information

MDPI

Article

Visual Saliency Based Just Noticeable Difference Estimation in DWT Domain

Chunxing Wang [1], Xiaoyue Han [1], Wenbo Wan [2,*], Jing Li [3], Jiande Sun [2] and Meiling Xu [1]

[1] School of Physics and Electronics, Shandong Normal University, Jinan 250014, China; cxwang@sdnu.edu.cn (C.W.); hanxiaoyue2018@163.com (X.H.); xumeiling1@stu.sdnu.edu.cn (M.X.)

[2] School of Information and Engineering, Shandong Normal University, Jinan 250014, China; jiandesun@163.com

[3] School of Mechanical and Electrical Engineering, Shandong Management University, Jinan 250100, China; lijingjdsun@163.com

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

Received: 19 June 2018; Accepted: 13 July 2018; Published: 20 July 2018

Abstract: It has been known that human visual systems (HVSs) can be applied to describe the underlying masking properties for the image processing. In general, HVS can only perceive small changes in a scene when they are greater than the just noticeable distortion (JND) threshold. Recently, the cognitive resources of huma visual attention mechanisms are limited, which can not concentrate on all stimuli. To be specific, only more important stimuli will react from the mechanisms. When it comes to visual attention mechanisms, we need to introduce the visual saliency to model the human perception more accurately. In this paper, we presents a new wavelet-based JND estimation method that takes into account the interrelationship between visual saliency and JND threshold. In the experimental part, we verify it from both subjective and objective aspects. In addition, the experimental results show that extracting the saliency map of the image in the discrete wavelet transform (DWT) domain and then modulating its JND threshold is better than the non-modulated JND effect.

Keywords: just noticeable distortion (JND); visual saliency; discrete wavelet transform (DWT)

1. Introduction

With the fast development of visual sensitivity of a human visual system (HVS) related applications (e.g., image/video compression, watermarking, information hiding, visual quality assessment), there is an increasingly significant demand on incorporating perceptual characteristics into these applications for improved performance. It is well known that the just noticeable distortion (JND), which can describe the minimum visibility threshold of the HVS on visual contents, is widely used for visual redundancy estimation.

In the literature, there are two categories to classify the existing JND models. One is the pixel-wise JND models, and the JND estimation is directly obtained on the original image/video. The other category is the subband-domain JND models, which are calculated in a compressed domain, such as wavelet or discrete cosine transformation (DCT) domain. In this paper, we focus on the wavelet-based JND model, as wavelet decompositions provide invertible multiscale transforms enabling two-dimensional transformation. The usage of wavelet decompositions for perceptual image processing has commonly been used in JPEG2000 encoding and AVS video coding standard.

In general, a typical JND model mainly includes three major effect factors, which are the contrast sensitivity function (CSF), luminance adaptation (LA) effect and contrast masking (CM) effect. Several wavelet-based JND models have been proposed in the past twenty years. For example, the JND model proposed in References [1,2].

However, research shows that the existing JND models give every region in the image equal attention, and so the sensitivity of the HVS is lower in low-attention-level regions. Visual saliency is responsible for defining which areas of a visual content will attract more attention of the human visual system. In general, distortion occurring in an area that attracts the viewer's attention is more annoying than that in other areas.

Consequently, visual attention has become a visual feature in the information processing mechanism of a human visual system. The detection and analysis of the visual attention area have become an important research aspect in the field of image processing. It provides important clues for solving the problem of the gap between the image content and the high-level semantics. In the past few decades, many researchers have proposed many computational models for visual attention successively [3–5]. In this work, we investigate visual attention on the mechanisms of the JND estimations.

After recent years of development, the research on the model of perceived characteristics of a human visual system of images can be divided into two categories [6]. The first category is based on a visual attention model, such as the saliency detection model. Researchers try to find the objects or objects of interest in the image through the saliency detection algorithm to form a saliency map, which uses the saliency map to perform various operations on the image. The second category is based on the visual sensitivity model, such as the minimum perceptible distortion model JND, the minimum perceptible distortion model focusing on the discovery and quantification of the visual perception of image redundancy. The JND model mainly uses the visual masking mechanism of the human eyes. When the coding distortion is less than the human eye sensitivity threshold, the human eye can not perceive it.

For a fixed image, the human brain always pays more attention to high attention areas. Therefore, visual saliency requires adjustment of the visual sensitivity of different areas [6]. The saliency map precisely reflects the distribution of the attention of the human brain. Visual significance can adjust all visual perception levels, including increasing or decreasing actual visual sensitivity. Therefore, the JND profile of an image area and non-saliency area must be adjusted separately.

In order to better apply human eye characteristics to image processing, based on the existing theoretical models of a human visual system, the saliency detection algorithm representing the visual attention model is combined with the minimum perceivable distortion model to propose a new algorithm based on saliency region-first discernible distortion model. The model detects the saliency of the image by using the saliency detection algorithm, and calculates the minimum detectable distortion threshold of the image using the minimum detectable distortion model. Then, the minimum perceivable distortion threshold is adjusted with saliency: the minimum detectable distortion threshold is reduced for areas with more salience, and the minimum detectable distortion threshold is increased for areas with less salience. In the existing literature, there is very little with a combination of visual saliency and the JND model. The more prominent methods are proposed in [6,7]. The JND thresholds of the methods in both of the two documents are all tuned by a fixed set of linear significance adjustment functions. Recently, Reference [8] proposed a saliency-modulated JND model in the discrete cosine transform (DCT) domain, in which the JND threshold is determined by the DCT-based JND model whose size is adjusted by two nonlinear modulation functions for the visual significance of a given image pixel. The results of [8] show that the method performs better than both [6,7]. The effect of visual attention on visual acuity is represented by a mathematical expression that is a perceptual quality significance map (PQSM).

This paper presents a new saliency modulation JND model for better image processing. JND refers to the threshold of visual sensitivity within the visual range, where the position and spatial extent within the visual range of interest are represented by the visual significance. Thus, this gives us inspiration to consider modulatory effects in JND thresholds [6,7]. A new JND model with visual saliency modulation is proposed in this paper. The saliency-modulated JND (SJND) model we propose is a subband JND model and we perform the saliency map and the JND model separately in the wavelet domain. Then, the saliency map is used to modulate the JND model according to the nonlinear saliency modulation function. Our method in the paper is different from the Niu method [7] in three aspects. First, our various transformations are performed in the wavelet domain, and the Niu method is performed in the DCT domain. Second, we utilize two nonlinear functions to modulate, whereas two linear functions are used to modulate the JND thresholds in the Niu method. Third, we propose a systematic and automatic framework to calculate the parameters while the parameters of the linear functions in the Niu method were set experimentally. All of these differences make our approach more advantageous.

The rest of this paper is organized as follows. In Section 2, we talked about the related work. The proposed method is described in Section 3, which combined the novel JND model with a visual saliency model. Then, the experiments' results are provided to demonstrate the superior performance of the proposed method. Finally, we provide the conclusions for this paper.

2. Related Work

2.1. Wavelet Transform

It has been shown by psychophysical study that, when we see images, HVS performs a multi-scale frequency analysis [9]. By definition, a wavelet is an oscillating and decaying function whose integral is zero. The wavelet transform of two-dimensional signals involves recursive filtering and sub-sampling. The signal $f(x,y)$ with side $M \times N$ can be defined [9]:

$$[W_{\varphi}(j_0,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f\varphi_{j_0,m,n}] \tag{1}$$

and

$$[W_{\psi}^{i}(j_0,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f\psi_{j_0,m,n}^{i}], \tag{2}$$

where $\varphi(x)$ and $\psi(x)$ are the scaling function and the wavelet function, respectively; j_0 is an arbitrary initial scale; $i = \{H, V, D\}$. The essence of wavelet decomposition is that the image is divided into two parts, namely the high frequency part and low frequency part. The low frequency part we denoted as LL, which usually contains the signal of the main information, and the high frequency part is associated with noise and disturbance. The horizontal high-frequency information (LH), the vertical high-frequency information (HL) and the diagonal high-frequency information (HH) compose the high frequency part. The high frequency coefficient consists of three coefficients: horizontal detail coefficients, vertical detail coefficients and diagonal detail coefficients. These three coefficients are represented by H, V and D, respectively.

According to the needs of analysis, the low-frequency part can be decomposed continuously, so that the image of the lower-frequency part and the image of the higher-frequency part can be obtained.

2.2. JND Estimations

In recent years, the JND model has drawn much attention in the field of image processing and has been used in video coding and video transmission based on visual characteristics, digital watermarking, image quality evaluation and display technology. Several JND models have been proposed. These JND models can be divided into two types: pixel domain JND model and frequency domain JND model(subband JND model). The pixel domain JND model is based on the characteristics of pixel values to establish the model. The frequency domain JND model is based on the pixel value in the transform domain characteristics of the model established, the usual transform domain, the DCT domain and the wavelet domain.

The pixel-based JND threshold in [10,11] is determined by two factors, background brightness masking and spatial contrast masking. Yang's JND model [12] is an improvement of the JND model in [11], which deduces the overlap effect between the background brightness masking and the spatial contrast masking, and obtains a more accurate JND model.

Although the pixel-domain JND model can also describe HVS, satisfactory results can not be achieved without considering the spatial contrast function, which describes the human eye's sensitivity to various frequency components. Because of this, the spatial contrast function is later integrated into the JND model. Since then, hot research has been shifted to subband JND models. An early subband JND model was described in [13]. The threshold of the model is determined by the subbands of luminance and texture masking.

Ahumada and Peterson [14] proposed a JND model within the DCT domain that has been referenced many times. In this model, the threshold of each DCT component is determined by the spatial contrast sensitivity function, which became the basis for many later JND models. The DCTune model [15] was an improvement on the Ahumada model, which introduced the brightness masking feature and the contrast masking feature into the Ahumada model to make the model better match the visual redundancy. In the document [16], the DCTune model was modified to work with a foveal area instead of a single pixel and was used in perceptual image coding. Recently, Bae [17] proposed a novel DCT-based luminance adaptation-JND model that takes into account its frequency characteristics. The letter also revealed that the luminance adaptation effect of HVS depends not only on background luminance but also on frequency in the DCT domain.

In the wavelet domain, it is necessary to examine the influence of a human visual system on image frequency sensitivity, brightness sensitivity and texture masking on JND. Since people's sensitivity to noise is not the same in different frequency bands, and human sensitivity to noise is different even in different directions of the same frequency band, when calculating frequency sensitivity, the influence of frequency bands, directions, etc. on the human visual system must be considered. The most common visual model in the wavelet domain is the JND model described in [1]. In the literature [2], the JND model was also based on the wavelet domain, and it constructed a mathematical model for DWT noise detection thresholds including orientation, and displaying visual resolution. Existing models rarely take saliency feature as a consideration, so, based on these studies, we propose a new JND model in wavelet domain, which takes saliency feature into account.

2.3. Visual Saliency Model

When you see an image, the human senses always automatically focus on areas that stimulate the optic nerve and then focus on the rest. The area of interest refers to the fixed point or object that human observers notice at first glance. Visually, people are always able to quickly focus on their own interests, according to the visual space of a variety of mutation information such as sports, flicker, sharp edges, and the basic information of the object. The use of selective attention strategies and active perception solves computational complexity issues, and use the human visual system to automatically generate a saliency map of the target image or video sequence. One of the basic principles of saliency detection is to suppress low-frequency features to highlight areas with high-frequency features, where the detection process can be summarized as: giving salient features and then determining the salient directions of extraction by different feature comparison methods to extract significant clues, and finally use different combined methods to get the final saliency map.

Most of the existing image saliency detection methods are based on the spatial domain and the frequency domain. Itti et al. [3] proposed a saliency detection model based on the spatial domain in 1998. The model had three key points. First of all, as a whole, the model was bottom-up. The saliency map was calculated by using the three bottom features of image color, intensity and direction. Secondly, locally, the model decomposed the original image into multi-scale sequence of three channels of color, brightness and direction, used the center-surround method to obtain the feature maps of each channel, normalized the sequence of feature maps of each channel to form multiple conspicuity maps and, finally, the saliency map was linearly combined with the conspicuity maps. In 2013, Xu [18] proposed a spatial domain saliency detection algorithm. It is based on space constraint features and contrast color features, combined with multi-scale segmentation algorithms. Fang [19] proposed a visual saliency detection model and applied it to adaptive scaling of images. The model extracted the intensity, color, and texture features of the image from the DCT coefficients.The time-frequency characteristics of wavelet transforms make people put more and more research on saliency detection models based on DWT. Tian et al. [20] presented a salient point detector. The detector tested global variations and local ones based on wavelet transform. However, these models had drawbacks, they were either computationally expensive or had variable parameters. Recently, a saliency detection model using low-level feature calculations based on wavelet transform has been proposed [21], which exploits low-level features to obtain saliency maps. This method has achieved better results. Visual saliency is a fundamental problem in both cognitive and computational sciences, including computer vision. Li [22] used a visual saliency model and discovered that a high-quality visual saliency model can be learned from multiscale features extracted using deep convolutional neural networks (CNNs), which have had many successes in visual recognition tasks. Recently, Wang [23] proposed an unsupervised bottom-up saliency detection model by exploiting novel graph structure and background priors.

Our JND model is based on saliency modulation, which is modulated in wavelet domain by a nonlinear function and the function combines saliency features with the JND model well.

3. Proposed Model

The saliency-modulated JND model combines visual attention with the visual sensitivity of the human eye to represent the additional accurate perceptual redundancies of the digital image. Experiments show that it is a very effective JND model. We require a visibility threshold for each DWT coefficient, and here we use a saliency modulated JND profile. Figure 1 shows this process.

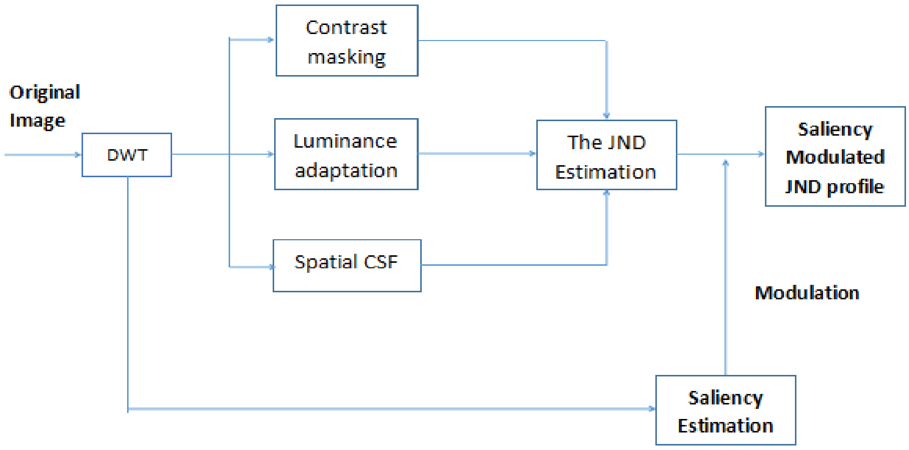

Figure 1. Proposed framework for SJND estimation in wavelet domain.

3.1. Overall JND Modeling

In image processing technology, the JND model based on the human visual system is commonly used, which provides a basic visual model that depends on the visual conditions and the characteristics of the image itself. The JND model that we propose in this article is a combination of three aspects, including spatial CSF, luminance masking and contrast masking.

3.1.1. Complete JND Estimation

$$JND(\lambda,\theta,i,j) = SF(\lambda,\theta,i,j)L(\lambda,\theta,i,j)T(\lambda,\theta,i,j). \tag{3}$$

Based on the above three elements, we can draw the JND threshold to be used in this article. $SF(\lambda,\theta,i,j)$, $L(\lambda,\theta,i,j)$ and $T(\lambda,\theta,i,j)$ represent the spatial CSF, luminance masking and contrast masking respectively. i and j are position coordinates.

3.1.2. Spatial CSF

Contrast sensitivity function is the most basic visual theory model. The model has nothing to do with the content of video images, just depending on the human eye to watch the video perspective. Our eyes have the band-pass feature in the spatial frequency domain.

The spatial CSF model is numerically interpreted as the reciprocal of the fundamental distortion threshold that each DWT coefficient can tolerate. We can calculate the base threshold according to the following formula:

$$SF(\lambda,\theta,i,j) = \begin{pmatrix} \sqrt{2}, & if\ \theta = HH; \\ 1, & otherwise \end{pmatrix} \frac{1}{H(f)(\lambda,\theta)}, \tag{4}$$

where $H(f)(\lambda,\theta)$ denotes CSF, $\frac{1}{H(f)(\lambda,\theta)}$ is given by [24] as the just perceptual weighting depending on spatial frequency, which describes minimally noticeable sensitivity. λ is the number of wavelet decomposition levels, θ is the direction of wavelet decomposition and $H(f)$ is given by [25], which proposed a widely adopted model for the CSF:

$$H(f) = 2.6(0.0192 + 0.0114f)e^{[-(0.0114)^{1.1}]}. \tag{5}$$

3.1.3. Luminance Masking

The luminance masking effect characterizes the human eye as less sensitive to the darker areas of the image than the lighter areas, and the human eye can more easily detect the brighter areas for the same intensity of noise and can easily ignore darker area distortions. The luminance masking effect only depends on the local features of the image, which is used to measure the effect of distortion perception under the condition of a constant pixel value as the background.

In a scene, the human eye is less sensitive to very bright or very dark areas, and many perceptual models are based on this. For example, in a DWT-based model, we can take advantage of the low-frequency portion of the image to account for local brightness. Inspired by it, Xie and Shen proposed a new model [26], which is obtained in the subband for a given level, including the level of reconstruction approximation, and assessed its local brightness. It can be expressed as

$$L(\lambda,\theta,i,j) = 1 + L'(\lambda,\theta,i,j), \tag{6}$$

$$L'(\lambda,\theta,i,j) = \begin{cases} 1 - x(\lambda,LL,i,j), & if\ \ x(\lambda,LL,i,j) < 0.5, \\ x(\lambda,LL,i,j), & otherwise, \end{cases} \tag{7}$$

where $x(\lambda, LL, i, j)$ is the coefficient value of the DWT at the level λ, (i, j) position of the LL subband. It should be noted that all the subbands mentioned above are normalized before calculating the local luminance masking. The normalized range is [0, 1]. We can see that when the image area is very bright or very dark, the local brightness factor will be the maximum.

3.1.4. Contrast Masking

The human eye has less tolerance to distortions in the edge regions and relatively insensitive to the texture regions, a phenomenon that is compared to the masking effect, which is usually related to the perceived degree of one signal in the presence of the other signal. That is, the visibility of the target component (the target) in the image will change with the presence of other components (the masker). The masking effect is strongest when the spatial frequencies, orientations and positions of the two components are the same.

The same pattern or noise, if placed in a highly textured area, is harder to find than a even area. In other words, the image texture can hide or mask another part of the pattern. The contrast masking effect can be modeled as [27]

$$T(\lambda,\theta,i,j) = T_{self}(\lambda,\theta,i,j)T_{neig}(\lambda,\theta,i,j), \tag{8}$$

where $T_{self}(\lambda,\theta,i,j)$ represents the self-contrast making adjustment factor at the location (λ,θ,i,j), and $T_{neig}(\lambda,\theta,i,j)$ represents the neighborhood contrast masking adjustment factor at the location (λ,θ,i,j).

The model for calculating $T_{self}(\lambda,\theta,i,j)$ is given by Teo and Heeger [28] and can be expressed as [29]

$$T_{self}(\lambda,\theta,i,j) = \max\left\{1, \left(\frac{|v(\lambda,\theta,i,j)|}{SF(\lambda,\theta)L(\lambda,\theta,i,j)}\right)^{\phi}\right\}, \tag{9}$$

where $v(\lambda,\theta,i,j)$ represents the DWT coefficient in position (λ,θ,i,j). For the LL subband, we can set ϕ to 0, and, for other subbands, set ϕ to 0.6.

To calculate $T_{neig}(\lambda,\theta,i,j)$, we must know that the signal determined by the wavelet coefficients in the reconstructed image (DWT basis function) is superimposed on other signals determined by the adjacent wavelet coefficients. Moreover, due to the uncertainty of the phase, spatially adjacent signals also cause some masking effects. The model for $T_{neig}(\lambda,\theta,i,j)$ can be expressed as

$$T_{neig}(\lambda,\theta,i,j) = \max\left\{1, \sum_{k\in neighbors\ of(\lambda,\theta,i,j)} \frac{\left|\frac{v_k}{SF(\lambda,\theta)L(\lambda,\theta,i,j)}\right|^{\delta}}{N_{i,j}}\right\}, \tag{10}$$

where the neighborhood is composed of coefficients in the same subband within the window centered at that location (i,j), $N_{i,j}$ is the number of coefficients of the neighborhood, v_k is the value of each neighborhood coefficient, and ζ is a constant that controls the degree of each neighborhood coefficient.

3.2. Saliency Estimation

About 75% of the information obtained by human sensory organs is visual information. Visually, people are always able to quickly focus on their own interests. The HVS automates the extraction of interesting content at all stages of visual processing, compressing or discarding a large amount of irrelevant information. Visual saliency analysis automatically generates a saliency map of this image to simulate the behavior of the human visual system when looking at an image. Here, we must emphasize that the saliency map in this article is not the use of eye-tracking data, but calculated according to specific formulas. Experiments show that this method is more in line with HVS and has a better effect on the modulation of the JND threshold.

Wavelet transform has the characteristics of multi-resolution analysis. It can show the characteristics of the image and fully reflect the huma visual characteristics. According to [21], the local saliency map is obtained by using the wavelet transform and the feature map, then the global saliency map is obtained based on the probability density function. Finally, the local saliency map and the global saliency map are combined to obtain the final saliency map.

Because we need to get more image contents, we will directly perform wavelet transform on the input image I^{ic} without a filter.

Using haar wavelet to decompose I^{ic},

$$[A_N^c, H_s^c, V_s^c, D_s^c] = WT_N(I^{ic}). \tag{11}$$

N is the largest decomposition level of DWT, $s \in \{1, ..., N\}$, c represents the channel of I^{ic} and $c \in \{L, a, b\}$. A_N^c, H_s^c, V_s^c and D_s^c respectively represent the approximate coefficients, the horizontal coefficients, the vertical coefficients and the diagonal coefficients of the s-th decomposition level in the c-th channel. H_s^c, V_s^c and D_s^c collectively referred to as the detail coefficients.

We use the coefficients of detail to get the feature map by IDWT:

$$f_s^c(x, y) = \frac{(IWT_s(H_s^c, V_s^c, D_s^c))^2}{\eta}. \tag{12}$$

$f_s^c(x, y)$ is the feature map of the scale s(s-th level) of the image c channel, η is the scaling factor, $\eta = 10^4$.

Next, use the feature map to calculate the global saliency map. First, we calculate the probability density function (PDF) based on a feature map:

$$p(f(x, y)) = \frac{e^{(-1/2(f(x,y)-\mu)^T \sum^{-1} (f(x,y)-\mu))}}{(2\pi)^{n/2} |\sum|^{1/2}} \tag{13}$$

with

$$\sum = E\left[(f(x, y) - \mu)(f(x, y) - \mu)^T\right]. \tag{14}$$

μ is the average vector that contains the average of each feature map, i.e., $\mu = E[f]$; T is the transpose operation; $\sum$ represents the $n \times n$ covariance matrix; n is the number of feature vectors, where it includes three color channels and N feature maps for each color channel; n is calculated by the formula $n = 3 \times N$. The covariance matrix's determinant is represented by $|\sum|$. Then, calculate the global saliency map according to Equation (15):

$$s_G(x, y) = (\log(p(f(x, y))^{-1}))^{1/2}. \tag{15}$$

After a global saliency map is calculated, we calculate a local saliency map according to feature map:

$$s_L(x, y) = (\sum_{s=1}^{N} \arg\max(f_s^L(x, y), f_s^a(x, y), f_s^b(x, y))). \tag{16}$$

$f_s^L(x, y)$, $f_s^a(x, y)$ and $f_s^b(x, y)$ are the feature maps of the s-th decomposition level in L, a and b channels, separately. $s_L(x, y)$ is the local saliency map we want to calculate.

We have the global saliency map and the local saliency map and thus combine the two maps to get the final saliency map $s(x, y)$, and this article uses it to modulate the JND model:

$$s(x, y) = M(s_L(x, y) \times e^{s_G(x,y)}). \tag{17}$$

$M(.)$ is used as the nonlinear normalization function to diminish the effect of amplification on the saliency map and it is calculated as

$$M(.) = (.)^{\ln \sqrt{2}} / \sqrt{2}. \tag{18}$$

3.3. Saliency Modulated JND Profile

When the image distortion is less than the JND threshold, it is not perceived by the human eyes. The human visual system is generally attracted by certain significant regions of the image and is likely to pay more attention to the visual details of the interest object. With the above features, the use of JND threshold to remove undetectable redundancy coefficients and the use of generated saliency maps to collaboratively guide JND threshold assignments help us to further reduce image distortion. By significantly adjusting the size of the JND threshold, the effect of improving image quality is achieved.

According to the previous discussion and the mechanism of visual attention, in a pair of images, the visibility threshold of visual attention area is smaller, while the visibility threshold of non-attention area is larger. Thus, we need to adjust the JND profile in the saliency and non-saliency areas of the image. The higher the visual saliency value, the higher the frequency corresponding to the turning point of the spatial contrast masking curve, and the decreasing adaptation tolerances will decrease.

Based on these statements, we propose a new saliency-modulated JND model according to the idea of [7]:

$$SJND(\lambda,\theta,s,i,j,\alpha,\beta) = JND(\lambda,\theta,i,j,SF'(\lambda,\theta))f(s,\beta), \tag{19}$$

where s is the saliency value of the position (x,y) of wavelet decomposition level λ; $JND(\lambda,\theta,i,j,SF'(\lambda,\theta))$ means the JND model; $SF'(\lambda,\theta)$ is obtained by modulating the original spatial frequency through a modulation function. Human eyes show a band-pass property in the spatial frequency domain. Because the modulation function in this paper is calculated based on the human visual characteristics and the saliency map of the image, the modulated spatial frequency can be more perfectly combined with the JND model than the unmodulated spatial frequency. It is calculated by

$$SF'(\lambda,\theta) = SF(\lambda,\theta)t(s,\alpha), \tag{20}$$

where α is a vector of four coefficients, $\alpha = [\alpha_1,\alpha_2,\alpha_3,\alpha_4]$; $t(s,\alpha)$ represents a sigmoid function defined as:

$$t(s,\alpha) = \alpha_1(\frac{1}{2} - \frac{1}{1+\exp(\alpha_2(s-\alpha_3))}) + \alpha_4. \tag{21}$$

Similar to α, β is a vector of four coefficients, $\beta = [\beta_1,\beta_2,\beta_3,\beta_4]$; $f(s,\beta)$ is a sigmoid function defined as:

$$f(s,\beta) = \beta_1(\frac{1}{2} - \frac{1}{1+\exp(\beta_2(s-\beta_3))}) + \beta_4. \tag{22}$$

3.4. Setting the Parameters of the SJND Model

We will now describe how the parameters α and β in the modulation function $t(s,\alpha)$ and $f(s,\beta)$ are set. DWT is performed on the image I, and the saliency modulated just noticeable distortion (SJND) value of each level is calculated according to the formula. We add a random noise to the DWT coefficients of I. Thus, the amplitude of the noise to the DWT coefficients is equal to the SJND threshold at that DWT index. Then, we take the inverse DWT of the I to obtain a noise-added image $I'(\omega)$, as shown in Equation (23) below. For simplicity, let $\omega = (\alpha,\beta)$:

$$I'(\omega) = DWT^{-1}\left(DWT(I) + f \times SJND(\lambda,\theta,i,j,s,\omega)\right). \tag{23}$$

$DWT(I)$ refers to the coefficients after wavelet transform of image I; f is a random matrix of +1 and -1 to avoid creating a fixed artificial spatial pattern; s is the corresponding saliency value; $SJND(\lambda,\theta,i,j,s,\omega)$ is the SJND threshold we obtained above.

We know that if the amount of change in the image is less than the SJND threshold, the difference between $I'(\omega)$ and I cannot be seen. Therefore, it is necessary to adjust the ω value to make the SJND larger, and it is less likely to notice the difference between $I'(\omega)$ and I. For this purpose, we design the following cost function. The purpose is to adjust ω to make the cost function smaller:

$$Q(I'(\omega)\,|I,s) = \frac{PSNR(I'(\omega)\,|I)}{\mu VQ(I'(\omega)\,|I,s)} \tag{24}$$

and

$$\omega^* = \arg\min_{\omega} Q(I'(\omega)\,|I,s), \tag{25}$$

where $PSNR(I'(\omega)\,|I)$ represents the PSNR of $I'(\omega)$ with respect to I; $VQ(I'(\omega)\,|I,s)$ is the visual quality score of $I'(\omega)$ with respect to I. For the perceptual quality metric, without loss of generality, we use the visual saliency-based index(VSI) [30] to represent the value of it. The VSI index is given as:

$$VSI = \frac{\sum_{x\in\Omega} S_{VS}(x)\cdot[S_G(x)]^{0.4}\cdot VS_m(x)}{\sum_{x\in\Omega} VS_m(x)},$$

where

$$VS_m(x) = \max(VS_1(x), VS_2(x)).$$

$VS_1(x)$ and $VS_2(x)$ are the saliency values of the original image and the noised image at position x, respectively. Ω means the whole spatial domain:

$$S_{VS} = \frac{2VS_1\cdot VS_2 + C_1}{VS_1^2 + VS_2^2 + C_1},$$

$$S_G = \frac{2G_1\cdot G_2 + C_2}{G_1^2 + G_2^2 + C_2}.$$

S_G is the similarity of gradient model. In addition, the gradient model G is calculated as [30]. C_1 and C_2 are positive experience values.

μ is a regularization term, which is set to 30 here [8]. It can be known from the above equation that, as the JND threshold increases, the PSNR of $I'(\omega)$ becomes smaller and $VQ(I'(\omega)\,|I,s)$ becomes larger, so the quality of $I'(\omega)$ is better. Therefore, to make the quality of the added noise image $I'(\omega)$ better, the cost function must be minimized, that is, adjusting ω to obtain the maximum SJND threshold.

The modulation function in [8] is linear, whereas the modulation function used in this paper is nonlinear. We have carried out this optimization procedure 60 times with a randomly starting point each time in the experiment. In 57 runs out of 60, the cost function values we get is equal to the minimum cost function, and in the other three runs, the obtained cost function values were above the minimum. That is to say, the optimization procedure yielded the same optimum point in 95% of the conducted runs. The values we get for the parameters of the two modulation functions are $\alpha = [-0.1, 4.98, 0.62, 0.95]$ and $\beta = [0.597, 4, 61, 0.71, 0.85]$. The values found are image independent.

4. Experimental Results and Performance Analysis

The tests were conducted on a Windows platform. The PC was equipped with an Intel(R) Pentium(R) CPU G630 and 8 GB of memory. Images are taken from the image saliency detection of the public database MSRA. The MSRA database provides 1000 natural images and the ground truth of these 1000 images, which are often used as a test and comparison for target detection.

In order to evaluate the SJND model proposed in this paper, we conducted the following experiments to verify.

4.1. Evaluating Saliency Estimation

In this experiment, we use the significance test described earlier in Section 3.2 to get the saliency map. In the saliency map, each pixel corresponds to a significant value, and the larger the saliency value, the more significantly it corresponds. In order to verify the performance of the algorithm in this paper, we compare the saliency map produced by this method with the saliency map generated by the MSNR approach [31].

We select three images from the MSRA [32] database which is commonly used for some saliency or target detection and generate their own salient images using the method of this article and MSNR (multi-scale statistical non-redundancy) approach, respectively. The results are shown in the following Figure 2. By contrast, we can see that the effect of the saliency map produced by this method is obviously better than the saliency map generated by the MSNR approach. From this, it can be concluded that the saliency detection model in this paper has the better performance. Moreover, compared with MSNR approach, this method also has greater advantages in terms of time consumption.

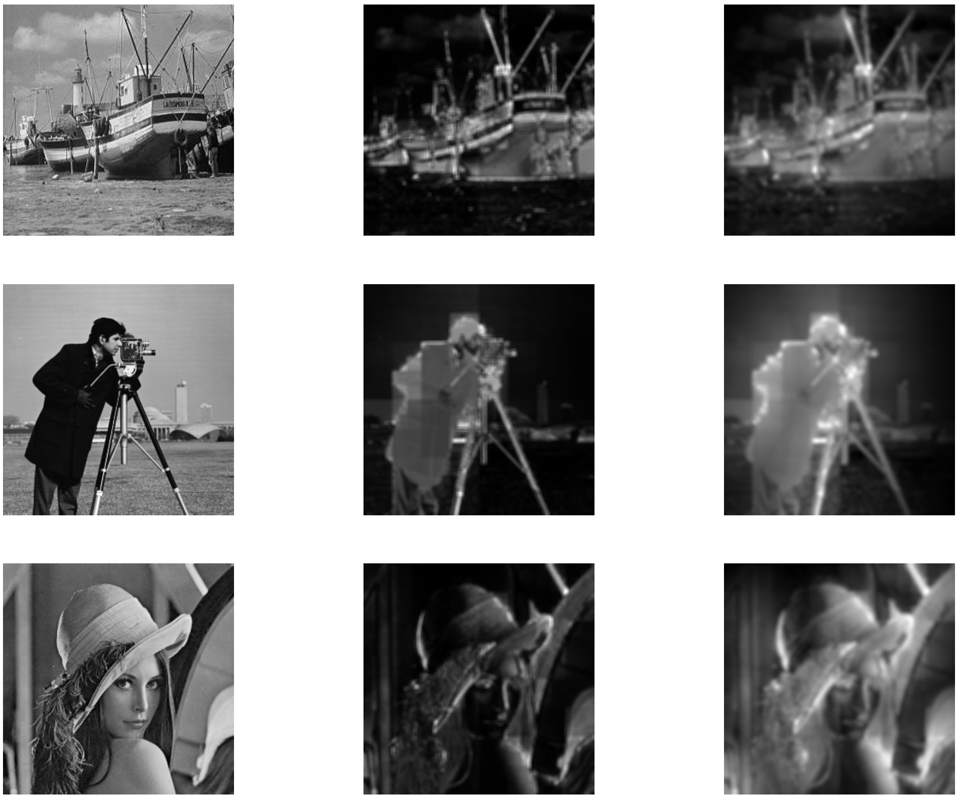

Figure 2. The saliency estimation result of the proposal method in comparison with the MSNR method. From left to right: the original image, the saliency map generated by MSNR method and the saliency map generated by proposal method.

As can be seen from the above, the JND model modulated by the saliency map distributes more noise in the non-significant area and less noise in the salient area, thereby improving the image quality. According to the experiment, the method of this paper can obtain a more accurate saliency map, so the noise can be distributed more accurately. That is to say, the salient map can better adjust the JND model and the generated image quality is better.

4.2. Evaluating Saliency Modulated JND Profile

4.2.1. The Objective Experiment

Based on previous papers, this paper proposes a method for modulating JND values using saliency map. First, we separately verify the image quality of the JND model proposed in [27,33] after modulation with saliency map, which uses our modulation function. We call them the Zebbiche method and Liu method in this paper.

In this section, we first design a subjective experiment to verify the Zebbiche JND model and its saliency modulated JND model. Prior to this, the intensity of the added noise or the control parameter was adjusted so that the PSNR values of the images generated by the two JND models were almost the same, so the less distorted the image, the better the model. PSNR is an objective criterion for evaluating images. In this paper, we adjust the PSNR value to 26 ± 0.01 dB because the distortion is just easier to see at this PSNR value. We selected a Two-Formed Choice (2AFC) method to compare the image quality produced by the two JND model, respectively. The 2AFC approach requires participants to choose one of two options. In this experiment, we asked the participants to choose one of two images with better image quality, where the two images were generated by the Zebbiche JND model and its saliency modulated JND model, respectively.

First of all, we select 20 images in the database, and then add noise to the images according to the two JND model we want to compare, that is, the Zebbiche JND model and its saliency modulated JND model. In each set of experiments, 24 participants were asked to observe two side-by-side images on a mid-gray background, with participants positioned vertically in the same direction and separated horizontally by a distance of 1 cm. The display time for each pair of images is 10 s, and, after 10 s, the screen will change to medium gray with a display time of 5 s. When displaying image pairs on the screen, we asked participants to answer the question of which of the two images looks better. Regardless of the level of certainty of the participants, a choice must be made, left or right. The premise of the experiment is that the participants do not know beforehand which method the images were generated from. In these 20 experiments, we randomly select 10 groups to put the images produced by the Zebbiche JND model on the left side of the screen and the other 10 groups on the right side of the screen to make the images generated by the saliency modulated JND model. The purpose of this is to avoid chance.

The experiment was run in a quiet classroom with 24 participants including 15 males and nine females between 22 and 25 years old. All participants had normal or corrected to normal vision. They are all related people in image processing specialties.

Results are shown in the following Table 1. The first column is the sequence of images, the second column and the third column are the number of times the Zebbiche JND model is selected and the visual saliency modulated Zebbiche JND model is selected. From this, it can be clearly seen that the number of methods for selecting the visual saliency modulated Zebbiche JND model is significantly more than that of the original unmodulated Zebbiche JND model. We test the statistical significance of the experimental results according to the two-sided chi-square test. The null hypothesis means that there is no preference either for the method presented by Zebbiche or for the saliency modulated Zebbiche JND model. The probability that the null hypothesis holds (also known as p-value) [8] is shown in the fourth column of the table. In doing experiments, empirically, the null hypothesis is rejected when $p < 0.05$. As shown in the table, at all $p < 0.05$, this indicates that the images produced by the two methods are of different quality, that is to say there is always one image in each image pair with better quality.

As can be seen from the Table 1, only 2 of the 20 experiments has p values greater than 0.05, so, in the remaining cases, it is clear that the image produced by the method of the saliency modulated Zebbiche JND model yields better image quality than the original unmodulated Zebbiche JND model. According to the above experiment, it can be concluded that most of the participants feel that the images produced by the SJND method are better, and few participants think that the unmodulated JND method produces better results. From this experiment, we can know that the saliency modulated JND model is better than an unmodulated JND model. This is because the saliency modulated Zebbiche JND model can allocate less noise to the salient areas of the image when the image is added with noise, while allocating more noise than the salient areas. Thus, our modulated method is effective. Figure 3 shows two images from Table 1 and their corresponding transformations with each method.

The same method can verify the Liu JND model and its saliency modulated JND model. The experimental results are shown in Table 2. It is observed that, in only three cases out of 20, the *p*-value is greater than 0.05. Then, we interviewed the two participants in the first group of experiments and the three participants in the second group of experiments. They said that they could not tell which of the images in the experiment was of better quality, so they randomly picked one of them.

Table 1. Comparing the Zebbiche JND model and its saliency modulated JND model based on the number of votes collected from 24 subjects.The first 10 images are from Group A, and the others are from Group B.

Image Number	Unmodulated	Modulated	*p*-Value
A1	14	36	0.0019
A2	15	35	0.0047
A3	20	30	0.1573
A4	13	37	0.0007
A5	10	40	0.0001
A6	8	42	0.0001
A7	17	33	0.0237
A8	12	38	0.0002
A9	10	40	0.0001
A10	15	35	0.0047
B1	11	39	0.0001
B2	14	36	0.0019
B3	28	22	0.3961
B4	9	41	0.0001
B5	16	34	0.0109
B6	12	38	0.0002
B7	14	36	0.0019
B8	11	39	0.0001
B9	15	35	0.0047
B10	13	37	0.0007
Total	277	723	0.0001

Table 2. Comparing the Liu JND model and its saliency modulated JND model based on the number of votes collected from 24 subjects. The first 10 images are from Group A, and the others are from Group B.

Image Number	Unmodulated	Modulated	*p*-Value
A1	10	40	0.0001
A2	13	37	0.0006
A3	20	30	0.1573
A4	15	35	0.0047
A5	11	39	0.0001
A6	14	36	0.0002
A7	26	24	0.7773
A8	16	34	0.0002
A9	9	41	0.0001
A10	11	39	0.0001
B1	14	36	0.0019
B2	17	33	0.0237
B3	15	35	0.0047
B4	30	20	0.1573
B5	16	34	0.0109
B6	14	36	0.0002
B7	16	34	0.0109
B8	12	38	0.0002
B9	10	40	0.0001
B10	8	42	0.0001
Total	297	703	0.0001

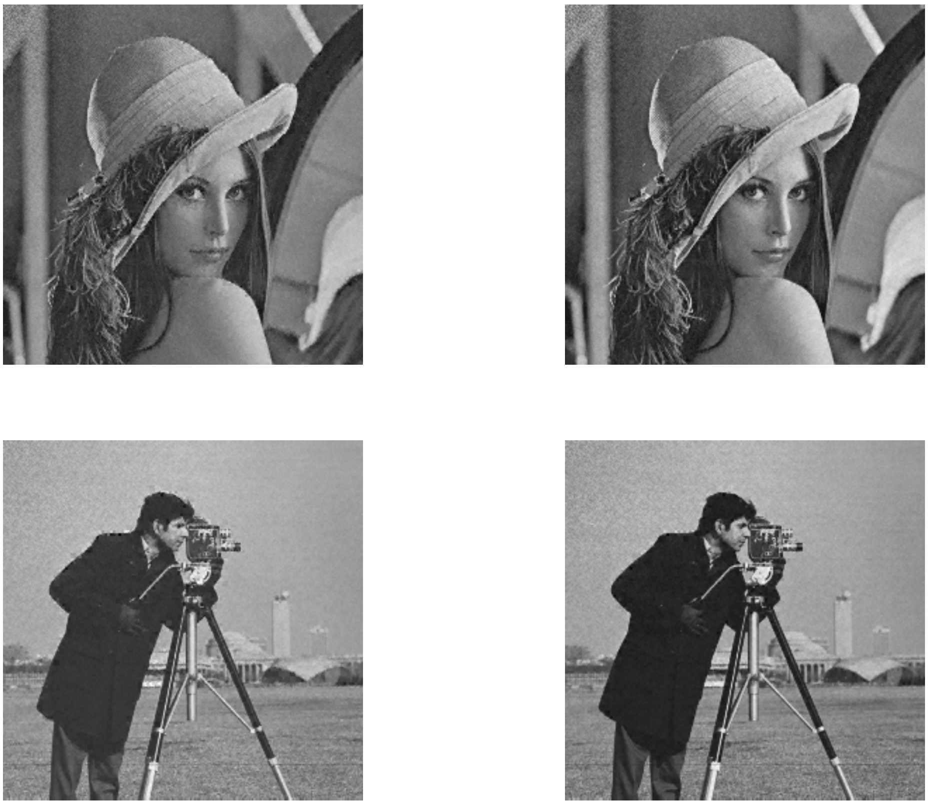

Figure 3. The images from Table 1 and their corresponding transformations with each method. From left to right: the image produced by the Zebbiche JND model and the image produced by the modulated Zebbiche JND model.

4.2.2. The PSNR

Through the above experiments, we can know that using the visual saliency map to modulate on the basis of the original JND model will get better image quality. Next, we continue to test the superiority of this method.

We verify the Liu JND model, K JND model [34] , Zebbiche JND model, Cui JND model [22] and our proposed saliency modulated JND model, respectively. According to these models, noise is added to the 20 images so that the VSI scores of the generated images are the same or similar. We can see that almost all of the images added with noise can not see the presence of noise. However, their PSNR values are not the same. Figure 4 shows five visual examples comparing the PSNR of the proposed method with the other five methods of the same visual quality. The PSNR value produced by the saliency modulated JND model in this paper is 29.7242. The PSNR value produced by the saliency modulated Zebbiche JND model is 30.1301. The PSNR value produced by the saliency modulated Cui JND model is 30.3502. The PSNR value produced by the saliency modulated K JND model is 30.6411. The PSNR value produced by the saliency modulated Liu JND model is 30.8715. The PSNR value in this paper is the smallest and lower than the PSNR value of the unmodulated JND model proposed in this paper. From the above experiment, it can be known that our modulated JND model has lower PSNR values than the other JND model. This shows that, although the saliency modulation model can distribute more noise in the non-saliency area and less noise in the salient area, in general, compared to the unmodulated JND model, the significant modulation in the JND under estimation, the image can add more noise energy.

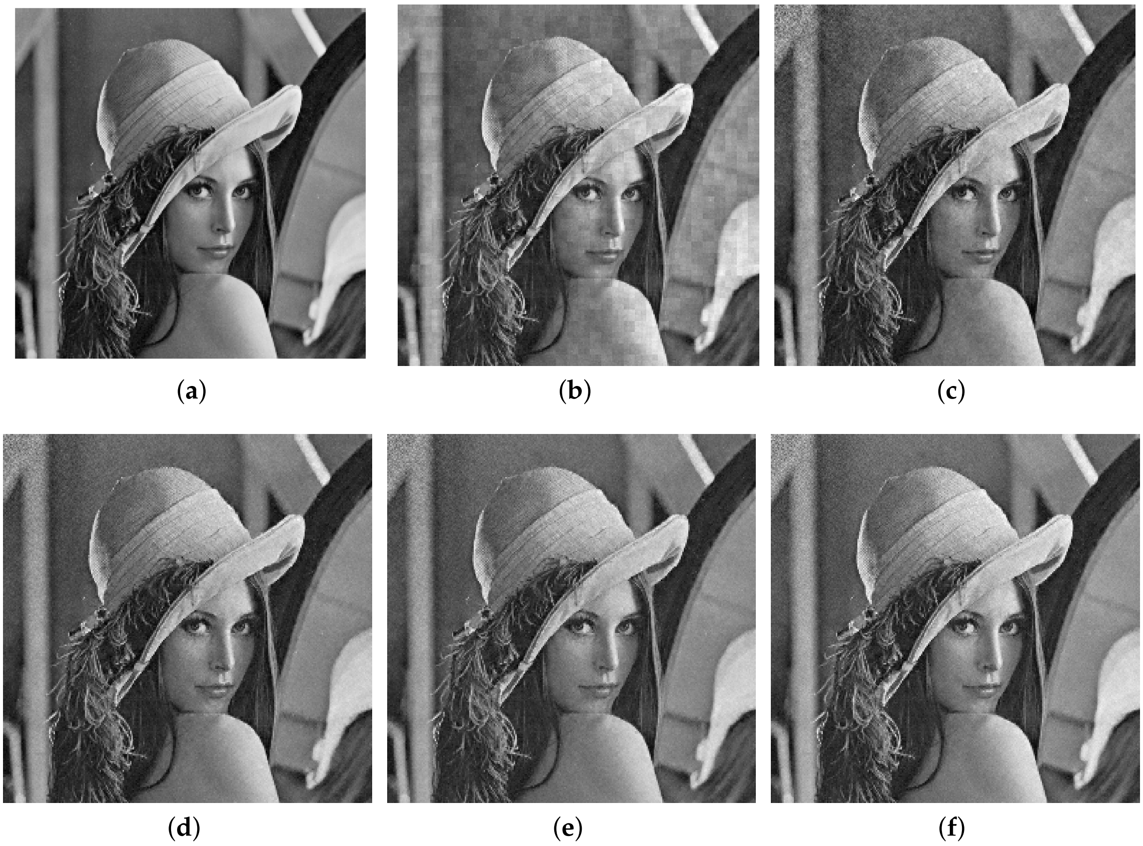
(**a**) (**b**) (**c**)
(**d**) (**e**) (**f**)

Figure 4. The VSI score of the four images is 0.95 ± 0.01. From left to right: (**a**) the original image; (**b**) the image produced by the saliency modulated Liu JND model; (**c**) the image produced by the saliency modulated K JND model; (**d**) the image produced by the saliency modulated Cui JND model; (**e**) the image produced by the saliency modulated Zebbiche JND model; (**f**) the image produced by the proposed saliency modulated method. The PSNR values of the last five pictures are 30.8715, 30.6411, 30.3502, 30.1301 and 29.7242.

5. Conclusions

In this paper, we propose a new saliency-modulated JND model in the DWT domain. Because wavelet transform has the characteristics of multi-resolution analysis, it can show that the image features fully reflect the human visual characteristics, so the method of this paper based on wavelet transform and its modulation is better. For the purpose of assigning different JND thresholds to different areas of the image according to the saliency, we add two nonlinear modulation functions to the proposed model. In order to verify the effectiveness of the proposed method, we conducted two experiments: one is a subjective experiment and the other is to compare PSNR values. Both experiments show that our SJND is much better than unmodified JND and other modulated JND model and it can provide a high distortion hiding capacity.

Author Contributions: Conceptualization, X.H.; Funding acquisition, W.W.; Methodology, X.H.; Writing-original draft, X.H.; Supervision, J.L. and J.S.; Investigation, M.X.; Writing-review and editing, C.W. and W.W.

Funding: This research received no external funding.

Acknowledgments: This work is partially supported by the Natural Science Foundation of China (No. 61601268), the Natural Science Foundation of China (U1736122), the Natural Science Foundation for Distinguished Young Scholars of Shandong Province (JQ201718), the Natural Science Foundation of Shandong Province (ZR2016FB12), the Key Research and Development Foundation of Shandong Province (2016GGX101009) and the Shandong Provincial Key Research and Development Plan (2017CXGC1504).

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

References

1. Xiao, L.; Wei, Z.H.; Wu, H.-Z. A digital watermarking in wavelet domain utilizing huma visual masking. *J. China Inst. Commun.* **2002**, *23*, 100–106.
2. Watson, A.B.; Yang, G.Y.; Solomon, J.A.; Villasenor, J. Visibility of wavelet quantization noise. *IEEE Trans. Image Process.* **1997**, *6*, 1164–1175. [CrossRef] [PubMed]
3. Itti, L.; Koch, C.; Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
4. Wang, H.; Wang, L.; Hu, X.; Tu, Q.; Men, A. Perceptual video coding based on saliency and Just Noticeable Distortion for H.265/HEVC. In Proceedings of the 2014 International Symposium on Wireless Personal Multimedia Communications (WPMC), Sydney, Australia, 7–10 September 2014; pp. 106–111.
5. Wang, C.; Zhang, T.; Wan, W.; Han, X.; Xu, M. A Novel STDM Watermarking Using Visual Saliency-Based JND Model. *Information* **2017**, *8*, 103. [CrossRef]
6. Lu, Z.; Lin, W.; Yang, X.; Ong, E.; Yao, S. Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation. *IEEE Trans. Image Process.* **2005**, *14*, 1928–1942. [PubMed]
7. Niu, Y.; Kyan, M.; Ma, L.; Beghdadi, A.; Krishnan, S. Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking. *Signal Process. Image Commun.* **2013**, *28*, 917–928. [CrossRef]
8. Hadizadeh, H. A saliency-modulated just-noticeable-distortion model with nonlinear saliency modulation functions. *Pattern Recogn. Lett.* **2016**, *84*, 49–55. [CrossRef]
9. Daubechies, I. Orthonormal Bases of Compactly Supported Wavelets II. Variations on a Theme. *SIAM J. Math. Anal.* **1993**, *24*, 499–519. [CrossRef]
10. Chou, C.H.; Li, Y.C. A perceptual tuned subband image coder based on the Measure of just-noticeable-distortion profile. *IEEE Trans. Circuits Syst. Video Technol.* **1996**, *5*, 467–476. [CrossRef]
11. Chin, Y.J.; Berger, T. *A Software-Only Videocodec Using Pixelwise Conditional Differential Replenishment and Perceptual Enhancements*; IEEE Press: Piscataway, NJ, USA, 1999.
12. Yang, X.; Lin, W.; Lu, Z.; Ong, E.; Yao, S. Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. *IEEE Trans. Circuits Syst. Video Technol.* **2005**, *15*, 742–752. [CrossRef]
13. Safranek, R.J.; Johnston, J.D. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Glasgow, UK, 23–26 May 1989; Volume 3, pp. 1945–1948.
14. Ahumada, A.J. Luminance-model-based DCT quantization for color image compression. In *SPIE 1666, Human Vision, Visual Processing, and Digital Display III*; International Society for Optics and Photonics: San Diego, CA, USA, 1992; pp. 365–374.
15. Watson, A.B. DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images. In Proceedings of the 9th Computing in Aerospace Conference, San Diego, CA, USA, 19 October 1993; pp. 946–949.
16. Tong, H.H.Y.; Venetsanopoulos, A.N. A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking. In Proceedings of the 1998 International Conference on Image Processing, ICIP98 (Cat. No. 98CB36269), Chicago, IL, USA, 4–7 October 1998; Volume 3, pp. 428–432.
17. 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.
18. Xu, L.; Li, H.; Zeng, L.; Ngan, K.N. Saliency detection using joint spatial-color constraint and multi-scale segmentation. *J. Vis. Commun. Image Represent.* **2013**, *24*, 465–476. [CrossRef]
19. Fang, Y.; Chen, Z.; Lin, W.; Lin, C.-W. Saliency detection in the compressed domain for adaptive image retargeting. *IEEE Trans. Image Process.* **2012**, *21*, 3888–3901. [CrossRef] [PubMed]
20. Tian, Q.; Sebe, N.; Lew, M.S.; Loupias, E.; Huang, T.S. Image retrieval using wavelet-based salient points. *J. Electron. Imaging* **2003**, *10*, 835–849.
21. Zeng, W.; Yang, M.; Cui, Z.; Al-Kabbany, A. An improved saliency detection using wavelet transform. In Proceedings of the 2015 IEEE International Conference on Communication Software and Networks (ICCSN), Chengdu, China, 6–7 June 2015.

22. Li, G.; Yu, Y. *Visual Saliency Detection Based on Multiscale Deep CNN Features*; IEEE Press: Piscataway, NJ, USA, 2016.
23. Wang, Q.; Zheng, W.; Piramuthu, R. GraB: Visual Saliency via Novel Graph Model and Background Priors. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 535–543.
24. Cui, L.; Li, W. Adaptive multiwavelet-based watermarking through JPW masking. *IEEE Trans. Image Process.* **2011**, *20*, 1047–1060. [PubMed]
25. Mannos, J.; Sakrison, D.J. *The Effects of a Visual Fidelity Criterion on the Encoding of Images*; IEEE Press: Piscataway, NJ, USA, 1974.
26. Xie, G.; Shen, H. Toward improved wavelet-based watermarking using the pixel-wise masking model. In Proceedings of the IEEE International Conference on Image Processing, Genova, Italy, 11–14 September 2005; p. I-689-92.
27. Liu, Z.; Karam, L.J.; Watson, A.B. JPEG2000 encoding with perceptual distortion control. In Proceedings of the 2003 International Conference on Image Processing (Cat. No. 03CH37429), Barcelona, Spain, 14–17 September 2003; Volume 1, p. I-637-40.
28. Teo, P.C.; Heeger, D.J. Perceptual image distortion. In Proceedings of the 1st International Conference onImage Processing, Austin, TX, USA, 13–16 November 1994; Volume 2179, pp. 127–141.
29. Hontsch, I.; Karam, L.J. Adaptive image coding with perceptual distortion control. *IEEE Trans. Image Process.* **2002**, *11*, 213–222. [CrossRef] [PubMed]
30. 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. [CrossRef] [PubMed]
31. Scharfenberger, C.; Jain, A.; Wong, A.; Fieguth, P. Image saliency detection via multi-scale statistical non-redundancy modeling. In Proceedings of the 2014 IEEE International Conference onImage Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4294–4298.
32. Liu, T.; Yuan, Z.; Sun, J.; Wang, J.; Zheng, N.; Tang, X.; Shum, H.Y. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 353–367. [PubMed]
33. Zebbiche, K.; Khelifi, F. Efficient wavelet-based perceptual watermark masking for robust fingerprint image watermarking. *IET Image Process.* **2014**, *8*, 23–32. [CrossRef]
34. Liu, K.C. Wavelet-based watermarking for color images through visual masking. *AEUE Int. J. Electron. Commun.* **2010**, *64*, 112–124. [CrossRef]