J. Vis. Commun. Image R. 24 (2013) 1316–1327

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Evaluation of Region-of-Interest coders using perceptual image quality assessments

J.C. Garcia-Alvarez [a,b,*], H. Führ [b], G. Castellanos-Dominguez [a]

[a] Universidad Nacional de Colombia, Signal Processing and Recognition Group, Manizales, Colombia
[b] RWTH Aachen, Lehrstuhl A für Mathematik, Germany

ARTICLE INFO

Article history:
Received 9 March 2012
Accepted 6 September 2013
Available online 19 September 2013

Keywords:
Region-of-Interest
Image coding
Wavelet
Distortion measure
Quality assessment
Perceptual evaluation
Human visual system
Mean-observed scores
Rate-distortion function

ABSTRACT

A perceptual measure emulates the human vision for image quality assessment. This paper illustrates the evaluation of Region-of-Interest (ROI) coders using perceptual image quality assessments. The goal of this evaluation is to characterize the coder performance by controlling the ROI quality. Perceptual measures are taken into account for evaluation since they behave as a human-made evaluation. Moreover, a perceptual assessment named *Wavelet Quality Index* (WQI), is introduced as another image coder evaluator. Proposed assessment aims at emulating the human vision by a weighted linear combination of three wavelet-based perceptual measures. We evaluate the following types of ROI-coders: those preserving the quality of ROI by coarse compression of background (*Max-Shift* coder), and those balancing the quality between ROI and background (*SCM-Shift*, and *BbB-Shift* coders). Using considered assessments for the performance evaluation of coders, results show a variation of evaluation by nature of measurement.

1. Introduction

For *store-and-forward* tele-medicine applications in emerging countries, a high-resolution image travels from a data source repository to another remote destination through channels with very reduced bandwidth [1]. Such a procedure causes undesired effects on transmitted data like latency and/or loss of information. Coders send an image at a reduced bandwidth providing less image data. Consequently, received image becomes distorted leading to low quality transmission. To overcome this issue, *Region-of-Interest* (ROI) coder is used to preserve, as much as possible, quality of specific image areas. After then, the coder sends rest of information with adequate quality over some continuous time intervals. Several applications employ the ROI coder: detail preservation [2], face detection (pattern recognition area) [3], and content search (content description area) [4], among others.

Conventionally, a coder performance evaluation consists of measuring image quality resulting after the carried out coding procedure. Specifically, the *image quality assessment* (IQA) evaluates the coder considering the following [5–7]: (i) the entire encoded image; (ii) the ROI-area; (iii) the difference between the ROI and complementary Background (BG) qualities. Although different objective measures had been suggested for implementing IQA, mostly, the *Peak Signal to Noise Ratio* is employed. Nonetheless, coders using this measure as control index do not adequately keep quality of decoded images, since they often produce visually distracting artifacts [6]. To avoid those added artifacts, perceptual measures have been recently introduced involving human vision models [8]. So, the IQA includes a perceptual measure that provides human-like visual evaluation of the coder performance, which any objective measure might not supply. Therefore, there is some preference in using perceptual methods, despite their higher degree of conceptual and computational complexity [9,10]. Since a given perceptual measure should be highly correlated to the scores made by human subjects, IQA assessment may differently evaluate the performance of analyzed coder on dependence on the used perceptual measure as well as the considered distortion model [8]. As a result, a perceptual assessment better characterizing the ROI-coder remains still as an open issue.

Abbreviations: IQA, Image quality assessment; ROI, Region of Interest; BG, Image background; PSNR, Peak Signal to Noise Ratio; MS-SSIM, Multi-Scale Structural SIMilarity; VIF, Visual Image Fidelity; MSE, Mean Squared Error; R-Q, Rate-Quality Function; QILV, Quality Index based on Local Variance; RF, Reflection Factor; VSNR, Visual Signal to Noise Ratio; DN, Divisive Normalization; WQI, Wavelet-based Quality Index; ρ, Measured/Quality Value; CC, Correlation coefficient; QND, Quality Normalized Difference; RBD, ROI-background difference.

* Corresponding author at: Universidad Nacional de Colombia, Signal Processing and Recognition Group, Manizales, Colombia.

E-mail addresses: jcgarciaa@bt.unal.edu.co (J.C. Garcia-Alvarez), fuehr@matha.rwth-aachen.de (H. Führ), cgcastellanosd@unal.edu.co (G. Castellanos-Dominguez).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.09.003

The present study aims to provide a ROI-coder performance evaluation by explicitly including perceptual image quality assessments. Namely, it comprises the proposed *Wavelet-based Quality Index* (WQI), which uses the assumed weighted linear combination of wavelet-based perceptual measures. Thus, the suggested perceptual assessment becomes function of several perceptual measures, demanding extra parameters that are extracted from the same representation [11,12]. Proposed approach focuses on computing a set of linearly combined weights rather than handling a more complex distortion model, as discussed in [13]. For validation purpose of proposed WQI assessment, we use the correlation coefficient that is calculated between quality values (extracted from an image set) and the data scores extracted from evaluations, given by human observers.

The agenda of present work is as follows: Section 2 presents an overview of the existing perceptual IQAs. Section 3 describes in detail the proposed assessment. In Section 4, an experimental methodology for performance evaluation of ROI-coder is stated using considered assessments. The evaluation results are presented in Section 5, and they are further discussed in Section 6, where the highlights of proposed assessment and future work are also provided.

2. Overview of image quality assessments

Mainly, a loss of quality of a distorted image should be expressed by a real-value measure. So, quality assessment of encoded image supplies coder performance evaluation by using either *Objective* or *Perceptual* measures. The former measure is distance-based, forcing it to be less adaptable or correlated to human-like evaluation, while the latter qualitative measure should be properly set to get a suitable distortion score.

Generally, any objective measure is calculated by using a single distance, which is commonly an image difference parameter itself. Examples of these measures are the *Mean Squared Error* (MSE), *Peak Signal to Noise Ratio* (PSNR), and *Czekanowski* distance, which are widely used since their calculation is directly done over a considered image and no other representation is required [14]. In turn, a perceptual measure, employing human-visual-system-based features, can be modeled as collection of selective channels in terms of frequency and orientation [15]. Yet, each channel that has different center frequency and orientation requires an additional number of operations for its description. In particular, *Human Visual System*, which is based on biological and psychological human-eye models as well, is employed for sensitivity parametrization. As a result, the time processing burden is increased and it leads to a severe delay in the quality assessment of large images. To cope with this issue, several IQA assessments had been proposed [16–18,8]. In particular, evaluation of JPEG2000 wavelet-based coder is performed in [19], based on human panel evaluations made during clinically relevant visual tasks. In [20], another recent extensive study is presented relating human quality assessments, including many images, distortion types, and number of human judgments per image. Though all these studies have influenced the use of the perceptual measure as control parameter on coders, they are just adequate where wavelet-based processing is suitable since some perceptual measures benefits of the same coding representation [15].

In this work, all considered IQA assessments are assumed to employ a fully-referenced measure that is calculated as the difference between a given reference (source) image, $\boldsymbol{X} \in \mathbb{R}^{M_1 \times M_2}$ of size $M_1 \times M_2$, and a distorted image, $\widetilde{\boldsymbol{X}} \in \mathbb{R}^{M_1 \times M_2}$, having the same size, where (M_1, M_2) is the image size set (width and height, respectively). Concrete perceptual measures for IQA, noted as $\rho_k \in \mathbb{R}^+$ (Variable k indexes each considered measure), are shown in Table 1, namely, Multi-Scale Structural SIMilarity (MS-SSIM) [21], Visual Image Fidelity (VIF) [22], *Quality Index based on Local Variance* (QILV) [23], *Reflection Factor* (RF) [24], *Visual Signal to Noise Ratio* (VSNR) [25], and *Divisive Normalization* (DN) [26].

The MS-SSIM that is a multi-scaled generalization of the structural similarity measure is given as follows [27]:

$$\rho_1 = [h_l]^{\alpha} \prod_{j=1}^{J} [h_c(j)]^{\beta_j} [h_s(j)]^{\gamma_j},$$

where $h_l \in \mathbb{R}, h_c \in \mathbb{R}$, and $h_s \in \mathbb{R}$ are the luminance, contrast, and structure comparison measures, respectively, between source, $\boldsymbol{X}$, and encoded, $\widetilde{\boldsymbol{X}}$, images; exponents $\alpha \in \mathbb{R}, \beta_j \in \mathbb{R}$, and $\gamma_j \in \mathbb{R}$ are the adjustment parameters for the jth scaled image, with $j \in J$, being J the total number of decomposition levels (number of sub-bands inside wavelet structure). A cross-scale calibration provides the required adjustment of exponent parameters as well as the scaling filter [21].

The VIF quantifies Shannon information, $I(\cdot)$, present in the encoded image to the source representation [22]:

$$\rho_2 = \frac{\sum_{j=1}^{J} I(\boldsymbol{c}^j; \widetilde{\boldsymbol{X}} | \varsigma^j)}{\sum_{j=1}^{J} I(\boldsymbol{c}^j; \boldsymbol{X} | \varsigma^j)},$$

where vector $\boldsymbol{c}^j \in \mathbb{R}^{M_1 \times M_2}$ holds the wavelet coefficients that are generated from a Gaussian random model and conditioned to a given structural human-based model, $\varsigma^j \in \mathbb{R}^{M_1 \times M_2}$; this vector is calculated from either source $\boldsymbol{X}$ or encoded $\widetilde{\boldsymbol{X}}$ image, for a particular j sub-band.

The QILV measure arises as a combination of local variance estimators, calculated as:

$$\rho_3 = \frac{2\mu_{V_{\boldsymbol{X}}}\mu_{V_{\widetilde{\boldsymbol{X}}}}}{\mu_{V_{\boldsymbol{X}}} + \mu_{V_{\widetilde{\boldsymbol{X}}}}} \frac{2\sigma_{V_{\boldsymbol{X}}}\sigma_{V_{\widetilde{\boldsymbol{X}}}}}{\sigma^2_{V_{\boldsymbol{X}}} + \sigma^2_{V_{\widetilde{\boldsymbol{X}}}}} \frac{\sigma_{V_{\boldsymbol{X}}V_{\widetilde{\boldsymbol{X}}}}}{\sigma V_{\boldsymbol{X}} + \sigma^2_{V_{\widetilde{\boldsymbol{X}}}}},$$

where $\mu_{V_{\boldsymbol{X}}} \in \mathbb{R}$ and $\sigma_{V_{\boldsymbol{X}}} \in \mathbb{R}^+$ are the expected value and standard deviation of the local variance, respectively, which are calculated for each element of image $\boldsymbol{X}$; $\sigma_{V_{\boldsymbol{X}}V_{\widetilde{\boldsymbol{X}}}} \in \mathbb{R}^+$ is the covariance between the local variances of the images $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$. QILV value ranges between $[0, 1]$, being a monotonically increasing measure, tending to one when quality increases.

Consecutively, the RF assessment is computed as:

$$\rho_4 = \frac{\sum_{i=1}^{M_v} \mid \delta_i \mid w_i}{\sum_{i=1}^{M_v} v_i},$$

where $v_i \in \mathbb{R}$ is each one of the M_v nonzero singular values estimated from Singular Value Decomposition method, carried out over the source image $\boldsymbol{X}$; $\delta_i \in \mathbb{R}$ is a measure consisting of each one of the vector elements calculated from the reliable difference factor between the original and the distorted image; and $w_i \in \mathbb{R}$ is each one of the elements of the normalized vector of singular values of the source image. RF value ranges within interval $[0, 1]$, monotonically increasing with higher distortion.

Next, the VSNR is a wavelet-based measure estimating visual fidelity in the form:

$$\rho_5 = 20 \log \frac{\xi(\boldsymbol{X})}{\xi(\boldsymbol{X} - \widetilde{\boldsymbol{X}})},$$

where values, $\xi(\boldsymbol{X}) \in \mathbb{R}^+$ and $\xi(\boldsymbol{X} - \widetilde{\boldsymbol{X}}) \in \mathbb{R}^+$, refer to the root-mean-squared contrast of source and error images, respectively. For calculation, image contrast requires the following [25]: contrast thresholds for distortion detection, a measure of the perceived contrast of the distortions, and a measure of the degree to which the distortions disrupt global contrast in the image.

Lastly, The DN measure consists of a contrast-based normalization of wavelet coefficients computed over the encoded image.