Perceptual Image Quality Assessment Combining Free-energy Principle and Sparse Representation

Yutao Liu[1], *Guangtao Zhai*[2], *Xianming Liu*[1] *and Debin Zhao*[1]
1.Department of Computer Science, Harbin Institute of Technology, Harbin, China
2.Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China
{liuyutao2008, zhaiguangtao, xmliu.hit}@gmail.com, dbzhao@hit.edu.cn

***Abstract*—Since the purpose of objective image quality assessment is to be consistent with subjective image quality assessment as highly as possible, the understanding of the mechanisms of human visual system will certainly benefit the study of objective image quality assessment. Recent developments in brain theory and neuroscience, particularly the free-energy principle, account for the perception and understanding of visual scenes. As the free-energy principle conjectures, the brain tries to generate the corresponding prediction for its encountered scene by an internal generative model. On the other hand, sparse representation is evidenced to resemble the neural response properties of simple cells in the primary visual cortex. Conjunctively, in this paper, we suppose the prediction manner of the internal generative model in free-energy principle follows sparse representation and propose an image quality metric accordingly. Experiments on LIVE, TID2008 and CSIQ image databases demonstrate the effectiveness of the proposed image quality metric. Noteworthily, our metric needs little information (only a single scalar) of the reference image and is training-free.**

I. Introduction

Nowadays, a large number of digital images enter people's life and become the important way of transferring and exchanging information. Nevertheless, the quality of these images is hard to guarantee. That is, the images often suffer from different kinds of distortions, such as noise, blurring, blockiness etc., which all degrade the image quality. Therefore, it is necessary to acquire the image quality for image applications. To this end, image quality assessment (IQA) is concerned and deeply studied in recent years. Generally, existing IQA methods fall into two categories, subjective assessment and objective assessment. Subjective assessment refers to evaluating the image quality from the subjective judgements of humans. This is the most reliable way of assessing the image quality since human beings are usually the ultimate receivers in image processing applications. However, subjective assessment is always expensive, cumbersome and hard to be embedded into real-time systems. Therefore, researchers endeavor to explore objective IQA methods. According to the availability of reference image, the objective assessment methods can be classified into full reference (FR), reduced-reference (RR) and no-reference (NR). For FR methods, the pristine or distortion-free image is referred when assessing the image quality, while this condition is rather ideal in practice. For RR methods, only partial information of the pristine image is needed, while NR methods can evaluate the image quality without any information of pristine images, which is closest to reality as there're actually no corresponding reference when we view images.

Past years have seen a lot of outstanding objective IQA methods which greatly promote the study of IQA. For FR methods, the mean-squared error (MSE) and its relevant peak signal-to-noise ratio (PSNR) were most popular owing to their low computational cost, high portability and clear physical meaning. While they were also found poorly consistent with subjective evaluation for the image quality. Hence, Wang et al. proposed the structural similarity index (SSIM) [1], which is based on the hypothesis that the human visual system (HVS) is highly adapted to extract the structural information from the visual scene. Therefore, measuring the structural similarity between the reference image and distorted image can provide a good estimation of the image quality. Later, some other superior FR IQA models [2] [3] were presented from different perspectives. For RR IQA, partial information of pristine/reference image is used in quality assessment. In [4], the entropies of wavelet coefficients of the reference image were compared with that of the distorted image to measure image quality. While in most cases of real applications, the reference image is often absent or unavailable. On this occasion, NR IQA is the only way to obtain the image quality. Usually, NR methods follow two stages, feature extraction and mapping process which maps the feature vectors onto the image quality level [5].

While the purpose of objective IQA is to mimic subjective IQA, consequently, the mechanisms of HVS should be introduced in designing the objective IQA algorithms. From this point, Zhai et al. in [6] proposed a new psychovisual image quality metric (FEDM) based on the free-energy principle in brain theory and neuroscience. In free-energy principle, the perception and understanding of an image is modeled as an active inference process, in which the brain tries to explain the image using an internal generative model. With this model, the brain generates predictions of those encountered scenes. However, there exists a discrepancy between the scene and its prediction and this discrepancy should be related to the quality of perceptions. For computational simplicity, the linear AR model was chosen to simulate the internal generative model in [6], while this lacks necessary considering of visual processing mechanism of the HVS. Therefore, a model that resemble perception behavior is needed to better simulate the internal generative model. As stated in [7], the receptive fields (RFs) of simple cells in mammalian primary visual cortex can be characterized as being spatially localized, oriented and bandpass. One approach to understanding such response properties of visual neurons is to investigate the relationship between the properties and the statistical structure of natural

978-1-4799-5341-7/16/$31.00 ©2016 IEEE

images in terms of efficient coding. Then the authors of [7] accounted that sparse representation for natural scenes earned similar results to those found in the primary visual cortex. Inspired by this, in this paper, we suppose the prediction manner of visual scenes by the internal generative model follows sparse representation which states that a signal can be represented by a linear combination of a small number of atoms in a dictionary. Similarly, the discrepancy between the visual scene and its sparse representation indicates the perceptual quality of the visual scene. Based on these analysis, we propose a perceptual image quality metric by combining the free-energy principle and sparse representation. The proposed quality metric belongs to RR IQA methods actually because partial information of the reference image is needed for quality estimation. However, the needed information of our method can be modeled as a single scalar (entropy of the prediction residuals) extracted from the reference image, which is negligible compared to the size of image data. Under this considering, our metric can be approximately regarded as an NR method. Noteworthily, our method is training-free which has better universality than the training-based methods. Experiments on LIVE, TID2008 and CSIQ image databases confirm the effectiveness of our image quality metric.

The remainder of this paper is organized as follows: Section II introduces the concept of the free-energy principle. Section III details the proposed perceptual image quality metric. Experimental results and analysis are presented in Section IV. Finally, we conclude this paper in Section V.

II. The Free-energy Principle

Our method is under the guidance of the free-energy principle. Therefore, we will specify the free-energy principle at first. As we mentioned before, the fundamental assumption in free-energy principle is that the cognitive process is governed by an internal generative model in the brain. With the model, the brain is able to actively infer predictions of meaningful information from visual scenes and reduce residual uncertainty at the meantime.

For operational amenability, it is assumed that the internal generative model $\mathcal{G}$ for visual perception is parametric, which explains visual scenes by adjusting the parameter vector $\mathbf{g}$. Specifically, given an image, its 'surprise' can be calculated by integrating the joint distribution $P(I,\mathbf{g})$ over the space of the internal model parameters $\mathbf{g}$ as:

$$-\log P(I) = -\log \int P(I,\mathbf{g}) d\mathbf{g}. \tag{1}$$

Then a dummy term $Q(\mathbf{g}|I)$ is integrated into both the denominator and numerator of the right part of equation (1) as follows:

$$-\log P(I) = -\log \int Q(\mathbf{g}|I) \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{2}$$

Here, $Q(\mathbf{g}|I)$ is an auxiliary posterior distribution of the model parameters given the image, It can be thought of as an approximate posterior to the true posterior of the model parameters $P(\mathbf{g}|I)$ calculated by the brain. The brain minimizes the discrepancy between the approximate posterior $Q(\mathbf{g}|I)$ and the true posterior $P(\mathbf{g}|I)$. Through Jensen's inequality, equation (2) changes to:

$$-\log P(I) \leq -\int Q(\mathbf{g}|I) \log \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{3}$$

Afterwards, the right side of equation (3) is defined as the free energy as follows:

$$F(\mathbf{g}) = -\int Q(\mathbf{g}|I) \log \frac{P(I,\mathbf{g})}{Q(\mathbf{g}|I)} d\mathbf{g}. \tag{4}$$

Obviously, the free energy defines an upper bound of 'surprise' for the given image I. Rearranging (4), we obtain:

$$\begin{aligned} F(\mathbf{g}) &= \int Q(\mathbf{g}|I) \log \frac{Q(\mathbf{g}|I)}{P(I,\mathbf{g})} d\mathbf{g} \\ &= \int Q(\mathbf{g}|I) \log Q(\mathbf{g}|I) d\mathbf{g} - \int Q(\mathbf{g}|I) \log P(I,\mathbf{g}) d\mathbf{g} \\ &= E_Q[-\log P(I,\mathbf{g})] - E_Q[-\log Q(\mathbf{g}|I)], \end{aligned} \tag{5}$$

which expresses the free energy as energy minus entropy, the first term represents taking the expectation of Gibbs free energy of the system that contains the image I and the model parameters $\mathbf{g}$, and the second term is the entropy of the approximate posterior density. For intuitive understanding, with $P(I,\mathbf{g}) = P(\mathbf{g}|I)P(I)$, equation (5) can be transferred to:

$$\begin{aligned} F(\mathbf{g}) &= \int Q(\mathbf{g}|I) \log \frac{Q(\mathbf{g}|I)}{P(\mathbf{g}|I)P(I)} d\mathbf{g} \\ &= -\log P(I) + \int Q(\mathbf{g}|I) \log \frac{Q(\mathbf{g}|I)}{P(\mathbf{g}|I)} d\mathbf{g} \\ &= -\log P(I) + \mathbf{KL}(Q(\mathbf{g}|I) \| P(\mathbf{g}|I)), \end{aligned} \tag{6}$$

where $\mathbf{KL}(\cdot)$ refers to the Kullback-Leibler divergence between the approximate posterior and the true posterior distributions and it's nonnegative. It is clearly seen that the free energy $F(\mathbf{g})$ is greater than or equal to the image 'surprise' -log$P(I)$. The brain tries to lower the divergence $\mathbf{KL}(Q(\mathbf{g}|I) \| P(\mathbf{g}|I))$ between the approximate posterior and its true posterior distributions when perceiving the image I. More details about the free-energy principle can be found in [6].

III. Perceptual Image Quality Metric

In the work of [6], the linear AR model was adopted to simulate the internal model which generates predictions of the images due to its computational simplicity. Consequentially, the visual processing mechanism of HVS is underestimated with AR prediction. Therefore, in this paper, we resort to other prediction method that is similar to the perception behavior of the visual system. As mentioned before, sparse representation for visual scenes resembles the neural response properties of simple cells in the primary visual cortex, this inspires us to suppose that the prediction manner of the internal generative model in free-energy principle follows sparse representation.

A. Sparse Representation

Sparse representation refers to representing a signal with a linear combination of a small number of atoms from a predefined or trained dictionary [8]. Specifically, given a signal $\mathbf{y} \in \mathbb{R}^n$ with an overcomplete dictionary matrix

$\mathbf{D} \in \mathbb{R}^{n \times K}$ that contains K columns, each column represents one prototype atom. Then the dictionary $\mathbf{D}$ can be denoted as $[\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3 ... \mathbf{d}_K]$. The signal $\mathbf{y}$ is represented as a sparse linear combination of the atoms in $\mathbf{D}$ as:

$$\mathbf{y} = \mathbf{D}\mathbf{x}, \tag{7}$$

or approximately represented as:

$$\mathbf{y} \approx \mathbf{D}\mathbf{x} \quad s.t. \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_p \leq \xi, \tag{8}$$

where $\mathbf{x} \in \mathbb{R}^K$ represents the vector that contains the representation coefficients. $\|\cdot\|_p$ is the l^p norm. What we concerned is finding fewest number of nonzero coefficients to represent the signal $\mathbf{y}$, namely requesting for the sparsest representation:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = \mathbf{D}\mathbf{x}, \tag{9}$$

where $\|\cdot\|_0$ is the l^0 norm, meaning the number of nonzero elements of a vector. However, l^0-minimization is an NP-hard problem, one approach is applying "pursuit algorithm" to find an approximate solution. Another alternative solution is to replace l^0 norm with l^1 norm and minimize the l^1 norm as:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \quad s.t. \quad \mathbf{y} = \mathbf{D}\mathbf{x}, \tag{10}$$

This equation can be further turned into an unconstrained optimization problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_1, \tag{11}$$

where λ is a positive constant balancing the importance of the reconstruction fidelity term and the sparse constraint term. This unconstrained optimization problem can be solved by iterative shrinkage/thresholding algorithm [9]. With the obtained coefficient vector $\mathbf{x}$ and the predefined dictionary $\mathbf{D}$, we can get the sparse representation of signal $\mathbf{y}$ accordingly.

B. The Perceptual Image Quality Index

The free-energy principle points out that the brain tends to generate predictions of external images through an internal generative model. While the prediction/representation of the input image can't reach the image itself. In other words, there indeed exists a discrepancy between the image and its model-predicted version and this discrepancy is believed to be closely related to the perceptual quality of the image. As defined in equation (4), free energy presents a discrepancy measure between the image data and its prediction. Therefore, we can measure the quality of the distorted image by checking the variance of its free energy. In other words, here free energy can be viewed as a quality-connected feature extracted from the reference image and distorted image respectively and we can acquire the quality of the distorted image through feature comparison. Based on these analysis, we denote the absolute difference between the reference image r and its distorted version d in free energy as the image quality index Q:

$$Q = |F(\mathbf{g}_d) - F(\mathbf{g}_r)|. \tag{12}$$

For intuitive observation, we show the computational process of the proposed image quality metric in Fig. 1. As can be seen, the reference image and the distorted image are firstly sparsely represented to get the predicted reference and distorted images

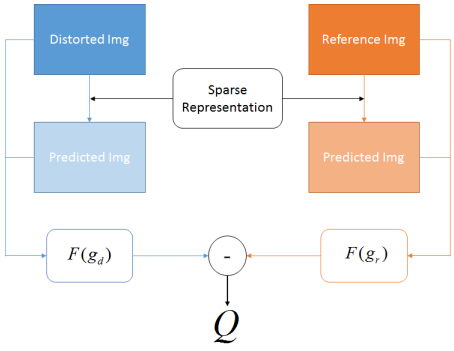

Fig. 1. The computational process of the proposed image quality metric.

respectively. This step is to simulate the prediction process of the internal generative model in the brain. The sparse representation follows the procedures described in III-A and the detailed configurations of sparse representation are given in section IV. After obtaining the predicted images, the free energies of the reference image and the distorted image are computed separately. The entropy of the residuals between the image and its predicted version is employed for the computation of free energy, as it can measure the uncertainty of the discrepancy between image and its predicted version in a simple way. At last, we take the absolute difference of the two free energies as the final image quality metric Q.

It should be noted that the lower Q is, the closer the perceptual quality is between the distorted image and its reference/distortion-free image, which means the perceptual quality of the distorted image is higher. Otherwise, if Q is higher, the perceptual quality of the reference image and its corresponding distorted image diverges more, which indicates the distorted image has poorer image quality. Summarily speaking, the lower Q is, the higher the quality of the distorted image is. Undoubtedly, the proposed quality index belongs to RR IQA method as partial information (the free energy) of the reference image is needed when calculating Q. However, here the free energy is simplistically represented with a single scalar (the entropy of the predicted residuals) which is negligible to the image data. From this point, our metric can be approximately regarded as a NR method.

IV. Experimental Results and Analysis

A. Experiment Configurations

The experiment configurations concentrate on the sparse representation for images. Specifically, we divide the image into 8×8 non-overlapped patches. Each patch is vectorized as the signal $\mathbf{y}$ and we seek its sparsest representation according to III-A described. The overcomplete DCT dictionary is employed as the predefined dictionary $\mathbf{D}$, the size of $\mathbf{D}$ is 64×128 with totally 128 atoms available for representing each patch. We restrict the maximum number of nonzero coefficients for representing each patch to 20. The orthogonal matching pursuit (OMP) algorithm [10] is utilized to find the representation coefficients.

B. Experimental Results and Comparison

We test the proposed image quality metric on three widely-used image databases, LIVE [11], TID2008 [12], CSIQ [13]

TABLE I. SROCC VALUES ON LIVE

Methods	**JP2K**	**JPEG**	**WN**	**Blur**	**FF**
PSNR	0.8898	0.8409	0.9853	0.7816	0.8903
SSIM	0.9528	0.9116	0.9694	0.9516	0.9553
BIQI	0.9187	0.8886	0.9903	0.9543	0.8205
NIQE	0.8977	0.8661	0.9716	0.9329	0.8644
DIIVINE	0.9025	0.7511	0.9878	0.9584	0.8592
FEDM	0.9145	0.8543	0.9153	0.7594	0.8230
Our's	0.9299	0.8875	0.9089	0.8948	0.8479

TABLE II. SROCC VALUES ON TID2008

Methods	**JP2K**	**JPEG**	**WN**	**Blur**
PSNR	0.8300	0.9011	0.9115	0.8682
SSIM	0.9723	0.9270	0.8310	0.9596
BIQI	0.6940	0.8443	0.7386	0.7468
NIQE	0.8964	0.8608	0.7797	0.8165
DIIVINE	0.8525	0.6309	0.8085	0.8237
FEDM	0.8162	0.7594	0.6855	0.7980
Our's	0.8991	0.8112	0.6552	0.9163

with the frequently encountered distortion types, JP2k, JPEG, Gaussian White Noise (WN), Blur and Fast fading (FF). The Spearman rank order correlation coefficient (SROCC) is calculated between the objective results given by the image quality metric and the subjective evaluation scores. The higher SROCC value means the objective metric is more consistent with subjective evaluation and the objective metric is superior. We compare our metric with a number of representative IQA metrics. Among them, PSNR, SSIM [1] are the two most influential FR methods. As our metric can be approximately regarded as a NR method, the classical NR methods, BIQI [14], DIIVINE [5], NIQE [15] are also included for comparison. BIQI and DIIVINE are training-based methods, while NIQE demands a prediction model trained from natural images. In addition, the free-energy based RR method FEDM [6] is also compared. We list the experimental results clearly in Table I, II and III. From the tables, our method exceeds FEDM in most cases which verifies that sparse representation is closer to the perception behavior than AR model and more reasonable to simulate the internal generative model as we supposed. With compared to the FR methods, our metric performs better than PSNR in some distortion types, while inferior to SSIM. Noted that our method needs just a number from the reference image, while the FR methods require the whole reference image for quality assessment. Though our method is training-free, it is still comparable even better than the competing NR methods over the three databases.

V. CONCLUSION

In this paper, we have proposed a perceptual image quality metric by combining the free-energy principle and sparse representation. The proposed metric was inspired by the exploration of the perception mechanism and information representation of the brain. In addition, it's conveniently computed and demands only a number from the reference image which can be embedded into the header file. Therefore, it's practical for image applications to measure the image quality. Experimental results confirmed the effectiveness of the proposed metric.

TABLE III. SROCC VALUES ON CSIQ

Methods	**JP2K**	**JPEG**	**WN**	**Blur**
PSNR	0.9361	0.8879	0.9363	0.9291
SSIM	0.9605	0.9543	0.8974	0.9609
BIQI	0.7084	0.8673	0.8794	0.7713
NIQE	0.9062	0.8832	0.8097	0.8945
DIIVINE	0.8304	0.7998	0.8662	0.8716
FEDM	0.8945	0.9166	0.8246	0.8522
Our's	0.9028	0.8886	0.8307	0.8909

ACKNOWLEDGEMENT

This work is supported by the Major State Basic Research Development Program of China (973 Program 2015CB351804), the National Science Foundation of China under Grants 61300110.

REFERENCES

[1] 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.

[2] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: a feature similarity index for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378–2386, 2011.

[4] R. Soundararajan and A. C. Bovik, "Rred indices: Reduced reference entropic differencing for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 517–526, 2012.

[5] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3350–3364, 2011.

[6] 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.

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

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

[9] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[10] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[11] L. C. H. R. Sheikh, Z. Wang and A. C. Bovik, "Live image quality assessment database release 2," 2006.

[12] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[13] E. C. Larson and D. Chandler, "Categorical image quality (csiq) database," *Online, http://vision. okstate. edu/csiq*, 2010.

[14] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 513–516, 2010.

[15] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 209–212, 2013.