

$$Cor(v_t, w_t) = \frac{1}{N} \sum_{t=0}^{N-1} (v_t w_t) , \tag{16}$$

where N is the length of w_t.

The value of eq. (16) may not be 0, i.e. its variance does not equal to 0. As pointed out in ref. [13], its variance affects the detection performance, and the smaller, the better.

In order to improve the detector performance, i.e. to decrease the variance of the estimator (16), we use a whitening filter prior to correlation. A parametric first-order FIR whitening filter $G(z) = 1 - aZ^{-1}$ is used as a whitening filter, and in the next section, we will statistically calculate the optimum value for a.

5 Optimal whitening filter

The value of eq. (16) directly affects the detection performance, so we will analyze it.

Theorem 1. Eq. (16) is a zero-mean variable, its variance is

$$Var[Cor(v_t, w_t)] = \frac{2}{N^2} \sum_{\Delta=1}^{N-1} (N - \Delta) \cdot R_v(\Delta) R_w(\Delta) + \frac{1}{N} R_v(0) R_w(0) ,$$

where $R_v(\Delta)$, $R_w(\Delta)$ are the auto-correlation functions of signal v_t and w_t.

Proof. The mean of $Cor(v_t, w_t)$ is

$$E[Cor(v_t, w_t)] = E\left[\frac{1}{N} \sum_{t=0}^{N-1} (v_t w_t) \right] = \frac{1}{N} \sum_{t=0}^{N-1} E[v_t \cdot w_t] = E[v_t] \cdot E[w_t] = 0 . \tag{17}$$

The variance of $Cor(v_t, w_t)$ is

$$\begin{aligned} & Var[Cor(v_t, w_t)] \\ &= E[(Cor - E[Cor])^2] = E[Cor^2] = E\left[\left(\frac{1}{N} \sum_{t=0}^{N-1} (v_t w_t) \right)^2 \right] \\ &= \frac{1}{N^2} E\left[\left(\sum_{m=0}^{N-1} (v_m w_m) \right) \cdot \left(\sum_{n=0}^{N-1} (v_n w_n) \right) \right] = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[v_m w_m v_n w_n)] \\ &= \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} (E[v_m v_n] E[w_m w_n]) = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} (R_v(m-n) \cdot R_w(m-n)), \end{aligned} \tag{18}$$

where $R_v(m-n)$, $R_w(m-n)$ are the autocorrelation functions (ACF) of v_t and v_w.

Let us simplify the double summation in eq. (18).

The double summations $\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} R_v(m-n)\cdot R_w(m-n)$ can be rewritten as a matrix:

$$\begin{matrix} R_v(0-0)R_w(0-0) & +R_v(0-1)R_w(0-1) & +R_v(0-2)R_w(0-2) & +\cdots & +R_v(0-N+1)R_w(0-N+1) \\ +R_v(1-0)R_w(1-0) & +R_v(1-1)R_w(1-1) & +R_v(1-2)R_w(1-2) & +\cdots & +R_v(1-N+1)R_w(1-N+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ +R_v(N-1-0)R_w(N-1-0) & +R_v(N-1-1)R_w(N-1-1) & +R_v(N-1-2)R_w(N-1-2) & +\cdots & +R_v(N-1-N+1)R_w(N-1-N+1) \end{matrix}$$

So
$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} R_v(\hat{m}-n)\cdot R_w(m-n) = \sum_{\Delta=-(N-1)}^{N-1} (N-|\Delta|)\cdot R_v(\Delta)R_w(\Delta). \tag{19}$$

Considering the property $R_v(-\Delta)=R_v(\Delta)$ and $R_w(-\Delta)=R_w(\Delta)$, then we get

$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} R_v(m-n)\cdot R_w(m-n) = 2\cdot\sum_{\Delta=1}^{N-1}(N-\Delta)\cdot R_v(\Delta)R_w(\Delta)+N\cdot R_v(0)R_w(0). \tag{20}$$

Eq. (18) becomes

$$Var\left[Cor(v_t,w_t)\right]=\frac{2}{N^2}\sum_{\Delta=1}^{N-1}(N-\Delta)\cdot R_v(\Delta)R_w(\Delta)+\frac{1}{N}R_v(0)R_w(0). \tag{21}$$

End of proof.

From the above equation, when N is a finite number, the variance usually does not equal to 0. The false alarm probability of detection depends on the value and obviously the smaller the better.

From the detection theory it follows that correlation detectors are optimum in the case of a Linear Time-Invariant, frequency non-disperse, additive white Gaussian noise channel[17]. However, in our system, the AAE series of subsequent frames are naturally highly correlated and the low-pass filtered watermark is also correlated. In order to achieve optimum or better detection, we use a whitening filter prior to correlation. Assume the whitening filter we use is

$$G(z)=1-aZ^{-1}. \tag{22}$$

Then we have the following theorem.

Theorem 2. After the whitening filter in eq. (22), the variance becomes

$$\begin{aligned} &Var\left[Cor(v'(t),w'(t))\right] \\ &=\frac{2}{N^2}\sum_{\Delta=1}^{N-1}[(N-\Delta)\cdot((1+a^2)R_v(\Delta)-aR_v(\Delta-1)-aR_v(\Delta+1)) \\ &\cdot((1+a^2)R_w(\Delta)-aR_w(\Delta-1)-aR_w(\Delta+1))] \\ &+\frac{1}{N}[((1+a^2)R_v(0)-2aR_v(1))\cdot((1+a^2)R_w(0)-2aR_w(1))]. \end{aligned}$$

Proof. After whitening, the AAE signal v_t and the watermark signal w_t become

$$v'_t = v_t - av_{t-1}, \tag{23}$$

$$w'_t = w_t - aw_{t-1}. \tag{24}$$

According to Theorem 1, we have

$$Var[Cor(v_t', w_t')] = \frac{2}{N^2}\sum_{\Delta=1}^{N-1}(N-\Delta)\cdot R_{v'}(\Delta)R_{w'}(\Delta) + \frac{1}{N}R_{v'}(0)R_{w'}(0). \tag{25}$$

Since

$$\begin{aligned} R_{v'}(\Delta) &= R_{v'}(m-n) = E[v_m' v_n'] = E[(v_m - av_{m-1})\cdot(v_n - av_{n-1})] \\ &= E[v_m v_n - av_{m-1}v_n - av_m v_{n-1} + a^2 v_{m-1}v_{n-1}] \\ &= (1+a^2)R_v(m-n) - aR_v(m-n-1) - aR_v(m-n+1) \\ &= (1+a^2)R_v(\Delta) - aR_v(\Delta-1) - aR_v(\Delta+1). \end{aligned} \tag{26}$$

In the same way,

$$R_{w'}(\Delta) = (1+a^2)R_w(\Delta) - aR_w(\Delta-1) - aR_w(\Delta+1). \tag{27}$$

So eq. (25) becomes

$$\begin{aligned} &Var[Cor(v_t', w_t')] \\ &= \frac{2}{N^2}\sum_{\Delta=1}^{N-1}[(N-\Delta)\cdot((1+a^2)R_v(\Delta) - aR_v(\Delta-1) - aR_v(\Delta+1)) \\ &\quad\cdot((1+a^2)R_w(\Delta) - aR_w(\Delta-1) - aR_w(\Delta+1))] \\ &\quad + \frac{1}{N}[((1+a^2)R_v(0) - 2aR_v(1))\cdot((1+a^2)R_w(0) - 2aR_w(1))]. \end{aligned} \tag{28}$$

End of proof.

In order to obtain an optimized result, that is to minimize the above equation, we have to derive the differential of the equation and obtain the optimum value for a. However, it is quite difficult to give a practical close form. We should simplify it.

Case 1. In most cases, the ACF of inter frame signals can be modeled as exponential curves[18]. The signal with such characters can be achieved by passing a white Gaussian noise through a first order IIR filter.

Assume variance-normalized ACF of signal v_t and w_t be $\rho_v(\Delta) = \alpha^{|\Delta|}$, $\rho_w(\Delta) = \beta^{|\Delta|}$, then eq. (28) becomes

$$\begin{aligned} &Var[Cor(v_t', w_t')] \\ &= \frac{R_v(0)R_w(0)}{N}\left\{\frac{2}{N}\cdot\frac{(\alpha\beta)^{N-1} - N + (N-1)(\alpha\beta)^{-1}}{[1-(\alpha\beta)^{-1}]^2}\right. \\ &\quad\cdot[(1+a^2) - a\cdot\alpha^{-1} - a\cdot\alpha]\cdot[(1+a^2) - a\cdot\beta^{-1} - a\cdot\beta] \\ &\quad\left. + (1+a^2-2a\alpha)(1+a^2-2a\beta)\right\}. \end{aligned} \tag{29}$$

Let

$$\begin{aligned} f(a) = &\frac{2}{N}\cdot\frac{(\alpha\beta)^{N-1} - N + (N-1)(\alpha\beta)^{-1}}{[1-(\alpha\beta)^{-1}]^2}\cdot[(1+a^2) - a\cdot\alpha^{-1} - a\cdot\alpha] \\ &\cdot[(1+a^2) - a\cdot\beta^{-1} - a\cdot\beta] + (1+a^2-2a\alpha)(1+a^2-2a\beta). \end{aligned} \tag{30}$$

Since $\frac{R_v(0)R_w(0)}{N}$ is a constant, in order to minimize $Var[Cor(v_t', w_t')]$, we only

need to minimize $f(a)$. The differential of $f(a)$ is

$$\frac{df(a)}{da}=\frac{2}{N}\cdot\frac{(\alpha\beta)^{N-1}-N+(N-1)(\alpha\beta)^{-1}}{[1-(\alpha\beta)^{-1}]^2}\cdot(2a-\alpha^{-1}-\alpha)$$
$$\cdot(1+a^2-a\cdot\beta^{-1}-a\cdot\beta)+\frac{2}{N}\cdot\frac{(\alpha\beta)^{N-1}-N+(N-1)(\alpha\beta)^{-1}}{[1-(\alpha\beta)^{-1}]^2}$$
$$\cdot(1+a^2-a\cdot\alpha^{-1}-a\cdot\alpha)\cdot(2a-\beta^{-1}-\beta)+(2a-2\alpha)$$
$$\cdot(1+a^2-2a\beta)+(1+a^2-2a\alpha)\cdot(2a-2\beta). \tag{31}$$

Let it equal to 0. Then we can obtain optimal value of a.

Though we can obtain a close form of optimal a, however, the expression is too complicated to have practical usage.

In order to have practical usages, we numerically calculate the optimal a with different parameters α and β. The relationships between optimal a and α are shown in Fig. 5.

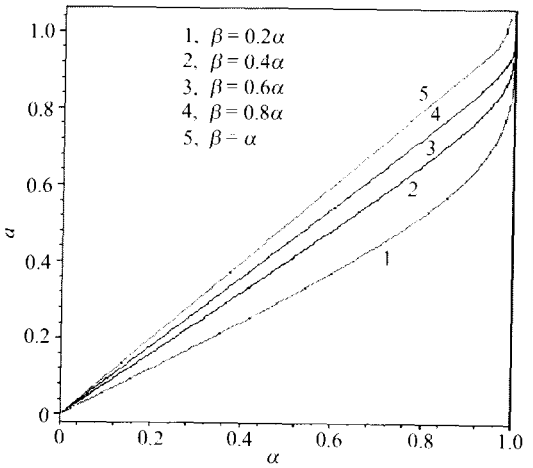

Fig. 5. Optimal a via α.

In most situations, the ACF of signals can be modeled as exponential curves. So we can choose the optimal whitening filter referring to Fig. 5.

Case 2. When $\rho_v(\Delta)=1$, then eq. (28) becomes

$$Var[Cor(v_t',w_t')]$$
$$=\frac{1}{N}R_v(0)(1-a)^2\cdot\left[\frac{2}{N}R_v(0)\sum_{\Delta=1}^{N-1}[((1+a^2)R_w(\Delta)-aR_w(\Delta-1)-aR_w(\Delta+1))]\right.$$
$$\left.+((1+a^2)R_w(0)-2aR_w(1))\right]. \tag{32}$$

Obviously $a=1$ make eq. (32) equal to 0, that is, $a=1$ minimize eq. (28).

This means that when $\rho_v(\Delta)=1$, then the optimal whitening filter is always $G(Z)=1-Z^{-1}$ in spite of any watermark sequence.

6 Experimental results

In the following experiments, we use a Philips standard movie and the standard "Saleman" as test videos, which are in YUV format and its Y component is at resolution of 128×128 pixels per frame and 8 bits per pixel. Philips movie consists of different scenes and contains big movements, it has 12300 frames. "Salesman" has 12572 frames and contains small movements.

In the experiments, $Am = 512$, $N = 4000$. Without any attacks, we can successfully extract the watermark, and the watermarked videos do not visually flick. The average PSNR of the whole watermarked Philips sequence $PSNR = 48.6$ dB and the average PSNR of "Salesman" is $PSNR = 49.1$ dB.

In the following, some experimental results are presented to test the scheme presented in the paper.

Experiment 1. In the experiment, we evaluate the design and the performance of the optimal whitening filter.

According to Theorem 2, we first calculate the normalized ACF of v_t for Philips movie, they are

$$\rho_v(0) = 1,\ \rho_v(1) = 0.9965,\ \rho_v(2) = 0.9919,\ \cdots$$

So $\rho_v(\Delta) \approx 1$, according to Case 2 of Theorem 2, the optimal whitening filter is $G(Z) = 1-Z^{-1}$, i.e. optimal $a = 1$.

In order to evaluate the performance of the designed whitening filter, we vary the value of a, calculate $Cor(v'_t, w'_t)$ and draw its distribution curves. By comparing the shape of curves, we can evaluate the performance. Fig. 6 summarizes the results.

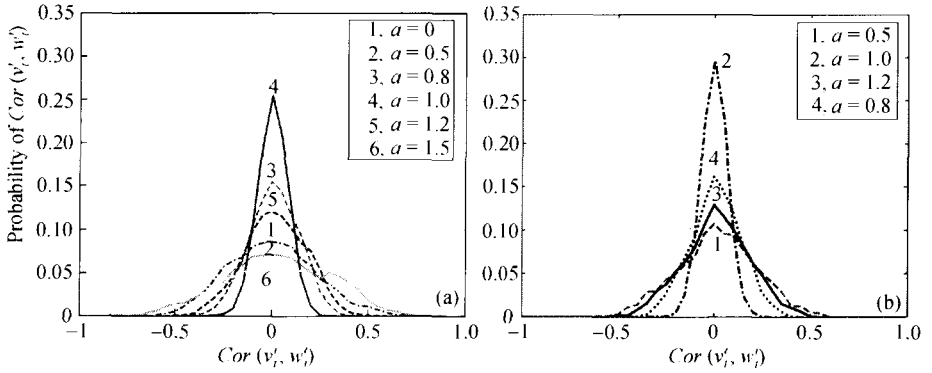

Fig. 6. The distribution curves of $Cor(v'_t, w'_t)$ using whitening filter with varying value of a. (a) For Philips test movie; (b) for "salesman".

Obviously, the situation with a=1 achieves the best performance in our case. The result coincides with our theory and design.

For "salesman", we get similar results.

Experiment 2. We attack the system by respectively rotating every frame 45 degree, by shifting 20×20 pixels, and by scaling as 64×64. In the first two attacks, we maintain

the frame size the same as the original 128×128, so some parts may be padded with black and some pixels are cropped.

Fig. 7 shows $Cor(v_t', w_t')$ along temporal axes in three cases. In rotation and shifting, the correlation value in the correct position decreases a little, but we can still extract the bits.

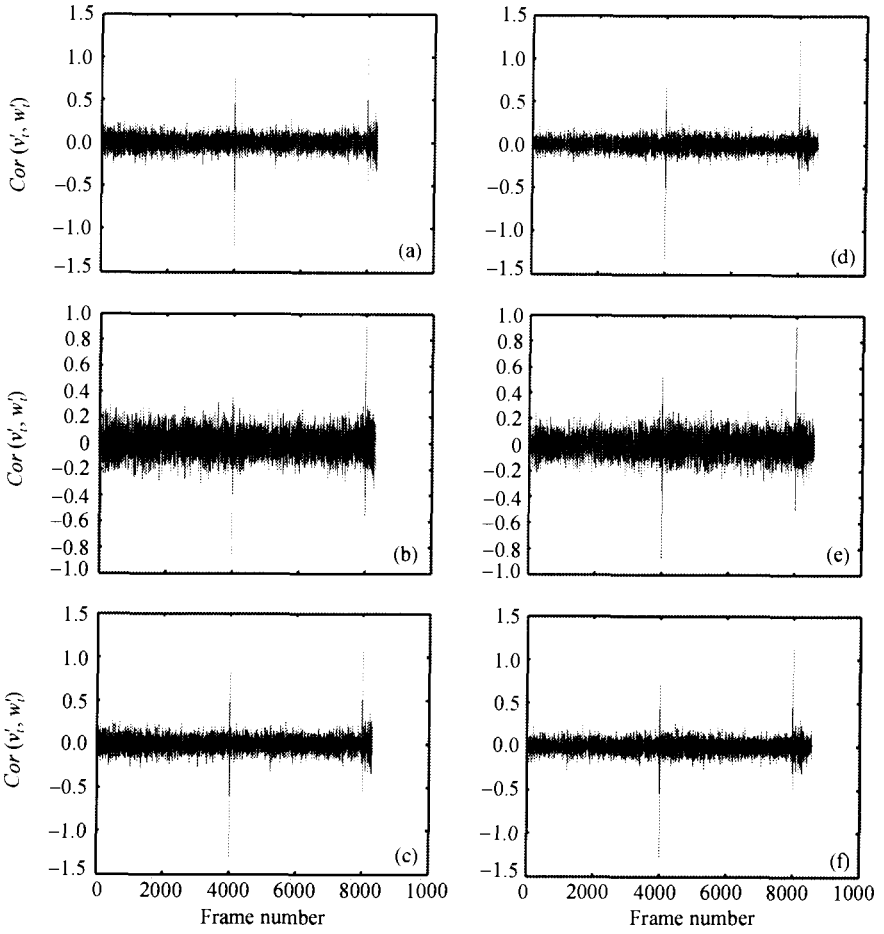

Fig. 7. The correlation after attacks. (a) For Philips after rotation 45 degree; (b) for Philips after shifting 20×20 pixels; (c) for Philips after scaled as 64×64; (d) for "salesman" after rotation 45 degree; (e) for " salesman" after shifting 20×20 pixels; (f) for "salesman" after scaled as 64×64.

Experiment 3. We also respectively attack the system by low-pass filtering AAE along the temporal axes with a low-pass filter $H(z) = 0.5 + 0.5z^{-1}$ and low-pass filtering every frame with a filter $h = \frac{1}{5}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. Our system still successfully extracts the watermark. Fig. 8 shows $Cor(v_t', w_t')$ along temporal axes in the two cases.

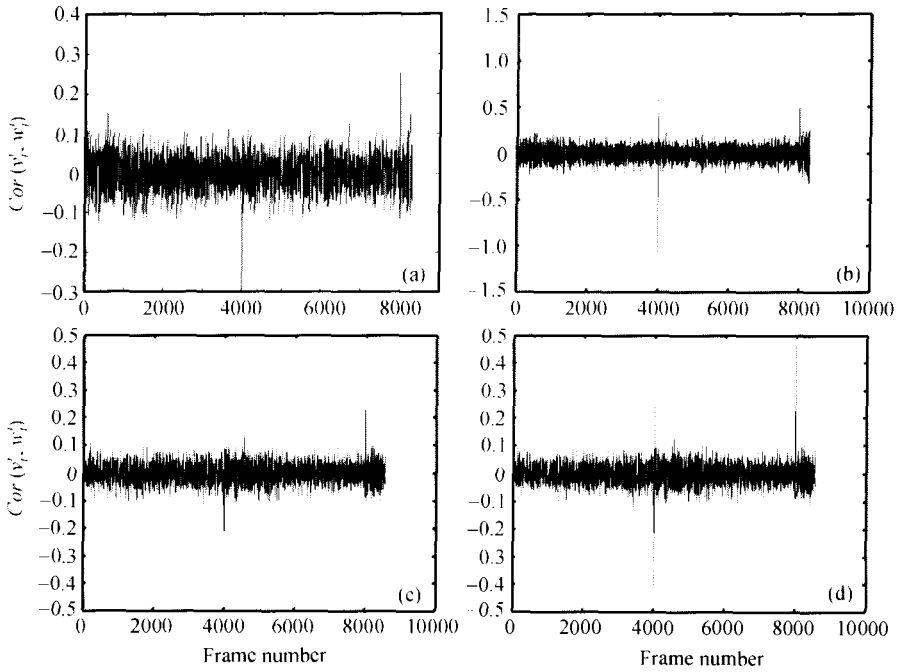

Fig. 8. $Cor(v_i', w_i')$ after temporal and spatial low-pass filtering. (a) For Philips after temporal low-pass filtering; (b) for Philips after spatial low-pass filtering; (c) for "salesman" after temporal low-pass filtering; (d) for "salesman" after spatial low-pass filtering.

7 Conclusion

The scheme presented in this paper has the following properties:

(1) The paper uses a new geometrical invariant, i.e. average AC energy (AAE) to combat arbitrary geometrical attacks.

(2) Since the scheme modifies only the AC energy instead of DC energy, the watermarked sequence does not visually spark.

(3) The paper proposes the optimal whitening filter and derives some useful results, which can significantly improve the detection performance.

(4) In order to access the watermark randomly from any frame and resist frame removal attack, synchronization bits are embedded alternately with the meaningful watermark.

Acknowledgements The author would like to thank Prof. Reginald L. Lagendijk (Delft University of Technology) and Prof. Ton Kalker (Philips Research, Eindhoven) for their constructive comments, discussion and suggestions. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 60373028 and 90604032), Specialized Research Fund for the Doctoral Program of Higher Education and the Program for New Century Excellent Talents in University.

References

1 Cox I J, Kilian J, Leighton T, et al. Secure spread spectrum watermarking for multimedia. Technical Report 95-10, NEC Research Institute. 1995

2 Schyndel R G van, Tirkel A Z, Osborne C F. A digital watermark. In: Proc IEEE Int Conf on Image Processing (ICIP'94). Austin, 1994, II: 86－90

3 Petitcolas F A P, Anderson R J, Kuhn M G. Attacks on copyright marking systems. In: Aucsmith D, ed. Proc Workshop Information Hiding. Berlin: Spriner, 1998. 15－17

4 Bas P, Chassery J M, Macq B. Geometrically invariant watermarking using feature points. IEEE Trans on Image Processing, 2002, 11(9): 1014－1028

5 Cox I J, Kilian J, Leighton T, Shamoon T. Secure spread spectrum watermarking for multimedia. IEEE Trans on Image Processing, 1997, 6(12): 1673-1687

6 Davoine F, Bas P, Hebert P A, Chassery J M. Watermarking et résistance aux déformations géométriques. In: Coresa99, 1999

7 Lin C Y, Chang S F. Distortion modeling and invariant extraction for digital image print-and-scan process. In: Proc Int Symp on Multimedia Information Processing (ISMIP 99), 1999

8 Lefebvre F, Gueluy A, Delannay D, Macq B. A print and scan optimized watermarking scheme. In: Dugelay J -L, Rose K, ed. Proc IEEE Fourth Workshop on Multimedia Signal Processing. Piscataway: IEEE, 2001. 511－516

9 Pereira S, Ruanaidh J, Deguillaume F, Cscurka G, Pun T. Template based recovery of fourier-based watermarks using log-polar and log-log maps. In: Proc Int Conf on Multimedia Computing and Systems. Piscataway: IEEE, 1999

10 Herrigel A, Voloshynovskiy S, Rytsar Y. The watermark template attack. In: Wong P W, Delp E J, eds. IS&T/SPIE'S 13th Annual Symposium, Electronic Imaging: Security and Water-marking of Multimedia Content III, 2001. 23－27

11 Kutter M. Watermarking resisting to translation, rotation and scaling. In: Proc SPIE Int Symp on Voice, Video, and Data Communication, 1998, 3528: 423—431

12 Oruanaidh J, Pun T. Rotation, scale and translation invariant spread spectrum digital image watermarking. Signal Processing, 1998, 66(3): 303－317

13 Haitsma Jaap, Kalker Ton. A watermarking scheme for digital cinema. In: Proc Int Conf for Image Processing. USA: IEEE, 2001. 487－489

14 Zhao Yao, Lagendijk R L. Video watermarking scheme resistant to geometric attacks. In: Proc Int Conf on Image Processing. USA: IEEE, 2002, 2: 145－148

15 Voloshynovskiy S, Deguillaume F, Pun T. Content adaptive watermarking based on a stochastic multiresolution image modeling. In: Proc European Signal Processing Conference (EUSIPCO2000), 2000

16 Cornsweet T N. Visual Perception. New York: Acadmic Press, 1970

17 Depover G, Kalker T, Linnartz J -P. Improved watermark detection reliability using filtering before correlation. In: Proc Int Conf for Image Processing. USA: IEEE, 1998. 430－434

18 Jayant N, Noll P. Digital Coding of Waveforms. NJ: Prentice Hall, 1984