Signal Processing: *Image Communication* 28 (2013) 917–928

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking

Yaqing Niu [a,d,*], Matthew Kyan [b], Lin Ma [c], Azeddine Beghdadi [a], Sridhar Krishnan [b]

[a] *L2TIlab, Institut Galilée, Université Paris 13, France*
[b] *Department of Electrical and Computer Engineering, Ryerson University, Canada*
[c] *Department of Electronic Engineering, The Chinese University of Hong Kong, China*
[d] *Information Engineering School, Communication University of China, China*

ARTICLE INFO

Available online 21 August 2012

Keywords:
Visual attention
Visual saliency
Just noticeable distortion (JND)
Image watermarking

ABSTRACT

Perceptual watermarking should take full advantage of the results from human visual system (HVS) studies. Just noticeable distortion (JND) gives us a way to model the HVS accurately. In this paper, another very important aspect affecting human perception, visual saliency, is introduced to modulate JND model. Based on the visual saliency's modulatory effect on JND model which incorporates visual attention's influence on visual sensitivity, the saliency modulated JND profile guided image watermarking scheme is proposed. The saliency modulated JND profile guided watermarking scheme, where the visual sensitivity 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. Experimental results confirm the improved performance of our saliency modulated JND profile guided watermarking scheme in terms of transparency and robustness. Our watermarking scheme is capable of shaping lower injected-watermark energy onto more sensitive regions and higher energy onto the less perceptually significant regions in the image, which yields better visual quality of the watermarked image. At the same time, the proposed saliency modulated JND profile guided image watermarking scheme is more robust compared to unmodulated JND profile guided image watermarking scheme.

1. Introduction

Digital image watermarking has emerged as a solution to the problem of copyright protection in the past decade. The strength of the watermarked signal is bounded by perceptual visibility. Thus, in order to maintain the image quality and at the same time increase the probability of the watermark detection, it is necessary to take the characteristic of human visual system (HVS) into consideration when engaging in watermarking research [1]. Visual sensitivity refers to the ability of human observers to detect distortion in visual field. Numerically, visual sensitivity can be regarded as the inverse of the just noticeable distortion (JND), which refers to the maximum distortion thresholds in pixels or subbands that the HVS does not perceive. Another aspect affecting human perception towards visual signal is visual attention which can enhance or reduce the actual visual sensitivity and consequently JND profile needs to be adjusted inside and outside the salient regions in images [2,3].

Two psychophysical concepts are referred in this paper: JND and visual saliency. The former tells us how much distortion we can tolerate and the latter expresses where our visual attention is the most attracted. Knowledge of distortion sensitivity is important because it determines

* Corresponding author. Postal address: 100 St. George Street, Room 623A, Toronto, Ontario, Canada M5S 3G3. Tel.: +1 416 978 1540; fax: +1 416 978 4811.
E-mail address: yaqing0930@gamil.com (Y. Niu).

0923-5965/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2012.07.009

scene-adaptive upper bounds on watermark insertion and visual saliency is important as well because it allows us to use sensitivity information wisely. Perceptually based image watermarking is implemented by applying models of the human visual system in order to determine the upper boundary for watermark embedding. The technique described in this paper assists image watermarking by producing a visual saliency modulated JND profile that can be used as a guide to optimize watermarking. Hence, the watermark should be embedded in such a way that below the visual saliency modulated JND profile and, then, robust and meanwhile hard to be perceived by the HVS. JND indicates the HVS' distortion threshold with fixation, and visual saliency gives the extent of visual attention. For an optimal perceptual image watermarking scheme, it is essential to consider visual saliency's modulatory effect on JND profile.

In this work, the saliency modulated JND profile guided image watermarking scheme is proposed to reflect the modulatory aftereffects of visual saliency on JND profile which can be used as the scene-adaptive upper bounds on watermark insertion. To obtain visual saliency modulated JND profile, we need the JND value of the image. From the previous work we have proposed an effective combined JND model [4,29]. We also need the factor of visual saliency, which tells us areas of importance in the scene. In our previous study [36], a very important aspect affecting human perception, visual saliency, is introduced to modulate JND profile. We modify the combined JND model [4,29] by the modulation of the latest efficient works on visual saliency detection [5]. Based on visual saliency's modulatory effect on JND profile [36], the saliency modulated JND profile guided image watermarking scheme is proposed in this study. Beyond our previous study [36] we construct a series of tests to observe the performance of visual saliency modulated JND profile guided image watermarking scheme in terms of watermark's visual quality, capacity and robustness. Experimental results confirm the improved performance of our saliency modulated JND profile guided watermarking scheme in terms of transparency and robustness. Our watermarking scheme is capable of shaping lower injected-watermark energy onto more sensitive regions and higher energy onto the less perceptually significant regions in the image, which yields better visual quality of the watermarked image. At the same time, the proposed saliency modulated JND profile guided image watermarking scheme is more robust compared to unmodulated JND profile guided image watermarking scheme [29]. Because the visual model combined with visual saliency's modulatory effect is fully used to determine image dependent upper bounds on watermark insertion, The saliency modulated JND profile guided watermarking scheme allows us to provide the maximum strength transparent watermark.

This paper is organized as follows. The next section reviews the related work on existing models for JND estimation, relevant exploration of visual saliency detection, current techniques combining visual saliency with JND profile, and JND based perceptual image watermarking. In Section 3, the details of the proposed saliency modulated JND profile are presented. And Section 4 gives the details of the proposed saliency modulated JND profile guided image watermarking scheme. The proposed scheme is tested against common attacks, such as additive Gaussian noise, valumetric scaling, geometric distortion, and JPEG compression in Section 5. A detailed comparison of the proposed saliency modulated JND profile guided watermarking scheme with unmodulated JND profile guided watermarking scheme is also presented in this β. Finally, the paper is concluded in Section 6.

2. Related research work

This section reviews the previous work related to the proposed visual saliency modulated JND profile. In Section 2.1, a review of the existing JND estimators is presented. Relevant researches on visual saliency detection are discussed in Section 2.2. Current techniques combining visual saliency with JND profile are introduced in Section 2.3. Section 2.4 introduces the current JND based perceptual image watermarking.

2.1. JND estimations

The existing computational JND models can be classified into two categories: pixel based and subband based. The pixel domain JND estimation is much computationally simpler; the subband domain JND estimator can be more precise.

JND estimation for still images has been relatively well developed. Several frequently used computational modules are discussed in [37]. An early perceptual threshold estimation in discrete cosine transform (DCT) domain was proposed by Ahumada [6], which gives the threshold for each DCT component by incorporating the spatial contrast sensitivity function (CSF). This scheme was improved by Watson [7] after the luminance adaptation effect had been added to the base threshold, and contrast masking [8] had been calculated as the elevation factor. In [9] 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 [14] incorporates new spatial CSF, luminance adaptation and contrast masking. Since motion is a specific feature of videos, temporal dimension needs to be taken into account for human perceptual visibility analysis. JND estimation for video sequences need to incorporate not only the spatial CSF, but the temporal CSF as well. A spatio-temporal CSF model was proposed by Kelly [10] from experiments on visibility thresholds under stabilized viewing conditions. Daly [11] extended Kelly's model to fit unconstrained natural viewing conditions with a consideration of eye movements. Based on Daly's model, Jia [12] estimated the JND thresholds for videos by combining other visual effects such as the luminance adaptation and contrast masking. An improved temporal modulation factor proposed by Zhenyu Wei [13,14] incorporates not only temporal CSF, but also the directionality of motion. From the previous work, we have proposed an effective combined JND model [4,29,31] for image

watermarking. We have also proposed a combined video-driven JND profile for video watermarking [15,16,30] which incorporates the temporal modulation factor, retinal velocity, luminance adaptation, and block classification.

2.2. Visual saliency detection

Human perception tends to firstly pick the regions in the imagery that stimulate the vision nerves the most before continuing to interpret the rest of the scene. These attended regions could correspond to either prominent objects in images or interesting actions in video sequences. Salient areas in natural scenes are generally regarded as the candidates of attention focus in human eyes. Visual attention analysis deals with detecting the regions of interest (ROI) in images and interesting actions in video sequences which are the most attractive to viewers. It simulates this human vision system behaviour by automatically producing saliency maps of the target image or video sequence.

Tresiman [17] proposed a theory which describes that visual attention has two stages. First, the parallel, fast, but simple *pre-attentive* process; and then, the serial, slow, but complex *attention* process. The first stage pre-attentively processes all incoming visual information equally and in a parallel fashion; the second stage filters and combines the extracted information. A set of low level visual features such as color, edges and motion is processed in parallel at pre-attentive stage. And then a limited-capacity process stage performs other more complex operations like face recognition, etc. [18]. The ability of human visual system to detect visual saliency is extremely fast and reliable. However, computational modeling of this basic intelligent behaviour still remains a challenge.

Several computational models have been proposed to simulate human's visual attention. Itti et al. [19] proposed a bottom-up model and built a system called Neuromorphic Vision C++ Toolkit (NVT). After that, following Rensink's theory [20], Walther extended this model, successfully applied it to object recognition tasks and created Saliency Tool Box (STB) [21]. However, the high computational cost and variable parameters are still the weaknesses of these models. Recently a simple and fast approach based on Fourier transform called spectral residual (SR) was proposed, which used SR of the amplitude spectrum to obtain the saliency map [5].

2.3. Combining visual saliency with JND

More computational resource of the human brain is allocated to high attentional areas than low attentional areas, and this is the reason of visual saliency's modulation on visual sensitivity in different areas [22]. The saliency map is designed to reflect the statistical allocation of the human brain's processing resource on local visual contents. Visual saliency modulates all levels of visual perception, including visual sensitivity. Visual saliency can enhance or reduce the actual visual sensitivity. Consequently JND profile needs to be adjusted inside and outside of the salient areas in images.

In [23] a computational model is proposed for incorporating a selectivity measure into the JND profile. In [22] perceptual quality significance map is proposed to reflect the modulatory after-effects of visual attention on visual sensitivity and visual quality evaluation.

2.4. Perceptual image watermarking

Previous perceptual image watermarking researches have only partially used the results of the HVS studies [1,24–27]. Many previous image watermarking algorithms utilize visual models to increase the robustness and transparency. The perceptual adjustment of the watermark is mainly based on Watson's spatial JND model [1,24,25,28]. An image-adaptive watermarking procedure based on Watson's spatial JND model was proposed in [25]. In [1], the DCT-based watermarking approach uses Watson's spatial JND model in which the threshold consists of spatial frequency sensitivity, luminance sensitivity and contrast masking. An energy modulated watermarking algorithm based on Watson's spatial JND model was proposed in [24]. During the modulation, Watson's perceptual model is used to restrict the modified magnitude of DCT coefficients. In [28], Watson's perceptual model and modified Watson's perceptual model were used to adaptively select the quantization step size for modifications to dither modulated Quantization Index Modulation (QIM). In the modified Watson model, the luminance masking is simply modified to scale linearly with valumetric scaling. The main drawback of utilizing Watson's visual models for images watermarking is that it does not satisfactorily provide the maximum strength transparent watermark. The obtained watermark is not optimal in terms of imperceptibility and robustness. We have proposed an effective combined JND model guided image watermarking scheme in DCT domain [4,29]. The watermarking scheme is based on the design of additional accurate JND visual model which is fully used to determine scene-adaptive upper bounds on watermark insertion.

3. Visual saliency modulated JND profile

Saliency modulated JND profile is an efficient measurement incorporating visual attention's influence on visual sensitivity of the human eye to represent the additional accurate perceptual redundancies for digital image. Here we compute the visibility threshold of each DCT coefficient with the saliency modulated JND profile which is illustrated by Fig. 1.

In the following, we proposed the effective saliency modulated JND profile in DCT domain. In Section 3.1, the JND estimation of image is presented. Saliency estimation is acquired in Section 3.2. Section 3.3 demonstrates the saliency modulated JND profile.

3.1. JND estimation

Just-noticeable distortion (JND) refers to the maximum distortion threshold that the HVS cannot perceive; therefore, it is an efficient model to represent the perceptual

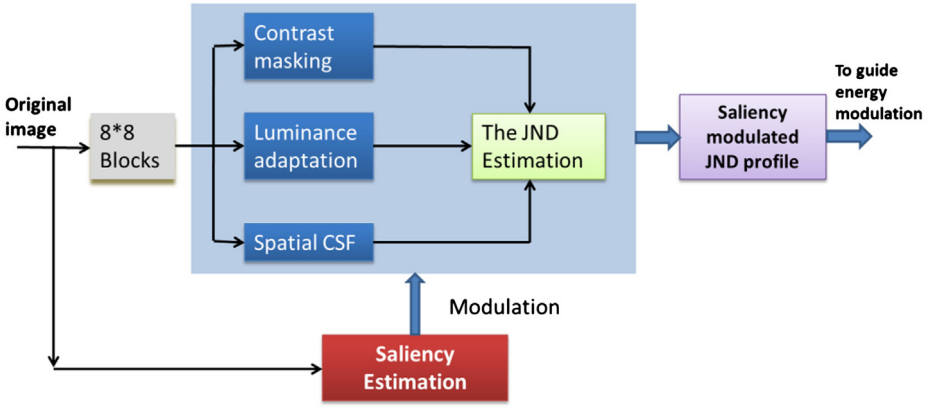

Fig. 1. Diagram of saliency modulated JND profile.

redundancies. Here we compute a combined JND estimation which incorporates Spatial CSF, luminance adaptation and contrast masking all together.

3.1.1. Spatial CSF

Human eyes show a band-pass property in the spatial frequency domain. The Spatial CSF model is the reciprocal of the base distortion threshold which can be tolerated for each DCT coefficient.

$$\phi_m = \begin{cases} \sqrt{1/N}, m=0 \\ \sqrt{2/N}, m>0 \end{cases} (m=i,j) \tag{1}$$

$$\omega_{ij} = \frac{1}{2N}\sqrt{\left(\frac{i}{\theta_x}\right)^2 + \left(\frac{j}{\theta_y}\right)^2} \tag{2}$$

$$\theta_h = 2\cdot\arctan\left(\frac{\Lambda_h}{2\cdot l}\right)(h=x,y) \tag{3}$$

$$\varphi_{ij} = \arcsin\left(\frac{2\omega_{i,0}\omega_{0,j}}{\omega^2_{ij}}\right) \tag{4}$$

$$T_{BASEs}(n,\omega_{ij},i,j) = s\cdot\frac{1}{\phi_i\phi_j}\cdot\frac{\exp(c\omega_{ij})/(a+b\omega_{ij}}{r+(1-r)\cdot\cos^2\varphi_{ij}} \tag{5}$$

Literature [14] shows that the base threshold for DCT domain T_{BASEs} corresponding to spatial CSF model can be expressed using Eq. (5). In Eq. (5), Φ_i and Φ_j are DCT normalization factors using Eq.(1), ω_{ij} is the spatial frequency using Eq. (2), and φ_{ij} stands for the directional angle of the corresponding DCT component using Eq. (4). T_{BASEs} is the base threshold corresponding to spatial CSF model and n is the index of a block in the image, i and j are the DCT coefficients' indices. $a=1.33$, $b=0.11$, $c=0.18$, $s=0.25$, $r=.0.6$. In Eqs. (1) and (2), N is the dimension of the DCT block. In Eq. (2), θ_x and θ_y are the horizontal and vertical visual angles of a pixel using Eq. (3). In Eq. (3), l is the viewing distance and Λ stands for the display width/length of a pixel on the monitor.

3.1.2. Luminance adaptation

The HVS is more sensitive to the noise in medium gray regions, so the visibility threshold is higher in very dark or very light regions. Because our base threshold is detected at the 128 intensity value, for other intensity values, a modification factor needs to be included. This effect is called the luminance adaptation effect

$$a_{Lum}(n) = \begin{cases} (60-I(n))/150+1 & I(n)\le 60 \\ 1 & 60<I(n)<170 \\ (I(n)-170)/425+1 & I(n)\ge 170 \end{cases} \tag{6}$$

The curve of the luminance adaptation factor is a U-shape which means the factor at the lower and higher intensity regions is larger than the middle intensity region. An empirical formula for the luminance adaptation factor a_{Lum} in [14] is shown using Eq. (6) where $I(n)$ is the average intensity value of the n_{th} block in the image.

3.1.3. Contrast masking

Contrast masking refers to the reduction in the visibility of one visual component in the presence of another one. The masking is strongest when both components are of the same spatial frequency, orientation, and location. To incorporate contrast masking effect, we employ contrast masking $a_{contrast}$ [8] measured using Eq. (7) where $C(n,i,j)$ is the (i,j)th DCT coefficient in the nth block in the image

$$a_{contrast}(n,i,j) = \max\left(1, \left(\frac{C(n,i,j)}{T_{BASE}(n,i,j)\cdot a_{Lum}(n)}\right)^{\varepsilon}\right) \tag{7}$$

To find a suitable value for ε, Zhang's perceptual model [9] proposed a much lower value 0.36, but Zhang's model tends to underestimate JND in edgy blocks. We exploit ε fixed to 0.7 as in Watson's perceptual model [8].

3.1.4. Complete JND estimation

$$T_{JND}(n,\omega_{ij},i,j) = T_{BASEs}(n,\omega_{ij},i,j)\cdot a_{Lum}(n)\cdot a_{contrast}(n,i,j) \tag{8}$$

The overall JND using Eq. (8) can be determined by the base threshold T_{BASEs}, the luminance adaptation factor a_{Lum} and the contrast masking factor $a_{Contrast}$. $T_{JND}(n,\omega_{ij},i,j)$ is the complete image-driven JND estimator which represents the additional accurate perceptual visibility threshold profile to guide watermarking for digital images.

3.2. Saliency estimation

Visual attention is one of the most important characteristics of human visual system. Salient areas in natural scenes are generally regarded as the candidates of attention focus in human eyes. Visual attention analysis simulates human vision system behavior by automatically producing saliency maps of the target image.

It was discovered that an image's spectral residual (SR) of the log amplitude spectrum represented its innovation. By using the exponential of SR instead of the original amplitude spectrum, the reconstruction of the image results in the saliency map [5]. The saliency estimation is carried out using this computational model

$$\mathcal{A}(f) = A(\mathfrak{F}[\mathcal{J}(x)]) \tag{9}$$

$$\mathcal{P}(f) = P(\mathfrak{F}[\mathcal{J}(x)]) \tag{10}$$

$$\mathcal{L}(f) = \log(\mathcal{A}(f)) \tag{11}$$

$$\mathcal{R}(f) = \mathcal{L}(f) - h_n(f) * \mathcal{L}(f) \tag{12}$$

$$\mathcal{S}(x) = g(x) * \mathfrak{F}^{-1}[\exp(\mathcal{R}(f) + \mathcal{P}(f))]^2 \tag{13}$$

In this computational model, the spectral residual $\mathcal{R}(f)$ contains the innovation of an image which can be obtained using Eq. (12). In Eq. (12), $\mathcal{L}(f)$ denotes the logarithm of amplitude spectrum $\mathcal{A}(f)$ using Eq. (11) and $h_n(f)$ is the average filter. Using inverse Fourier transform then squared, the saliency map in spatial domain is constructed. For better visual effects, we smoothed the saliency map $\mathcal{S}(x)$ with a Gaussian filter $g(x)$ using Eq. (13), where $\mathfrak{F}$ and $\mathfrak{F}^{-1}$ denote the Fourier transform and inverse Fourier transform, and $\mathcal{P}(f)$ denotes the phase spectrum of the image. In Eqs. (9) and (10), $\mathcal{A}$ denotes amplitude spectrum of the image $\mathcal{J}(x)$ and P denotes phase spectrum.

3.3. Saliency modulated JND profile

Visual saliency modulates all levels of visual perception, including visual sensitivity. Visual saliency can enhance or reduce the actual visual sensitivity. Consequently JND profile needs to be adjusted inside and outside of the salient areas in images.

The visual sensitivity is enhanced on salient spatial locations, due to the aftereffects of visual attention. The visibility thresholds in the salient areas are lower than the other nonattentional areas. With higher visual saliency value, the turning point of spatial contrast masking curve is pushed to higher frequencies, and the luminance adaptation tolerances are reduced. Itti et al. [32] reported that visual attention can elevate the sensitivity with spatial and temporal frequencies by 30%, the sensitivity with orientations by 40% and the sensitivity peak altitude by 5.2 dB. Itti's experimental results have been used in determining the actual maximum and minimum values of each modulation function

$$T^M_{JND}(n,\omega^M_{ij},i,j) = T_{JND}(n,\omega^M_{ij},i,j) * f^M_T(i,j) \tag{14}$$

$$\omega^M_{ij}(i,j) = \omega_{ij}(i,j) * f^M_{\omega_{ij}}(i,j) \tag{15}$$

$$f^M_T(i,j) = \begin{cases} \mathrm{M}^{\min}_{\mathrm{Lum}}, & s_{\mathrm{block}} \geq \mathrm{S}_{\max} \\ 1-(s_{\mathrm{block}}-0.1)*\phi_{\mathrm{Lum}}, & \mathrm{S}_{\min} < s_{\mathrm{block}} < \mathrm{S}_{\max} \\ \mathrm{M}^{\max}_{\mathrm{Lum}}, & s_{\mathrm{block}} \leq \mathrm{S}_{\min} \end{cases} \tag{16}$$

$$f^M_{\omega_{ij}}(i,j) = \begin{cases} \mathrm{M}^{\max}_{\omega}, & s_{\mathrm{block}} \geq \mathrm{S}_{\max} \\ 1+(s_{\mathrm{block}}-0.1)*\phi_{\omega}, & \mathrm{S}_{\min} < s_{\mathrm{block}} < \mathrm{S}_{\max} \\ \mathrm{M}^{\min}_{\omega}, & s_{\mathrm{block}} \leq \mathrm{S}_{\min} \end{cases} \tag{17}$$

The saliency modulated JND profile $T^M_{JND}(n,\omega^M_{ij},i,j)$ can be expressed using Eq. (14). In Eq. (14) $T_{JND}(n,\omega^M_{ij},i,j)$ is the JND estimator yielded by substituting ω_{ij} for ω^M_{ij} in Eq. (8), ω^M_{ij} is modulated ω_{ij} using Eq. (15), $f^M_T(i,j)$ and $f^M_{\omega_{ij}}(i,j)$ are the corresponding modulation functions using Eqs. (16) and (17). In Eqs. (16) and (17), s_{block} is the normalized saliency value of each 8×8 pixel size block which is calculated based on original saliency map, ϕ_{Lum} and ϕ_{ω} are factors to control the slope for each modulation which are set to 0.12 and 0.14, respectively. Based on our experiments, M^{max}_{Lum}, M^{min}_{Lum}, M^{max}_{ω} and M^{min}_{ω} are set to 1.0072, 0.9822, 1.0208 and 0.9916, respectively.

4. Visual saliency modulated JND profile guided image watermarking scheme

We exploit the saliency modulated JND profile T^M_{JND} guided watermarking scheme to embed and extract watermarking. The scheme first constructs a set of approximate energy sub-regions using the improved longest processing time (ILPT) algorithm [33], and then enforces an energy difference between every two sub-regions to embed watermarking bits [34,35] under the control of the saliency modulated JND profile T^M_{JND} proposed in Section 3.

The watermark bit string is embedded bit-by-bit in a set of regions (each region is composed of $2n$ 8×8 DCT blocks) of the original image. Each region is divided into two sub-regions (each sub-region is composed of n 8×8 DCT blocks). A single bit is embedded by modifying the energy of two sub-regions separately. However, for better imperceptibility, approximate energy sub-regions has to be constructed using ILPT, so that the original energy of each sub-region in one region are approximate. Each bit of the watermark bit string is embedded in its constructed bit-carrying-region. For instance, in Fig. 2 each bit is embedded in a region of $2n=16\times 8\times 8$ DCT blocks. The value of the bit is encoded by introducing an energy difference between the low frequency DCT coefficients of the top half of the region (denoted by sub-region A) containing in this case $n=8\times 8\times 8$ DCT blocks, and the bottom half (denoted by sub-region B) also containing $n=8\times 8\times 8$ DCT blocks. The number of watermark bits that can be embedded is determined by the number of blocks in a region.

4.1. The watermark embedding procedure

Diagram of saliency modulated JND profile guided watermark embedding is shown in Fig. 3. The embedding

procedure of the scheme is described as the following steps:

1. Decompose the original image into non-overlapping 8×8 blocks and compute the energy of the low-frequency DCT coefficients in the zigzag sequence.
2. Obtain approximate energy sub-regions by ILPT algorithm.
3. Map the index of the DCT blocks in a sub-region according to ILPT.
4. Use combined saliency modulated JND profile described in Section 3 to calculate the perceptual visibility threshold profile for DCT coefficients.
5. If the watermark to be embedded is 1, the energy of sub-region A should be increased (positive modulation) and the energy of sub-region B should be decreased (negative modulation). If the watermark to be embedded is 0, the energy of sub-region A should be decreased (negative modulation) and the energy of sub-region B should be increased (positive modulation). The energy of each sub-region is modified by adjusting the low-frequency DCT coefficients according to the saliency modulated JND profileusing Eq. (18), where $C(k,n,i,j)^m$ is the modified DCT coefficient, Sign(.) is the sign function, *PM* is positive modulation which means increase the energy and *NM* is negative modulation which means decrease the energy, $T^M_{JND}(n,i,f)$ is the perceptual visibility threshold by our saliency modulated JND profile and $f(.)$ can be expressed using Eq. (19)

$$C(n,i,j)^m = \begin{cases} C(n,i,j)+Sign(C(n,i,j))\cdot f(C(n,i,j),T^M_{JND}(n,i,j)), & PM \\ C(n,i,j)-Sign(C(n,i,j))\cdot f(C(n,i,j),T^M_{JND}(n,i,j)), & NM \end{cases} \tag{18}$$

$$f(C(n,i,j),T^M_{JND}(n,i,j)) = \begin{cases} 0, & if \quad C(n,i,j) < T^M_{JND}(n,i,j) \\ T^M_{JND}(n,i,j), & if \quad C(n,i,j) \geq T^M_{JND}(n,i,j) \end{cases} \tag{19}$$

6. Conduct IDCT to the energy modified result to obtain the watermark embedded image.

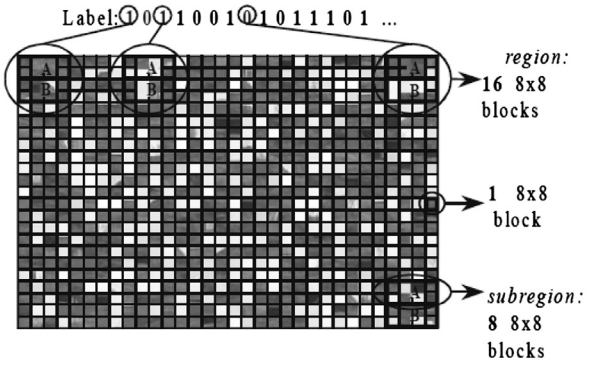

Fig. 2. Watermark bit corresponding to approximate energy sub-regions.

4.2. The watermark extraction procedure

Diagram of saliency modulated JND profile guided watermark extraction is shown in Fig. 4. The extraction procedure is described as follows:

1. Decompose the watermark embedded image into non-overlapping 8×8 blocks and compute the energy of the low-frequency DCT coefficients in the zigzag sequence.
2. Energy of each sub-region is calculated according to the index map.
3. Compare the energy of sub-region A with sub-region B. If the energy of sub-region A is greater than the energy of sub-region B, the watermark embedded is 1. If the energy of sub-region A is smaller than the energy of sub-region B, the watermark embedded is 0. The watermark is extracted accordingly.

5. Experimental results and performance analysis

The proposed algorithm is composed of two parts. The first part of the proposed algorithm is in Section 3 which derived the visual saliency modulated JND profile. And in Section 4 which is the second part of the proposed algorithm gives the details of how to utilize the visual

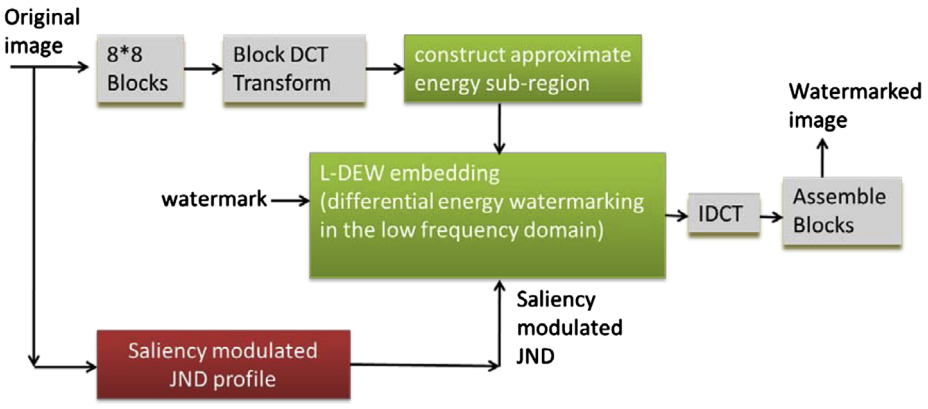

Fig. 3. Diagram of saliency modulated JND profile guided watermark embedding.

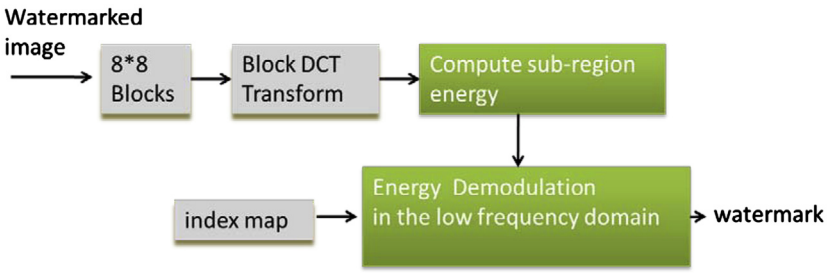

Fig. 4. Diagram of saliency modulated JND profile guided watermark extraction.

saliency modulated JND profile to guide image watermarking. To demonstrate the effectiveness of the proposed visual saliency modulated JND profile and its application in image watermarking, we performed experiments according to these two parts. For the first part, we performed experiments in Sections 5.1 and 5.2 to evaluate the performance of saliency estimation and saliency modulated JND profile, respectively. For the second part, we performed experiments in Section 5.3 to evaluate the performance of the proposed visual saliency modulated JND profile guided watermarking scheme focusing on the watermark's visual quality, capacity and robustness.

5.1. Evaluating saliency estimation

In this experiment, we follow the spectral residual method in Section 3.2 to get the saliency estimation with three main steps: (1) compute the saliency value at each pixel to form a saliency map. (2) Normalize these values in the saliency map. (3) Average values in each 8×8 block to attain the block based saliency estimation. Each pixel value of saliency map represents the weight for the pixel region. If a location has a larger saliency value in the saliency map (i.e., light area), it is more attention getting. We can average the saliency values of each 8×8 block and get a block based saliency estimation, which is used to modulate the JND profile. To evaluate the performance of saliency estimation, we also generate the saliency maps based on Itti's well-known theory [19] to compare with the spectral residual method.

Four images are chosen from salient object image database as test images for this experiment shown in Fig. 5 (kid, flower, couple and sandwich).

The saliency estimation results are shown in Fig. 5. From the saliency estimation result, we observe that the spectral residual method provides overall better performance than Itti's method. Computationally, the cost of the spectral residual method is relatively low – this brings considerable advantage for a saliency estimator, making it easier to implement on image watermarking. Compared with Itti's method, the computational consumption of the spectral residual method is parsimonious, providing a promising solution to modulate JND profile.

5.2. Evaluating saliency modulated JND profile

To evaluate the performance of the proposed visual saliency modulated JND profile against the original JND profile [4,29,31], noise is injected into images according to these two JND estimations. In this experiment, the generated JND profile can be used to guide noise shaping in image to evaluate the performance of different JND profiles. The original unmodulated JND profile was compared with the proposed visual saliency modulated JND estimator.

Kid image from salient object image database is used as a test image for this experiment. Noise is added to each DCT coefficients of the images using Eq. (20). Where f takes $+1$ or -1 randomly; $T(n,\omega_{if}\ i,\ j)$ represents the threshold determined by JND profile obtained via these two JND estimators, where $T(n,\ \omega_{if}\ i,\ j)$ is $T_{JND}(n,\ \omega_{ij},\ i,\ j)$ for the original unmodulated JND profile, or $T^{M}_{JND}(n,\ \omega^{M}_{ij},\ i,\ j)$ for the proposed visual saliency modulated JND profile; $C'(n,\ i,\ j)$ is the noise-injected DCT coefficient

$$C'(n,i,j) = C(n,i,j) + f \times T(n,\omega_{ij},i,j) \tag{20}$$

A more accurate JND profile is supposed to derive a noise injected image with better visual quality under the same level of noise (controlled by $T(n,\ \omega_{if},\ i,\ j)$), because it is capable of shaping more noise onto the less perceptually significant regions in the image. For a convincing evaluation of the proposed visual saliency modulated JND estimator for still images, we tested in two aspects. One is to measure how much the visual content variations are after these two JND estimation guided noise injection and the second is to assess the quality of each noise injected image. PSNR is used here just to denote the injected noise level under different test conditions. On the other hand, the subjective viewing was used to assess the quality of the resultant visual content. A better JND profile allows higher injected-noise energy (corresponding to lower PSNR) without jeopardizing picture quality (measured by the subjective viewing). With the same PSNR, the JND profile relating to a better subjective visual quality is better. Alternatively, with the same perceptual visual quality, the JND profile relating to a lower PSNR is better.

5.2.1. The difference of two JND estimations

To compare the original unmodulated JND profile and the visual saliency modulated JND profile, noise is injected into kid image according to these two JND estimations.

The original kid image and the difference map of these two generated JND estimation of kid image is shown in Fig. 6. Fig. 6 (a) shows the original kid image while Fig. 6(b)

Fig. 5. The saliency estimation result of the spectral residual method in comparison with Itti's method. (a) The original image, (b) saliency map generated by spectral residual and (c) saliency map generated by Itti's method.

shows the difference map between the unmodulated JND estimation and the visual saliency modulated JND estimation. We can see from Fig. 6(b) that at the more sensitive region (with high saliency value from Fig. 5) the visual saliency modulated JND estimation value is smaller compared to the unmodulated JND estimation value, while at the nonsensitive region (with low saliency value from Fig. 5) the visual saliency modulated JND estimation value is larger compared to the unmodulated JND estimation value. This result confirms that the visual saliency modulated JND estimation is capable of shaping more noise onto the less perceptually significant regions and less noise onto the more perceptually significant regions in the image.

5.2.2. *The PSNR*

Noise is injected into kid image according to these two JND estimations. It is observed that the noise hardly noticeable in two resultant images. The visual saliency modulated JND estimation yields the PSNR=27.59 dB; the original unmodulated JND estimation yields the PSNR=27.61 dB. The PSNR result reflects that although the visual saliency modulated JND estimation shaping more noise onto the less

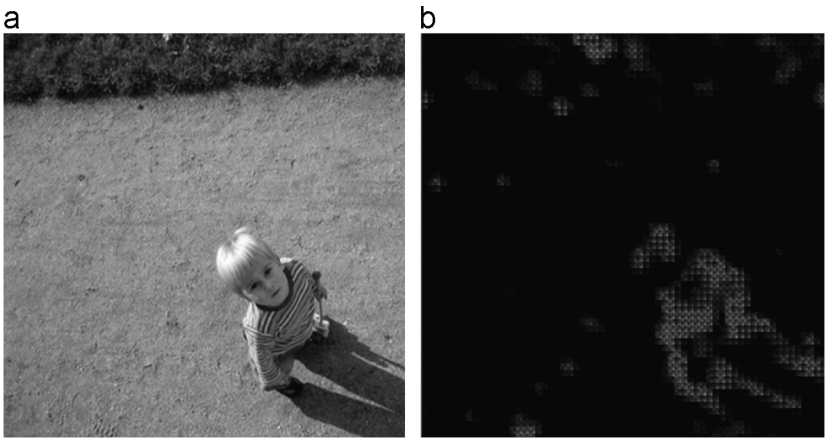

Fig. 6. (a) The original image and (b) the difference map between the unmodulated JND estimation and the visual saliency modulated JND estimation.

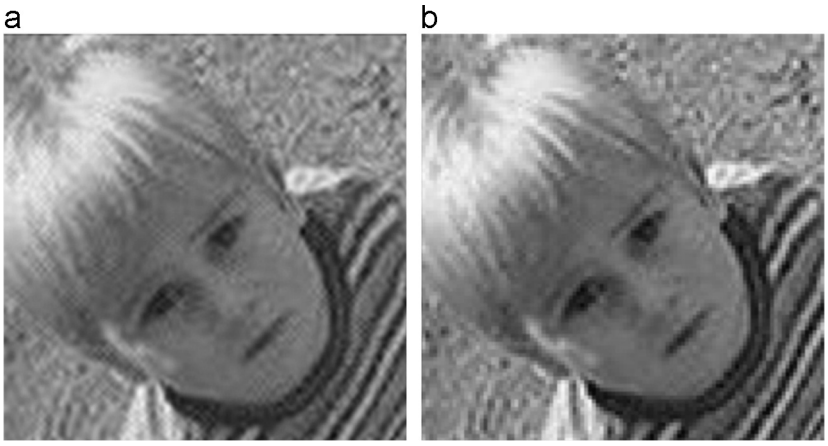

Fig. 7. Close-up view of the effect of noise injection over the most sensitive region. (a) The original unmodulated JND and (b) the visual saliency modulated JND.

sensitive regions and less noise onto the more sensitive regions, the overall injected-noise energy under the saliency modulated JND estimation is a little bit higher compared to the original unmodulated JND estimation.

5.2.3. *Perceptual visual quality*

Fig. 7 gives a close-up view in the most sensitive region (with highest saliency values) for better visualization, and, as can be seen, saliency modulated JND estimation yields better visual quality in the noise-injected images. The perceptual visual quality result reflects that the visual saliency modulated JND estimation shaping less noise onto the more sensitive regions.

From the experimental results in Sections 5.2.1–5.2.3, we can see saliency modulated JND estimation correlates with the HVS better with the evidence of being capable of yielding higher injected-noise energy (we can see lowest PSNR in Section 5.2.2) with better picture quality (in Fig. 7). And from the result in Section 5.2.1, the visual saliency modulated JND estimation is capable of shaping more noise onto the less perceptually significant regions and less noise onto the more perceptually significant regions yielding better perceptual quality of more sensitive regions in the image.

5.3. *Evaluating visual saliency modulated JND profile guided watermarking scheme*

We construct a series of tests to observe the performance of visual saliency modulated JND profile guided image watermarking scheme in terms of watermark's visual quality, capacity and robustness. The 512×512 kid image is used for experiments. The original watermark is a binary image of the logo of Communication University of China with size 32×64.

5.3.1. *Perceptual visual quality*

As discussed in Section 5.2.2, the visual saliency modulated JND profile correlates with the human visual system very well. We can use our profile to guide noise shaping in each DCT coefficients of the images yet the difference is hardly noticeable. To embed watermark with the guiding of our profile is similar to the process of noise shaping. As described in Section 4, not all the coefficients

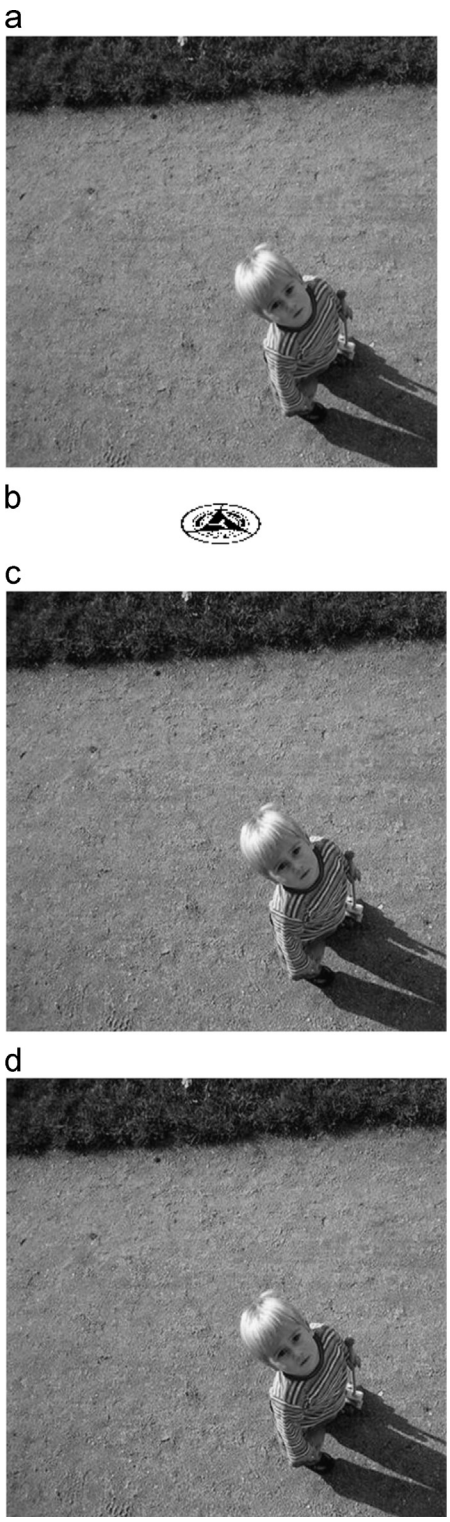

Fig. 8. (a) Original image, (b) watermark, (c) watermarked image by unmodulated JND estimation and (d) watermarked image by visual saliency modulated JND estimation.

are modified with the consideration of Lossy Compression so the Visual quality is even better than the test result in Section 5.2.2. Fig. 8(a), (c), and (d) show the kid image before and after watermark insertion guided by the original unmodulated JND estimation and the visual saliency modulated JND estimation. Fig. 8(b) shows the watermark. We can see no obvious degradation in Fig. 8(c) and (d) whose PSNR are 36.81 dB and 36.79 dB, respectively. That reflects the visual saliency modulated JND estimation allowing higher injected-watermark energy without introducing noticeable visual distortions. The visual saliency modulated JND estimation is capable of shaping more injected-watermark energy onto the less perceptually significant regions and less injected-watermark energy onto the more perceptually significant regions.

5.3.2. Estimation of capacity

The watermark bit string is embedded bit-by-bit in a set of regions of the original image. Each region is divided into two sub-regions. A single bit is embedded by modifying the energy of two sub-regions separately. For each 512×512 image the number of bits can be embedded is determined by the number of 8×8 DCT blocks in a region. We set the number of blocks at 2 in each region in the following experiments. That means each 512×512 image would embed 2048 bits. In Section 5.3 we focused on testing that for the same number of embedded bits (2048 bits) which watermarking scheme is more robust based on the relevant JND estimation while retaining the watermark transparency.

5.3.3. Estimation of robustness

In practice, watermarked content has to face a variety of distortions before reaching the detector. We present robustness results with different attacks like Gaussian noise, JPEG compression, valumetric scaling and geometric distortion. Robustness results of algorithm based on unmodulated JND estimation was compared with results of algorithm based on the visual saliency JND estimation shown in Figs. 9–11 and Table 1. For each category of distortion, the watermarked images were modified with a varying magnitude of distortion and the

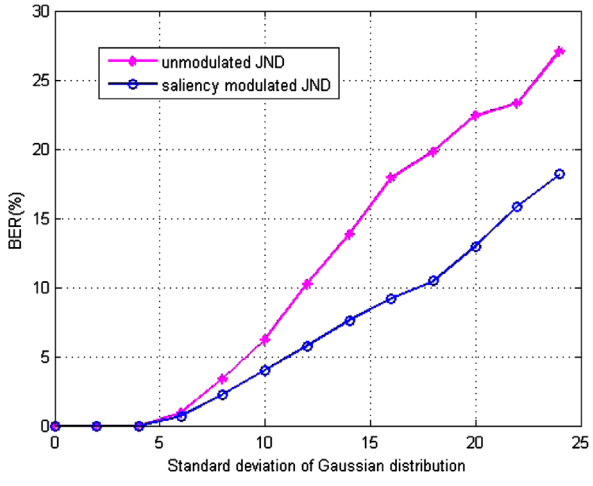

Fig. 9. Robustness versus Gaussian noise.

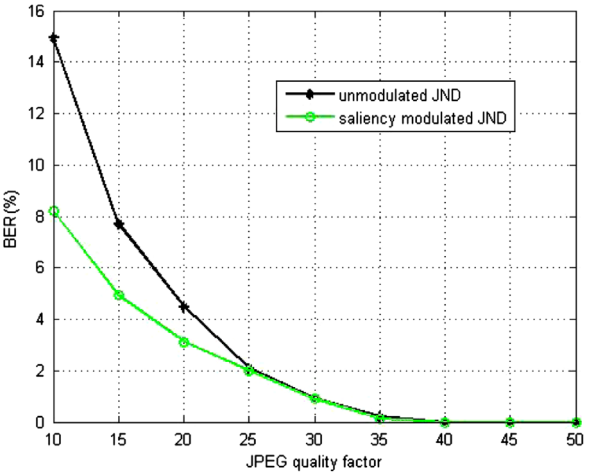

Fig. 10. Robustness versus JPEG compression.

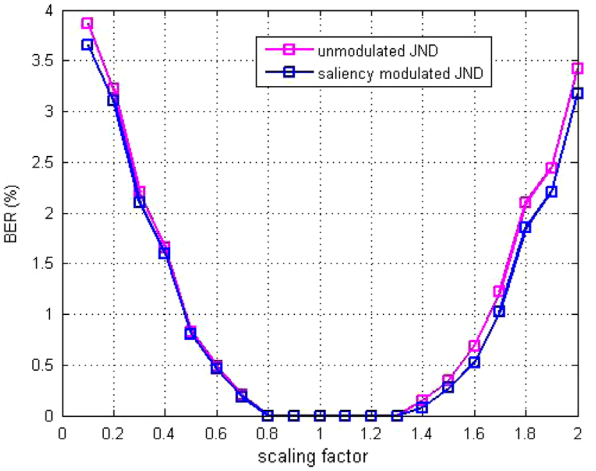

Fig. 11. Robustness versus valumetric scaling.

Table 1
Bit error rate after some geometric distortions.

Attack	Bit error rate (by unmodulated JND estimation) (%)	Bit error rate (by saliency modulated JND estimation) (%)
1_Row_1_Col_removed	1.90	1.85
1_Row_5_Col_removed	1.71	1.67
5_Row_1_Col_removed	1.71	1.67
5_Row_17_Col_removed	2.10	2.04
17_Row_5_Col_removed	1.81	1.77

bit error rate (BER) of the extracted watermark was then computed.

The normal distributed noise with mean 0 and standard deviation σ is added to the watermarked images, where σ varies from 0 to 25. The experimental results are presented in Fig. 9. From the robustness results with Gaussian noise in Fig. 9, the watermarking scheme based on the visual saliency modulated JND estimation performs better than algorithms based on unmodulated JND estimation. Fig. 10 demonstrates that the visual saliency modulated JND estimation based watermarking scheme's performance against Gaussian noise has an average bit error rate value 4.5% lower than algorithm based on unmodulated JND estimation. Following the practice, we considered the watermarking scheme to be robust if at least 80% of the watermarks were correctly retrieved, i.e. the BER is below 20%. If we establish a threshold at a bit error rate of 20%, then the algorithms based on unmodulated JND estimations robust against Gaussian noise with standard deviation, σ, of only 18. In contrast, the watermarking scheme based on the visual saliency modulated JND estimation is robust to standard deviations up to 24.

The JPEG compression is implemented to the watermarked images, where JPEG quality factor varies from 10 to 50. The experimental results are presented in Fig. 10. From the robustness results with JPEG compression in Fig. 10, the watermarking scheme based on the visual saliency modulated JND estimation performs better than algorithms based on unmodulated JND estimation. Fig. 11 demonstrates that the visual saliency modulated JND estimation based watermarking scheme's performance against JPEG compression has an average BER value 1.2% lower than algorithm based on unmodulated JND estimation. We can see that if the JPEG quality factor is decreased to 10, there will be about 14.9% bit error rate introduced by unmodulated JND estimation; while only 8.1% is based on the visual saliency modulated JND estimation. As the JPEG quality factor decreases, the gap between the bit errors of unmodulated JND estimation compared to visual saliency modulated JND estimation increases.

From the robustness results with valumetric scaling in Fig. 11 and geometric distortion in Table 1, the watermarking scheme based on saliency modulated JND estimation performs slightly better than algorithm based on unmodulated JND estimation. The experiment shown in Fig. 11 reduced the image intensities as scaling factor varied from 1 to 0.1, and increased the image intensities as scaling factor varied from 1 to 2. Fig. 11 demonstrates that the saliency modulated JND Estimation guided watermarking scheme's performance against valumetric scaling has an average bit error rate value 0.1% lower than algorithm based on unmodulated JND estimation. Table 1 demonstrates that the saliency modulated JND estimation guided watermarking scheme performances slightly better than algorithm based on unmodulated JND estimation in all cases where some geometric distortions were applied.

We have presented the results of our investigation on the performance of the saliency modulated JND estimation guided watermarking scheme in terms of watermark visual quality, capacity and robustness. The improvement evident above is due to the saliency modulated JND estimation for still images, which correlates with the human visual system better than unmodulated JND estimation, shaping less noise onto more sensitive regions and more noise onto the less perceptually significant regions in the image, at the same time obtain better robustness in image watermarking while retaining the watermark transparency.

6. Conclusions

In this work, the saliency modulated JND profile guided image watermarking scheme is proposed to reflect the modulatory aftereffects of visual saliency on JND profile which can be used as the scene-adaptive upper bounds on watermark insertion. The saliency modulated JND profile guided watermarking scheme allows us to provide the maximum strength transparent watermark. Experimental results confirm that our saliency modulated JND profile guided watermarking scheme provides overall better performance than the unmodulated JND profile guided image watermarking scheme. The proposed watermarking scheme is capable of shaping lower injected-watermark energy onto more sensitive regions and higher energy onto the less perceptually significant regions in the image, meanwhile, the proposed saliency modulated JND profile guided image watermarking scheme is more robust than unmodulated JND profile guided image watermarking scheme [29].

References

[1] R.B. Wolfgang, C.I. Podilchuk, E.J. Delp, Perceptual watermarks for digital images and video, in: Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information, vol. 87, 1999, pp. 1108–1126.
[2] Harold L. Hawkins, Steven A. Hillyard, Steven J. Luck, Mustapha Mouloua, Cathryn J. Downing, Donald P. Woodward, Visual attention modulates signal detectability, Journal of Experimental Psychology: Human Perception and Performance 16 (November (4)) (1990) 802–811.
[3] M. Carrasco, C. Penpeci-Talgar, M. Eckstein, Spatial covert attention increases contrast sensitivity across the csf: support for signal enhancement, Vision Research 40 (10–12) (2000) 1203–1215.
[4] Y.Q. Niu, J.B. Liu, S. Krishnan, Q. Zhang, Combined just noticeable difference model guided image watermarking, in: IEEE International Conference on Multimedia and Expo 2010 Workshops, July 2010.
[5] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, Proceedings of the CVPR (2007).
[6] A.J. Ahumada Jr, H.A. Peterson, Luminance-model-based DCT quantization for color image compression, Proceedings of the SPIE 1666 (1992) 365–374.
[7] A.B. Watson, DC Tune: a technique for visual optimization of DCT quantization matrices for individual images, in: Society for Information Display Digest of Technical Papers XXIV, 1993, pp. 946–949.
[8] G.E. Legge, A power law for contrast discrimination, Vision Research 21 (1981) 457–467.
[9] X.K. Zhang, W.S. Lin, P. Xue, Improved estimation for just-noticeable visual distortion, Signal Processing 85 (4) (2005) 795–808.
[10] D.H. Kelly, Motion and vision. II. Stabilized spatio-temporal threshold surface, Journal of the Optical Society of America 69 (1979) 1340–1349.
[11] S. Daly, Engineering observations from spatio velocity and spatiotemporal visual models, Proceedings of the SPIE 3299 (1998) 180–219.
[12] Y. Jia, W. Lin, A.A. Kassim, Estimating just noticeable distortion for video, IEEE Transactions on Circuits and Systems for Video Technology 16 (7) (2006) 820–829.
[13] Z. Wei, K.N. Ngan, A temporal just-noticeable distortion profile for video in DCT domain, in: Proceedings of the 15th IEEE International Conference on Image Processing, 2008, pp. 1336–1339.
[14] Z. Wei, K.N. Ngan, Spatial just noticeable distortion profile for image in DCT domain, in: Proceedings of the IEEE International Conference on Multimedia and Expo, 2008, pp. 925–928.
[15] Y.Q. Niu, Y. Zhang, S. Krishnan, Q. Zhang, A video-driven just noticeable distortion profile for watermarking, in: Proceedings of the 2009 International Conference on Engineering Management and Service Sciences, 2009.
[16] Y.Q. Niu, J.B. Liu, S. Krishnan,Q. Zhang, Spatio-temporal just noticeable distortion model guided video watermarking, in: Proceedings of the 2009 IEEE Pacific-Rim Conference on Multimedia, 2009.
[17] A. Treisman, G. Gelade, A feature-integration theory of attention, Cognitive Psychology 12 (1) (1980) 97–136.
[18] J. Wolfe, Guided search 2.0: A revised model of guided search, Psychonomic Bulletin & Review 1 (2) (1994) 202–238.
[19] L. Itti, C. Koch, E. Niebur, et al., A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11) (1998) 1254–1259.
[20] R. Rensink, Seeing, sensing, and scrutinizing, Vision Research 40 (10–12) (2000) 1469–1487.
[21] D. Walther, C. Koch, Modeling attention to salient proto-objects, Neural Networks 19 (2006) 1395–1407.
[22] Z. Lu, W. Lin, X. Yang, E. Ong, S. Yao, Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation, IEEE Transactions on Image Processing 14 (11) (2005) 1928–1942.
[23] Z. Lu, X. Lin, X. Yang, E. Ong, S. Yao, Spatial selectivity modulated just-noticeable-distortion profile for video, in: IEEE Transactions on Image Processing, Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.
[24] H.F. Ling, Z.D. Lu, F.H. Zou, R.X. Li, An energy modulated watermarking algorithm based on Watson perceptual model, Journal of Software 17 (5) (2006) 1124–1132.
[25] C.I. Podilchuk, W. Zeng, Image-adaptive watermarking using visual models, Proceedings of the IEEE 16 (1998) 525–539.
[26] J. Huang, Y.Q. Shi, Adaptive image watermarking scheme based on visual masking, Electronics Letters 34 (1998) 748–750.
[27] M.S. Kankanhalli, K.R. Ramakrishnan, Content based watermarking of images, Proceedings of the Sixth International ACM Conference on Multimedia, 1998, pp. 61–70.
[28] L. Qiao, I.J. Cox, Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation watermarking, IEEE Transactions on Information Forensics and Security 2 (2) (2007).
[29] Y.Q. Niu, M. Kyan, S. Krishnan, Q. Zhang, A combined just noticeable distortion model guided image watermarking, Signal, Image and Video Processing, Journal no. 11760, ISSN: 1863–1703.
[30] Y.Q. Niu, S. Krishnan, Q. Zhang, Spatio-temporal just noticeable distortion model guided video watermarking, in: Special Issue on Intelligent Multimedia Security and Forensics, International Journal of Digital Crime and Forensics (IJDCF), ISSN: 1941-6210 1703.
[31] Y.Q. Niu, M. Kyan, S. Krishnan, Q. Zhang, A combined just noticeable distortion model guided image watermarking, in: Proceedings of International Conference on Digital Content, Multimedia Technology and its Applications, 2010.
[32] L. Itti, J. Braun, C. Koch, Modeling the modulatory effect of attention on human spatial vision, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds), Advances in Neural Information Processing Systems, vol. 14, MIT Press, Cambridge, MA, 2002.
[33] F.H. Zou, Research of Robust Video Watermarking Algorithms and Related Techniques, A Dissertation Submitted for the Degree of Doctor of Philosophy in Engineering, Huazhong University of Science & Technology, 2006.
[34] G.C. Langelaar, R.L. Lagendijk, Optimal differential energy watermarking of DCT encoded images and video, IEEE Transactions on Image Processing 10 (1) (2001) 148–158.
[35] H.F. Ling, Z.D. Lu, F.H. Zou, Improved differential energy watermarking (IDEW) algorithm for DCT-encoded imaged and video, in: Proceedings of the Seventh International Conference on Signal Processing, 2004, pp. 2326–2329.
[36] Yaqing Niu, Matthew Kyan, Lin Ma, Azeddine Beghdadi, Sridhar Krishnan, A visual saliency modulated just noticeable distortion profile for image watermarking, in: The 2011 European Signal Processing Conference (EUSIPCO2011), Barcelona, August 2011.
[37] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: A survey, Journal of Visual Communication and Image Representation 22 (4) (2011) 297–312.