

Non-Shift Edge Based Ratio (NSER): An Image Quality Assessment Metric Based on Early Vision Features

Min Zhang, Xuanqin Mou, and Lei Zhang

Abstract—How to evaluate the image perceptual quality is a fundamental problem in image and video processing, and various methods have been proposed for image quality assessment (IQA). This letter presents a novel IQA metric, which is based on the image primitive features produced in the earliest processing stage of human visual system. The procedures involved in the proposed method include computing the response of classical receptive fields, zero-crossing detection, and non-shift edge based ratio (NSER) calculation. The proposed IQA metric is very simple but very effective. The experimental results on benchmark databases show that the NSER index has very high consistency with the psychological evaluation, performing much better than most state-of-the-art IQA metrics.

Index Terms—Image quality assessment (IQA), non-shift edge (NSE), non-shift edge based ratio (NSER), zero-crossing.

I. INTRODUCTION

IMAGE quality assessment (IQA) has been attracting much research attention because of its importance in various image/video processing applications. Conventional IQA indices such as mean squared error (MSE), signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR) have been used for a long time but they do not conform well to the human visual perception. In the past decades, much effort has been made on full reference (FR) IQA based on known models of human visual system (HVS). Perceived visual errors are computed based on the visual error sensitivities to various factors of HVS, such as brightness, contrast sensitivity, visual attention, etc. [1]. Later, some typical metrics, such as the Sarnoff JNDmatrix visual discrimination model (VDM) [2], the noise quality measure (NQM) [3], and the wavelet based visual signal to noise ratio (VSNR) [4] have been proposed. Instead of using the known HVS models, some IQA methods attempt to directly model the property of perceptual distortion of HVS, including the universal image quality index (UQI) [5], the Structural-SIMilarity (SSIM) index [6] and its multiscale counterpart MS-SSIM [7], the Information Fidelity

Criteria (IFC) based on Natural Scene Statistics (NSS) [8], and the Visual Information Fidelity (VIF) [9].

The SSIM index has been widely recognized as a benchmark for IQA. Its motivation is that a measure of image structural change can provide a good approximation to the perceived distortion. Based on this, in SSIM the image quality is evaluated by computing the structural similarity in terms of the mean, variance, and covariance of the reference and distorted images. As we know, the structures of an image can be extracted into many modalities at different stages in the visual pathway from retina to primary cortical area (V1), and the earliest human visual process is carried out by ganglion and lateral geniculate nucleus (LGN) neurons with the classical receptive field (CRF), which can be modeled by the Laplacian of Gaussian (LOG) function, to produce the basic primitive structures of images [10]. According to Marr's theory [10], those structural primitives are crucial to represent image semantic information in late HVS process. In particular, most information of the basic primitives is carried by the edge locations where zero-crossings occur [10]. Therefore, the edge related information can be useful for IQA tasks. In our previous work [13], we proposed an FR IQA metric called QMESS which considers multiscale edge structural similarities, low frequency distortion and stochastic noise in the metric design. The edge structural similarity was defined as how much edge points keeping their original positions after distortion and the edge point was detected by finding modulus maxima from wavelet decomposition. Our recent work found that above edge structural similarity can be used lonely as a good IQA metric if the edge is detected from the earliest visual feature by zero-crossing detection. Meanwhile, some other works have also been proposed to incorporate the edge information into the task of IQA [14] and perceiving image utility [15]. In this letter, we provide a new IQA metric called Non-Shift Edge based Ratio (NSER), which is well supported by the experimental results.

The rest of the letter is organized as follows. Section II presents the proposed NSER metric. Section III presents the experimental results and discussions. Section IV concludes the letter.

II. NON-SHIFT EDGE BASED RATIO (NSER): A NEW IMAGE QUALITY METRIC

A. Non-Shift Edges

In the visual pathway of mammals, ganglion and LGN neurons have evolved to response the rapid change in light intensity

Manuscript received January 06, 2011; revised March 03, 2011; accepted March 04, 2011. Date of publication March 14, 2011; date of current version March 28, 2011. This work was supported in part by the NSFC under Grants 90920003, 60472004, and 60551003, and by the the Hong Kong General Research Fund (PolyU 5330/07E). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vicky Zhao.

M. Zhang and X. Mou are with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China (e-mail: xqmou@mail.xjtu.edu.cn).

L. Zhang are with the Department of Computing, Hong Kong Polytechnic University, Hong Kong.

Corresponding Author. Email: xqmou@mail.xjtu.edu.cn.

Digital Object Identifier 10.1109/LSP.2011.2127473

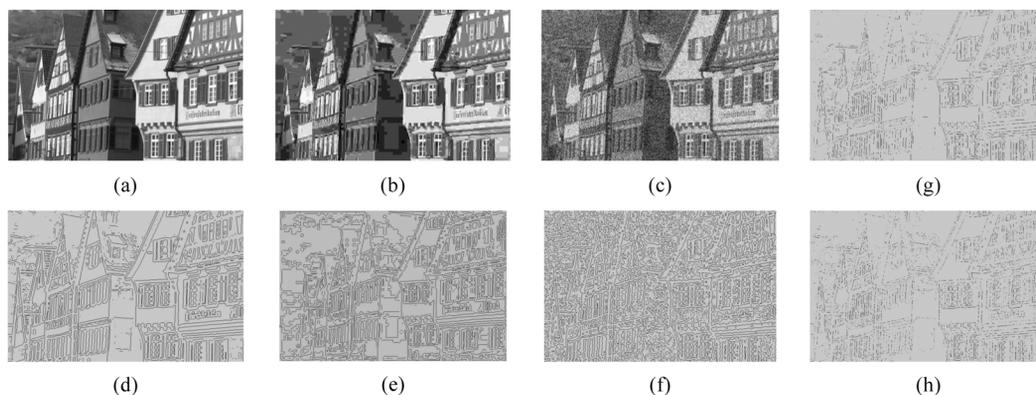


Fig. 1. An example of NSE. (a) Reference image in the LIVE database, and (d) is its edge map extracted by the LOG filter (here the scale factor σ is 2.6 and the threshold for edge detection is 0.5). (b) and (c) Distorted images of (a) in the LIVE database. (b) Compressed image by JPEG with DMOS value 59.01, and (e) is its edge map. (c) Noisy image with additive white noise and its DMOS value is 56.68. (f) Edge map of (c). (g) and (h) NSE maps of (e) and (f) according to (d), respectively. We see that although (e) and (f) are very different, their NSE maps are similar, which is identical to the objective DMOS evaluation scores.

on different scales by their unique center surrounded CRF [10]. From a computational point of view, the response of CRF can be modeled by a series of hierarchical filters, i.e., the LOG filters [10], which are defined as

$$\nabla^2 G(x, y, \sigma) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

where $G(x, y, \sigma)$ is the well known Gaussian kernel and σ is the scale factor.

The above exported basic primitive structures are mainly conveyed to the primary visual cortex for further processing to produce the raw primal sketches, such as bars, boundaries, blobs, etc, to represent the scene. According to Marr's theory [10], the pivotal information existed in those basic primitives is carried by the locations where zero-crossing occurs, and it was conjectured that an image may be (nearly) completely represented by these zero-crossings on multiple scales. Computational vision researchers have later studied to which degree an image can be reconstructed from its edges or zero-crossings [11], [12]. Although the late works showed that zero-crossings alone cannot form a complete representation of an image, they indeed carry out most of the semantic information in an image. In this letter, we propose to use the earliest vision features, more specifically, zero-crossing edges only, to measure the difference between reference and distortion images.

The information carried by the edges defined as the zero-crossings is represented by their spatial locations in the image. When an image is distorted from the original one, the positions of edge points will change accordingly. The more serious the distortion is, the higher the degree the edge positions will change. Therefore, it is a straightforward idea to compare the edge maps of the reference and distorted images to measure their difference. Here we assume that the reference and distorted images are well registered, which is commonly assumed in IQA research.

Since it is not easy to pair the edge points in the reference and distorted images, comparing the locations of the same edge point in the reference and distorted images is difficult. Considering the fact that the significant edges in an image won't easily change their spatial locations when the image is distorted, we

turn to investigate the edges that stay in their original locations after the image is deteriorated, and we call those edges Non-Shift Edges (NSE), whose map is defined as below:

$$\text{NSE}(C, D) = E_C \cap E_D. \quad (2)$$

Here, C and D denote the reference and distorted images, respectively, and E_C and E_D are the edge maps of them. An edge map is a binary image, where "1" denotes an edge point and "0" denote a non-edge point. Obviously, the NSE map can be calculated by the "AND" operation of the two binary edge maps, denoted by $E_C \cap E_D$.

We use an example to illustrate how NSE represents the information shared by the reference and the distorted images. Fig. 1(a) is a reference image from the LIVE database at <http://live.ece.utexas.edu/research/quality>. Two types of distorted images of Fig. 1(a) with similar objective quality evaluation indexed by DMOS are shown in Fig. 1(b) (JPEG compression) and Fig. 1(c) (additive white noise), respectively. The edge maps of the three images are shown underneath them in the second row of Fig. 1, while the NSE maps of the two distorted image are shown in the fourth column of Fig. 1. We see that although the edge maps are very different, interestingly the NSE maps of the two distorted images are very similar to each other, which is identical to the objective quality evaluation. This example clearly demonstrates that NSE can be used to evaluate the perceptual quality of distorted images.

B. Non-Shift Edge Based Ratio (NSER)

The variation of the number of edge points in NSE can be used to measure the image quality. Clearly, the more serious the distortion is, the fewer points the NSE map will have. By considering the different contents in different images, the number of edge points in NSE should be normalized by that in the reference image. Hence, a novel and simple FR IQA metric, namely Non-Shift Edge based Ratio (NSER), is constructed in this letter. The proposed method consists of mainly four steps, which are as follows.

- Step 1) For the given reference (C) and distorted (D) images, the LOG filter defined in (1) on different scales σ_i are applied to generate the outputs C_i and D_i ($1 \leq i \leq N$), respectively.

TABLE I
LINEAR CORRELATION COEFFICIENT (CC) AFTER NONLINEAR REGRESSION

CC	NSER	VIF	VSNR	IFC	MS-SSIM	SSIM	NQM	UQI	PSNR
IVC	0.9080	0.9028	0.8032	0.908	0.8934	0.9119	0.8498	0.8325	0.7199
LIVE	0.9395	0.9598	0.9231	0.9248	0.9430	0.9449	0.9122	0.8974	0.8721
MICT	0.9278	0.9144	0.871	0.8434	0.8935	0.8887	0.8955	0.7208	0.6426
A57	0.8557	0.6158	0.9472	0.4548	0.8504	0.8017	0.802	0.5578	0.6587
TID2008	0.7957	0.8090	0.6820	0.7359	0.8425	0.7732	0.6135	0.6623	0.5309
CSIQ	0.9473	0.9277	0.8002	0.8381	0.8998	0.8613	0.7433	0.8304	0.8001
MEAN	0.8957	0.8549	0.8378	0.7817	0.8871	0.8637	0.8027	0.7502	0.7041

TABLE II
SPEARMAN RANK ORDER CORRELATION COEFFICIENT (SROCC)

SROCC	NSER	VIF	VSNR	IFC	MS-SSIM	SSIM	NQM	UQI	PSNR
IVC	0.8981	0.8966	0.7983	0.8978	0.8847	0.9018	0.8347	0.8267	0.6885
LIVE	0.9419	0.9631	0.9274	0.9234	0.9445	0.9479	0.9086	0.8933	0.8755
MICT	0.9247	0.9086	0.8614	0.8387	0.8864	0.8794	0.8911	0.6969	0.613
A57	0.8249	0.6223	0.9355	0.3185	0.8394	0.8066	0.7981	0.4253	0.6189
TID2008	0.7404	0.7496	0.7046	0.5692	0.8528	0.7749	0.6243	0.6004	0.5245
CSIQ	0.9337	0.9193	0.8106	0.7482	0.9138	0.8756	0.7402	0.8098	0.8057
MEAN	0.8773	0.8433	0.8396	0.7160	0.8869	0.8644	0.7995	0.7087	0.6877

- Step 2) Edge points are identified by locating the zero-crossings in C_i and D_i on all N scales according to (2), and on each scale σ_i , the same threshold TH_i is used to detect the edges in C_i and D_i . The edge maps of reference and distorted images on scale i are denoted as E_{C_i} and E_{D_i} , respectively.
- Step 3) The NSE map at scale σ_i is detected by $E_{C_i} \cap E_{D_i}$. Denote by $\|\bullet\|$ the operator that counts the number of edge points in an NSE map. A ratio to measure the similarity between C_i and D_i , denoted by p_i , is defined as

$$p_i = \frac{\|E_{C_i} \cap E_{D_i}\|}{\|E_{C_i}\|}. \quad (3)$$

Clearly the range of p_i is within $[0, 1]$. The more similar the two edge maps are to each other, the higher the value of p_i is. Obviously, the degradation of the distorted image is related to the value of $1 - p_i$.

- Step 4) $1 - p_i$ across all scales are combined by their product to yield the overall measurement. And we use the logarithmic function to improve its linear correlation to the subjective score. The use of this function is based on our experiment study:

$$\text{NSER}(C, D) = - \sum_{i=1}^N \log_{10}(1 - p_i). \quad (4)$$

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Settings

Six publicly available IQA databases are used to evaluate the proposed NSER metric. They are the LIVE database at <http://live.ece.utexas.edu/research/quality>, the IVC database at <http://www2.irccyn.ecnantes.fr/ivcdb/>, the A57 database at <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>, the MICT database at <http://mict.eng.u-toyama.ac.jp/mict/index2.html>, the CSIQ database at <http://vision.okstate.edu/csiq/>, and the TID2008 database [16]. According to the VQEG Phase-II

[17], a five-parameter nonlinear function mapping between the objective score and the subjective scores is defined as follows:

$$\text{Quality}(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5. \quad (5)$$

The proposed algorithm is compared with seven state-of-the-art IQA metrics of different classes: VIF [9] and IFC [8] which are based on the information theory framework, SSIM [6], MS-SSIM [7] and UQI [5] which are based on the structural distortion, VSNR [4] and NQM [3] which are based on the HVS model, as well as the L2 distance based PSNR. All of them work on the luminance component only. The used codes of all metrics are available at http://foulard.ece.cornell.edu/gaubatz/metrix_mux/, except for SSIM which is available at <http://www.ece.uwaterloo.ca/~z70wang/research/ssim/ssim.m>. For the proposed algorithm, five LOG filters are used and the scales σ of these filters are selected as $[0.5 \ 1.3 \ 2.6 \ 5.2 \ 10.4]$. The thresholds for zero-crossing detection on these scales are selected as $[0.6 \ 0.4 \ 0.2 \ 0.08 \ 0.02]$. It should be noted that the above parameter selection is not optimized, and we empirically found that NSER is not sensitive to parameters in a wide range.

B. Experimental Results and Discussions

Tables I and II summarize the validation results by the competing IQA metrics. To better compare the performance of these metrics, the mean score across all the databases for each metric is listed in the last row of Tables I and II, respectively. The best three metrics producing the greatest correlations for each database are marked in bold. As can be seen from the two Tables, the proposed method provides relatively stable results across all the six databases. MS-SSIM, SSIM, and VIF are also noticeable.

Based on the mean linear correlation coefficient (CC), the best three metrics are NSER (0.8957), MS-SSIM (0.8871), and SSIM (0.8637), while based on the mean spearman rank order correlation coefficient (SROCC), the best three are

MS-SSIM (0.8869), NSER (0.8773), and SSIM (0.8644). This demonstrates that NSER and MS-SSIM are more robust than other metrics. More specifically, NSER is among the best three metrics on IVC, MICT, A57, TID2008, and CSIQ databases in terms of CC, and on IVC, MICT, A57, and CSIQ in terms of SROCC. Even on LIVE database on which NSER ranks the fourth in terms of both CC and SROCC, its index (CC: 0.9395, SROCC: 0.9419) is very close to that of MS-SSIM (rank third, CC: 0.9430, SROCC: 0.9445) and SSIM (rank second, CC: 0.9449, SROCC: 0.9479). Overall, as an FR IQA metric, NSER is better than most of state-of-the-art metrics.

From the experimental results, we can see that both MS-SSIM and NSER perform very well across all the databases. But they are very different from each other. Specifically, MS-SSIM mimics functionally the IQA of HVS to build the metric [6], [7], while NSER uses only the early vision features (i.e., edges) in the IQA metric design. On the other hand, MS-SSIM includes three distortion components: luminance, contrast and the so called structural-similarity, among which the structural-similarity is the core factor, while NSER uses only the binary edge maps to measure the image quality in the form of NSE that can be considered as the “structural-similarity” in some sense. Interestingly, by using only the primitive zero-crossings, NSER still achieves comparable performance to MS-SSIM. This fact shows that zero-crossings can be very effective and efficient for IQA. The NSE detection actually selects the most significant features in the reference and distorted images, and it eliminates much information redundancy in the image. The information lost in the binary edge detection process is not so important for IQA.

A natural scene is constituted by spatially distributed structure features. The pixels belonging to a structure are related to each other with a specific intensity distribution, and the information the structure carries is hidden behind this distribution. When an image is deteriorated, the structure varies, and the distribution varies accordingly. This is why the structural similarity indexes [5]–[7] and the information fidelity criteria [8], [9] work well for IQA. On the other hand, the image structure features used by the above IQA metrics are constructed from the basic primitive signals generated by ganglion and LNG neurons, and by Marr’s theory [10], the information existed in the basic primitive signals can be represented by the zero-crossings and their spatial distribution. Thus the structural variation caused by the image distortion will lead to the change of spatial distribution of zero-crossings, and this change can be expressed and measured by using the so-called NSE map and NSER metric. In summary, the proposed NSER is able to measure the structural change of images in statistics, and our experiments showed that it is a simple but effective index to measure the perceptual distortion of image structures.

IV. CONCLUSION

This letter presented a novel image quality assessment (IQA) metric, namely the Non-Shift Edge based Ratio (NSER), which operates on the low-level early vision features, more specifically zero-crossing edges according to Marr’s theory. The framework of the proposed metric is straightforward and very simple; however, extensively experiments indicated that the proposed metric works robustly across different IQA databases. It achieves better performance than or comparable performance to state-of-the-art IQA metrics, such as MS-SSIM. The work in this letter reveals the high correlation between visual perception of image quality and the so called NSE map.

REFERENCES

- [1] S. Daly, “The visible difference predictor: An algorithm for the assessment of image fidelity,” in *Proc. SPIE*, 1992, vol. 1616, pp. 2–15.
- [2] “Sarnoff Corporation,” JNDmetrix Technology [Online]. Available: http://www.sarnoff.com/productservices/video_vision/jndmetrix/
- [3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 636–650, Apr. 2000.
- [4] D. M. Chandler and S. S. Hemami, “VSNR: A wavelet-based visual signal-to-noise ratio for natural images,” *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [5] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Sep. 2002.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error measurement to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” in *Proc. IEEE Conf. Signals, Systems, and Computers*, 2003, pp. 1398–1402.
- [8] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
- [9] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [10] D. Marr, *Vision*. New York: W. H. Freeman, 1980.
- [11] R. Hummel and R. Moniot, “Reconstruction from zero-crossings in scale-space,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2111–2130, Dec. 1989.
- [12] J. Elder, “Are edges incomplete?,” *Int. J. Comput. Vis.*, vol. 34, no. 2/3, pp. 97–122, 1999.
- [13] M. Zhang and X. Q. Mou, “A psychovisual image quality metric based on multi-scale edge structure similarity,” in *Proc. ICIP*, 2008, pp. 381–384.
- [14] G. H. Chen, C. L. Yang, and S. L. Xie, “Gradient-based structural similarity for image quality assessment,” in *Proc. ICIP*, 2006, pp. 2929–2932.
- [15] D. M. Rouse and S. S. Hemami, “Natural image utility assessment using image contours,” in *Proc. ICIP*, 2009, pp. 2217–2220.
- [16] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, “Color image database for evaluation of image quality metrics,” in *Proc. MMSP*, Cairns, Australia, 2008, pp. 403–408.
- [17] “VQEG,” Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment Apr. 2000 [Online]. Available: <http://www.vqeg.org/>