Internal Generative Mechanism Inspired Reduced Reference Image Quality Assessment with Entropy of Primitive

Shanshe Wang[†], Shiqi Wang[#], Ke Gu[§], Xiaoqiang Guo[*], Siwei Ma[†],Wen Gao[†]
[†]Institute of Digital Media, Peking University, Beijing, China
[#]Department of Computer Science, City University of Hong Kong, Hong Kong, China
[§]Beijing University of Technology,Beijing, China
[*]Academy of Broadcasting Science, SAPPRFT, Beijing, China
[†]{sswang,swma,wgao}@pku.edu.cn,[#]shiqwang@cityu.edu.hk, [§] guke@bjut.edu.cn,
[*]guoxiaoqiang@abs.ac.cn

***Abstract*—In this paper, we propose a novel reduced-reference (RR) image quality assessment (IQA) algorithm based on the internal generative mechanism, which suggests that the human visual system (HVS) can actively predict the primary visual information and avoid the uncertainty. Specifically, the explanation of the visual scene is formulated as the process of sparse representation. In particular, the entropy of primitive accounts for the primary visual information and the discrepancy between the image signal and its best sparse description is regarded as the uncertainty in perception. As such, the combined feature that can summarize the primary visual information and uncertainty in sparse domain is required to be transmitted in the RR-IQA framework. Comparative studies of the proposed reduced reference metric is conduced on both single and multiple distortion databases, and experimental results demonstrate that the proposed metric can achieve high correlation with the human perception by only sending ignorable additional information.**

***Index Terms*—Reduced-reference, image quality assessment, sparse representation, internal generative mechanism, entropy-of-primitive**

I. Introduction

Developing accurate, efficient and reliable objective image quality assessment (IQA) algorithms is crucial as they can be applied in evaluating, controlling, and optimizing the visual quality of the multimedia systems [1], [2]. In the literature, most IQA algorithms assume that the "pristine" signal is fully available, which are also termed as full reference (FR) IQA. However, they are not always practical as the reference signals may not be available at the receiver side. Therefore, designing robust IQA algorithms that rely on much less information from the reference image is another focus of research. These algorithms can be categorized into reduced reference (RR) and no reference (NR) algorithms. For NR IQA, due to the absence of the reference image information, they are usually less efficient in providing a high correlation with the subjective quality evaluations. Fortunately, the RR IQA can achieve a good tradeoff between the FR and NR algorithms, as they can predict the image quality in terms of a few features, which are extracted and transmitted from the sender to the receiver end.

Recently, many RR-IQA algorithms have been proposed by extracting global features to represent the image quality. In [3], the distance between the probability distributions of wavelet coefficients has been employed to evaluate the perceptual quality. Based on the principle that the image distortions may make the image unnatural, in [4] the authors predicted the image quality by measuring the destruction of "naturalness". In [5], the DCT coefficients are reorganized into several representative subbands and the image quality is evaluated based on the city-block distance (ROCB). In [6], the orientation selectivity based RR-IQA method was proposed. Without the reference signal, the SSIM index is predicted from the statistical features from a multiscale multiorientation divisive normalization transform [7]. In [8], the distance between the structural degradation information of the original and distorted images is used to quantify the image quality.

The free-energy principle presented in [9] and [10] formulates the internal generative mechanism (IGM) in inferring the input scenes. The underlying idea of the IGM is in accordance with the Bayesian brain hypothesis [11], which states that the brain explains their sensations according to the inherent priori knowledge. According to the philosophy that the IGM actively predicts the primary visual information and attempts to avoid the residual uncertainty/disorder, in [12], [13] the IGM based image quality methods in spatial domain were proposed. In general, a powerful model in describing the natural scene can better explain the IGM principle, such that it is curial to employ an appropriate model in the natural scene synthesis. Sparse representation is an emerging and powerful method in describing the visual signals based on the sparsity and redundancy of their representations [14]. It is also interesting to find that the primitive, or the basis in sparse representation has the properties of spatially localized, oriented and bandpass, which are closely related to the characteristics of receptive fields of simple cells [15]. As such, it is efficient in dealing with rich, varied and directional information contained in natural scene [16]–[20]. In [21], [22], the visual uncertainty in sparse domain was applied in IQA. However, the primarily visual information which is indispensable has been ignored. In [23], the entropy of primitive (EoP) was proposed to characterize the visual information in sparse domain, and a series applications [23]–[26], such as just noticeable estimation and image quality assessment have been presented.

In this work, we employ the sparse representation to quantify both the primary information and uncertainty, and thereby propose a reduced-reference image quality assessment metric. Specifically, the difference between the image signal and its best sparse description is measured as the uncertainty. The self information of the primitives, which is characterized by EoP in describing the natural scene, is employed to quantify the orderly/primary information. They are further combined to acquire the final quality score. In this metric, only one number is needed to transmitted, which has a fairly low RR data rate. As a result of this advantage, we further demonstrate its good potential in visual communications applications for quality monitoring.

The remainder of this paper is organized as follows. In Section II, we introduce the IGM and demonstrate the sparse representation in describing the natural scene. Section III elaborates the reduced reference quality assessment metric. Experimental results are given in Section IV. Finally, Section V concludes this paper.

II. IGM and Sparse Representation

This section explains the IGM in detail and introduces the relationship between IGM and sparse representation. Subsequently, the sparse representation is discussed, where the typical dictionary training and sparse coding methods are introduced.

A. Internal Generative Mechanism

The IGM [27], [28] claims that the brain can best explain the input and suppress the surprise by changing the model states. In other words, given an visual input, an minimization process is performed by fitting the internal generative model to the external sensory states. As such, for the evaluation of visual quality, both the primary information that can be explained by the internal generative model and the uncertainty play critical roles based on the assumption of IGM. In particular, the primary visual information accounts for the the information content extraction, such that distortions regarding such information may lead to unreliable image understanding. By contrast, the uncertainty indicates the unpredictable information which cannot be efficiently explained by the human visual system (HVS). As such, the combination of them provides a practical way for quality evaluation. In order to quantify the primary visual information and uncertainty, the model that best explains the input information is of sufficient interest. Here, we adopt the sparse representation, which is a poweful model in simulating the perception of natural scene. The detailed introduction of sparse representation is introduced in Section II-B.

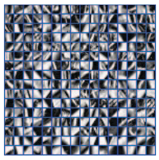

Fig. 1. Illustration of the dictionary learned from the image *Lena*.

B. Sparse Representation

Sparse representation is formulated based on the popular Sparseland model. In particular, given the input visual signal x ($x \in \mathbb{R}^n$), it is approximated by a linear combination of primitives in an over-complete dictionary. As such, this process can be formulated as $\forall x, x \approx \Psi\alpha$ and $\|\alpha\|_0 \ll n$. Here Ψ ($\Psi \in \mathbb{R}^{n\times k}$) represents the dictionary that is composed of primitives, and α ($\alpha \in \mathbb{R}^k$) denotes the coefficients for sparse representation. Moreover, $\|\cdot\|_0$ is the notation of l_0 norm. Generally speaking, the dictionary Ψ is redundant to x, such that $k > n$.

The dictionary Ψ is critical to sparse representation. To learn this dictionary, the K-SVD method [29] is adopted. In particular, K-SVD consists of two stages, including sparse representation and dictionary updating. To generate the training samples, the input image X is divided into patches $x_1, x_2, \ldots, x_i, \ldots$. Based on the assumption of the Sparseland model, the dictionary that can lead to efficient representation of the given training samples with the sparsity constraint is obtained, which can be formulated by,

$$(\Psi, \{\alpha_i\}) = \underset{\Psi,\{\alpha_i\}}{\arg\min} \sum_k \|x_i - \Psi\alpha_i\|_2{}^2, \text{s.t.} \|\alpha_i\|_0 < L. \quad (1)$$

A typical dictionary learned from the image *Lena* is given in Fig. 1. As such, given the dictionary, the sparse coefficients of the new input patch can be obtained. Here, we adopt the OMP scheme [30] to achieve the sparse representation.

III. Proposed RR-IQA Model

In this work, the sparse representation is treated as the explanation process of the input natural scene, and serves as the groundwork for the proposed RR-IQA model. More specifically, the proposed RR-IQA model is composed of two components, the primary visual information and uncertainty. The primary visual information is characterized with the EoP [23], and the uncertainty is described by the residuals between the input signal and the composite signal with sparse representation.

A. Primary Visual Information

The primary visual information can be extracted from the content that can be understood from the natural scene. This is in consistent with the design philosophy of EoP, which evaluates the visual information given the input natural image. Here, we assume that ϕ_j^i indicates how many times that the j^{th}

primitive is selected in the i^{th} iteration of the *OMP* process, and Φ_j^i denotes the total number of the j^{th} primitive selected in the previous i iterations. As such, we can have the following relationship

$$\Phi_j^i = \sum\nolimits_{t=1}^{i} \phi_j^t. \tag{2}$$

Therefore, the probability for the primitive j is given by

$$p^i\left(j\right) = \frac{\Phi_j^i}{\sum\limits_t \Phi_t^i}, \tag{3}$$

which denotes the primitive distribution in all of the previous i iterations. Given the distribution, the EoP can be computed based on the *Shannon* Theory, which is given by,

$$EoP_i(X) = -\sum_{j=1}^{k} p^i\left(j\right) \log p^i\left(j\right), \tag{4}$$

where X denotes the input image and $EoP_i(X)$ refers to the EoP of the input image X corresponding to the i iteration. In general, it is an inherent attribute of the image, indicating how much information it contains in the sparse domain. Instead of the pixel level entropy calculation, the EoP evaluates the visual information with sparse representation, based on the assumption that the interpretation process of visual information behaves like sparse coding. As such, the information in sparse domain is applied to characterize the primary information in the RR-IQA model.

B. Uncertainty Characterization

The visual uncertainty is characterized by the free energy, which characterizes the discrepancy between the input visual signal and the best interpretation of such signal [9], [10]. In [13], the free energy in spatial domain is derived, and the entropy of the difference between the input image and the autoregressive (AR) model prediction is applied to approximate the free energy, such that the visual quality can be quantified by the free energy variation. Here, we adopt the sparse representation, such that a more powerful model is applied in interpreting how HVS perceives images. In principle, free energy based on sparse representation presents a discrepancy measure between the image data and its sparse representation [21], [22]. In particular, assuming the reconstructed image after the sparse representation is $\bar{X}$, the free energy is characterized by the entropy of the difference between the original and reconstructed signal, which is given by

$$F_i(X) = E(X - \bar{X}_i) \tag{5}$$

where the function E denotes the calculation of the entropy. Again, i denotes the index for sparse representation iteration. As such, when treating the sparse representation as the image interpretation process, the uncertainty in sparse domain can be computed after sparse representation.

C. Proposed IQA Model

The proposed IQA model combines the primary information and uncertainty in sparse domain together. Assuming the original image is X and the distorted version at the receiver end is Y, for RR-IQA, the features extracted from X are transmitted via a channel and compared with those of Y. In particular, the feature is calculated as the combination of the primary visual information and the amount of uncertainty. As such, the RR-IQA model is given by

$$Q = EoP_i(X) \cdot F_i(X) - EoP_i(Y) \cdot F_i(Y). \tag{6}$$

For practical implementations, the images X and Y are downsampled for the high efficiency sparse coding. Moreover, the parameter i is set to be 5, which approximately lies in the saturation range of EoP [23]. Therefore, only one feature is required to be transmitted to the receiver side. This has greatly reduced the transmission overhead.

IV. Experimental Results

The performance of the proposed scheme is firstly validated via a popular single distortion IQA database-LIVE database [31]. Five typical distortions are included in the database, including JPEG2000, JPEG, White Noise, Gaussian Blur and Fastfading. To access the IQA accuracy, five evaluation criterions are adopted, including Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SRCC), mean absolute error (MAE), Root mean-squared error (RMSE), and Kendall's rank correlation coefficient (KRCC). In particular, higher PLCC, SRCC and KRCC but lower RMSE and MAE values indicate better IQA metric and vice-versa. Moreover, four RR-IQA and two FR-IQA methods are used for comparisons. The RR-IQA metrics include the RRVIF [12], OSVP [6], ROCB [5] and WNISM [3], and the FR-IQA metrics are PSNR and SSIM [32]. The performance comparisons are listed in Table I. It is shown that our scheme is superior to these state-of-the-art RR-IQA metrics. Moreover, it is also worth noting that very low bit rate features are required to be transmitted, such that negligible transmission overhead is introduced. As such, we can conclude that the proposed scheme can achieve promising quality prediction accuracy with the highly compact feature. Moreover, from the experimental results, we can see that the proposed scheme is superior to PSNR, and achieves close performance to SSIM. This further provides useful evidence that our proposed scheme can achieve satisfactory performance with very low bit rate for practical RR-IQA.

Moreover, we evaluate the multiple distortion scenario when the image is firstly blurred and then JPEG compressed. This is a practical case in image acquisition, as the captured image may suffer multiple distortions simultaneously. The performance comparisons are shown in Table II. Again, we can see that the performance of the proposed scheme is superior to these RR-IQA models. As such, our proposed method is not only efficient in dealing with single distorted image, but is also effective at multiply distorted ones.

TABLE I
PERFORMANCE EVALUATION BASED ON LIVE IMAGE DATASET.

	PLCC	SRCC	KRCC	RMSE	MAE
SSIM	0.9042	0.9104	0.7311	11.669	9.228
PSNR	0.8723	0.8756	0.6865	13.360	10.509
RRVIF	0.7543	0.7246	0.5438	17.937	13.675
OSVP	0.8201	0.8218	0.6275	15.633	12.261
ROCB	0.8866	0.8822	0.6966	12.636	9.853
WNISM	0.7512	0.7599	0.5697	18.035	14.020
Proposed	**0.9146**	**0.9157**	**0.7418**	**11.050**	**8.731**

TABLE II
PERFORMANCE EVALUATION OF MULTIPLY DISTORTED IMAGES (BLUR PLUS JPEG).

	PLCC	SRCC	KRCC	RMSE	MAE
SSIM	0.8003	0.7443	0.5430	11.489	9.167
PSNR	0.7429	0.6621	0.4775	12.826	10.417
RRVIF	0.8297	0.7218	0.5114	10.695	8.647
OSVP	0.8584	0.7976	0.5880	9.830	7.874
ROCB	0.8516	0.7708	0.5691	10.043	7.907
WNISM	0.8346	0.7683	0.5624	10.554	8.422
Proposed	**0.8836**	**0.8474**	**0.6441**	**8.971**	**7.188**

V. CONCLUSION

We have specifically developed a RR-IQA model in sparse domain that can automatically predict the image quality based on the philosophy of IGM. The features obtained based on the IGM principle, which characterize the primary visual information and uncertainty are combined to evaluate the visual quality at the receiver end. The proposed scheme enjoys low RR transmission rate, which can be regarded as negligible in image transmission. Experimental results show that the proposed method is well correlated with subjective evaluations of single and multiple distorted images.

ACKNOWLEDGMENT

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2015AA015903) National Natural Science Foundation of China (61632001, 61571017, 61421062), and the Top-Notch Young Talents Program of China, which are gratefully acknowledged.

REFERENCES

[1] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[2] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on ssim-inspired divisive normalization," *Image Processing, IEEE Transactions on*, vol. 22, no. 4, pp. 1418–1429, 2013.

[3] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Electronic Imaging 2005*, 2005, pp. 149–159.

[4] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1680–1689, 2006.

[5] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized dct-based image representation," *Multimedia, IEEE Transactions on*, vol. 13, no. 4, pp. 824–829, 2011.

[6] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18–29, 2016.

[7] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3378–3389, 2012.

[8] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *ISCAS*. IEEE, 2013, pp. 1095–1098.

[9] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology-Paris*, vol. 100, no. 1, pp. 70–87, 2006.

[10] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[11] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *TRENDS in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.

[12] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.

[13] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 41–52, 2012.

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

[15] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[16] A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.* Springer, 2009.

[17] X. Zhang, W. Lin, S. Ma, S. Wang, and W. Gao, "Rate-distortion based sparse coding for image set compression," in *Visual Communications and Image Processing (VCIP), 2015*. IEEE, 2015, pp. 1–4.

[18] X. Zhang, W. Lin, R. Xiong, X. Liu, S. Ma, and W. Gao, "Low-rank decomposition-based restoration of compressed images via adaptive noise estimation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4158–4171, 2016.

[19] L. Xie, X. Zhang, S. Wang, X. Zhang, and S. Ma, "Quality assessment of tone-mapped images based on sparse representation," in *ISCAS*. IEEE, 2016, pp. 2218–2221.

[20] S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, "Nonlocal in-loop filter: The way toward next-generation video coding?" *IEEE MultiMedia*, vol. 23, no. 2, pp. 16–26, 2016.

[21] Y. Liu, G. Zhai, X. Liu, and D. Zhao, "Perceptual image quality assessment combining free-energy principle and sparse representation," in *ISCAS*. IEEE, 2016, pp. 1586–1589.

[22] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Transactions on Multimedia*, 2017.

[23] S. Ma, X. Zhang, S. Wang, J. Zhang, H. Sun, and W. Gao, "Entropy of primitive: From sparse representation to visual information evaluation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 2, pp. 249–260, 2017.

[24] W. Shi, F. Jiang, and D. Zhao, "Image entropy of primitive and visual quality assessment," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2087–2091.

[25] Z. Wan, Y. Liu, F. Qi, and D. Zhao, "Reduced reference image quality assessment based on entropy of classified primitives," in *Data Compression Conference (DCC), 2017*. IEEE, 2017, pp. 231–240.

[26] S. Wang, X. Zhang, S. Ma, and W. Gao, "Reduced reference image quality assessment using entropy of primitives," in *Picture Coding Symposium (PCS), 2013*, Conference Proceedings, pp. 193–196.

[27] K. J. Friston, J. Daunizeau, and S. J. Kiebel, "Reinforcement learning or active inference?" *PloS one*, vol. 4, no. 7, p. e6421, 2009.

[28] R. L. Gregory, "Perceptions as hypotheses," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 290, no. 1038, pp. 181–197, 1980.

[29] M. Aharon, M. Elad, and A. Bruckstein, "KSVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[30] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[31] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2005.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.