This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

Energy-efficient Images

Hadi Hadizadeh

***Abstract*—In this paper, a novel method is presented for producing energy-efficient images, i.e., images that consume less electrical energy on energy-adaptive displays, yet have the same, or very similar perceptual quality to their original images. The proposed method relies on the fact the the energy consumption of pixels in modern energy-adaptive displays like OLED displays is directly proportional to the luminance of the pixels. Hence, in this paper to reduce the energy consumption of an image while at the same time preserving its perceptual quality, it is proposed to reduce the luminance of the pixels in the image by one just-noticeable-difference (JND) threshold. To determine the JND thresholds, an adaptive saliency-modulated JND (SJND) model is developed. In the proposed model, the JND thresholds of each block in the given image are elevated by two non-linear saliency modulation functions using the visual saliency of the block. The parameters of the saliency modulation functions are estimated through an adaptive optimization framework, which utilizes a state-of-the-art saliency-based objective image quality assessment (IQA) method. To evaluate the proposed methods, a set of subjective experiments were conducted, and the real energy consumption of the produced energy-efficient images were measured by an accurate power monitor equipment on an OLED display. The obtained experimental results demonstrated that, on average, the proposed method is able to reduce the energy consumption by about 14.1% while preserving the perceptual quality of the displayed images.**

***Index Terms*—power reduction, energy-adaptive displays, just noticeable distortion, visual saliency, OLED**

I. Introduction

With the advancement of power-hungary IT technologies and the rapid proliferation of mobile devices, the demands for energy-efficient and sustainable electronic devices are tremendously increasing due to the limited batteries capacity as well as the raising concerns about global warming and climate changes. For this reason, the area of "green computing" has been an active research field in the recent years [1], [2].

Displays are known as the main energy consumer in electronic devices such as computers and mobile devices, often consuming more than half the laptop or handheld system's total electrical energy [1], [3], [4], [5], [6]. Conventional thin film transistor liquid crystal displays (TFT LCDs) consume a considerable amount of energy due to using a uniform backlight system, which is always 'ON' even when displaying a completely dark image. On the other hand, emerging emissive display technologies such as organic light-emitting diode (OLED) displays and dual-layer high dynamic range (HDR) displays (e.g., Dolby's Pro-monitors with backlight modulation), consume energy in a more controllable and efficient manner [3], [7]. Such displays use an array of LEDs whose brightness can be controlled individually based on the content of the displayed images. In these displays, the energy consumption of a pixel is related to its luminance [5], [6], and no backlight is required. Hence, the power consumption of such displays is directly related to the displayed images because pixels with different colors and luminances consume different amount of power on these displays. This feature enables the design of energy-adaptive displays, and motivates researchers and engineers to develop energy-aware methods for power reduction in the emerging displays [1], [5], [6].

In this paper, a novel method is presented for producing energy-efficient images, i.e., images that their perceptual quality is the same as (or similar to) the perceptual quality of their original images while at the same time they consume less electrical energy (power) than their original versions when displayed on energy-adaptive displays like OLED displays. The main idea behind the proposed method is that there exists perceptual redundancies in images, which cause some details in an image not to be perceived by the human visual system (HVS) [8], [9]. For instance, the luminance of a pixel in an image can be altered without being noticed by the viewer. On the other hand, the energy consumption of the energy-adaptive displays such as OLED displays is directly proportional to the luminance of the pixels in the displayed images [6]. Hence, by exploiting such perceptual redundancies, it is possible to reduce the energy consumption of the displayed images by altering the luminance of their pixels in such a way that they are perceived as indistinguishable from their original images.

The perceptual redundancies in images are mainly due to the psychophysical properies of the HVS, and the visual attention (VA) mechanism of the human brain [8], [10], [11]. The first concept is related to the well-known fact that the HVS cannot sense fine-scale variations of visual signals below the so-called *just-noticeable-distortion (JND) thresholds* due to several physiological and psychological mechanisms in the eyes and the brain [8], [10]. In other words, JND corresponds to the minimum visibility threshold below which no changes can be perceived by the HVS. JND can also be defined as the reciprocal of the so-called *contast sensitivity function* (CSF), which expresses the ability of the HVS to discern between contrast differences at various spatial or temporal frequencies [10]. There exists many different methods for JND estimation in digital images in both spatial and frequency domains [9], [12], [13]. On the other hand, VA provides a mechanism for selection of particular aspects of a visual scene that are most relevant to our ongoing behavior while eliminating interference from irrelevant visual data in the background so as to reduce the computational load on the brain [11].

In the literature, several computational models of human visual attention have been developed [14], [15], [16], [17]. According to the current knowledge, it is believed that VA is driven by "visual saliency", which is a measure of propensity for drawing our attention to a particular aspect or region in a

Hadi Hadizadeh is with the Quchan University of Advanced Technology, Quchan, Khorasan Razavi, Iran. e-mail: (h.hadizadeh@qiet.ac.ir).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

visual scene. In a visual scene, a region is said to be "visually salient" if it possesses certain characteristics, which make it stand out from its surrounding regions and draw our attention to it. The existing computational models of VA for still images are able to produce a *visual saliency map* by which salient regions and also gaze locations in the image can be predicted automatically [14].

It is known that visual saliency affects the visual sensitivity or JND thresholds. For instance, it is known that JND thresholds in attended regions are smaller than the JND thresholds in un-attended regions [18], [19], [20]. This means that, visual saliency modulates JND thresholds. Hence, to estimate JND thresholds more accurately, it is reasonable to consider the modulatory effects of visual saliency on JND thresholds.

In the first part of this paper, an adaptive saliency-modulated JND (SJND) model in the discrete cosine transform (DCT) domain is developed, which is able to generate a JND map for a given image based on the visual saliency map of the image. The developed SJND model is then used to produce energy-efficient images. The proposed SJND model is built upon an existing spatial JND model in the block-based DCT domain. To consider the modulatory effects of visual saliency on JND thresholds, the JND thresholds of each block in the given image are elevated based on two saturating non-linear modulation functions whose input is the visual saliency of the block. The parameters of these non-linear functions are estimated adaptively from the given image via the following optimization procedure. Given an image and its saliency map, a random noise is first generated such that its amplitude is determined by the proposed SJND model with some initial values for its parameters. The generated noise is then added to the image. After that, the parameters of the employed saliency modulation functions are tuned in an optimization loop such that the perceptual quality of the noise-injected image becomes equal or very close to the perceptual quality of the original image while its peak signal-to-noise ratio (PSNR) value with respect to the original image is reduced as much as possible. Note that under the same perceptual quality, the lower the PSNR, the more accurate is the JND model because it indicates that the JND model is able to shape more noise onto the less perceptually-significant regions in the image. PSNR is used here to measure the injected-noise level under different test conditions. To measure the perceptual quality of an image, any arbitrary image quality assessment metric can be used here. In this paper, the *visual saliency-induced index* (VSI) [21] is used, which can be considered as the state-of-the-art method for measuring the perceptual image quality using the visual saliency information.

Note that although JND estimation for static images has been relatively well-developed [12], [13], only few JND models exist, which consider the interplay between JND and visual saliency [18], [22]. Actually, the most recent method in this regard is the one proposed in [22], which in this paper, it is referred to as the "Niu's method". The proposed SJND model in this paper differs from the Niu's method in two aspects. First, in the Niu's method, two linear functions are used to modulate the JND thresholds based on the saliency of a block whereas in the proposed method, two saturating non-linear functions are utilized for this purpose. Second, in the Niu's method, the parameters of the linear functions are fixed and their values were found experimentally while in the method proposed here, a systematic and automatic framework is proposed for calculating the parameters of the employed functions for any given image in an content-adaptive manner. The motivation for using non-linear modulation functions is twofolds. First, there are many saturating non-linear functions in biophysics [23], [8]. Hence, it is postulated that the modulatory effects of visual saliency on JND thresholds may be modeled by non-linear functions better than linear functions. Second, a non-linear function has a more flexibility in describing data as it has more free parameters to tune.

After developing the SJND model, it is utilized to generate the energy-efficient images as follows. Given the SJND threshold of each pixel in a given image, the luminance of each pixel in the image is reduced by its SJND threshold so as to reduce its energy consumption. By doing so for all pixels in the image, an energy-efficient image is produced whose perceptual quality is the same as (or similar to) the perceptual quality of the original image.

To evaluate the proposed methods in this paper, a set of subjective experiments were conducted. Also, the real power consumption of the produced images on an OLED display was measured by an accurate power monitor equipment. The obtained results indicated that the proposed method for producing energy-efficient images is able to reduce the energy consumption by about 14.1% on average, while maintaining the same perceptual quality.

The paper is organized as follows. In Section II, some background information are briefly reviewed first. The proposed methods are then presented in Section III. The experimental results are given in Section IV followed by conclusions in Section V.

II. Background

A. Energy consumption in energy-adaptive displays

The consumed energy in energy-adaptive displays such as OLED displays is proportional to the number of 'ON' pixels, and the brightness of their red (R), green (G), and blue (B) components, or simply their luminance [7], [3], [5], [6]. Therefore, different colors consume different amounts of electrical energy in these displays. For instance, in [5], the actual power consumption of a single OLED pixel was measured as a function of the intensity level in the linear (non-gamma-corrected) RGB format as an indication of luminance. The results are shown in Fig. 1. As seen from this figure, the relation between the power consumption and luminance is linear. Hence, the lower the luminance of a pixel, the lower its energy consumption and vice versa.

Note that various hardware techniques, such as ambient-based backlight modulation combined with histogram analysis, and LCD compensation with backlight reduction, can also be used to achieve energy savings [24], [5], [6] but such methods usually do not preserve the perceptual quality of the displayed images. However, in this paper the focus is on the pixel-level energy consumption, and a content-based method is presented

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

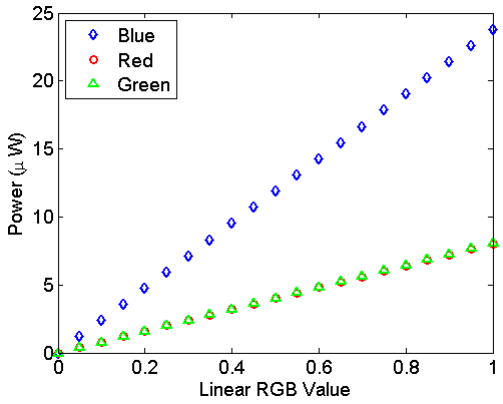

Fig. 1. Intensity level vs. real power consumption for the linear RGB components of an OLED pixel [5].

to reduce the energy consumption of the displayed images while preserving their perceptual quality.

B. Psychophysical properties of the HVS

Over the last decades, psychophysical research has been one of the major lines of research for understanding the properties and functionalities of the HVS. Psychophysical research infers the HVS properties based on experiments using human subjects [10]. For example, the HVS can be characterized by finding *contrast detection thresholds* for detection of various spatial patterns such as sine-wave gratings. The contrast detection threshold for a given target is defined as the minimum contrast necessary for a human observer to visually detect the target. Many researchers have measured the contrast detection thresholds for various spatial frequencies [12], [13]. It was discovered that contrast sensitivity varies with the spatial frequency of a grating, and such a relationship is usually described by a bandpass *contrast sensitivity function* (CSF) with peak sensitivity near 4-6 cycles/degree [8].

It is also known that the HVS sensitivities are changed due to *visual masking*. Visual masking refers to the perceptual phenomenon in which the presence of a masking signal reduces a subject's ability to detect another signal. Two important visual masking effects are *luminance adaptation* and *contrast masking* [13]. Luminance adaptation accounts for the masking from local background luminance, while contrast masking accounts for the visibility reduction of one contrast pattern at the presence of another contrast pattern [8], [10].

III. The Proposed Methods

In this section, the proposed methods are presented. In Section III-A, the proposed adaptive SJND model is first presented. The proposed energy-efficient method is then described in Section III-C. A flowchart of the proposed methods is depicted in Fig. 2.

A. The proposed adaptive SJND model

The proposed adaptive SJND model is built upon the spatial JND model proposed in [25] in the blocked DCT domain,

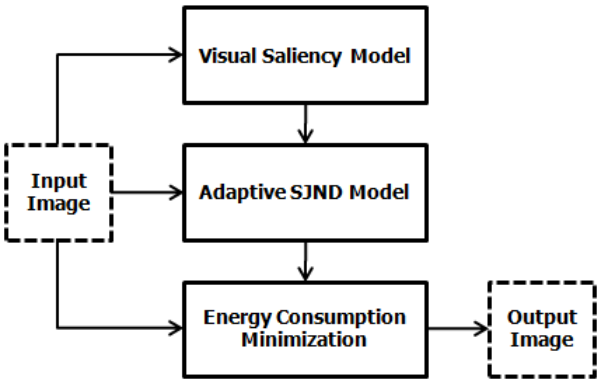

Fig. 2. The flowchart of the proposed method.

and adds two non-linear modulation functions to it. The modulation functions are utilized to elevate the spatial JND thresholds of a block in the image based on the visual saliency of the block. In the sequel, the spatial JND model in [25] is introduced in Section III-A1. The proposed adaptive SJND model is then described in Section III-A2.

1) A description of the employed spatial JND model: The employed spatial JND model [25] is implemented in the blocked DCT domain, and considers the effect of the following three important properties of the HVS: CSF, luminance adaptation, and contrast masking.

To take the effect of spatial CSF into consideration, a base JND threshold is estimated in the spatial JND model as follows [25]. The luminance channel of a given input image is first partitioned into $N \times N$ blocks, and then the DCT transform of all blocks are computed. In this paper, the value of N is set to 8. After that, the base JND threshold for each $N \times N$ block is computed using the following formula [25]:

$$T_b(n,i,j,\omega_{ij}) = \frac{e\exp(c\omega_{ij})}{(\phi_i\phi_j)(a+b\omega_{ij})(d+(1-d)cos^2\psi_{ij})}, \quad (1)$$

where i,j are the horizontal and vertical indices of the DCT coefficient at location (i,j) for $i,j = 0,\cdots,N-1$, and $a = 1.33$, $b = 0.11$, $c = 0.18$, $d = 0.6$, $e = 0.25$. Also, ϕ_i and ϕ_j are defined as:

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

and the spatial frequency of the (i,j)-th DCT coefficient is defined as $\omega_{ij} = \frac{1}{2N}\sqrt{\left(\frac{i}{\theta_x}\right)^2 + \left(\frac{j}{\theta_y}\right)^2}$, where θ_x and θ_y are the horizontal and vertical visual angles of a pixel, and they can be calculated using the following equation based on the viewing distance l and the actual width and height of a pixel on the monitor, P_x and P_y:

$$\theta_t = 2\arctan\left(\frac{P_t}{2l}\right), (t = x,y). \quad (3)$$

Also, ψ_{ij} in (1) denotes the directional angle of the (i,j)-th DCT coefficient, and it is defined as $\psi_{ij} = \arcsin\left(\frac{2\omega_{i,0}\omega_{0,j}}{\omega_{ij}^2}\right)$. The term $1/(d+(1-d)cos^2\psi_{ij})$ in (1) accounts for the so-called *oblique* effect [8], which models the fact that the human

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

visual sensitivity in the horizontal and vertical directions (when i or j equals to 0) is higher than other directions (e.g. when $i = j$).

The base JND threshold in (1) is obtained at the intensity value of 128, i.e. on a medium gray background. But it is known that the visibility threshold in very dark and very bright regions is lower than in medium gray regions [13], [25]. This effect is known as the *luminance adaptation* effect. Thus, to consider this effect, the base JND threshold is scaled for higher and lower intensity values using a modification factor T_l, which is usually estimated by the the following experimental U-shape function [25]:

$$T_l(n) = \begin{cases} (60 - \bar{I}(n))/150 + 1 & \bar{I}(n) \leq 60 \\ 1 & 60 < \bar{I}(n) < 170, \\ (\bar{I}(n) - 170)/425 + 1 & \bar{I}(n) \geq 170 \end{cases} \quad (4)$$

where $\bar{I}(n)$ denotes the average intensity of the n-th block in the image.

The last component in the employed spatial JND model takes the effect of contrast masking into consideration. Contrast masking refers to the reduction in the visibility of one visual signal (contrast pattern) in the presence of another one. This masking is strongest when both visual signals are of the same spatial frequency, orientation, and location. For this purpose, the contrast masking level is estimated as follows, and it is used to scale the base JND threshold [26], [22]:

$$T_c(n, i, j, \omega_{ij}) = \max\Big(1, \Big(\frac{C(n, i, j)}{T_b(n, i, j, \omega_{ij}) T_l(n)}\Big)^{\epsilon}\Big), \quad (5)$$

where $C(n, i, j)$ is the (i, j)-th DCT coefficient of the n-th block, and ϵ is an exponent, which lies between 0 and 1. Similar to [26], the value of ϵ was set to 0.7.

Finally, based on (1), (4), and (5), the spatial JND model is constructed by scaling the base JND threshold by the luminance adaptation and contrast masking factors as follows:

$$T_{JND}(n, i, j, \omega_{ij}) = T_b(n, i, j, \omega_{ij}) T_l(n) T_c(n, i, j, \omega_{ij}). \quad (6)$$

The reason for using a DCT-based JND model is twofolds. First, most image and video compression standards like JPEG, H.264, and H.265 use blocked DCT [27], and so hardware implementation of the proposed method becomes easier and faster when using DCT. Second, in the DCT domain, CSF can easily be incorporated into the JND profile, which allows a better modeling of the HVS. In fact, following the seminal work by Ahumada and Peterson [12], the DCT-based JND estimation has been very attractive for many researchers, and many different DCT-based JND models like the above-mentioned model have been developed [13], [9], [28].

2) Developing the adaptive SJND model: As mentioned earlier, due to the visual attention mechanism, the visibility thresholds in salient regions are lower than that of the visibility thresholds in non-salient regions. It is also known that with higher visual saliency, the turning point of spatial contrast masking curve is pushed to higher frequencies [19], [18]. To take these issues into consideration, given the visual saliency map of an image, a saliency-modulated JND model is proposed as follows:

$$T_{SJND}(n, s_n, \boldsymbol{\alpha}, \boldsymbol{\beta}) = T_{JND}\Big(n, i, j, \hat{\omega}_{ij}(\boldsymbol{\alpha})\Big) h(s_n, \boldsymbol{\beta}), \quad (7)$$

where s_n denotes the normalized saliency of the n-th block, $T_{JND}\Big(n, i, j, \hat{\omega}_{ij}(\boldsymbol{\alpha})\Big)$ is the JND model defined in (6), and $\hat{\omega}_{ij}(\boldsymbol{\alpha})$ is a frequency modulation function defined as:

$$\hat{\omega}_{ij}(\boldsymbol{\alpha}) = \omega_{ij}\ g(s_n, \boldsymbol{\alpha}), \quad (8)$$

where $g(s_n, \boldsymbol{\alpha})$ is a sigmoid function defined by four coefficients in vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ as follows:

$$g(s_n, \boldsymbol{\alpha}) = \alpha_1\Big(\frac{1}{2} - \frac{1}{1 + \exp(\alpha_2(s_n - \alpha_3))}\Big) + \alpha_4, \quad (9)$$

Also, $h(s_n, \boldsymbol{\beta})$ is a sigmoid function defined by four coefficients in vector $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]$ as follows:

$$h(s_n, \boldsymbol{\beta}) = \beta_1\Big(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(s_n - \beta_3))}\Big) + \beta_4. \quad (10)$$

The functions g and h are used to modulate the JND thresholds based on the normalized visual saliency of a block. Note that the spatial JND model defined in (6) is also used in the Niu's method [22], and the format of (7) is similar to the SJND model proposed in the Niu's method. However, the modulation functions in the Niu's method are linear and their parameters are fixed, and their values were found experimentally, but the proposed modulation functions here are non-linear, and as will be discussed in the next section, their parameters are estimated adaptively within a systematic framework based on the image content and its saliency map. As will be shown in Section IV, the proposed adaptive SJND model outperforms the non-adaptive SJND model proposed in the Niu's method.

3) Estimating the parameters of the adaptive SJND model: In this section, a method is presented to estimate the parameters of the adaptive SJND model defined in (7) for a given input color image $\mathbf{I}$ whose luminance channel is denoted by $\mathbf{R}$. First, $\mathbf{R}$ is partitioned into $N \times N$ non-overlapping blocks. Let $\mathbf{S}$ be the visual saliency map of $\mathbf{I}$, $\mathbf{R}(n)$ be the n-th block in $\mathbf{R}$, and $\mathbf{R}(n, i, j)$ be the (i, j)-th element in $\mathbf{R}(n)$.

Suppose that the parameters of the two modulation functions $g(s_n, \boldsymbol{\alpha})$ and $h(s_n, \boldsymbol{\beta})$ (i.e. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) are set to some initial values. For simplicity, let define $\boldsymbol{\theta} = (\boldsymbol{\alpha},\ \boldsymbol{\beta})$. A $N \times N$ DCT is applied to every block in $\mathbf{R}$, and the SJND thresholds of each block are computed based on $\boldsymbol{\theta}$. Let $\mathbf{T}_{SJND}(n, s_n, \boldsymbol{\theta})$ be a $N \times N$ matrix whose value at location (i, j) is equal to the SJND threshold at the (i, j)-th DCT coefficient of $\mathbf{R}(n)$.

If the parameters of the proposed SJND model are estimated properly, then increasing or decreasing the value of the DCT coefficients of each block by their corresponding SJND thresholds does not change the perceptual quality of the block. To measure the accuracy of the JND models, usually a random bipolar noise is added to the image, where the amplitude of the noise is determined by the JND model, and it is either positive or negative. Obviously, if the viewers can distinguish the resultant noisy image from the original pristine image, then the performance of the JND model is not good and vice versa. Here, to estimate the parameters of the proposed adaptive SJND model, a similar methodology is used. Specifically, a random noise is added to the DCT coefficients of $\mathbf{R}(n)$ such that the amplitude of the noise at each DCT index is equal to the SJND threshold at that DCT index. The inverse DCT of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

the resultant block is then computed to obtain $\mathbf{D}(n,\boldsymbol{\theta})$ as the noisy version of $\mathbf{R}(n)$ as follows:

$$\mathbf{D}(n,\boldsymbol{\theta}) = \text{DCT}^{-1}\Big(DCT\big(\mathbf{R}(n)\big) + \eta\ \mathbf{f} \odot \mathbf{T}_{SJND}(n, s_n, \boldsymbol{\theta})\Big), \tag{11}$$

where $\mathbf{f}$ is a random $N \times N$ matrix whose elements are either +1 or -1, $\odot$ denotes the pixel-wise multiplication, and η is a control parameter to adjust the noise energy as measured by mean squared error (MSE) or PSNR. This parameter provides a way for comparing various JND models. Under the same noise energy, the model that provides a better perceptual quality is a better model. The reason for using $\mathbf{f}$ is to avoid creating a fixed artificial spatial pattern. By computing (11) for all blocks in $\mathbf{R}$, a noisy luminance image $\mathbf{D}(\boldsymbol{\theta})$ is produced, and by adding the chromatic channels of $\mathbf{I}$ to $\mathbf{D}(\boldsymbol{\theta})$, the noisy version of $\mathbf{I}$ is obtained as $\hat{\mathbf{I}}(\boldsymbol{\theta})$.

Note that if the SJND thresholds are estimated properly, then $\hat{\mathbf{I}}(\boldsymbol{\theta})$ should not perceptually be distinguishable from $\mathbf{I}$. This can be achieved if $\boldsymbol{\theta}$ is estimated such that $\mathbf{D}(\boldsymbol{\theta})$ becomes perceptually indistinguishable from $\mathbf{R}$. For this purpose, the following cost function is defined, and $\boldsymbol{\theta}$ is found such that this cost function is minimized:

$$Q(\mathbf{D}(\boldsymbol{\theta})|\mathbf{I},\mathbf{S}) = \text{PSNR}\big(\mathbf{D}(\boldsymbol{\theta})|\mathbf{R}\big) - \lambda\ VQ\big(\hat{\mathbf{I}}(\boldsymbol{\theta})|\mathbf{I},\mathbf{S}\big), \tag{12}$$

and

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\text{argmin}}\ Q(\mathbf{D}(\boldsymbol{\theta})|\mathbf{I},\mathbf{S}), \tag{13}$$

where $\text{PSNR}\big(\mathbf{D}(\boldsymbol{\theta})|\mathbf{R}\big)$ is the PSNR of $\mathbf{D}(\boldsymbol{\theta})$ with respect to $\mathbf{R}$, and $VQ\big(\hat{\mathbf{I}}(\boldsymbol{\theta})|\mathbf{I},\mathbf{S}\big)$ is the visual quality score of $\hat{\mathbf{I}}(\boldsymbol{\theta})$ with respect to $\mathbf{I}$ as estimated by an arbitrary full-reference objective image quality assessment (FR-IQA) method that takes the saliency information $\mathbf{S}$ into account. λ in (12) is a regularization term. Note that the larger the SJND thresholds, the lower the PSNR of $\mathbf{D}(\boldsymbol{\theta})$. Also, the higher the value of $VQ\big(\hat{\mathbf{I}}(\boldsymbol{\theta})|\mathbf{I},\mathbf{S}\big)$, the better the perceptual quality of $\hat{\mathbf{I}}(\boldsymbol{\theta})$. Hence, the minimization of the defined cost function allows us to find the maximum SJND thresholds by which the perceptual quality of the resultant noisy images is remained close to the perceptual quality of the corresponding original images. Note that the HVS is more sensitive to luminance changes than to chroma changes [8], and this is the reason for using only the luminance channel in the proposed framework.

For the perceptual quality metric, without loss of generality, the recently-proposed *visual saliency-based index* (VSI) [21] is employed because of two reasons. First, this FR-IQA method considers the visual saliency information for perceptual image quality prediction. Second, the experimental results reported in [21] indicated that this method achieves a better image quality prediction accuracy as compared to several state-of-the-art objective IQA methods while maintaining a moderate computational complexity.

B. Pixel-wise SJND estimation

Using (7), the SJND thresholds of a given block in the DCT domain can be estimated. However, for the energy consumption reduction method presented in this paper, it is needed to estimate the SJND thresholds in the pixel domain. For this purpose, it is proposed to use the conversion method proposed in [29], [30] to obtain the SJND thresholds in the pixel domain from the SJND thresholds in the DCT domain as follows. Let $\mathbf{U}(n)$ be the 2-D DCT of $\mathbf{R}(n)$, and $t_{SJND}(n,i,j) = T_{SJND}(n,i,i,\boldsymbol{\theta}^*)$ be the SJND threshold of the (i,j)-th DCT coefficient in the n-th block. First, $t'_{SJND}(n,i,j)$ is defined as follows [29]:

$$t'_{SJND}(n,i,j) = \begin{cases} \zeta(n,i,j)t_{SJND}(n,i,j) & |\mathbf{U}(n,i,j)| \geq t_{SJND}(n,i,j) \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

where $\zeta(n,i,j) = sign\Big(\mathbf{U}(n,i,j)\Big)$. By computing $t'_{SJND}(n,i,j)$ for $i,j = 0,\cdots N-1$, a $N \times N$ matrix $\mathbf{T}'_{SJND}(n)$ is obtained. By taking the inverse DCT (IDCT) of $\mathbf{T}'_{SJND}(n)$, a $N \times N$ matrix $\mathbf{T}^p_{SJND}(n)$ is obtained whose elements are the pixel-level SJND thresholds for the n-th block as follows:

$$\mathbf{T}^p_{SJND}(n) = IDCT\Big(\mathbf{T}'_{SJND}(n)\Big). \tag{15}$$

C. The proposed method for energy consumption reduction

In Section II-A, it was mentioned that the power consumption of pixels in OLED displays is directly proportional to their luminance value. Hence, by reducing the luminance of a color, the energy consumption of the color is reduced. In particular, if we reduce the luminance of a color by one JND threshold, the resultant color can be still perceptually indistinguishable from the initial color. Based on these facts, a method for reducing the energy consumption of a color image when displayed on an energy-adaptive display is proposed in this section.

Consider a color image $\mathbf{I}$ of size $W \times H$ pixels. Let $\mathbf{r} = (x,y)$ denote the pixel location within $\mathbf{I}$, and $\mathbf{C}(\mathbf{r}) = \big(Y(\mathbf{r}), Cb(\mathbf{r}), Cr(\mathbf{r})\big)$ be the color of the pixel at location $\mathbf{r}$ in the YCbCr perceptual color space [31], where the luminance channel is denoted by Y, and the chroma channels are denoted by Cb and Cr. Let $SJND_Y(\mathbf{r})$ be the SJND threshold of the pixel at location $\mathbf{r}$ as estimated in the pixel domain by the proposed adaptive SJND model from the luminance (Y) channel of $\mathbf{I}$. Given $SJND_Y(\mathbf{r})$, a new color $\mathbf{C}^-(\mathbf{r})$ is generated from $\mathbf{C}(\mathbf{r})$ by subtracting $SJND_Y(\mathbf{r})$ from the luminance component of $\mathbf{C}(\mathbf{r})$ (for all $\mathbf{r} \in \mathbf{I}$) as follows:

$$\mathbf{C}^-(\mathbf{r}) = \Big(Y(\mathbf{r}) - SJND_Y(\mathbf{r}), Cb(\mathbf{r}), Cr(\mathbf{r})\Big). \tag{16}$$

The new color can be considered perceptually indistinguishable from $\mathbf{C}(\mathbf{r})$, since its chroma components are the same as those of $\mathbf{C}(\mathbf{r})$, and the difference between its luminance component and the luminance component of $\mathbf{C}(\mathbf{r})$ does not exceed the SJND threshold. However, since the luminance value of the new color is lower than the luminance value of the original color, the energy consumption of it is also lower than that of the original color. By repeating this process for all pixels in the given input image, a new energy-efficient image is produced whose perceptual quality is the same as (or close to) the perceptual quality of the original image while at the same time it consumes less power when displayed on an adaptive-energy display like an OLED display.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

IV. Experimental Results

Two evaluate the proposed methods in this paper, a set of different experiments were conducted. First, a number of experiments were performed to evaluate the proposed adaptive SJND model. The proposed method for producing the energy-efficient images was then evaluated. The details of each experiment are described in the following sections.

A. Testing the proposed adaptive SJND model

In order to implement and evaluate the proposed adaptive SJND model, the MIT eye-tracking dataset [32] was utilized. This dataset contains 1003 RGB images of resolution 1024×768 pixels with their associated ground truth eye-tracking data (fixation points) of 15 subjects. This dataset contains various images from landscape images to portrait images. Also, in this dataset, there exists a fixation map for each image, where the fixation maps were produced based on the fixation points in the original images. These fixation maps can provide the true visual saliency of different regions in the image because they were obtained by the real eye-tracking data. Hence, for the experiments described in this section, these fixation maps were used so as to remove the need for using a separate method for visual saliency estimation. However, in practice, any arbitrary visual saliency method can be used for this purpose. For example, in the next section, the SDSP method proposed in [17] is used for visual saliency computation when testing the proposed energy-efficient images.

The cost function in (12) was minimized using a multidimensional downhill simplex [33], [34] to obtain $\boldsymbol{\theta}^*$ for each given image. Also, the value of η was set to 1. Note that the typical PSNR values for images with moderate to good quality are usually between 35 and 40. On the other hand, the quality scores predicted by VSI are between zero and one. Therefore, it was found experimentally that setting the value of λ in (12) to 30 works good to treat both terms in (12) almost equally.

In the literature, one of the most popular approaches for comparing two JND models on a given image is as follows [29], [18]. First, a random bipolar noise is added to the image, where the amplitude of the noise is deterimend by the JND thresholds estimated by one of the JND models. This produces a noisy (distorted) image. After that, a similar noise is added to the original image whose amplitude is determined by the JND thresholds estimated by the other JND model multiplied by a scaling parameter. The scaling parameter is set such that the PSNR of the second noisy image becomes equal to the PSNR of the first noisy image. The noisy images are then subjectively compared with each other. Under the same level of noise energy (or PSNR), the noisy image with better perceptual quality determines the better JND model. Equivalently, at the same perceptual quality, the model that produces a noisy image with lower PSNR is the better model as it can tolerate more distortion at a specific perceptual quality level.

To evaluate the accuracy of the proposed adaptive SJND model as compared to the Niu's method [22] using the above-mentioned comparison methodology, a subjective experiment was performed. In the subjective experiment, a Two Alternative Forced Choice (2AFC) method [35] was used to compare subjective image quality. In 2AFC, the participant is asked to make a choice between two alternatives, in our case the noisy image produced by the proposed adaptive SJND model and the noisy image produced by the non-adaptive method proposed by Niu et al. This way of comparing image quality is less susceptible to measurement noise [36] than quality ratings based on scale, such as Mean Opinion Score (MOS) and Double Stimulus Continuous Quality Scale (DSCQS) [37], because in this approach test subjects directly compare qualities of two images or videos without having to map them to numerical values, which allows them to focus on quality assessment [7]. In fact, this approach has long been used in the psychophysics community to measure detection thresholds for various psychophysical attributes with minimial cognitive load for the subjects. Also, it has been employed in many image and video quality assessment methods [7], [38].

To select images for performing the subjective experiment, the cost function values of all images in the dataset were first sorted. Ten images with the lowest cost function value were then picked as Group A, and also 10 images with the highest cost function value were chosen as Group B. A random noise was then injected to these 20 images using both the proposed adaptive SJND method and the Niu's method. Similar to η in the proposed method, there is a control parameter in the Niu's method by which we can control the energy of the injected noise. The value of this control parameter was manually set so that the PSNR of images produced by the Niu's method becomes equal or very close to the PSNR of the corresponding images produced by the proposed method within 1% difference. Also, for a fair comparison, the actual fixation maps provided in the dataset were used for providing visual saliency information for both methods. At the same PSNR, the method which collects more votes from the participants will be the winner.

In each trial, participants were looking at two side by side images (in the same vertical position, and separated by 1 cm horizontally) on a mid-gray background as shown in Fig. 3. Each image pair was shown for 5 seconds. After this presentation, a mid-gray blank screen was shown for 5 seconds. During this period, participants were asked to indicate on an answer sheet, which of the two images looks better (Left or Right). They were asked to answer either Left or Right for each image pair, regardless of how certain they were of their response. Participants did not know which image was obtained by the proposed method and which was produced by the Niu's method. Randomly chosen half of the trials had the image produced by the proposed method presented on the left side of the screen and the other half on the right side, in order to counteract side bias in the responses. This gave a total of 40 trials (duplicated to balance left and right presentation).

The experiment was run in a quiet classroom with 24 participants (14 males and 10 females of age between 19 and 22). All participants had normal or corrected-to-normal vision. A 17-inch Dell monitor P170S with maximum brightness 250 cd/m^2, and resolution 1024×768 pixels was used. The brightness and contrast of the monitor were set to 50%. The size of all images were reduced by a factor of 1.16 using a bicubic interpolation algorithm to fit the screen. The actual

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

Fig. 3. In each 2AFC trial, participants were looking at two side by side images on a mid-gray background.

Fig. 4. A visual example showing the accuracy of the proposed method as compared to the Niu's method. Top row: the original image (left) and its saliency map (right). Bottom row: the image produced by the Niu's method (left) with (PSNR = 39.13 dB, VSI = 0.9878) and the image produced by the proposed method (right) with (PSNR = 37.43 dB, VSI = 0.9875). Note that the noisy images are not very distinguishable from the original image as the amplitude of the added noise has been determined by the JND thresholds.

size of the displayed images on the screen were 17.5×12.3 centimeters. The illumination in the room was in the range 106-132 Lux. The distance between the monitors and the subjects was fixed at 50 cm. Each participant was familiarized with the task before the start of the expeirment via a short printed instruction sheet. The total length of the experiment for each participant was approximately 7 minutes.

The results are shown in Table I, where the number of responses that showed preference for the Niu's method (n_1) and the proposed adaptive SJND method (n_2) were indicated in the second and third column, respectively. Note that $n_1 + n_2 = M$, where M is the total number of the subjects. Similar to [38], a two-sided Pearson's chi-square (χ^2) test [39] was used to examine the statistical significance of the results. The null hypothesis is that there is no preference for either the proposed method or the Niu's method. Under this hypothesis, the expected number of votes is $e_1 = e_2 = 24$ for both cases. The probability that the null hypothesis holds (the so-called p-value) is also indicated in the table. Note that the Pearson's chi-square test statistic is computed as follows:

$$\chi^2 = \sum_{t=1}^{2} \frac{(n_t - e_t)^2}{e_t}, \tag{17}$$

where using it a p-value can be found from χ^2 distribution tables or graphs [39]. In experimental sciences, as a rule of thumb, the null hypothesis is rejected when $p < 0.05$. For this to happen, χ^2 needs to be sufficiently large, i.e., the observed votes n_1 or n_2 need to deviate sufficiently from their expected values e_1 or e_2 under the null hypothesis. When this happens in Table I, it means that the two methods cannot be considered to produce the same subjective quality, since one of them has obtained a statistically significantly higher number of votes, and therefore seems to provide better image quality.

As seen in Table I, in only 5 out of 24 trials the p-value is greater than 0.05 - these are indicated in bold typeface. In all other cases, subjects showed a statistically significant preference for the images produced by the proposed adaptive SJND model. Looking across all trials (i.e., summing up all the votes for the two options), the results show that participants have preferred the images produced by the proposed adaptive SJND model more than the images produced by the Niu's method (283 vs. 677 votes) with overall $p = 0.0001$, which is a very statistically significant result, because the odds of it occurring by chance are 1 in 10000. Also, it is observed that the higher preference to the proposed method is more vivid in Group A as compared to Group B. This demonstrates that the lower the value of the proposed cost function, the better.

To evaluate the distortion-hiding capacity of the proposed adaptive SJND method, again a random bipolar noise was added to the 20 images described above using both the proposed adaptive SJND model and the Niu's method. The noise energy was manually controlled such that the VSI score of the images produced by the proposed method becomes equal or very close to the VSI score of the corresponding images produced by the Niu's method. Under the same perceptual quality (as measured by VSI), the PSNR of the distorted images produced by the two methods were compared with each other. It was observed that, on average, the PSNR of the images produced by the proposed method is about 1.7 dB lower than the PSNR of the images produced by the Niu's method. This states that at the same perceptual quality, the proposed model is able to tolerate (hide) more distortion than the Niu's method. Figure 4 shows two of such noise-contaminated images. In this figure, the image on the left was produced by the Niu's method, and the other one was produced by the proposed method. It can be seen that the perceptual quality of the two images is almost the same, however, the PSNR of the image produced by the proposed method is about 1.7 dB lower than the PSNR of the image produced by the Niu's method. This shows that the proposed method has a better noise-hiding capacity due to its better estimation of JND thresholds, which is very desirable for our purpose. Please also note that the perceptual quality of the images produced by both methods are very close to the perceptual quality of the original image, which is expectable because the amplitude of the injected noise in these images does not exceed the JND thresholds estimated by these methods. So, the injected noise in these examples is not very vivid. For a better visual comparison of the two methods, a much stronger noise (distortion) was intentionally injected to the original image shown in Fig. 5 using each of the two methods. The PSNR of the distorted images produced by both methods were matched to 18.8 dB. As seen from these visual results, the distorted image produced by the proposed adaptive SJND model has a better subjective quality than the distorted image produced by the Niu's method, especially in the salient regions as desired.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

TABLE I
COMPARING THE PROPOSED SJND MODEL WITH THE NIU'S 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	Niu's	Proposed	p-value
A1	10	38	0.0001
A2	14	34	0.0039
A3	16	32	0.0209
A4	13	35	0.0015
A5	19	29	**0.1489**
A6	8	40	0.0001
A7	9	39	0.0001
A8	12	36	0.0005
A9	10	38	0.0001
A10	15	33	0.0094
B1	25	23	**0.7728**
B2	17	31	0.0433
B3	18	30	**0.0833**
B4	8	40	0.0001
B5	13	35	0.0015
B6	16	32	0.0209
B7	19	29	**0.1489**
B8	8	40	0.0001
B9	14	34	0.0039
B10	19	29	**0.1489**
Total	283	677	0.0001

TABLE II
COMPARING THE PERCEPTUAL QUALITY OF THE IMAGES PRODUCED BY THE PROPOSED SJND MODEL WITH THEIR CORRESPONDING ORIGINAL IMAGES 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	Original	Proposed	p-value
A1	25	23	**0.7728**
A2	26	22	**0.6312**
A3	23	25	**0.7728**
A4	27	21	**0.1103**
A5	25	23	**0.7728**
A6	28	20	**0.1015**
A7	26	22	**0.6312**
A8	27	21	**0.1103**
A9	25	23	**0.7728**
A10	26	22	**0.6312**
B1	26	22	**0.6312**
B2	30	18	**0.0833**
B3	33	15	0.0094
B4	29	19	**0.1489**
B5	26	22	**0.6312**
B6	30	18	**0.0833**
B7	25	23	**0.7728**
B8	25	23	**0.7728**
B9	35	13	0.0015
B10	32	16	0.0209
Total	549	411	**0.0689**

To further demonstrate the accuracy of the proposed adaptive SJND model, another similar subjective experiment was performed. This time the subjective quality of the noise-injected images produced by the proposed adaptive SJND model was compared with their corresponding original images. The results are reported in Table II. As seen from the results in this table, in most cases the noisy images produced by the proposed method are not distinguishable from their original images, which states that the proposed adaptive SJND model is able to shape noise in an imperceptible manner as desired.

Fig. 5. A visual example comparing the subjective quality of the distorted images produced by the proposed adaptive SJND model and the Niu's method at a very low PSNR of 18.8 dB. From left to right and top to bottom: the original image, the fixation map, the distorted image produced by the Niu's method, the distorted image produced by the proposed adaptive SJND model.

B. Testing the proposed method for producing energy-efficient images

To test the accuracy of the proposed method for producing energy-efficient images, 50 images were randomly chosen from the MIT database, and the SDSP method proposed in [17] was employed to obtain the saliency maps of these images. The proposed method was then applied on these images to generate their energy-efficient images. The motivation for choosing the SDSP method was the promising results provided in [21] for this method. In fact, in [21], it was reported that the SDSP method outperforms eight prominent state-of-the-art methods for visual saliency computation including the popular IKN model [14]. Once again note that, without loss of generality, any another arbitrary visual saliency computation method can also be used here.

After that, another subjective experiment was performed to compare the quality of the produced energy-efficient images with their corresponding original images. The setup of the experiment, the procedure, and the participants were the same as in the first subjective experiment described in the previous section but this time, a Samsung Galaxy TabPro S 12" was used for displaying the images. This tablet uses a 12-inch OLED display, and so it was a good choice for measuring the power savings provided by the proposed method.

To accurately measure the real power consumption of each image during displaying it on the tablet, the battery of the tablet was removed, and a Monsoon power monitor [40] was connected to supply power to the tablet directly. For this purpose, during the power consumption evaluation period, each image was displayed in the full-screen mode of the tablet for 5 seconds at the highest brightness level of the display, and all unnecessary system services that may have caused any significant power consumption noise, including all radio and network communications (WiFi, 4G, Bluetooth, NFC), background services, and power optimization applications were disabled. The employed power monitor device supports a sampling frequency up to 5 KHz, and it is able to provide

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2690523, IEEE Transactions on Image Processing

real time power consumption measurements.

The obtained results for these experiments are shown graphically in Fig. 6 for a simpler and faster comparison. In the left image in this figure, the number of votes for the original images and the proposed energy-efficient images (collected from the 24 participants) are shown using a stacked horizontal bar plot. The Pearson's chi-squared test was again used for checking the statistical significance of the results. The null hypothesis was that the energy-efficient images look the same as their original images. For each tested image, the corresponding p-value of the collected votes was computed. The cases in which the p-value is less than 0.05 are marked by a '+' on their corresponding bar. For these cases, the null hypothesis is rejected, which means that the produced energy-efficient images do not look the same as their original images but for the other cases, they look the same as their original images. As seen from these results, in 40 cases out of 50 (i.e., in 80% of the cases), the energy-efficient images produced by the proposed method look the same as their original images. In fact, the overall p-value across all the 50 images was 0.0710, which means that the produced energy-efficient images cannot be distinguished from their original images perceptually, which is very desirable.

The resultant energy reduction percentage for each tested image is shown in the right graph in Fig. 6. Here, the energy reduction percentage is computed as $\Delta E = (E_2 - E_1)/E_1 \times 100$, where E_1 and E_2 are the total power consumption of the original image, and its energy-efficient version, respectively. As seen from these results, the average energy savings provided by the proposed energy-efficient images is about 14.1% at the same perceptual quality. This energy reduction amount can translate to many MegaWatts when all displays around the world are taken into account, which would be very considerable from the green computing point of view.

Fig. 7 shows two visual examples comparing the original color images with their corresponding energy-efficient images produced by the proposed method. These energy-efficient images consume about 14% less power than their original images. As seen from these examples, the perceptual quality of the energy-efficient images is very similar to the perceptual quality of their original images, which is very desirable as it complies well with the main objective of the proposed method.

The average processing time (optimization + image creation) for generating the energy-aware version of a given 1024×768 RGB image in the MIT dataset on an Intel Core i5 CPU @ 2.5 GHz with 6 GB RAM with an unoptimized code in 64-bit Matlab was about 5.8 seconds. Hence, the proposed method in its current form would be more suitable for server-side implementation or image on demand applications, where the energy-aware version of any arbitrary image can be generated once offline, and used indefinitely for all users. As a future work, further optimization is needed to make the proposed method suitable for real-time applications.

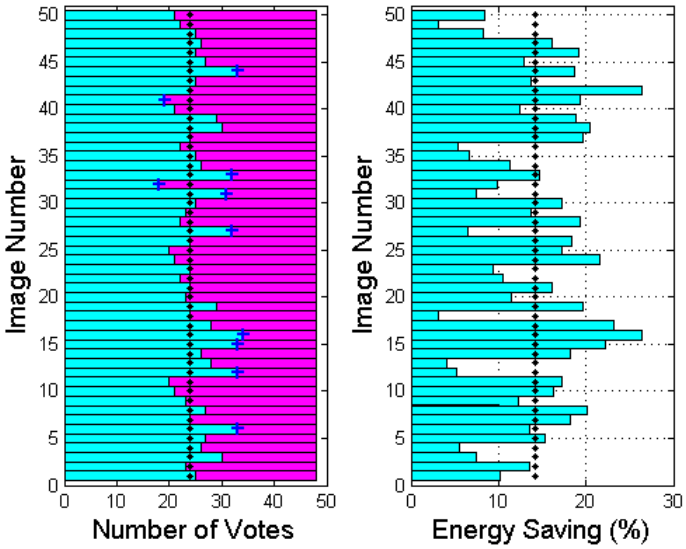

Fig. 6. (Left): The results of the subjective experiment for comparing the energy-efficient images with their corresponding original images based the number of votes collected from 24 participants. The cyan (left) bars show the number of votes for the original images while the magenta (right) bars show the number of votes for the energy-efficient images. The '+' marks show the cases in which the corresponding p-value is less than 0.05, and the vertical line shows the line on which the number of votes is equal to 24. Note that the total number of votes for each image is 48. (Right): The energy saving percent (ΔE) for each of the 50 tested images. On the vertical line $\Delta E = 14.1$.

Fig. 7. Two visual examples showing the quality of the proposed energy-efficient images. Left column: the original images. Right column: the energy-efficient images produced by the proposed method with $\Delta E = 14.0\%$ (top) and $\Delta E = 14.2\%$ (bottom).

V. Conclusions

In this paper, a novel framework was proposed for producing energy-efficient images that consume less electrical energy than their original images when displayed on energy-adaptive displays such as the new generation OLED displays, yet have the same or similar perceptual quality. In the proposed method, to reduce the overall energy consumption of a given image, the luminance of each pixel in the image is reduced by one JND threshold. To estimate the JND thresholds, a novel adaptive saliency-modulated JND (SJND) model was proposed. The proposed adaptive SJND model considers the interplay between the visibility thresholds of different pixels and their visual saliency. The parameters of this model are estimated adaptively based on the image content and its visual

saliency map. To evaluate the proposed method, a set of subjective experiments were conducted. Experimental results indicated that the proposed adaptive SJND model achieves a high accuracy in JND estimation for static images, and also it provides a high distortion-hiding capacity, which is very desirable in many practical applications such as watermarking, image compression, and image quality prediction. Also, the results obtained by real power measurements on an OLED display demonstrated that the proposed method for producing the energy-efficient images is able to reduce the energy consumption by about 14.1% on average at the same perceptual quality of the original images. This amount of saving per image can translate to many MegaWatts when all displays in the world are taken into account, which would be very considerable from the global warming point of view.

REFERENCES

[1] P. Ranganathan, E. Geelhoed, M. Manahan, and K. Nicholas, "Energy-aware user interfaces and energy-adaptive displays," *Computer*, vol. 39, no. 3, pp. 31–38.

[2] J. Wu, J. Thompson, H. Zhang, and D. C. Kilper, "Green communications and computing networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 148–149.

[3] J. Chuang, D. Weiskopf, and T. Moller, "Energy aware color sets," *Computer Graphics Forum (Proceedings of Eurographics Conference 2009)*, vol. 28, no. 2, pp. 203–211, Apr. 2009.

[4] S. Murugesan, "Harnessing green IT: Principles and practices," *IT Professional*, vol. 10, no. 1, pp. 24–33, 2008.

[5] M. Dong, Y.-S. K. Choi, and L. Zhong, "Power modeling of graphical user interfaces on OLED displays," *DAC 2009*, pp. 652–657.

[6] Y. Liu, M. Xiao, M. Zhang, X. Li, M. Dong, Z. Ma, Z. Li, and S. Chen, "GoCAD: GPU-assisted online content-adaptive display power saving for mobile devices in internet streaming," *Proceed. of the 25th Int. Conf. on World Wide Web*, pp. 1329–1338, 2016.

[7] H. Hadizadeh, I. V. Bajić, P. Saeedi, and S. Daly, "Good looking green images," *IEEE ICIP*, pp. 3177–3180.

[8] A. B. Watson, *Digital Images and Human Vision*. The MIT press, 1993.

[9] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. on Image Proc.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.

[10] F. A. A. Kingdom, *Psychophysics: A Practical Introduction*. Academic press, 2009.

[11] L. Itti, G. Rees, and J. K. Tsotsos, *Neurobiology of Attention*. Academic Press, 2005.

[12] A. Ahumada and H. Peterson, "Luminance-model-based DCT quantization for color image compression," *Vision Visual Process. Digital Display III*, pp. 365–374, 1992.

[13] A. Watson, "DCTune: a technique for visual optimization of dct quantization matrices for individual images," *Soc. Inf. Display Digest Tech. Papers XXIV*, pp. 946–949, 1993.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1254–1259, Nov. 1998.

[15] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, pp. 1489–1506, 2000.

[16] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, 2004.

[17] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," *Proc. IEEE Int. Conf. Image Process.*, pp. 171–175, Sep.

[18] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attentions modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, pp. 1928–1942, 2005.

[19] L. Itti, J. Braun, and C. Koch, "Modeling the modulatory effect of attention on human spatial vision," in *Advances in Neural Information Processing Systems (NIPS*2001), Vol. 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, Aug 2002, pp. 1247–1254.

[20] K. Herrmann, D. J. Heeger, and M. Carrasco, "Feature-based attention enhances performance by increasing response gain," *Vision Research*, vol. 74, pp. 10–20, 2012.

[21] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, p. 42704281, 2014.

[22] Y. Niu, M. Kyan, L. Ma, A. Beghdadi, and S. Krishnan, "Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking," *Signal Processing: Image Communication*, vol. 28, pp. 917–928, 2013.

[23] W. Kienzle, M. O. Franz, B. Scholkopf, and F. A. Wichmann, "Center-surround patterns emerge as optimal predictors for human saccade targets," *Journal of Vision*, vol. 9, p. 115, 2009.

[24] L. Kerofsky and S. Daly, "Brightness preservation for LCD backlight reduction," *SID Annual Technical Digest*, 2006.

[25] Z. Wei and K. N. Ngan, "Spatial just noticeable distortion profile for image in DCT domain," *Proc. IEEE International Conference on Multimedia and Expo (ICME10)*, pp. 925 – 928, 2008.

[26] G. E. Legge, "A power law for contrast discrimination," *Vision Research*, vol. 21, pp. 457–467, 1981.

[27] I. E. Richardson, *The H.264 Advanced Video Compression Standard*. Wiley, 2010.

[28] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 15, no. 6, pp. 745–752, Jun. 2005.

[29] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *J. Vis. Commun. Image R.*, vol. 19, pp. 30–41.

[30] B. J. Walter, S. N. Pattanaik, and D. P. Greenberg, "Using perceptual texture masking for efficient image synthesis," *Computer Graphics Forum*, vol. 21, no. 3, pp. 393–400.

[31] G. Sharma, *Digital Color Imaging Handbook*. CRC Press, 2002.

[32] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[33] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jun. 1965.

[34] L. Itti, C. Koch, and J. Braun, "Revisiting spatial vision: Toward a unifying model," *Journal of the Optical Society of America, JOSA-A*, vol. 17, no. 11, pp. 1899–1917, Nov 2000.

[35] M. M. Taylor and C. D. Creelman, "PEST: Efficient estimates on probability functions," *J. Acoustical Society of America*, vol. 41, pp. 782–787, 1967.

[36] E. Poulton, *Bias in Quantifying Judgments*. Hove, UK: Erlbaum, 1989.

[37] ITU-R, "Recommendation bt.500-8: Methodology for the subjective assessment of the quality of television pictures," ITU, Tech. Rep., 1998.

[38] H. Hadizadeh and I. V. Bajić, "Saliency-aware video compression," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 19–33, Jan. 2014.

[39] J. T. McClave and T. Sincich, *Statistics*. Prentice Hall, Upper Saddle River, NJ, 9th edition, 2003.

[40] "Monsoon power monitor," http://www.msoon.com.

Hadi Hadizadeh received the B.Sc.Eng. degree in electronic engineering from the Shahrood University of Technology, Shahrood, Iran, in 2005, the M.S. degree in electrical engineering from the Iran University of Science and Technology, Tehran, Iran, in 2008, and the Ph.D. degree in engineering science from Simon Fraser University, Burnaby, BC, Canada, in 2013. He is currently an Assistant Professor at the Quchan University of Advanced Technology, Quchan, Khorasan Razavi, Iran. His current research interests include perceptual image/video coding, visual attention modeling, error resilient video transmission, image/video processing, computer vision, multimedia communication, and machine learning. He was a recipient of the Best Paper Runner-up Award at ICME 2012 in Melbourne, Australia and the Microsoft Research and Canon Information Systems Research Australia Student Travel Grant for ICME 2012. In 2013, he was serving as the Vice Chair of the Vancouver Chapter of the IEEE Signal Processing Society.