# LETTER Image Quality Assessment Based on Multi-Order Visual Comparison

# Fei ZHOU<sup>†a)</sup>, Member, Wen SUN<sup>†b)</sup>, and Qingmin LIAO<sup>†c)</sup>, Nonmembers

**SUMMARY** A new scheme based on multi-order visual comparison is proposed for full-reference image quality assessment. Inspired by the observation that various image derivatives have great but different effects on visual perception, we perform respective comparison on different orders of image derivatives. To obtain an overall image quality score, we adaptively integrate the results of different comparisons via a perceptioninspired strategy. Experimental results on public databases demonstrate that the proposed method is more competitive than some state-of-the-art methods, benchmarked against subjective assessment given by human beings.

key words: image quality assessment, image derivatives, multi-order visual comparison

# 1. Introduction

Image quality assessment (IQA) has attracted increasing interest due to its importance in image acquisition, transmission, display, etc. Although subjective evaluation is the most reliable way of IQA, it is time-consuming, laborious, and expensive. Hence, it is necessary to develop objective IQA metrics that can automatically measure image quality and well approximate subjective scores. According to the availability of a reference image, objective metrics can be classified as full-reference, no-reference, and reduced-reference methods [1]–[4]. In this letter, we focus on the problem of full-reference IQA.

The most conventional metrics involving meansquared error and peak signal-to-noise ratio (PSNR) have been widely criticized for not correlating well with subjective fidelity ratings. To address this problem, many efforts have been made on designing bottom-up models to imitate human visual system (HVS), e.g., visual signal-to-noise ratio (VSNR) [1]. However, most models are simplified based on a number of strong assumptions due to the complexity of HVS. Recently, many researchers prefer to regard HVS as a black box. That is, it is reasonable to achieve IQA by measuring up-bottom similarities. Inspired by this view, structural similarity (SSIM) index [2], feature similarity (FSIM) index [3], and gradient similarity (GSIM) index [4] are designed for full-reference IQA. Therein, the key problem is

DOI: 10.1587/transinf.E97.D.1379

how to design and measure the up-bottom similarity. In this Letter, the motivation is from our recent work in the field of super-resolution [5], which shows that image details can be well represented by high-order image derivatives. Accordingly, we measure the up-bottom similarity by multi-order visual comparison. The visual comparisons are performed on different image derivatives and then pooled into a single similarity index. Experimental results demonstrate the superiority of our method over some state-of-the-art methods.

#### 2. Proposed Scheme

# 2.1 Multi-Order Visual Comparison

It is observed that visual responses to different orders of image derivatives are nontrivial and distinct. To be specific, lower-order information mainly acts on overall perception while higher-order information generally determines the visual perception of image details [5]. Therefore, the measurement of up-bottom similarity should compare multi-order information as well as distinguish them. In this work, we investigate zero-order, first-order, and second-order image derivatives. Given a test image f, second-order information is calculated as second-order derivatives of f:

$$f_{x2}(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2},$$
  

$$f_{y2}(x, y) = \frac{\partial^2 f(x, y)}{\partial y^2},$$
  

$$f_{xy}(x, y) = \frac{\partial^2 f(x, y)}{\partial xy},$$
  
(1)

where *x* and *y* represent abscissas and ordinates respectively, and  $f_{x2}$ ,  $f_{y2}$ , and  $f_{xy}$  denote the second-order information of *f*. To simplify the formulas in this Letter, we generally omit the arguments of a function after the definition, e.g.,  $f_{x2}(x, y)$ will be abbreviated as  $f_{x2}$ . To get first-order information, we remove second-order derivatives from image gradients:

$$f_{x1}(x,y) = \frac{\partial f(x,y)}{\partial x} - \sqrt{f_{x2}^2 + f_{xy}^2},$$
  

$$f_{y1}(x,y) = \frac{\partial f(x,y)}{\partial y} - \sqrt{f_{y2}^2 + f_{xy}^2},$$
(2)

where  $f_{x1}$  and  $f_{y1}$  denote the first-order information of f. Similarly, zero-order information  $f_0$  is given by

$$f_0(x,y) = f(x,y) - \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2}.$$
 (3)

Manuscript received November 18, 2013.

<sup>&</sup>lt;sup>†</sup>The authors are with Shenzhen Key Laboratory of Information Science and Technology, Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, P. R. China.

a) E-mail: flyingzhou@sz.tsinghua.edu.cn

b) E-mail: 1213755392@qq.com

c) E-mail: liaoqm@tsinghua.edu.cn

In the proposed scheme, to compute of image gradients (first-order image derivatives) along horizontal and vertical directions, we convolve images with a pair of Scharr operators. And second-order image derivatives can be estimated in a similar way by convolving image gradients with Scharr operators. For the reference image g, we can also define its multi-order visual information in the same way, and denote them as  $g_{x2}$ ,  $g_{y2}$ ,  $g_{xy}$ ,  $g_{x1}$ ,  $g_{y1}$ , and  $g_0$  respectively.

Based on Eqs. (1)–(3), we define the measurement functions for the respective comparison of multi-order information as follows:

$$s_{0}(f,g,x,y) = \frac{2 \cdot \left|\overline{f_{0}} \cdot \overline{g_{0}}\right| + C_{0}}{\overline{f_{0}}^{2} + \overline{g_{0}}^{2} + C_{0}} \cdot \frac{2 \cdot \sigma_{f0} \cdot \sigma_{g0} + C_{0}}{\sigma_{f0}^{2} + \sigma_{g0}^{2} + C_{0}},$$

$$s_{1}(f,g,x,y) = \frac{2 \cdot |f_{x1} \cdot g_{x1}| + C_{1}}{f_{x1}^{2} + g_{x1}^{2} + C_{1}} \cdot \frac{2 \cdot |f_{y1} \cdot g_{y1}| + C_{1}}{f_{y1}^{2} + g_{y1}^{2} + C_{1}},$$

$$s_{2}(f,g,x,y) = \frac{2 \cdot |f_{x2} \cdot g_{x2}| + C_{2}}{f_{x2}^{2} + g_{x2}^{2} + C_{2}} \cdot \frac{2 \cdot |f_{y2} \cdot g_{y2}| + C_{2}}{f_{y2}^{2} + g_{y2}^{2} + C_{2}} \cdot \frac{2 \cdot |f_{xy} \cdot g_{xy}| + C_{2}}{f_{y2}^{2} + g_{y2}^{2} + C_{2}},$$

$$(4)$$

where  $s_0$ ,  $s_1$ , and  $s_2$  are the visual comparisons of zeroorder, first-order, and second-order information,  $\bar{f}_0$ ,  $\bar{g}_0$  and  $\sigma_{f0}$ ,  $\sigma_{g0}$  are the means and standard deviations of the patch centred at (x, y) in  $f_0$  and  $g_0$  respectively, and  $C_0$ ,  $C_1$ , and  $C_2$  are constants for the stability to avoid a nearly zero denominator. Specifically,  $C_0$ ,  $C_1$ , and  $C_2$  are respectively calculated as  $(K_1 \times L)^2$ ,  $(K_2 \times L)^2$ , and  $(K_3 \times L)^2$ , where *L* is the dynamic range of pixel values (255 for 8-bit grayscale image),  $K_1$ ,  $K_2$ ,  $K_3$  are constants much smaller than 1. The values of  $K_1$ ,  $K_2$ , and  $K_3$  will be provided in experimental part.

About Eq. (4), we can further declare the following points: First, it is easy to verify that all the expressions in Eq. (4) are consistent with the masking effect of HVS [2], [4]. Secondly, the expression  $s_n$  (*n* is 0, 1, or 2) is a symmetric metric which ranges from 0 to 1. And it achieves the maximum value 1 if and only if  $n^{\text{th}}$ -order information of the test and reference image is identical. Thirdly, larger value of  $s_n(f, g, x, y)$  implies higher quality at the position of (x, y) in terms of  $n^{\text{th}}$ -order information.

#### 2.2 Pooling

To assess image quality, a single index is necessary. However, the measurement functions in Eq. (4) are performed on respective order in a pixel-wise fashion. Therefore, it is required to pool them into an overall score. To achieve this, we first need to combine  $s_0$ ,  $s_1$ , and  $s_2$ . It has been proven that the simultaneous existence of multiple distortion components at a given position will mask the perception of each other [4]. In other words, the smaller distortion will be masked by the larger distortion. Hence, we obtain the combined measurement function *s* by

$$s(f, g, x, y) = \frac{1}{N_1} \cdot \left( \sum_{n \neq 0} s_n \cdot s_0 + \sum_{n \neq 1} s_n \cdot s_1 + \sum_{n \neq 2} s_n \cdot s_2 \right),$$
(5)

where  $N_1$  is a normalization constant to ensure the sum of weights equal to 1. Finally, we integrate the measurements *s* at every pixel position into a single score. Since humans are more sensitive to severely distorted regions, we calculate the overall image quality *q* as

$$q(f,g) = \frac{1}{N_2} \sum_{x} \sum_{y} (1 - s(f,g,x,y)) \cdot s(f,g,x,y), \quad (6)$$

where  $N_2$  is also a normalization constant. In both (5) and (6), smaller distortions result in smaller weights. However, the inspirations are different. The weights in (5) are based on the masking effect of HVS while the weights in (6) are inspired by the visual attention.

## 3. Experimental Results

Experiments are conducted on three publicly available and subject-rated databases, known as TID2008 [6], LIVE [7], and MICT [8]. In TID2008, there are 25 original images and 1700 test images with 17 types of distortions. And LIVE database has 29 reference images and 779 distorted images, including five distortion types. MICT database contains 14 original images and 168 distorted images with two types of distortions. Moreover, mean opinion score (MOS) is available for TID2008 and MICT while differential mean opinion score (DMOS) is provided in LIVE. Following guidelines of the Video Quality Experts Group [9], we use a five-parameter logistic function to map the objective predictions q to the subjective scores. The used function has the form of

$$p(q) = \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (q - \beta_3))}\right) + \beta_4 \cdot q + \beta_5,$$
(7)

where the model parameters  $\beta_i$  (i = 1, 2, 3, 4, 5) are chosen to minimize the squared error between the subjective scores and fitted objective scores. In our experiments, the small constants  $K_1$ ,  $K_2$ ,  $K_3$  mentioned in Sect. 2.1 are set to 0.01, 0.1, and 0.1, respectively. It is worthwhile to notice that  $K_2$  and  $K_3$  are larger than  $K_1$ . The reasons are twofold: First, the dynamic range of gradients is larger than that of pixel values. Secondly, image derivatives smaller than just noticeable difference would not be perceived by human beings. Therefore, in order to avoid over-estimating visual distortions on the regions with small derivatives, constants  $C_2$ and  $C_3$  cannot be very small. In Fig. 1, we illustrate the scatter plots of the proposed IQA scheme.

To verify the effectiveness of the proposed method, we compare our predictions with those of PSNR, VSNR [1], SSIM [2], FSIM [3], and GSIM [4]. Quantitative comparisons are based on four criteria, including Spearman rank-order correlation coefficient (SROCC), Kendall rank-order correlation coefficient (KROCC), Pearson linear correlation



**Fig. 1** Scatter plots of subjective scores versus our object prediction *q* on different databases: (a) TID2008, (b) LIVE, (c) MICT.

coefficient (PLCC), and root mean-squared error (RMSE) between mapped objective scores and subjective scores. The first two criteria are used to evaluate prediction monotonicity and the other two can measure prediction accuracy [9]. Larger SROCC, KROCC, PLCC, and smaller RMSE signify better performances. The quantitative comparisons are shown in Table 1, from which we can find that the proposed method outperforms the compared methods on all the databases.

# 4. Conclusion

In this letter, we propose a full-reference IQA which per-

 Table 1
 Performance comparisons for IQA methods.

Database	Criteria	Methods					
		PSNR	VSNR[1]	SSIM[2]	FSIM[3]	GSIM[4]	Ours
TID2008	SROCC	0.5794	0.7049	0.7749	0.8805	0.8554	0.9013
	KROCC	0.4210	0.5345	0.5768	0.6946	0.6651	0.7292
	PLCC	0.5726	0.6823	0.7732	0.8738	0.8462	0.8956
	RMSE	1.1103	0.9810	0.8511	0.6525	0.7151	0.6053
LIVE	SROCC	0.8756	0.9280	0.9479	0.9634	0.9554	0.9719
	KROCC	0.6865	0.7625	0.7963	0.8337	0.8131	0.8468
	PLCC	0.8723	0.9237	0.9430	0.9597	0.9437	0.9681
	RMSE	13.360	10.469	9.0956	7.6780	9.0376	7.0802
MICT	SROCC	0.6130	0.8614	0.8795	0.9059	0.9233	0.9354
	KROCC	.04447	0.6762	0.6939	0.7302	0.7541	0.7671
	PLCC	0.6426	0.8710	0.8887	0.9078	0.9287	0.9413
	RMSE	0.9588	0.6147	0.5738	0.5248	0.4640	0.4350

forms visual comparisons of multi-order information. Inspired by the property of visual responses to different orders of image derivatives, we derive multi-order information from various image derivatives. The overall objective score is evaluated by an adaptive and perceptually inspired pooling. Experiments on three well-known databases have confirmed the effectiveness of the proposed method in comparison with the state-of-the-art methods.

#### Acknowledgments

The authors would like to thank all the editors and reviewers for their work on this letter. This work was in part supported by the Natural Science Foundation of China under Grant 61271393 and 61301183, and in part by China Postdoctoral Science Foundation under Grant 2013M540947.

# References

- M. Chandler and S. Hemami, "VSNR: A wavelet-based visual signalto-noise ratio for natural images," IEEE Trans. Image Process., vol.16, no.9, pp.2284–2298, 2007.
- [2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
- [3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, 2011.
- [4] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," IEEE Trans. Image Process., vol.21, no.4, pp.1500–1512, 2012.
- [5] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image superresolution using multisurface fitting," IEEE Trans. Image Process., vol.21, no.7, pp.3312–3318, 2012.
- [6] N. Ponomarenko and K, Egiazarian, "Tampere image database 2008 TID2008," http://www.ponomarenko.info/tid2008.htm/.
- [7] H.R. Sheikh, K. Seshadrinathan, A.K. Moorthy, Z. Wang, A.C. Bovik, and L.K. Cormack, "Live image quality assessment databases release," http://live.ece.utexas.edu/research/quality/.
- [8] Y. Horita, K. Shibata, Y. Kawayoke, and Z.M. Parvez Sazzad, "MICT Image Quality Evaluation Database," http://mict.eng.u-toyama.ac.jp/ mict/index2.html.
- [9] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org.