An Integrated Technique for Video Watermarking

Chien-Chuan Ko, and Bo-Zhi Yang

Dep.of Computer Science and Information Engineering of National Chia Yi University
Email: kocc@mail.ncyu.edu.tw

Abstract

The major goal of this study is to embed watermarks into a video file. The proposed scheme embeds watermarks into image block which contains features with high intensity, high texture, and fast motion, and improves the robustness of the embedded watermarks based on spatial domain. Since human visual system can’t sense the variations of these feature regions within the video frames, an adaptive watermarking technique was developed to embed watermark into these regions. To this end, a digital video is first divided into various frames consisting of several blocks which are transformed by discrete wavelet transform. Spread spectrum incorporated with just notice difference is utilized to embed the watermark into the feature blocks or non-feature blocks. The proposed method can resist attacks performed by linear transformations including average, frame reduction, and frame shuttle, and obtain better performance in comparison with traditional schemes.

1. Introduction

In recent years, intellectual property protection using digital watermarking has become one of very attractive research field. The authentication of the digital products not only offers the traditional registered trademark directly, but provides a good protection mechanism- "digital watermark" [1].

To authenticate the digital media, a digital watermarking technique based on feature domain, spread spectrum, and discrete wavelet transform was proposed to embed a watermark into a static image even into a video file. The proposed approach not only can meet requirements for a robust watermark, but also detects malicious tamper with effectively. In comparison with traditional techniques, the proposed approaches can obtain better performance for some malicious attacks. Above all, the watermark detection and extraction does not need the original video.

Furthermore, it can embed different watermarks into different shot based on a simple shot detection scheme.

In general, watermark embedding domain can be categorized into three domains: frequency domain, spatial domain, and feature domain [2]. In order to obtain better robustness and resist possible malicious attacks performed by image processing operations, P. M. J. Rongen et al. [3] proposed a watermark embedding scheme based on feature domain by detecting salient points. In 2003, Qing-Ming Ge et al. [4] proposed a method based on DCT which identify regions with high texture and fast motion, and hided a watermark into the second, the third and the fourth bit-plane. Basically, it can also be categorized as feature-domain watermarking. The major reason for adopting feature domain watermarking is due to the fact that blocks containing high brightness, high textured, and fast motion are perceptually most insignificant; hence they are well suited for watermark embedding.

In this paper, we proposed a mixed watermark-embedding method which detects the features in the DWT domain, but embeds the watermark in the spatial domain in terms of the detected features. The video watermarking method is an improved and an extension version of the above DCT feature methodology, but it has better performance compared to the traditional approach. First, DWT is adopted to detect spatial feature and frequency feature so that it can be more suitable for human visual system. Second, brightness feature is added so as to detect abrupt intensity changes within each block. Third, spread spectrum technology [5] is used to detect the watermark without original media. Fourth, JND check [6] enables the transparency of the watermark, adaptively controls the scaling factor, and improves the image quality.

2. Methodology

In this paper, we exploit discrete wavelet transform, feature domain and statistic calculation such as entropy
to detect feature blocks with high brightness, high texture and fast motion.

2.1 Main Architecture

The stages of the video watermarking system including DWT, feature block, JND, spread spectrum, watermark embedding and watermark detection is shown in Fig. 1. The feature detection stage is developed based on brightness, texture, and motion. Furthermore, this system exploits spread spectrum incorporated with JND to develop an adaptive spread spectrum embedding technique and finally uses correlation function to detect the watermark response.

Fig. 1. Schema of the proposed video watermarking system.

2.2 Detection of Feature Block

As mentioned above, the watermark embedding was performed alternatively to assure the security of the watermark. To this end, it is necessary to enhance the watermark signal in the specific block based on feature domain. Here, blocks on each frame are performed by block-based DWT and then are determined based on the selected features. The feature extraction exploits brightness energy, texture energy and motion energy to determine the candidate blocks. Three thresholds corresponding to these energy components including TH_B, TH_T, and TH_M are empirically determined. If each energy component is greater than the corresponding threshold, the candidate blocks are considered as feature blocks which provide more strong evidences for embedding the watermark signal. These energy components are illustrated as follows.

Assume that there are \( f \) frames in the digital video and each frame has a spatial resolution of \( m \times n \) pixels. The \( i \)-th frame is denoted as \( V_i \), where \( i = 0 \) to \( f - 1 \). \( V_i \) is decomposed into a non-overlapping block with \( p \times q \) coefficients, and each block \( j \) within the frame \( i \) is denoted as \( B_{ij} \), where \( j = 0 \) to \( [m \times n / p \times q - 1] \). The coefficient in each block \( B_{ij} \) is transformed by using discrete wavelet transform based on tri-level decomposition. The transferred blocks are denoted as \( b_{ij} \). The coefficient in each transferred block is scanned by \( z \)-scan order and is defined as \( c_{ij,k} \). As a result, there are ten sub-bands in each transferred block including LL3, HL3, LH3, HH3, LL2, LH2, HH2, HL1, LH1 and HH1 respectively.

2.2.1 Measurement of brightness feature

Tri-level DWT is performed on the selected blocks of each frame in the video sequence. Ten sub-bands \{LHz, HLz, HHz, LL3, \( z = 1 \sim 3 \)\} are hence obtained, where \( z \) is the order of the sub-band. Transferred coefficients corresponding to these sub-bands are recorded with a \( z \)-scan order, and are denoted as \( c_{ij,k} \), where \( i \) is the \( i \)-th frame, \( j \) is the \( j \)-th block, and \( k \) is the \( k \)-th coefficient in terms of the \( z \)-scan order, respectively.

Rearranging \( c_{ij,k} \) by spectrum sequence, all coefficients are denoted as \( e_{ij,k}(z) \), where \( i \) is frame number, \( j \) is block number, and \( k \) denotes each sub-band in the \( z \)-scan order. A sub-band contains a horizontal region (LHz), a vertical region (HLz), and a diagonal region (HHz), or it may only contain a low-pass region (LLz).

According to the definition of multi-resolution based on DWT, LL3 coefficients can denote the most important feature of the selected block within each frame. Therefore, the brightness energy can be determined by

\[
BE(i,j) = \frac{64}{p \times q} \sum_{k=0}^{64} e_{ij,k}(3) \quad \text{........................................(1)}
\]

2.2.2 Measurement of texture feature

Wavelet analysis is a useful tool for description of image features. With respect to frequency domain, wavelet domain can not only preserve properties of frequency domain, but also properties of spatial domain. Chang T. et al. proposed wavelet tree to construct architecture of texture analysis [8]. To measure the complexity of the texture energy, the distribution of edges needs to be identified. The block may be classified as a block with complicated texture if there are numbers of edges in the frame and the distribution of these edges is non-uniform. To identify edge point and its distribution, edge detection is an intuitive image processing operation. Most of edge detection schemes detect all sharp transitions of image intensity. To describe complexity of texture in wavelet
domain, *Haar* function is selected as a basis function for the wavelet transform.

According to the nature of multi-resolution derived from wavelet decomposition, LH1, and HL1 represent a horizontal gradient, and a vertical gradient, respectively; HH1 has both natures of them. Therefore, LH1 and HL1 denote a vertical texture, and a horizontal texture, respectively; HH1 contains both natures of them. With multi-scale edge detection and complexity of texture analysis, texture energy can be measured by the variance of absolute value of LH1, HL1 and HH1 coefficients in each block. The detection of the texture energy of each block is defined by Eq. 2, where $M_{ij}$ is the mean value of first order sub-band coefficients of each block. The mean value $M_{ij}$ is denoted by Eq. 3.

$$ TE(i,j) = \frac{4}{3 \times p \times q} \sum_{k=0}^{(3p+q+1)} (|c_{i,j,k}(1)| - M_{ij})^2 \quad (2) $$

$$ M_{ij} = \frac{4}{3 \times p \times q} \sum_{k=0}^{(3p+q+1)} c_{i,j,k}(1) \quad (3) $$

This paper adopts *Haar* function to perform horizontal and vertical 2z decomposition for each block. The decomposition operations produce many wavelet blocks. Since entropy is a good measure of uncertainty, the entropy based on wavelet function $TE2(i,j)$ can be exploited to determine the uniformity of the texture. And, it can be defined by equation (4), where the range of $z$ spreads from level 1 to level 3, but not includes LL3 sub-band. Here, $Ph$ is the probability of the wavelet coefficient $h$ which spreads from -255 to 255. In practice the larger entropy is, the more uncertainty is. The measure also describes the number of varieties for the tested textures.

$$ TE(i,j) = - \sum_{b=0}^{3p+1} P(c_{i,j,k}(b)) \log(P(c_{i,j,k}(b))) \quad (4) $$

In order to efficiently detect the texture energy, we calculate a weighted texture energy $TE(i,j)$ by Eq. (5) incorporated with $TE1(i,j)$ and $TE2(i,j)$, where $i$ is i-th frame, $j$ is j-th block, $k$ is k-th block, $w1$, $w2$ are the weight of $TE1(i,j)$, $TE2(i,j)$ respectively, and $w1+w2=1$. If the texture energy is larger than the preset threshold $TH_T$, this block will be considered as a high textured block. Here, we use empirical value $w1=0.3$ and $w2=0.7$ as a weighted combination.

$$ TE(i,j) = w1 \times TE1(i,j) + w2 \times TE2(i,j) \quad (5) $$

### 2.2.2.3 Measurement of motion feature

Applying multi-resolution with $2^z$ dividing in the horizontal and vertical directions generates numbers of wavelet blocks with size of $p \times q$ pixels; then we can construct wavelet function. This stage adopts wavelet LL3 coefficients $(e_{i,j,k}(3))$ to measure the motion of the corresponding frames. Because LL3 can represent the most important region of an image, this paper adopts the difference of corresponding blocks within the video frames. This method is simple to describe the motion of objects within these frames. If a block size contains $p \times q$ pixels and $e_{ij,k}(3)$ represents LL3 wavelet coefficients, Eq. 7 denotes the motion energy $ME1(i,j)$ and this measure denotes the first evaluation of motion at local regions, where $k \in \{L\}$, $i$ is the frame number, $j$ is the block number and $k$ is the z-scan number of each sub-band.

$$ ME(i,j) = \frac{64}{p \times q} \sum_{k=0}^{(p \times q \times 1)} |e_{i,j,k}(3) - e_{i-1,j,k}(3)| \quad (7) $$

This stage also adopts ratio-log image to generate block-based wavelet LL3 coefficients, and can detect object motion efficiently. Eq. 7 represents the measure of ratio-log using LL3 coefficients to obtain motion energy of $ME2(i,j)$, and selects this measurement as the second measure of the motion energy.

$$ ME(i,j) = \frac{64}{p \times q} \sum_{k=0}^{(p \times q \times 64)} \log(e_{i,j,k}(3)/e_{i-1,j,k}(3)) \quad (7) $$

Because wavelet entropy can represent the variety of texture, this paper exploits it to evaluate the motion changes. Eq. 8 measures the difference of wavelet entropy $ME3(i,j)$; it is selected as the third measure of the feature. Since entropy can describe the varieties of edge distribution, we exploit it as a measure to detect motion changes based on texture feature.

$$ ME(i,j) = ME1(i,j) - ME2(i,j) \quad (8) $$

Finally, this study adopts weighted texture energy $ME(i,j)$ by Eq. 9 in order to determine blocks with motion feature, where $i$ is the i-th frame, $j$ is the j-th block, $w1$, $w2$ and $w3$ are the weight of $ME1(i,j)$, $ME2(i,j)$, $ME3(i,j)$ respectively, and $w1+w2+w3=1$. If the motion energy is greater than a preset threshold $TH_M$, this block will be considered as a fast motion block. Here, we choose $w1=0.2$, $w2=0.3$ and $w3=0.5$ as weights for measurement. The threshold corresponding to the energy is determined to identify the status of the selected block (i.e. a static block or a motion block).

$$ ME(i,j) = w1 \times ME1(i,j) + w2 \times ME2(i,j) + w3 \times ME3(i,j) \quad (9) $$

### 2.3 Watermark Embedding and Detection

Watermark embedding and detection process is illustrated in Fig. 2. To create an invisible watermark, a color watermark signal needs to be encrypted by a random generator seed, and used the seed as a security key to produce a string $S$ with a length of $r \times s \times 3 \times 8$ bits, where $r$, $s$ is height, and width of the watermarked image, respectively. The bit sequence $S$ is denoted as $S = \{s_i | s_i = 0 or 1, i =1 \text{ to } r \times s \times 3 \times 8\}$. Here, the Rabin public key encryption system is adopted due to its high security and simplicity to generate the watermark [5].
To generate a key sequence, we first choose two large prime numbers \( u \) and \( v \), and calculate \( N = u \times v \). The Rabin public key system encrypts a plaintext \( W \) to a ciphertext \( S \) by
\[
S = W^2 \mod N \quad \text{....................................................... (10)}
\]
To get the plaintext \( W \) from the ciphertext \( S \), \( u \) and \( v \) must be determined first.

**Fig. 2.** Stages of Watermark embedding and detection

However, plaintext \( W \) must be equal to one of the roots in order to get the key sequence. According to the Rabin public key system aforementioned, a pseudo random number generator with security can be designed by [5], where \( X_0 \) is the seed of the random number generator, \( N \) equals to \( u \times v \), and \( N \) is a very large number. \( R \) is a modulus factor to minimize the range domain. The binary bit sequence \( S \) is converted into the hash image \( h_k \) with a bipolar form \( h_k \in \{-1, 1\} \) which is generated by using the Rabin public key system [5].

### 2.3.1 Watermark Embedding

The stages of the entire watermark embedding are illustrated in Fig. 2. Each original frame is divided into RGB channel, respectively, and is transformed with block-based DWT. JND of each original frame \( i \) is then measured using the approach proposed by Barni et al. [6]. Each JND coefficient \( k \) for each block \( j \) of a frame \( i \) is defined as \( \text{JND}_{i,j,k} \). This stage detects the feature in DWT domain in terms of the feature domain mentioned above. The adaptability is controlled by calculating the sum of JND coefficients divided from the absolute sum of frame coefficients. An initialized scaling factor \( J A_i \) with JND check is determined by
\[
J A_i = \frac{\sum_{j=0}^{\text{StartBlock}} \sum_{k=0}^{\text{EndCoefficient}} \text{JND}_{i,j,k}}{\sum_{j=0}^{\text{StartBlock}} \sum_{k=0}^{\text{EndCoefficient}} |V_{i,j,k}|} \quad \text{....................................................... (11)}
\]
where \( i \) is the frame number.

When we obtain the scaling factor of \( J A_i \), all frames’ initialized scaling factors are defined. The embedding algorithm is described as below.

For \( i = \text{StartFrame} \) to \( \text{EndFrame} \)
For \( j = \text{StartBlock} \) to \( \text{EndBlock} \)
If \( V_{i,j,k} \in \text{Feature Blocks} \)
\[
\alpha_{i,j} = R H \times J A_i
\]
Else
\[
\alpha_{i,j} = R L \times J A_i
\]
End if
For \( k = \text{StartCoefficient} \) to \( \text{EndCoefficient} \)
Embed video block coefficients \( V_{i,j,k} \)
Next \( k \)
Next \( j \)
Next \( i \)

**Fig. 3.** Watermark embedding algorithm based on JND

After the evaluation of energy of feature regions, feature scaling factor \( RH \) and non-feature scaling factor are determined; \( RH \) is greater than or equal to \( RL \). The coefficient \( \alpha_i \) is adaptively determined depending on the status of the selected block. For a feature block, we exploit \( \alpha_{i,j} = RH \times J A_i \) to embed the watermark signal into spatial coefficients of each frame. For each non-feature block, we use \( \alpha_{i,j} = RL \times J A_i \) to embed the watermark signal into spatial coefficients of each frame. Eq. 12 shows the embedding process based on spread spectrum, where \( V_{i,j,k} \) is the original video signal, \( V_{i,j,k}' \) is the watermarked video signal, \( w_{i,j,k} \) is the repeated hash image sequence \( h_k \). Finally, the watermark signal \( w_{i,j,k} \) is then embedded into LSB (bit-plane 0) of the embedded frame coefficient \( V_{i,j,k} \) in order to detect malicious tamper with.
\[
V_{i,j,k}' = V_{i,j,k} + \alpha_{i,j} |V_{i,j,k}| w_{i,j,k} \quad \text{....................................................... (12)}
\]

### 2.3.2 Watermark Detection

To authenticate and detect the watermark, a correlation coefficient \( z \) is measured by Eq. 13, where \( w \) is a row vector of a repeated hash image sequence \( h_k \) with a size of \( m \times n \) bits, \( T \) represents a transpose operation, \( \vec{V}_i \) is a video signal which transforms i-th frame coefficients into a bit sequence, and \( M \) equals to video frame signal size of \( m \times n \). If \( z \)’s value is greater than the preset threshold \( TH_W \) which is calculated by Eq. 14, the watermark is demonstrated to be detected. For the sake of performance evaluation of watermark, this study adopts normalized correlation (NC) as shown in Eq. 15 as a measure of watermark detection, where \( w_i' \) is the bit sequence, which is reconstructed.
from the hash image sequence \( h_i \) for each video frame \( i \).

\[
Z_i = \frac{w^T(V_i)}{M} \tag{13}
\]

\[
\text{TH}_W = \frac{\sum_{j=0}^{\min} \sum_{k=0}^{\max} a_{j,k} |V_{i,j,k}| w_{j,k}}{2M} \tag{14}
\]

\[
\text{NG} = \frac{w^T w}{w^T w} \tag{15}
\]

3. Experimental results

As shown in Fig. 3, the table tennis video was first tested to authenticate the proposed architecture. Each frame of the video contains a size of 352×240 pixels. The watermark is the “msn” icon with a size of 27×27 pixels. The feature detection stage based on DWT can be performed to detect blocks with high energy; these blocks are shown in Fig. 4. Because feature regions have higher energy, using spread spectrum for watermarking embedding in these regions can increase robustness. Fig. 5 shows transparency and distribution of the watermark signal. The left one is an original frame, the middle one is the watermarked frame and the right one is the absolute difference of the watermark signal. The left one is an original frame, the middle one is the watermarked frame and the right one is the absolute difference of the watermark signal. Fig. 6 (a)-(i) depict PSNR values under all possible attacks including: (a) JPEG compression (10%), (b) sharpening, (c) median filter, (d) blurring, (e) drop noise, (f) brightness adjustment, (g) cropping, (h) Gaussian noise, and (i) MPEG compression (3Mbps). Fig. 7 shows that the watermark can be detected under possible attacks because the detector can response peak values and all are higher than the threshold. Fig. 8 depicts the detection caused by tampering with.

In comparison with the methods proposed by Q.M. Ge et al. [4] and X.W. Kong et al. [12], performed by many kind of attacks, in Fig. 8, the PSNR values are almost the same between Kong’s and ours, but Ge’s is smaller. In Table 1, the NC values are measured with Kong’s, GE’s and the proposed method. Our method performs a better quality than GE’s. In JPEG compression, NC value of our method is 0.36 and is smaller than GE’s (0.45) and Kong’s (0.55). In the aspect of MPEG compression, our NC value is 0.84 and is greater than GE’s (0.20) and Kong’s (0.66). For frame-average attack, the NC of our method is 0.99, and is better than Kong’s method (0.95) and GE’s method (0.44). In sharpening operation, NC of our method is almost the same as Kong’s, but is smaller than Kong’s in a median filter and blur operation. It also illustrates that our method may obtain worse performance in the operation of smoothing. In the drop-noise attack, brightness-adjustment attack, and cropping attack, our method has better quality and high fidelity.
In this paper, we proposed a watermark embedding scheme based on feature domain and spread spectrum for video data. This method can resist attacks of image processing such as JPEG compression, sharpening, median filter, blur, drop noise, brightness adjustment, cropping, Gaussian noise, and MPEG compression etc. It has the capability of detecting malicious tampering with. It can controls image quality and adjusts scaling factor for each frame using JND check adaptively. Moreover, our method can efficiently increase the capacity, and the robustness of the embedded watermark.

References


4. Conclusion