# Robust Image Watermarking via Perceptual Structural Regularity-based JND Model

**Chunxing Wang[1], Meiling Xu[1], Wenbo Wan[2*], Jian Wang[2], Lili Meng[2], Jing Li[3] and Jiande Sun[2]**
[1] School of Physics and Electronics, Shandong Normal University
Jinan, CHN
[2] School of Information and Engineering, Shandong Normal University
Jinan, CHN
[3] School of Mechanical and Electrical Engneering, Shandong Management University
Jinan, CHN
*Corresponding author: Wenbo Wan
[e-mail: wanwenbo@sdnu.edu.cn]

## Abstract

A better tradeoff between robustness and invisibility will be realized by using the just noticeable (JND) model into the quantization-based watermarking scheme. The JND model is usually used to describe the perception characteristics of human visual systems (HVS). According to the research of cognitive science, HVS can adaptively extract the structure features of an image. However, the existing JND models in the watermarking scheme do not consider the structure features. Therefore, a novel JND model is proposed , which includes three aspects: contrast sensitivity function, luminance adaptation, and contrast masking (CM). In this model, the CM effect is modeled by analyzing the direction features and texture complexity, which meets the human visual perception characteristics and matches well with the spread transform dither modulation (STDM) watermarking framework by employing a new method to measure edge intensity. Compared with the other existing JND models, the proposed JND model based on structural regularity is more efficient and applicable in the STDM watermarking scheme. In terms of the experimental results, the proposed scheme performs better than the other watermarking scheme based on the existing JND models.

***Keywords:*** STDM, robust JND model, watermarking robustness, structural regularity, direction energy.

## 1. Introduction

**W**ith the rapid development of network technology and digital products, information hiding technologies such as steganography, image hashing technology [1-3] and digital watermarking technology have emerged to protect multimedia information security. Digital watermarking is an important branch of information hiding technology, and has become a prevailing solution to protect multimedia data. Digital watermarking technology embed the watermark information into multimedia information carriers, which didn't degrade the perceptual quality and withstand common attacks. The technology must satisfy robustness, security, fidelity and watermark capacity [4]. Security refers to demanding the embed information be secret to decry. Robustness refers to the ability of a system to possess an effective watermark after common attacks. Fidelity means that after embedding the watermark into an image, the visual quality of the watermarked image has no change compared with the original image. The watermark capacity refers the amount of information that can be embedded into the host image. Among these requirements, increasing the watermark strength is common way to enhance robustness at the cost of loosing fidelity. Therefore, there is a non-negligible issue for researchers to realize better tradeoff between robustness and fidelity.

Spread transform dither modulation (STDM), proposed by Chen and Wornell, is an important extension of quantization index modulation (QIM) watermarking algorithms, which probably obtains better tradeoff between robustness and fidelity than the other existing watermarking schemes [5]. In recent years, to enhance the robustness of the watermarking algorithms more effectively, many researchers have integrated the perceptual knowledge of human visual systems (HVS) into the watermark mechanism, then the watermark strength can be adjusted adaptively by utilizing the visual redundancy. Just noticeable distortion (JND) refers that the HVS cannot perceive the distortion below a certain threshold because of the existence of various visual masking effects, and is an efficient method to model the perceptual redundancies [6,7].

JND model in this paper assist the STDM watermarking framework in maximizing robustness while keeping the fidelity. There are still some limitations with the previous subband-based JND models that were used in digital watermarking systems to adjust the embedding strength, which are mainly reflected in two aspects. On one hand, most of the existing models cannot match well with the STDM framework. This is mainly because the perceptual slacks obtained from the watermarked image is inconsistent with that from the original image. On the other hand, the JND model does not reflect the real perception characteristics of the HVS well. Consequently, it is necessary to construct an effective perceptual model which estimates the perceptual slacks accurately and efficiently so as to enhance the performance of STDM watermarking scheme.

The research on primary visual cortex demonstrates that the neuronal cells in the V1 region are distributed in a precise and orderly manner. And these neuronal cells are arranged in stripe pattern with certain direction so that V1 area can extract structure features. The human brain can adaptively extract the structure regularities for signal perception and understanding [8–10]. It indicates that structural regularity is one of the important characteristics of visual perception. In the scene perception, humans are more likely to perceive the regularity regions than the irregularity regions. So inspired by the this mechanism, we consider that the

sensitivity of human eyes to local area is affected by the structure of the local area.

In this paper, a novel structural regularity-based JND model that meets human visual perception characteristics is proposed to generate the perceptual slack vector in the STDM watermarking scheme, which achieves better tradeoff between robustness and fidelity. The contributions that we have made in this paper can be summarized as follows:

1.      The HVS is more sensitive to the distortion in plain or orderly regions than that in texture or disorderly regions. To ensure the results of block classification consistent with human visual characteristics, we present two new methods of block classification which can classify DCT blocks effectively and accurately.

2.      In order to avoid the edge density in the JND model affected by the watermark embedding and to measure the edge density more accurately, a new edge density calculation method by using the un-watermarked AC coefficients is proposed.

3.      A new method for block classification and a new edge density calculation method are used to measure the contrast masking (CM) effects in the JND model. A new structural regularity-based perceptual JND model is established, which takes into account three factors, namely, contrast sensitivity function (CSF), luminance adaptation (LA) effect and CM effect. Compared with the latest JND models, the effectiveness of the proposed model in the STDM watermarking framework is verified.

4.      The proposed JND model is applied in the STDM watermarking framework to obtain optimum rate-distortion-robustness tradeoff. The experiments are conducted to compare the performance of the proposed STDM scheme with other existing STDM watermarking algorithms. In terms of experimental results, the proposed STDM watermarking algorithm based on the proposed JND model has excellent performance.

The rest sections of this paper are organized as follows: The next section reviews the related work on existing JND estimation models and the traditional perceptual STDM algorithms. A robust DCT-based JND model for STDM watermarking framework is established and the proposed perceptual JND model-based STDM watermarking scheme is described in Section 3. The Section 4 of this paper presents the experimental results and analysis, and Section 5 concludes this paper.

## 2. Related Work

As we all know, the spread transform dither modulation (STDM) is more robust to re-quantization. Li et al. [11] proposed robust algorithms which incorporate Watson's model into STDM framework and proved that perceptual STDM is one of the most effective schemes to embed multi-bit watermark in an image by quantization approach.

### 2.1 Spread Transform Dither Modulation

STDM which is an classical extension of quantization index modulation (QIM) plays a significant part in the quantization-based watermarking scheme [5]. By using the dither modulation (DM) quantizer, STDM modulates the projection of the host vector along a given random direction to achieve the embedding watermark information. With QIM and spread spectrum system cooperating, this scheme provides significant improvements in effectiveness and robustness.

In the embedding procedure, the projections $X_u$ is obtained by projecting the host vector onto a random vector $u$ and is modulated on the basis of the message bit $m$. The watermarked can be obtained as:

$$Y = X + [Q(X_u, \Delta, m, \delta) - X_u] \cdot u, m \in \{0,1\} \tag{1}$$

where the quantization function $Q(\cdot)$ is represented as

$$Q(X_u, \Delta, m, \delta) = \Delta \cdot round\left(\frac{X_u + \delta}{\Delta}\right) - \delta, m \in \{0,1\} \tag{2}$$

where $\Delta$ and $\delta$ denote quantization step and dither signal, respectively. In the process of extracting watermark, the received vector $X'$ may be distorted because of channel transmission, and the projection $X_u'$ is obtained by projecting the distortion vector $X'$ onto the given vector $u$. The watermark bit $m'$ is extracted on the basis of the minimum distance detector as:

$$m' = \arg\min_{b \in \{0,1\}} \left| X_u' - Q(X_u', b, \Delta, \delta) \right| \tag{3}$$

## 2.2 Perceptual STDM with JND Model

To avoid the noise which is introduced by quantization exceeding the distortion visibility threshold, the perceptual model is applied into watermarking framework. Watson proposed a JND model to estimate the minimum distortion value that can be perceived by human visual systems in DCT domain [12].

As mentioned above, the distortion caused by watermark embedding in STDM scheme actually happens on the random direction $u$. Therefore, the slack vector $s$ is also projected onto the random direction $u$ and then we can calculate the visual redundancy of host vector $X$ in the random direction $u$. The reason why the quantization step $\Delta = 2s_u$ is because the maximum quantization error of the quantizer is $\Delta/2$ [5][17]. It's easy to found that the slack vector $s$ calculated by each image are unique. So the slack vector $s$ obtained from the host image is not the same as that from the watermarked image. In order to avoid this mismatch phenomenon and ensure that the image distortion caused by watermark embedding is more correlates with human visual characteristics, a more accurate and robust JND model for STDM framework is necessary.

## 2.3 JND Estimation

In recent years, the analysis of human visual system has always been research focus. Applying this knowledge to digital image analysis can help researchers to understand image features. Here, we describe four state-of-the-art DCT-based perceptual JND models.

In [12], the classic DCT-based JND model which was put forward by Watson consists of masking component based on luminance, a sensitivity function and contrast masking. A frequency sensitivity table is defined in this model, and each value represents the minimum DCT coefficients that can be perceived when there is no noise. Lower value indicates that the human visual system is more sensitive to its frequency. M. Kim et al. [13,16] confirmed that

frequency characteristics are also major determinants that affect the threshold of JND by performing psychophysical experiment, and then proposed a novel DCT-based JND model which contains three parts: the base threshold, luminance adaptation, and contrast masking. The model provides a more precise CSF, and the spatial frequency is used as one of determining factor to estimate the LA and CM effects. In [14], the authors have put forward a DCT-based JND model which extracts the frequency characteristics from the corresponding psychophysical experiment results, and this model is the product of three parts: the base threshold, the masking factors for LA and CM. In this model, there is a new approach to calculate edge strength and edge density. More recently, a orientation regularity-based JND model in which use the orientation features as one of essential factors in analyzing visual content is proposed in [15]. The authors adopt frequency texture energy and orientation regularity to classify DCT blocks into five types. Based on block types and their HVS sensitivities, the authors can measure the CM effect. As mentioned above, compared with the conventional Watson model, other model do a better job in simulating human visual characteristics. However, applying the Kim's model or the model of [15] to the STDM framework exists mismatch problem, which is mainly because the JND value estimated from the watermarked image is distinguished with that from the initial image. The model of [14] is applied into the STDM framework, which can solve the mismatch problem. Nevertheless, the structural characteristics are not considered in this model so that the visual redundancies cannot be estimated well. Therefore, a more precise and robust perceptual model for the STDM watermarking framework is necessary to generate the perception slack vector.

## 3. The Proposed Robust Perceptual Model-based Watermarking Scheme

In the STDM watermarking framework, the high-precision JND model is introduced to accurately measure the visual redundancy of images and to improve the quality of the watermarked images. In this section, we first introduce a new robust structural regularity -based JND model in which the sensitivity of HVS to image structure is considered as an important factor, and then we applied the perceptual JND model to estimate the quantization step size of each sample in STDM-based watermarking embedding and detection procedures.

### 3.1 The Proposed Perceptual Just Noticeable Distortion Model

A novel JND estimation model based on structural regularity is proposed in this article, which is proposed as a model for the joint effect of contrast masking, luminance adaptation, and contrast sensitivity function. We focus on the JND model for the fixed-size $8 \times 8$ DCT block. For the $n$'th block in the DCT domain, the JND threshold $T_{jnd}(n, x, y)$ of position $(x, y)$ is expressed as

$$T_{jnd}(n, x, y) = T_B \cdot F_{LA} \cdot F_M \tag{4}$$

Where $F_{LA}$ is the modulation factor from the luminance adaptation, $F_M$ is another modulation factor from the contrast masking, $T_B$ refers the base threshold.

### 3.1.1 Contrast Sensitivity Function

$T_B$ reflects the spatial CSF and is only variant to spatial frequencies [18]. The base threshold $T_B$ for each spatial frequency $w_{x,y}$ can be calculated by

$$T_B(n,x,y) = \frac{1}{4\phi_x\phi_y} \cdot \frac{\exp(0.18w_{x,y})/1.33 + 0.11w_{x,y}}{0.6 + 0.4\cos^2\varphi_{x,y}} \tag{5}$$

where $\varphi_{x,y}$ is the direction angle and can be represented as Equation (6), $\phi_x$ and $\phi_y$ are the normalization factors and are calculated as Equation (7)

$$\varphi_{x,y} = \arcsin\left(\frac{2w_{x,0} \cdot w_{0,y}}{w_{x,y}^2}\right) \tag{6}$$

$$\phi_v = \begin{cases} \sqrt{1/8}, v = 0 \\ \sqrt{2/8}, v > 0 \end{cases} \tag{7}$$

$w_{x,y}$ represents the cycles per degree in spatial frequency for the $(x, y)$-th DCT coefficient and is defined as

$$w_{x,y} = \frac{1}{2 \times 8}\sqrt{(x/\theta_h)^2 + (y/\theta_v)^2} \tag{8}$$

where $\theta_h/\theta_v$ is the horizontal/ vertical angles of a pixel.

### 3.1.2 Luminance Adaptation Effect

The luminance masking threshold usually depends on the background brightness of a local image area; the brighter the background is, the higher the masking value is [19]. The modulation luminance adaptation factor $F_{LA}$ which we refers the model of [19] can be expressed as

$$F_{LA} = \begin{cases} (60 - \bar{u})/150 + 1, & \bar{u} \le 60 \\ 1, & 60 < \bar{u} < 170 \\ (\bar{u} - 170)/425 + 1, & \bar{u} \ge 170 \end{cases} \tag{9}$$

where $\bar{u}$ is the average intensity of the DCT block.

### 3.1.3 Contrast Masking Effect

The CM factor is usually determined by the edge density. There exists a phenomenon: some areas include many structural regularity edges or orderly texture structure information but they may be classified into block type with high texture complexity. In fact, the HVS is sensitive to the structural regularity information [20]. This fact indicates that the structural regularity is one of the important factors to estimate contrast masking effect. Therefore, we proposed a new CM model by analyzing the structure features and calculating the texture complexity.

### A. Direction Energy

The regularity which indicates the change of content structure along diverse directions can be accurately reflected by AC coefficients [21]. Consequently, we decide to choose the AC coefficients $AC_{0,1}$, $AC_{1,0}$ and $AC_{1,1}$ (the subscripts indicate the index within the DCT block) to measure the energy along horizontal, vertical and diagonal directions, respectively. We define the horizontal energy $E_{hor}$ of a block is as

$$E_{hor}(i,j) = \left| AC_{0,1}(i,j) \right| \tag{10}$$

where $(i, j)$ denote the position of the block. Similarly, we defines the vertical energy $E_{ver}$ and the diagonal energy $E_{dia}$ which are calculated as follows, respectively

$$E_{ver}(i,j) = \left| AC_{1,0}(i,j) \right| \tag{11}$$

$$E_{dia}(i,j) = \left| AC_{1,1}(i,j) \right| \tag{12}$$

The maximum value $E_{max}$ and the median value $E_{med}$ of the direction energy of the block are obtained by comparing the horizontal energy $E_{hor}$, the vertical energy $E_{ver}$ and the diagonal energy $E_{dia}$ of the blocks. In order to determine whether a block has the dominant direction, the ratio $E_r$ is defined and is calculated as

$$E_r = \frac{E_{max}}{E_{med} + \gamma} \tag{13}$$

where $\gamma$ denotes a small constant value and is given by $\gamma = (0.03 \times L)^2$, and $L$ denotes the gray level of the image.

## B. Edge Pixel Density

The edge density $u_{edge}$ which can be obtained by detection operator goes in hand with the CM factor [22]. In the commonly used edge detectors, the Canny operator is one of the powerful operators which can provide good and reliable performance. The $u_{edge}$ detected by Canny operator is given by

$$u_{edge} = \left( 1/N^2 \right) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} Edge_{i,j} \tag{14}$$

where $Edge_{i,j}$ and N denote an edge value (0 is non edge, 1 is edge pixel) in an edge map and the size of DCT blocks, respectively.

But in the JND model based on watermarking scheme, we usually adopt the Canny operator to detect the edge density. The Canny operator results in a mismatch problem: the CM factor obtained from the watermarking image is inconsistent with that from the host image without any attacks. To settle the mismatch problem, the model of [14] used two AC coefficients to measure the edge strength of a block. Although this method can avoid the edge computation using Canny operator, the texture complexity of a block in the diagonal direction was not considered. In order to calculate the CM factors more accurately and avoid the mismatch problem, the AC coefficients $AC_{0,1}$, $AC_{1,0}$ and $AC_{1,1}$ are selected to measure the texture complexity of a block in the horizontal, vertical and diagonal directions. We define the new edge strength of a block as

$$S_{AC}(i,j) = \left| AC_{0,1}(i,j) \right| + \left| AC_{1,0}(i,j) \right| + \left| AC_{1,1}(i,j) \right| \tag{15}$$

where $(i, j)$ denotes the position of the block. Thus new edge density $u_{EDGE}$ is given by

$$u_{EDGE} = \ln \left[ 1 + \eta \cdot \frac{S_{ac} - \min(S_{ac})}{\max(S_{ac}) - \min(S_{ac})} \right] \tag{16}$$

To make the value of $u_{EDGE}$ closer to the edge density $u_{edge}$ value calculated by Canny operator, we carried out experiment to obtain the optimal value of the parameter $\eta$.

$$\eta = \arg \min_{\eta \in [0.01, 0.81]} D \qquad (17)$$

$$D = \left| u_{edge} - u_{EDGE} \right|^2 \qquad (18)$$

where $\eta$ changes from 0.01 to 0.81 with the step 0.02. When D obtains the minimum value, the corresponding parameter $\eta$ is the optimal value. According to the experimental results, the parameter $\eta$ is set as 0.47 and its corresponding the minimum value D of Euclidean distance is 0.9753. In [14], the parameter is empirically set as 0.57 and its corresponding D obtained by this experiment is 1.1786. This proved that the edge density value calculated by Equation (16) is closer to that obtained by Canny operator. The curve fitting of parameter $\eta$ is shown in **Fig.1**.
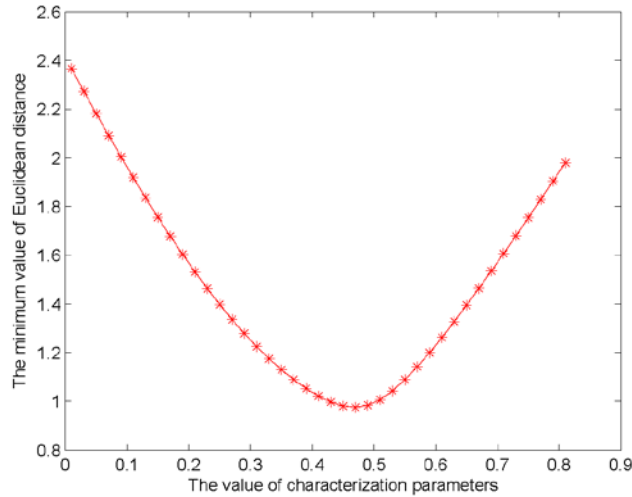


**Fig. 1.** The curve fitting graph of parameter $\eta$.

## C. CM estimation model based on Block Classification

As mentioned, to raise the accuracy of block classification, the structure is also seen as one of the significant factors for the CM model. In this section, we propose a new CM model which employs two methods of block classification to obtain the classification results of blocks more accurately.

As shown in the **Algorithm**, to begin with, we analyze the structure features of blocks. For the current block, by calculating $E_{hor}$, $E_{ver}$ and $E_{dia}$, the maximum value $E_{max}$ and median value $E_{med}$ of direction energy and their ratio $E_r$ can be obtained, by Equations (10), (11), (12) and (13). The value of $E_r$ not exceeding a certain threshold $T_1$ means that the current block has no dominant direction.

---

**ALGORITHM**: Block Classification $(I, T_1, T_2)$

**Input**: $I$: Host image, $T_1, T_2$: Threshold.
**Output**: Block type: (plain 'p', orderly-texture 'order-t', disorderly-texture 'dis-t').
**Begin**
  (1) The host image $I$ is divided into 8×8 non-overlapping block.
  (2) Each blocks are converted into frequencies using DCT, and DC coefficients are obtained.
  (3) The horizontal, vertical, diagonal energy $(E_{hor}, E_{ver}, E_{dia})$ of a block are computed by
     Equation (10-12).
  (4) By comparing $E_{hor}, E_{ver}, E_{dia}$, the maximum and median value $(E_{max}$ and $E_{med})$ are
     obtained, then the ratio $E_r$ is computed by Equation (13).
  (5) The edge density $u_{EDGE}$ is computed by Equation (16).
    **for** each DCT block **do**
       **if** $E_r \leq T_1 \| E_{max} < (DC/8)$ **then**
         the block belongs to disorderly block
       **end**
       **if** $E_r > T_1 \&\& E_{max} \geq (DC/8)$ **then**
         **if** $E_{max} = E_{hor}$ **then**
           the block belongs to the horizontal of orderly block
         **else if** $E_{max} = E_{ver}$ **then**
           the block belongs to the vertical of orderly block
         **else**
           the block belongs to the diagonal of orderly block
         **end**
       **end**
       **if** $u_{EDGE} \leq T_2$ **then**
         Block type= 'p'
       **else**
         **if** the block belongs to disorderly block **then**
           Block type='dis-t'
         **else**
           Block type='order-t'
         **end**
       **end**
    **end**
**end**

---

If the value of $E_r$ does, the average value of direction energy of the current block is higher than the value of dominant direction energy, which indicates that the current block has almost no direction information. In the two cases above, the current block belongs to disorderly block type. Nevertheless, the direction corresponding to the maximum value $E_{max}$ will be the primary directions of the current block and so the current block belongs to orderly block type. The method of block classification based on direction energy can effectively distinguish the orderly-texture blocks from the regions with high texture complexity. Then, the edge density $u_{EDGE}$ of the current block is calculated by Equation (16). Assuming that the value of $u_{EDGE}$ is less than a certain threshold $T_2$, then the current block has lower texture complexity, therefore the current block is the type of plain block. Otherwise, the current block belongs to texture block type. Furthermore, If the current block is texture block and orderly block, the current block belongs to orderly-texture block type. If the current block is texture block and disorderly block, the current block is the type of disorderly-texture block. Finally, by analyzing the direction features and calculating the edge pixel density, all the blocks can be classified into three types:

plain, disorderly-texture and orderly-texture. In experiment, the threshold $T_1$ and $T_2$ are empirically given by 2 and 0.19, respectively.

As far as contrast masking effect, human eyes have diverse sensitivities to various kinds of block types. The HVS is more susceptible to the plain regions than the orderly-texture areas and is less susceptible to the disorderly-texture regions. The contrast masking values are also variant with the sensitivity of the HVS. Therefore, to reflect the varieties of contrast masking effect, the masking factor for vertical, horizontal and diagonal blocks can be given as follows:

$$M_{hor}(x, y) = 0.8 - \nabla \cdot y \tag{19}$$

$$M_{ver}(x, y) = 0.8 - \nabla \cdot x \tag{20}$$

$$M_{dia}(x, y) = 0.8 - \nabla \cdot k, k = \max(x, y) \tag{21}$$

where $x$ and $y$ are the DCT sub-band indices. $\nabla$ denotes constant value and is given by 0.1.

Last, by merging with the human visual sensitivity into the spatial frequency, the CM factor for each block type can be given by

$$F_M = \begin{cases} 1, \text{ for plain block} \\ 1 + \begin{cases} M_{hor}, \text{ for horizon of the orderly} - \text{texture block} \\ M_{ver}, \text{ for vertical of the orderly} - \text{texture block} \\ M_{dia}, \text{ for diagonal of the orderly} - \text{texture block} \end{cases} \\ 2.25, \text{ for} (x^2 + y^2 \leq 16) \text{ in disorderly} - \text{texture block} \\ 1.25, \text{ for} (x^2 + y^2 > 16) \text{ in disorderly} - \text{texture block} \end{cases} \tag{22}$$

## 3.2 Proposed JND Model-based STDM Watermarking Scheme

The proposed structural regularity-based JND model is applied to modulate quantization step in the STDM watermarking scheme. And the proposed watermarking scheme includes two parts, embedding and extraction procedure. **Fig. 2** illustrates the embedding and extraction steps of the watermarking scheme.

Here, taking Lena image as an example, the procedures of the watermark embedding and extraction are shown as follows.

Watermark embedding procedure

1. In order to obtain DCT coefficients, the Lena image will be divided into $8 \times 8$ sub-band blocks, and we will perform the DCT transform. Then the coefficients are scanned by zigzag arrangement. A part of coefficients (the fourth, sixth to tenth and thirteenth coefficients) are obtained by DCT sequence to form a host vector $x$.

2. According to the proposed perceptual JND model, Perceptual redundancy vector $s$ is obtained.

3. The host vector $x$ and the perceptual redundancy vector $s$ are mapped onto the given projection vector $u$ (which can be used as key) to generate projections $x_u$ and $s_u$. The quantization step $\Delta$ is calculated from $s_u$ can be multiplied by the embedding strength in practice.

4. One bit of watermark message $m$ is embedded into the host vector projection $x_u$ in terms of Equation (1). Then the modified DCT coefficients are transformed to obtain a

watermarked image.

Watermark extraction procedure

1. A host vector $x'$ is acquired by the same way as step 1 of watermark embedding procedure.
2. Perceptual redundancy vector $s'$ is obtained by the proposed perceptual JND model.
3. The host vector $x'$ and the perceptual redundancy vector $s'$ are mapped onto the given projection vector $u$ to generate projections $x_u'$ and $s_u'$, respectively. The quantization step $\Delta$ is calculated from $s_u'$ and can be multiplied by the embedding strength.
4. According to Equation (3), the watermark $m'$ is extracted by using the DM detector.
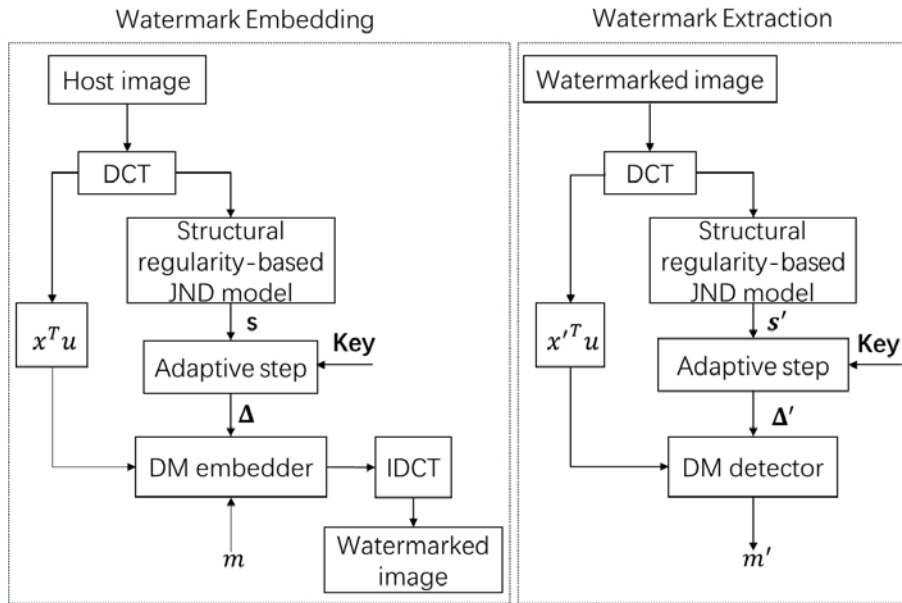


**Fig. 2.** Flowchart of proposed watermarking scheme.

## 4. Experimental Results

### 4.1. Evaluation of The Proposed Structural Regularity-based JND Model

A better JND model can tolerate more noise in the DCT coefficients of the image at the same perceptual quality. Therefore, we evaluate the performance of the proposed JND model by injecting noise into the DCT coefficients based on the corresponding JND value.

$$C'(n, x, y) = C(n, x, y) + \lambda \cdot JND(n, x, y) \tag{23}$$

where $C'(n, x, y)$ indicates the modified DCT coefficient; $\lambda$ takes random value of +1 or -1; $JND(n, x, y)$ denotes the threshold obtained from Watson's model [12], Kim's model [13], the model of [14], the model of [15] and the proposed model, respectively.

To verify the performance of the proposed JND model, eight standard images, 'Camera', 'Columbia', 'Crowd', 'Elaine', 'Lena', 'Peppers', 'Plane' and 'Woman', were used as testing images [23]. The size of the eight testing images are $256 \times 256$ shown in **Fig. 3**.
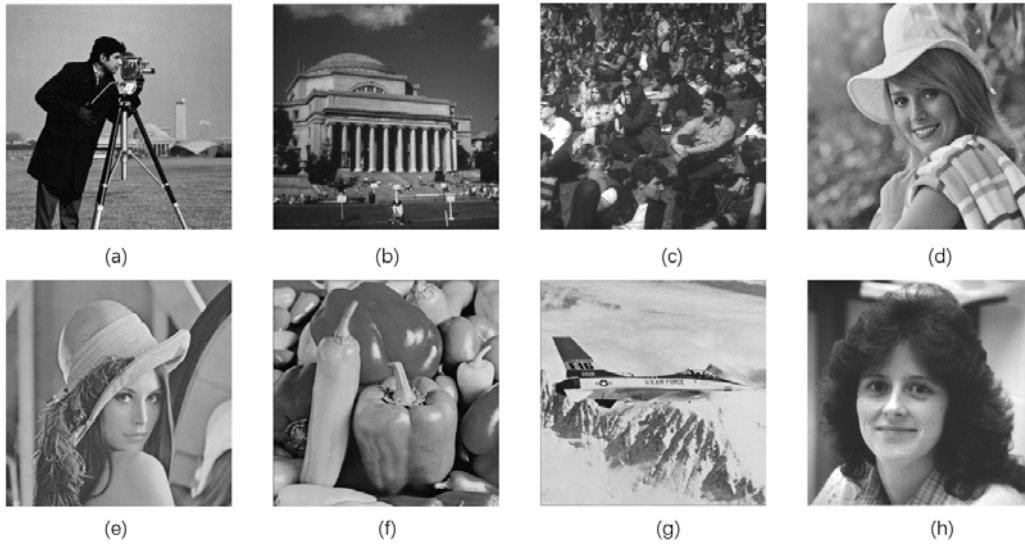
**Fig. 3.** Original cover images: Camera, Columbia, Crowd, Elaine, Lena, Peppers, Plane,Woman.

In this experiment, the structural similarity index measurement (SSIM) was employed to weigh the similarity between the original image and the distorted image obtained by injecting noise according to JND value generated by JND model. The SSIM index is calculated on the blocks x and y of an image [26].

$$SSIM\left(x, y\right) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)'} \tag{24}$$

With:

$\mu_x$, $\mu_y$ are the mean of $x$ and $y$, respectively;

$\sigma_x^2$, $\sigma_y^2$ are the variance of $x$ and $y$, respectively;

$\sigma_{xy}$ is the covariance of $x$ and $y$;

$c_1 = \left(k_1 L\right)^2$, $c_2 = \left(k_2 L\right)^2$ are two variables to stabilize the division with weak denominator; $L$ is the dynamic range of the pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$ by default.

At a same perceived quality, the better JND model is, the lower SSIM value will be yield. To clearly and intuitively describe the results of the JND comparison, we present the results in form of histogram **Fig. 4**. As shown in **Fig. 4**, the average SSIM values of Watson's model, Kim's model, the model of [14], the model of [15] and the proposed model are 0.7235, 0.5487, 0.5156, 0.3812 and 0.4811, respectively. The results show that the SSIM value of our proposed model is lower than Watson's model and Kim's model, which means that our proposed model can withstand more distortion than Watson's model and Kim's model. At the same time, we found that the average SSIM value of the proposed model is lower than the model of [14]. This is because the DCT coefficients $AC_{0,1}\left(x, y\right)$, $AC_{1,0}\left(x, y\right)$ and $AC_{1,1}\left(x, y\right)$ are used to calculate the edge intensity of (x, y)'th block in the proposed model, whereas the model of [14] only uses two AC coefficients $AC_{0,1}\left(x, y\right)$ and $AC_{1,0}\left(x, y\right)$. It can be observed from **Fig. 4** that the average SSIM of the proposed model is higher than the

model of [15]. Because all the DCT coefficients are used to calculate the frequency texture energy and direction feature energy in the model of [15]. Although only three AC coefficients cannot accurately reflect the image block's texture complexity, it can be closer to the result obtained by using the edge detection operator. Taken together, the classification results obtained by using the above two classification methods are almost the same as those in [15]. The following results proved that the proposed model is more efficient and applicable in the STDM watermarking framework.
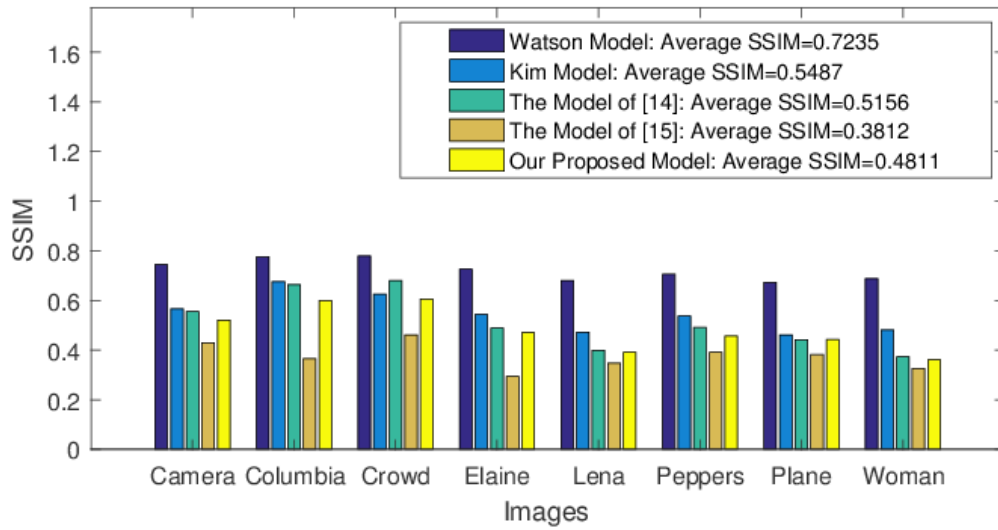


**Fig. 4.** SSIM comparison of different JND models

## 4.2. Evaluation of Different JND Models for STDM Watermarking Algorithms

This experiment is used to compare the performance of different JND models used in the same STDM watermarking scheme [12–15]. The each of testing images is standard image with a dimension of 256×256 obtained from the USC-SIPI image database [23]. A random message (1024 bits) is embedded into each original image. The fourth, sixth to tenth and thirteenth coefficients are selected from the DCT sequence to form a host vector and embedded one bit in it. The bit error rate (BER) is calculated for comparison, when the SSIM value between a testing image and the watermarked image is 0.982. To test the robustness of various models, several signal processing attacks are employed to attack the watermarked image.

**Table 1.** BER of noise attacks in STDM scheme with different JND models

| Attack types | Parameters | Watson Model | Kim Model | Model of [14] | Model of [15] | Proposed Model |
|---|---|---|---|---|---|---|
| Gaussian noise | 0.0003 | 0.0566 | 0.0488 | 0.0332 | 0.1338 | **0.0273** |
| | 0.0006 | 0.1055 | 0.1006 | 0.0830 | 0.2139 | **0.0645** |
| | 0.0009 | 0.1426 | 0.1523 | 0.1307 | 0.2471 | **0.1221** |
| Salt & pepper noise | 0.0003 | 0.0146 | 0.0234 | **0.0068** | 0.0137 | **0.0068** |
| | 0.0006 | 0.0186 | 0.0264 | **0.0146** | 0.0225 | **0.0146** |
| | 0.0009 | 0.0225 | 0.0313 | 0.0225 | 0.0254 | **0.0166** |

**Table 2.** BER of JPEG attacks in STDM scheme with different JND models

| Attack types | Parameters | Watson Model | Kim Model | Model of [14] | Model of [15] | Proposed Model |
|---|---|---|---|---|---|---|
| JPEG | 100 | 0.0059 | 0.0127 | **0** | 0.0078 | **0** |
| | 80 | 0.0127 | 0.0176 | **0** | 0.1201 | **0** |
| | 60 | 0.0469 | 0.0273 | 0.0049 | 0.1973 | **0.0029** |
| | 40 | 0.1133 | 0.0635 | 0.0459 | 0.2852 | **0.0420** |
| | 20 | 0.2847 | **0.2734** | 0.3806 | 0.4453 | 0.2861 |

**Table 3.** BER of scaling attacks in STDM scheme with different JND models

| Attack types | Parameters | Watson Model | Kim Model | Model of [14] | Model of [15] | Proposed Model |
|---|---|---|---|---|---|---|
| Volumetric scaling | 0.1 | 0.4072 | 0.4980 | 0.0039 | 0.3184 | **0.0029** |
| | 0.4 | 0.1279 | 0.3350 | **0** | 0.2021 | **0** |
| | 0.7 | 0.0332 | 0.1191 | **0** | 0.0957 | **0** |
| | 1.0 | 0.0059 | 0.0127 | **0** | 0.0088 | **0** |
| | 1.3 | **0.0166** | 0.1162 | 0.0225 | 0.0928 | 0.0293 |

**Table 4.** BER of filtering attacks in STDM scheme with different JND models

| Attack types | Parameters | Watson Model | Kim Model | Model of [14] | Model of [15] | Proposed Model |
|---|---|---|---|---|---|---|
| Median filtering | 3×3 | 0.0850 | 0.0762 | **0.0566** | 0.1172 | 0.0586 |
| | 5×5 | 0.3428 | 0.3555 | 0.3496 | 0.3564 | **0.3369** |
| Gaussian filtering | 3×3 | 0.0439 | 0.0547 | 0.0557 | 0.1797 | **0.0400** |

If adding noise in the image transmission, the watermarked image is attacked easily and inevitably. The results of the watermarked image attacked by adding Gaussian noise and adding Salt-and-Pepper noise are shown in **Table 1**. The watermarked image was tested against Gaussian noise with mean zero and different variance values. Compared with the other JND models, the proposed model always has the lowest BER for different noise intensities. This indicates that the proposed model performed much better than others.

The watermark robustness against the attack of JPEG compression is an important performance to be evaluated. As shown in **Table 2**, different performance emerges in the five JND models within the STDM watermarking algorithms show about JPEG compression attacks. The average BER of Watson's, Kim's, the model of [14], the model of [15] and the proposed model are 0.0927, 0.0789, 0.0863, 0.2111 and 0.0662, respectively.

When the watermarked image is under volumetric scaling attack with different scaling factors, the BER values of five JND models are presented in **Table 3**. For volumetric scaling attack with scaling factor ranging from 0.1 to 1, the performance of the proposed model is better than that of other models. For volumetric scaling attack with scaling factor 1.3, the performance of the proposed model is not the best, but the BER of the proposed model is still much lower than the model of [15].

As one of the classical attacks, filtering attack is significant. For example, the median filtering and Gaussian filtering are usually used to attack the watermarked image. **Table 4** presents the comparison results of filtering. For filtering with median filter (3×3), the BER value of the proposed model is 0.2% higher than the model of [14], but lower than the rest of models. For Gaussian filtering and filtering with median filter (5×5), the proposed model has the lowest BER values. In summary, the STDM watermarking scheme based on the proposed

mod performs excellently.

## 4.3. Evaluation of The Proposed Model-based STDM Watermarking Algorithms

In this section, the superior performance of the proposed scheme is demonstrated by comparing its functionality with that of the popular STDM watermarking schemes. We report the results that were performed on the test image "Lena" (size 256×256), and a random message of length 1024 bits was embedded into original Lena image.



(a)                                                          (b)

**Fig. 5.** Watermark insertion in Lena image. (a) Original image, (b) Watermarked image

**Fig. 5** shows the host image and the watermarked image without any attacks. The similarity between the original image (a) and the watermarked image (b) is measured by Equation (24). Generally, a larger SSIM indicates that the watermarked image resembles the host image more closely. The SSIM value between the host image (a) and the watermarked image (b) is 0.9893, which means that the watermarking scheme makes the watermark imperceptible.

The eight images in **Fig. 3** are used to test the execution time required for each STDM watermarking scheme, which is implemented as program codes running on a Matlab® working environment. **Table 5** presents the average time necessitated by different schemes.

**Table 5.** Average execution time per image (in second) for the compared watermarking schemes

| Process | Method | | | | |
|---------|--------|--------|--------|------|-----------------|
|         | AdpWM  | OptiWM | RDWMm  | RW   | Proposed scheme |
| Embedding | 1.009 | 0.940 | 1.223 | 0.997 | 0.763 |
| Extraction | 0.535 | 0.601 | 0.855 | 0.198 | 0.413 |

To further prove that the proposed model-based STDM watermarking scheme has better robustness, the proposed scheme was compared with the previous STDM watermarking schemes, referring to AdpWM [24], OptiWM [11], RDMWm [17], RW [25].

**Table 6.** BER of noise attacks for different STDM watermarking schemes

| Attack types | Parameters | AdpWM | OptiWM | RDWMm | RW | Proposed Scheme |
|--------------|------------|--------|--------|--------|--------|-----------------|
| Gaussian noise | 0.0003 | 0.0566 | 0.1006 | 0.0352 | 0.0527 | **0.0273** |
|  | 0.0006 | 0.1084 | 0.1602 | 0.0879 | 0.1475 | **0.0645** |
|  | 0.0009 | 0.1420 | 0.2129 | 0.1406 | 0.2197 | **0.1221** |
| Salt & pepper noise | 0.0003 | 0.0146 | 0.0605 | 0.0098 | 0.0098 | **0.0068** |
|  | 0.0006 | 0.0186 | 0.0635 | **0.0146** | 0.0195 | **0.0146** |
|  | 0.0009 | 0.0225 | 0.0654 | 0.0215 | 0.0303 | **0.0166** |

**Table 7.** BER of JPEG attacks for different STDM watermarking schemes

| Attack types | Parameters | AdpWM | OptiWM | RDWMm | RW | Proposed Scheme |
|---|---|---|---|---|---|---|
| JPEG | 100 | 0.0059 | 0.0547 | **0** | **0** | **0** |
| | 80 | 0.0127 | 0.0547 | **0** | **0** | **0** |
| | 60 | 0.0479 | 0.0840 | 0.0088 | 0.0059 | **0.0029** |
| | 40 | 0.1143 | 0.1200 | 0.0547 | 0.0771 | **0.0420** |
| | 20 | 0.2847 | 0.3574 | 0.2667 | 0.3301 | **0.2661** |

**Table 8.** BER of scaling attacks for different STDM watermarking schemes

| Attack types | Parameters | AdpWM | OptiWM | RDWMm | RW | Proposed Scheme |
|---|---|---|---|---|---|---|
| Volumetric scaling | 0.1 | 0.0352 | 0.0664 | 0.0668 | 0.4584 | **0.0029** |
| | 0.4 | 0.0098 | 0.0557 | **0** | 0.3594 | **0** |
| | 0.7 | 0.0059 | 0.0488 | **0** | 0.2090 | **0** |
| | 1.0 | 0.0059 | 0.0508 | **0** | **0** | **0** |
| | 1.3 | 0.0215 | 0.0586 | 0.0225 | 0.2188 | 0.0293 |

**Table 9.** BER of filtering attacks for different STDM watermarking schemes

| Attack types | Parameters | AdpWM | OptiWM | RDWMm | RW | Proposed Scheme |
|---|---|---|---|---|---|---|
| Median filtering | 3×3 | **0.1357** | 0.1914 | 0.1455 | 0.2109 | 0.1409 |
| | 5×5 | **0.3730** | 0.4297 | 0.3975 | 0.4023 | 0.3779 |
| Gaussian filtering | 3×3 | 0.0439 | 0.0752 | **0.0381** | 0.0664 | 0.0400 |

**Table 6** reflects the different performance of five different STDM watermarking algorithms for Gaussian noise and adding Salt-and-Pepper noise attacks. The OptiWM and RW have bad performance. However, the scheme in this paper did not exceed 12.3% for the Gaussian noise with a variance of $0.9 \times 10^{-3}$ and performs better than the other schemes in the Gaussian noise-adding attacks. For the same intensity of Salt-and-Pepper attack, the proposed scheme has the lowest BER values by comparing with other STDM schemes.

**Table 7** shows the sensitivity to JPEG compression. It is easy to show that the scheme in this paper has a better performance than the other schemes in robustness against JPEG compression. The average BER generated by OptiWM exceeds 13.4%, and other schemes has average BER lower than 9%. It is worth to notice that the average BER of the proposed scheme is 6.22%. The resistance of five schemes to volumetric scaling attacks are presented in **Table 8**. RW has weak resistance to volumetric scaling attacks. The average BER of OptiWM exceeds 5.6%, and the average BER of AdpWM and RDMWm are no lower than 1.5%. The proposed scheme has lowest average BER value which is 0.64%.

**Table 9** shows the response to median filtering and Gaussian attack. For filtering with median filter, AdpWM shows the best performance. The average BER of the proposed scheme is 0.51% higher than AdpWM, but lower than the rest of schemes. And for filtering with the Gaussian filter, RDWMm has the lowest BER value, and the proposed scheme is 0.19% higher than RDWMm. The BER of other schemes are higher than that of the proposed scheme.

Combining all of the experimental results, the proposed scheme has the best performance.

## 5. Conclusion

In this paper, a new perceptual JND model that meets human visual perception characteristic was introduced, which can effectively measure the masking effect. Using this perceptual model in the STDM watermarking scheme can not only meet the human visual perception characteristics, but also achieve better balance between robustness and fidelity. Specifically, new measurement methods for blocks classification and edge strength evaluation are proposed, which can estimate the perception slacks more effectively and accurately. Experimental results show that the proposed scheme is more robust than the watermarking algorithms based on the existing JND models with the uniform fidelity and has a superior performance than the previously proposed perceptual STDM framework.

In the proposed scheme, the structural regularity of image is an important factor to establish the proposed JND model. For an image with poor structural regularity, the performance of the proposed scheme may not be satisfactory. Therefore, for an image, the measurement of structural strength is an urgent problem to be solved. According to the structural strength of an image, we can make an adjustment on the scheme, which is the focus of future research.

## References

[1]   C. Qin, X. Chen, D. Ye, J. Wang, X. Sun, "A novel image hashing scheme with perceptual robustness using block truncation coding," *Information Sciences*, vol. 361-362, pp. 84-99, September 2016. Article (CrossRef Link)

[2]   C. Qin, X. Chen, X. Luo, X. Zhang, X. Sun, "Perceptual image hashing via dual-cross pattern encoding and salient structure detection," *Information Sciences*, vol. 423, pp. 284-302, January 2018. Article (CrossRef Link)

[3]   V. Monga, M.K. Mihcak, "Robust and secure image hashing via non-negative matrix factorizations," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, September 2007. Article (CrossRef Link)

[4]   S. A. Parah, J. A. Sheikh, N. A. Loan, "Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing," *Digital Signal Processing*, vol. 53, pp. 11-24, 2016. Article (CrossRef Link)

[5]   B. Chen, G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001. Article (CrossRef Link)

[6]   J. Zong, L. Meng, H. Zhang, W. Wan, "Jnd-based multiple description image coding," *KSII Transactions on Internet and Information system*, vol. 11, no. 8, pp. 3935-3949, August 2017. Article (CrossRef Link)

[7]   A. B. Watson, "DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images," *Society for Information Display Digest of Technical Papers XXIV*, pp. 946-949, 1993. Article (CrossRef Link)

[8]   J. Sun, X. Liu, W. Wan, J. Li, D. Zhao, H. Zhang, "Video hashing based on appearance and attention features fusion via DBN," *Neurocompuing*, vol. 213, no. 12, pp. 84-94, November 2016. Article (CrossRef Link)

[9]   N. B. Turkbrowne, J. Junge, B. J. Scholl, "The automaticity of visual statistical learning," *Journal of Experimental Psychology: General*, vol. 134, no. 4, pp. 552-564, November 2005. Article (CrossRef Link)

[10]  J. R. Saffran, E. D. Thiessen, "Pattern induction by infant language learners," *Developmental Psychology*, vol. 39, no. 3, pp. 484-494, 2003. Article (CrossRef Link)

[11]  Q. Li, I. J. Cox, "Improved spread transform dither modulation by using a perceptual model: robustness to amplitude scaling and JPEG compression," *IEEE Int. Conf. Acoust. Speech Signal Process*, vol. 2, pp. 185-188, 2007. Article (CrossRef Link)

[12] A. B. Watson, "DCT quantization matrices optimized for individual images," *Proc. SPIE*, vol. 1913, pp. 202-216, 1993. Article (CrossRef Link)

[13] S. H. Bae, M. Kim, "A Novel DCT-based JND model for luminance adaptation effect in DCT frequency," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 893-896, July 2013. Article (CrossRef Link)

[14] W. Wan, J. Liu, J. Sun et al, "Improved logarithmic spread transform dither modulation using a robust perceptual model," *Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13481-13502, November 2016. Article (CrossRef Link)

[15] W. Wan, J. Wu, X. Xie et al, "A novel just noticeable difference model via orientation regularity in DCT domain," *IEEE Access*, vol. 5, pp. 22953-22964, April 2017. Article (CrossRef Link)

[16] S. H. Bae, M. Kim, "A new DCT-based JND model monochrome images for contrast masking effects with texture complexity and frequency," in *Proc. of IEEE Int. Conf. Image Process*, pp. 431-434, 2013. Article (CrossRef Link)

[17] X. Li, J. Liu, J. Sun et al, "Step-projection-based spread transform dither modulation," *IET Information Security*, vol. 5, no. 3, pp. 170-180, 2011. Article (CrossRef Link)

[18] J. Wu, W. Wan, G. Shi, "Content complexity based just noticeable difference estimation in DCT domain," in *Proc. of Signal and Information Processing Association Annual Summit and Conference*, pp. 1-5, 2017. Article (CrossRef Link)

[19] Z. Wei, K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337-346, February 2009. Article (CrossRef Link)

[20] J. Wu, G. Shi, W. Lin et al, "Just noticeable difference estimation for images with free-energy principle," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1705-1710, June 2013. Article (CrossRef Link)

[21] Y. Zhong, H. Zhang, A. K. Jain, "Automatic caption localization in compressed video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 4, pp. 385-392, 2000. Article (CrossRef Link)

[22] J. Canny, "A computational approach to edge detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986. Article (CrossRef Link)

[23] USC-SIPI Image Database, Available online: http://sipi.usc.edu/database/ (accessed on 11 Mar. 2018) Article (CrossRef Link)

[24] L. Ma, "Adaptive spread-transform dither modulation using a new perceptual model for color image," *IEICE Trans. Inf. Sys*, vol. E93. D, no. 4, pp. 843-857, 2010 Article (CrossRef Link)

[25] Q. Li, G. Doerr, I. J. Cox, "Spread transform dither modulation using a perceptual model," in *Proc. of IEEE Workshop on Multimedia Signal Processing*, pp. 98-102, 2006. Article (CrossRef Link)

[26] X. Z. Xie, C. C. Lin, C. C. Chang, "Data hiding based on a two-layer turtle shell matrix," *Symmetry*, vol. 10, no. 2, pp. 47, 2018. Article (CrossRef Link)

**Chunxing Wang** received Ph.D. degree in Shandong Normal University, Jinan, China in 2010. Now he is a Professor with the School of Physics and Electronic Science, Shandong Normal University. His current research interests include image processing and digital watermarking.

**Meiling Xu** is a graduate student in school of Physics and Electronic Science, Shandong Normal University. Her interests include image processing and image/ video watermarking.

**Wenbo Wan** received Ph.D. degree in Shandong University, Jinan, China in 2015. Now he is a Lecturer with the School of Information Science and Engineering, Shandong Normal University. His research interests include image/video processing and image/video watermarking.
E-mail: wanwenbo@sdnu.edu.cn

**Jian Wang** received the B.E. degree from Qingdao University, Qingdao China, in 2012. He received the M.E. and D.E. degree from Ritsumeikan University, Kusatsu, Japan in 2014 and 2018, respectively. He is currently a lecturer at Shandong Normal University, Jinan, China. His research interests include image processing and analysis, machine learning, and computer vision.

**Lili Meng** received the B.E. degree from Shandong University, Jinan, China, in 2005, and received the PhD degree from Beijing Jiaotong University, Beijing, China, in 2013, supervised by Prof. Yao Zhao. During 2010.03~2010.06, she visited the National Kaohsiung University of Applied Sciences, Taiwan. During 2010.09~ 2011.09, she was a Visiting Student at Simon Fraser University, Canada. Since January 2014, she has joined the School of Information Science and Engineering, Shandong Normal University, Jinan, China.

**Jing Li** is a Ph.D. student in Shandong Normal University, Jinan, China. She is also a lecturer with School of Mechanical and Electrical Engineering, Shandong Management University. Her research interests include machine learning, multimedia processing and retrieval, etc.

**Jiande Sun** received the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2000 and 2005, respectively. From September 2008 to August 2009, he was a Visiting Researcher with the Institute of Telecommunications System, Technical University of Berlin, Berlin, Germany. From October 2010 to December 2012, he was a Post-Doctoral Researcher with the Institute of Digital Media, Peking University, Beijing, China, and with the State Key Laboratory of Digital-Media Technology, Hisense Group, respectively. From July 2014 to August 2015, he was a DAAD Visiting Researcher with Technical University of Berlin and University of Konstanz, Germany. From October 2015 to November 2016, he was a Visiting Researcher with the Language Technology Institute, School of Computer Science, Carnegie Mellon University, USA. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University. He has published more than 60 journal and conference papers. He is the co-author of two books. His current research interests include multimedia content analysis, video hashing, gaze tracking, image/video watermarking, 2D to 3D conversion, and so on.
E-mail: jiandesun@sdnu.edu.cn