

Article

# A Novel STDM Watermarking Using Visual Saliency-Based JND Model

Chunxing Wang <sup>1</sup>, Teng Zhang <sup>1</sup>, Wenbo Wan <sup>2,3,\*</sup> , Xiaoyue Han <sup>1</sup> and Meiling Xu <sup>1</sup>

<sup>1</sup> School of Physics and Electronics, Shandong Normal University, Jinan 250014, China; cxwang@sdnu.edu.cn (C.W.); zhangteng2@stu.sdnu.edu.cn (T.Z.); hanxiaoyue@stu.sdnu.edu.cn (X.H.); xumeiling1@stu.sdnu.edu.cn (M.X.)

<sup>2</sup> School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

<sup>3</sup> Institute of Data Science and Technology, Shandong Normal University, Jinan 250014, China

\* Correspondence: wanwenbo@sdnu.edu.cn; Tel.: +86-183-5311-3687

Received: 27 July 2017; Accepted: 23 August 2017; Published: 25 August 2017

**Abstract:** The just noticeable distortion (JND) model plays an important role in measuring the visual visibility for spread transform dither modulation (STDM) watermarking. However, the existing JND model characterizes the suprathreshold distortions with an equal saliency level. Visual saliency (VS) has been widely studied by psychologists and computer scientists during the last decade, where the distortions are more likely to be noticeable to any viewer. With this consideration, we proposed a novel STDM watermarking method for a monochrome image by exploiting a visual saliency-based JND model. In our proposed JND model, a simple VS model is employed as a feature to reflect the importance of a local region and compute the final JND map. Extensive experiments performed on the classic image databases demonstrate that the proposed watermarking scheme works better in terms of the robustness than other related methods.

**Keywords:** watermarking; JND; STDM; visual saliency

## 1. Introduction

With the development of imaging devices, such as digital cameras, smartphones, and medical imaging equipments, our world has been witnessing a tremendous growth in the quantity, availability, and importance of images. Image recognition, image retrieval, information hiding, fingerprint detection, and watermarking have caused more and more scholars' concerns [1].

In recent years, with rapid growth within content management systems, blind watermarking schemes (i.e., algorithms in which the watermark can be extracted without resorting to the original unmarked signal) provide a robust and imperceptible technique to enhance image security. Among the well-designed watermarking schemes, fidelity and robustness are the most conflicting issues. Usually, we can increase the watermark strength to improve robustness at the expense of losing fidelity, and vice versa. Consequently, how to maintain a balance between imperceptibility and robustness is important for a sophisticated watermarking algorithm [2].

In 2008, human visual system based adaptive digital image watermarking, proposed by Huiyan Qi [3], is an adaptive image watermarking algorithm. In this scheme, the spatial masking is built according to the image features, and the quality of the watermarked image using the proposed adaptive masking is much better than the one without using the adaptive masking. Recently, the adaptive nature of the proposed moment-based watermarking method, owing to the optimized fuzzy inference systems embodying prior knowledge, is proposed by G.A. Papakostas [4], which controls the amount of watermark information each moment coefficient is capable to carry. Spread transform dither modulation (STDM), proposed by Chen and Wornel [5], is a typical one owing to its advantages in implementation and computational flexibility. To achieve a better

rate-distortion-robustness trade-off, it is necessary to take the characteristic of the human visual system (HVS) into consideration. Various STDM-based watermarking algorithms incorporate the perceptual knowledge of HVS to improve fidelity by regulating the watermark strength [6]. The just noticeable distortion (JND) represents distortion perceived by the HVS and determine the watermark strength. In 2006, I. Cox et al. [7] first proposed a new STDM watermarking using Watson's perceptual JND model [8]. In their framework, the projection vectors used in STDM are assigned as the slacks computed by Watson's perceptual model, so as to ensure that more changes are directed to coefficients with larger perceptual slacks. However, it is not robust enough for a volumetric scaling attack. Subsequently, an improved method was proposed [9], where the perceptual model is not only used to determine the projection vector but also used to select the quantization step size. However, it must use many DCT coefficients for one-bit embedding and a low embedding rate can be resulted. In [10], an additional block classification based contrast masking and luminance adaptation was considered by Zhang for digital images. A spatial JND model proposed by Zhenyu Wei [11] incorporates new spatial contrast sensitivity function (CSF), luminance adaptation, and contrast masking. L. Ma proposed to compute the quantization step size by adopting the projection vector and the perceptual slacks from Watson's JND model [12]. Based on the luminance effect, which was only part of Watson's model, X. Li et al [13] proposed a step-projection based scheme that can ensure the values of quantization step size used in the watermark embedder and detector. More recently, Tang [14] presented an improved STDM watermarking scheme based on a more sophisticated luminance-based JND model [15].

Recently, the visual saliency (VS) model has been attracting tremendous attention [16]. As a consequence of evolution, most vertebrates, including humans, have a remarkable ability to automatically pay more attention to salient regions of the visual scene. As we all know, perceptual watermarking should take full advantage of the results from human visual system (HVS) studies. JND gives us a way to model the HVS accurately. VS is another very important aspect affecting human perception. Intuitively, VS and JND are intrinsically related because both of them depend on how the HVS perceives an image and suprathreshold distortions can be a strong attractor of visual attention [17]. Visual saliency can enhance or reduce actual visual sensitivity. Consequently, JND needs to be adjusted to the salient areas in images. Salient areas are generally regarded as the attention focus in human eyes. The visual saliency modulated JND model guided watermarking scheme, where the JND model combined with visual saliency's modulatory effect is fully used to determine image-dependent upper bounds on watermark insertion, allows us to provide the maximum strength transparent watermark. It is widely accepted that incorporating VS features appropriately can benefit JND metric. Ling et al. [18] proposed a simple information hiding algorithm based on Stentiford visual saliency model [19]. Niu et al. [20] proposed visual saliency to modulate just noticeable distortion (JND) profile and guide watermark embedding. Wan et al. [21], presented a watermarking algorithm based on the logarithmic spread transform dither modulation framework (STDM). The authors use the Watson's model [8] to get the JND threshold, and the AC coefficients are used to compute visual saliency map. However, the simple visual saliency model used can not achieve better prediction performance, and the existing VS-based JND models cannot provide maximum performance in terms of robustness while maintaining the fidelity constraint for a practical watermarking framework.

In this paper, we present a novel STDM watermarking scheme based on an effective visual saliency-based JND model. Inspired by our previous watermarking method [14], a new luminance-based JND model [22], which is better correlated with the real visual perception characteristics of HVS, is applied within the VS-based JND model. We claim that the VS map can be used as a feature map to characterize the JND model. The underlying reason is that perceptible JND thresholds can lead to measurable changes of images' VS maps. Consequently, a simple yet very effective VS model is used for the JND task. In our work, the new proposed VS-based JND model is implemented in the STDM watermarking framework. Experimental results show that the proposed watermarking scheme has a superior performance in comparison with existing schemes.

The rest of this paper is organized as follows. Section 2 provides a brief introduction of the conventional perceptual STDM schemes. The robust VS-based JND models for the STDM watermarking framework are proposed in Section 3. Section 4 incorporates the proposed model to design an improved STDM-based watermarking scheme. In Section 5, experimental results are provided to demonstrate the superior performance of the proposed scheme. Finally, Section 6 summarizes the paper.

## 2. Perceptual Spread Transform Dither Modulation

STDm is an important extension of quantized index modulation and uses dither modulation (DM) to modulate the projection of host vector along a given direction. It can provide superior performance compared to DM. It has both the effectiveness of QIM and the robustness of a spread-spectrum system.

STDm differs from regular DM. The host vector  $x$  is first projected onto the random direction vector  $u$  to obtain the projection  $x_u$ , so we can use the traditional dither modulation to embed the watermark bit  $m$  to modulate  $x_u$ . Finally, we get the watermarking information  $y$ :

$$y = x + (Q(x^T u, m, \Delta, \delta) - x^T u) \cdot u, m \in (0, 1) \quad (1)$$

where the quantization modulation  $Q(\cdot)$  is expressed as

$$Q(x^T u, m, \Delta, \delta) = \Delta \cdot \text{round} \left( \frac{x_u + \delta}{\Delta} \right) - \delta, m \in (0, 1) \quad (2)$$

where  $m$  represents the watermark bit that is generated by the pseudo random sequence with number "0" and "1".  $\delta$  is the dither signal corresponding to the message bit  $m$ , which is a pseudo-random number distributed uniformly in  $[-\Delta/2, \Delta/2]$ .  $\Delta$  is the quantization step that will be detailed later.

In the extraction procedure, the received vector  $y'$  can be attacked and projected onto the direction vector  $u$  such that we can get  $y'_u$ . Then the watermark  $m'$  is detected according to the minimum distance detector as follows:

$$m' = \arg \min_{b \in \{0,1\}} |y'_u - Q(y'_u, b, \Delta, \delta)| \quad (3)$$

In general, the motivation of using a perceptual JND model to perform the STDm watermarking scheme is to embed the message bit  $m$  into the host signal  $x$ . The error introduced by the quantization should not exceed the distortion visibility thresholds (slack)  $s$ . Therefore, the watermark will not become perceptible. Here, we provide a new visual saliency-based JND model to the  $s$  associated with each DCT coefficient within the typical STDm watermarking framework.

## 3. Visual Saliency-Based JND Model

In this section, we present the visual saliency-based JND model. The Human Visual System (HVS) is sensitive to many salient features that lead to attention being drawn towards specific regions in a scene and it is a well studied topic in psychology and biology [23]. Visual saliency model is based on the principle of HVS. According to HVS, the different areas of the image have different degrees, so we set different weights for an image. Image processing technology, combined with the visual saliency, can reflect the subjective initiative of the system more accurately. The visual saliency model is of great significance and broad application in many media information processing fields, e.g., segmentation [24], target recognition [25], image target relocation [26], and so on. The proposed model, the visual saliency-based JND model (as Figure 1 shown), is an efficient model to represent the perceptual redundancies. This model includes the contrast sensitivity function (CSF), luminance (LA) effect, and visual saliency (VS) effect. It is defined as:

$$JND(n, i, j) = s \cdot N \cdot M_{base} \cdot M_{LA} \cdot M_{VS} \tag{4}$$

where  $s$  is intended to account for the summation effect of individual JND thresholds over a spatial neighborhood for the visual system and is set to 0.14.  $N$  is the dimension of DCT.  $n$  is the index of a DCT block (we often set  $n = 8$ ) and  $(i, j)$  is the position of  $(i, j)$ -th DCT coefficient. The parameters  $M_{base}$ ,  $M_{LA}$ , and  $M_{VS}$  will be detailed later.

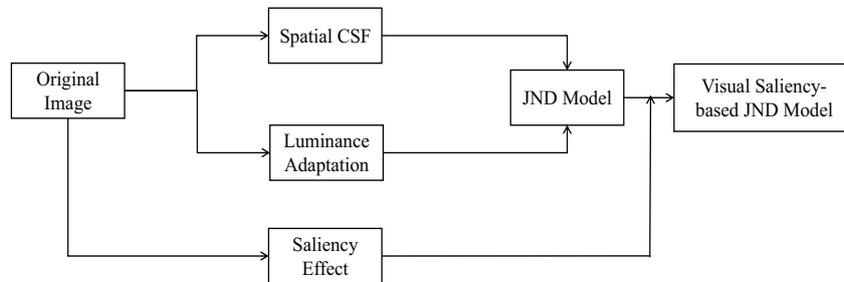


Figure 1. The visual saliency-based JND model.

### 3.1. Spatial CSF Effect

The human visual system is a multi-channel structure that breaks the input image into different sensory components. Each sensory channel has its own threshold (called the visual threshold). If the excitation value is lower than the visual threshold of the channel, the human eye can not feel the incentive. The visual threshold prevents the damage below the threshold from being perceived. On the other hand, the masking increases the visual threshold. The main features of the HVS show that HVS has a band-pass property. It is more sensitive to the noise injected in the DCT basis function along the horizontal and vertical directions than the diagonal direction in spatial frequency. The spatial CSF model describes the sensitivity of human vision for each DCT coefficient.

The base threshold  $M_{base}$  is generated by spatial CSF based on a uniform background image and can be given by considering the oblique effect [27] as

$$M_{base}(\omega_{i,j}, \varphi_{i,j}) = (M_d(\omega_{i,j}) - M_v(\omega_{i,j})) \cdot \sin(\varphi_{i,j})^2 + M_v(\omega_{i,j}) \tag{5}$$

where  $M_d(\omega_{i,j})$  and  $M_v(\omega_{i,j})$  are found as

$$\begin{cases} M_d(\omega_{i,j}) = 0.0293 \cdot \omega_{i,j}^2 + (-0.1382) \cdot \omega_{i,j} + 1.75 \\ M_v(\omega_{i,j}) = 0.0238 \cdot \omega_{i,j}^2 + (-0.1771) \cdot \omega_{i,j} + 1.75 \end{cases} \tag{6}$$

where the  $\omega_{i,j}$  is cycle per degree (cpd) in spatial frequency for the  $(i, j)$ -th DCT coefficient and is given by

$$\omega_{i,j} = \sqrt{i^2 + j^2} / 2N\theta \tag{7}$$

$$\theta = \tan^{-1}[\frac{1}{2} \cdot R_{VH} \cdot H] \tag{8}$$

where  $\theta$  indicates the horizontal/vertical length of a pixel in degrees of visual angle [28],  $R_{VH}$  is the ratio of the viewing distance to the screen height,  $H$  is the number of pixels in the screen height, and  $\varphi_{i,j}$  stands for the direction angle of the corresponding DCT component, which is expressed as

$$\varphi_{i,j} = \sin^{-1}(2 \cdot \omega_{i,0} \cdot \omega_{0,j} / \omega_{i,j}^2) \tag{9}$$

### 3.2. Luminance Adaptation Effect

The LA factor remains sensitive to volumetric scaling since the average intensity does not scale linearly with amplitude scaling, so we need the average intensity to scale linearly with volumetric scaling for robustness. We introduce  $\mu_p$  to describe the pixel intensity and it is expressed as

$$\mu_p = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(i, j)}{KN^2} \cdot \frac{128}{C_0} \tag{10}$$

where  $N$  is the DCT block size,  $I(i, j)$  is the pixel intensity at the position  $(i, j)$  of the block,  $K$  is the maximum pixel intensity, and  $C_0$  denotes the mean intensity of the whole image.

A novel empirical luminance adaptation factor  $M_{LA}$  that employed both the cycles per degree (cpd)  $\omega_{i,j}$  for spatial frequencies and the average intensity value of the block  $\mu_p$  can be formulated as

$$M_{LA}(\omega_{i,j}, \mu_p) = \begin{cases} 1 + (M_{0,1}(\omega_{i,j}) - 1) \left| \frac{\mu_p - 0.3}{0.2} \right|^{0.8} & \mu_p \leq 0.3 \\ 1 + (M_{0,9}(\omega_{i,j}) - 1) \left| \frac{\mu_p - 0.3}{0.6} \right|^{0.6} & \mu_p > 0.3 \end{cases} \tag{11}$$

where the  $M_{0,1}(\omega_{i,j})$  and  $M_{0,9}(\omega_{i,j})$  are empirically set as

$$\begin{cases} M_{0,1}(\omega_{i,j}) = 2.468 \times 10^{-4} \omega_{i,j}^2 + 4.466 \times 10^{-3} \omega_{i,j} + 1.14 \\ M_{0,9}(\omega_{i,j}) = 1.230 \times 10^{-3} \omega_{i,j}^2 + 1.433 \times 10^{-2} \omega_{i,j} + 1.34 \end{cases} \tag{12}$$

### 3.3. Visual Saliency Effect

This part describes  $M_{VS}$  in detail, which fuses the prior information for calculating the significant value distribution of the image.

The proposed visual saliency models are mostly based on a bottom-up framework, but these algorithms have high computational complexity and low calculation efficiency. Here, we use a simple visual saliency model [29], which is driven by low-level visual stimulus in the scene, such as frequency prior and location prior. Watermarking algorithms are for grayscale images mostly. Thus, the color prior in the expression doesn't need to be considered.

Firstly, we resort to band-pass filtering [30] for saliency detection. With respect to the band-pass filter, we adopt the log-Gabor filter [31].

The frequency prior  $I_f$  is defined as

$$I_f = IFFT(FFT(I) \times G) \tag{13}$$

where  $I$  is the carrier image ( $I = (i, j) \in R^2$ ) and  $G$  is the transfer function of a log-Gabor filter in the frequency domain.

Several previous studies have demonstrated that objects near the image center are more attractive to people [32]. That implies locations near the center of the image will be more likely to be "salient" than the ones far away from the center. This prior can be simply and effectively modeled as a Gaussian map. Suppose  $C$  is the center of the image. Then, the "location saliency" at  $X$  under the "location prior" can be expressed as a Gaussian map

$$I_l = \exp\left(-\frac{\|X - C\|_2^2}{\sigma_D^2}\right) \tag{14}$$

As discussed in Ref. [29],  $\sigma_D$  is the experience parameter and often set to 114. Then we can get the saliency value  $I_{VS}$ ,

$$I_{VS} = I_f \times I_l \tag{15}$$

Consequently, the VS value of each block can be shown,

$$\mu_{VS} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I_{VS}(i, j)}{N^2} \tag{16}$$

where  $N$  is the DCT block size.  $I_{VS}(i, j)$  is the visual saliency at the position  $(i, j)$  of the block, and  $\mu_{VS}$  is the final block-based VS map. The saliency value distribution of the image obtained by the VS model is a gray scale image with a range of  $[0, 1]$ . The closer pixel value is to “1”, the higher degree of saliency is. The closer the pixel value is to the area, the higher the degree of saliency. We set a threshold (set to 0.3) to binarize the VS map into “saliency” area and “non-saliency” area. HVS is more sensitive to changes in the “saliency” area so it is easier to perceive distortion of the image. According to the image area of the “saliency” or “non-saliency” area, we take different weights to modulate JND model. In the “saliency” area of the image, the quantization step size is reduced. However, in the “non-saliency” area, the quantization step is increased. By this way, we can reduce the distortion.

Then the  $M_{VS}$  can be described as:

$$M_{VS} = \begin{cases} 1, \mu_{VS} \geq 0.3 \\ 1.15, \mu_{VS} < 0.3 \end{cases} \tag{17}$$

where  $M_{VS}$  is the parameter of  $JND(n, i, j)$  in Equation (4) and  $\mu_{VS}$  is the final block-based VS map as Equation (16) shown.

#### 4. STDM Watermarking Using Visual Saliency-Based JND Model

The motivation of using a perceptual model to perform the watermarking scheme is to embed the message bit  $m$  into the host signal  $x$  and the error introduced by the quantization should not exceed the distortion visibility thresholds (called slack)  $s$ , otherwise the watermark will become perceptible. In practice, for image signals, our proposed perceptual model can be served as the foundation to calculate  $s$  as follows:

$$s(n, i, j) = JND(n, i, j) \tag{18}$$

where  $JND(n, i, j)$  is defined as Equation (4). The maximum imperceptible changes  $s$  in the direction of  $u$  can be given as  $s_u$ . Since the maximum quantization error of the quantizer is  $\Delta/2$ , the quantization step  $\Delta$  in the STDM quantizer is given by  $\Delta = 2s_u$  [13].

##### 4.1. Watermark Embedding Procedure

As shown in Figure 2, a new method is adopted in this section. Luminance STDM watermarking using visual attention-based JND model is composed of embedding and detection procedures.

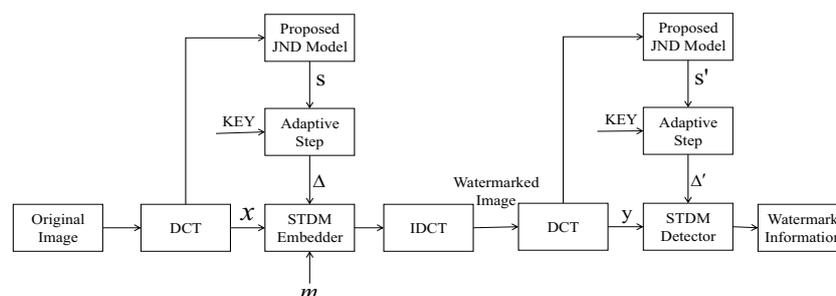


Figure 2. The proposed watermarking scheme.

The watermark embedding process as follows:

1. Calculate the saliency value of each pixel in the carrier image, according to Equations (13)–(17), and binarize the gray scale image according to the threshold to obtain a region of “saliency” area and “non-saliency” area of the image.
2. Divide the carrier image into eight by eight blocks and perform a DCT transform to determine the DCT coefficients. The coefficients are scanned by zig-zag arrangement. Then select the four to ten DCT coefficients (in zig-zag-scanned order after the eight by eight block DCT on each image) to form a single vector which denoted as the host vector  $x$ .
3. The JND coefficients are calculated according to Equation (4). According to step (1), the JND coefficients belonging to the “saliency” or the “non-saliency” areas are multiplied by the different modulation factors to obtain the coefficients and form the visual redundancy vector  $s$ .
4. Following Equation (18), we can get the perceptual slack vector  $s$ . Then the host vector  $x$  and the perceptual slack vector  $s$  are projected onto the given projection vector  $u$ , which is set as the KEY, to generate the projections  $x_u$  and  $s_u$ . Then we can obtain the quantization step size  $\Delta$  via  $s_u$ , which can be multiplied by the embedding strength in practice.
5. One bit of the watermark “ $m$ ” is embedded in the host projection  $x_u$ .
6. Finally, the modified coefficients are transformed to obtain the watermarked image.

#### 4.2. Watermark Detection Procedure

Watermarked images are likely to be subjected to some signal processing attacks during propagation such as salt and pepper noise, Gaussian noise, and JPEG compression. At the detection, the specific steps of the watermark extraction process are as follows:

1. Calculate the saliency value of each pixel in the carrier image, according to Equations (13)–(17), and binarize the gray scale image according to the threshold to obtain a region of “saliency” area and “non-saliency” area of the image.
2. Divide the carrier image into eight by eight blocks and perform DCT transform to determine the DCT coefficients. The coefficients are scanned by zig-zag arrangement. Then select the four to ten DCT coefficients (in zig-zag-scanned order after the eight by eight block DCT on each image) to form a single vector which denoted as the host vector  $y$ .
3. The JND coefficients are calculated according to Equation (4). According to step (1), the JND coefficients belonging to the “saliency” or the “non-saliency” area are multiplied by the different modulation factors to obtain the coefficients and form the visual redundancy vector  $s'$ .
4. The host vector  $x$  and the perceptual slack vector  $s'$  are projected onto the given projection vector  $u$ , which is set as the KEY to generate the projections  $x'_u$  and  $s'_u$ . Then we can obtain the quantization step size  $\Delta$  via  $s'_u$ , which can be multiplied by the embedding strength in practice.
5. Use the STDM detector to extract the watermark message  $m'$  according to Equation (3).

## 5. Experimental Results and Analysis

To evaluate the performance of our proposed scheme, we used standard images with dimensions of  $256 \times 256$  from the USC-SIPI image database [33]. A random binary message of length 1024 bits was embedded into each image. Specifically, we selected the four to ten DCT coefficients (in zig-zag-scanned order after the eight by eight block DCT on each image) to form the host vector and embedded one bit in it. The bit error rate (BER) was computed for comparison purposes.

The experiments were conducted to compare the performance of the proposed scheme and other proposed STDM improvements, termed as STDM-RW [7], STDM-AdpWM [12], STDM-RDMWm [13], and LSTDM-WM [34]. Meanwhile, three kinds of attacks (Gaussian noise with mean zero variance ranging from 0 to 15; JPEG compression, where the JPEG quality factor varies from 20 to 100; and volumetric scaling attacks that can reduce the image intensities as scaling factor varies from 0.1 to 1.5) were used to verify the performance of the proposed models.

5.1. Experiment of Robustness with SSIM = 0.982

The test images were watermarked with a uniform fidelity, a fixed structural similarity index (SSIM) of 0.982. SSIM is a measure of the similarity of two indicators of the image. The overall SSIM image quality index for two images is computed by averaging the SSIM index  $SSIM(x, y)$  values computed for small patches of the two images [35].

$$SSIM(x, y) = \frac{(2\bar{x}\bar{y} + C_1)(2\sigma_{xy} + C_2)}{(\bar{x}^2 + \bar{y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{19}$$

where  $\bar{x}$  and  $\bar{y}$  represent the mean of the luminance of the carrier image and the watermarked image,  $\sigma_x$  and  $\sigma_y$  represent the variances of the luminance of the carrier image and the watermarked image, and  $\sigma_{xy}$  represents the covariance of the above both.

Figure 3 shows the response to additive white Gaussian noise. The STDM-AdpWM and STDM-RW perform significantly worse because of the mismatch problem. Our proposed scheme did not exceed 15% for the Gaussian noise with variance  $1.5 \times 10^{-3}$ . Our proposed scheme has average BER values 2% lower than the LSTDM-WM. Obviously, our scheme outperforms others in the noise-adding-attacks in particular.

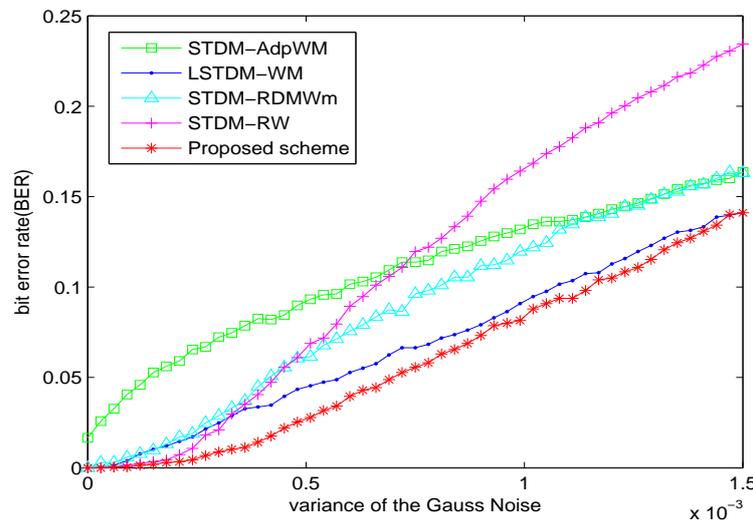


Figure 3. BER versus Gaussian noise for different watermarking algorithms with SSIM = 0.982.

The sensitivity to JPEG compression is demonstrated in Figure 4. From the robustness results, the STDM-AdpWM performs significantly worse. The STDM-RW performs worse than the LSTDM-WM. Our proposed scheme, on the other hand, has average BER values 3% lower than the STDM-RDMWm and 2% lower than the LSTDM-WM. Our model outperforms other models in robustness against JPEG compression.

As shown in Figure 5, all the schemes, except STDM-RW, do have robustness to volumetric scaling due to the mismatch problem within the watermark embedding. Although the other three algorithms showed passable robustness to this attack, our proposed scheme has the best performance.

Moreover, robustness of the above schemes is tested under some common attacks with the  $SSIM = 0.982$  and the results are depicted in Table 1. From the average values, we can see our scheme outperforms other schemes and has good robustness for some common attacks.

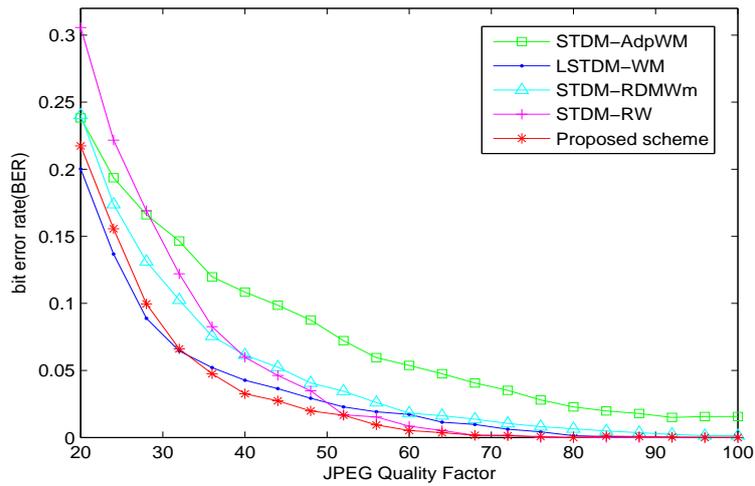


Figure 4. BER versus JPEG compression for different watermarking algorithms with SSIM = 0.982.

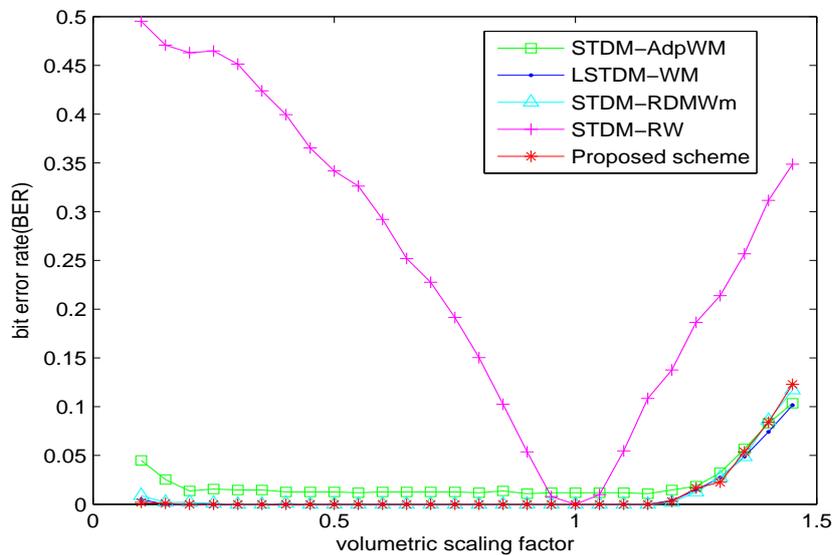


Figure 5. BER versus volumetric scaling factor for different watermarking algorithms with SSIM = 0.982.

Table 1. The BER comparisons for some common attacks with SSIM = 0.982.

Attack	STDM-RW	STDM-AdpWM	STDM-RDMWm	LSTDW-WM	Proposed
Salt-and-peppers noise 0.015	0.2881	0.2292	0.2393	0.2185	0.2093
Wiener filtering 3 × 3	0.1611	0.1264	0.1889	0.1533	0.1189
Median filtering 3 × 3	0.1631	0.1201	0.1194	0.1631	0.1130
Average	0.2041	0.1585	0.1825	0.1783	0.1471

5.2. Experiment of Robustness with VSI = 0.982

The scheme described in this paper is based on visual saliency JND model. Therefore, based on the general image quality evaluation standard PSNR and SSIM, we adopt VSI fusing the visual

attention characteristic to quality image [16]. In the visual saliency-induced index (VSI), the role of VS is twofold. First, VS is used as a feature when computing the local quality map of the distorted image. Second, when pooling the quality score, VS is employed as a weighting function to reflect the importance of a local region. Calculate the distribution maps of the original image and watermarked image of the significant distribution map through the VS model. Extensive experiments performed on four large-scale benchmark databases demonstrate that the IQA index VSI works better in terms of the prediction accuracy than all state-of-the-art IQA indices. The VSI index is given as

$$VSI = \frac{\sum_{i=1}^M \sum_{j=1}^N S_{VS}(i, j) \cdot [S_G(i, j)]^{0.4} \cdot \max[VS_1(i, j), VS_2(i, j)]}{\sum_{i=1}^M \sum_{j=1}^N \max[VS_1(i, j), VS_2(i, j)]} \quad (20)$$

This paper only considers the quality evaluation of gray image, so the color feature of the original text is not taken into account.  $VS_1$  and  $VS_2$  are the VS map of the carrier and distorted image.  $S_{VS}$  is the similarity between  $VS_1$  and  $VS_2$ .  $S_G$  is the similarity of gradient model.

$$S_{VS} = \frac{2VS_1 \cdot VS_2 + C_1}{VS_1^2 + VS_2^2 + C_1} S_G = \frac{2G_1 \cdot G_2 + C_2}{G_1^2 + G_2^2 + C_2} \quad (21)$$

The gradient model  $G$  is calculated as Ref. [16].  $C_1$  and  $C_2$  are positive experience values.

The test images were watermarked with a uniform fidelity, a fixed VSI of 0.982. The bit error rate (BER) is computed for comparison purposes. The comparisons of the test results of BER under various attacks are given as follows.

Figure 6 shows the bit-error-rate (BER) as a function of the additive white Gaussian noise strength. Results form both our STDM watermarking using visual saliency-based JND model and other STDM algorithms are given as follows. Our proposed scheme does not exceed 8% for the Gaussian noise with variance  $1.5 \times 10^{-3}$   $v$  and outperforms other schemes for the Gaussian noise clearly, especially being 6% and 2% lower than the STDM-RDMWm scheme and the LSTDM-WM scheme, respectively.

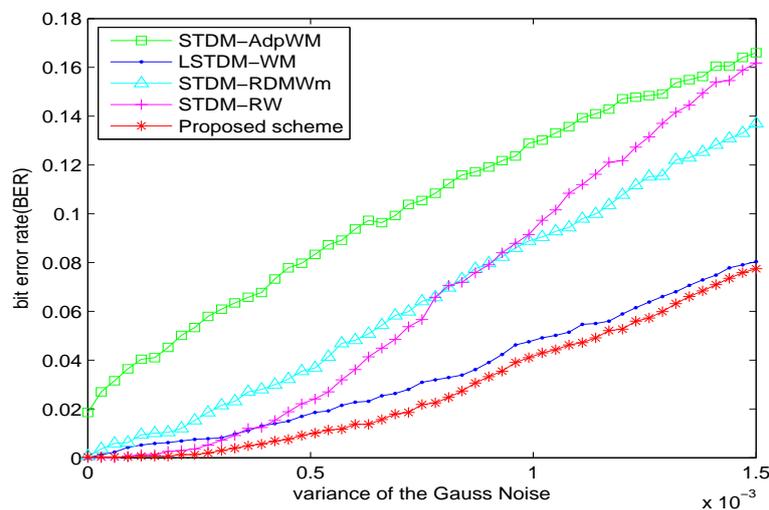


Figure 6. BER versus Gaussian noise for different watermarking algorithms with VSI = 0.982.

The JPEG compression is implemented to the watermarked images. The experimental results are demonstrated in Figure 7. From the robustness results, the proposed scheme outperforms STDM-RW [7], STDM-AdpWM [12], STDM-RDMWm [13], and LSTDM-WM [34] schemes. Our proposed scheme has average BER values 3% lower than the STDM-RW and STDM-RDMWm schemes. The superior performance of the proposed scheme is achieved by the superior robustness properties of our proposed STDM scheme.

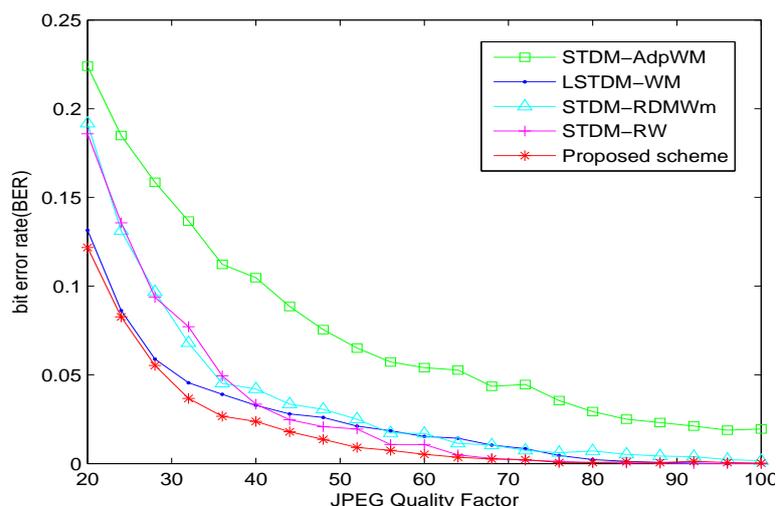


Figure 7. BER versus JPEG compression for different watermarking algorithms with VSI = 0.982.

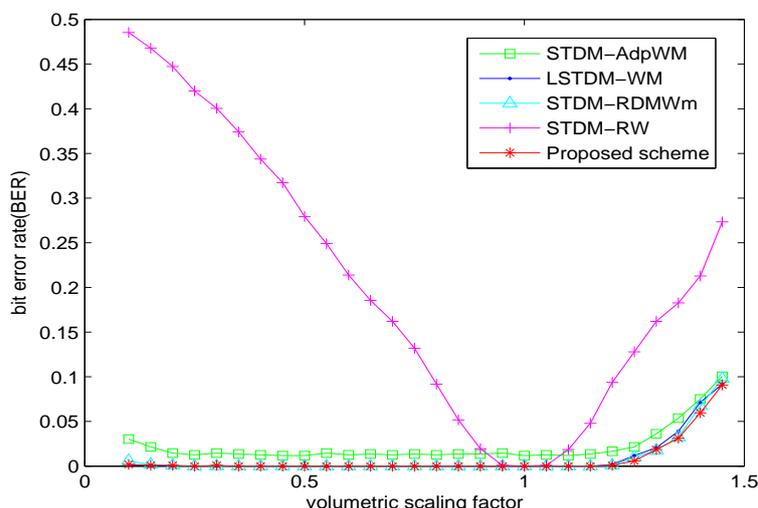


Figure 8. BER versus volumetric scaling factor for different watermarking algorithms with VSI = 0.982.

From the robustness results with volumetric scaling in Figure 8, the watermarking scheme based on visual saliency JND model performs better than others. The proposed method has good robustness for volumetric scaling attack.

Table 2. The BER comparisons for some common attacks with VSI = 0.982.

Attack	STDM-RW	STDM-AdpWM	STDM-RDMWm	LSTDM-WM	Proposed
Salt-and-peppers noise 0.015	0.2129	0.1885	0.1885	0.1221	0.1145
Wiener filtering 3 × 3	0.0869	0.0938	0.0549	0.1064	0.0479
Median filtering 3 × 3	0.1143	0.1094	0.0840	0.1038	0.0645
Average	0.1380	0.1306	0.1091	0.1107	0.0756

Similar to the above, robustness of the above schemes is tested under some common attacks with the  $VSI = 0.982$  and the results are depicted in Table 2. From the average values, we can see our scheme outperforms other schemes and has good robustness for some common attacks.

## 6. Conclusions

We have proposed a novel STDM watermarking based on visual saliency-based JND model, which gives us an improved way to model the HVS in the watermarking algorithms. According to the psychologists research, visual saliency plays an important role in the JND map analyses. In this regard, the spatial CSF, luminance adaption effect and visual saliency effect are introduced to calculate the slacks at the watermark embedder and detector. In order to extract the saliency information, a simple but more effective VS model is introduced and the visual attention effect is investigated for the JND model. Then a comprehensive JND model is defined with the combinations of the VS feature. Finally, by using the new JND model, a novel perceptual STDM watermarking scheme with a better trade-off between robustness and fidelity is presented. Experiments determine that our proposal produces powerful resistance against common attacks compared to other STDM watermarking algorithms.

**Acknowledgments:** This work is partially supported by the Natural Science Foundation of China (No. 61601268), Natural Science Foundation of Shandong Province (ZR2016FB12, ZR2014FM012), Key Research and Development Foundation of Shandong Province (2016GGX101009).

**Author Contributions:** Chunxing Wang and Wenbo Wan conceived and designed the study. Teng Zhang performed the experiments and wrote the paper. Wenbo Wan, Xiaoyue Han and Meiling Xu reviewed and edited the manuscript. All authors read and approved the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kang, H.; Iwamura, K. Information Hiding Method Using Best DCT and Wavelet Coefficients and Its Watermark Competition. *Entropy* **2015**, *17*, 1218–1235.
2. Abdullatif, M.; Zeki, A.M.; Chebil, J.; Gunawan, T.S. Properties of digital image watermarking. In Proceedings of the 2013 IEEE 9th International Colloquium on Signal Processing and Its Applications, Kuala Lumpur, Malaysia, 8–10 March 2013; pp. 235–240.
3. Qi, H.; Zheng, D.; Zhao, J. Human visual system based adaptive digital image watermarking. *Signal Process.* **2008**, *88*, 174–188.
4. Papakostas, G.; Tsougenis, E.; Koulouriotis, D. Fuzzy knowledge-based adaptive image watermarking by the method of moments. *Complex Intell. Syst.* **2016**, *2*, 205–220.
5. Chen, B.; Wornell, G.W. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theor.* **2001**, *47*, 1423–1443.
6. Yu, D.; Ma, L.; Wang, G.; Lu, H. Adaptive spread-transform dither modulation using an improved luminance-masked threshold. In Proceedings of the 15th IEEE International Conference on Image Processing, ICIP 2008, San Diego, CA, USA, 12–15 October 2008, pp. 449–452.
7. Li, Q.; Doerr, G.; Cox, I.J. Spread Transform Dither Modulation using a Perceptual Model. In Proceedings of the 2006 IEEE Workshop on Multimedia Signal Processing, Victoria, BC, Canada, 3–6 October 2006, pp. 98–102.
8. Watson, A.B. DCT quantization matrices optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV, Proceedings of the IS and T/SPIE's Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, USA, 31 January–5 February 1993; Allebach, J.P., Rogowitz, B.E., Eds.; SPIE: Bellingham, WA, USA, 1993.
9. Li, Q.; Cox, I.J. Improved Spread Transform Dither Modulation using a Perceptual Model: Robustness to Amplitude Scaling and JPEG Compression. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007, pp. II-185–II-188.
10. Zhang, X.H.; Lin, W.S.; Xue, P. Improved estimation for just-noticeable visual distortion. *Signal Process.* **2004**, *85*, 795–808.

11. Wei, Z.; Ngan, K.N. Spatial just noticeable distortion profile for image in DCT domain. In Proceedings of the 2008 IEEE International Conference on Multimedia and Expo, Hannover, Germany, 23–26 June 2008, pp. 925–928.
12. Ma, L.; Yu, D.; Wei, G.; Tian, J.; Lu, H. Adaptive Spread-Transform Dither Modulation Using a New Perceptual Model for Color Image Watermarking. *IEICE Trans. Inf. Syst.* **2010**, *93*, 843–857.
13. Li, X.; Liu, J.; Sun, J.; Yang, X.; Liu, W. Step-projection-based spread transform dither modulation. *IET Inf. Secur.* **2011**, *5*, 170–180.
14. Tang, W.; Wan, W.; Liu, J.; Sun, J. Improved Spread Transform Dither Modulation Using Luminance-Based JND Model. In *Image and Graphics, Proceedings of the 8th International Conference on Image and Graphics, ICIG 2015, Tianjin, China, 13–16 August 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 430–437.
15. Bae, S.H.; Kim, M. A Novel DCT-Based JND Model for Luminance Adaptation Effect in DCT Frequency. *IEEE Signal Process. Lett.* **2013**, *20*, 893–896.
16. Zhang, L.; Shen, Y.; Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **2014**, *23*, 4270–4281.
17. Gu, K.; Zhai, G.; Yang, X.; Chen, L.; Zhang, W. Nonlinear additive model based saliency map weighting strategy for image quality assessment. In Proceedings of the 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP), Banff, AB, Canada, 17–19 September 2012, pp. 313–318.
18. Ling, J.; Liu, J.; Sun, J.; Sun, X. Visual model based iterative AQIM watermark algorithm. *Acta Electron. Sin.* **2010**, *38*, 151–155.
19. Amrutha, I.; Shylaja, S.; Natarajan, S.; Murthy, K. A smart automatic thumbnail cropping based on attention driven regions of interest extraction. In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, Seoul, Korea, 24–26 November 2009; pp. 957–962.
20. Niu, Y.; Todd, R.M.; Anderson, A.K. Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Front. Psychol.* **2012**, *3*, 336.
21. Wan, W.; Liu, J.; Sun, J.; Ge, C.; Nie, X. Logarithmic STDM watermarking using visual saliency-based JND model. *Electron. Lett.* **2015**, *51*, 758–760.
22. Lubin, J. A human vision system model for objective picture quality measurements. In Proceedings of the 1997. International Broadcasting Convention, Amsterdam, Netherlands, 12–16 September 1997; pp. 498–503.
23. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **1980**, *12*, 97–136.
24. Han, J.; Ngan, K.N.; Li, M.; Zhang, H.J. Unsupervised extraction of visual attention objects in color images. *IEEE Trans. Circuits Syst. Video Technol.* **2006**, *16*, 141–145.
25. Rutishauser, U.; Walther, D.; Koch, C.; Perona, P. Is bottom-up attention useful for object recognition? In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; pp. II-37–II-44.
26. Fang, Y.; Lin, W.; Chen, Z.; Tsai, C.M.; Lin, C.W. A Video Saliency Detection Model in Compressed Domain. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 27–38.
27. Rust, B.W.; Rushmeier, H.E. A New Representation of the Contrast Sensitivity Function for Human Vision. In Proceedings of the International Conference on Imaging Science, System, Technology, Las Vegas, NV, USA, 30 June–3 July 1997.
28. Ahumada, A.J., Jr.; Peterson, H.A. Luminance-model-based DCT quantization for color image compression. *Proc. SPIE* **1992**, *1666*, 365–374.
29. Zhang, L.; Gu, Z.; Li, H. SDSP: A novel saliency detection method by combining simple priors. In Proceedings of the 2013 20th IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 15–18 September 2013; pp. 171–175.
30. Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S. Frequency-tuned salient region detection. In Proceedings of the CVPR 2009 IEEE Conference on Computer vision and pattern recognition, Miami, FL, USA, 20–25 June 2009; pp. 1597–1604.
31. Field, D.J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **1987**, *4*, 2379–2394.
32. Judd, T.; Ehinger, K.; Durand, F.; Torralba, A. Learning to predict where humans look. In Proceedings of the 2009 IEEE 12th international conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 2106–2113.

33. USC-SIPI Image Database. Available online: <http://sipi.usc.edu/database/> (accessed on 20 August 2017).
34. Wan, W.; Liu, J.; Sun, J.; Yang, X.; Nie, X.; Wang, F. Logarithmic spread-transform dither modulation watermarking based on perceptual model. In Proceedings of the 2013 20th IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 15–18 September 2013; pp. 4522–4526.
35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).