Correspondences

Reduced-Reference Image Quality Assessment With Visual Information Fidelity

Jinjian Wu, Weisi Lin, *Senior Member, IEEE*, Guangming Shi, *Senior Member, IEEE*, and Anmin Liu

***Abstract*—Reduced-reference (RR) image quality assessment (IQA) aims to use less data about the reference image and achieve higher evaluation accuracy. Recent research on brain theory suggests that the human visual system (HVS) actively predicts the primary visual information and tries to avoid the residual uncertainty for image perception and understanding. Therefore, the perceptual quality relies to the information fidelities of the primary visual information and the residual uncertainty. In this paper, we propose a novel RR IQA index based on visual information fidelity. We advocate that distortions on the primary visual information mainly disturb image understanding, and distortions on the residual uncertainty mainly change the comfort of perception. We separately compute the quantities of the primary visual information and the residual uncertainty of an image. Then the fidelities of the two types of information are separately evaluated for quality assessment. Experimental results demonstrate that the proposed index uses few data (30 bits) and achieves high consistency with human perception.**

***Index Terms*—Image quality assessment, reduced-reference, internal generative mechanism, information fidelity.**

I. Introduction

Objective image quality assessment (IQA) plays an important role in image and video processing, such as in information compression, transmission, restoration and display [1]. During the last decade, a lot of IQA indices have been introduced. Most of them are full-reference (FR) methods which require the whole reference image for quality evaluation [2]. However, the reference images are not always available, and no-reference (NR) IQA indices are expected. Because of the varied image contents and the individual distortion types, the NR quality evaluation with no prior knowledge is an extremely difficult task [3].

As a compromise between FR and NR, reduced-reference (RR) IQA indices are designed to evaluate the perceptual quality by using partial information of the reference images. A successful RR IQA index is expected to use less data of reference images and achieve higher evaluation accuracy [4]. To this end, some representative global features are extracted for quality evaluation. In [3], a wavelet-domain natural image statistic metric (WNISM) is introduced. Under the assumption that most real world image distortions disturb image statistics, WNISM evaluates the perceptual quality by computing the distance between the probability distributions of wavelet coefficients. In [5], according to the distribution of wavelet coefficients, geometric information is extracted for quality assessment. In addition, by analyzing the DCT coefficient distributions, Ma *et al.* [6] reorganized the DCT coefficients into several representative subbands and evaluated the perceptual quality based on their city-block distance (ROCB). Recently, Soundararajan and Bovik [7] suggested to measure the quality degradation according to the entropies of wavelet coefficients and introduced a Reduced Reference Entropic Differences (RRED) algorithm. All of these algorithms are performed on subband domain and promote our understanding on quality assessment.

It is well known that the HVS is highly sensitive to spatial information (e.g., luminance contrast and structure) of an input image. In this letter, we try to directly analyze the quality degradation within spatial domain according to the spatial features which the HVS is highly sensitive to. Recent researches on brain theory (i.e., the Bayesian brain theory [8] and the free-energy principle [9]) suggest that the brain works with an internal generative mechanism (IGM) for visual perception and image understanding. With an input scene, the IGM actively predicts the primary visual information and attempts to avoid the residual uncertainty/disorder [9]. This mechanism reveals the inner processing of visual signals and prompts the cue for quality assessment. According to this mechanism, Zhai *et al.* [10] introduced a free-energy-based distortion metric (FEDM) which measures the quality based on the change of the residual uncertainty. We advocate that distortions on the primary visual information and the residual uncertainty result in different visual content degradations. Therefore, we need to discriminately evaluate the degradations on the two types of information.

In this letter, we introduce a novel RR IQA index by discriminately evaluating the visual content fidelities of the primary visual information and the residual uncertainty. Inspired by the active prediction of the IGM, an autoregressive (AR) model is employed to predict the visual content and to decompose the input image into two portions, the orderly portion and the disorderly portion. The orderly portion possesses the primary visual information of the input scene, which will be further processed by the HVS for image understanding and recognition [11]. The disorderly portion consists of the residual uncertainty, which will be avoided by the HVS for further processing [9]. Therefore, we separately evaluate the information fidelities on the two portions. Firstly, the quantities of information of the two portions are computed (each feature is quantized into 15 bits) according to information theory. Then, we respectively evaluate the information fidelities on the two portions, and combine the two results to acquire the overall quality score.

The rest of this letter is organized as follows. In Section II, we give the detailed description of the proposed index. Then, the performance of the proposed index is demonstrated in Section III. Finally, conclusions are drawn in Section IV.

II. The Proposed RR IQA Index

In this section, we will introduce the proposed RR IQA index in detail, and its deployment is illustrated in Fig. 1. Inspired by the IGM theory, we decompose the input image into two portions, the orderly visual information and the disorderly uncertainty. Then the quantity

Manuscript received June 24, 2012; revised September 30, 2012 and December 28, 2012; accepted January 07, 2013. Date of publication June 04, 2013; date of current version October 11, 2013. This work was supported by the Major State Basic Research Development Program of China (973 Program) (No. 2013CB329402), NSF of China (No. 61033004, 61070138, 61072104, and 61227004), and the Fundamental Research Funds for the Central Universities (No. K50513100005 and K5051202034).. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sheng-Wei (Kuan-Ta) Chen.

J. Wu and G. Shi are with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, School of Electronic Engineering, Xidian University, Xi'an, China (e-mail: jinjian.wu@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

W. Lin (corresponding author) and A. Liu are with the School of Computer Engineering, Nanyang Technological University, Nanyang 639798, Singapore (e-mail: wslin@ntu.edu.sg; liua0002@ntu.edu.sg).

Digital Object Identifier 10.1109/TMM.2013.2266093

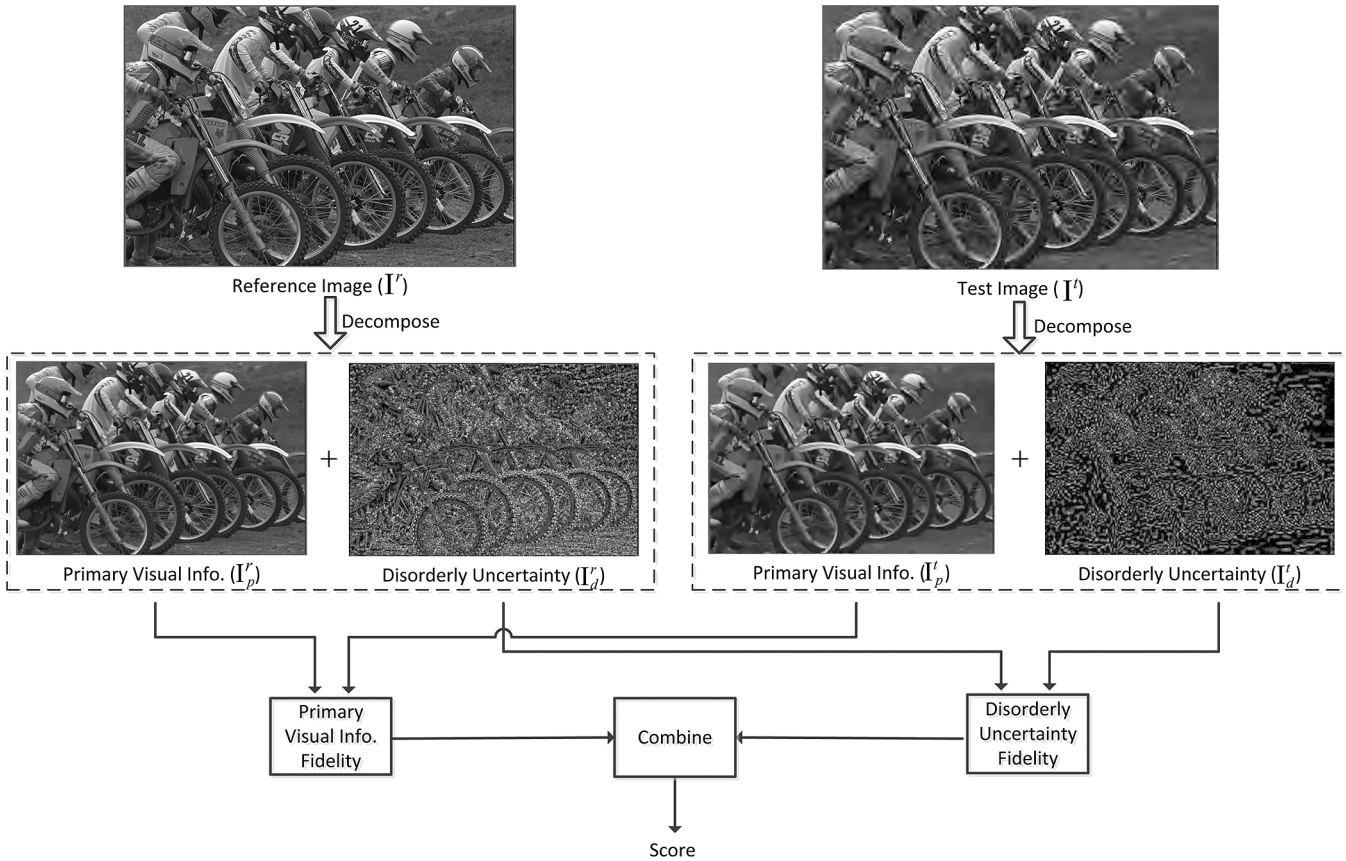

Fig. 1. Deployment of the proposed index. I^r is the original image and I^t is the contaminated image. Pixel values of the two disorderly uncertainty images are scaled to [0, 255] for a clear view.

of information of each portion is computed. Finally, we evaluate the visual quality based on the information fidelity.

A. Image Understanding Within the IGM

As an active signal processing system, the HVS helps us to understand the colorful world. Recent research on brain theory suggests that the HVS has an IGM for visual perception and understanding [8], [10]. To the retinal stimuli, the IGM actively predicts and explains their sensations according to the inherent priori knowledge, and tries to avoid the remaining uncertainty for further processing [9]. Inspired by the active prediction in the IGM, we suppose to decompose an input image into two portions, which we call the orderly and disorderly portions, for quality assessment. Both the Bayesian brain theory [8] and the free-energy principle [9] indicate that the IGM optimizes the input scene by minimizing the prediction error. From the perspective of pixel, the IGM tries to accurately predict the value of pixel x with minimum error. It is well known that pixels are highly correlated with their surroundings and jointly carry structural information [12]. Therefore, these correlated pixels possess high inter-pixel redundancy, and we can accurately predict pixels based on the relationships among them. In other words, according to the relationships between a central pixels and its surrounding pixels, the prediction error of the central pixel can be minimized by highlighting these highly correlated pixels on the prediction procedure. To this end, an AR based prediction model is introduced in [13], in which the central pixel x is predicted based on the structure similarity with its surrounding pixels x_i.

$$x_p = \sum_{x_i \in \mathcal{X}} \mathcal{C}_i x_i + \varepsilon, \tag{1}$$

where x_p is the predicted value of pixel x, $\mathcal{X}$ being a local surroundings of x, $\mathcal{C}_i$ being the structural related adjusting parameter, and ε is white noise. More details about (1) can be found in [13]. With the help of (1), an input image is decomposed into the orderly portion with x_p and the disorderly portion with $x - x_p$. As shown in Fig. 1, the orderly portion (I_p^r and I_p^t) possesses the primary visual information of an image (I^r and I^t), and the disorderly portion (I_d^r and I_d^t) possesses the residual uncertainty. Furthermore, the two portions contain different visual information and play different roles on image perception and understanding.

Visual quality is closely related to the information fidelities on the orderly and disorderly portions. Distortions on the orderly portion disturb the prediction of the primary visual information, such as blurring the structure and degrading the edge, which directly impact on image understanding. While distortions on the disorderly portion have little interference on the prediction of primary visual information, which mainly arouse uncomfortable perception and have limited effect on image understanding [12]. As shown in Fig. 1, distortions on the orderly portion I_p^t blur the structure of the motorbike and the words on the clothes. Comparing with the degradation on the orderly portion I_p^t, the HVS takes less care about the degradation on the disorderly portion I_d^t (further analysis of distortions on the two portions will be given in Section III-A). Therefore, distortions on the orderly and disorderly portions have different effects and we should discriminatively evaluate the information fidelities on the two portions.

B. Information Fidelity Based Quality Assessment

The IGM performs as an active inference system which perceives an input scene by adjusting the inner configuration [14]. Hence, we as-

0	0	0	0	0
1	3	8	3	1
0	0	0	0	0
-1	-3	-8	-3	-1
0	0	0	0	0

(a)

0	0	1	0	0
0	8	3	0	0
1	3	0	-3	-1
0	0	-3	-8	0
0	0	-1	0	0

(b)

0	0	1	0	0
0	0	3	8	0
-1	-3	0	3	1
0	-8	-3	0	0
0	0	-1	0	0

(c)

0	1	0	-1	0
0	3	0	-3	0
0	8	0	-8	0
0	3	0	-3	0
0	1	0	-1	0

(d)

Fig. 2. Edge filters for four directions.

sume the IGM as a parametric system ($\mathcal{B}$) for visual stimuli processing, within which the quantity of information (E) of an input image I can be computed as

$$E(I/\mathcal{B}) = -\log\ P(I/\mathcal{B}), \tag{2}$$

where $P(I/\mathcal{B})$ is the conditional probability of I given $\mathcal{B}$.

As discussed in the above subsection, an image I is decomposed into two portions (i.e., I_p and I_d) for processing. Since the two portions contain different visual information, it is reasonable to suppose that their contents are independent. The representation of the image in the HVS can be regarded as the union representations of the two portions. According to information theory [15], (2) can be rewrote as

$$\begin{aligned} E(I/\mathcal{B}) &= -\log\ P(I_p, I_d/\mathcal{B}) = -\log(P(I_p/\mathcal{B})P(I_d/\mathcal{B})) \\ &= -\log\ P(I_p/\mathcal{B}) - \log\ P(I_d/\mathcal{B}) \\ &= E(I_p/\mathcal{B}) + E(I_d/\mathcal{B}), \end{aligned} \tag{3}$$

where $E(I_p/\mathcal{B})$ and $E(I_d/\mathcal{B})$ are the quantities of information of I_p and I_d, respectively.

The HVS is highly adapted to extract the luminance contrast, which can effectively represent the primary visual content and is often used to measure the degradation of image structure [16], [17]. Therefore, we employ the luminance contrast map ($\mathcal{E}$) to represent the primary visual information (I_p) in the HVS ($\mathcal{B}$), which is computed as [17],

$$\mathcal{E}(x) = \max_{k=1,\ldots,4} Grad_k(x), \tag{4}$$

$$Grad_k = |\varphi \nabla_k * I_p|, \tag{5}$$

where ∇_k are four directional filters, as shown in Fig. 2, $\varphi = 1/16$, and symbol $*$ denotes the convolution operation.

With (4) and (5), the contrast map of the primary visual information (i.e., I_p^r and I_p^t) is acquired. Then, the probability distribution of the contrast map is calculated based on the intensity of the contrast. Finally, according to Shannon Entropy [15], the quantity of information of the primary visual information (i.e., $E(I_p^r/\mathcal{B})$ and $E(I_p^t/\mathcal{B})$) is obtained. The fidelity of the primary visual information (i.e., between I_p^r and I_p^t) is measured as

$$Q_p = \frac{2E\left(I_p^r/\mathcal{B}\right) E\left(I_p^t/\mathcal{B}\right)}{\left(E\left(I_p^r/\mathcal{B}\right)\right)^2 + \left(E\left(I_p^t/\mathcal{B}\right)\right)^2}. \tag{6}$$

Since the disorderly portion consists of residual uncertainty which is independent from the primary visual information, its intensity directly represents the uncertain degree [9]. Thus, we employ the intensity energy of the disorderly portion to represent its quantity of information,

$$E(I_d/\mathcal{B}) = \frac{1}{N}\sum_{x=1}^{N} I_d(x)^2, \tag{7}$$

where N is the total number of pixels in image I_d. The information fidelity between the disorderly portions of the reference and test images is measured as

$$Q_d = \frac{2E\left(I_d^r/\mathcal{B}\right) E\left(I_d^t/\mathcal{B}\right)}{\left(E\left(I_d^r/\mathcal{B}\right)\right)^2 + \left(E\left(I_d^t/\mathcal{B}\right)\right)^2}, \tag{8}$$

where $(I_d^r/\mathcal{B})$ and $(I_d^t/\mathcal{B})$ are the quantities of information of I_d^r and I_d^t, respectively.

Combining the two evaluation results, Q_p and Q_d, we deduce the overall visual quality $\mathcal{S}$ as follows

$$\mathcal{S} = Q_p^{\alpha} Q_d^{\beta}, \tag{9}$$

where α and β denote the relative importance of the two parts. Since Q_p represents the fidelity of the primary visual information, it is more important than Q_d which represents the fidelity of the residual uncertainty, in this letter, we simply set $\alpha = 0.9$ and $\beta = 0.1$ (more discussion about the two parameters will be given in Section III).

III. Experimental Results

In this section, we firstly illustrate the effectiveness of the proposed index. Then we verify the proposed index by comparing it with three latest RR IQA indices and two classic FR IQA indices. In the experiment, the quantities of information of orderly and disorderly portions are firstly quantized into [0, 255.99] (only retains 2 digits after the decimal point, and therefore the value can be represented within 15 bits) for further processing.

A. Analysis on the Proposed Index

The proposed RR IQA index is based on visual information fidelity, which includes the content degradations on both of the primary visual information and the disorderly uncertainty. Fig. 3 shows an example of the proposed RR IQA index (considering the resolution of the screen, we crop a part of the bikes image for a clear view). From the left to the right columns, they are reference images, white noise (WN) contaminated images ($\mathrm{MSE} = 220$), and JPEG2000 (JP2K) contaminated images ($\mathrm{MSE} = 202$).

Different types of distortion result in different visual content degradations. As shown in Fig. 3(b), the white noise mainly increases the uncertainty/disorder of the image. Since the HVS can effectively filter out the white noise and actively predict the primary visual information, the WN in Fig. 3(b) mainly causes uncomfortable sensation and has little effect on image understanding. By comparing Fig. 3(a) with (b), it can be seen that Fig. 3(b) possesses almost as much primary visual information as Fig. 3(a), such as the detailed structure of the tyre and the words on the steel tube. In summary, the WN in Fig. 3(b) increases the disorderly uncertainty and has little degradation on the primary visual information. However, the JP2K distortion mainly causes degradation on image structure. As shown in Fig. 3(c), the structure of the tyre is severely blurred, and the words on the steel tube has almost been erased. Though with a similar level of error energy (represented by MSE), the primary visual information in Fig. 3(c) is much more severely distorted than that in Fig. 3(b). As a result, Fig. 3(c) (with $\mathrm{DMOS} = 42.2$) gains a lower perceptual quality than Fig. 3(b) (with $\mathrm{DMOS} = 38.0$).

The proposed index discriminatively measures the fidelities on the primary visual information and the disorderly uncertainty. As shown in Fig. 3, the input image is firstly decomposed into two portions, the orderly images locate at the second row and the disorderly images locate at the bottom row. Under the WN, the orderly image Fig. 3(e) is highly similar with Fig. 3(d) and their entropy are almost the same (Fig. 3(d) with $H_r^p = 28.88$ and Fig. 3(e) with $H_t^p = 28.35$). However, with the JP2K distortion, the orderly image (Fig. 3(f)) is much different from Fig. 3(d), and its entropy ($H_t^p = 25.83$) is much smaller than Fig. 3(d). On the other hand, the white noise increases the disorderly degree of Fig. 3(b), and the energy of its disorderly uncertainty portion (Fig. 3(h) with $H_r^d = 19.22$) is larger than that of the reference image (Fig. 3(g) with $H_r^d = 14.20$). While with the blurring effect from the JP2k distortion, the disorderly uncertainty is decreased and its corresponding energy is also decreased (Fig. 3(i) with $H_r^d = 12.39$).

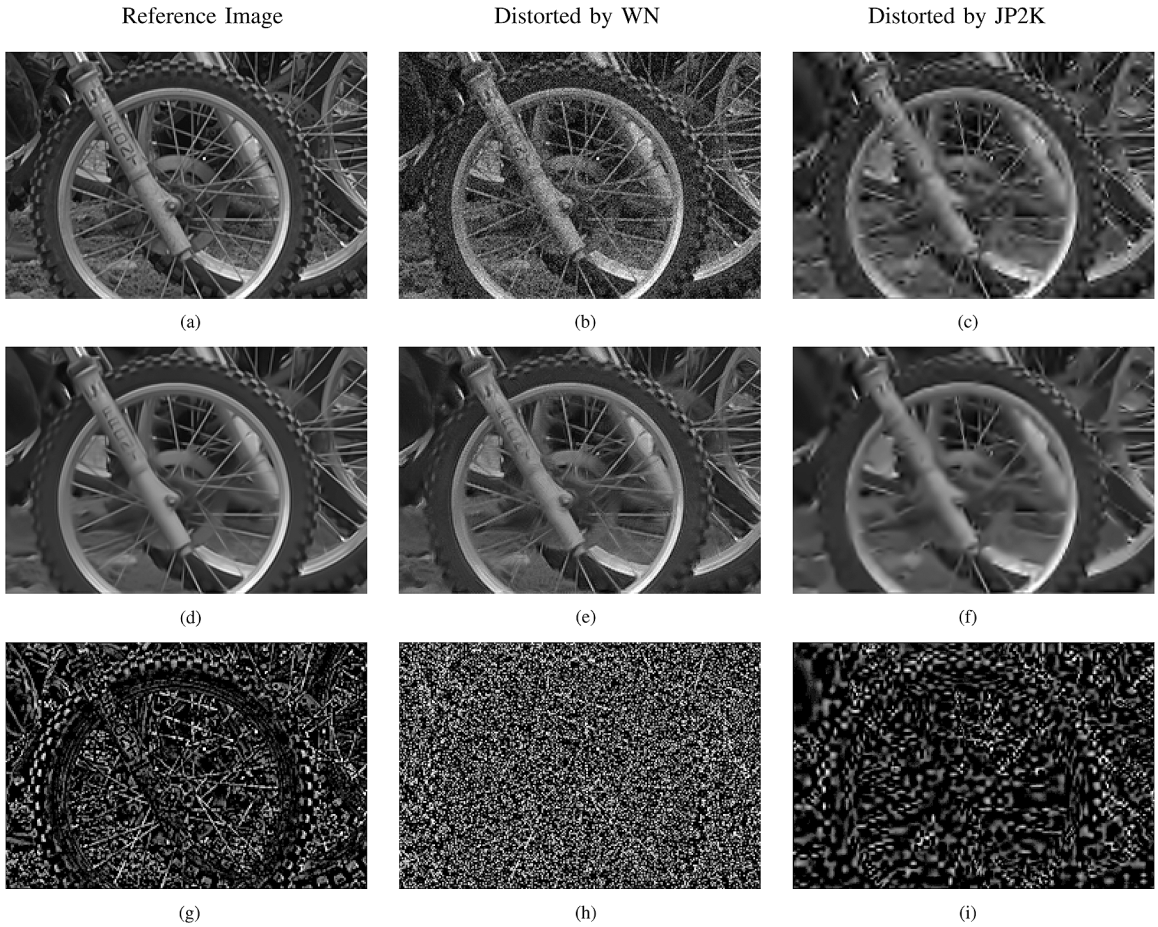

Fig. 3. Visual information degradation analysis. From the top to the end rows, they are the original images, the primary visual information portions, and the disorderly uncertainty portions (their pixel values are scaled to [0, 255] for a clear view), respectively. (a) $\mathrm{MSE} = 0$, (b) $\mathrm{MSE} = 220$, (c) $\mathrm{MSE} = 212$, (d) $H_p^r = 28.88$, (e) $H_p^t = 28.35$, (f) $H_p^t = 25.83$, (g) $H_d^r = 14.20$, (h) $H_d^t = 19.22$, (i) $H_d^t = 12.39$.

TABLE I
PERFORMANCE OF IQA INDICES ON LIVE DATABASE

Distortion Type	Algo. / Criteria	RR					FR	
		proposed	FEDM	WNISM	RRED	RRED*	PSNR	MS-SSIM
Reference Data		30	32	162	32	$32\times\frac{N}{36}$	$8\times N$	$8\times N$
JPEG2000	CC	0.932	0.921	0.924	0.930	0.963	0.896	0.957
	SRCC	0.950	0.915	0.920	0.923	0.958	0.889	0.953
	RMSE	5.881	6.316	6.176	9.280	6.811	7.180	4.691
JPEG	CC	0.895	0.875	0.876	0.831	0.979	0.860	0.943
	SRCC	0.885	0.854	0.851	0.836	0.976	0.841	0.942
	RMSE	7.148	7.733	7.711	17.74	6.448	8.169	5.327
White Noise	CC	0.957	0.925	0.890	0.926	0.985	0.982	0.974
	SRCC	0.946	0.915	0.870	0.916	0.978	0.985	0.973
	RMSE	4.658	6.065	7.286	10.58	4.901	2.989	3.648
Gaussian Blur	CC	0.955	0.902	0.888	0.953	0.969	0.784	0.955
	SRCC	0.961	0.931	0.915	0.953	0.968	0.782	0.954
	RMSE	4.660	6.783	7.222	5.622	4.604	9.766	4.685
Fast Fading	CC	0.944	0.875	0.925	0.922	0.941	0.890	0.947
	SRCC	0.941	0.852	0.923	0.916	0.943	0.890	0.947
	RMSE	5.424	7.961	6.245	11.03	9.618	7.514	5.302
Overall	CC	0.725	–	0.710	0.764	0.939	0.872	0.943
	SRCC	0.732	–	0.703	0.765	0.943	0.876	0.945
	RMSE	17.15	–	18.43	17.61	9.433	13.37	9.09

With discriminative measurement on information fidelities of the orderly portion and the disorderly portion, the proposed RR IQA can accurately evaluate the degradations of the two contaminated images (i.e., Fig. 3(b) and (c)). According to the proposed index, the WN contaminated image (Fig. 3(b)) mainly increases the energy of the disorderly portion and has little effect on the orderly portion; but the JP2K noise contaminated image (Fig. 3(c)) has severe degradation on the orderly portion, meanwhile it slightly decreases the energy of the disorderly portion. As a result, the measurement results from the proposed index (Fig. 3(b) (with $\mathcal{S} = 0.954$) has a better quality than Fig. 3(c) (with $\mathcal{S} = 0.937$)) are consistent with the subjective evaluation results (represented by the DMOS values).

B. Performance Comparison

In order to make a comprehensive analysis, we verify the proposed RR IQA index on two large databases: LIVE database [18], which is comprised of five prevailing distortion types (i.e., JPEG2000, JPEG, white noise, Gaussian blur and fast fading) across 799 distorted images; and TID database [19], which is comprised of 17 types of distortion across 1700 distorted images. Furthermore, three well known and/or latest RR IQA indices: RRED [7], FEDM [10] and WNISM [3]; and two classical FR IQA indices: PSNR and MS-SSIM [2] are chosen for comparison.

According to the performance evaluation standard proposed by the Video Quality Experts Group (VQEG) [20], three performance criteria, which are Spearman rank-order correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC), and root mean squared error (RMSE), are adopted to evaluate the performance of the IQA methods. A better IQA index has higher SRCC and PLCC, while lower RMSE values.

A successful RR IQA index is expected to use less data of reference images and achieve higher evaluation accuracy. Therefore, the reference data of these IQA indices is firstly listed in Table I. As can be seen, the reference data of the proposed index, FEDM, WNISM, a single scalar based RRED, and large scalars based RRED^* are 30, 32, 162, 32, and $32\times\frac{N}{36}$ (where N is the size of the image in pixels). For fair comparison (the reference data of RRED^* are much larger than those of the proposed index, and the reference data of RRED are similar with those of the proposed index), we mainly compare the proposed index with the single scalar based RRED.

The evaluation results on LIVE database and TID database are listed in Tables I and II. As shown in Table I, comparing with the three RR IQA indices (i.e., FEDM, WNISM, RRED), the proposed RR IQA index performs the best on all of the five types of distortion. In addition, the proposed RR IQA index outperforms the FR PSNR index on four out of five types of distortion (i.e., JPEG2000, JPEG, Gaussian blur and fast fading); it performs similarly with the FR MS-SSIM index on Gaussian blur and fast fading, and a little worse on the other three types of distortions. On TID database (as shown in Table II), the proposed index performs much better on fourteen out of seventeen types of distortion than WNISM and RRED, and is comparable with the two FR-IQA indices (i.e., PSNR and MS-SSIM) on almost all of the types of distortion. In summary, the proposed index performs very well on each type of distortion.

The overall performance of the proposed index on the whole database is also given in Tables I and II. On LIVE database, the overall performances of the proposed index is similar with WNISM and a little worse than RRED. On TID database, the overall performance of the three RR-IQA indices are similar (the proposed index is slightly better than WNISM and RRED). With further analysis, we have found that different types of distortion generate different degradation on the orderly and disorderly portions, and the two portions should not be combined with two fixed parameters (i.e., α and β). Therefore, the two parameters should be determined based on distortion type. We can improved the overall performance on LIVE database with the help of classification procedure (as used in [21]). For example, we randomly select 15 reference images (29 reference images in total in LIVE database) and their corresponding distorted images for training. Through the training, the distortion classifier is acquired. Meanwhile, the α and β for each type of distortion are obtained according to the minimized error between the computed scores and the DMOS values. Then, the remaining test images are classified and evaluated.

TABLE II
SRCC VALUES OF IQA INDICES ON TID DATABASE

Distortion type	Proposed	WNSIM	RRED	PSNR	MS-SSIM
awgn	0.884	0.603	0.702	0.911	0.809
awgn-color	0.884	0.604	0.684	0.907	0.806
spatial corr-noise	0.875	0.599	0.712	0.923	0.820
masked noise	0.893	0.633	0.744	0.849	0.816
high-fre-noise	0.932	0.708	0.794	0.932	0.869
impulse noise	0.742	0.593	0.549	0.918	0.687
quantization noise	0.781	0.679	0.591	0.870	0.854
gblur	0.942	0.871	0.935	0.868	0.961
denoising	0.948	0.864	0.925	0.938	0.957
jpg-comp	0.599	0.834	0.819	0.901	0.935
jpg2k-comp	0.928	0.935	0.948	0.830	0.974
jpg-trans-error	0.891	0.875	0.782	0.767	0.874
jpg2k-trans-error	0.837	0.691	0.628	0.777	0.853
pattern-noise	0.589	0.452	0.277	0.593	0.734
block-distortion	0.595	0.590	0.691	0.585	0.762
mean shift	0.573	0.292	0.418	0.697	0.737
contrast	0.775	0.701	0.723	0.613	0.640
overall	0.528	0.512	0.521	0.553	0.853

We conducted such procedure for 100 times (to ensure the algorithm is robust) and the average SRCC value for LIVE database is 0.867. However, with much more distortion types, the classification method introduced in [21] is not efficient enough for the TID database, and further studies are needed in this regard.

IV. CONCLUSION

In this letter, inspired by the recent brain theory, a visual information fidelity based RR IQA index is introduced. The IGM theory indicates that the HVS actively predicts the primary visual information and tries to avoid the residual uncertainty for image perception and understanding. Therefore, distortions separately degrade the primary visual information and change the disorderly uncertainty, and we should discriminatively evaluate their degradations. We firstly decompose an image into predicted/orderly and uncertain/disorderly portions. Then according to information theory, we compute the quantities of visual information of the two portions, respectively. Finally, we separately evaluate the information fidelities on the two portions and combine the two results to acquire the overall quality. Experimental results demonstrate that the proposed RR IQA index needs few bits to achieve high consistency with human perception.

REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
[2] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Conf. Record 37th Asilomar Conf. Signals, Systems and Computers, 2003*, 2003, vol. 2, pp. 1398–1402, Vol. 2.
[3] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *SPIE*, 2005, vol. 5666, pp. 149–159.
[4] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Select. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.
[5] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.
[6] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.
[7] R. Soundararajan and A. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[8] D. C. Knill and R. Pouget, "The bayesian brain: The role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, pp. 712–719, 2004.
[9] K. Friston, "The free-energy principle: A unified brain theory?," *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, Feb. 2010.
[10] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free energy principle," *IEEE Trans. Image Process.*, to be published.
[11] R. Sternberg, *Cognitive Psychology*, 3rd ed. Belmont, CA, USA: Thomson Wadsworth, Aug. 2003.
[12] J. Wu, F. Qi, and G. Shi, "Self-similarity based structural regularity for just noticeable difference estimation," *J. Visual Commun. Image Represent.*, vol. 23, no. 6, pp. 845–852, 2012.
[13] X. Zhang and X. Wu, "Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.
[14] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *J. Physiol., Paris*, vol. 100, no. 1–3, pp. 70–87, Sep. 2006.
[15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
[16] G.-H. Chen, C.-L. Yang, L.-M. Po, and S.-L. Xie, "Edge-Based structural similarity for image quality assessment," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2006 (ICASSP 2006)*, May 2006, vol. 2, p. II.
[17] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment base on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
[18] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack, Image and Video Quality Assessment Research at Live, 2004. [Online]. Available: http://live.ece.utexas.edu/research/quality/.
[19] N. Ponomarenko and K. Egiazarian, Tampere Image Database 2008 TID2008, 2008. [Online]. Available: http://www.ponomarenko.info/tid2008.htm.
[20] Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment II, 2003, video Quality Expert Group (VQEG). [Online]. Available: http://www.vqeg.org/.
[21] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

Just Noticeable Difference Estimation for Images With Free-Energy Principle

Jinjian Wu, Guangming Shi, *Senior Member, IEEE*, Weisi Lin, *Senior Member, IEEE*, Anmin Liu, and Fei Qi

***Abstract*—In this paper, we introduce a novel just noticeable difference (JND) estimation model based on the unified brain theory, namely the free-energy principle. The existing pixel-based JND models mainly consider the orderly factors and always underestimate the JND threshold of the disorderly region. Recent research indicates that the human visual system (HVS) actively predicts the orderly information and avoids the residual disorderly uncertainty for image perception and understanding. Thus, we suggest that there exists disorderly concealment effect which results in high JND threshold of the disorderly region. Beginning with the Bayesian inference, we deduce an autoregressive model to imitate the active prediction of the HVS. Then, we estimate the disorderly concealment effect for the novel JND model. Experimental results confirm that the proposed JND model outperforms the relevant existing ones. Furthermore, we apply the proposed JND model in image compression, and around 15% of bit rate can be reduced without jeopardizing the perceptual quality.**

***Index Terms*—Autoregressive (AR) model, disorder, free energy, internal generative mechanism (IGM), just noticeable difference (JND).**

I. Introduction

As an important part of the central nervous system, the human visual system (HVS) helps us to know the outside world by processing the visual detail. Rather than literally translating the input scenes, the HVS actively infers the input scenes with an internal generative mechanism (IGM) [1]–[3]. Recently, a free-energy principle has been introduced, which tries to formulate the IGM and intends to provide a unified account for human action, perception, and learning [4], [5].

The underlying idea of the free-energy principle is that all adaptive biological agents resist a natural tendency to disorder [5]. In other words, by optimizing the configuration of the IGM, e.g., adaptively adjusting the way it samples the environment and/or the internal encodement of the information [4], [6], the HVS tries to extract as much information as possible to minimize the uncertainty of an input scene. The principle indicates that the HVS cannot fully process all of the sensation information and tries to avoid some surprises (i.e., information with uncertainties, which are usually from disorderly regions) [5], [7]. Such disorderly concealment effect reveals a key limitation of the human perception which should be properly considered in the just noticeable difference (JND) estimation models. The JND threshold, which refers to the minimum visibility threshold of the HVS, is useful in perceptual image/video processing systems [8]–[10].

Manuscript received August 22, 2012; revised October 29, 2012, January 25, 2013; accepted March 11, 2013. Date of publication June 13, 2013; date of current version October 11, 2013. This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329402, the National Science Foundation of China under Grants 61033004, 61070138, 61072104, and 61227004, and the Fundamental Research Funds for the Central Universities under Grant K50513100005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yiannis Andreopolous.

J. Wu, G. Shi, and Fei Qi are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: jinjian.wu@mail.xidian.edu.cn; gmshi@xidian.edu.cn; fred.qi@ieee.org).

W. Lin and A. Liu are with the School of Computer Engineering, Nanyang Technological University, Nanyang 639798, Singapore. (e-mail: wslin@ntu.edu.sg; liua0002@ntu.edu.sg).

Digital Object Identifier 10.1109/TMM.2013.2268053