Entropy of Primitive: A Top-Down Methodology for Evaluating the Perceptual Visual Information

Xiang Zhang[1], *Shiqi Wang*[2], *Siwei Ma*[2], *Shaohui Liu*[1], *Wen Gao*[2]
[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2]Institute of Digital Media, Peking University, Beijing 100871, China

ABSTRACT

In this paper, we aim at evaluating the perceptual visual information based on a novel top-down methodology: entropy of primitive (EoP). The EoP is determined by the distribution of the atoms in describing an image, and is demonstrated to exhibit closely correlation with the perceptual image quality. Based on the visual information evaluation, we further demonstrate that the EoP is effective in predicting the perceptual lossless of natural images. Inspired by this observation, in order to distinguish whether the loss of input signal is visual noticeable to human visual system (HVS), we introduce the EoP based perceptual lossless profile (PLP). Extensive experiments verify that, the proposed EoP based perceptual lossless profile can efficiently measure the minimum noticeable visual information distortion and achieve better performance compared to the-state-of-the-art just-noticeable difference (JND) profile.

***Index Terms*—**Entropy of primitive, visual information, just-noticeable difference, perceptual lossless profile.

1. INTRODUCTION

The human visual system (HVS) allows human beings to perceive visual information from the outside world, and the psychological process of visual information is known as visual perception. As the ultimate receiver of the images and videos is the HVS, the visual information plays an important role in the fields of image or video processing. For example, image restoration such as denoising and debluring aim to restore the visual information as much as possible, and image and video compression are meant to convey reproductions of visual information at the receiver side with the constraint of communication bandwidth. Therefore, there is an urgent need to effectively measure how much the visual information is contained in an image by a practical way. Basically, the visual information is highly relevant with the image quality; and in recent years, it is noticed that the traditional mean square error (MSE) or mean absolute difference (MAD) cannot well correlate the visual quality [1], which motivated many researchers involved in developing efficient image quality assessment algorithms, such as structural similarity (SSIM) [2], visual information fidelity [3], feature similarity [4] and internal generative mechanisms (IGM) [5]. However, these metrics are designed to measure the visual information loss given the distorted and original images, while the inherent visual information cannot be directly reflected.

In addition to these above mentioned supra-threshold image quality assessment algorithms, the just-noticeable difference (JND) for natural images has also been thoroughly studied [6]-[9]. JND refers to the minimum visual loss results from physiological and psychophysical phenomena in the HVS. As long as the distortion level is below the JND threshold, the signal distortion will not be perceived by the HVS. In the previous work [6]-[9], the JND model is classified into two categories, spatial-domain JND and the transform-domain JND. Both of them are taking advantage of the characteristics of the HVS, such as contrast sensitivity function (CSF), luminance adaptation, contrast masking and texture masking, etc. Therefore, these kinds of methodologies can be classified into the "bottom-up" method, which builds a computational system that functions the same way as the HVS.

It is well-known that the HVS is such a sophisticated system and simulating the HVS completely is almost impossible at present. However, currently all of the "bottom-up" methods are based on a simplified HVS, which may not be accurate in general. In view of this, a "top-down" philosophy is proposed in this paper for visual information evaluation. This methodology is built on the statistical properties of the entropy of primitive (EoP), which has been shown to be highly correlated with the perceptual cues [10]. In this paper, the relationship between EoP and visual information is further studied. Moreover, it is observed that when the EoP curve tends to be stable, the reconstructed image has no perceptual difference with the original image. Based on this observation, we propose a perceptual lossless profile (PLP) in generating the perceptual lossless image. It is demonstrated that the proposed PLP can accurately predict the perceptual lossless information.

The rest of this paper is organized as follows: In section 2, the core concept of EoP is introduced, and we analyze the correlation between EoP and the perceptual visual quality. The concept of visual information and a novel EoP based method for measuring the visual information is presented in section 3. In section 4, the perceptual lossless profile is then proposed. Section 5 provides the experimental results and analyses. Finally, we conclude our paper in section 6.

2. ENTROPY OF PRIMITIVE (EOP)

The image primitive coding is based on Sparse-Land model [11]. In this model, for each signal $x \in \mathbb{R}^d$, x can be represented approximately as $x \approx D\alpha$, where D ($D \in \mathbb{R}^{d \times k}$) is an over-complete dictionary containing k primitives. Since $k \gg d$, the representation vector α is sparse, indicating that $\|\alpha\|_0 \ll k$, where the notion $\|\bullet\|_0$ represents the l_0 norm in this paper.

In this work, we apply the K-SVD algorithm to train the content-adaptive dictionary [12][13]. The algorithm consists of two procedures: sparse coding and dictionary updating. The dictionary D and the sparse representation vector α are obtained by this objective function:

$$(D, \alpha) = \underset{D,\alpha}{argmin} \sum_k \|x - D\alpha\|_2{}^2 \quad s.t. \quad \|\alpha\|_0 < L. \tag{1}$$

For an input image X, we partition it into small patches $x_1, x_2, \ldots, x_N$, which are used as samples to train dictionary D by the K-SVD algorithm. After obtaining the dictionary D, the sparse representation vectors $\{\alpha_i\}$ for each patch $\{x_i\}$ are calculated by the Orthogonal Matching Pursuit (OMP) algorithm [14]. Fig. 1 illustrates an example of the dictionary trained by the 8x8 patches partitioned from Lena image.

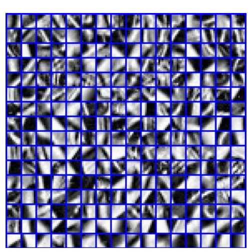


Fig. 1. Dictionary trained with overlapped image patches from Lena image.

After getting the sparse representation vectors $\{\alpha_i\}$, the times of every image primitives used for representing the patches can be obtained, which are denoted by $t_1, t_2, \ldots, t_k$. Formally, the sparse representation vector α_i has one donation to the $t_j$when the j^{th} coefficient of α_i is nonzero. According to the Shannon Theory, the entropy of primitive (EoP) is defined as [10]:

$$EoP = -\sum_{i=1}^{k} p_i \cdot log\, p_i \;\; (p_i \neq 0), \tag{2}$$

where p_i denotes the probability as follows:

$$p_i = \frac{t_i}{\sum_{j=1}^{k} t_j}. \tag{3}$$

We have shown that the EoP has some interesting statistical properties based on a large number of statistics. The EoP value increases with the number of primitives L, and tends to converge to stable values, as illustrated in Fig. 2. Moreover, EoP also has close correlation with the image

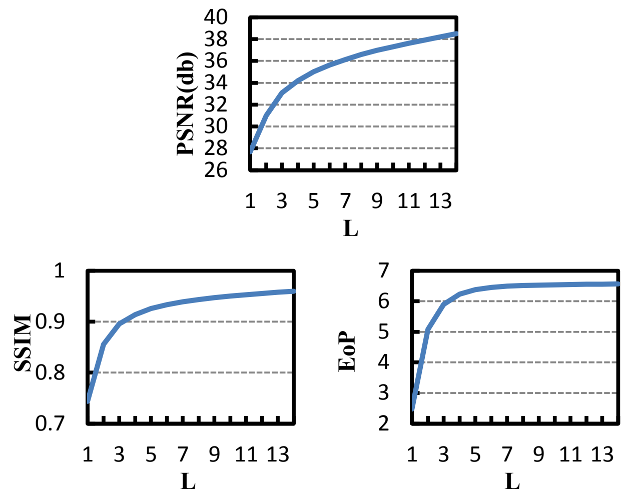


Fig. 2. PSNR, SSIM and EoP of Lena image in terms of L.

quality, as demonstrated in Fig. 2 for the SSIM and PSNR curves.

In order to quantitively evaluate the correlation between the EoP and perceptual visual quality, we calculate the Pearson Linear correlation coefficient (PLCC) to measure the correlation between the EoP and the widely employed quality measures SSIM and PSNR. Firstly, a nonlinear model, for a better fit for all data, is applied to the PSNR-EoP and SSIM-EoP respectively. The non-linear model is represented by:

$$y = a_1 \left[\frac{1}{2} - \frac{1}{1 + e^{a_2(x - a_3)}}\right] + a_4 x + a_5, \tag{4}$$

where a_1 to a_5 are model parameters using a nonlinear regression process. We use the PSNR-EoP $\{p_i, e_i\}$ pairs and SSIM-EoP$\{s_i, e_i\}$ pairs as the input of the nonlinear model to train the model parameters $a_1$to a_5. Then the PLCC can be calculated as follows:

$$\text{PLCC} = \frac{\sum_i (y_i - \bar{y})(e_i - \bar{e})}{\sqrt{\sum_i (y_i - \bar{y})^2 \sum_i (e_i - \bar{e})^2}}. \tag{5}$$

In Fig. 3, we show the PLCC of PSNR-EoP and SSIM-EoP respectively, with 30 frequently-used nature images, most of which are chosen from the LIVE database [18]. It is observed that for all the images the PLCC of SSIM-EoP is higher that of PSNR-EoP. It reveals that EoP is highly relevant to the perceptual quality, which motivated us to employ EoP in evaluating the visual information.

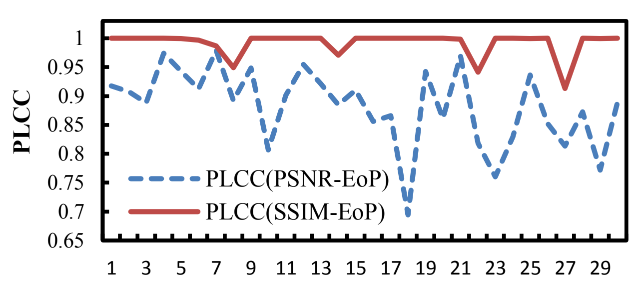
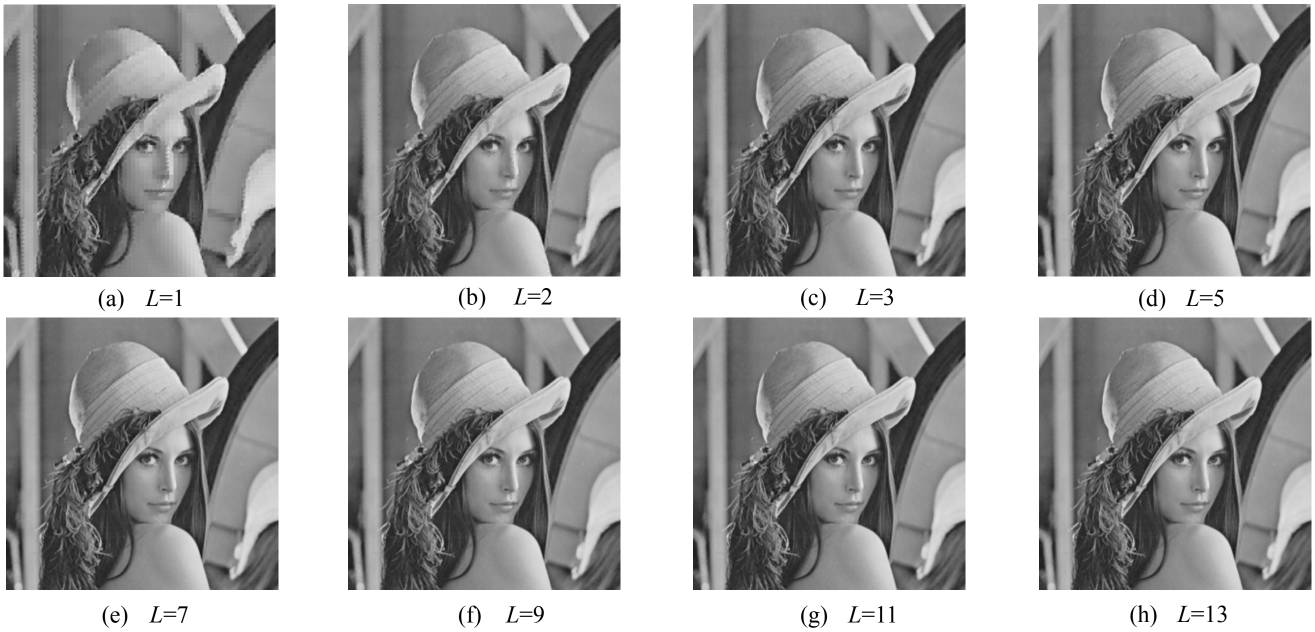


Fig.3. PLCC between PSNR/SSIM and EoP with 30 images.



(a) L=1 (b) L=2 (c) L=3 (d) L=5

(e) L=7 (f) L=9 (g) L=11 (h) L=13

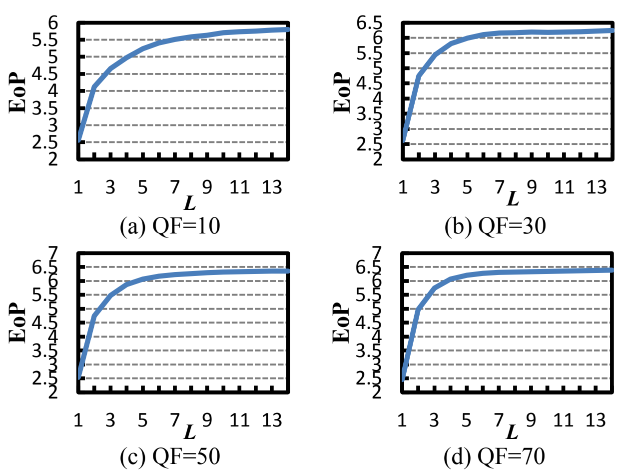
Fig. 4. Reconstructed images with different values of L.

3. VISUAL INFORMATION EVALUATION

.

As such, it is concluded that the EoP is an effective methodology to evaluate the perceptual visual information. With the increasing values of L, the perceptual quality of reconstructed image becomes better, as shown in Fig. 4. When the EoP curve goes flat after L increases up to 5 in Fig. 2, it is observed that the HVS cannot sense the visual differences anymore, even if distortions are still present in the reconstructed image.

In other words, after EoP value becomes steady, the loss of perceptual visual information is tolerable to HVS. Therefore, in this section, we propose an EoP based visual information evaluation algorithm. This algorithm can be applied not only to original images, but also to the corrupted images.



(a) QF=10 (b) QF=30

(c) QF=50 (d) QF=70

Fig. 5. EoP curves of distorted Lena images with different QF.

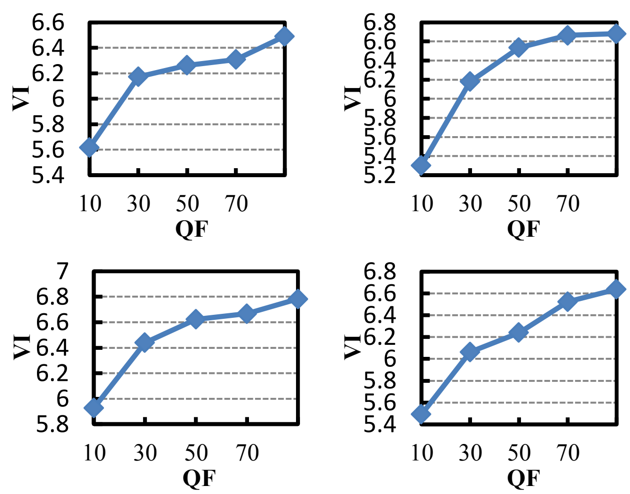
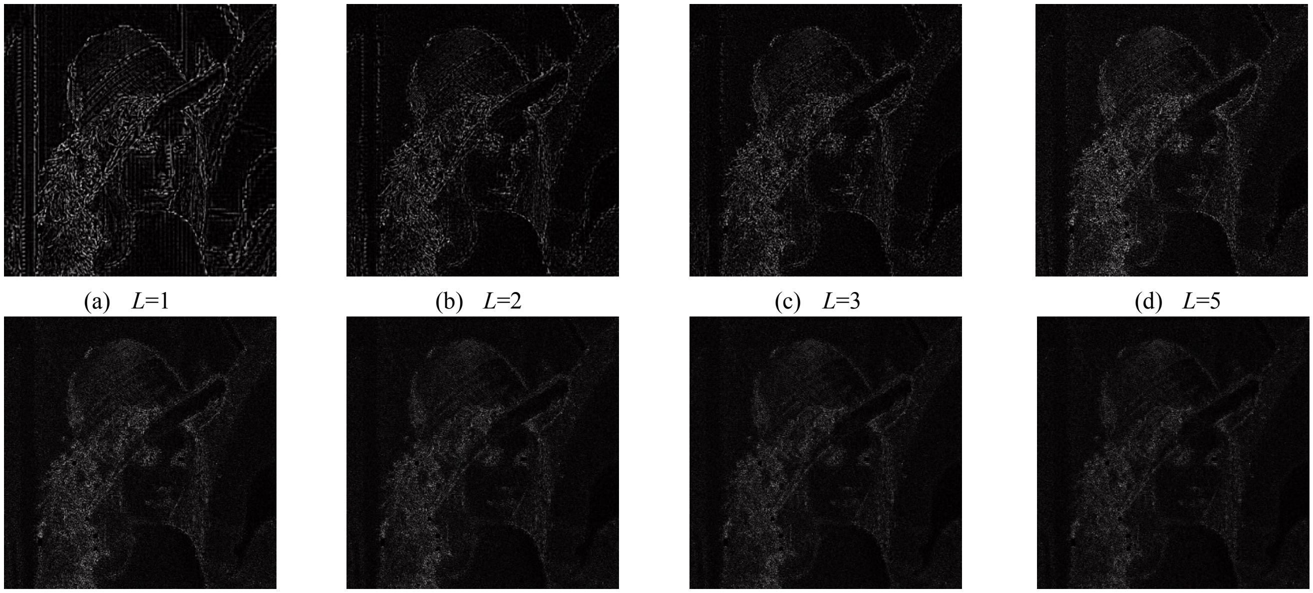


Fig. 6. The relationship between visual information and different distortion level (right most is the original) of different images.

Let X denote the input image and EoP_i represents the EoP value with $L = i$, where $1 \leq i \leq N$. N is set to be 14 in this work as larger L will not bring any further quality improvement in sparse representation. Then we define the threshold $\tilde{l}$ indicating that the visual information at this stage is equal to the original image. The threshold is defined as follows:

$$\tilde{l} = \underset{i}{argmin}\, i, \quad s.t. \ \frac{EoP_i - EoP_{i-1}}{\max_j(EoP_j) - \min_j(EoP_j)} < \varepsilon, \qquad (6)$$

where ε is a constant value and set to be 0.01 in this work. This definition is inspired from the statistical properties of EoP [10] and implies that the EoP curve will become relatively stable for $L \geq \tilde{l}$. For example, for Lena image, $\tilde{l}$ is



(a) L=1 (b) L=2 (c) L=3 (d) L=5

(e) L=7 (f) L=9 (g) L=11 (h) L=13

Fig.7. Scaled EoP based JND maps with different L.

calculated to be 7, and as shown in Fig. 2, EoP becomes stable after this value.

Finally, the visual information of the image X is defined to be the EoP at the threshold $\tilde{l}$:

$$VI = EoP_{\tilde{l}}. \quad (7)$$

To further investigate the properties of the visual information, we use different quality factors (QF) to JPEG compress the original Lena image. Fig. 5 illustrates the EoP curves of Lena image with different QF values. It is noted that larger QF values corresponds to higher quality. It is observed that these images share the similar shape of EoP curves. However, the maximum EoP values are different, which implies that they have different visual information. Moreover, in the visual information for several original and compressed images are shown in Fig. 6, from which we can conclude that better perceptual quality corresponds to higher visual information. With the increase of QF, the reconstructed image has better perceptual visual quality, and the VI grows monotonically to reach the VI of original image. Therefore, our proposed scheme is reliable for evaluating the visual information in natural images.

4. PERCEPTUAL LOSSLESS PROFILE

In this section, we further demonstrate the application of the visual information evaluation scheme. We define the perceptual lossless profile $\tilde{X}$ of an input image as the reconstructed one by $\tilde{l}$ primitives using the OMP algorithm. The perceptual lossless profile means that it has no visual difference compared with the input one. It is a powerful tool to estimate the maximum difference between distorted and original image that is tolerable for the HVS, which is also common known as just-noticeable difference (JND).

Therefore, the proposed EoP based JND model is defined as follows,

$$EJND[i,j] = abs\left(\tilde{X}[i,j] - X[i,j]\right). \quad (8)$$

The notion abs(•) represents the absolute value. $EJND[i,j]$ represents the maximum range that the luminance of $X[i,j]$ can change without being noticed by the HVS. In another word, if the luminance of $X[i,j]$ is varied beyond the threshold $EJND[i,j]$, this variation is then noticeable.

The EoP based JND map of the *Lena* image with different L value is demonstrated in Fig. 7. With the increasing values of L, the error energy in the EoP based JND map decreases gradually until the L reaches 7, which inspires us to employ the reconstructed image by 7 primitives as the perceptual lossless profile.

5. EXPERIMENTAL RESULTS

In this section, we have conducted several experiments to verify our proposed perceptual lossless profile model. The experiments are based on a two-alternative forced choice (2AFC) approach, which has widely been applied in verifying video processing algorithms [15][16][17]. Specifically, in 2AFC a subject is shown an image pair and forced to choose the better-quality one by his/her visual feelings. The said pair consists of the original input image, and either an image distorted by the traditional JND injected [7][8] noise or EoP based JND injected noise.

This method for evaluating JND is to inject JND-guided noise into images:

$$\hat{X}(i,j) = X(i,j) + \gamma \cdot S^{random}(i,j) \cdot M(i,j), \quad (9)$$

where $X(i,j)$ is the pixel in original image, γ is the parameter of the noise level, $S^{random}(i,j)$ equals either +1 or -1 randomly, $M(i,j)$ represents the JND value in pixel(i,j). And $\hat{X}$ is the noise-injected image. The parameter γ must be adjusted to make sure that different JND-guided noise has the same error energy (the same PSNR), and the better perceptual image quality means the better JND profile. The EoP based JND noised image can be obtained in the same way.

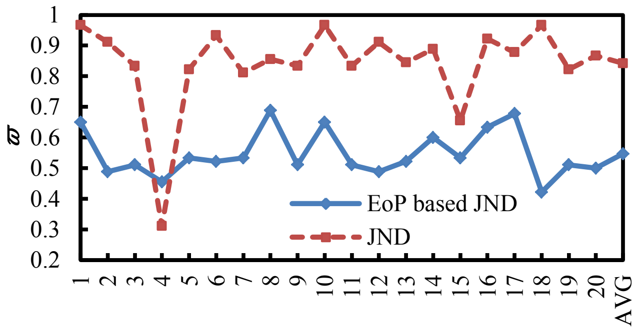
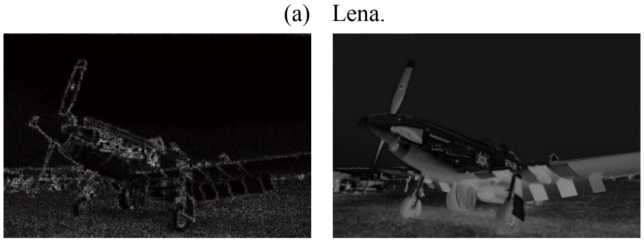


Fig. 8. Results for contrast experiment of EoP based JND and traditional JND. (The percentage ϖ and the 20 observers, the right-most value is the average).

(a) Lena.



(b) Plane.

Fig. 9. Comparison of the EoP based JND map (left) and the traditional JND map (right).

These experimental results are shown in Fig. 8, which reveals the percentage ϖ of the correct choice of the original image. It is evident that the average ϖ of EoP based JND model is very close to 50% (54.7%), that means the HVS cannot distinguish the original and the EoP based JND-noised images. On the other hand, the average ϖ of traditional JND model is much higher than 50% (84.2%). It indicates that the observers can easily distinguish the distortions. From these results, we concluded that the EoP based JND profile performs better than the traditional JND profile.

To further demonstrate the differences between EoP based JND and traditional JND model, the JND map and the JND injected images are illustrated in Fig. 9 and Fig. 10, respectively. It is shown that the traditional JND model may overestimate the maximum tolerable visual information, especially in the flat and texture area, like the body of the plane and the hair of Lena. Hence it causes bad visual quality in the reconstructed images. However, in our model, the edge and texture area are retained with better perceptual quality.

In order to further verify the hypothesis that the proposed EoP based JND is the maximum loss threshold of the visual information, we conducted another 2AFC experiment. In each pair, one of them is the EoP based JND-noised image created by applying the aforementioned method, and the other one is reconstructed with L=15 primitives with OMP algorithm. Fig. 11 shows the results of this experiment. It is observed that the average percentage ϖ is approximate to 0.5 (0.47), from which we can reach the conclusion that the HVS cannot sense the differences between the proposed EoP based JND injected images and the one reconstructed with more primitives in sparse representation.

(b) Lena

(c) Plane

Fig. 10. Comparison of EoP based JND-noised image (left) and traditional JND-noised image (right).

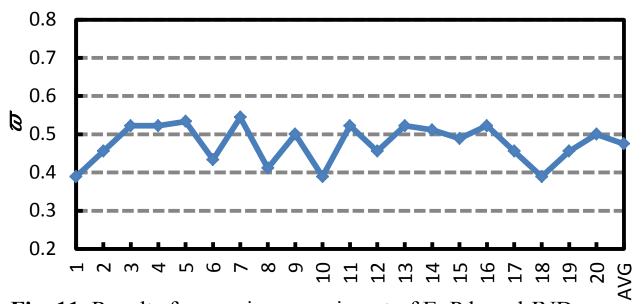


Fig. 11. Results for proving experiment of EoP based JND. (The percentage ϖ and the 20 observers, the right-most value is the average).

6. CONCLUSIONS

In this paper, we propose a top-down methodology to evaluate the visual information. The novelty of this paper lies in employing the entropy of primitive for the visual information evaluation and defining a perceptual lossless profile based on the visual information. Extensive subjective experiments are conducted to verify the accuracy and efficiency of the proposed scheme. It also shows good potential in the field of perceptual coding and quality assessment.

7. ACKNOWLEDGEMENT

This research is supported by the 973 program (2009CB320900), the 863 program (2012AA011505) and the National Science Foundation of China (61121002, 61103088, 61272386), which are gratefully acknowledged.

8. REFERENCES

[1] B. Girod, “What’s wrong with mean-squared error,” in Digital Images and Human Vision. the MIT press, 1993, pp. 207–220.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: From error visibility to structural similarity,” IEEE Trans.on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] H. R. Sheikh and A. C. Bovik, “Image information and visual quality, ”IEEE Trans. on Image Processing, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[4] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” IEEE Trans. on Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[5] J. Wu, W. Lin, G. Shi, A. Liu, “Perceptual quality metric with internal generative mechanism”, IEEE Trans. on Image Processing, vol. 22, no. 1, pp. 43–54, Jan. 2013.

[6] X. Yang, W. Lin, Z. Lu, E. Ong and S. Yao, “Motion-compensated Residue Pre-processing in Video Coding Based on Just-noticeable-distortion Profile,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 15, no. 6, pp. 742-750, Jun. 2005.

[7] X. Zhang, W. Lin and P. Xue, “Just-Noticeable Difference Estimation with Pixels in Images,” Journal of Visual Communication and Image Representation, vol. 19, no. 1, pp. 30-41, 2008.

[8] W. Lin, “Computational Models for Just-noticeable Difference, Chapter 9” in Digital Video Image Quality and Perceptual Coding, eds. H. R. Wu and K. R. Rao, CRC Press, 2006.

[9] X. Zhang, W. Lin and P. Xue, “Improved Estimation for Just-noticeable Visual Distortion,” Signal Processing, vol. 85, no. 4, pp.795-808, Apr. 2005.

[10] J. Zhang, S. Ma, R. Xiong, D. Zhao and W. Gao, “Image Primitive Coding and Visual Quality Assessment,” PCM 2012, vol. 7674, pp. 674-685, 2012.

[11] M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing", Springer, 2010.

[12] R. Rubinstein, T. Peleg and M. Elad, “Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model,” IEEE Trans. on Signal Processing, vol. 61, no. 3, pp. 661-677, March 2013.

[13] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," the IEEE Trans. on Signal Processing, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.

[14] J. A. Troppand A. C. Gilbert, “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit”, IEEE Trans. on information theory, Vol. 53, No. 12, Dec. 2007.

[15] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, “SSIM-Motivated Rate-Distortion Optimization for Video Coding,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 4, pp. 516-529, 2012.

[16] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, “Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization.” IEEE Transactions on Image Processing 22(4): 1418-1429, 2013.

[17] Z. Wang and E. P. Simoncelli, “Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities,” Journal of Vision, vol. 8, no. 12, pp. 1-13, Sept. 2008.

[18] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack. (2006). Image and Video Quality Assessment Research at LIVE [Online]. Available: http://live.ece.utexas.edu/research/quality/