Pattern Recognition Letters 84 (2016) 49–55

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A saliency-modulated just-noticeable-distortion model with non-linear saliency modulation functions☆

Hadi Hadizadeh

Quchan University of Advanced Technology, Quchan, Khorasan Razavi, Iran

ARTICLE INFO

Article history:
Received 23 February 2016
Available online 26 August 2016

Keywords:
Just noticeable distortion
Visual saliency
DCT

ABSTRACT

It is well-known that the human visual system (HVS) cannot sense small variations of visual signals below the so-called just-noticeable distortion (JND) thresholds due to their underlying spatial/temporal masking properties. It is also known that the visual attention mechanism of the human brain can enhance or reduce visual sensitivity. In other words, the visual attention has modulatory effects on JND thresholds. The current knowledge also states that the visual attention is mainly driven by visual saliency in an automatic and involuntary manner. In this paper we present a saliency-modulated JND (SJND) model for static images in the discrete cosine transform (DCT) domain. In the proposed model, the JND thresholds of each block in a given image are elevated by two non-linear modulation functions using the visual saliency of the block. The parameters of the saliency modulation functions are obtained through an optimization framework, which utilizes a state-of-the-art saliency-based objective image quality assessment method. To evaluate the proposed SJND model, two subjective experiments were conducted. The obtained experimental results demonstrated that the proposed method achieves a high accuracy in JND estimation, and also it provides a high distortion-hiding capacity.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

It is well-known that the Human Visual System (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 [1,2]. 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 contrast sensitivity function (CSF), which expresses the ability of the HVS to discern between contrast differences at various spatial and temporal frequencies [2].

In visual stimuli such as digital images, JND thresholds may vary due to several *visual masking* effects [1]. For instance, it is known that the visibility of one visual component may be reduced in the presence of another visual component. This implies that a lot of perceptual redundancies may exist in static images and videos. Various applications can exploit such redundancies to achieve a certain goal. For example, in the context of image or video compression, fewer bits can be allocated to blocks in which higher distortion can be tolerated [3], or in watermarking applications, the watermark signal can safely be embedded in regions with larger JND thresholds to achieve imperceptibility [4].

In the literature, several models have been proposed to estimate JND thresholds in both spatial and temporal domains for static images and videos [3,5–8]. Most of these models have been obtained through extensive psychophysical experiments on simple visual stimuli such as Gabor gratings of various spatial or temporal frequencies [1]. In such experiments, the subject is usually instructed to attend at a specific point in the screen at which a grating of a certain spatial/temporal frequency is displayed on a uniform gray background. The contrast of the grating is gradually increased from a low level (e.g. zero), and the subject must respond whenever he/she is able to distinguish the grating from the background. At that level of contrast, the visual sensitivity is obtained by taking the reciprocal of the obtained contrast value. However, several studies have shown that the sensitivity of the HVS is lower in un-attended regions of a visual scene [9,10]. This is because of a behavioral and cognitive process in the human brain called *the visual attention mechanism* [11].

Visual attention (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 [12]. Over the last decades, visual attention has been studied extensively, and several computational models of VA have

☆ This paper has been recommended for acceptance by Eckart Michaelsen.
E-mail address: h.hadizadeh@qiet.ac.ir

http://dx.doi.org/10.1016/j.patrec.2016.08.011
0167-8655/© 2016 Elsevier B.V. All rights reserved.

been developed [13–15]. According to the current knowledge, it is believed that VA is driven by "visual saliency", which is a measure of propensity for drawing VA. In a visual scene, a region is said to be visually salient if it possess certain characteristics, which make it stand out from its surrounding regions and draw our attention to it. The existing computational models of VA for static images are able to produce a *saliency map* by which salient regions and also gaze locations in the image can be predicted automatically [13].

With this introduction, it is observed that both JND and visual saliency provide perceptual redundancies, and also they are related to each other. JND indicates the visual sensitivity thresholds within attended regions, and visual saliency specifies the location and spatial extent of the attended regions. On the other hand, it is known that the visual sensitivity thresholds in attended regions are lower than in other un-attended regions [9,10,16]. In other words, visual saliency modulates JND thresholds. Therefore, it is reasonable to consider the modulatory effects of visual saliency on JND thresholds [4,16]. For this purpose, we propose a *saliency-modulated JND (SJND) model* in this paper.

Our 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, we elevate the JND thresholds of each block in the given image based on two saturating non-linear modulation functions, which their input is the visual saliency of the block. The parameters of these non-linear functions are obtained through an optimization procedure, which mimics a system-identification framework. The reason is that in the employed optimization procedure, a fixed model is considered for the problem, and it is fitted to some real data. Such an approach is very common in many neurophysiological studies [17]. For example, when modeling response properties of neurons, a generic model is fitted to experimental spike data such that it describes the stimulus–response relationship of the neurons. Here, we apply a similar idea to obtain a saliency-modulated JND model.

Given a set of images with their saliency maps generated by real fixation data, we first generate a random noise such that its amplitude is determined by the proposed SJND model. We then add the generated noise to all the images. 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 images becomes equal or very close to the perceptual quality of the original images while their peak signal-to-noise ratio (PSNR) value with respect to the original images 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 metric can be used. In this paper we use the *visual saliency-induced index* (VSI) [18], which is the state-of-the-art method for measuring the perceptual image quality using visual saliency information.

Note that although JND estimation for static images has been relatively well-developed, only few JND models exist, which consider the interplay between JND and visual saliency. To the best of our knowledge, the most recent method in this regard is the one proposed in [4], which in this paper, we refer to it as the "Niu's method". Our proposed method 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 we utilize two saturating non-linear functions for this purpose. Second, in the Niu's method, the parameters of the linear functions were set experimentally while in the method proposed here, we present a systematic and automatic framework for calculating the parameters of the employed functions. Our motivation for using a non-linear modulation function is twofolds. First, there are many saturating non-linear functions in biophysics [1,17]. Hence, we postulate that the modulatory effects of visual saliency on JND thresholds may be modeled by such a function better than a simple linear function. Second, a non-linear function has a more flexibility in describing data as it has more free parameters to tune.

It is worth pointing out that the proposed method in this paper is different from our previous works in [19,20]. In our previous works we utilized the concept of visual saliency for either video compression or video error concealment without using or proposing a model for JND estimation whereas in this paper we propose a model for JND estimation in still images using the concept of visual saliency.

To evaluate the proposed method, we performed a set of subjective experiments. The obtained results indicated that the proposed method achieves a high accuracy in JND estimation, and also it provides a better distortion-hiding capacity than the Niu's method at the same perceptual quality.

This paper is organized as follows. The related works are briefly reviewed in Section 2. The proposed SJND model is presented in Section 3. The experimental results are given in Section 4, followed by conclusions in Section 5.

2. Related works

In the literature various methods for JND estimation have been developed for both static images and videos [3–7,16]. These methods are mainly based upon spatial/temporal CSF, background luminance adaptation and contrast masking. Luminance adaptation accounts for the masking from local background luminance, while contrast masking accounts for the visibility reduction of one visual component at the presence of another one [1,2].

The JND thresholds can be determined for either pixels [3] or subbands [6]. Subband-based JND estimators are of particular interest since the CSF can be more easily incorporated in the frequency domain, and also most images are subband-coded. Prominent subband-based JND models are the ones proposed in [5] and [6]. For instance, Watson proposed a DCT-based JND model called DCTune [6] in which the JND thresholds for each DCT component are estimated as the product of a base threshold and an elevation factor. The base threshold is determined by the spatial CSF and luminance adaptation, while the elevation factor is determined by contrast masking. DCTune model was later improved by several researchers. For example, in [7], the DCTune model was improved to estimate the JND thresholds in very dark and very bright regions better.

Although there are various models for JND estimation, there are very few JND models that account for the effect of visual saliency on visibility thresholds. Notable methods in this regard are the ones presented in [4,16]. In the method proposed in [16], the visual attention's modulatory aftereffects on visual sensitivity are considered through a numerical expression called *the perceptual quality significance map* (PQSM). The PQSM is designed to reflect the statistical allocation of the human brain's processing resource on local visual contents, and it is obtained by the product of the influence of motion suppression and the visual saliency of the given video frame/image. In this method, the PQSM is used to scale the JND thresholds estimated by any arbitrary JND model through a set of linear modulation functions. The other more recent method in this regard is the saliency-modulated JND model proposed in [4]. This method is built upon a combined JND model in the DCT domain, and similar to [16], the estimated JND thresholds are scaled by linear modulation functions. In both of the above mentioned

Download English Version:

https://daneshyari.com/en/article/6940867

Download Persian Version:

https://daneshyari.com/article/6940867

Daneshyari.com