J. Vis. Commun. Image R. 40 (2016) 335–344

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Blind quality index for camera images with natural scene statistics and patch-based sharpness assessment ☆

Lijuan Tang [a,b], Leida Li [a], Ke Gu [c], Xingming Sun [d], Jianying Zhang [a,*]

[a] School of Info. and Elec. Eng., China Univ. of Mining and Technology, China
[b] School of Info. and Elec. Eng., Jiangsu Vocational College of Business, China
[c] Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, China
[d] School of Compu. and Soft., Nanjing Univ. of Information Science and Technology, China

ARTICLE INFO

Article history:
Received 10 March 2016
Revised 11 July 2016
Accepted 12 July 2016
Available online 12 July 2016

Keywords:
Image quality assessment (IQA)
Camera images
Blind/no-reference (NR)
Natural scene statistics (NSS)
Local sharpness
Free energy theory
Structural degradation model

ABSTRACT

The current image quality metrics work on the assumption that an image contains single and simulated distortions which are not representative of real camera images. In this paper we address the problem of quality assessment of camera images from two respects, natural scene statistics (NSS) and local sharpness, and associated three types of features. The first type of four features measures the naturalness of an image, inspired by a recent finding that there exists high correlation between structural degradation information and free energy entropy on natural scene images and this regulation will be gradually devastated as more distortions are introduced. The second type of four features originates from an observation concerning the NSS that a broad spectrum of statistics of distorted images can be caught by the generalized Gaussian distribution (GGD). Both the two types of features above belong to the NSS-based models, but they come from the considerations of local auto-regression (AR) and global histogram, respectively. The third type of three features focuses on estimating the local sharpness, which works by computing log-energies in discrete wavelet transform domain. Finally our quality metric is achieved via a SVR-based machine learning tool, and its performance is proved to be statistically better than state-of-the-art competitors on the CID2013 database dedicated to the quality assessment of camera images.

1. Introduction

With the soaring development of mobile devices and network, an enormous amount of images are being presented to users every moment. It is challenging to evaluate and control the quality of digital photographs. At the same time, a supreme effort is still made by camera manufacturers to improve the quality of photography. As thus, it is in an urgent pursuit of finding ways to automatically predict the perceptual quality of camera images.

In the past few years, as for the issue of image quality assessment (IQA), many objective metrics of the ability to faithfully evaluate the quality of distorted images have been developed with applications to compression [1–3], transmission [4], enhancement [5], tone mapping [6] and image forensics [7–9]. If the distortion-free image which distorted image can be compared with is available, the metric is called full-reference (FR) IQA [10]. But in most cases only the distorted image is available, this type of IQA models are called no-reference (NR) IQA. Furthermore, according to the requirement of prior knowledge of the images or their distortions, current NR IQA algorithms also can be further classified into two categories, namely general-purpose metrics and distortion-specific metrics. Typical distortion-specific blind quality measures are devoted to ringing effect [11], blockiness [12], noise [13], sharpness/blurriness [14–22], etc. Ferzli et al. proposed a blur metric by integrating just noticeable blur into a probability summation model to evaluate the amount of blurriness in distorted images, dubbed as just-noticeable blur metric (JNB) [15]. Inspired by JNB, Narvekar et al. pooled the localized probability of blur detection by means of a cumulative probability of blur detection to measure distorted images [16]. In [17], the slope of the magnitude spectrum and the total spatial variation is used to create sharpness map to be used to predict image blurriness. Thereafter, Vu et al. [18] used the log-energies in high frequency wavelet subbands to predict the global and local sharpness of distorted image. Very recently, few attempts to estimate camera images estimations have been made. Nuutinen et al. [23,24] tried to search for efficient feature sets for predicting visual quality of real photographs. In [25], the authors proposed an approach by utilizing NSS modeling as well as the

☆ This paper has been recommended for acceptance by M.T. Sun.
* Corresponding author.
E-mail address: zhangjianyingiqa@163.com (J. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2016.07.007
1047-3203/

consumer-centric, quality aware interpretable features for real consumer-type images quality prediction. They also presented a framework [26] for blind quality consumer content images evaluation.

In recent years, general-purpose NR-IQA metrics have been an active research field [27–33,35,36]. In [27], the authors proposed a two-step framework to evaluate a distorted image based on natural scene statistics (NSS), in which the first step is to estimate the existence of distortion types in the image and the second one is to evaluate the distorted image through each of these distortions. Natural image quality evaluator (NIQE) [28] was devised to predict the deviations from statistical regularities observed in natural images without any prior knowledge of the images or their distortion types. Also inspired by some underlying statistics, Saad et al. [29] used discrete cosine transform coefficients to extract features, and then predicted image quality scores with a simple Bayesian inference approach. Scene statistics of locally normalized luminance coefficients was used by blind/referenceless image spatial quality evaluator (BRISQUE) [30] to measure possible losses of "naturalness" in the image referable to the presence of distortions and provide an overall quality measure of the distorted images.

Although aforementioned metrics perform well on the popular databases such as the LIVE database [37], they do not perform as well on real photographs which are subjected to many different distortion sources and types. Because these image quality metrics are based on the assumption that an image contains single or simulated distortions which are not representative of what one encounters in practical real scenarios [25]. Camera images contain more practical distortions unlike most distortions present in the popular databases.

Compared with the previous works, to the best of our knowledge, this paper is the first to propose to a modular framework for IQA of camera images based on the NSS regulation and local sharpness assessment. And furthermore, the proposed blind quality index for camera images (BQIC) has acquired a substantially high performance, it is the only metric with the correlation coefficient of beyond 0.8 in both linear and monotonic performance indices.

The paper is structured as follows. In Section 2, we present the details of the BQIC algorithm. Section 3 provides performance measures and comparisons of our BQIC and state-of-the-art blind quality metrics on the CID2013 database [38] dedicated to the IQA of camera images. General conclusions and future works are given in Section 4.

2. Proposed blind quality measure

Selecting appropriate features plays an important part in IQA. The features of the proposed metric consists of three parts. The flowchart of the proposed NR IQA metric is outlined in Fig. 1.

The first group of features is extracted based on the free-energy principle, which is recently developed in brain theory and neuroscience [34], and structure degradation measurement [39]. The free-energy principle operates under the assumption that there always exists a difference between an input genuine visual signal and its processed one by human brain. Human perceptual process is manipulated by an internal generative model, which can infer predictions of the input visual signal and avoid the residual uncertainty information. On this basis, the psychovisual quality of a scene is defined by both the scene itself and the output of the internal generative model. It differs from most traditional methods which are based on signal decomposition.

To facilitate operation, we assume that the internal generative model G for visual perception is parametric, which infers perceived scenes by adjusting the parameter vector $\mathbf{x}$. Given an input visual signal I, the joint distribution $p(I,\mathbf{x})$ over the space of model parameters $\mathbf{x}$ can compute the "surprise"[1] information (measured by entropy) of the given image. The joint distribution function can be computed as follows:

$$-\log p(I) = -\log \int p(I,\mathbf{x})d\mathbf{x} \tag{1}$$

However, the joint distribution $p(I,\mathbf{x})$ is difficult to be directly measured according to current knowledge. As thus, a dummy term $q(\mathbf{x}|I)$ that is an auxiliary posterior distribution of the model parameters given the image is brought into both the numerator and the denominator. So we can rewrite Eq. (1) to be

$$-\log p(I) = -\log \int q(\mathbf{x}|I)\frac{p(I,\mathbf{x})}{q(\mathbf{x}|I)}d\mathbf{x} \tag{2}$$

Next, Jensen's inequality is used to apply to Eq. (2), and we have

$$-\log p(I) \leqslant -\int q(\mathbf{x}|I)\log\frac{p(I,\mathbf{x})}{q(\mathbf{x}|I)}d\mathbf{x} \tag{3}$$

and the right side of Eq. (3) is defined as the free energy:

$$J(\mathbf{x}) = -\int q(\mathbf{x}|I)\log\frac{p(I,\mathbf{x})}{q(\mathbf{x}|I)}d\mathbf{x} \tag{4}$$

Eq. (4) expresses the free energy $J(\mathbf{x})$ to be energy minus entropy. And the free energy estimation of the image I can be expressed by

$$F(I) = J(\hat{\mathbf{x}}) \quad \text{with} \quad \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} J(\mathbf{x}) \tag{5}$$

Any quantitative application of Eq. (5) operates under the assumption that the brain generative model excites. A model with higher expressive power approximates the brain better but incurs higher computational complexity. In this paper we choose the linear AR model as the generative model for its effectiveness and simplicity to describe natural scenes [21,35]. The AR model is defined as

$$y_n = \boldsymbol{\chi}^k(y_n)\boldsymbol{\theta} + \varepsilon_n \tag{6}$$

where y_n is a pixel of the distorted image, $\boldsymbol{\chi}^k(y_n)$ is a row-vector consisting of k nearest neighbors of y_n, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ is a vector of AR coefficients, and ε_n is the error term. Then, the predicted version of the input distorted visual signal I can be estimated by $\boldsymbol{\chi}^k(y_n)\cdot\boldsymbol{\theta}_{opt}$, where $\boldsymbol{\theta}_{opt}$ is the optimal estimate of AR parameters for y_n based on the least square method. Consequently, the estimated AR parameters can be used to represent the distribution of the model parameters $q(\mathbf{x}|I)$. In order to demonstrate the distribution of model parameters exhibits a center-peaked appearance, a natural image and its posterior distribution of the model parameters $q(\mathbf{x}|I)$ are shown in Fig. 2. According to [40], the total description length of I with the kth-order AR model can be expressed by

$$L(\hat{\mathbf{x}}) = -\log p(I|\hat{\mathbf{x}}) + \frac{k}{2}\log M \tag{7}$$

where M is the number of pixels. And in the large sample limit $M \to \infty$, the free energy is the total description length:

$$J(\hat{\mathbf{x}}) = -\log p(I|\hat{\mathbf{x}}) + \frac{k}{2}\log M \quad \text{with} \quad M \to \infty \tag{8}$$

We choose a fixed-model order, and thus the second term $\frac{k}{2}\log M$ is constant and can be ignored in the quality evaluation.

[1] The free energy principle works on the assumption that all biological agents resist the natural tendency to disorder in an ever-changing environment. Therefore, it suggests that biological agents can somehow violate the second law of thermodynamic by keeping their internal states at low entropy level to maintain themselves within some physiological bounds. This process is to avoid encountering "surprise" under different environment [34].