Quality Assessment of Contrast-Altered Images

Min Liu[†], *Ke Gu*[§], *Guangtao Zhai*[†], *Jiantao Zhou*[♭], *and Weisi Lin*[§]
[†]Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, China
[§]School of Computer Engineering, Nanyang Technological University, Singapore
[♭]Department of Computer and Information Science, University of Macau

***Abstract*—In image / video systems, the contrast adjustment which manages to enhance the visual quality is nowadays an important research topic. Yet very limited efforts have been devoted to the exploration of image quality assessment (IQA) for contrast adjustment. To address the problem, this paper proposes a novel reduced-reference (RR) IQA metric with the integration of bottom-up and top-down strategies. The former one stems from the recently revealed free energy theory which tells that the human visual system always seeks to understand an input image by the uncertainty removal, while the latter one is towards using the symmetric K-L divergence to compare the histogram of the contrast-altered image with that of the reference image. The bottom-up and top-down strategies are lastly combined to derive the Reduced-reference Contrast-altered Image Quality Measure (RCIQM). A comparison with numerous existing IQA models is conducted on contrast related CID2013, CCID2014, CSIQ, TID2008 and TID2013 databases, and results validate the superiority of the proposed technique.**[1]

***Index Terms*—Contrast alteration, image quality assessment (IQA), reduced-reference (RR), hybrid parametric and non-parametric model (HPNP), bottom-up, top-down**

I. Introduction

The importance of visual media, which in most conditions are provided to human consumers, have been realized lately. As the users' requirements for high-quality images / videos are increasingly rising, a reliable system to evaluate, control and improve the users' quality of experience (QoE) is highly required. This gives rise to the demand of faithful metrics of image quality assessment (IQA) for predicting the quality in accordance with human visual perception [1].

With respect to the accessibility of the original references, objective IQA metrics are mostly classified into three types: 1) *full-reference* (FR); 2) *reduced-reference* (RR); 3) *no-reference* (NR). Depending on the supposition of structural variations being extremely vital in quality perception, the last few years have witnessed the emergence of a vast majority of FR IQA metrics [2-6]. Under the condition of partial reference image or several extracted features being available as side information, RR-IQA has a broader range of practical scenarios. Guided by the recent discovery of free energy theory [7], we lately designed the free energy based distortion metric (FEDM) [8] by simulating the internal generative model of human brain to detect input visual signals. Exploiting a set of filters and valid pooling strategy, structural degradation model (SDM) [9] has managed to modify FR SSIM into valid RR IQA techniques with only a few numbers as RR information.

Despite the prosperity and successfulness of IQA studies, very limited efforts have been devoted to the field of IQA with contrast change [10]. Moreover, existing IQA algorithms do not work validly in this field. As a matter of fact, contrast is an important research topic [11], which has practical applications such as contrast enhancement technologies [12-13]. This motivates the design of a novel dedicated contrast-changed image database (CID2013) [14], including 400 contrast-changed images by mean shifting and four kinds of transfer mappings, and its advanced version (CCID2014) [15].

In this paper we further dig into the issue of contrast-changed IQA, and develop a new RR IQA model with the combination of bottom-up and top-down strategies. Relative to the frequently seen distortion types, e.g. JPEG / JPEG2000 compressions, the human visual sensation of image contrast (mainly including brightness and contrast alteration) is more prone to the aesthetic quality assessment [16], and therefore inclines to the measurement in visual and psychological fields. A recently revealed free energy principle illustrates that the HVS always tries to perceive a visual signal by reducing the uncertain portion and measures the psychovisual quality as the agreement between an image and its output of the internal generative model. With this, we evaluate the visual quality of contrast-altered images in the bottom-up model based on the internal generative mechanism, which is constructed by the non-parametric autoregressive (AR) model via perceptual information for weighting.

On the other hand, as pointed out in several existing contrast enhancement methods [12], the histogram modification can result in the contrast adjustment and largely influence users' experiences. The top-down strategy aims to compare two distances between histograms; one is of the contrast-adjusted image and its original counterpart, and the other is of the contrast-altered image and the one created from the original image through histogram equalization. The Kullback-Leibler (K-L) divergence, one of the most popular information-theoretic "distances" comparing two probability distributions, is naturally taken into account. But the K-L divergence is non-symmetric and brings unstable results in calculation. So we use the symmetrized and smoothed Jensen-Shannon (JS) divergence [17] to compute the two distances stated above. Finally, the bottom-up and top-down strategies are combined to develop the Reduced-reference Contrast-changed Image Quality Measure (RCIQM), whose superiority is verified over

[1]This work was supported in part by NSFC (61025005, 61371146, 61221001, 61390514, 61402547) and Macau Science and Technology Development Fund under grant FDCT/046/2014/A1.

Authorized licensed use limited to: Universidade de Macau. Downloaded on July 06,2022 at 09:07:08 UTC from IEEE Xplore. Restrictions apply.

existing visual quality evaluators.

The remainder of this paper is organized as follows. In Section II, we combine the bottom-up and top-down models to derive the RCIQM metric. Section IV conducts comparative studies of our measure with numerous existing FR- and RR-IQA methods on CID2013 [14], CCID2014 [15], CSIQ [18], TID2008 [19], TID2013 [20] databases, and then reports and discusses the results. Section IV concludes this paper.

II. Proposed RCIQM Metric

Although there exist a large quantity of IQA techniques, none of them is able to acquire satisfied correlation performance. Therefore, we in this paper concentrate on the IQA of contrast alteration.

A. Bottom-Up Strategy

In the research of aesthetic quality assessment [21], it is argued that people would prefer a visual signal with balanced lighting and proper contrast. Furthermore, in comparison to the typical distortion types, e.g. image / video coding, the HVS perception to image contrast, which is affected by luminance and contrast variation, is more prone to the assessment of aesthetic quality and thus to the visual and psychological measurement [16]. So we first establish the bottom-up model based on the free energy principle, which generates an approximate estimation of the psychovisual quality [8].

According to the analysis in [8], there exists a gap between the external input signal and its generative-model-explainable part highly connecting to the psychovisual quality, which can be used to assess contrast-changed images. The free energy is the gap (i.e. the error map) between the input visual signal and its output best explanation inferred by the internal generative model. In the error map, larger-value regions are what cannot be well explained by the generative model, whereas smaller-value pixels are what can be easily described. This error map is obtained by minimizing free energy.

The internal generative model is defined to be a new hybrid parametric and non-parametric (HPNP) model, which fuses the linear AR model with the bi-lateral filtering. The first AR model is simple and it can simulate a wide range of natural scenes by varying its parameters. Particularly, the AR model is expressed by

$$y_i = \mathcal{Y}^k(y_i)\boldsymbol{\alpha} + \varepsilon_i \tag{1}$$

where y_i is the value of a pixel at location x_i, $\mathcal{Y}^k(y_i)$ defines k member neighborhood vector of y_i, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_k)^T$ is a vector of AR coefficients, and ε_i is a difference term between truth values and predictions. We use the method in [8] to find the solution of $\boldsymbol{\alpha}$ for each pixel.

The AR model is sometimes unstable at image edges. So we further take advantage of the bi-lateral filtering [22], which is a non-linear filtering of good edge-preserving ability. Also, the bi-lateral filtering just has two variables, making it convenient to control. We define this filtering by

$$y_i = \mathcal{Y}^k(y_i)\boldsymbol{\beta} + \varepsilon'_i \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)^T$ is a vector of bi-lateral filtering coefficients, and ε'_i is an error term. The $\boldsymbol{\beta}$ is manipulated by the spatial Euclidean distance between x_i and x_j as well as the photometric distance between y_i and y_j, referring to the definition in [22].

In the following, the HPNP model fuses the merits of both parametric AR model and non-parametric bi-lateral filtering, and thus generate the estimation of $\bar{y}_i$ to be

$$\bar{y}_i = \gamma \cdot \mathcal{Y}^k(y_i)\hat{\boldsymbol{\alpha}} + (1-\gamma) \cdot \mathcal{Y}^k(y_i)\boldsymbol{\beta} \tag{3}$$

where γ is used for adjusting the relative contribution of the AR model and the bi-lateral filtering. Generally, salient regions attract much attention and thus highly influence the visual quality. This paper incorporates the luminance, contrast and structural information, as defined in [2], for weighting:

$$w_i = l(y_i, \bar{y}_i) \cdot c(y_i, \bar{y}_i) \cdot s(y_i, \bar{y}_i), \tag{4}$$

and the estimation error of the gap between the real scene and brain's prediction for the local pixel at x_i is evaluated by

$$\bar{e}_i = w_i(y_i - \bar{y}_i). \tag{5}$$

For the original image I_o, the point-wise error $\bar{e}_i$ can be computed using Eqs. (1)-(5) to get the error map E_o. The free energy of this error map is measured by entropy:

$$H(E_o) = -\sum p_i(E_o) \log p_i(E_o) \tag{6}$$

where $p_i(E_o)$ is the probability density of grayscale i in the error map E_o. With the same manner, we measure entropy of $H(E_c)$ for the contrast-changed image I_c. The psychovisual quality of I_c compared to I_o within the bottom-up strategy is finally defined as their difference:

$$Q_{bu} = H(E_o) - H(E_c). \tag{7}$$

In practice, the kernel of the bottom-up strategy lies in the HPNP model for approximating the internal generative model in human brain. Images with high contrast and visual quality usually have an abundant number of valuable details. Our HPNP model is of different description abilities between low- and high-complexity visual signals. For a fixed input image with its free energy $H(E_o)$, the positive contrast change will increase the visual quality by revealing undiscernible details. This renders the designed HPNP model inefficient to characterize the contrast-altered image, and thus makes its free energy $H(E_c)$ higher than $H(E_o)$ and Q_{bu} lower than zero. On the contrary, the negative contrast change will decrease the visual quality by concealing details, which leads to the associated free energy $H(E_c)$ smaller than $H(E_o)$ and Q_{bu} larger than zero.

B. Top-Down Strategy

One of important applications related to contrast alteration is the familiar contrast enhancement technology, which can be treated as the positive contrast change for validly advancing the contrast and raising the visual quality of an input image. Broadly speaking, contrast enhancement aims to generates a

Authorized licensed use limited to: Universidade de Macau. Downloaded on July 06,2022 at 09:07:08 UTC from IEEE Xplore. Restrictions apply.

more visually-pleasing or informative image or both. Typical viewers describe the enhanced images as if a curtain of fog has been removed from the picture.

In [12], the authors have provided some suggestions to contrast enhancement, namely the enhanced image should be not far from its original one, and they smartly provided a compromise scheme. Instead of using the uniformly distributed histogram $\mathbf{h_u}$ as the target histogram, their goal is to find a modified histogram $\tilde{\mathbf{h}}$ that is near to $\mathbf{h_u}$ as desired, but also not far from the original image histogram $\mathbf{h_o}$. This is a bi-criteria optimization problem, and can be formulated as a weighted sum of the two objectives:

$$\tilde{\mathbf{h}} = \arg\min_{\mathbf{h}} \|\mathbf{h} - \mathbf{h_o}\| + \phi\|\mathbf{h} - \mathbf{h_u}\| \tag{8}$$

where $\tilde{\mathbf{h}}$, $\mathbf{h}$, $\mathbf{h_o}$, $\mathbf{h_u} \in \mathbb{R}^{256\times 1}$, and ϕ is a control parameter varying over $[0, \infty)$.

Enlightened by Eq. (8), we in the top-down strategy focus our attention on measuring two distances; one is between the histogram $\mathbf{h_c}$ of the contrast-changed image and $\mathbf{h_o}$, and the other is between $\mathbf{h_c}$ and $\mathbf{h_u}$. Nonetheless, the construction of the top-down model is not straightforward. Firstly, we find that $\mathbf{h_u}$ is not a good choice, since the histograms of most images cannot be distributed uniformly after HE due to various image contents or scenes. Instead, this paper applies the equalized histogram $\mathbf{h_e}$ that is produced from $\mathbf{h_o}$ using HE. Secondly, it is important to note that the free energy in the bottom-up strategy is measured by entropy, so we had better evaluate the aforesaid two distances with the same dimension for the combination of bottom-up and top-down models to predict the visual quality score of the contrast-adjusted image. The K-L divergence is of the expected dimension. Given two probability densities p_0 and p_1, the K-L divergence is defined as

$$\mathcal{D}_{KL}(p_1\|p_0) = \int p_1(x)\log\frac{p_1(x)}{p_0(x)}dx. \tag{9}$$

This K-L divergence is however non-symmetric and easy to bring some troubles in real applications. Simple examples illustrate that the ordering of the arguments in the K-L distance might yield substantially different results. We resort to the symmetric K-L divergence accordingly. In [17], the authors have summarized many symmetric forms of K-L divergence, e.g. algebraic mean and geometric mean. Here we consider using the symmetrized and smoothed Jensen-Shannon (JS) divergence as follows:

$$\mathcal{D}_{JS}(p_0, p_1) = \frac{1}{2}\mathcal{D}_{KL}(p_0\|\bar{p}) + \frac{1}{2}\mathcal{D}_{KL}(p_1\|\bar{p}) \tag{10}$$

with $\bar{p} = \frac{1}{2}(p_0 + p_1)$.

As a consequence, given three probability densities p_o, p_e and p_c for an original image and its HE and contrast-altered counterparts, the quality of I_c compared to I_o within the top-down part is determined by

$$Q_{td} = \mathcal{D}_{JS}(p_c, p_o) + s\mathcal{D}_{JS}(p_c, p_e) \tag{11}$$

where s is a fixed weighting parameter for altering the relative importance between the above two distances. The analyses

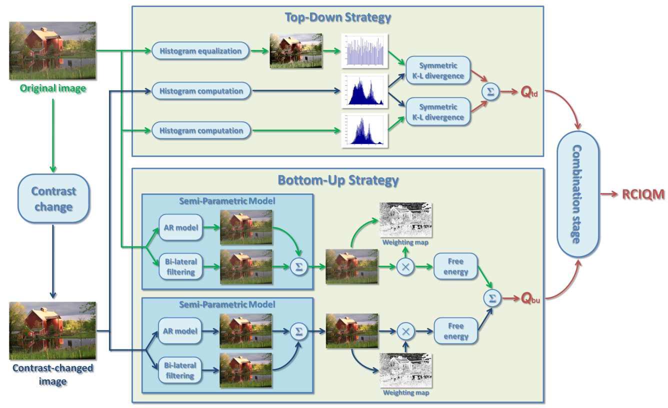

Fig. 1. The flowchart of the proposed RCIQM algorithm.

in the histogram modification method point out that proper-contrast images should be a good tradeoff between the original image histogram and the uniformly distributed one. Our top-down model is properly developed for this, and it can thereby judge the quality levels of contrast-changed images.

C. The Combination Stage

Popular contrast enhancement technologies are devoted to highlighting undiscernible details [13] or redistributing image histogram [12]. Given an image, the former bottom-up model aims to estimate how much detailed information is contained, while the latter top-down model is to measure whether the histogram is properly distributed. From the viewpoint of working, these two models play complementary roles. Hence we fuse bottom-up and top-down strategies to approximate the HVS perception to the contrast-altered image quality. Since the quality measures based on the two models are of the same dimension (i.e. entropy) in our research, they can be directly integrated. The RCIQM is lastly defined to be a simply linear function combining the two quality predictions in bottom-up and top-down parts:

$$\text{RCIQM} = Q_{bu} + tQ_{td} \tag{12}$$

where t is a constant weight that is used to control the relative contribution between the bottom-up and top-down strategies. All the parameters used in the proposed RCIQM model have fixed values. We present the flowchart in Fig. 1 for helping readers to readily understand the RCIQM metric. Our source code will be released soon.

Furthermore, we want to discuss why the proposed RCIQM is a RR IQA metric. In the bottom-up model, the RR feature only includes one single number of the free energy $H(E_o)$, and two histograms $\mathbf{h_o}$ and $\mathbf{h_e}$ are required to transmitted as the ancillary information in the top-down part. In reality, $\mathbf{h_e}$ is the output of the equalized $\mathbf{h_o}$. So the RR information used in RCIQM just includes $H(E_o)$ and $\mathbf{h_o}$ (totally 257 numbers), far less than the size of the original image. Besides, a supplementary specification is that, according to the convention, this paper utilizes different signs (e.g. p_o and $\mathbf{h_o}$, p_e and $\mathbf{h_e}$) but with the same meaning.

Authorized licensed use limited to: Universidade de Macau. Downloaded on July 06,2022 at 09:07:08 UTC from IEEE Xplore. Restrictions apply.

TABLE I
PERFORMANCE COMPARISONS ON THE FIVE CONTRAST RELATED DATABASES. WE HIGHLIGHT THE TOP THREE METRICS.

Metrics	Type	CID2013 (400)		CCID2014 (655)		CSIQ (116)		TID2008 (200)		TID2013 (250)		Weighted average	
		PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
PSNR	FR	0.6503	0.6649	0.6832	0.6743	0.9002	0.8621	0.5131	0.5207	0.5071	**0.5425**	0.6425	0.6462
SSIM	FR	0.8119	0.8132	0.8256	**0.8136**	0.7450	0.7397	0.5057	0.4877	0.5658	0.4905	0.7369	0.7182
GSI	FR	0.8353	0.8372	0.8073	0.7768	0.9325	0.9354	0.6739	0.5126	0.6665	0.4985	0.7850	0.7275
LTG	FR	**0.8656**	**0.8605**	**0.8384**	0.7901	0.9560	0.9414	0.6795	0.4655	**0.6749**	0.4639	**0.8087**	**0.7279**
SDM	RR	0.7158	0.6145	0.7360	0.6733	0.9175	0.9141	**0.7817**	**0.7378**	0.5831	0.3482	0.7261	0.6338
RIQMC	RR	**0.8995**	**0.9005**	**0.8726**	**0.8465**	**0.9652**	**0.9579**	**0.8585**	**0.8095**	**0.8651**	**0.8044**	**0.8829**	**0.8567**
RCIQM	RR	**0.9187**	**0.9203**	**0.8845**	**0.8565**	**0.9645**	**0.9569**	**0.8807**	**0.8578**	**0.8866**	**0.8541**	**0.8985**	**0.8792**

III. EXPERIMENTAL RESULTS

In this paper, using the five contrast related CID2013, CCID2014, CSIQ, TID2008, and TID2013 databases, we validate the proposed RCIQM algorithm and compare with an enormous number of classical and state-of-the-art IQA metrics: 1) Classical FR IQA: PSNR and SSIM [2]; 2) State-of-the-art FR IQA: GSI [4] and LTG [5]; 3) Recently devised RR IQA: SDM [9] and RIQMC [15].

First we compute the objective prediction scores of each testing IQA models, and use the nonlinear regression to map those scores to subjective ratings with the five-parameter logistic function [23]:

$$q(\varepsilon) = \phi_1\left(\frac{1}{2} - \frac{1}{1+e^{\phi_2(\varepsilon-\phi_3)}}\right) + \phi_4\varepsilon + \phi_5 \qquad (13)$$

where ε and $q(\varepsilon)$ respectively indicate the input score and the mapped score, and ϕ_j $(j = 1, \ldots, 5)$ are free parameters to be decided. We then make use of Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SRCC) to compute the performance. A value close to 1 for PLCC and SRCC means superior correlation with subjective opinions. Table I presents the performance of our RCIQM and six models. Across the five databases, we further calculate the database size-weighted average, whose results are also reported in Table I.

Referring to the nature of our proposed metric and the results in Table I, we give two conclusions. First, it is apparent that our metric achieves very exciting result on each database and the average. We notice that only the proposed RCIQM technique has acquired the SRCC values greater than 0.92 on the CID2013 database, and larger than 0.85 on the large-scale CCID2014, TID2008 and TID2013 databases. Although a few IQA models (e.g. FEDM) do well in the CSIQ database, our RCIQM is also of the highest performance, even higher than 0.95 in the accuracy and monotonic measure.

Second, as compared to FR- and RR-IQA algorithms tested in this paper, it can be readily observed that the proposed RCIQM is of the optimal performance on average, clearly better than the second-place RIQMC and third-place LTG methods. In fact, almost all FR and RR IQA methods assume that the reference image is prefect. But there exist contrast-changed images produced by the positive contrast alteration have better quality than their original ones, and this leads to serious deterioration in the performance of FR and RR IQA techniques when assessing contrast-altered images.

IV. CONCLUSION

In this paper, we have introduced a new RCIQM metric with bottom-up and top-down strategies. Considering that the visual quality of the contrast-altered image is highly connected to the psychovisual mechanism in human brain, the bottom-up strategy applies the HPNP model with perceptual information for weighting. The top-down strategy compares the histogram of the contrast-changed image with those of the original and equalized ones with K-L divergence. Results verify the superiority of our RCIQM over state-of-the-art competitors.

REFERENCES

[1] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *JVCIR*, vol. 22, no. 4, pp. 297-312, May 2011.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE T-IP*, 2004.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE T-IP*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.

[4] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE T-IP*, vol. 21, no. 4, pp. 1500-1512, Apr. 2012.

[5] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," *in Proc. IEEE Int. Conf. Image Process.*, pp. 506-510, Oct. 2014.

[6] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, and W. Zhang, "Content-weighted mean-squared error for quality assessment of compressed images," *Signal Image and Video Processing*, pp. 1-8, Sept. 2015.

[7] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, pp. 127-138, 2010.

[8] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE T-IP*, vol. 21, pp. 41-52, Jan. 2012.

[9] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," *in Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 1095-1098, May 2013.

[10] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE T-CSVT*, vol. 15, no. 7, pp. 900-909, Jul. 2005.

[11] E. Peli, "Contrast in complex images," *JOSA*, vol. 7, pp. 2032-2040, Oct. 1990.

[12] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE T-IP*, 2009.

[13] X. Wu, "A linear programming approach for optimal contrast-tone mapping," *IEEE T-IP*, vol. 20, no. 5, pp. 1262-1272, May 2011.

[14] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Subjective and objective quality assessment for images with contrast change," *in Proc. IEEE Int. Conf. Image Process.*, pp. 383-387, Sept. 2013.

[15] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybernetics*, vol. 46, no. 1, pp. 284-297, Jan. 2016.

[16] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," *in IEEE CVPR*, pp. 419-426, Jun. 2006.

[17] D. H. Johnson and S. Sinanović, "Symmetrizing the Kullback-Leibler distance," *IEEE T-IT*, 2001.

[18] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," [Online], Available: http://vision.okstate.edu/csiq

[19] N. Ponomarenko *et al.*, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, 2009.

[20] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Sig. Process.: Image Commun.*, vol. 30, pp. 57-55, Jan. 2015.

[21] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE T-MM*, vol. 15, no. 8, pp. 1930-1943, Dec. 2013.

[22] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *in IEEE ICCV*, pp. 836-846, Jan. 1998.

[23] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000, http://www.vqeg.org/.

Authorized licensed use limited to: Universidade de Macau. Downloaded on July 06,2022 at 09:07:08 UTC from IEEE Xplore. Restrictions apply.