Robust video watermarking algorithm for H.264/AVC based on JND model

Weiwei Zhang[1*], Xin Li[1], Yuzhao Zhang[1], Ru Zhang[2] and Lixin Zheng[1]
[1] College of Engineering, Huaqiao University
Quanzhou, Fujian - CHN
[e-mail: weiweizh@hqu.edu.cn]
[2] National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications
Beijing, CHN
[e-mail: 2412565488@qq.com]
*Corresponding author: Weiwei Zhang

Received May 28, 2016; revised December 18, 2016; accepted February 22, 2017; published May 31, 2017

Abstract

With the purpose of copyright protection for digital video, a novel H.264/AVC watermarking algorithm based on JND model is proposed. Firstly, according to the characteristics of human visual system, a new and more accurate JND model is proposed to determine watermark embedding strength by considering the luminance masking, contrast masking and spatial frequency sensitivity function. Secondly, a new embedding strategy for H.264/AVC watermarking is proposed based on an analysis on the drift error of energy distribution. We argue that more robustness can be achieved if watermarks are embedded in middle and high components of 4×4 integer DCT since these components are more stable than dc and low components when drift error occurs. Finally, according to different characteristics of middle and high components, the watermarks are embedded using different algorithms, respectively. Experimental results demonstrate that the proposed watermarking algorithm not only meets the imperceptibility and robustness requirements, but also has a high embedding capacity.

Keywords: H.264/AVC, JND model, embedding strategy, robust watermarking

This work was supported by the Natural Science Foundation of China (Grant No. 61372107), the Research Foundation of Huaqiao University (Grant No. 13BS415), Science and Technology Development Foundation of Quanzhou City (Grant No. 2014Z112); The Natural Science Foundation of Fujian Province (Grant No. 2015J05125) ; Special Foundation of Fujian Provincial Department of Science and Technology (Grant No. 2013H2002).

https://doi.org/10.3837/tiis.2017.05.024 **ISSN : 1976-7277**

1. Introduction

In recent years, with the rapid development of HD digital TV and VOD (video on demand) technology accompanied by the popularity of multimedia sharing site, video multimedia has increasingly become a significant part in people's daily life. Thus, efficient methods for copyright protection of digital video have been research hotspots in the academic field of multimedia security [1]. Digital watermarking technology is one of the most effective and efficient techniques to protect copyright under the open network circumstance [2, 3]. Because of the large amount of data, video needs to be compressed in the process of storing and transmission. H.264/AVC is designed for higher video compression efficiency and also to be network friendly. It has been widely used for various video applications such as video broadcasting, video conference and video streaming etc [4]. Thus, it makes the demand for video watermarking based on H.264/AVC increasingly urgent.

Designing a feasible robust watermarking scheme can be regarded as an optimization problem with these contradictory goals: better robustness and higher imperceptibility [5]. In order to balance robustness and imperceptibility, it is necessary to take the characteristic of human visual system (HVS) and embedding strategy into consideration when engaging in watermarking research.

By taking into account the HVS properties in the design of watermarking algorithms, we can maximize robustness when placing a perceptual constraint on transparency. Human eyes cannot sense any changes below the just noticeable difference (JND) threshold around a pixel due to their underlying spatial/temporal sensitivity and masking properties [6]. By means of JND thresholds, the maximum strength of a watermark can be varied to be coarser in areas with higher JND threshold values and finer otherwise. In this respect, several perceptual models have been successfully exploited for watermarking. Wolfgang et al [7] used the visual models developed in [8] for still images to find the JND for each DCT coefficient of digital images and video, based on frequency sensitivity, luminance masking, and contrast masking. They embedded the watermark only in those coefficients that exceeded their JNDs. Simitopoulos et al [9] proposed a watermarking scheme that embeds the watermark directly into the MPEG stream. To make the watermark imperceptible, perceptual analysis [8] and block classification techniques were combined. Another very important aspect affecting human perception, visual saliency, was introduced to modulate JND model in [10]. The authors proposed the saliency modulated JND profile guided image watermarking scheme based on the visual saliency's modulatory effect on JND model which incorporates visual attention's influence on visual sensitivity. Unfortunately, the JND model based on 8×8 floating-point DCT of the proposed watermarking algorithms [7-10] cannot be applied directly to H.264/AVC because different integer DCT are applied in H.264/AVC instead of the floating-point DCT.

Up to now, the compressed-domain video watermarking embedding position schemes based on H.264/AVC can be generally classified into three groups: the first depends on the DCT coefficients [11-13], the second relies on the motion vectors [14-16] and the third works on the entropy coding domain [17,18]. Watermarking DCT coefficients has the advantage of being more robust to attacks and distortions [19]. But the key issue is where should watermarks be embedded in 4×4 integer DCT domain in order for the invisible video watermarks to be robust. Many papers [20-23] in the literature agree that watermarks should be embedded in perceptually significant components, such as dc and low components in 4×4 integer DCT

domain for H.264/AVC video. However, our analysis and experiments indicate that the watermarks may be more robust if they are embedded in middle and high components.

In this paper, we propose a novel H.264/AVC video watermarking algorithm based on a modified JND model which includes luminance masking, contrast masking and spatial frequency sensitivity function. Then, we propose a new embedding strategy based on 4×4 integer DCT for H.264/AVC video watermarking. Finally, we use different algorithms to embed the watermarks in the middle and high components, respectively.

The remainder of this paper is organized as follows. In Section 2, we derive the JND model for a 4×4 integer DCT. The watermark embedding and extraction schemes are presented in Section 3. Experimental results are given in Section 4. Section 5 concludes the paper.

2. JND model based on H.264

HVS cannot perceive all the changes in the images due to its underlying physiological and psychological mechanism, so a lot of perceptual redundancies exist in the processed images and videos [27]. Perceptual watermarking should take full advantage of the results from HVS studies. JND gives us a way to model the HVS accurately. A few recently published papers have discussed embedding a watermark in the H.264 video based on the JND model. Noorkami et al [24] employed a human visual model adapted for a 4×4 discrete cosine transform block to increase the payload and robustness while limiting visual distortion. In their proposed scheme, a key-dependent algorithm was used to select a subset of the coefficients that have visual watermarking capacity. Su et al [25] take into account the human visual characteristics to guarantee the imperceptibility of watermark signals in their proposed H.264/AVC watermarking algorithm. Li et al [23] proposed an H.264/AVC HDTV watermarking method that is robust to camcorder recording. Because contents in the consecutive frames of a video are almost identical, they embedded the copyright information by fine-tuning the luminance relationship of the consecutive frames. However, the proposed schemes [23,24,25] only used the luminance masking to established a JND model. Wang et al. [26] proposed a 4×4 block DCT JND model which includes a spatial frequency sensitivity function, a luminance masking adaptive factor, and a contrast masking factor. In their proposed JND model, the space frequency sensitivity function is directly obtained from the modified JND model proposed by Wei [27]. They simply change the block size N, the parameters used in determining block type. However, as Naccari et al [28] pointed out, Wei's [27] parameter of space frequency sensitivity function of JND model, carried out by means of subjective tests, regards only the 8×8 floating point DCT. Therefore, an extension of these models for the 4×4 and 8×8 integer DCT used in H.264/AVC would require new subjective testing. That means the accuracy of JND threshold for H.264/AVC watermarking proposed by Wang [26] needs to be further improved.

JND in video signals depends on spatial masking effects which is mainly due to the following three aspects [28]: 1) the type of frequency representation (the transform) used; 2) the luminance variations; and 3) the presence of patterns such as textured regions which boost masking effects. In this paper, we propose a new JND model based on 4×4 integer DCT, which is given by

$$T_{JND} = T_{BASIC} \times F_{lum}(k,n) \times F_{contrast}(k,n,i,j) \tag{1}$$

where k is the index number of frames of a video sequence, n is a block index in the $k-th$ frame, and i and j indices of the DCT coefficients. T_{BASIC} and $F_{lum}(k,n)$ denote space frequency masking factor and adaptive luminance masking factor respectively. The $F_{contrast}(k,n,i,j)$ represents the contrast masking factor.

In [29], the default perceptual matrices adopted in the H.264/AVC reference software were empirically designed for 4×4 integer DCT. The intra-coding perceptual matrix is defined as:

$$JND_{band}^{intra}(\mathrm{i,j})=\begin{bmatrix} 6 & 13 & 20 & 28 \\ 13 & 20 & 28 & 32 \\ 20 & 28 & 32 & 37 \\ 28 & 32 & 37 & 42 \end{bmatrix}/16 \tag{2}$$

Hence, the frequency masking factor T_{BASIC} in this paper is shown as follows:

$$T_{BASIC}=\alpha \bullet \begin{bmatrix} 6 & 13 & 20 & 28 \\ 13 & 20 & 28 & 32 \\ 20 & 28 & 32 & 37 \\ 28 & 32 & 37 & 42 \end{bmatrix} \tag{3}$$

where α is the adjustment parameter of the frequency masking factor with the empirical value of 0.5, which is empirically decided by the user.

The HVS is more sensitive to the noise in medium gray regions, so the visibility threshold is higher in very dark or very light regions. Because our base threshold is detected at the 128 intensity value, for other intensity values, a modification factor needs to be included. This effect is called the luminance variations. The curve of the luminance variations is a U-shape which means the factor at the lower and higher intensity regions is larger than the middle intensity region, an approximation of this curve is proposed in [27,28] as follows:

$$F_{lum}(k,n)=\begin{cases} 1+[60-\bar{I}(k,n)]/150 & \bar{I}(k,n)\le 60 \\ 1 & 60<\bar{I}(k,n)<170 \\ 1+[\bar{I}(k,n)-170]/425 & \bar{I}(k,n)\ge 170 \end{cases} \tag{4}$$

where $\bar{I}(k,n)$ represents the luminance intensity for the $n-th$ block in the $k-th$ frame.

The contrast masking factor is determined by the edge energy of the image. On the one hand, noise is less visible in the regions where texture energy is high. On the other hand, the human eye is more sensitive to the change of the smooth and edge areas. Therefore, the contrast masking factor should be determined according to the texture energy of the image. As we know, Canny operator is a very effective edge detector for a given image, so we classify the

type of each 4×4 block according to Canny operations on each block. The block type is determined by (5).

$$\text{block type} = \begin{cases} \text{Plane}, & \rho \le \beta, \\ \text{Edge}, & \rho \le \beta \le \chi, \\ \text{Texture}, & \rho > \chi. \end{cases} \tag{5}$$

where ρ is the number of edge pixels in a given block with the edge map generated by the Canny operator. In our experiments, it is found that $\beta = 2$ and $\chi = 5$ work well for most images tried.

This pattern masking mechanism varies the JND threshold values. The experiment in [27] measures these variations for masker signals such as Gabor or sinusoidal patterns. The measurements led to a nonlinear relationship between the JND thresholds and the normalized masking energy, defined as the squared ratio between the masker (i.e., the image) and the masking signal (i.e., the quantization error) as follows:

$$F_{contrast}(n,i,j) = \begin{cases} \psi, & \text{for } (i^2 + j^2) \le 4 \text{ in Plane and Edge block,} \\ \psi \bullet \min(4, \max(1, (\dfrac{C(n,i,j)}{T_{BASIC} \bullet F_{lum}(k,n)})^{0.36})), & \text{others.} \end{cases} \tag{6}$$

where k is the index number of frames of a video sequence, n is a block index in the $k-th$ frame, and i and j indices of the DCT coefficients. $C(n,i,j)$ represents the 4×4 integer DCT coefficients of the current blocks. ψ is determined by the following:

$$\psi = \begin{cases} 1, & \text{for Plane and Edge block,} \\ 2.25, & \text{for } (i^2+j^2) \le 4 \text{ in Texture block,} \\ 1.25, & \text{for } (i^2+j^2) > 4 \text{ in Texture block.} \end{cases} \tag{7}$$

where i and j indices of the DCT coefficients.

3. Watermark embedding and extraction

3.1 Watermark embedding strategy

Cox et al. [30] argue that watermarks should be embedded in the perceptually significant components in order for the watermarks to be robust. Huang et al. [31] indicate that the watermarks may be more robust if they are embedded in dc components, according to their quantitative analysis and experimental results on the magnitude of DCT components. Some researchers [20,21,22,23] directly applied these conclusions made by [30] and [31] for invisible image watermarking to video watermarking based on H.264/AVC. We argue that

middle and high components in the 4×4 integer DCT based on H.264/AVC are more suitable for watermarking than any dc and low components.

Since re-encoding by H.264/AVC is a common video processing method, we explore the effect of re-encoding on 4×4 integer DCT coefficients. We assume that Y is the matrix of unscaled DCT coefficients corresponding to the residual block R. The 4×4 integer DCT between Y and R can be expressed in a matrix form as follows:

$$Y = C_f R C_f^T \tag{8}$$

where

$$C_f = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{9}$$

and R is defined by

$$R = X - X_p \tag{10}$$

where X represents the original 4×4 image matrix and X_p is the best prediction 4×4 image matrix.

Then, the post-scaling and quantization process is given as follows:

$$\overline{Y} = (Y .\times PF)./Q_{step} = C_f R C_f^T .\times PF./Q_{step} \tag{11}$$

where $.\times$ and $./$ denote the point multiplication and point division operation in matrix operation respectively and Q_{step} is the quantizer step size after first encoding. PF matrix is defined as follows:

$$PF = \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix}, a = \frac{1}{2}, b = \sqrt{\frac{2}{5}} \tag{12}$$

We assume that $R^{'}$ is the reconstructed 4×4 image residual matrix after inverse integer DCT and quantization. Then, we can obtain the reconstructed 4×4 image matrix by

$$X^{'} = (X_p + R^{'}) \tag{13}$$

According to (8), the integer DCT coefficients generated in the re-encoding process are as follows:

$$\overline{Y} = C_f(X^{'} - X_p^{'})C_f^T = C_f(X_p + R^{'} - X_p^{'})C_f^T \tag{14}$$

where $X_p^{'}$ represents the best prediction 4×4 image matrix during re-encoding process.

The post-scaling and quantization process during re-encoding is shown as follow:

$$\overline{Y^{'}} = ((C_f(X_p + R^{'} - X_p^{'})C_f^T).\times PF)./Q_{step}^{'} \tag{15}$$

where $Q_{step}^{'}$ represents the quantization step size during re-encoding.

The difference of quantified integer DCT coefficients during the first encoding and re-encoding is as follows:

$$\begin{aligned}\Delta Y &= \overline{Y} - \overline{Y^{'}} \\ &= C_f R C_f^T.\times PF./Q_{step} - ((C_f(X_p + R^{'} - X_p^{'}).\times PF)./Q_{step}^{'}) \\ &= C_f(R./Q_{step} - R^{'}./Q_{step}^{'} + (X_p - X_p^{'})./Q_{step}^{'})C_f{}^T.\times PF\end{aligned} \tag{16}$$

We let

$$D = (R./Q_{step} - R^{'}./Q_{step}^{'}) + (X_p - X_p^{'})./Q_{step}^{'} \tag{17}$$

Then we can get the difference of quantified integer DCT coefficients as follows:

$$\Delta Y = C_f D C_f^T.\times PF \tag{18}$$

From (17), we know that D is composed of two parts: $(R./Q_{step} - R^{'}./Q_{step}^{'})$ is the quantization error during the first encoding, and $(X_p - X_p^{'})./Q_{step}^{'}$ represents the difference of the best prediction 4×4 image matrix between the first encoding and re-encoding. As we known, the intra-prediction of H.264/AVC itself is not reversible. That is, the re-encoding of the intra pictures does not guarantee the same results as before. The intra-prediction modes in many blocks and the coefficient values are changed in re-encoding [32], which means $(X_p - X_p^{'})$ is not equal to 0, normally.

Then, equation (18) shows that ΔY is obtained using the 4×4 integer transform of D,which means that the energy of D is mainly distributed in the dc and low component in the ΔY . Therefore, dc and low components are easily influenced by the quantization error and the difference of the best prediction 4×4 image matrix between the first encoding and re-encoding. That means when embedding the same strength of watermark in integer DCT coefficients, the robustness of the dc and low is lower than middle and high components. The following experiment will prove the correctness of our conclusion.

5 video sequences in CIF format foreman, mobile, mother, hall, and flower are used to be test samples. The standard video sequences are encoded into 100 frames and with an intra-period of 10 frames. The H.264 encoder used a fixed quantization step parameter of 28 for all I frames. **Fig. 1** shows the proportion of the changed quantized DCT coefficients of the 5 video sequences after re-encoding. **Fig. 1** shows that the change ratio of dc coefficient and low frequency is higher than other ac coefficients. **Fig. 2** shows that when embedding the same strength watermark, there is higher error rate in dc and low components.

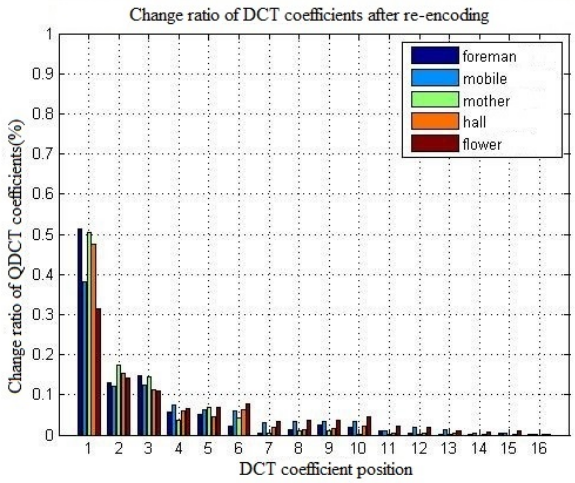


Fig. 1. The change ratio of DCT coeffcients after re-encoding

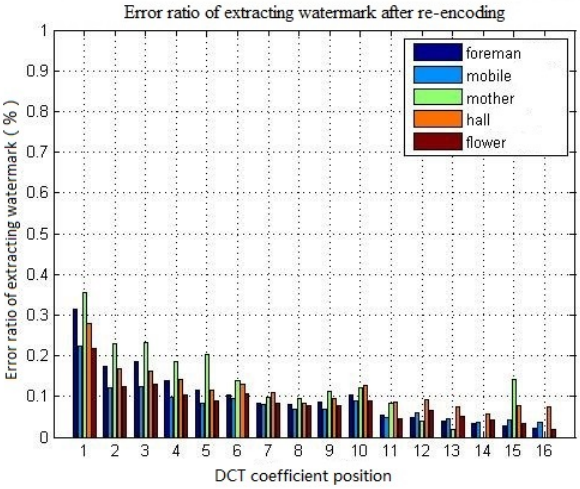


Fig. 2. Error ratio of extraction watermark after re-encoding

3.2 Watermark embedding

From the above analysis, we will choose the middle and high components which exceed the JND threshold to embed the watermark. There are mainly two kinds of watermark embedding algorithms in DCT domain for H.264/AVC. One kind of algorithm [19,21,33] changes the parity of DCT coefficients according to the watermark information, which is generally used in the low and middle component of DCT coefficient. This kind of algorithm has little impact on video quality, but poor robustness. Another kind of algorithm [25,34,35] increases the DCT coefficient or sets it equal to zero according to the watermark information. This kind of algorithm is generally used for high frequency components and has better robustness. Therefore, we apply two kinds of algorithms to embed the watermark to middle and high component, respectively. The flowchart of watermark embedding process is showed in **Fig. 3.**

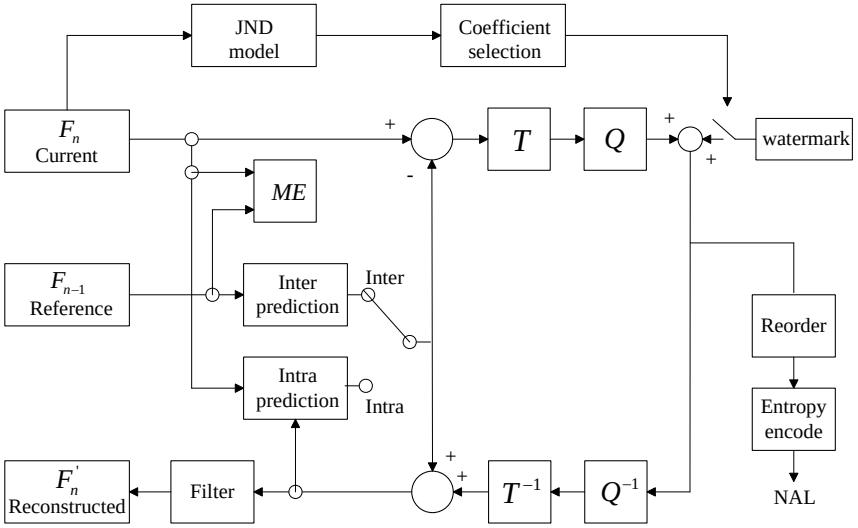


Fig. 3. The flowchart of watermark embedding process

The watermark embedding process is shown as follows:

step 1 Divide the current I-frame of video into 4×4 blocks, and implement integer DCT on each block. Then, calculate every JND threshold value of the integer DCT coefficient.

step 2 Analyze whether the JND threshold is larger than the quantized value. Since embedding watermarks in non-zero DCT coefficients will not significantly increase the bit rate, therefore, we analyze whether coefficient is greater than zero. If both of these two conditions are met, we embed the watermark. Then we set the watermark embedding position as a secret key.

When embedding the watermark in the middle frequency of the DCT coefficients, the algorithm is as follows:

If $W_i = 0$

$$cw_i' = \begin{cases} cw_i - 1, & if \quad \mathrm{mod}(cw_i, 2) = 1, \\ cw_i, & if \quad \mathrm{mod}(cw_i, 2) = 0, \end{cases} \tag{19}$$

If $W_i = 1$

$$cw_i^{'} = \begin{cases} cw_i - 1, & if \quad \mathrm{mod}(cw_i, 2) = 0, \\ cw_i, & if \quad \mathrm{mod}(cw_i, 2) = 1, \end{cases} \tag{20}$$

When in the high frequency, the algorithm is as follows:

$$cw_i^{'} = \begin{cases} \max\{cw_i, W_i\}, & if \quad W = 1, \\ 0, & if \quad W = 0, \end{cases} \tag{21}$$

where $cw_i^{'}$ is the DCT coefficient after embedding the watermark, cw_i is the DCT coefficients before watermarking, W_i is the spread spectrum watermark information.

step 3 Circulate Step1 and Step2, until all of the watermark information are embedded.

3.3 Watermark extraction

The process of extracting digital watermark is as follows:

step 1 Use the secret key to find the actual location of watermarks and extract the spread spectrum watermark information as follows:

In the middle frequency.

$$W_i^{'} = \begin{cases} 0, & if \quad \mathrm{mod}(cw_i^{'}, 2) = 0, \\ 1, & if \quad \mathrm{mod}(cw_i^{'}, 2) = 1. \end{cases} \tag{22}$$

while in the high frequency component:

$$W_i^{'} = \begin{cases} 0, & if \quad \left|cw_i^{'}\right| > 0, \\ 1, & else. \end{cases} \tag{23}$$

where $cw_i^{'}$ is the DCT coefficient after embedding the watermark, $W_i^{'}$ is the embedded spread spectrum watermark information.

step 2 Circulate Step1 and Step2, until all of the watermark information are extracted.

4. Experimental results and analysis

The experiment presents the performance of JND model and video watermarking. The proposed video watermarking algorithm has been implemented in the H.264/AVC reference software version JM12.4. The test sequences include *Foreman, Mobile, Coastguard, Hall, and Flower* (CIF (352×288) format). The video sequences are encoded into 100 frames at the frame rate 15frame/s and with an intra-period of 10. The GOP structure comprises IPP…,

compliant to the *Baseline Profile* of H.264/AVC. The H.264 encoder uses a fixed quantization step parameter of 28 for all I frames.

4.1 Comparison and analysis of JND model

JND models can avoid generating values larger than the actual HVS thresholds. A better JND model should yield larger JND values at a fixed perceptual quality. To evaluate the performance of the JND models, noise is added to each DCT coefficient in an image according to the JND values generated by the JND profiles:

$$C^{'}(n,i,j) = C(n,i,j) + f \times \alpha \times JND(n,i,j) \tag{24}$$

where $C^{'}(n,i,j)$ is the $n-th$ DCT coefficient in the image in which the JND noise is added and $C(n,i,j)$ represents the $n-th$ DCT coefficient in the image. f takes the value of +1 or -1 randomly. α is the noise intensity adjustment factor for JND.

(1) Objective evaluation

In this experiment, five different 512×512 grayscale images were chosen for testing. The PSNR is used to measure the distortion of the image under a variety of JND model noise. At the same perceived quality, a better model will achieve higher JND thresholds and result in lower PSNR. **Table 1** shows the comparison results of the PSNRs under the same conditions of the three JND profiles which have the same image quality by adjusting α. It can be found that the proposed method can tolerate more distortion among three models.

Table 1. Applications in each class

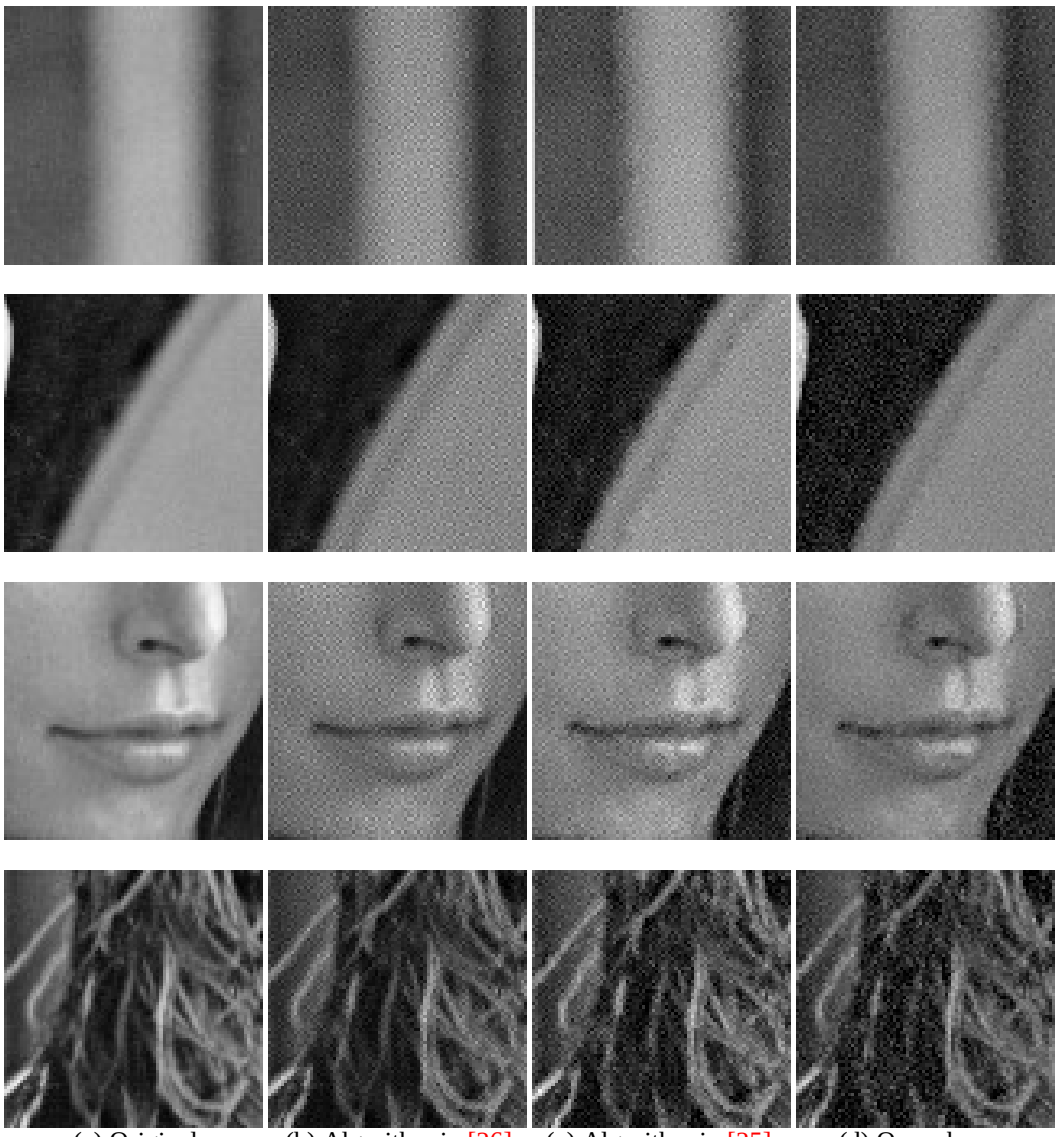
Image	Algorithm in [25] (dB)	Algorithm in [26] (dB)	Our scheme (dB)
Lena	31.8186	31.1472	30.7094
Baboon	31.9401	29.9835	28.9907
Barbara	31.8775	30.4589	29.8342
Man	31.8637	30.7364	29.9917
Peper	31.7803	31.3033	29.6914

Fig. 4 shows the noise-contaminated Lena and Barbara images created by three JND models. From **Fig. 4** (a), (b), (c) and (g), (h), (I), we can see that the human eye cannot see the distortion caused by the three kinds of model of the JND noise under 30dB. It illustrates the validity of the three models. Otherwise, **Fig. 4** (d) and (j) showcase that the noise of difference images under the model [25] are obvious on edge, smooth or textured area.

(a) (b) (c)

Fig. 4. Lena and Barbara images added JND noise with 30dB PSNR. (a)-(c), (g)-(i) respectively are Lena and Barbara images added JND noise images. (d)-(f), (j)-(l) respectively represent the different image between the original image and the image with the three kinds of JND noise.

From **Fig. 4** (e) and (k), it can be seen that the model in [26] has more noise in edge and texture. This is because the model of [26] added the Canny operator in contrast masking factor. The human eye is relatively insensitive to these areas, so the model of [26] is more consistent with the human visual system than the model of [25]. But the noise in smooth regions of the [26] model is obvious, which is caused by the frequency sensitivity factor of JND model. Since the [26] directly adopts frequency sensitivity function parameters for floating point 8×8 DCT transform proposed by Wei et al [27], it will cause more obvious noise in the 4×4 model. Noise of **Fig. 4** (f) and (l) in the edge and texture areas is more obvious, but the smooth area noise is relatively minor. **Fig. 4** illustrates that the 4×4 JND model proposed in this paper is more consistent with human visual characteristics than the models of [25] and [26].



(a) Original (b) Algorithm in [26] (c) Algorithm in [25] (d) Our scheme

Fig. 5. Comparison on the Lena image in detail under JND noise with 30 dB PSNR.

Fig. 5 is the details of Lena image which were added to the three JND model noise respectively. It can be seen from **Fig. 5** that the model presented in this paper has smaller noise in smooth areas than that of literature [25] and [26], but larger noise in texture and edge area than that of literature [25] and [26].

(2) Subjective evaluation

For a more convincing evaluation, the subjective viewing tests were conducted based on "Adjectival categorical judgement methods". In each test, two images were juxtaposed on the screen (left is the noise-injected image and right is the original image). Twenty subjects (ten are experts in image processing field and ten are naive) were asked to give quantitative scores for all the image pairs, using the continuous quality comparison scale shown in **Table 2**. In this experiment, the text equipment is a LEN1152 L197 Wide monitor. The viewing distance is four times that of the image height.

In the experiment, the five different images each had added one of the three JND model noises, respectively. But the observer was unaware of which kind of JND noise had been added in the images. The observer scored according to their perception of the difference between the original image and the image with JND noise. The experiment uses the widely used average evaluation score MOS (Mean Opinion Score) method, as shown in **Table 2**. Reviewers gave a specific score according to this standard. **Table 3** is the MOS average value obtained from 5 images with three different JND noises.

Table 2. Subjective score of image quality

Score	Describe
5	No damage or interference can be detected on the image.
4	There is slightly detectable damage or interference on the image.
3	There is obviously detectable damage or interference on the image.
2	There is relatively serious damage or interfere on the image.
1	There is very serious damage or interference on the image.

Table 3. Subjective average scores of 5 images added three kinds of JND noise

Image	Algorithm in [26] (dB)	Algorithm in [25] (dB)	Our scheme (dB)
Lena	4.17	4.67	4.75
Baboon	4.50	4.75	4.83
Barbara	3.42	3.58	3.83
Man	4.08	4.58	4.42
Peper	3.50	3.67	4

As seen from the results in **Table 3**, for most images, the proposed spatial masking procedure gains the highest scores except the “Man” image. The reason is that the Man’s face in the image contains many edge regions which is the region of interest, whereas our proposed algorithm and literature [26] enhance the JND noise in this region. Consequently, the subjective quality of the proposed algorithm and literature [26] is worse than that of literature [25]. However, the average score of the proposed procedure still outperforms [25] and [26], subjects showed a statistically significant preference for the images produced by the proposed JND model.

4.2 Analysis on watermark invisibility

Fig. 6 and **Fig. 7** demonstrate that the watermark embedded with the proposed algorithm is invisible, where **Fig. 6** (a) and **Fig. 7** (a) are the original 30th frames of “Foreman” and “Container” video sequences, respectively. And **Fig. 6** (b) and **Fig. 7** (b) are watermarked 30th frames, respectively. In the experiments, no visible artifacts can be observed in all of the test video sequences.

(a) original 30th frame of Foreman sequence (b) watermarked 30th frame of Foreman sequence

Fig. 6. Comparison of Foreman test sequences before and after embedding watermark (the rate of 768kbit / s)

(a) original 30th frame of Foreman sequence (b) watermarked 30th frame of Foreman sequence

Fig. 7. Comparison of Container test sequences before and after embedding watermark (the rate of 768kbit / s)

Table 4 demonstrates the comparison results in terms of bit-rate increase, perceptual quality using the PSNR, and the achieved capacity for different tested video sequences between the proposed algorithm and [36].

Table 4. Applications in each class

Class	**PSNR(dB)**			**Bit-rate(kbits/s)**			**The number of embedded watermark (bit)**	
Tested video	Original	Algorithm in [36]	Our scheme	Original	Algorithm in [36]	Our scheme	Algorithm in [36]	Our scheme
foreman	37.49	37.39	37.22	283.90	291.79	283.63	3960	6256
mobile	35.43	35.38	35.07	1260.20	1265.60	1276.03	3960	23128
coastguard	35.45	35.41	35.28	865.95	871.79	862.95	3960	9064
hall	38.30	38.20	38.07	206.89	214.07	206.86	3960	4392
flower	36.44	36.40	36.03	1096.27	1103.22	1115.25	3960	23020

As can be seen from the **Table 4**, the average watermark capacity of each I frame of 5 test video sequences of our proposed algorithm is 1317 bits. The watermark capacity of [36] is 396 bits. In [36], watermark is embedded in a diagonal coefficient of 4×4 integer DCT coefficients of a macroblock, which means a macroblock is only embedded with one bit of watermark information. However, the proposed algorithm in this paper uses the JND model to guide watermark embedding to increase the watermark capacity. The average embedding capacity of our algorithm is about 3.33 times as much as that of [36].

As for the impact on visual quality, PSNR is used to access the video quality. **Table 4** shows that the PSNR decreased by 0.288dB in this paper and 0.066dB in [36] after embedding watermarks, respectively. However, as Noorkami [24] points out, PSNR cannot objectively measure the quality of video image based on JND model. The reason why PSNR of this paper has a greater decline than that of literature [36] is that the capacity of the proposed algorithm is significantly higher than that of [36].

As we can see from **Table 4**, bit-rate increases by 6.29kbits/s in the algorithm of this paper and 6.64 kbits/s in that of [36]. The algorithm of this paper not only has more watermark embedding capacity but also less increased bit-rate than that of [36]. This is because the watermark is embedded in non-zero quantization DCT coefficients in this paper. However, the watermark is embedded in a fixed diagonal coefficient in [36], which may make the original zero coefficients become non-zero coefficients. It affects bit-rate deeply.

4.3 Analysis on watermark invisibility

To test the robustness of the proposed algorithm, the watermarked videos are re-encoded which changes prediction mode and quantized DCT coefficient. Common signal processing attacks have been applied, including additive Gaussian noise, salt and pepper noise and Gaussian low-pass filtering. **Fig. 8** shows the examples of attacked Forman. The visual quality decreases significantly, especially under Gaussian noise.

(a) QP=28 re-encoding (b) Gauss noise

(c) Salt and pepper noise (d) Gauss low pass filter

Fig. 8. Foreman sequence first frame image under various attacks

Table 5 and **Table 4** are under the same experimental condition. In the case of obtaining watermarking capacity, impact of watermark embedding on PSNR and bit-rate increase from **Table 4**, **Table 5** shows the robustness of the watermark. In **Table 5**, we set Gaussian noise with variance 0.0005, salt and pepper noise with intensity 0.01 and Gaussian low pass filter with 5×5 mask.

The values in **Table 5** are the correct ratio of watermark extraction. That is:

$$sim = \frac{N_{\mathrm{d}}}{N_e} \tag{25}$$

where N_{d} represents the number of watermark bits extracted correctly from the video sequences, N_{e} is the total number of watermark bits.

Table 5. Comparison of robustness of watermarking

	Recoding QP=28		**Gauss noise attack(Mean is 0, the variance is 0.0005)**		**Salt and pepper noise (Noise density is 0.01)**		**Gaussian low-pass lter(5×5,0.3)**	
Tested video	Algorithm in [36]	Our scheme	Algorithm in [36]	Our scheme	Algorithm in [36]	Our scheme	Algorithm in [36]	Our scheme
Foreman	0.9702	0.9967	0.8394	0.8661	0.8533	0.9650	0.9646	0.9958
mobile	0.9596	0.9924	0.7891	0.8097	0.8689	0.9625	0.9540	0.9911
coastguard	0.9798	0.9928	0.7500	0.7855	0.8078	0.9620	0.9765	0.9902
hall	0.9712	0.9902	0.7889	0.7532	0.7843	0.9565	0.9672	0.9879
flower	0.9629	0.9958	0.7500	0.7826	0.8078	0.9624	0.9588	0.9944

As can be seen from the **Table 5**, robustness of watermarking algorithm in this paper is better than that in [36]. The reason is that we use a different embedding algorithm according to different characteristics of middle and high components in 4×4 integer DCT. Another reason is that we embed watermark in the non-zero quantization DCT coefficients.

5. Conclusion

In this paper, we propose a robust H.264/AVC video watermarking algorithm based on JND model. In order to embed watermarks invisibly, a new and more accurate JND model is proposed. Then, after analysis on the drift error of energy distribution, we propose that watermarks should be embedded in middle and high components of 4×4 integer DCT. Finally, in order to improve the robustness, the watermarks are embedded using different algorithms in middle and high components, respectively. Experimental results demonstrate that the watermarked video sequences have good visual quality and satisfactory robustness against common video attacks and high watermarking capacity.

References

[1] F. Fund, S.A Hosseini, S.S Panwar, "Under a cloud of uncertainty: legal questions affecting internet storage and transmission of copyright-protected video content," *IEEE Network*, vol.30, no.2, pp.32-38, 2016. Article (CrossRef Link)

[2] X. Jiang, T. Sun, Y. Zhou, W. Wang, Y.Q Shi, "A Robust H. 264/AVC Video Watermarking Scheme with Drift Compensation," *The Scientific World Journal*, vol.3, no. 802347, pp.1-13, 2015. Article (CrossRef Link)

[3] L. Tian, N. Zheng, J. Xue, C. Li, "Authentication and copyright protection watermarking scheme for H.264 based on visual saliency and secret sharing," *Multimedia Tools and Applications*, vol.74, no.9, pp.2991-3011, 2015. Article (CrossRef Link)

[4] X. Zhu, C.W Chen, "A Joint Source-Channel Adaptive Scheme for Wireless H. 264/AVC Video Authentication," *IEEE Transactions on Information Forensics and Security*, vol.11, no.1, pp.141-153, 2016. Article (CrossRef Link)

[5] P.B Nguyen, A. Beghdadi, M. Luong, "Perceptual watermarking using a new Just-Noticeable-Difference model," *Signal Processing: Image Communication*, vol.28, no.10, pp.1506-1525, 2013. Article (CrossRef Link)

[6] N. Jayant, J. Johnston, R. Safranek, "Signal compression based on models of human perception," in *Proc. of IEEE*, pp.1385-1422, 1993. Article (CrossRef Link)

[7] R.B Wolfgang, C.I Podilchuk, E.J Delp, "Perceptual watermarks for digital images and video," in *Proc. of IEEE*, pp.1108-1126, 1999. Article (CrossRef Link)

[8] A.B Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. of SPIE 1913, Human Vision, Visual Processing, and Digital Display*, pp.202-216, 1913. Article (CrossRef Link)

[9] D. Simitopoulos, S.A Tasftaris, N.V Boulgouris, M. G. Strintzis, "Compressed-domain video watermarking of MPEG streams," in *Proc. of IEEE International Conference on Multimedia & Expo*, pp.569-572, 2002. Article (CrossRef Link)

[10] Y. Niu, M. Kyan, L. Ma, A. Beghdadi, S. Krishnan, "Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking," *Signal Processing: Image Communication*, vol.28, no.8, pp.917-928, 2013. Article (CrossRef Link)

[11] A.M. Joshi, V. Mishra, R.M. Patrikar, "Design of real-time video watermarking based on Integer DCT for H.264 encoder," *International Journal of Electronics*, vol.102, no.1, pp.141-155, 2015. Article (CrossRef Link)

[12] A. Mansouri, A.M. Aznaveh, F. Torkamani, F. Kurugollu, "A low complexity video watermarking in H.264 compressed domain," *IEEE Transactions on Information Forensics and Security*, vol.5, no.4, pp.649-657, 2010. Article (CrossRef Link)

[13] S. Li, J. Fu, P. Liu, Y. Jiang, "An information hiding approach based on integer transform coefficient and virtual space encoding for H.264/AVC," *Circuits Systems and Signal Processing*, vol.34, no.11, pp.3573-3594, 2015. Article (CrossRef Link)

[14] S. Gaj, A.S. Patel, A. Sur, "Object based watermarking for H.264/AVC video resistant to rst attacks," *Multimedia Tools and Applications*, vol.75, no.6, pp.3053-3080, 2016. Article (CrossRef Link)
[15] D.W. Xu, R.D. Wang, "Efficient reversible data hiding in encrypted H.264/AVC videos," *Journal of Electronic Image*, vol.23, no.5, 053022, 2014. Article (CrossRef Link)
[16] J.M. Thiesse, J. Jung, M. Antonini, "Rate distortion data hiding of motion vector competition information in chroma and luma samples for video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.21, no.6, pp.729-741, 2011. Article (CrossRef Link)
[17] D.W. Xu, R.D. Wang, Y.Q. Shi, "Data hiding in encrypted H.264/AVC video streams by codeword substitution," *IEEE Transactions on Information Forensics and Security*, vol.9, no.4, pp.596-606, 2014. Article (CrossRef Link)
[18] T. Stutz, F. Autrusseau, A. Uhl, "Non-Blind structure-preserving substitution watermarking of H.264/CAVLC inter-frames," *IEEE Transactions on Multimedia*, vol.16, no.5, pp.1337-1349, 2014. Article (CrossRef Link)
[19] C.H. Wu, Y. Zheng, W.H. Ip, C.Y. Chana, K.L. Yunga, Z.M. Lu, "A flexible H.264/AVC compressed video watermarking scheme using particle swarm optimization based dither modulation," *International Journal of Electronics and Communications*, vol.65, no.1, pp.27-36, 2011. Article (CrossRef Link)
[20] A.M. Joshi, V. Mishra, R.M. Patrikar, "FPGA prototyping of video watermarking for ownership verification based on H.264/AVC," *Multimedia Tools and Applications*, vol.75, no.6, pp.3121-3144, 2016. Article (CrossRef Link)
[21] Y.L. He, G.B. Yang, N.B. Zhu, "A real-time dual watermarking algorithm of H.264/AVC video stream for video-on-demand service," *AEU-International Journal of Electronics and Communications*, vol.66, no.4, pp.305-312, 2012. Article (CrossRef Link)
[22] M.A. Ali, E.A. Edirisinghe, "Watermarking H.264/AVC by modifying DC coefficients," in *Proc. of International Conference on CyberWorlds*, pp 241-245, 2009. Article (CrossRef Link)
[23] L. Li, Z.H. Dong, J.F. Lu, J.P. Dai, Q.R. Huang, C.C. Chang et al., "An H.264/AVC HDTV watermarking algorithm robust to camcorder recording," *Journal of Visual Communication and Image Representation*, vol.26, no.C, pp.1-8, 2015. Article (CrossRef Link)
[24] M. Nookami, R.M. Mersereau, "Digital watermarking in P-Frames with controlled video bit-rate increase," *IEEE Transactions on Information Forensics and Security*, vol.3, no.3, pp.441-455, 2008. Article (CrossRef Link)
[25] Su PC, Wu CS, Chen IF, C.Y. Wu, Y.C. Wu, "A practical design of digital video watermarking in H.264/AVC for content authentication," *Signal Processing: Image Communication*, vol.26, no.8, pp.413-426, 2011. Article (CrossRef Link)
[26] D.W. Wang, S.J. Huang, G.R. Feng, "Perceptual differential energy watermarking for H.264/AVC," *Multimedia Tools and Applications*, vol.60, no.3, pp.537-550, 2012. Article (CrossRef Link)
[27] Z. Wei, K.N. Ngan, "Spatial-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Transactions on Circuits and System for Video Technology*, vol.19, no.3, pp.337-346, 2009. Article (CrossRef Link)
[28] M. Naccari, F. Pereira, "Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment," *IEEE Transactions on Circuits and System for Video Technology*, vol.21, no.6, pp.766-782, 2011. Article (CrossRef Link)
[29] T. Suzuki, K. Sato, Y. Yagasaki, "Weighting matrix for JVT codec," Document JVT-C053rl, *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG*. Article (CrossRef Link)
[30] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol.6, no.12, pp.1673-1687, 1997. Article (CrossRef Link)
[31] J.W. Huang, Y.Q. Shi, Y. Shi, "Embedding image watermarks in dc components," *IEEE Transactions on Circuits and System for Video Technology*, vol.10, no.6, pp.974-979, 2000. Article (CrossRef Link)

[32] D.W. Kim, Y.G. Choi, H.S. Kim, J.S. Yoo, H.J. Choi, Y.H. Seo, "The problems in digital water marking into intra-frames of H.264/AVC," *Image and Vision Computing*, vol.28, no.8, pp.1220-1228, 2010.Article (CrossRef Link)
[33] M. Noorkami, R.M. Mersereau, "Compressed-domain video watermarking for H.264," in *Proc. of International Conference on Image Processing (ICIP)*, pp.1229-1232, 2005. Article (CrossRef Link)
[34] C. Langelaarg, R.R. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Transactions on Image Processing*, vol.10, no.1, pp.148-158, 2001. Article (CrossRef Link)
[35] P.C. Su, M.L. Li, I.F. Chen, "A content-adaptive digital watermarking scheme in H.264/AVC compressed videos," in *Proc. of International Conference on Intelligent Information Hiding & Multimedia Signal Processing*, pp.849-852, 2008. Article (CrossRef Link)
[36] G. Qiu, P. Marziliano, A.T.S. Ho, D. He, "A hybrid watermarking scheme for H.264/AVC video," in *Proc. of the 17th International Conference on Pattern Recognition*, pp.865-868, 2004. Article (CrossRef Link)

Weiwei Zhang received his B.S. and M.S. degree from Huaqiao University, Quanzhou, Fujian, in 2006 and 2009, respectively and Ph.D. degree in information security from Beijing University of Posts and Telecommunications, Beijing, P.R. China, in 2013. He is currently a lecturer at the College of Engineering of Huaqiao University. His research interests include copyright protection for digital content, big data security and information security.

Xin Li received his B.S. and M.S. degree in Internet of Things Engineering from Huaqiao University, Quanzhou, Fujian, in 2014 and 2017, respectively. He is currently pursuing a doctoral degree in information security in Beijing University of Posts and Telecommunications, Beijing, P.R. China. His research interests include database security and network security.

Yuzhao Zhang received his B.S. degree in radio engineering from Nanjing Institute of Technology, Nanjing, Jiangsu, in 1981. He is currently an associate professor at the College of Engineering of Huaqiao University. His research interests include embedded system security and the communication quality of Internet of things.

Ru Zhang received her Ph.D. degree in Computer Application Technology in 2003. She is a Prof. at Computer College, Beijing University of Posts and Telecommunications. She researches on Information Security in the state key laboratory of networking and switching technology in Beijing University of Posts and Telecommunications. Her interests include digital watermark, cryptography and multimedia authentication. She was awarded a national second prize and two provincial prizes.

Lixin Zheng received his B.S. and M.S. degree from Huaqiao University, Quanzhou, Fujian, in 1987 and 1990, respectively and Ph.D. degree in Prediction and Control from Tianjin University, Tianjin, P.R. China, in 2002. He is currently a professor at the College of Engineering of Huaqiao University. His research interests include motion control and machine vision.