Digital Signal Processing ••• (••••) •••–•••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Free-energy principle inspired visual quality assessment: An overview

Guangtao Zhai *, Xiongkuo Min, Ning Liu

Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Free energy principle
Image quality assessment
Generative model
Visual quality
Image processing

ABSTRACT

The free energy principle was proposed several years ago as a unified justification for some brain theories. The process of human perception, cognition, action, and learning can be well explained using the free energy principle. The free energy principle suggests that the human perception and understanding of a given scene can be modeled as an active inference process, and the human brain tries to explain the scene using an internal generative model. The discrepancy between the given image or view and its best internal generative model explainable part is upper bounded by the free energy of the inference process. It was then conjectured that perceptual quality of the input image is closely related to free energy value of the process. Following this framework, dozens of visual quality assessment techniques have been proposed in the last few years and many have achieved state of the art performance. In this paper, we first give an overview of the free energy principle and then review the free energy principle inspired visual quality assessment metrics with a comparison in terms of algorithm design and performance.

1. Introduction

Vision is the most important information source for human beings. Human visual system (HVS) allows us to perceive the outside world by using the light induced retinal stimulation. However, it should be noted that the retinal stimulation is by no means what we finally see, since our brain works a series of psychological inference on the inputs. Hermann von Helmholtz pointed out in 1860s that "vision is an outcome of inferences, it is a process of making assumptions and drawing conclusions from (partial) sensory data" [1,2]. It is now widely believed by the researchers in psychology, cognitive science and neuroscience that the physiological and psychological mechanisms of perception in human brain are intrinsic interactions between the retinal stimuli and human vision. In the fields of perceptual quality assessment, it was realized that the design of image quality metrics with purely signal processing approach, i.e. without consideration of the intrinsic interactions, is fundamentally limited in its resembling of the working mechanism of the HVS [3].

In order to model the active inference process in the HVS, we resort to the free energy principle [4,5] which was proposed as a unification of several brain theories about human perception, learning and action. The second law of thermodynamics asserts that an isolated system tends to a disorder state and its entropy will increase over time, until reaching a maximum value at the equilibrium state. However, all biological agents try to maintain their internal states at low entropy level, as a prerequisite of being alive. The free energy principle suggests that the living agents achieve this goal by minimizing the free energy of the process, which turns out to be an upper bound of the total "surprise" encountered in different environments [4].

The free energy principle points out that free energy can be evaluated by a biological agent using its internal generative model and external sensory inputs. For visual perception, the brain gives predictions of those encountered scenes in a constructive manner with the internal generative model. The generative model can be decomposed into likelihood multiplies a prior and visual perception is therefore a process of inverting the likelihood towards the posterior possibilities of the given scene. Since a brain cannot possibly hold a universal generative model for all visual scenes, there will always be a discrepancy between the external visual input and its generative model explainable part. And the psychovisual quality of a scene can be defined using the agreement between the scene and its predicted version using the internal generative model that best describes the scene. Studies show that the primal visual system processes the visual inputs in multi-channel multi-resolution manner [6]. And many successful quality assessment algorithms are based on HVS plausible decomposition of visual signals [7]. It was argued that the free energy principle inspired visual quality metric differs fundamentally from this decomposition based methods and works in a "synthesis" way with visual signal inputs [3].

* Corresponding author.
E-mail addresses: zhaiguangtao@sjtu.edu.cn (G. Zhai), minxiongkuo@sjtu.edu.cn (X. Min).

https://doi.org/10.1016/j.dsp.2019.02.017
1051-2004/

A key task to the design of free energy principle inspired visual quality metric, is the construction of the internal generative model. A good generative model should be able to approximate any given visual input with high precision, mimicing the working mechanism of the human brain. Following a pioneer work in [3], many algorithms used linear autoregressive (AR) model as an approximation of the optimal internal generative model. For example, AR model has been used as the internal generative model in full-reference (FR) quality metrics [8,9], reduced-reference (RR) metrics [10,11] no-reference (NR) metrics [12,13] and comparative quality assessment [14]. However, it is noticed that under most circumstances, the AR model used in those quality metrics is not "optimal" in the statistical sense because the model order and its 2D shape is fixed. A more favorable AR model [15] can be designed using the minimum description length principle [16]. Dictionary learning and sparse representation was proposed as alternatives for improved accuracy [17–19].

The internal generative model predicted image is generally considered as the "ordered" portion of the scene in the sense that an "order" is inherently or explicitly depicted by the model. Meanwhile, the discrepancy between the model explainable part and the input, i.e. the model prediction residual is deemed as the "disordered" portion of the scene. In the view of dividing the image into ordered- and disordered-parts, the idea of computing the model prediction and discrepancy can also be fulfilled without using an explicit model. For instance, a bilateral filter [20] can be used to separate image into two layers with plain (ordered) and detailed (disordered) information [21–24].

The rest of the paper is organized as follows. In Section 2, we review the free energy principle and relate it to psychovisual quality assessment. The free energy principle inspired visual quality assessment algorithms are surveyed and compared in Section 3. Finally, Section 4 concludes the paper with some discussions.

2. Free energy and psychovisual quality assessment

2.1. The free energy principle [3]

As introduced in Section 1, the internal generative model is of crucial importance to the implementation of the free energy principle. We defined the parametric internal generative model as $\mathcal{G}$, which explains perceived scenes by adjusting the vector $\boldsymbol{\theta}$ of parameters. Given a visual stimuli (an image) I, the 'surprise' can be computed by integrating over the space of model parameters $\boldsymbol{\theta}$ the joint distribution $P(I,\boldsymbol{\theta}|\mathcal{G})$

$$-\log P(I|\mathcal{G}) = -\log \int P(I,\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta}. \tag{1}$$

Since the joint distribution of the parameter and the image $P(I,\boldsymbol{\theta}|\mathcal{G})$ is beyond our current knowledge of how our brain is working, we introduce an auxiliary posterior term of the model parameters given the image $Q(\boldsymbol{\theta}|I)$ into both the denominator and numerator in (1) and have

$$-\log P(I|\mathcal{G}) = -\log \int Q(\boldsymbol{\theta}|I)\frac{P(I,\boldsymbol{\theta}|\mathcal{G})}{Q(\boldsymbol{\theta}|I)}d\boldsymbol{\theta}. \tag{2}$$

$Q(\boldsymbol{\theta}|I)$ is an approximate to the true posterior of the model parameters $P(\boldsymbol{\theta}|I,\mathcal{G})$ that can be calculated by the brain. The brain minimizes the difference between the approximate posterior $Q(\boldsymbol{\theta}|I,\mathcal{G})$ and the true posterior $P(\boldsymbol{\theta}|I,\mathcal{G})$ by adjusting the parameters $\boldsymbol{\theta}$ to best explain I. $Q(\boldsymbol{\theta}|I)$ is also called the recognition density in Bayesian brain theory [25]. See [26–29] for in-depth analysis and examples.

In our analysis below, the dependency on the generative model $\mathcal{G}$ is dropped for simplicity. Note that in (2), the negative 'surprise' $\log P(I|\mathcal{G})$ is the log-evidence of the image data I given the model. So, the minimization of surprise equals the maximization of model evidence. Using Jensen's inequality, from (2) we have

$$-\log P(I) \le -\int Q(\boldsymbol{\theta}|I)\log\frac{P(I,\boldsymbol{\theta})}{Q(\boldsymbol{\theta}|I)}d\boldsymbol{\theta}. \tag{3}$$

Following the tradition in statistical physics and thermodynamics [30], the right hand side of (3) is defined as the free energy

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

Obviously, since $-\log P(I) \le F(\boldsymbol{\theta})$, free energy $F(\boldsymbol{\theta})$ defines an upper bound of the 'surprise' for the image data I. Rearranging (4) we have

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

which expresses the free energy as *energy* minus *entropy*. $-\log P(I,\boldsymbol{\theta})$ defines the Gibbs free energy of the system consisting of the image data I and the generative model $\mathcal{G}$ (manifested through model parameters $\boldsymbol{\theta}$). And $E_Q[-\log P(I,\boldsymbol{\theta})]$ gives energy term's expectation over different approximate recognition densities. $E_Q[-\log Q(\boldsymbol{\theta}|I)]$ is simply the entropy of the recognition density.

In practice, since (4) and (5) are not directly computable because the integration over joint distribution $P(I,\boldsymbol{\theta})$ can be intractably complex. Noticing that $P(I,\boldsymbol{\theta}) = P(\boldsymbol{\theta}|I)P(I)$, (4) can be rewritten as

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

According to Gibbs' inequality, the Kullback-Leibler divergence between the recognition posterior and true posterior parameter distribution in (6) is non-negative, i.e. $KL(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}|I)) \ge 0$, with equality if and only if $Q(\boldsymbol{\theta}|I) = P(\boldsymbol{\theta}|I)$. Here the free energy $F(\boldsymbol{\theta})$ defines a strict upper bound of the 'surprise' or negative model evidence. As indicated in (6), for fixed image data I, the free energy is suppressed by minimizing the divergence term. In other words, the brain tries to lower the divergence $KL(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}|I))$ between the approximate recognition density of model parameters and its true posterior density when perceiving a given scene.

Alternatively, we may let $P(I,\boldsymbol{\theta}) = P(I|\boldsymbol{\theta})P(\boldsymbol{\theta})$ in (4) and have

$$\begin{aligned} F(\boldsymbol{\theta}) &= \int Q(\boldsymbol{\theta}|I)\log\frac{Q(\boldsymbol{\theta}|I)}{P(I|\boldsymbol{\theta})P(\boldsymbol{\theta})}d\boldsymbol{\theta} \\ &= \int Q(\boldsymbol{\theta}|I)\log\frac{Q(\boldsymbol{\theta}|I)}{P(\boldsymbol{\theta})}d\boldsymbol{\theta} \\ &\quad -\int Q(\boldsymbol{\theta}|I)\log P(I|\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= KL(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta})) + E_Q[\log P(I|\boldsymbol{\theta})] \end{aligned} \tag{7}$$

where $KL(Q(\boldsymbol{\theta}|I)\|P(\boldsymbol{\theta}))$ measures the distance between the recognition density and the true prior density of the model parameters, and it attains zeros only when $Q(\boldsymbol{\theta}|I) = P(\boldsymbol{\theta})$. $E_Q[\log P(I|\boldsymbol{\theta})]$ measures the average likelihood of the data over the approximating posterior density.

Please cite this article in press as: G. Zhai et al., Free-energy principle inspired visual quality assessment: An overview, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.02.017

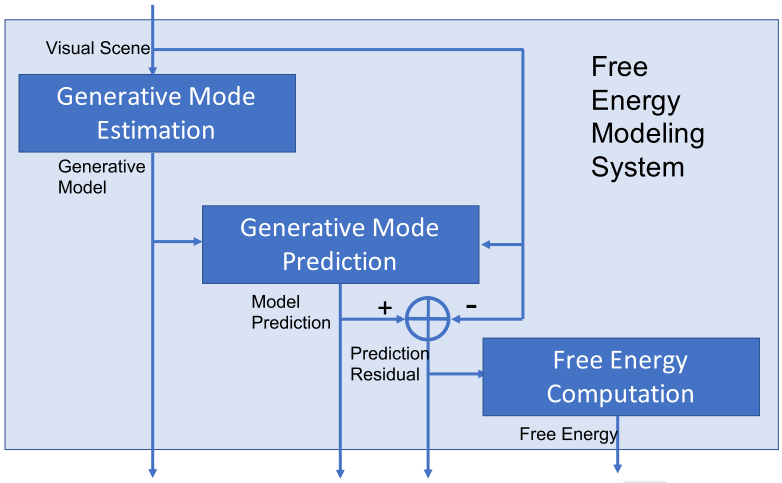

Fig. 1. A general framework of free energy modeling.

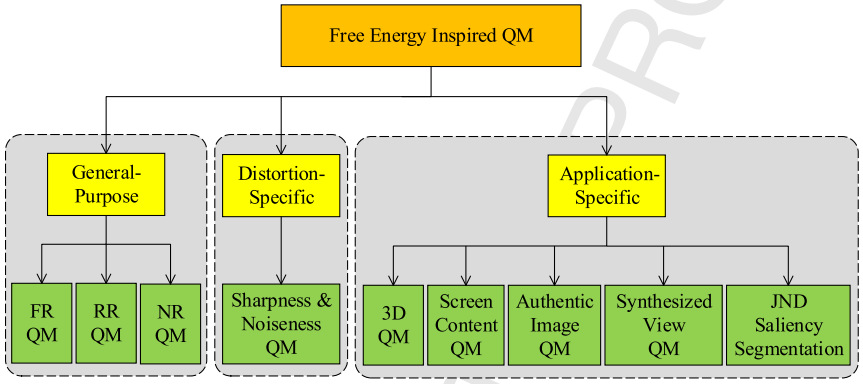

Fig. 2. A categorization of the free energy inspired visual quality measures.

2.2. Free energy principle and psychovisual quality assessment

The free energy principle governs visual perception and the free energy value indicates how well an image can be explained by a generative model. Since quality assessment is an "inherent" task performed by viewers during visual perception, there is a reason to believe that the free energy principle can be used for visual quality assessment. A straightforward implementation is to relate visual quality to the free energy value itself. For the scenario of full-reference quality assessment, with a reference image I_r and its distorted version I_d, their quality difference can be measured as the absolute difference in free energy, i.e.

$$\mathcal{D}(I_d, I_r) = |F(\hat{\boldsymbol{\theta}}_d) - F(\hat{\boldsymbol{\theta}}_r)| \tag{8}$$

with $\hat{\boldsymbol{\theta}}_r = \arg\min_{\boldsymbol{\theta}} F(\boldsymbol{\theta}|\mathcal{G}, I_r)$, $\hat{\boldsymbol{\theta}}_d = \arg\min_{\boldsymbol{\theta}} F(\boldsymbol{\theta}|\mathcal{G}, I_d)$. Note that if the free energy value of the original image is pre-computed, the method in (8) also qualifies for a reduced-reference metric. And this reduced-reference metric is extremely efficient in terms of the auxiliary information needed, as the value $F(\hat{\boldsymbol{\theta}}_r)$ can be expressed as a single float number.

The measure in (8) is a distortion measure in the sense that the value decreases as the quality of I_d increasing. Clearly, the measure enjoys the merits of being bounded and symmetric. It is easy to show that $0 \leq \mathcal{D}(I_d, I_r) \leq M$, with M being largest possible free energy values, e.g. it equals to the number of bits to encode a pixel if an image is completely random, i.e., $M = \lceil \log_2 L \rceil$ for an image of L intensity levels.

For the scenario of no-reference quality assessment, the free energy value $F(\hat{\boldsymbol{\theta}}_d)$ of the distorted image I_d can be used as a proxy of visual quality directly.

$$\mathcal{Q}(I_d) = F(\hat{\boldsymbol{\theta}}_d), \tag{9}$$

where the image I_d is corrupted by a degradation that alters the free energy value in a single direction, i.e. it either increases or decreases the free energy $\hat{\boldsymbol{\theta}}_r$ of I_r. For instance, image blurring and noise injection decreases and increases the free energy, respectively. And those two types of distortions cannot be mixed when using the no-reference metric. For complicated distortions, such as those from JPEG compression can eliminate useful image details while adding spurious structures, it cannot be handled effectively by this straightforward scheme.

A general framework of free energy modeling system is illustrated in Fig. 1. Input of the modeling process is the visual scene, and possible outputs of the system includes the estimated internal generative model, the model prediction, the prediction residual, and the free energy level. In those two most straightforward implementations of free energy principle inspired quality metrics, only the free energy levels are used. Some more complicated implementations based on other outputs will be reviewed in the next section.

3. Free energy inspired visual quality measures

We first give an overview of the free energy inspired visual quality measures. The algorithms are categorized in Fig. 2 according to the application domains, i.e. general purpose, distortion specific and application specific. Table 1 further gives details of the

Please cite this article in press as: G. Zhai et al., Free-energy principle inspired visual quality assessment: An overview, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.02.017

Table 1
Summarization of free energy inspired quality measures.

Category	Application	Algorithm & Ref.	Generative Model	Key idea
General-Purpose	FR VQA	IGM [8]	AR	(Dis)orderly content similarity
		FePVQ [31]	Structure/texture separation	Structure/texture similarity
		D-VICOM [32]	Least square decomposition	(Dis)orderly gradient similarity
		FEA-PSNR [9]	AR	Adjusted PSNR
	RR VQA	FEDM [3]	AR	Difference of residual entropy
		Wu et al. [10]	AR	(Dis)orderly content fidelity
		FSI [17,18]	AR and sparse representation	Difference of residual entropy
		MCFRM [11]	Sparse representation	Difference of residual entropy
		FEMJ [33]	JPEG & JP2K Compression	Difference of residual entropy
	NR VQA	NFEQM [3]	AR	Residual entropy
		NFSDM [12]	AR	Residual entropy and others
		NFERM [13]	AR	Residual entropy and others
	Comparative VQA	C-IQA [14]	AR	Cross modeling
Distortion-Specific	Sharpness	ARISM [34]	AR	AR parameters
		Han et al. [19]	Sparse representation	Difference of residual entropy
	Noiseness	Zhao et al. [35]	AR	Residual entropy
Application-Specific	3D VQA	Zhu et al. [36]	AR	Residual entropy and others
		Zhu et al. [37]	AR	Residual entropy and others
	Screen VQA	SVQI [21]	AR & bi-lateral filter	Global/local structure similarity
		SIQE [22]	AR & bi-lateral filter	Residual entropy and others
		Jakhetiya et al. [23]	AR & bi-lateral filter	Std of the residual
		MPS [38]	Sparse representation	Rough/smooth region features
		Che et al. [24]	AR & guided filter	Residual entropy and others
	Authentic VQA	Liu et al. [39]	AR	Residual entropy and others
		BQIC [40]	AR	Residual entropy and others
		Zhu et al. [41]	AR	Residual entropy and others
	Synthesized VQA	APT [42]	AR	Residual thresholding
		Jakhetiya et al. [23]	AR & bi-lateral filter	Std of the residual
		JETC [43]	JPEG compression	Residual thresholding
	Saliency	FES [44]	AR & bi-lateral filter	Residual entropy
	JND	Wu et al. [45]	AR	Prediction residual
	Segmentation	Feng et al. [46]	AR	(Dis)orderly content segmentation

algorithms including the choice of internal generative model and the key ideas.

3.1. General-purpose visual quality measures

General-purpose quality measures are assumed to be able to handle various kinds of distortions and applications. According to the availability of the distortion-free reference image, visual quality measures can be categorized as full-reference, reduced-reference, and no-reference. The readers can refer to Table 1 for an overview of details.

3.1.1. Full-reference visual quality measures

As discussed in Section 2, the full-reference measures generally follow a general framework illustrated in Fig. 3. First the reference and distorted images are approximated using the generative model respectively. And both images are therefore decomposed into model approximated ordered part. Then similarities between two images' ordered and disordered parts are measured separately before being integrated to give the final image quality.

Following the framework described above, Wu et al. [8] proposed a FR perceptual quality metric with AR internal generative model. The AR model was utilized to decompose the input scene into ordered (i.e. model explainable) part and disordered (i.e. prediction residual) part. The prediction part is believed to hold primary visual information, thus structural similarity measure [47] was utilized to measure the distortion. And the disordered prediction residual part mainly corresponds to uncertain information, and peak signal-to-noise ratio (PSNR) measure was utilized to measure the distortion. These two distortion measurements were integrated as the final quality.

Xu et al. [31] proposed a free energy principle inspired FR video quality metric. They followed the same idea of separating the image frame into ordered and disordered parts. To incorporate temporary information, they included a motion strength factor which was derived based on the human visual speed perception (HVSP), and thus the free-energy principle was extended to spatiotemporal domain for video quality assessment. To reduce the computational complexity, they used the relative total variation model rather than AR model to approximate the internal generative model, and block-wise motion vectors were used to represent the video motion. Besides the video quality metric, the authors also discussed the application of the proposed metric in perceptual rate distortion optimization (RDO).

Instead of decomposing the image directly, Di Claudio et al. [32] decomposed the gradient of the distorted image as a predicted version of the reference image's gradient plus the residual signal in gradient domain. The detail loss was then identified by the attenuation of the predicted gradient into a small observation window, and the presence of spurious detail was identified by the gradient residual observed into a small window. Then two metrics were proposed to measure the perceptual impact of detail losses and spurious details, which were finally integrated into the overall quality.

Liu and Zhai [9] introduced a free energy adjusted PSNR (FEA-PSNR) metric for FR visual quality assessment (VQA). The proposed measure was based on the observation that the deviation between subjective score and PSNR for each type of distortions can be systematically captured by perceptual complexity of the target image. The free energy modeling was utilized to measure perceptual complexity of an image. PSNR was improved by a linear score mapping process considering image free energy and distortion type.

Please cite this article in press as: G. Zhai et al., Free-energy principle inspired visual quality assessment: An overview, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.02.017

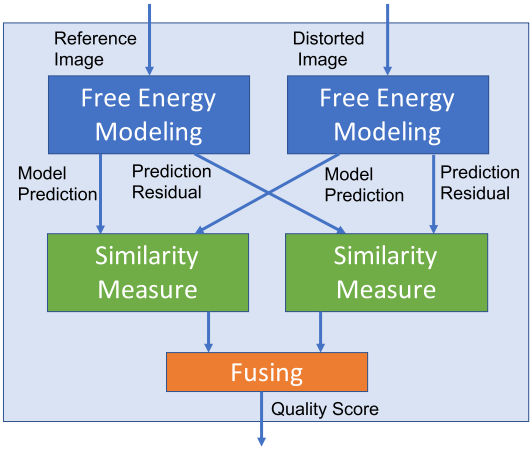

Fig. 3. A framework of the free energy inspired full-reference visual quality measures.

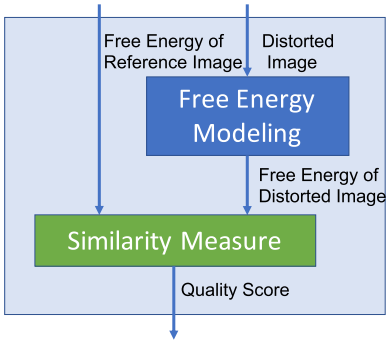

Fig. 4. A framework of the free energy inspired reduced-reference visual quality measures.

3.1.2. *Reduced-reference and no-reference visual quality measures*

Frameworks of the free energy inspired reduced-reference and no-reference visual quality measures have been given in Fig. 4 and Fig. 5, respectively. In a pioneering study [3], Zhai et al. developed a RR free-energy-based distortion metric (FEDM) and a no-reference free-energy-based quality metric (NFEQM) following the most straightforward idea outlined in Section 2. A practical implementation of the internal generative model using AR model, that has been widely used in the following works, was also proposed in [3].

Wu et al. [10] also used AR model as the internal generative model to decompose the image into primary visual information (the model prediction) and uncertainty part (prediction residual). The local binary patter (LBP) histogram of the two parts were used as features to compared the relative quality between the original and distorted images. The proposed method is a RR quality measure because that only the extracted free energy based LBP features are needed in the final metric.

Liu et al. [17,18] proposed a RR Free-energy principle and Sparse representation-based Index (FSI) for image quality assessment. FSI followed the same framework as FEDM [3] by measuring image quality using the difference between the entropies of the prediction discrepancies. FSI utilizes sparse representation rather than the AR model to simulate the internal generative model.

Zhu et al. [11] introduced a reduced reference quality metric named MCFRM (Multi-Channel Free-energy based Reduced-reference quality Metric). MCFRM first decomposes the input images using a two-level discrete Haar wavelet transform. Similar to FSI, MCFRM then utilizes sparse representation to simulate the internal generative model and extract the free-energy based features. The free-energy based features are then integrated to the final quality score.

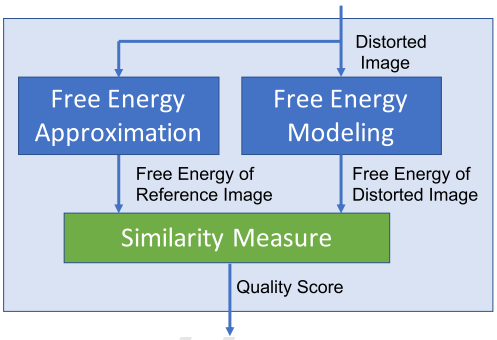

Fig. 5. A framework of the free energy inspired no-reference visual quality measures.

To reduce the computational burden of estimating parameters of the AR model, Gu et al. [33] proposed to use JPEG/JPEG2000 compression to approximate the model prediction process. The idea was that since JPEG/JPEG2000 compression at low bit rate removes fine image details in a similar way as the prediction process, it can be used to approximate the internal generative model.

Gu et al. [12] developed no-reference free energy and structural degradation based distortion metric (NFSDM) by integrating NFEQM and the structural degradation model (SDM) which measures structural degradation information by comparing the distorted image and its blurred version. And later a linear dependency between free energy value and SDM was noticed and NFSDM was modified as a NR quality metric named NR free energy based robust metric (NFERM) and achieved state of the art performance [13]. NFERM used three groups of features: features in NFSDM, similarities between the distorted and predicted images in terms of HVS-inspired features (such as structure and gradient features), and natural scene statistic (NSS) features of the mean subtracted contrast normalized coefficients.

3.1.3. *Comparative visual quality assessment*

A particularly interesting problem that has not been widely recognized in the research of visual quality assessment is how to compare the relative quality of two images of the same content but with different type and/or level of distortions. This problem was defined as the comparative image quality assessment (CIQA) problem [14]. CIQA differs from FR and RR tasks because the two images can both contain distortions. CIQA contrasts with NR problem in that the strong prior knowledge that two images are of the same contents must be used. In the work of [14], two AR generative models were estimated from the input image pairs, and yet both self and cross-over model prediction was performed. In other words, the model estimated from one image was used to explain both itself and the other image. And the free energy levels of the self- and cross-modeling process were compared to determine the relative quality of the image pair, see Fig. 6. A basic assumption of the algorithm was that the image with higher quality contains a model that can better explain the other image.

3.2. *Distortion-specific visual quality measures*

In some application scenarios, the distortion types can be assumed as known for the quality assessment task. Distortion-specific measures are designed for such scenarios, and these measures generally analyze the artifacts introduced by the distortion process using specific features. Due to the high correlation between free energy level and the perceived image complexity, some free energy inspired sharpness or noisiness measures were pro-

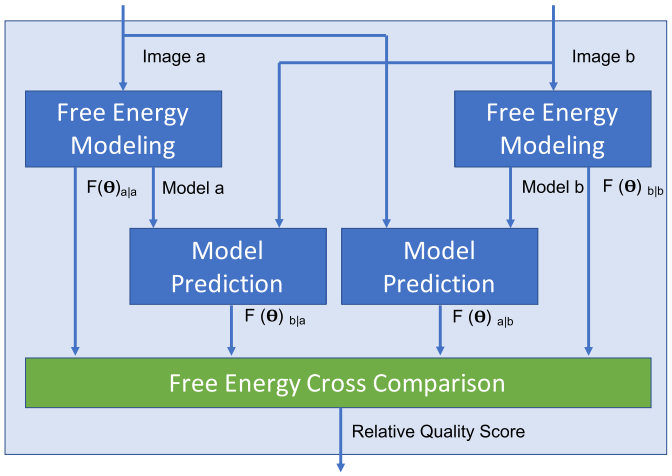

Fig. 6. A framework of the free energy inspired comparative visual quality measures.

posed. Readers can refer to Table 1 for a brief overview of these measures.

A NR AR model-based image sharpness metric (ARISM) was proposed in [34]. In ARISM, the AR model used in NFEQM [3] was adopted. The final ARISM metric was established via the analysis of AR model parameters. The algorithm first calculates the energy- and contrast-differences using the locally estimated AR coefficients, and then measures the image sharpness with percentile pooling. Color information is also incorporated to enhance the above metric.

Han et al. [19] proposed a free energy principle inspired RR measure to evaluate the quality of blurred images. Sparse representation was utilized to approximate the internal generative model. Visual saliency is incorporated to weight the metric locally for improved performance.

Zhao et al. [35] proposed a free energy principle inspired NR quality measure for noisy images following the same idea as NFEQM. The major difference was that it used a local median filter as an approximate to the internal generative model. The proposed algorithm is very efficient because the median filtering process is far less complicated than AR modeling.

3.3. Application-specific visual quality measures

Besides the abovementioned general visual quality measures, free energy principle also inspired the design of quality assessment metrics for specific applications, such as 3D content, synthesized views, authentic distortions, screen content quality evaluation and saliency detection. Readers can refer to Table 1 for a brief overview.

3.3.1. 3D visual quality measures

Zhu et al. [36] introduced a blind quality measure for stereo-pairs within the framework of free energy modeling. NFERM [13] was used to assess the quality of the 2D views, and then binocular rivalry level was estimated using model prediction procee to determine the relative importance of the left and right views. The final quality is a weighted average of the qualities of two views. The work was later modified in [37] with 2D image quality computed from traditional 2D VQA measures and the internal generative model predicted version were used to determine the inter-view weights together with a local tuning using a visual attention model.

3.3.2. Visual quality measures for screen content

The computer-generated screen content images (SCIs) are found to be quite different from natural scene images (NSIs) in many statistical characteristics [48]. Thus many VQA measures have been proposed specifically for SCIs or for both NSIs and SCIs in recent years [49,50].

Gu et al. [21] introduced a FR structural variation based quality index (SVQI) by combining the measurements of variations in global structures and local structures. The global structures variations are measured as the free energy change between the reference and distorted images. And both AR and bi-lateral filter were used to approximate the internal generative model.

A screen image quality evaluator (SIQE) [22] was proposed by integrating image complexity, screen content statistics, global brightness quality, and sharpness. Among which, the image complexity term was measured as the free energy level of the image. Similar to SVQI, both AR model and bi-lateral filter were used.

Jakhetiya et al. [23] introduced a RR VQA measure for SCIs based on a perceptually relevant generative model using ordinary least squares (OLS)-based autoregression. The OLS AR model was claimed to be better than the naive AR model in that the textual regions were better modeled and can therefore be emphasized in the final pooling stage of the quality metric.

Wu et al. [38] partitioned the SCIs into rough and smooth regions in gradient domain. Then a most preferred structure (MPS) metric was derived to represent the "meaningful" part of each image patch following the basic idea of the free energy principle using sparse representation. And the generative model predicted image was considered as more meaningful than the prediction residual because most useful structures are retained. Finally, the MPSs of the rough and smooth regions were pooled and integrated into the final quality.

Che et al. [24] designed a RR SCI quality measure by dividing SCI into text layer and pictorial layer, which were then evaluated by respective quality metrics. In both quality metrics, free energy level for the image layer was used as a feature.

3.3.3. Visual quality measures for authentic distortions

In many realistic settings, the distortions may not be of a single type and the authentic (realistic) distortions also became an important topic for visual quality assessment. Liu et al. [39] designed a NR quality measure for camera captured images. First, the low-level statistical features were extracted using the locally mean subtracted and contrast normalized coefficients. Then high-level free energy features which measures the divergence between the image and its internal generative model predicted version were extracted and integrated with low level features via neural network towards the final quality metric.

Tang et al. [40] integrated three groups of features for camera captured realistic image quality assessment, including natural scene statistics of the mean subtracted contrast normalized coefficients, patch-based image sharpness measure, and the free energy feature.

Zhu et al. [41] conducted a study concerning the visual quality assessment under various viewing conditions. They set up a system to record what the eyes see from the screen and built a viewing environment-changed image database. An objective method was also designed to assess different viewing environments. The free energy principle was utilized to model an image that human brain perceives using the prediction of the generative model.

3.3.4. Visual quality measures for synthesized views

View synthesis is very important in many applications such as virtual/augmented/mixed reality, free viewpoint video, and light field display. Many view synthesis methods have been proposed these years. And these methods introduce different visual distortions and result in quality degradation. Therefore, view synthesis quality measures are needed to quantify the synthesis quality. Gu et al. [42] designed a free energy inspired NR quality measure for depth image-based rendering (DIBR)-synthesized images using the

Please cite this article in press as: G. Zhai et al., Free-energy principle inspired visual quality assessment: An overview, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.02.017

AR model-based local image description. The proposed measure was designed based on the observation that the residual error between a DIBR-synthesized image and its predicted version with a AR model captures the geometry distortions well. This idea was improved by Qiao et al. [43] by replacing the AR model with an image compression module that is similar to [33] for improved efficiency.

The RR SCI quality measure designed by Jakhetiya et al. [23] also works for synthesized views. This is because that the free energy principle also applies to 3D contents and the OLS AR model separate 3D synthesized images into the predicted part and the disorder residual part that can be treated independently in the quality assessment stage.

3.4. The application of free energy in other image processing techniques

Besides visual quality assessment, the free energy principle has also inspired some other visual perceptual related image processing tools such as saliency detection and just noticeable difference (JND) estimation.

Wu et al. [45] introduced a JND estimation model inspired by the free energy principle. The JND threshold was estimated using a masking model of the disordered information (internal generative model prediction residual) of the visual scene.

A free energy inspired saliency (FES) algorithm was proposed in [44]. FES was motivated by the fact that the free energy principle directly relates free energy level to the amount of "surprise" of a given scene. AR and bi-lateral filtering was used as the generative model.

Feng et al. [46] proposed a medical brain image segmentation method driven by the internal generative model. They conduct Otsu thresholding [51] on both the original image and the model predicted version for segmentation. Then a regrouping measure was designed to refine the segmentation result.

3.5. Performance of free energy inspired visual quality measures

It is interesting to list and compare the performance of the free energy inspired visual quality measures. The literatures generally follow a common practice to conduct quality measures performance evaluation. As suggested by the Video Quality Experts Group (VQEG) [52], performance of objective assessment algorithms can be evaluated from three aspects: prediction accuracy, prediction monotonicity and prediction consistency. To reduce the nonlinearity of the prediction quality scores, first a five-parameter logistic function is used to map the objective quality scores nonlinearly [53–55]

$$q = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (s - \beta_3)}} \right) + \beta_4 s + \beta_5, \tag{10}$$

where $\{\beta_i | i = 1, 2, ..., 5\}$ are parameters determined via curve fitting, s and q are the predicted and mapped quality scores.

Then the consistency between the mapped quality scores and the subjective ratings can be measured from the following aspects to evaluate the quality metrics comprehensively [53–55].

- Spearman Rank-order Correlation Coefficient (SRCC). It computes the monotonicity while ignoring the relative distance between the data:

$$\mathrm{SRCC} = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}, \tag{11}$$

  where d_i denotes the difference between the i-th image's ranks in subjective and objective evaluations, and N is the number of images in the testing database.

- Kendall Rank-order Correlation Coefficient (KRCC):

$$\mathrm{KRCC} = \frac{N_c - N_d}{0.5(N-1)N}, \tag{12}$$

  where N_c and N_d express the numbers of concordant and discordant pairs in the testing data, respectively.

- Pearson Linear Correlation Coefficient (PLCC), which is a prediction accuracy measure:

$$\mathrm{PLCC} = \frac{\sum_i (q_i - \overline{q}) \cdot (o_i - \overline{o})}{\sqrt{\sum_i (q_i - \overline{q})^2 \cdot (o_i - \overline{o})^2}}, \tag{13}$$

  where o_i and q_i are the i-th image's subjective rating and the converted objective score after nonlinear mapping; $\overline{o}$ and $\overline{q}$ are mean values of o_i and q_i.

- Root Mean Square Rror (RMSE), which calculates the difference between the predicted and ground-truth quality scores:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)^2} \tag{14}$$

  where o_i and p_i are the i-th image's subjective quality rating and the mapped objective score.

Among the above four evaluation criteria, SRCC and KRCC measure the prediction monotonicity, PLCC estimates the linearity and consistency, RMSE evaluate the prediction accuracy. Higher SRCC, KRCC, PLCC and lower RMSE indicate better performance of the quality measures.

We collect the source codes of the general-purpose free-energy inspired visual quality measures, re-run all quality measures on the relevant databases, and compare their performance under the same testing procedures described above. Among all reviewed general-purpose free-energy inspired visual quality measures, the source codes of IGM [8], FEDM [3], FSI [17,18], MCFRM [11], NFSDM [3] and NFERM [13] are available. Table 2 lists their performances on the widely used image quality databases in the literature, including LIVE [56], CSIQ [57], TID2008 [58], and TID2013 [59]. For the training based measures, they are trained on the LIVE database and tested on other databases, and their performances on the LIVE database are not listed. While for other distortion-specific and application-specific quality measures, we do not report the performance in this survey due to the high diversity of the benchmarking datasets.

It can be noticed in Table 2 that some free energy inspired visual quality metrics indeed achieve state of the art performance. For almost all metrics, the correlation coefficients are pretty high on small scale databases of LIVE and CSIQ. Meanwhile, results on large scale databases TID2008 and TID2013 are still low. On most databases, the quality metrics can achieve higher performance if the more reference information is given, which is consistent with the results given in the literatures.

3.6. Analyses of free energy inspired visual quality measures

We further give some analyses to have a better understanding of the free energy inspired visual quality measures.

3.6.1. Comparison with deep visual quality measures

Thanks to the most recent advancement of deep learning, some deep learning based visual quality measures have been proposed these years [60–63], and pretty impressive performance improvements have been achieved in this area. As described in the literatures [60–63], these deep visual quality measures rely on large

Table 2
Performance of free energy inspired image quality assessment algorithms.

Dataset	Criteria	FR	RR			NR	
		IGM	FEDM	FSI	MCFRM	NFSDM	NFERM
LIVE	SRCC	0.9576	0.7947	0.8826	-	0.9274	–
	KRCC	0.8238	0.5964	0.6957	–	0.7776	–
	PLCC	0.9567	0.7976	0.8808	–	0.9285	–
	RMSE	7.9550	16.479	12.937	–	10.147	–
CSIQ	SRCC	0.9324	0.7230	0.7759	0.6846	0.5133	0.6609
	KRCC	0.7775	0.5247	0.5977	0.5063	0.3601	0.4937
	PLCC	0.9185	0.7210	0.8063	0.7667	0.6350	0.7522
	RMSE	0.1038	0.1819	0.1553	0.1686	0.2028	0.1730
TID2008	SRCC	0.8899	0.2870	0.4069	0.4134	0.2636	0.3064
	KRCC	0.7101	0.1688	0.2864	0.2930	0.2054	0.2298
	PLCC	0.8855	0.4707	0.5882	0.5803	0.4281	0.4902
	RMSE	0.6236	1.1840	1.0853	1.0928	1.2128	1.1696
TID2013	SRCC	0.8096	0.3977	0.3977	0.4471	0.2855	0.3479
	KRCC	0.6395	0.2764	0.2764	0.3160	0.2112	0.2375
	PLCC	0.8562	0.5840	0.5840	0.5926	0.3956	0.4882
	RMSE	0.6406	1.0063	1.0063	0.9986	1.1385	1.0819
Time (seconds)		9.3519	45.863	3.2327	4.8535	21.805	25.050

training data and they show large advantages over traditional visual quality measures in performance benchmarking. But they may not have good generalizability, especially when there is no similar data in the training set. What is more, these deep quality measures are purely data-driven and thus do not have good interpretability. Different from the deep visual quality measures, the free energy inspired visual quality measures are inspired by the psychological mechanisms of perception in human brain, i.e., the free energy principle. They are more close to the working mechanism of the HVS, and thus have better generalizability and interpretability. They are less dependent on training data, but they may fall short in performance benchmarking.

3.6.2. Computational complexity of free energy inspired visual quality measures

We test the computational complexity of all compared algorithms, and report the average running time (seconds/image) for 100 pairs of images with a fixed resolution of 512×512 in Table 2. The algorithms are tested with MATLAB R2016a operating on a computer with Intel Core i7-6700K CPU @4.00 GHz and 32 GB RAM. The running time includes all feature extraction and quality prediction time. For all competitors, we use the original implementations provided by the authors. It is observed that the free energy inspired visual quality measures whose simulated internal generative model is the AR model are quite computationally complex. FSI and MCFRM are based on sparse representation, and they are more computationally efficient.

3.6.3. Illustration of free energy inspired visual quality measures

To have a better understanding of the free energy inspired visual quality measures, we illustrate some results of various stages of these measures in Fig. 7. For both reference and distorted images, the AR model is used to predict the original scenes, and then the input scene is decomposed into a model approximated ordered part (2nd column in Fig. 7) and a disordered model approximation residual part (3rd column in Fig. 7). Note that the approximated ordered part has described the main contents and structures of the input scene, and the disordered model approximation residual part has described the image textures and details of the input scene. The entropy of the approximation residual can be calculated as the free energy of the scene. This free energy and the relevant decomposed images are then can be used to construct FR, RR, NR, and comparative visual quality measures according to the frameworks given in Fig. 3-Fig. 6.

3.6.4. Categorization of the internal generative models

In free energy inspired quality measures, the perception and understanding of an image is modeled as an active inference process, in which the brain tries to explain the scene using an internal generative model. The perceptual quality is closely related to how accurately visual sensory data can be explained by the internal generative model. Therefore, how to select the proper internal generative model to simulate the human brain generative model is a critical issue for free-energy inspired quality measures. Generally, the following models can act as the generative model

- Auto-regressive model: Auto-regressive model is a representation of a type of random process in statistics and signal processing. To achieve satisfying performance, the AR model needs to be piecewise. This local adaptability makes the model more expressive, and hence closer to the internal brain generative model in functionality. AR model achieves encouraging performance on evaluating the perceptual quality. It is observed in Table 1 that the majority of free energy inspired quality measures adopt AR as the generative model. The drawback of AR model is that it requires a higher computational cost compared to other generative model, such as bi-lateral filter.
- Bi-lateral filter: A bi-lateral filter is a non-liner, edge-preserving, and noise-reducing smoothing filter for images. It replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels. This generative model can preserve the sharp edges well, and achieves a satisfying performance on perceiving structural information of image. However, the bi-lateral filter model will introduce some types of artifacts, e.g. staircase effect (intensity plateaus that lead to images appearing like cartoons) and gradient reversal (introduce some false edges in the image).
- 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. The receptive fields of simple cells in mammalian primary visual cortex can be characterized as being spatially localized, oriented and bandpass. Sparse representation is qualified to resemble these neural response characteristics well, achieves a superiority for approximating the internal generative model. In addition, sparse representation has a faster computational time at the cost of the higher storage cost due to the large dictionary.

Fig. 7. Illustrations of various stages of free energy inspired visual quality measures. 1st column: original images; 2nd column: AR predicted images; 3rd column: prediction residuals (scaled to [0,255] for better illustration. 1st row: reference image; 2nd row: Gaussian blurred image.)

- Compression (JPEG, JPEG2000, etc): Besides the above models, image compression can also act as the generative model. Similar to the above models, image compression also acts as the same function of preserving the main image contents and removing image details, and the input image can be decomposed into a ordered image content part and a disordered image residual part. Owning to the availability of image codecs in various libraries, it is easy and fast to use image compression as the generative model, but it may not be so biologically plausible like the above models.

Note that different generative models have their advantages and disadvantages, and we may need to select a proper one to construct specific free energy inspired quality measure. Table 1 has summarized the used generative models of the reviewed measures.

4. Conclusion and discussions

This survey gives an overview of the free energy principle inspired visual quality metrics. The free energy brain theory was briefly introduced with the idea of its application in visual quality assessment was reviewed. The free energy principle inspired visual quality assessment algorithms in general purpose tasks i.e., full-reference, reduced-reference and no-reference visual quality assessment and distortion/application-specific tasks was then listed and compared in terms of basic ideas and performance.

Free energy modeling has been an emerging and powerful tool in the field of visual quality assessment. A few dozens of related algorithms have proposed in the last few years and some achieved state of the art performance. It was noticed from the analysis of existing free energy inspired visual quality metrics that the internal generative model is of key importance to the whole algorithm. The autoregressive model was proposed in a pioneer research and, despite of its high complexity, has been widely used in many following works. Although there were some attempts of designing novel generative models, e.g. bilateral filtering, many works were based on heuristic observations and therefore lack of theoretical justification. The perceptually more plausible and relevant internal generative model, e.g. epistemological automata, is needed to further improve the precision of modeling and the performance of those quality metrics.

Conflict of interest statement

None declared.

Acknowledgment

The authors want to thank the support from the National Science Foundation of China under Grants 61831015 and 61527804.

References

[1] H. von Helmholtz, J.P.C. Southall, Treatise on Physiological Optics, vol. 3, Courier Corporation, 2005.
[2] R.J. Sternberg, K. Sternberg, Cognitive Psychology, Nelson Education, 2016.
[3] G. Zhai, X. Wu, X. Yang, W. Lin, W. Zhang, A psychovisual quality metric in free-energy principle, IEEE Trans. Image Process. 21 (1) (2012) 41–52.
[4] K. Friston, J. Kilner, L. Harrison, A free energy principle for the brain, J. Physiol. (Paris) 100 (1–3) (2006) 70–87.
[5] K. Friston, The free-energy principle: a unified brain theory?, Nat. Rev. Neurosci. 11 (2) (2010) 127.
[6] D. Hubel, T. Wiesel, Receptive fields and functional architecture in the cat's visual cortex, J. Neurosci. 160 (1962) 106–154.
[7] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: a survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[8] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 22 (1) (2013) 43–54.
[9] N. Liu, G. Zhai, Free energy adjusted peak signal to noise ratio (fea-psnr) for image quality assessment, Sensing and Imaging 18 (1) (2017) 11.
[10] J. Wu, W. Lin, G. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimed. 15 (7) (2013) 1700–1705.
[11] W. Zhu, G. Zhai, X. Min, M. Hu, J. Liu, G. Guo, X. Yang, Multi-channel decomposition in tandem with free-energy principle for reduced-reference image quality assessment, IEEE Trans. Multimed. (2019).
[12] K. Gu, G. Zhai, X. Yang, W. Zhang, L. Liang, No-reference image quality assessment metric by combining free energy theory and structural degradation model, in: IEEE International Conference on Multimedia and Expo, 2013, pp. 1–6.
[13] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimed. 17 (1) (2015) 50–63.
[14] G. Zhai, A. Kaup, Comparative image quality assessment using free energy minimization, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 1884–1888.
[15] X. Wu, G. Zhai, X. Yang, W. Zhang, Adaptive sequential prediction of multidimensional signals with applications to lossless image coding, IEEE Trans. Image Process. 20 (1) (2011) 36–42.
[16] M.H. Hansen, B. Yu, Model selection and the principle of minimum description length, J. Am. Stat. Assoc. 96 (454) (2001) 746–774.
[17] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimed. 20 (2) (2018) 379–391.
[18] Y. Liu, G. Zhai, X. Liu, D. Zhao, Perceptual image quality assessment combining free-energy principle and sparse representation, in: IEEE International Symposium on Circuits and Systems, 2016, pp. 1586–1589.
[19] Z. Han, G. Zhai, Y. Liu, K. Gu, X. Zhang, A reduced-reference quality assessment scheme for blurred images, in: IEEE International Conference on Visual Communications and Image Processing, 2016, pp. 1–4.
[20] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision, 1998, pp. 839–846.

[21] K. Gu, J. Qiao, X. Min, G. Yue, L. Weisi, D. Thalmann, Evaluating quality of screen content images via structural variation analysis, IEEE Trans. Vis. Comput. Graph. 24 (10) (2018) 2689–2701.
[22] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, A.C. Bovik, No-reference quality assessment of screen content pictures, IEEE Trans. Image Process. 26 (8) (2017) 4005–4018.
[23] V. Jakhetiya, K. Gu, W. Lin, Q. Li, S.P. Jaiswal, A prediction backed model for quality assessment of screen content and 3-d synthesized images, IEEE Trans. Ind. Inform. 14 (2) (2018) 652–660.
[24] Z. Che, G. Zhai, K. Gu, P. Le Callet, Reduced-reference quality metric for screen content image, in: IEEE International Conference on Image Processing, 2017, pp. 1852–1856.
[25] D.C. Knill, A. Pouget, The bayesian brain: the role of uncertainty in neural coding and computation, Trends Neurosci. 27 (12) (2004) 712–719.
[26] G.E. Hinton, D. Van Camp, Keeping the neural networks simple by minimizing the description length of the weights, in: ACM Conference on Computational Learning theory, 1993, pp. 5–13.
[27] D.J. MacKay, et al., Ensemble learning and evidence maximization, in: Advances in Neural Information Processing Systems, 1995.
[28] W. Penny, S. Roberts, Bayesian methods for autoregressive models, in: IEEE Workshop on Neural Networks for Signal Processing, vol. 1, 2000, pp. 125–134.
[29] S.J. Roberts, W.D. Penny, Variational bayes for generalized autoregressive models, IEEE Trans. Signal Process. 50 (9) (2002) 2245–2257.
[30] R. Feynman, Statistical Mechanics: A Set of Lectures, 1998.
[31] L. Xu, W. Lin, L. Ma, Y. Zhang, Y. Fang, K.N. Ngan, S. Li, Y. Yan, Free-energy principle inspired video quality metric and its use in video coding, IEEE Trans. Multimed. 18 (4) (2016) 590–602.
[32] E.D. Di Claudio, G. Jacovitti, A detail-based method for linear full reference image quality prediction, IEEE Trans. Image Process. 27 (1) (2018) 179–193.
[33] K. Gu, G. Zhai, M. Liu, X. Yang, W. Zhang, Using jpeg and jpeg2000 compressions for fast image quality metrics based on free energy theory, in: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2014, pp. 1–5.
[34] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, IEEE Trans. Image Process. 24 (10) (2015) 3218–3231.
[35] Y. Zhao, Y. Liu, F. Jiang, X. Liu, D. Zhao, Fast noisy image quality assessment based on free-energy principle, in: International Forum on Digital TV and Wireless Multimedia Communications, 2017, pp. 290–299.
[36] Y. Zhu, G. Zhai, K. Gu, M. Liu, Blindly evaluating stereoscopic image quality with free-energy principle, in: IEEE International Symposium on Circuits and Systems, 2016, pp. 2222–2225.
[37] Y. Zhu, G. Zhai, K. Gu, Z. Che, D. Li, Stereoscopic image quality assessment with the dual-weight model, in: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2016, pp. 1–6.
[38] J. Wu, H. Li, Z. Xia, Z. Xia, Screen content image quality assessment based on the most preferred structure feature, J. Electron. Imaging 27 (3) (2018) 033025.
[39] Y. Liu, K. Gu, S. Wang, D. Zhao, W. Gao, Blind quality assessment of camera images based on low-level and high-level statistical features, IEEE Trans. Multimed. (2019).
[40] L. Tang, L. Li, K. Gu, X. Sun, J. Zhang, Blind quality index for camera images with natural scene statistics and patch-based sharpness assessment, J. Vis. Commun. Image Represent. 40 (2016) 335–344.
[41] Y. Zhu, G. Zhai, K. Gu, Z. Che, Closing the gap: visual quality assessment considering viewing conditions, in: International Conference on Quality of Multimedia Experience, 2016, pp. 1–6.
[42] K. Gu, V. Jakhetiya, J.-F. Qiao, X. Li, W. Lin, D. Thalmann, Model-based referenceless quality metric of 3d synthesized images using local image description, IEEE Trans. Image Process. 27 (1) (2018) 394–405.
[43] J. Qiao, M. Liu, S. Li, Z. He, Z. Yang, Highly efficient quality assessment of 3d-synthesized views based on compression technology, IEEE Access. (2019).
[44] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, Visual saliency detection with free energy theory, IEEE Signal Process. Lett. 22 (10) (2015) 1552–1555.
[45] J. Wu, G. Shi, W. Lin, A. Liu, F. Qi, Just noticeable difference estimation for images with free-energy principle, IEEE Trans. Multimed. 15 (7) (2013) 1705–1710.
[46] Y. Feng, X. Shen, H. Chen, X. Zhang, Internal generative mechanism based otsu multilevel thresholding segmentation for medical brain images, in: Pacific Rim Conference on Multimedia, 2015, pp. 3–12.
[47] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[48] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, W. Lin, Unified blind quality assessment of compressed natural, graphic, and screen content images, IEEE Trans. Image Process. 26 (11) (2017) 5462–5474.
[49] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, C.W. Chen, Blind quality assessment based on pseudo-reference image, IEEE Trans. Multimed. 20 (8) (2018) 2049–2062.
[50] X. Min, G. Zhai, K. Gu, Y. Liu, X. Yang, Blind image quality estimation via distortion aggravation, IEEE Trans. Broadcast. 64 (2) (2018) 508–517.
[51] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[52] VQEG, Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, [Online], http://www.vqeg.org/, 2003.
[53] X. Min, G. Zhai, K. Gu, X. Yang, X. Guan, Objective quality evaluation of dehazed images, IEEE Trans. Intell. Transp. Syst. (2019).
[54] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, W. Zhang, Quality evaluation of image dehazing methods using synthetic hazy images, IEEE Trans. Multimed. (2019).
[55] X. Min, K. Gu, G. Zhai, M. Hu, X. Yang, Saliency-induced reduced-reference quality index for natural scene and screen content images, Signal Process. 145 (2018) 127–136.
[56] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE image quality assessment database release 2, [Online], available http://live.ece.utexas.edu/research/quality.
[57] E. Larson, D. Chandler, Consumer subjective image quality database, http://vision.okstate.edu/csiq/.
[58] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008-a database for evaluation of full-reference visual quality assessment metrics, Adv. Modern Radioelectron. 10 (4) (2009) 30–45.
[59] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Image database TID2013: peculiarities, results and perspectives, Signal Process. Image Commun. 30 (2015) 57–77.
[60] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1733–1740.
[61] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE J. Sel. Top. Signal Process. 11 (1) (2017) 206–220.
[62] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2018) 206–219.
[63] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks, IEEE Trans. Image Process. 27 (3) (2018) 1202–1213.

Guangtao Zhai received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. From 2006 to 2007, he was a Student Intern with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen Nurem- berg, Erlangen, Germany. Since 2012, he has been with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, where he is currently a Professor. His research interests include multimedia signal processing and perceptual signal processing. He was a recipient of the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012 and the Best Paper Award of IEEE TRANSACTIONS ON MULTIMEDIA in 2018. His students received Best Student Paper Awards at the PCS 2015 and the IEEE ICME 2016. He and his students received the Grand Prize of the ICME 2018 Grand Challenge on Saliency 360!

Xiongkuo Min received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018. From 2016 to 2017, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a Post-Doctoral Fellow with Shanghai Jiao Tong University. His research interests include image quality assessment, visual attention modeling, and perceptual signal processing. He received the Best Student Paper Award from the IEEE ICME 2016.

Ning Liu received Ph.D degree from Shanghai Jiao Tong University in 2010 and is now an associate professor in the School of Electronics, Information and Electrical Engineering, Shanghai Jiao Tong University.

Please cite this article in press as: G. Zhai et al., Free-energy principle inspired visual quality assessment: An overview, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.02.017