J. Vis. Commun. Image R. 40 (2016) 600–610

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Perceptually-lossless image coding based on foveated-JND and H.265/HEVC ☆

Yiming Li, Hongyi Liu, Zhenzhong Chen *

School of Remote Sensing and Information Engineering, Wuhan University, Luoyu Road 129, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 6 January 2016
Revised 29 July 2016
Accepted 29 July 2016
Available online 16 August 2016

Keywords:
H.265/HEVC
Foveated JND
Lossless
Image coding
Perceptual coding

ABSTRACT

Removing perceptual redundancy plays an important role in image compression. In this paper we develop a foveated just-noticeable-difference (FJND) model to quantify the perceptual redundancy in the image and integrate it in the H.265/HEVC intra encoding framework to provide a perceptually lossless image coding solution. Different to the conventional JND models, our proposed FJND model considers the relationship between contrast masking effect and the foveation properties of HVS. Furthermore, to achieving the perceptually lossless coding, the FJND model is integrated in the H.265/HEVC framework by determining the quantization parameter to ensure that the resulting distortion is no larger than the FJND threshold. The experiments demonstrate that the proposed method effectively improves the compression performance.

1. Introduction

How to remove redundancy in the image for effective image compression has played an important issue in image applications. JPEG [1] or JPEG2000 [2] have been developed for efficient image compression. However, lossless or near lossless images are expected in many applications, such as remote sensing image, medical image, etc. Some lossless image compression algorithms have been developed, such as JPEG_LS [3] or JPEG2000 [2,4]. Since in some applications people just need perceptually lossless images [5,6], removing perceptual redundancy in the image to provide perceptually lossless image coding is desired that can further improve the efficiency of the compression.

In this paper, we propose a perceptually lossless image compression method based on H.265/HEVC intra coding technique and the foveated Just-Noticeable-Difference (FJND) model. H.265/HEVC, which means high efficiency video coding, was jointly developed by ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) as ISO/IEC 23008-2 MEPG-H Part 2 and ITU-T H.265 [7]. As the latest video compression standard, compared with H.264/AVC [8], H.265/HEVC can improve the compression performance up to 50% in compression ratio whilst achieving the same perceptual quality [9,10]. It can not only compress videos but also images efficiently [11]. As during the standardization, H.265/HEVC has adopted some effective coding tools for efficient image compression [7]. So we utilize its intra coding functionality as our still image coding basis. With this efficient coding framework, we aim to integrate a perceptual model which can effectively quantify the perceptual redundancy in the image therefore we can use it in perceptually lossless image compression.

During these past few years, several perceptual redundancy quantification models have been proposed with the better understanding of HVS such as JND model for visual sensitivity [12]. Foveated JND model has been further developed by designing a JND model with the considerations on human's visual foveation characteristics, while the conventional JND measures the visibility threshold under the assumption that visual acuity remains constant over the whole image [13–18]. It is well-known that HVS is not only related to spatio-temporal frequency response, but is also space-variant. To better exploit the space-variant properties of HVS, Chen et al. [19] proposed a foveated JND model based on psycho-visual experiments. Due to the nonuniform density of photoreceptor cells on retina, the visual acuity for pixel decreases as the distance between the pixel itself and fixation point increases, therefore leads to higher JND threshold. The connection between visibility and retinal eccentricity has been examined by combining a foveation model with spatial and temporal JND models [19].

Taking into consideration the fact that the FJND model in [19] is developed based on the luminance masking effect of HVS, while both luminance and contrast masking effects are important HVS characteristics [20], we aim to further exploit the foveation

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: zzchen@whu.edu.cn (Z. Chen).

http://dx.doi.org/10.1016/j.jvcir.2016.07.025
1047-3203/

property of contrast masking effect due to the increased retinal eccentricity. We have conducted psycho-visual experiments to model the contrast masking effect with different retinal eccentricity. Given the fixation point, a square noise area, where noises of same amplitude exists, is randomly displayed on the screen based on predefined eccentricity, background, and contrast. The visual threshold is acquired by changing the amplitude of noise using staircase method. Based on the experiments, we then propose our foveated contrast masking effect model, which measures the visibility threshold of the contrast masking effect with corresponding retinal eccentricity. We also improve a new spatial JND model with the proposed model, and compare the performance of different models. Moreover, we propose to apply the foveated JND model to describe the smallest detectable difference of the image. Then, we propose a perceptually lossless image coder based on H.265/HEVC and FJND to compress the image. If we guarantee the maximum distortion in the compressed image is smaller than the detectable difference by the FJND model provided, we can obtain the perceptually lossless image.

The rest of the paper is organized as follows. In Section 2, we introduce the psycho-visual experiments to model the contrast masking effect as well as the foveation effect, and build the foveated contrast masking effect based JND model. Integrating the foveated JND model in H.265/HEVC for perceptually lossless image coding is presented in Section 3. The experiments are shown in Section 4. Finally, Section 5 concludes the paper.

2. Foveated contrast masking model

2.1. Background information

The determination of visibility threshold of images involves many factors such as contrast masking effects, luminance masking effects, overlapping effects and so on [13,14,21,22]. It is suggested that visibility threshold increases with increased contrast [23]. Due to the non-uniform distribution of photoreceptor cells on human retina, the human visual system has different visual acuity to perceive image details. According to [19], when the visual stimulus is projected on the fovea, where exists highest density of senor cones, it is perceived at the highest resolution. The density of cone and retinal ganglion cells decreased gradually as the eccentricity increases, which results in descending visual acuity. Thus, a contrast masking model incorporating with the foveation model will lead to a more precise estimation of the capability of HVS to discriminate image differences.

2.2. Modeling contrast masking and foveation effect

To explore the relationship between contrast masking effect and foveation effect, we have conducted the experiments in laboratory to obtain the visibility threshold for further analysis. In our experiments, we used a DELL U3014 30in LCD monitor with resolution of 2560×1600 to display test images. Due to the fact that earlier JND models are based on experimental results using the CRT monitor, whose Gamma characteristics differ from that of the LCD monitor, we used the corresponding mapping function (same as that in [24]) to calibrate digital luminance of different monitors. The calculated noise threshold (in digital form) was first mapped to physical luminance, and then converted to the digital luminance level of the LCD display.

Based on the discussion in subSection 2.1, we have designed an image pattern to incorporate the contrast masking effect with the retina eccentricity. During the experiments, a test image generated using the predefined contrast *eh*, background *bg* and eccentricity *e*, was shown on the screen. As shown in Fig. 1, the test image

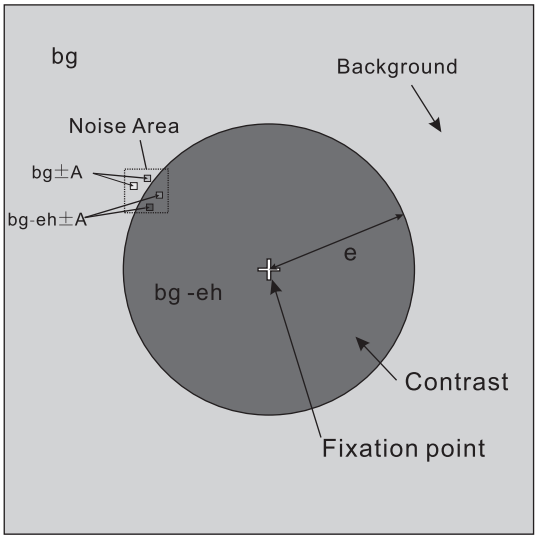

Fig. 1. Image pattern used in experiments exploring the contrast masking effect.

contained a circular area whose radius was calculated as the product of viewing distance and eccentricity. The luminance of outside and inside area was respectively set as bg and $bg - eh$, so that the test image maintained a constant contrast on the boundary. A square of 30×30 pixels, in which existed noise of same amplitude A, was randomly displayed on the boundary. Therefore, in the square, the luminance of pixels outside boundary was either $bg + A$ or $bg - A$, while that of pixels inside boundary was either $bg - eh + A$ or $bg - eh - A$. In both cases the luminance was also bounded with maximum and minimum luminance values, i.e., 0 and 255. The center of the image was intentionally designed as the fixation point by setting its luminance as 0 or 255 for the purpose of helping subjects to fix their attention.

At the beginning of the test, the noise amplitude was set as 0, which means the noise area was invisible to subjects. With their attention fixed at the image center, subjects were first asked to increase the noise amplitude by 1 each time until the noise area became visible and the corresponding amplitude was recorded as A_1. The amplitude was subsequently adjusted to $A_1 + \Delta A$, which was slightly higher than A_1. Then subjects conducted a reversed procedure that they decreased the amplitude from $A_1 + \Delta A$ until the noise became invisible and the corresponding amplitude was recoded as A_2. The procedure, as shown in Fig. 2, could be repeated multiple times to obtain more sample results A_i. The final noise threshold was then calculated as the average of A_i.

The visibility thresholds according to the different eccentricity and contrast are shown in Fig. 3. With the contrast increasing, the visibility threshold increases. This could be explained by conventional Spatial JND [13], which explores a linear relationship between the visibility threshold and contrast. Moreover, the visibility threshold increases rapidly as the eccentricity increases.

2.3. Foveated contrast masking model

We aim to build a model to estimate the relationship between the contrast masking effect and foveation effect. A spatial JND

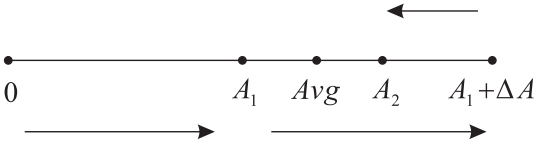

Fig. 2. The staircase method. *Avg* denotes the average of A_i.

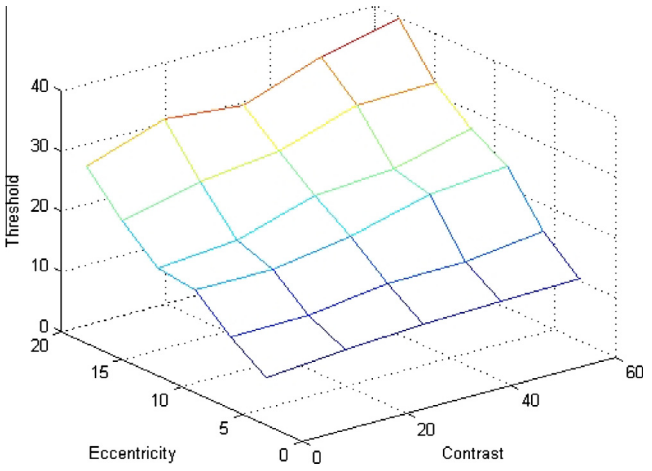

Fig. 3. Experimental results based on the contrast and eccentricity.

model [22], which considers the luminance adaption, spatial masking, and disorderly concealment effect, is adopted in this paper. The contrast masking effect is described in [22] as:

$$f_1(x,y) = [0.01B(x,y) + 11.5][0.01G(x,y) - 1] - 12 \tag{1}$$

where f_1 estimates the spatial masking effect, $B(x,y)$ and $G(x,y)$ represent background and contrast of point (x,y).

Given the fixation point (x_f, y_f) and viewing distance v, the retinal eccentricity of a certain point $P(x,y)$ can be calculated as:

$$e(v,P) = \tan^{-1}\left(\frac{d}{v}\right) \tag{2}$$

where $d = \sqrt{(x-x_f)^2 + (y-y_f)^2}$.

Table 1
Fitting precision.

RMSE	R-square	Adjusted R-square
0.4332	0.981	0.981

An analytical model was proposed in [25] to measure the contrast sensitivity as a function of eccentricity, which can be defined as:

$$CS(f,e) = \frac{1}{CT(f,e)} \tag{3}$$

where $CT(f,e) = CT_0 \exp(\chi \cdot f \cdot \frac{e+e_2}{e_2})$. $CS(f,e)$ and $CT(f,e)$ denote contrast sensitivity and contrast threshold, respectively. f is spatial frequency (cycles/degree), e is the retinal eccentricity (degree). The parameters CT_0, χ and e_2 are the minimum contrast threshold, the spatial frequency decay constant, and the half-resolution eccentricity constant, respectively. These quantities are model parameters set to $CT_0 = \frac{1}{64}, \chi = 0.106$ and $e_2 = 2.3$.

The cutoff frequency f_c can be obtained by setting CT to 1.0:

$$f_c(e) = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\chi(e+e_2)} \tag{4}$$

In [26], the display cutoff frequency should be half of the display resolution r:

$$f_d(v) = \frac{r}{2} \approx \frac{1}{2} \cdot \frac{\pi v}{180} \tag{5}$$

Combining $f_c(e)$ and $f_d(v)$, the cutoff frequency for any given location P can be refined as:

$$f_m(v,P) = \min(f_c(e(v,P)), f_d(v)) \tag{6}$$

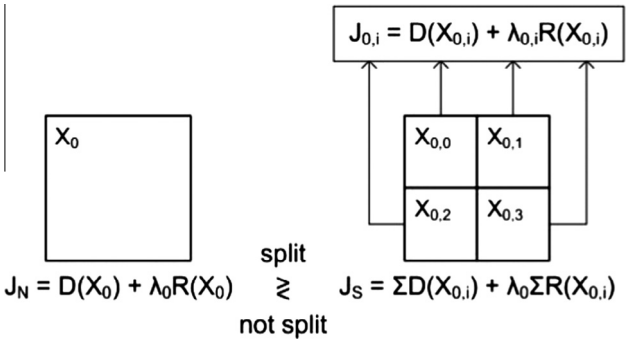

Fig. 5. The split flow by RDO [27].

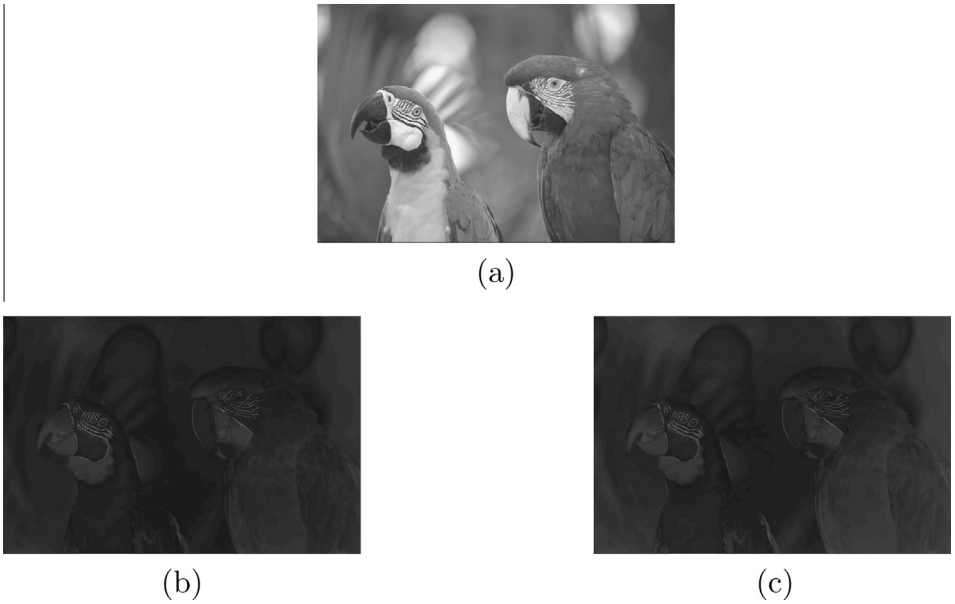

Fig. 4. (a) Original image. (b) Map of SJND (×5 for display). (c) Map of FJND (×5 for display).

So we obtained the normalized contrast sensitivity:

$$S_f = \frac{f_m(v,P)}{f_m(v,0)} \tag{7}$$

Finally, we model the contrast masking effect and foveation effect based on S_f and contrast *eh*.

For each threshold obtained from previous experiments, we first calculate the corresponding f_1 based on its background and contrast. In order to measure the foveation effect, we then compute the ratio between experimental threshold and its corresponding f_1. For simplicity, the ratios of different backgrounds are averaged to acquire mean ratios. The mean ratios are then fitted as a function of contrast and eccentricity. Thus, the foveation model of contrast masking effect can be defined as:

$$m_1(S_f, eh) = \left(\frac{1+a}{S_f+a}\right)^{\eta(eh)} \tag{8}$$

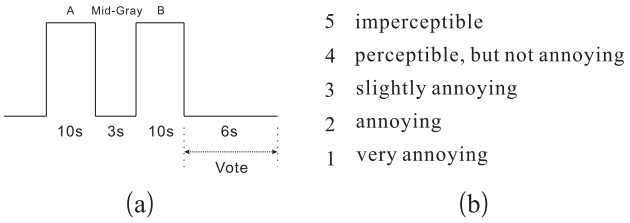

Fig. 7. (a) DSIS presentation structure. (b) Grading scale.

$$\eta(eh) = \left(\log\left(\frac{eh}{255}+1\right)+b\right)^2 + c \tag{9}$$

where S_f denotes the normalized contrast sensitivity, and *eh* measures the contrast of each pixel in the image. According to the fitting results, a, b and c have been set to 0.05, 1.5 and 0.66, respectively. The fitting precision is presented in Table 1.

Fig. 6. The test picture set.

Table 2
Subjective scores for noise-injected images.

Image	PSNR			Score	
	Chen's [19]	Wu's [22]	Ours	Avg	DP
1	34.59	32.61	31.31	5.00	0.00
2	34.28	34.2	32.23	4.76	0.18
3	35.39	35.69	33.51	4.71	0.18
4	35.31	34.97	33.18	4.82	0.12
5	32.71	32.27	30.69	5.00	0.00
6	35.31	34.05	32.16	4.88	0.12
7	36.05	35.96	33.64	4.94	0.06
8	32.78	32.75	31.04	4.71	0.29
9	37.15	36.9	34.84	4.82	0.18
10	36.55	36.18	34.13	4.88	0.12
11	34.21	33.46	32.01	4.94	0.06
12	36.59	36.05	33.67	4.82	0.12
13	33.91	29.77	28.93	4.94	0.06
14	33.82	32.94	31.45	4.94	0.06
15	32.71	33.32	31.45	4.71	0.24
16	35.4	35.07	32.99	5.00	0.00
17	32.59	33.25	31.65	4.94	0.06
18	32.88	31.85	30.64	4.94	0.06
19	34.93	34.57	32.95	4.94	0.06
20	34.6	34.42	32.51	4.94	0.06
21	35.66	33.94	32.29	4.76	0.24
22	35.9	34.33	32.79	4.82	0.12
23	35.18	35.98	33.53	4.82	0.18
24	34.41	32.89	31.45	5.00	0.00

Based on the foveation model describing the contrast masking effect and adopted JND model, we develop the foveated contrast masking effect model as:

$$F_1 = f_1 \cdot m_1 \tag{10}$$

where f_1 measures the contrast masking effect, as defined in (1), m_1 is the proposed foveation model, which calculates the foveation effect of contrast masking. The proposed model measures the perceptible distortion threshold when human visual attention is fixed at a certain location of image.

2.4. Foveated just noticeable difference model

Since the Spatial JND model consists of not only contrast masking, but also luminance masking and other masking effects, We aim to integrate our foveated contrast masking model to the JND model [22], and validate its effectiveness. The adopted model computes the JND value in orderly and disorderly parts, which are estimated by the nonlinear additivity model for masking (NAMM) [14] procedure and disorderly concealment effect (DCE) procedure, respectively. For luminance masking, we adopt Chen's foveated model [19] to estimate the foveation effect. The foveated luminance masking effect can be measured as:

$$F_2 = f_2 \cdot m_2 \tag{11}$$

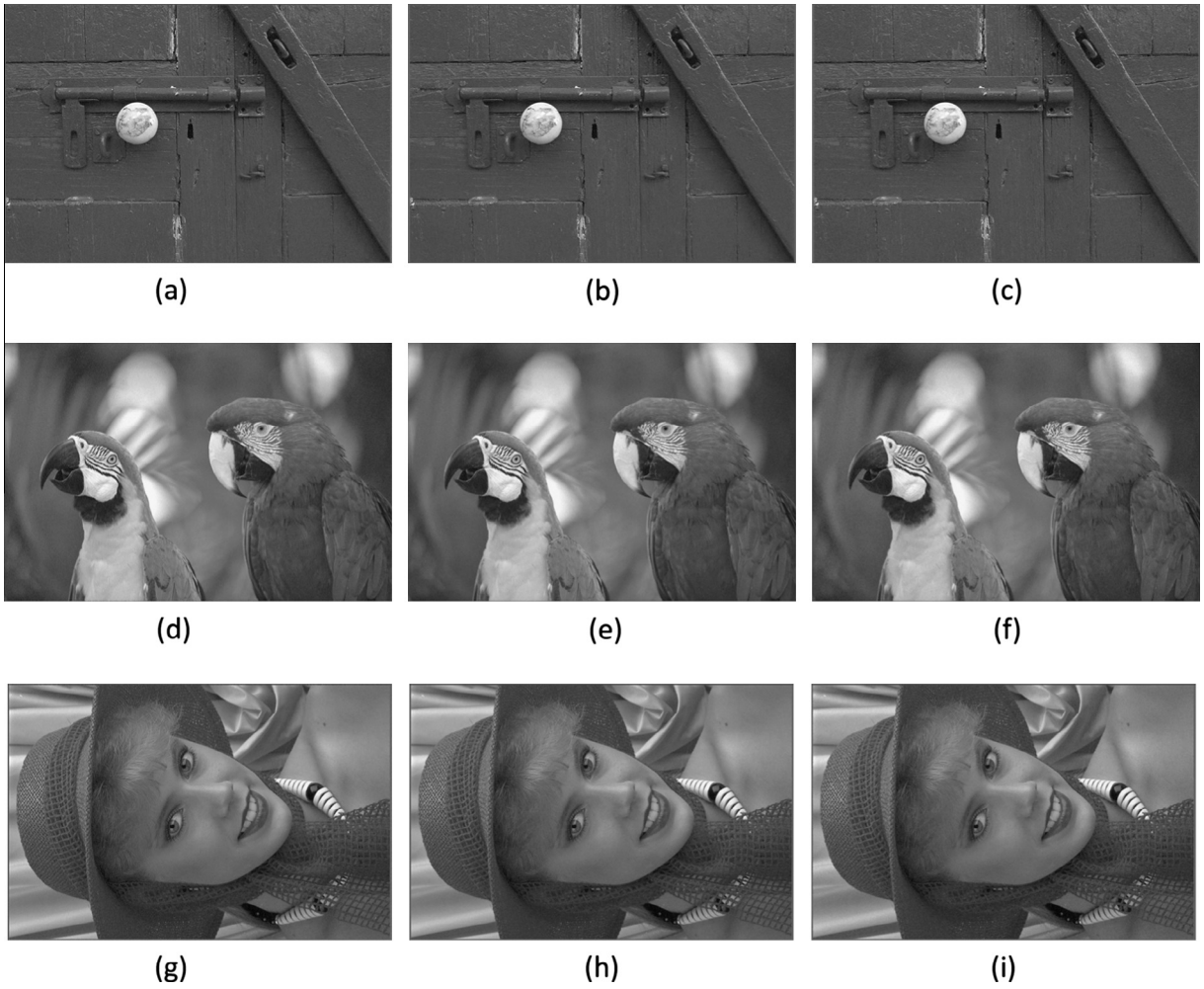

Fig. 8. Comparisons of result images with noise of different models. (a), (d) and (g) are from Chen's model, (b), (e) and (h) are from Wu's model, and (c), (f) and (i) are from our model.

where f_2 measures the luminance masking effect and m_2 is the coefficient that estimates the foveation effect in [19]. Taking the foveation effect into consideration, the JND for orderly parts of the image is refined as:

$$FJND_p = F_1 + F_2 - C_p^{gr} \cdot \min\{F_1, F_2\} \quad (12)$$

Combining the JND for disorderly parts of the image, the final JND can be refined as:

$$FJND = FJND_p + JND_d - C_f^{gr} \cdot \min\{FJND_p, JND_d\} \quad (13)$$

where JND_d represents the disorderly concealment effect, which is defined in [22]. C_p^{gr} and C_f^{gr} are gain reduction parameters that measures the overlapping effects, and are set as 0.8 empirically.

As shown in Fig. 4, we compare the proposed foveated JND model with the conventional spatial JND model to directly display the difference of application performance between these two methods. Compared with the original image (Fig. 4(a)), both images contaminated with noise presents no visual difference when visual attention is fixed at the calculated visual attention point (x_f, y_f). According to the noise distribution map of the proposed model (Fig. 4(c)), the periphery areas with contrast pattern (such as roof, pillar, trees) are injected with noise of higher amplitudes. This strategy of injecting noise coincides well with the foveation effect of human visual system. The conventional spatial JND, as presented in Fig. 4(b), however, does not take the decrease of visual acuity into consideration, which is improved in our model.

3. Application of the FJND model in H.265/HEVC image coding

3.1. Quantifying distortion visibility by foveated JND

After we have developed the modified foveated JND model which measures the smallest detectable difference between two signals therefore it can be utilized to quantify the perceivable distortion in the noise contaminated images. Since the purpose of image coding is to achieve the lowest bit rate whilst guarantee the desired subjective quality according to our HVS [14,27], foveated JND thresholds could hence be used to determine optimum quantization step sizes for different parts of image for perceptually-lossless image coding.

From Fig. 4, we can see that foveated JND map may be lighter than conventional JND map in some area which means it can tolerate more distortion. As mentioned above in Section 2, foveated JND enhances the SJND model by accounting for the relationship between visibility and eccentricity and can better exploit the perceptual redundancy in the area of less visual significance to observers. Thus we may have the consideration that if we use foveated JND model to guide our compression rather than SJND model or even no perceptual information we may save more bits without perceptual lossless.

3.2. Rate-distortion optimization

In H.265/HEVC intra coding, the image is divided into blocks which range from 64×64 to 8×8. The encoder adaptively choose different depth to split the units, which ranges from 0 to 3. In addition, if the depth is 3, the CU block can continually be divided into 4 blocks and these are the smallest size of PU, which size are all 4×4 pixels. When the image is divided into 64×64 size of unit which depth is 0 initially, the encoder choose the rough mode decision (RMD) and most probable mode (MPM) first. In the chosen RMD + MPM list (3 + MPM or 8 + MPM), each element indicates one of the probable prediction mode and the encoder choose the best mode from them by rate-distortion optimization (RDO). Then, it pre-split the unit into four 32×32. In each split block, it choose the best prediction mode again and further split. These aforementioned steps repeat up to the block size is divided into 4×4 pixels. To reckon if these block should be split like these pre-split mode or use the non-split mode, the encoder should also calculate the rate-distortion (R-D) cost of split/non-spilt mode, by rate distortion optimization (RDO), the encoder choose whether split or not at this prediction mode. The steps are shown in Fig. 5. The proposed

Table 3
Experimental results.

	bpp (bits per pixel)				PSNR_Y (dB)				Time (s)		Bpp saving			Mean
	Chen's [19] model result	Wu's [22] model result	Our model result	HM lossless	Chen's model result	Wu's model result	Our result	HM lossless	Proposed algorithm	HM origin	Vs Chen's model (%)	Vs Wu's model (%)	Vs HM lossless (%)	Comparison scale of proposed
1	2.44	1.93	1.70	6.53	40.9179	38.5583	37.3782	–	12.74	5.568	30.29	11.94	73.97	0.00
2	0.95	0.91	0.77	5.29	40.5584	40.2706	39.0884	–	10.166	5.011	19.46	15.23	85.49	0.00
3	0.84	0.73	0.59	4.26	43.103	42.5605	41.5676	–	10.663	4.583	29.81	19.43	86.17	0.00
4	1.24	1.14	0.92	5.31	40.9923	40.8598	39.4786	–	11.991	5.354	25.48	19.27	82.60	0.00
5	1.88	1.74	1.49	6.87	39.1224	38.3081	36.9993	–	12.896	6.297	20.67	14.11	78.28	0.00
6	1.78	1.47	1.28	5.89	41.4893	39.9061	38.6725	–	12.355	5.485	28.22	13.07	78.30	0.00
7	0.93	0.87	0.68	4.71	42.5728	42.2779	40.8608	–	11.115	4.49	27.19	22.10	85.58	0.00
8	2.29	1.97	1.68	6.85	39.9796	38.6923	37.3596	–	13.812	5.488	26.41	14.81	75.46	0.00
9	1.36	1.25	0.75	4.87	43.9941	42.8808	41.1714	–	10.952	4.329	44.91	40.12	84.57	0.00
10	1.30	1.21	0.84	4.95	42.9721	42.2515	40.6466	–	10.718	4.622	35.25	30.11	82.99	0.00
11	1.72	1.41	1.11	5.62	41.5535	39.9241	38.7802	–	12.739	5.24	35.12	21.19	80.20	0.00
12	1.19	0.94	0.70	4.75	43.3071	41.8086	40.5574	–	11.601	4.854	40.85	25.01	85.22	0.00
13	2.90	2.25	2.04	7.70	39.4216	36.2659	35.3252	–	14.075	6.67	29.78	9.55	73.54	0.00
14	2.04	1.69	1.43	6.45	40.2015	38.9605	37.7493	–	13.053	5.958	29.79	15.42	77.83	0.00
15	0.88	0.81	0.71	4.85	40.6888	40.3711	39.4944	–	11.091	4.581	19.10	12.45	85.31	0.00
16	1.29	1.07	0.88	5.16	41.4775	40.7481	39.4733	–	11.251	4.346	32.10	17.49	82.97	0.00
17	1.26	1.12	0.94	5.14	41.7508	40.9613	39.9394	–	10.214	4.799	24.98	15.86	81.69	0.00
18	1.67	1.54	1.38	6.77	39.0871	38.126	37.2143	–	14.162	5.532	16.98	10.02	79.58	0.00
19	1.65	1.42	1.09	5.70	41.2674	40.3806	38.7646	–	14.435	4.725	34.05	23.20	80.93	0.00
20	0.84	0.78	0.65	4.00	42.1398	41.7303	40.3419	–	11.126	4.512	22.23	16.03	83.70	0.00
21	1.73	1.55	1.11	5.77	41.3868	39.8027	38.3495	–	13.281	5.462	35.67	28.15	80.76	0.00
22	1.77	1.47	1.20	6.15	41.4172	39.9894	38.8779	–	12.468	5.623	32.30	18.39	80.53	0.00
23	0.85	0.71	0.54	4.63	43.4026	42.6472	41.5702	–	10.378	3.984	36.77	24.74	88.43	0.00
24	2.08	1.60	1.39	5.96	41.9178	39.4625	38.2049	–	12.912	5.404	33.21	13.30	76.74	0.00
Average											29.61	18.79	81.29	

perceptually lossless image coding is developed based on this framework. We assume human cannot perceive the distortion below the foveated JND threshold. Therefore, we calculate the FJND for each pixel in each block and determine the quantization parameter by ensuring that the resulting distortion is no larger than the FJND threshold. As (14) shown, if we obtain the higher QP value whilst satisfying the distortion constrains we can reduce more bits but still get the perceptually lossless image.

$$\max\{QP\}\ s.t.\ \begin{cases} SSE_{(i,j,N)} \leqslant FJND^2_{(i,j,N)} \\ \sum_{j=1}^{N}\sum_{i=1}^{N}[count(|y_{(i,j)} - y'_{(i,j)}| - FJND_{(i,j)}) > 0] \leqslant T \end{cases} \tag{14}$$

where T is in terms of the threshold of different block size. After performing the experiment, the value of T can be limited into a suitable domain as [1,4]. If the CU size is 64×64, we may assume T is 4. SSE means sum of squared errors of the block, which can be calculated as follows.

$$SSE = \sum_{j=1}^{N}\sum_{i=1}^{N}(y_{(i,j)} - y'_{(i,j)})^2 \tag{15}$$

We use a binary-search algorithm to find the optimal QP value and distortion. The distortion is no larger than the value provided by the foveated JND model at the same position, so that we can get both minimum bit rate and conform the foveated JND model. The algorithm is shown in Algorithm 1.

Algorithm 1. *QP* calculation

Input:
original image: I;
Foveated JND map of the image: S;
parameter to control threshold:
$begQP, midQP, lastQP, Count(i,j,N), T$;
Output:
final QP value for each $CU(i,j,N) : finalQP(i,j,N)$
set parameter
$begQP \leftarrow$ *original HEVC QP value*
set *lastQP*
for each $CU(i,j,N)$ in I **do**
  $FJND(i,j,N)^2 \leftarrow$ sum of squared value of pixels in (i,j,N) block in S
  set $Count(i,j,N) \leftarrow$ count of $D(i,j)$ larger than $FJND(i,j)$ in (i,j,N) block in S
  while $begQP \leqslant lastQP$ **do**
    set $midQP = \frac{begQP+lastQP}{2}$
    get the distortion $D(i,j,,N,midQP)$ of this $CU(i,j,N)$ in this QP by H.265/HEVC framework
    if $(D(i,j,N,midQP) == FJND(i,j,N)^2) or (Count(i,j,N) == T)$ **then**
      break;
    end if
    if $(D(i,j,N,midQP) < FJND(i,j,N)^2) and (Count(i,j,N) < T)$ **then**
      set $begQP = midQP + 1$
    else
      set $midQP = midQP - 1$
      set $lastQP = midQP$
    endif
  endwhile
  $finalQP(i,j) \leftarrow midQP$
end for

We do these operations in each depth which range from 0 to 3. We notice that there are 2 kinds of RDO process in these steps, one is RDO for mode decision and the other one is RDO for split/non-split decision. They all need be modified since our proposed optimization problem is different.

As we know, RDO in H.265/HEVC considers the optimization problem that minimize distortion D and subject to rate R smaller than R_{target}. While our purpose is to minimize rate R and subject to D smaller than *FJND*. Thus different from original H.265/HEVC RDO formula as (16) shown, λ in our modified RDO formula is a multiplier of D.

$$\min\{J\} \quad J = D + \lambda \cdot R \tag{16}$$

In addition, in the original rate-distortion optimization, the component D is described by SSE or SAD, it cannot consider the perceptual sensitivity of each CTU and cannot reflect the distortion by FJND model. Moreover, original λ is calculated by initial QP, thus an optimal λ_{opt} which is different from the initial one is also need be recalculated. The ideal perceptual rate-distortion optimization should be modified as (17).

$$\min\{J\} \quad J = R + \lambda_{opt} \cdot D_{FJND} \tag{17}$$

where D_{FJND} can be calculated as (18).

$$D_{FJND} = \max(|y_{(i,j)} - y'_{(i,j)}| - FJND_{(i,j)}, 0) \tag{18}$$

Since D_{FJND} is always equal or close to zero by our aforementioned process. Thus, the proposed perceptual based rate-distortion cost can be represented as follow finally.

$$\min\{J\} \quad J = R \tag{19}$$

So that we can notice that the mode decision strategy to choose the best mode or to determine the best CU split could just by rate instead of the original RDO that we can get the best balance between the perceptual distortion and bitrate.

4. Experiments

4.1. Validation tests for FJND

To evaluate the performance of the proposed contrast masking model, subjective validation tests have been conducted. The tests were performed in the laboratory with normal lighting. The display systems consisted of DELL U2412M 24 in. LCD monitors with resolution of 1920×1200. The viewing distance was approximately three times of the image width. Twenty-four images from Kodak Lossless True Color Image Suite [28] were chosen for the validation tests, as Fig. 6 shown. The true color images were firstly converted to grayscale images for further use. The validation tests were based on the double-stimulus impairment scale (DSIS) method recommended in ITU-R BT.500 [29], in which original images were used as references. The presentation structure is shown in Fig. 7(a), one stimulus image (Image A) is the original image and the other (Image B) is the distorted image. Assessors were first presented with unimpaired reference, then with the same picture impaired. The duration of each image was 10 s, and the break between two images was 3 s. If the difference between image A and image B was imperceptible, they should score 5, otherwise, referring the grading scale listed in Fig. 7(b). We adopted the bipolar pseudo-random noise injection strategy, i.e., each pixel in the image was randomly added or subtracted with noises with the maximum tolerable amplitude indicated by the propose model F_1. 17 subjects (with or corrected to normal vision) participated in the tests.

At the end of the validation tests, the mean score for each test image was calculated. As recommended in ITU-R BT.500, β_2 test was conducted to determine whether a subject should be rejected.

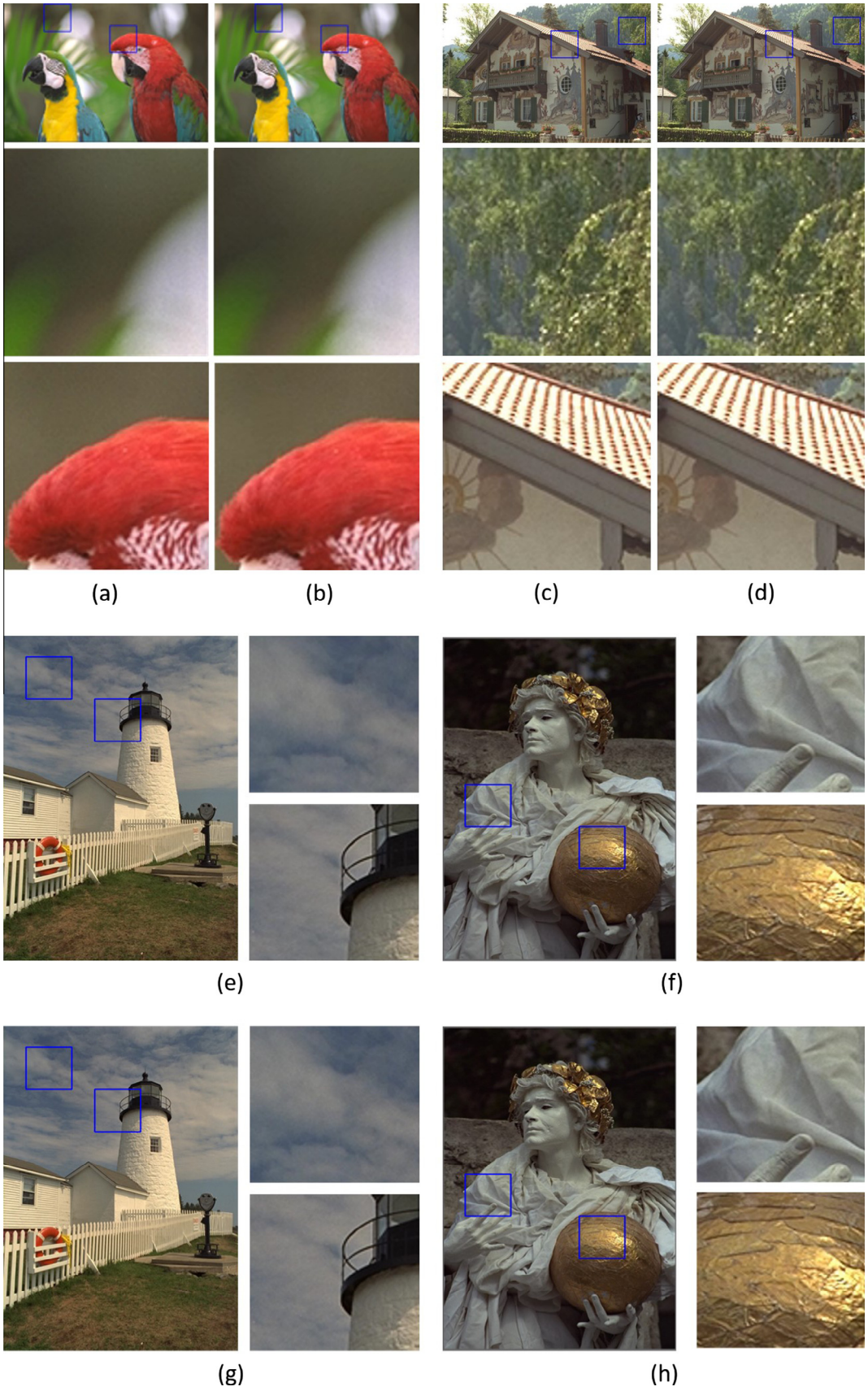

Fig. 9. The comparisons. (a), (c), (e) and (f) are HM lossless results and (b), (d), (g) and (h) are results of the proposed method.

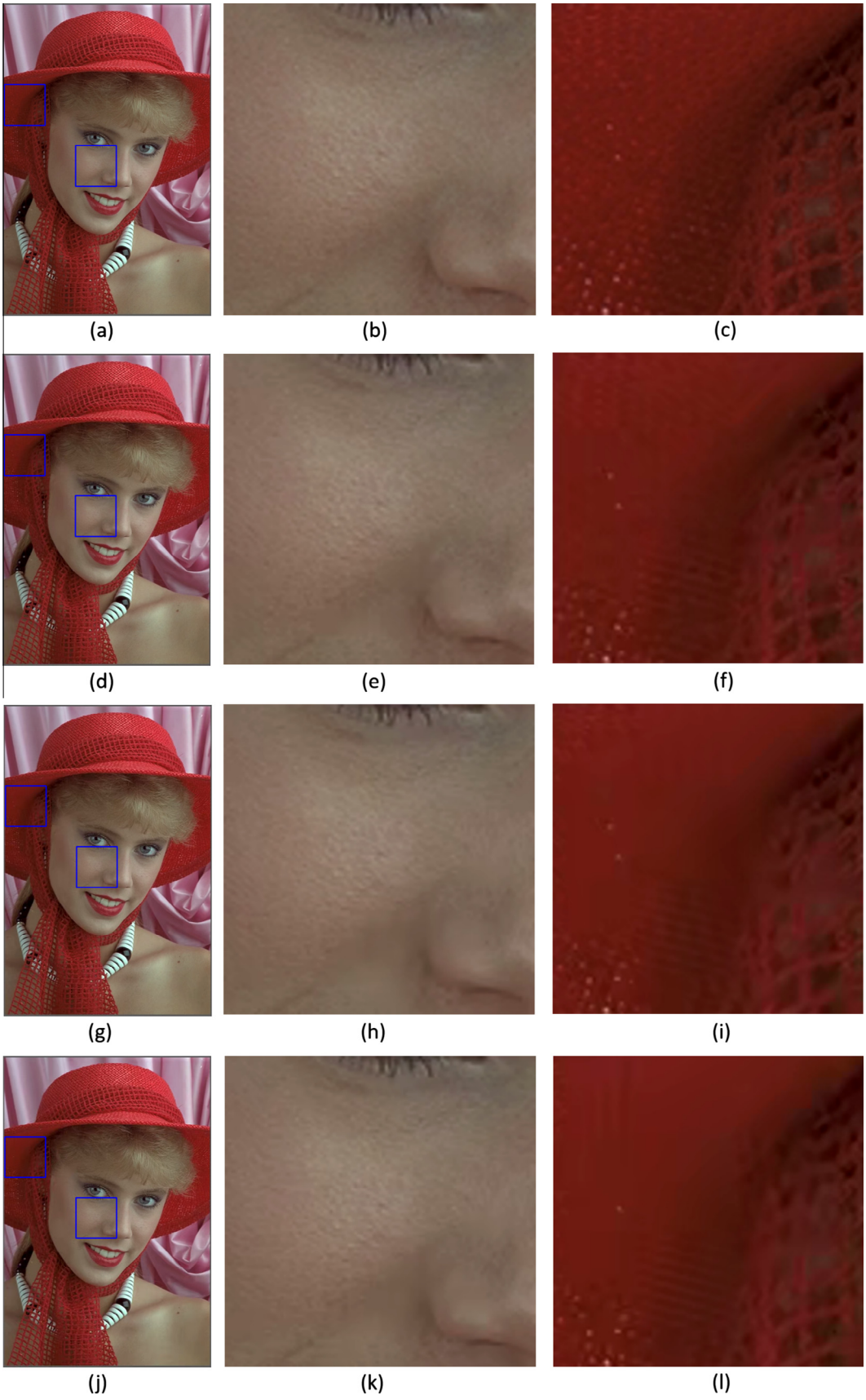

Fig. 10. The comparisons. (a)-(c) are original pictures. (d)-(f) are results of Chen's model. (g)-(i) are results of Wu's model. (j)-(l) are results of our model.

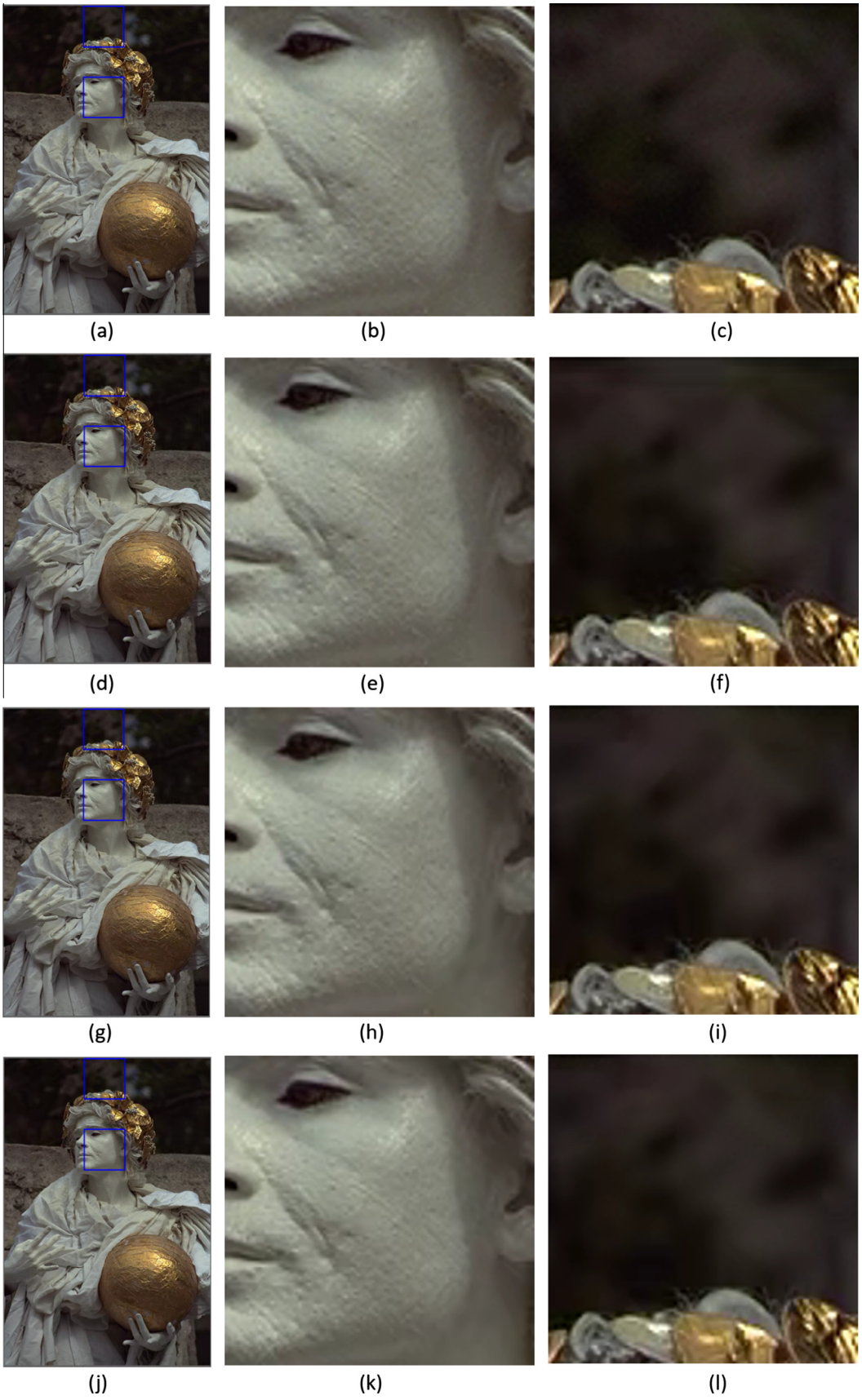

Fig. 11. The comparisons. (a)-(c) are original pictures. (d)-(f) are results of Chen's model. (g)-(i) are results of Wu's model. (j)-(l) are results of our model.

According to the results in the screening process, no subject was rejected. We also provide the detection probability (the percentage of subjects perceiving difference), for which 50% probability is a standard level for JND.

Subjective scores and PNSR comparisons of images with noise of Chen's model [19], Wu's model [22] and our FJND model (13) are shown in Table 2 and Fig. 8.

The results show that the injected noise based on our contrast masking model is almost invisible. According to the PSNR comparisons, the PSNR of distorted image with noise calculated by our FJND model is relatively smaller than other compared models. This is because our FJND model incorporates both foveation effect and disorderly concealment effect, which results in higher amplitude noises at the periphery of image.

4.2. Image compression based on foveated JND model

Furthermore, to compare the effectiveness of the proposed perceptual lossless coding algorithm, the original H.265/HEVC and method which based on the proposed FJND model are also implemented. Subjective tests have also been conducted to verify the performance of our proposed algorithm. The original HM method is used by setting pass quantification and close deblocking filter to represent lossless comparison. We also used the aforementioned Kodak Lossless True Color Image Suite as test image since it is the standard JPEG test image. The resolution of these pictures are 512×768 or 768×512.

The subjective test protocol we used is the double-stimulus continuous quality scale (DSCQS) protocol which is recommended in Rec. ITU-R BT.500 [29]. Ten female and ten male with normal vision have participated in this test and asked to vote these pictures. The vote is to compare each two stimulus pictures A and B, stimulus A is randomly the original H.265/HEVC result or the picture compressed by our method and stimulus B is the other result. We showed the 24 original H.265/HEVC results and the proposed method's results and the observers were asked to vote by the DSCQS protocol. The score range from -3 to 3 that -3 means B is much worse than A, and 0 means B is of the same quality as A. Table 3 shows the results. According to the experiments, our proposed method can save bit rate as 81.29% in average. From the Table 3 and Fig. 9, we can see though it has much bit rate reduction, the perceptual area has no visual quality loss.

In the experiment, we also compare our FJND model with other perceptual models by using our proposed H.265/HEVC perceptual lossless algorithm. As Table 3, Figs. 10 and 11 shown, we can see that at the position that we may focus, all of the reconstruction images' quality is the same and it is similar as original image. Whilst, in the area we do not focus on or the area can tolerate more distortions, the corresponding image quality may decrease but does not cause visual degradation for browsing the whole image. In addition, we can see that by using our proposed FJND model, it can save more bits than other perceptual models.

5. Conclusion

In this paper, we propose a foveated contrast masking model based on psycho-visual experiments, which models the foveation effect under the condition of contrast masking. Experiments show that our model can better exploit the perceptual redundancy by injecting higher and imperceptible distortion in the image. Furthermore, we use this model to provide the perceptually-lossless H.265/HEVC coding. The experiments show that the proposed method has up to 81.29% bit rate reduction without perceptual quality loss. The future work will focus on extending the proposed method to mobile applications where the viewing distance, screen resolution, indoor/outdoor environments bring more challenges.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 61471273), National Hightech R&D Program of China (863 Program, 2015AA015903), and Natural Science Foundation of Hubei Province of China (No. 2015CFA053).

References

[1] Information technology digital compression and coding of continuous-tone still images: requirements and guidelines, ISO/IEC 10918-1, ITU-T Rec. 81, 1994.
[2] A. Skodras, C. Christopoulos, T. Ebrahimi, The JPEG 2000 still image compression standard, IEEE Signal Process. Mag. 18 (5) (2001) 36–58.
[3] Information technology – lossless and near-lossless compression of continuous-tone still images: baseline, ISO/IEC 14495-1, ITU-T.87, 1998.
[4] M.W. Marcellin, M.J. Gormish, A. Bilgin, M.P. Boliek, An overview of JPEG-2000, in: Proc. IEEE Data Compression Conference (DCC), 2000, pp. 523–541.
[5] I. Höntsch, L.J. Karam, Adaptive image coding with perceptual distortion control, IEEE Trans. Image Process. 11 (3) (2002) 213–222.
[6] T. Acharya, L.J. Karam, F. Marino, Compression of color images based on a 2-dimensional discrete wavelet transform yielding a perceptually lossless image, US Patent 6,154,493, Nov. 28, 2000.
[7] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, IEEE Trans. Circ. Syst. Video Technol. 22 (12) (2012) 1649–1668.
[8] T. Wiegand, Draft I U-T recommendation and final draft international standard of joint video specification, ITU-T Rec. H.264, ISO/IEC 14496-10 AVC, 2003.
[9] T.K. Tan, M. Mrak, V. Baroncini, N. Ramzan, HEVC verification test results, in: Joint Collab. Team Video Coding (JCT-VC), JCTVC-Q204, 2014.
[10] J.-R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, T. Wiegand, Comparison of the coding efficiency of video coding standards including high efficiency video coding (HEVC), IEEE Trans. Circ. Syst. Video Technol. 22 (12) (2012) 1669–1684.
[11] T. Nguyen, D. Marpe, Performance analysis of HEVC-based intra coding for still image compression, in: Proc. Picture Coding Symposium (PCS), 2012, pp. 233–236.
[12] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models of human perception, Proc. IEEE 81 (10) (1993) 1385–1422.
[13] C.-H. Chou, Y.-C. Li, A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile, IEEE Trans. Circ. Syst. Video Technol. 5 (6) (1995) 467–476.
[14] X. Yang, W. Ling, Z. Lu, E.P. Ong, S. Yao, Just noticeable distortion model and its applications in video coding, Signal Process.: Image Commun. 20 (7) (2005) 662–680.
[15] R.J. Safranek, J.D. Johnston, A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1989, pp. 1945–1948.
[16] A.B. Watson, G.Y. Yang, J. Solomon, J. Villasenor, et al., Visibility of wavelet quantization noise, IEEE Trans. Image Process. 6 (8) (1997) 1164–1175.
[17] A.B. Watson, DCTune: a technique for visual optimization of DCT quantization matrices for individual images, Soc. Info. Display Dig. Tech. Papers, vol. 24, 1993. pp. 946-946.
[18] J. Lubin, A visual discrimination model for imaging system design and evaluation, Vision Models for Target Detection and Recognition, vol. 2, 1995, pp. 245–357.
[19] Z. Chen, C. Guillemot, Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model, IEEE Trans. Circ. Syst. Video Technol. 20 (6) (2010) 806–819.
[20] A.J. Schofield, M.A. Georgeson, Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour, Vision Res. 39 (16) (1999) 2697–2716.
[21] A. Liu, W. Lin, M. Paul, C. Deng, F. Zhang, Just noticeable difference for images with decomposition model for separating edge and textured regions, IEEE Trans. Circ. Syst. Video Technol. 20 (11) (2010) 1648–1652.
[22] J. Wu, G. Shi, W. Lin, A. Liu, F. Qi, Just noticeable difference estimation for images with free-energy principle, IEEE Trans. Multimedia 15 (7) (2013) 1705–1710.
[23] G.E. Legge, J.M. Foley, Contrast masking in human vision, J. Opt. Soc. Am. 70 (12) (1980) 1458–1471.
[24] Y. Zhao, Z. Chen, C. Zhu, Y.-P. Tan, L. Yu, Binocular just-noticeable-difference model for stereoscopic images, IEEE Signal Process. Lett. 18 (1) (2011) 19–22.
[25] W.S. Geisler, J.S. Perry, Real-time foveated multiresolution system for low-bandwidth video communication, in: Proc. SPIE, 1998, pp. 294–305.
[26] Z. Wang, L. Lu, A.C. Bovik, Foveation scalable video coding with automatic fixation selection, IEEE Trans. Image Process. 12 (2) (2003) 243–254.
[27] C. Yeo, H.L. Tan, Y.H. Tan, Ssim-based adaptive quantization in HEVC, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2013, pp. 1690–1694.
[28] R. Franzen, kodak lossless true color image suite, 2004. See <http://r0k.us/graphics/kodak>.
[29] Methodology for the subjective assessment of the quality of television pictures, ITU-R Rec. BT.500-11 Geneva, Switzerland, 2002.