Neurocomputing 275 (2018) 366–376

Contents lists available at ScienceDirect

Neurocomputing

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

Just Noticeable Difference for natural images using RMS contrast and feed-back mechanism

Vinit Jakhetiya[a,*], Weisi Lin[c], Sunil Jaiswal[b], Ke Gu[d], Sharath Chandra Guntuku[c]

[a] Department of Computer Science Engineering, Bennett University, Greater Noida
[b] Hong Kong University of Science and Technology, Hong Kong
[c] School of Computer Engineering, Nanyang Technological University, 639798 Singapore
[d] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 20 November 2016
Revised 8 April 2017
Accepted 15 August 2017
Available online 7 September 2017

Keywords:
Just Noticeable Difference
Contrast sensitivity
RMS contrast
Feedback mechanism
Bilateral filter

ABSTRACT

Contrast Sensitivity (CS), Luminance Adaptation (LA) and Contrast Masking (CM) are important contributing factors for Just Noticeable Difference (JND) in images. Most of the existing pixel domain JND algorithms are based only on LA and CM. Research shows that the human vision depends significantly on CS, and an underlying assumption in the existing algorithms is that CS cannot be estimated in the pixel domain JND algorithms. However, in the case of natural images, this assumption is not true. Studies on human vision suggest that CS can be estimated using the Root Mean Square (RMS) contrast in the pixel domain. With this in perspective, we propose the first pixel-based JND algorithm that includes a very important component of the human vision, namely CS by measuring RMS contrast. This RMS contrast is combined with LA and CM to form a comprehensive pixel-domain model to efficiently estimate JND in the low frequency regions. For high frequency regions (edge and texture), a feedback mechanism is proposed to efficiently alleviate the over- and under-estimation of CM. To facilitate this, a prediction based algorithm is used to classify an image into low (smooth) and high frequency regions. This feed-back mechanism is based on the relationship between the CS and RMS contrast. Experiments validate that the proposed JND algorithm efficiently matches with human perception and produces significantly better results when compared to existing pixel domain JND algorithms.

1. Introduction

Just Noticeable Difference (JND) is a visibility threshold below which change in an image cannot be sensed by the human visual system. In the human vision (HV), perceived information highly depends upon signal characteristics such as spatial frequency and contrast of signal [1–5]. In general, a change in a signal which is imperceptible to 75% of viewers can be defined as the JND value for the corresponding signal.

JND profiles are used in several multimedia applications. Information which cannot be sensed by the eyes can be removed with the guidance of JND and does not require to be coded in a bit-stream or this removed information help to enhance the accuracy of the image quality assessment matrices. Therefore, JND profiles are extensively used for multimedia coding [6,8,14,18,20]. The authors of [16,17,25,26] estimated video and image quality, respectively, using JND profiles. The JND profiles are also used for watermarking [15,19]. In the same line, researchers have also used JND profiles to guide visual signal enhancement [13,21]. Recently, JND profiles received a lot of attention and these profiles are used in different multimedia applications, such as, Fang et al. [27] used important component of JND (namely contrast sensitivity) for the saliency detection and authors of [28,29] used it for the depth sensation enhancement. Interestingly, JND profiles also have the ability to guide the seam carving [30] and enhancement of backlight-scaled images [31]. Such widespread use of JND reveals the significance of developing more accurate models to enhance the accuracy of multimedia applications.

Contrast sensitivity (CS), Luminance Adaptation (LA) and Contrast Masking (CM) are important contributing factors for JND in images. LA is the ability of human vision to adapt with the change in luminance and estimation of the LA is based upon the psychological experiments [20], and CM, which refers to the visibility

* Corresponding author.
E-mail addresses: vinit.jakhetiya@bennett.edu.in (V. Jakhetiya), wslin@ntu.edu.sg (W. Lin), spjaiswal@connect.ust.hk (S. Jaiswal), guke@bjut.edu.cn (K. Gu), sharathg@sas.upenn.edu (S.C. Guntuku).

http://dx.doi.org/10.1016/j.neucom.2017.08.031

Table 1
Comparison between existing pixel domain JND models [6–8,20].

Attribute	Chou et al. [20]	NAMM [8]	Liu et al. [7]	Wu et al. [6]
Luminance Adaptation or Texture Masking	Yes	No	No	No
Luminance Adaptation and Contrast Masking	No	Yes	Yes	Yes
Image classification	No	No	Yes	No
Contrast sensitivity	No	No	No	No
CM based upon edge strength	No	Yes	Yes	No
CM based upon prediction errors	No	No	No	Yes
Under and/or over-estimation of CM	Yes	Yes	Yes	Yes

reduction of one signal in the presence of other signals, is estimated based upon the edge and texture strength [6–8,20]. CS, on the other hand, gives an idea of our visual system's ability to differentiate an object, from its background. In a broad sense, JND algorithms can be divided into two categories: pixel domain and sub-band. As the HV highly depends upon CS, existing sub-band JND algorithms estimate CS using the Contrast Sensitivity Function (CSF) in sub-bands [22]. However, existing pixel domain JND estimation algorithms do not have such capability due to the lack of an effective way to account for CS in pixel domain. For this reason, many sub-band JND algorithms have been proposed in the literature, while only a few pixel domain JND algorithms have been proposed. In many applications (such as image and video coding [6–8], enhancement [13] and quality assessment [17]), a direct pixel domain JND estimator is preferred. The importance of CS can be seen by the fact that model in [24] has to convert a pixel domain model into the sub-band, in order to consider CS. Therefore in this paper, we concentrate on pixel domain JND estimation which includes CS.

1.1. Related work

As aforementioned, a few pixel domain JND models have been previously proposed in the literature [6–8,20]. In general, HV is sensitive to the slowly varying regions, such as smooth regions and regions with weak edges, and small changes in these regions can be easily identified by HV. While HV is less sensitive to the high frequency regions and changes made in these regions are being unnoticed by HV. With this view, in existing JND algorithms, the main components for JND estimation are LA and CM [8] for low and high spatial frequency regions, respectively. The first attempt for pixel domain JND estimation was made in [20], in which JND threshold was estimated using either LA or contrast (texture) masking. In the same line, Yang et al. [8] proposed a JND model called the Non-linear Additivity Model for Masking (NAMM), which considers both components. Liu et al. [7] proposed an algorithm, in which an image is decomposed so that edge masking and texture masking are calculated separately for the edge and texture regions. Recently, Wu et al. [6] proposed a JND algorithm based upon the free-energy principle. In this model, prediction errors are used as the CM for the JND estimation for textural regions, while for the other regions, JND is estimated using the existing NAMM [8]. In general, these algorithms are not specifically designed for any applications. However, authors of these algorithms have applied corresponding models to remove the perceptual redundancy of images and/or videos for the coding purpose [6–8]. The main attributes of these existing pixel domain JND models [6–8,20] are summarized in Table 1.

From Table 1, one can observe that existing JND algorithms have two major issues:

(1) All the above described pixel domain JND algorithms [6–8,20] only use LA and CM for JND estimation. As such, these algorithms cannot include the effect of CS, which in turn leads to visible artefacts in images, especially in the smooth regions.

(2) Furthermore, in high spatial frequency areas (edge and texture), CM estimation in the existing algorithms is only based upon edge and texture strength [7,8,20] or prediction errors [6], which may lead to under- or over-estimation [22] of the CM. The changes made in the signal guided by the over estimated CM can be easily sensed by human visual system [1,2,22]. In these situations, inaccurate JND estimation can severely affect the efficiency of the multimedia applications.

To overcome the above-mentioned problems associated with the existing pixel domain JND algorithms, we propose a comprehensive and efficient pixel domain JND algorithm in which, we merge the effect of CS (by measuring RMS contrast) with LA and CM for estimating JND. We also propose a novel feedback mechanism, which efficiently alleviates the over- and under- estimation of the CM in high spatial frequency regions (such as edge and texture). The decomposition of an image into smooth, edge and texture regions is based on prediction errors of the input image. From experiments, it is validated that the proposed algorithm produces significantly better results as compared to existing pixel domain JND algorithms and efficiently matches with the human perception.

In short, the main contribution of this paper is to propose the first pixel domain JND algorithm to include the most important component of the HV, namely CS, which is combined with LA and CM for accurate JND estimation. This accurate estimation of JND can help to enhance the efficiency of the several multimedia applications.

The rest of the paper is organized as follows. Section 2 describes both the proposed decomposition of an image into smooth, edge and texture regions using prediction errors, and the JND algorithm. The comparison of the proposed algorithm with the existing state-of-the-art algorithms and results of subjective tests are provided in Section 3, and the discussions and concluding remarks are given in Sections 4 and 5, respectively.

2. Proposed contrast sensitivity and feedback mechanism based JND algorithm

In the proposed algorithm, we try to build a computational JND model which efficiently matches with human perception [1–5]. The relationship between the CS and spatial frequency is represented using the parabolic curve (CSF) [1,2]. In general for natural images, the HV is more sensitive towards the low spatial frequency regions when compared to high spatial frequency regions [1–8]. Therefore, JND should be higher in the high frequency regions as compared to low frequency regions [6–8]. Accordingly, we decompose an image into low and high frequency areas, based upon the prediction errors of the input image. For the low frequency areas, we propose a new algorithm which takes into account the effect of LA, CM and CS (RMS contrast), while for high frequency regions we propose to have a feed-back mechanism based upon the RMS contrast, which can efficiently control the CM.

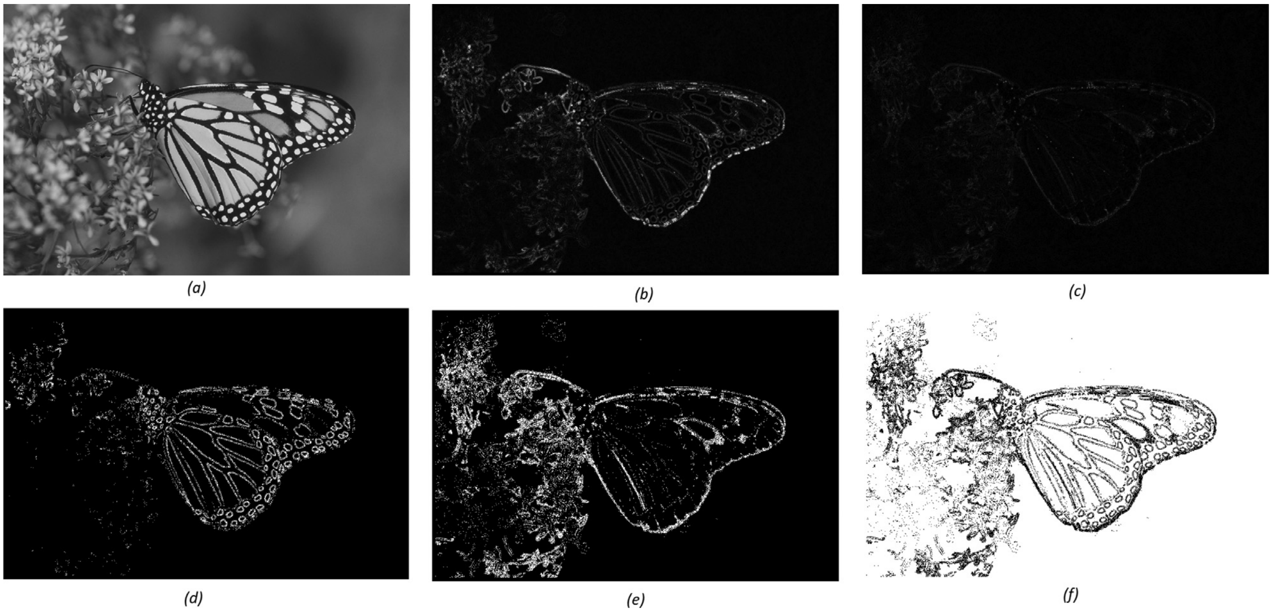

Fig. 1. Image decomposition based upon the prediction errors for the 'Monarch' image: (a) original image, prediction residual errors of (b) bilinear interpolation, (c) observation model based bilateral filter [9,10], and classified (d) edge, (e) texture and (f) smooth regions of the 'Monarch' image. In these binary images (d), (e) and (f), white pixels represent, edge, texture and smooth pixels, respectively, in the 'Monarch' image.

2.1. Image classification in smooth, edge and texture regions

Edge-directed prediction algorithms (such as the perceptually motivated [12] predictor and bilateral filter based interpolation [10] and deinterlacing [9] algorithms) can efficiently predict regular edges and smooth pixels, but these algorithms cannot predict texture information [6]. On the other hand, a simple bilinear interpolation algorithm cannot predict both edges and texture pixels, but it can predict smooth pixels efficiently. With this view, we decompose an image into smooth, edge and texture regions [7] by combining two prediction algorithms. We use the Observation model based Bilateral Filter (OBF) [9] as the edge directed prediction algorithm, as the OBF is computationally very simple and also has very good edge preserving capability. The OBF predicts the input image (B) as

$$\hat{P}_1 = \frac{\lambda B + \sum_{k \in N} w_k A_k}{\lambda + \sum_{k \in N} w_k} \tag{1}$$

where A_k and w_k are the pixels in the neighboring 3×3 window N and corresponding weights, respectively, while, λ is the parameter which controls the prediction accuracy. A higher value of λ increases the prediction accuracy and some of the texture pixels will be efficiently predicted. In this situation, these texture pixels will be wrongly classified as edge pixels. On the other hand, a lower value of λ results in poor prediction and some of the edge pixels will be wrongly treated as texture pixels. Taking this into account, in our experiments, we chose λ to be 0.3. These weights (w_k) are calculated based upon the pixel gradient and radiometric distance, as suggested by Hung et al. [10]. The bilinear predictor is used for the second prediction ($\hat{P}_2$) using pixels in the neighboring 3×3 window with equal weights. The prediction errors for these predictors are calculated as;

$$e_1 = |B - \hat{P}_1|; \;\; e_2 = |B - \hat{P}_2|. \tag{2}$$

As both of the predictors cannot adapt in the textural area and subsequently, both prediction errors (e_1 and e_2) for such kinds of pixels have a high value (as shown in Fig. 1 (b) and (c)). However for pixels with edges, the OBF predictor can adapt accurately but the bilinear interpolation based predictor fails. Hence, based on these observations, we decompose the pixels of an image into smooth, edge and texture pixels based on the prediction errors. The classification strategy is as follows:

Table 2
Image classification based upon the prediction errors of OBF (e_1) and bilinear interpolation (e_2).

Condition	OBF based (e_1)	Bilinear based (e_2)	Classification
I	$e_1 \geq Th_1$	$e_2 \geq Th_2$	Texture Pixel
II	$e_1 < Th_1$	$e_2 \geq Th_2$	Edge Pixel
III	$e_1 < Th_1$	$e_2 < Th_2$	Smooth Pixel
IV	$e_1 \geq Th_1$	$e_2 < Th_2$	Smooth Pixel

(1) If both the predictors can predict a pixel and produce small prediction errors, it suggests that this pixel is in a smooth region, condition III in Table 2.
(2) If both the predictors cannot predict a pixel, it suggests that this pixel belongs to the textural region, condition I, in Table 2.
(3) If predictor 1 can predict it efficiently and predictor 2 cannot, it suggests that this pixel belongs to the edge region, condition II, in Table 2.

The classification results for the 'Monarch' image is shown in Fig. 1. From experiments, we chose both thresholds Th_1 and Th_2 (for the corresponding predictors) to be 6. In the proposed algorithm, the pixels which cannot be classified as either edge or texture (condition IV) are classified as smooth pixels (only a few pixels fall into this category). The motivation of our prediction based decomposition algorithm is not to propose a competitive classification algorithm but to classify pixels in such a way that none of the smooth pixels are classified into edge or texture pixels. As CM should be quite high for high frequency regions as compared to smooth regions. If a smooth pixel is wrongly classified, it will have high CM and changes made in signal guided by this CM can be easily sensed by the HV [1–5], thereby decreasing the efficiency of the JND model.

In the proposed algorithm, these two prediction errors play a two fold role: (1). decomposition of the input image into smooth, edge and texture regions and (2). estimating the CM [6] for the edge and texture pixels (details are given in Section 2.3).

2.2. Low frequency regions

The existing pixel domain JND algorithms only consider LA and CM for estimating the JND value. As such, these algorithms [6–8,20] cannot estimate the CS due to lack of efficient way of estimating it in the pixel domain. So these algorithms assume that CS does not play an important role in the HV in the low frequency areas. Contrary to this argument, research on the HV clearly suggests that CS plays a very important role in achieving contrast invariance (ability of the human visual system to adapt to the change in contrast) [2]. Studies on the HV suggest that CS also can be estimated in the pixel domain using RMS contrast for natural images [1,4], and even to that RMS contrast has a better ability of representing the human vision for natural images [3] than the contrast sensitivity function. To address this issue, we combine the RMS contrast with LA and CM to efficiently estimate the JND value, which matches with human perception. The RMS contrast [1,3,4] is defined as;

$$C_{rms}(i) = \frac{\sigma_{Lum}(i)}{\mu_{Lum}(i)} \tag{3}$$

where $\sigma_{Lum}(i)$ and $\mu_{Lum}(i)$ are the standard deviation and mean of the ith pixel in the smooth area, respectively. In the proposed algorithm, we use a 5×5 window to calculate the $\sigma_{Lum}(i)$ and $\mu_{Lum}(i)$. We propose the JND model for low frequency regions as;

$$JND_{smooth}(i) = \alpha(i) \times NAMM(i) \tag{4}$$

where $\alpha(i)$ and *NAMM* are the contribution of the RMS contrast and non-linear additivity model for masking [8], respectively. This contribution $\alpha(i)$ is estimated using the relationship between the RMS contrast and CS [1], and this relationship was obtained from the psychological experiments on the HV in [1]. The *NAMM* [8] is defined as

$$LA(i) = \begin{cases} 17 \times (1 - \sqrt{(\frac{B(i)}{127}}) & \text{if } B(i) \leq 127 \\ \frac{3}{128} \times (B(i) - 127) + 3 & \text{else} \end{cases} \tag{5}$$

$$SM(i) = [0.01B(i) + 11.5]\,[0.01G(i) - 1] - 12 \tag{6}$$

$$NAMM(i) = LA(i) + SM(i) - 0.3 \times min\{LA(i), SM(i)\} \tag{7}$$

In the *NAMM* [8] $LA(i)$, $SM(i)$, $B(i)$ and $G(i)$ are the Luminance Adaptation, Spatial Masking, background luminance of the ith pixel and maximum height of edge in the 5×5 neighborhood, respectively. More details about the *NAMM* can be found in [8].

From Eq. (3), it can be observed that RMS contrast can have a higher value because of either the low mean (μ) value (in smooth regions with low luminance) or high value of the standard deviation (σ). In dark (low luminance) smooth regions, the RMS contrast is quite high and the contrast control gain mechanism in the HV cannot adapt to changes in such regions [1,5]. Therefore, CS has a high value [1,5] and a slight change in these regions can be noticed by the HV. Hence, the JND should decrease with the increment of RMS contrast (because of the low mean μ) in the smooth regions.

On the other hand, if a signal has a high standard deviation (σ), the change in the signal can not be identified by the HV, as the contrast gain mechanism in the HV can adapt to changes in such regions [1,2,5]. So, once the standard deviation is higher than the threshold, the CS becomes quite low and JND can have a higher value. We empirically determined this threshold to ideally be 10. By considering the above mentioned arguments, we model the contribution of RMS contrast in JND estimation (α) as

$$\alpha(i) = \begin{cases} e^{-(C_{rms}(i)-\gamma)^2} - 0.3 & \text{if } \sigma \leq 10 \\ 1.4 & \text{else} \end{cases} \tag{8}$$

where γ is the mean value of the RMS contrast (C_{rms}) and, as suggested by the HV models for smooth regions, we chose the γ value to be 0.01.

We call the proposed JND model for the smooth region as the RMS–NAMM, and this algorithm has the ability of incorporating the important component of the HV, CS by measuring RMS contrast. This inclusion of RMS contrast significantly improves the NAMM [8], and the proposed RMS–NAMM is able to produce more visually appealing results as compared to the existing pixel domain JND algorithms [6–8] (details are given in the experimental results, Section 3), in the smooth regions.

2.3. High frequency regions

In HV, the contrast threshold which is the minimum contrast required to see a target reliably and it is a reciprocal of the contrast sensitivity [4], does not increase monotonically with the RMS contrast [1]. In fact, the contrast threshold increases with RMS contrast, and above 0.4 RMS contrast, the contrast threshold starts to decrease. Even, with the RMS contrast value at 0.8, the contrast threshold becomes approximately similar to the contrast threshold at 0 RMS contrast [1]. In the situation of RMS contrast above 0.4, with a high amount of change in the signal, the contrast control gain mechanism of the HV cannot adapt efficiently and this change can be sensed by the HV.

On the other hand, if the RMS contrast is below 0.4 [1], the high amount of change in the signal cannot be sensed by the HV, as HV can attenuate this change [1,2,4] by adjusting the contrast control gain. The JND in the high frequency regions is mainly dominated by CM [6–8]. The above given arguments suggest the need for inclusion of RMS contrast with the CM. Hence, in the proposed algorithm, we modulate CM to efficiently control the JND estimation in the high frequency regions.

In short, the contrast gain mechanism of the HV can adapt and alleviate the effect of contrast change (contrast adaptation) if the RMS contrast is below 0.4, and above this RMS contrast, the ability to adapt to change in the signal decreases [1,2]. The contrast control gain mechanism provides a degree of contrast invariance (ability of the human visual system to adapt to the change in contrast) [2]. In view of this, we inspired by the HV models [1,4] we propose a feed-back mechanism based upon the RMS contrast, which can control the CM. We modulate the CM for estimating the JND in high frequency areas as;

$$JND_{HF}(i) = CM(i) \times e^{-(C_{rms}(i)-0.75)/(\sigma_{HF})} - C \tag{9}$$

where $CM(i)$ is the contrast masking of the ith pixel, which is estimated using the prediction errors [6] ($CM(i) = e_1(i)$ if the pixel belongs to the texture category and $CM(i) = e_2(i)$, otherwise), and C and σ_{HF} are the constant values. From experiments, for the texture pixels we chose C and σ_{HF} to be 0.1 and 2, respectively. As the HV is more sensitive to edge regions as compared to texture ones [23], CM should have a lower value in such regions. With this view, for the edge pixels these values were chosen to be 0.8 and 8, respectively. Here, $C_{rms}(i)$ is the RMS contrast (3) of the input image.

To estimate the efficient JND_{HF}, the CM is controlled by the feed-back parameter, as shown in Fig. 2. This feed-back parameter enhances the effect of prediction errors (CM) if the RMS contrast is below 0.4 and consequently, the JND can have a higher value. On the other hand, the feed-back parameter dampens the effect of prediction errors if the RMS contrast of the input image reaches the limit at which the humans can sense the change [1] in the signal. In this situation, the CM should be reduced. Hence, this feed-back parameter modulates the CM, in such a way that the HV cannot sense any change.

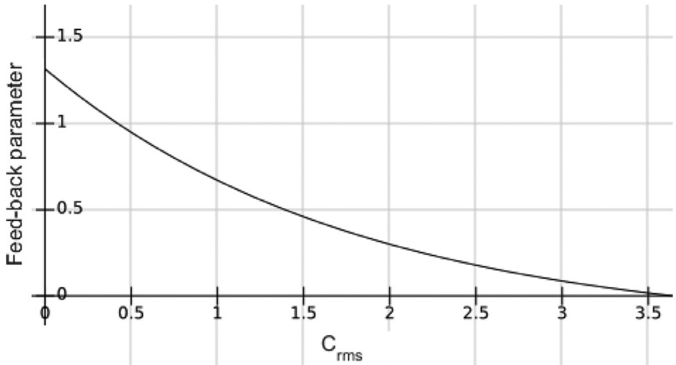

Fig. 2. Proposed relationship between the RMS contrast and feed-back parameter for texture regions.

The proposed algorithm is the first pixel domain JND algorithm which is able to modulate the effect of CM by incorporating the essential component of the HV, contrast sensitivity (by measuring RMS contrast). For this reason, the proposed algorithm can efficiently alleviate the under- or over- estimation of the CM and produces better results (more details are given in the experimental results, Section 3) as compared to the existing pixel domain JND algorithms. The proposed algorithm, especially performs very well when over estimation of CM occurs, as in this situation the change in the signal guided by the CM can be easily sensed and the output image, therefore looks visually unappealing.

2.4. Overview of JND estimation model

In the proposed algorithm, we consider three steps for efficiently estimating the JND value (as shown in Fig. 3). These three steps are the following: (1) classification of the input image into smooth, edge and textural regions; (2) the proposed RMS–NAMM for the smooth pixels; and (3) the feed-back mechanism for the edge and texture pixels. As discussed earlier, for smooth pixels, the JND value is mainly dominated by the LA and the CM does not contribute much, while for the edge and texture pixels, both factors (LA and CM) contribute significantly [8,20]. So, in order to calculate the final JND value, we have adopted the non-linear NAMM [8] to remove the overlapping between JND_{smooth} and JND_{HF}. The final JND value is calculated as:

$$JND(i) = JND_{smooth}(i) + JND_{HF}(i) - 0.3 \times min\{JND_{smooth}(i), JND_{HF}(i)\} \quad (10)$$

3. Experimental results

In this paper, we propose a comprehensive JND estimation model based upon the CS (RMS contrast) and feed-back mechanism. The proposed algorithm is compared with the state-of-the-art pixel domain JND algorithms, such as the NAMM [8], Liu's et al., method [7] and Wu's et al. method [6]. To compare the proposed algorithm with the existing pixel domain JND algorithms, noise is injected (shaped) into the image and this noise shaping is guided by the corresponding JND (of each algorithm). The noise shaping is done as follows:

$$\hat{B}_i = B_i + \beta \times \Delta_i \times JND_i \qquad \Delta_i \in \{-1, +1\} \quad (11)$$

For the fair comparison, all algorithms should inject noise with the same energy (in terms of MSE), and this is achieved by regulating the parameter β. A better JND model should be able to hide (inject) a higher amount of noise than others while being insensitive to the human visual system. Alternatively, with the same amount of noise injection, the noise injected images using a better JND model should look more visually appealing than other noise injected images. Similar to the previous studies [6–8,20], we also injected same amount of noise and checked the accuracy of the proposed algorithm.

Table 3
Scoring strategy for the subjective on-line test.

Score	0	1	2	3
Description	Same quality	Moderately better	Better	Much better

3.1. Subjective testing methodology

In order to compare the proposed JND algorithm with the existing pixel domain JND algorithms, an on-line subjective viewing test was conducted on 12 widely used test images. This subjective test was similar to the DSCQS (Double Stimulus Continuous Quality Scale) type II (ITU-R BT.500–11) standard [11]. The only difference between the DSCQS standard and method used to validate the proposed algorithm is that in DSCQS standard testing is performed in the laboratory conditions, while in the proposed algorithm, we chose to have an on-line subjective evaluation. The outcome of the on-line subjective test is shown in Table 4. For each case, two images were randomly shown side by side and subjects were asked to compare these images and give the better image a score according to the description in Table 3 (as shown in Fig. 4). One of these two images was a noise injected image according to the proposed algorithm and the other was with the existing JND algorithm [6–8]. The viewing conditions are same as those followed in [6,7]. The test conditions are different than the used in [6,7], as in these algorithms subjective tests were performed in the laboratory conditions. In order to have a more practical comparison of the proposed algorithms with existing algorithms, we preferred on-line subjective testing over the subjective testing done in the labora-

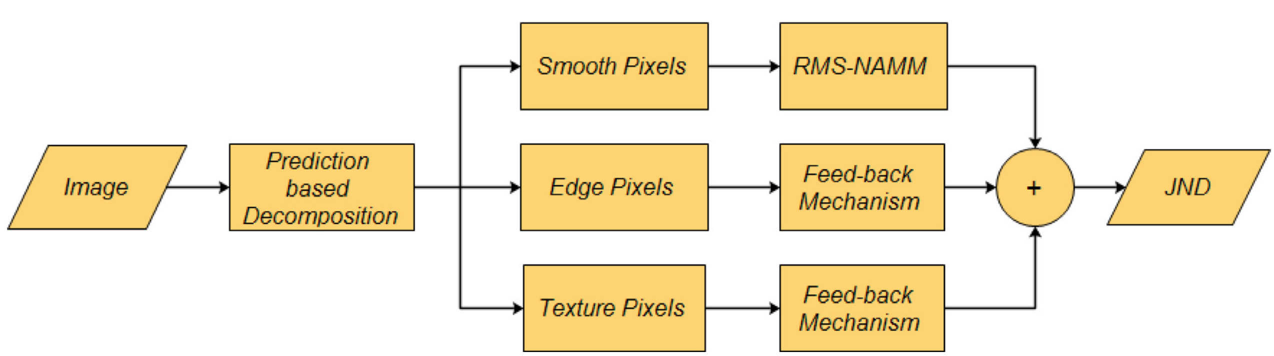

Fig. 3. Brief overview of the proposed algorithm.

Table 4
Comparison results of the on-line subjective test for the noise injected images, guided with the existing and proposed JND models.

Image	MSE	Mean			Standard deviation		
		Prop. Vs [8]	Prop. Vs [7]	Prop. Vs [6]	Prop. Vs [8]	Prop. Vs [7]	Prop. Vs [6]
Machine	142.75	0.730	0.324	0.568	0.932	0.884	1.094
Barbara	152.32	1.757	1.162	1.757	0.955	1.014	0.895
Bike	305.41	1.081	0.892	0.054	1.233	0.966	1.268
Cameraman	97.53	1.757	0.676	1.162	1.090	1.292	0.958
Computer	67.76	1.811	0.784	0.892	0.844	1.158	0.809
Lena	52.00	1.162	0.216	0.540	1.014	1.004	0.869
Monarch	40.92	0.541	0.540	0.243	1.117	0.803	0.956
Peppers	52.65	0.892	0.459	0.432	0.737	0.836	0.689
Pirate	87.44	0.757	0.730	0.568	1.234	1.018	0.929
Ruler	509.6	1.459	0.703	1.162	1.260	1.102	1.482
Target	196.83	1.757	1.081	1.405	1.342	1.422	1.518
Tire	151.08	2.108	0.892	1.216	0.737	1.286	1.084
Average		1.318	0.705	0.833	1.046	1.065	1.046

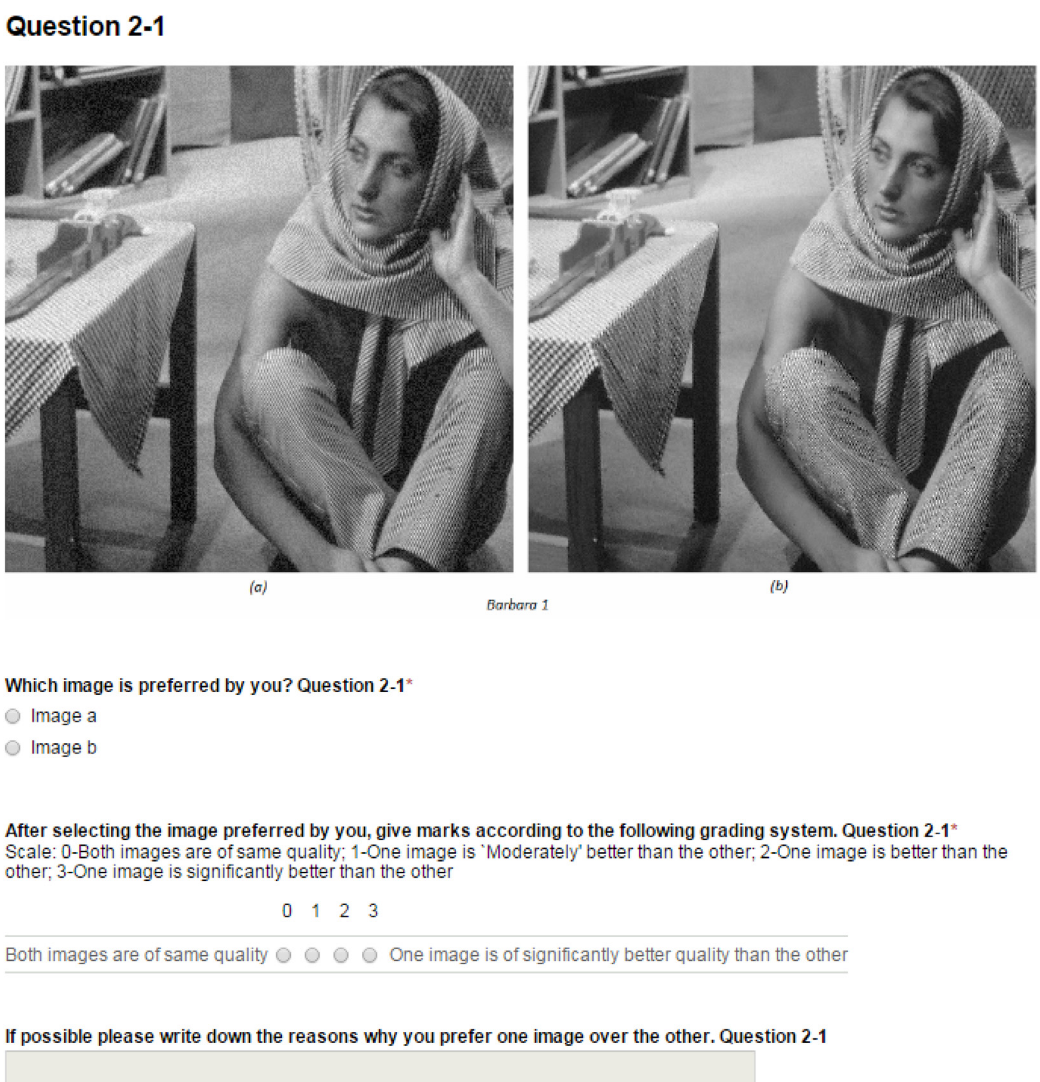

Fig. 4. The graphical interface of the on-line subjective evaluation. One of these two 'Barbara' images is a noise injected image according to the proposed algorithm and the other is with the existing JND algorithm [6–8].

tory conditions. As screen properties such as size and contrast of the display screen can not be controlled during the on-line testing.

In the subjective test, 37 subjects were asked to judge the images, and among these subjects, there was a good mix of image processing experts and non-experts (as shown in Fig. 5 (e)). Each subject is asked to perform 36 assessments (12 images and for each image, a comparison between the proposed algorithm and 3 existing algorithms [6–8]). During the on-line evaluation, we have asked five questions to the subjects, these five questions give information about the device used during the subjective evaluation, gender, age, education level and expertise in the image processing of the subjects.

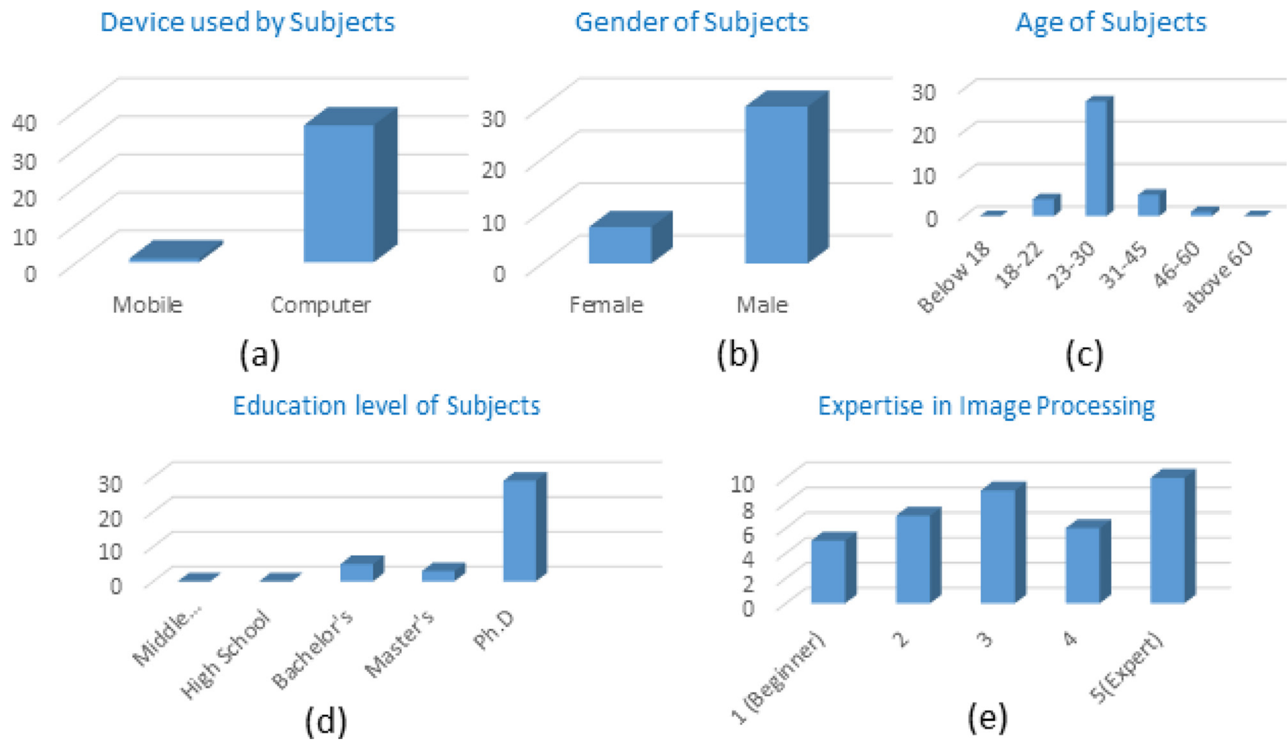

Fig. 5. Distribution of the device used by subjects (a), gender (b), age (c), education level (d) and expertise in image processing (e) of the 37 subjects.

This on-line subjective was done using the Google form and we have asked 37 subjects to fill this Google form. These 37 subjects were from the Singapore, South Korea, Hong Kong, USA, France, and India. In the on-line subjective test, we have requested subjects to spend less than 5 s on each pair of images and viewing distance should be between 14” and 24”. One session typically takes 5–15 min. The distributions of the age, gender, education level, device and expertise in image processing of the subjects are shown in the Fig. 5.

3.2. Quantitative results

The comparison results of the subjective test are shown in Table 4. The mean value is the average score of the 37 subjects and a positive or negative mean score shows whether the proposed algorithm is better or worse than the other algorithms, respectively. For example, the mean value (of scores given by 37 subjects) for ‘Barbara’ image is 1.757 when noise injected images guided by the proposed algorithm and Wu’s et al. method [6] are shown side by side during the on-line evaluation (as shown in Fig. 4), which suggests that the noise injected image using the proposed algorithm was chosen to be moderately better or better than Wu’s et al. method [6], during on-line evaluation. For all of the 12 images, the proposed algorithm achieves a positive mean value, which clearly suggests that it produces more visually appealing noise shaped images than the other existing pixel domain JND algorithms [6–8] with the same amount of noise injection.

In the proposed algorithm, the standard deviation is used to verify the consistency of the subjects during the on-line subjective evaluation and sustainability of the method chosen to evaluate the proposed algorithm. The Lower standard deviation, suggests similar scores were given by the 37 subjects for a particular image during the on-line evaluation and subjects were consistent, and vice-versa. The average standard deviation of responses given by the 37 subjects for 12 images is close to 1 (as shown in Table 4) when noise injected images using the proposed algorithm and state-of-the-art algorithms [6–8] were shown side by side. It suggests that throughout the subjective experiment, subjects were consistent and the majority of the subjects chose the same image as better than the other and gave a nearly similar quality score.

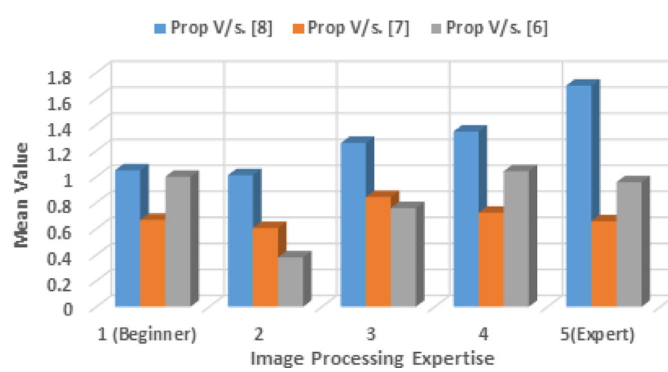

Fig. 6. The dependency of the performance of the proposed algorithm on expertise in image processing of subjects. Here horizontal axis represents the image processing expertise of 37 subjects, while the vertical axis shows the mean value of scores given by subjects.

In order to show that the proposed algorithm has better noise shaping capability and this capability can be perceived by the subjects with different levels of expertise in image processing, in Fig. 6 dependency of the performance of the proposed algorithm on expertise in image processing of subjects is shown. From Fig. 6, one can observe two important aspects, (1), even the subjects with no experience in image processing can perceive that the proposed algorithm has better noise shaping capability with respect to the existing algorithms [6–8]; (2), subjects which are experts in the image processing can perceive more difference between the noise

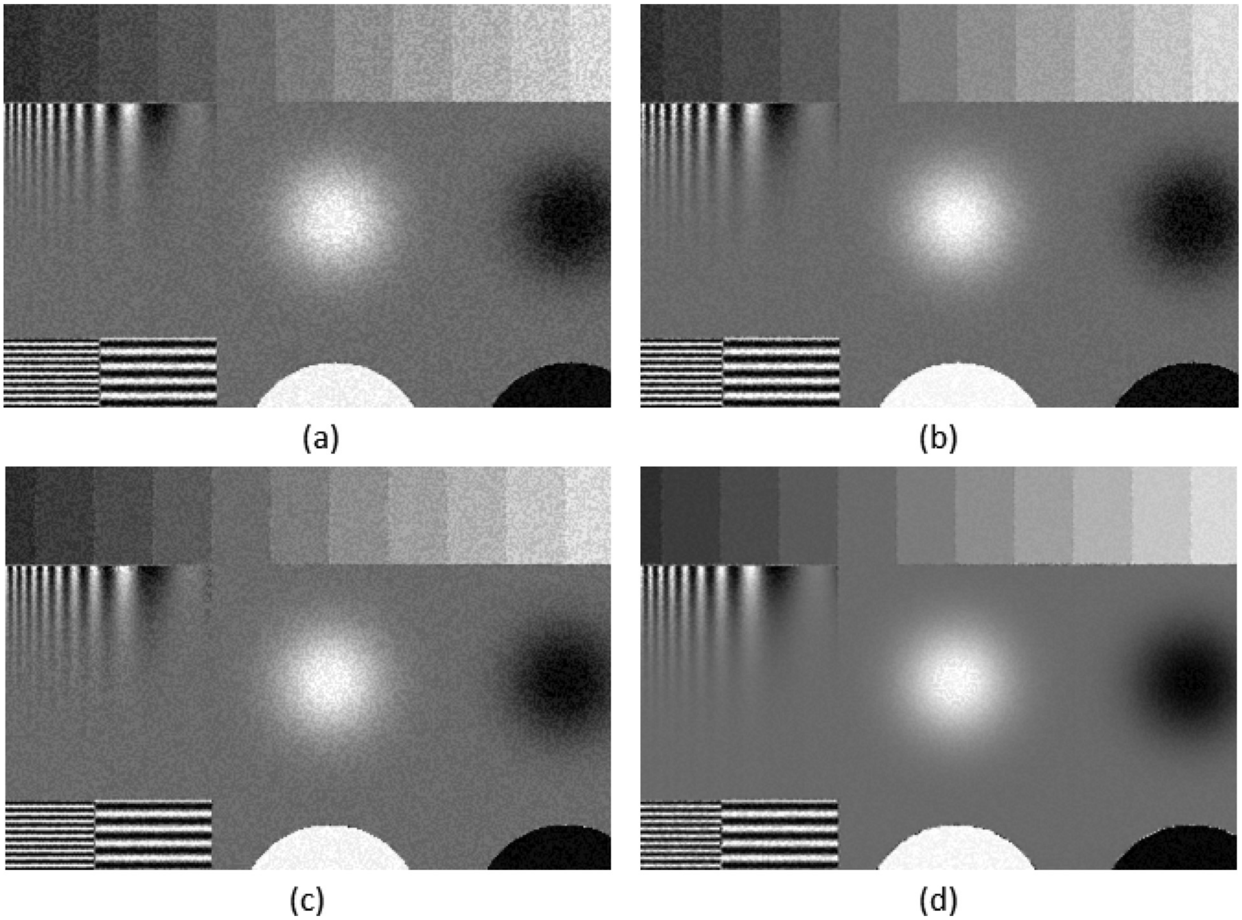

Fig. 7. Comparison of subjective quality for cropped noise injected (MSE=196.83) 'Target' image guided by the (a) NAMM [8], (b) Liu's et al. method [7] (c) Wu's et al. method [6] and, (d) the proposed algorithm.

shaped images guided by the proposed algorithm and existing algorithms.

3.3. Qualitative results

In order to show the contribution and efficiency of each individual component of the proposed algorithm, we show three subjective experimental results for the RMS–NAMM, feedback mechanism and, overall proposed JND model in Figs. 7–9, respectively.

To show the efficiency of the proposed RMS–NAMM, we injected similar amount of noise (in terms of MSE) guided by the proposed algorithm and existing algorithms [6–8]. In these three algorithms guided noise is mainly dominated by the LA and spatial masking (SM), only the LA and both LA and SM in smooth regions, as suggested by the NAMM [8], Liu's et al. model [7] and Wu's et al. model [6], respectively. Similarly, in smooth regions, in the proposed algorithm noise injection is mostly due to the RMS–NAMM and feed-back mechanism has a little impact. This noise injection is based on (11) and a similar amount of noise is injected by controlling the parameter β. From Fig. 7, it can be observed that the existing algorithms cannot shape the noise efficiently and the injected noise is visible. This effect is quite visible in the smooth areas, which have low luminance. These experimental results confirm that inclusion of CS can significantly improve the existing NAMM [8]. The proposed JND model based upon the RMS-contrast can efficiently represent the HV and out-performs the existing pixel domain JND algorithms [6–8] in the smooth areas.

The existing pixel domain JND algorithms [6–8] do not have any feed-back mechanism, and CM estimation is only based upon either edge strength [7,8] or prediction error [6]. Therefore, these algorithms may suffer from under and over estimation of CM [22]. In Fig. 8, noise injected 'Ruler' image is shown, in which noise is guided using the proposed algorithm and existing algorithms [6–8]. It can be observed from Fig. 8 that the NAMM [8] and Liu's et al. method [7] produces visually appealing noise injected images (this noise injection is mainly guided by the CM, in high frequency regions) but a much lower amount of noise has been injected (MSE = 36.35 for the NAMM [8] and MSE = 45.77 for Liu's et al. model [7]). This is a case of under estimation of CM, as more noise can be shaped without being noticed. On the other hand, free-energy principle based Wu's et al. method [6] cannot predict the high-frequency regions [9,12] and produce large prediction errors and subsequently inject a lot of noise in such regions, which can be easily perceived by our human vision. This is a condition of over-estimation of contrast masking, as shown in Fig. 8 (c). Our proposed algorithm, based upon the feed-back mechanism, can alleviate the problem of over- and under-estimation of the CM (as shown in Fig. 8. (d)) and estimate the maximum value of the CM. These experimental results confirm the need for the feed-back mechanism, in high frequency regions.

In Fig. 9, the efficiency of the overall proposed JND model (10) (which includes the effect of both the RMS–NAMM and feed-back mechanism) is shown. The 'Barbara' image with the same

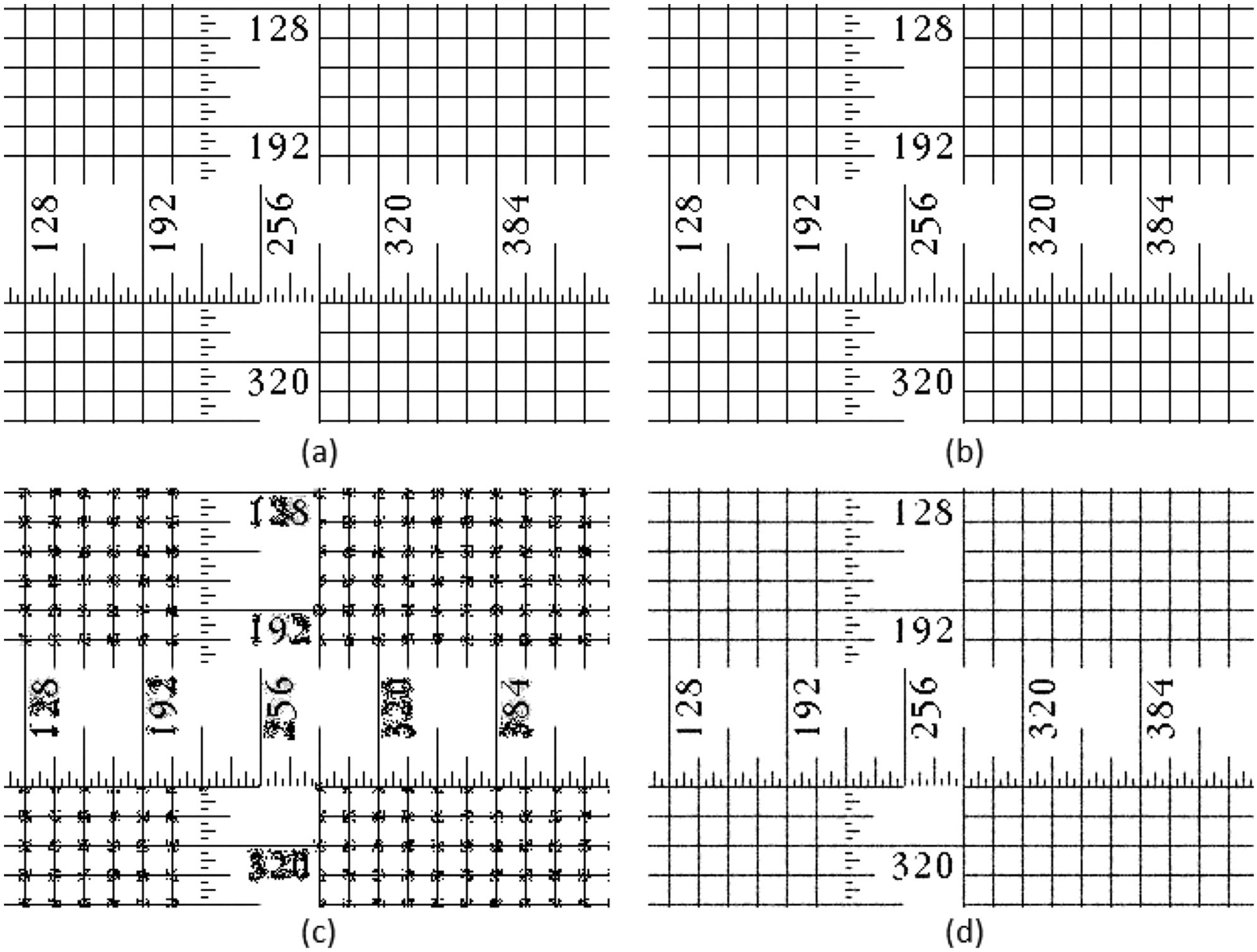

Fig. 8. Over and under estimation of the contrast masking for cropped 'Ruler' image. Noise injected images with (a) NAMM [8] (MSE 36.35), (b) Liu's et al. method [7] (MSE 45.77), (c) Wu's et al. method [6] (MSE 3486.94) and (d) proposed algorithm (MSE 290.25), respectively.

amount (equal MSE) of noise injected are shown in Fig. 9, and this noise injection is guided by the proposed and existing pixel domain JND [6–8] algorithms. It can be observed from Fig. 9 that the noise injected images guided by the existing JND algorithms look noisy in the smooth areas, specifically in the low luminance regions. During the subjective evaluation, we also asked subjects to give their reasons behind preferring one image over others. Some of the subjects during the on-line evaluation gave comments about noise in the smooth areas (one user commented, "Some images looks noisy." and another user said, "Preferred one is less noisy. Seems like the noise is more noticeable on flat and dark region"). Furthermore, the proposed algorithm also considers the effect of CS, and hence, the noise injected image corresponding to the proposed JND model is relatively noise-free and visually more appealing, as shown in Fig. 9 (d). The proposed algorithm uses the RMS contrast and feedback mechanism to efficiently estimate the JND value, and because of the RMS contrast, the proposed algorithm injects a slightly lower amount of noise into smooth areas (as shown in Fig. 7), while the maximum amount of noise is injected in the high frequency regions, which cannot be identified by the HV.

The above discussed and shown experimental results confirm the importance of CS and the need for the feed-back mechanism for estimating the efficient JND value in the pixel domain. From these experimental results, it can be concluded that the proposed JND model produces significantly better results as compared to the existing state-of-the-art pixel domain JND algorithms [6–8] and the proposed algorithm best matches with human perception.

4. Discussions

This work presents a new JND algorithm using the RMS contrast and feed-back mechanism. We feel, this work would be quite useful for the researchers working in the area of image quality assessment and water-marking, especially for quantitatively judging the quality of Screen Content (SC) images [32], as SC images contain a lot of rapidly varying edges (e.g. textual regions [32]) and in such situations algorithms in [6] and [7,8] over and under estimate the CM, respectively. These arguments can be validated by seeing the recent research on IQA of SC images [32,34], which suggests that free-energy principle based algorithms [25,33] can not predict image quality efficiently, due to the over-estimation of CM in textual regions. Our proposed algorithm has a few limitations that should be improved in the future, such as:

1. The performance of the proposed algorithm is nearly similar to the performance of Wu's et al. method [6] in high spatial frequency regions, when input image does not have any rapidly varying edges which result into no over-estimation of CM.
2. The proposed algorithm is specifically designed to estimate JND values for the natural images. In future, we will try to create a generic JND algorithm, which can suit the characteristics of the diverse variety of images (such as natural, depth, graphical, tone-mapped) to efficiently estimate the JND thresholds.

Fig. 9. Comparison of the proposed JND model with state-of-the-art pixel domain JND algorithms. These four rows show four respective noise injected, 'Barbara' images (MSE = 152.32), with (a) the NAMM [8], (b) Liu's et al. method [7], (c) Wu's et al. method [6] and (d) the proposed algorithm.

All of these requirements pose new challenges to the researchers working in the area of JND and IQA. In future, we will attempt to integrate proposed JND model with the IQA, especially for assessing the quality of SC and graphic images, as these areas are relatively less explored [32,34].

5. Conclusion

In this paper, we have presented the first pixel domain JND algorithm that includes a fundamental component of human vision, namely contrast sensitivity (via RMS contrast). Therefore, an oversight in the existing relevant models is addressed. In the proposed algorithm, we consider Luminance Adaptation, Contrast Masking and RMS contrast to efficiently estimate the JND value in the smooth regions. For edge and texture regions, we have also proposed a feed-back mechanism based upon the RMS contrast to efficiently control the CM, and a prediction based algorithm is used to classify an image into smooth, edge and texture regions. From experiments, it has been validated that the proposed algorithm efficiently matches with human perception and performs much better than the state-of-the-art pixel domain JND algorithms.

References

[1] P.J. Bex, S.G. Solomon, S.C. Dakin, Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure, J. Vis. 9.10 (2009) 1.
[2] V. Mante, et al., Independence of luminance and contrast in natural scenes and in the early visual system, Nat. Neurosci. 8 (2005) 1690–1697.
[3] R.L. Levien, Resolution-dependence of perceived contrast of textures, in: Proceedings of the Electronic Imaging 2004, International Society for Optics and Photonics, 2003.
[4] D.G. Pelli, P. Bex, Measuring contrast sensitivity, J. Vis. Res. 90 (2013) 10–14.
[5] A.M. Haun, E. Peli, Perceived contrast in complex images, J. Vis. 13.13 (2013) 3.
[6] J. Wu, et al., Just noticeable difference estimation for images with free-energy principle, IEEE Trans. Multimed. 7 (2013) 1705–1710.
[7] A. Liu, et al., Just noticeable difference for images with decomposition model for separating edge and textured regions, IEEE Trans. Circuits Syst. Video Technol. 11 (2010) 1648–1652.
[8] X.K. Yang, et al., Just noticeable distortion model and its applications in video coding, Sign. Process. Image Commun. 7 (2005) 662–680.
[9] V. Jakhetiya, et al., Fast and efficient intra-frame deinterlacing using observation model based bilateral filter, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 5819–5823.
[10] K.W. Hung, W.C. Siu, Fastimage interpolation using bilateral filter, IET Image Process. 7 (2012) 877–890.
[11] ITU Radiocommunication Assembly, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2003 International Telecommunication Union.
[12] V. Jakhetiya, W. Lin, S.P. Jaiswal, S.C. Guntuku, O.C. Au, Maximum a posterior and perceptually motivated reconstruction algorithm: a generic framework, IEEE Trans. Multimed. 19 (1) (2017) 93–106.
[13] S.W. Jung, J.Y. Jeong, S.J. Ko, Sharpness enhancement of stereo images using binocular just-noticeable difference, IEEE Trans. Image Process. 3 (2012) 1191–1199.
[14] C.H. Chou, K.C. Liu, Colour image compression based on the measure of just noticeable colour difference, IET Image Process. 6 (2008) 304–322.
[15] Y. Niu, et al., Combined just noticeable difference model guided image watermarking, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2010, pp. 1679–1684.
[16] Q. Feng, et al., Stereoscopic video quality assessment based on stereo just-noticeable difference model, in: Proceedings of the IEEE International Conference on Image Processing, 2013, pp. 34–38.
[17] S.A. Fezza, et al., Stereoscopic image quality metric based on local entropy and binocular just noticeable difference, in: Proceedings of the IEEE International Conference on Image Processing, 2014, pp. 2002–2006.
[18] H. Wang, X. Qian, G. Liu, Inter mode decision based on just noticeable difference profile, in: Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 297–300.
[19] M. Bouchakour, G. Jeannic, F. Autrusseau, JND mask adaptation for wavelet domain watermarking, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2008, pp. 201–204.
[20] C.H. Chou, Y.C. Li, A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile, IEEE Trans. Circuits Syst Video Technol. 6 (1995) 467–476.
[21] W. Lin, Y. Gai, A.A. Kassim, A study on perceptual impact of edge sharpness in images, IEE Proc. Vis. Image Signal Process. 2 (2) (2006) 215–223.
[22] X.H. Zhang, W.S. Lin, P. Xue, Improved estimation for just-noticeable visual distortion, Signal Process. 4 (2005) 795–808.
[23] M.P. Eckert, AP. Bradley, Perceptual quality metrics applied to still image compression, Signal Process. 3 (1998) 177–200.
[24] X. Zhang, W. Lin, P. Xue, Just-noticeable difference estimation with pixels in images, J. Vis. Commun. Image Represent. 1 (2008) 30–41.
[25] J. Wu, et al., Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimed. 7 (2013) 1700–1705.
[26] S. Wang, et al., Adaptive watermarking and tree structure based image quality estimation, IEEE Trans. Multimed. 2 (2014) 311–325.
[27] Y. Fang, et al., Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum, IEEE Trans. Multimed. 1 (2012) 187–198.
[28] S.W. Jung, S.J. Ko, Depth sensation enhancement using the just noticeable depth difference, IEEE Trans. Image Process. 8 (2012) 3624–3637.
[29] J. Lei, et al., Depth sensation enhancement for multiple virtual view rendering, IEEE Trans. Multimed. 4 (2015) 457–469.
[30] J. Shen, et al., Depth-aware image seam carving, IEEE Trans. Cybern. 5 (2013) 1453–1461.
[31] T.S. Huang, et al., Enhancement of backlight-scaled images, IEEE Trans. Image Process. 12 (2013) 4587–4597.
[32] H. Yang, Y. Fang, W. Lin, Perceptual quality assessment of screen content images, IEEE Trans. Image Process. 11 (2015) 4408–4421.
[33] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 1 (2013) 43–54.
[34] J. Qian, L. Tang, V. Jakhetiya, Z. Xia, K. Gu, H. Lu, Towards efficient blind quality evaluation of screen content images based on edge-preserving filter, Electronics Letters 53 (9) (2017) 592–594.

Vinit Jakhetiya received the B.Tech. degree in computer and communication engineering from LNM Institute of Information Technology, India, in 2011, and the Ph.D. degree in electronics and computer engineering from Hong Kong University of Science and Technology, Hong Kong, in 2016. From January 2015 to December 2015, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, and later he joined as a Project Officer at the same university. Currently, he is working as an assistant professor, Bennett University, India. His research interests include image/video processing, image quality assessment, and visual perceptual modeling.

Weisi Lin received the Ph.D. degree from Kings College, London University, London, U.K., in 1993. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, and served as a Lab Head of Visual Processing, Institute for Infocomm Research. He authors over 300 scholarly publications, holds 7 patents, and receives over 4 million dollars in research grant funding. He has maintained active long-term working relationship with a number of companies. His research interests include image processing, video compression, perceptual visual and audio modeling, computer vision, and multimedia communication. He served as an Associate Editor of IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, IEEE Signal Processing Letters, and Journal of Visual Communication and Image Representation. He is also on six IEEE Technical Committees and Technical Program Committees of a number of international conferences. He was the Lead Guest Editor for a special issue on perceptual signal processing of the IEEE Journal Of Selected Topics in Signal Processing in 2012. He is a Chartered Engineer in the U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He Co-Chaired the IEEE Multimedia Communications Technical Committee (MMTC) special interest group on quality of experience. He was an Elected Distinguished Lecturer of APSIPA in 2012.

Sunil P. Jaiswal received the B. Tech degree from the LNM Institute of Information Technology (LNMIIT), Jaipur, India, in 2012. He is currently pursuing the Ph.D. degree in electronic and computer engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong since 2012. Currently, he is a research internee in Technicolor R & D, Rennes, France. During his Ph.D. degree, he also used to be a visiting scholar in the visual analytics and imaging lab in Stony Brook University, NY, USA, and SUNY Korea in 2015/2016. He coauthored the best paper award in student category in I2MTC'2011. His research interest includes multispectral imaging, 3-D medical imaging, Image quality assessment, and Deep Learning.

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is the reviewer for IEEE T-IP, T-CSVT, T-MM, T-CYB, SPL, Neurocomputing, SPIC, JVCI, SIViP, IET-IP, etc. From July to November 2014, he was a visiting student at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. From December 2014 to January 2015, he was a visiting student at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include quality assessment, visual saliency detection and contrast enhancement.

Sharath Chandra Guntuku received B.E.(Hons.) in computer science from Birla Institute of Technology and Science—Pilani, India, in 2013, and Ph.D. degree in computer science at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include personality computing, user modeling, and multimedia analysis.