A NOVEL POOLING STRATEGY FOR FULL REFERENCE IMAGE QUALITY ASSESSMENT BASED ON HARMONIC MEANS

*Xinghao Ding, Zheng Zhang, Xin Chen, Yue Huang**

Department of Communication Engineering, Xiamen University, Fujian, China
yhuang2010@xmu.edu.cn

ABSTRACT

The most perceptual Full Reference Image Quality Assessment metrics (FR-IQA) shared a common two-step model; local quality measurement, and pooling. In this letter, a novel pooling strategy based on harmonic mean is proposed to predict the final quality score in FR-IQA. In contrast to arithmetic mean, the harmonic mean tends to emphasize the contributions from the local severely distorted regions or pixels in the definition of assessment function using reciprocal transformation. It is derived from the observations that humans visual attention is mostly affected with the region having severely distorted points or regions. In addition, the relationship of subjective visual quality with the quality score against different levels of distortion in the images is described as a non-linear procedure by introducing another reciprocal transformation in harmonic mean. The proposed pooling strategy is applied to some popular FR-IQA metrics, including SSIM, GSSIM, and FSIM. The experimental results have demonstrated that the metrics with proposed pooling strategy have better performances compared to the standard versions, especially on the images with small but seriously distorted regions. The proposed pooling strategy is computationally very efficient since only one averaging operation and two reciprocal transformations are required.

Index Terms— Image quality assessment, full reference, pooling strategy, harmonic mean

1. INTRODUCTION

According to the availability of a reference image, image quality assessment (IQA) models can be classified into full reference (FR) methods, reduced reference (RR) methods, and no reference (NR) methods [1][2][3]. This paper particularly focuses on the FR methods.Among the existing FR-IQA methods, most studies can be summarized into a general two-step frame-work; a local quality measurement step followed by the pooling stage [4][5], as illustrated in Fig.1. In the first step, usually a local quality map is computed from the feature maps of both the reference and the distorted image. In the second step, some popular pooling strategies are used to obtain the final quality score. The average pooling method is considered as an efficient and effective approach. It assumes that the distortion of each pixel has equal contribution to the image quality. For example, the famous structural similarity index (SSIM) [6][7] constructs three feature similarity maps at first stage, and use average pooling to predict the final score. It is obvious that the assumption ignores visual attention of the human vision system (HVS), which significantly depends on the seriously distorted points and regions. Zhang et al. [5] has recently proposed another pooling strategy that uses standard deviation to predict the image quality and achieve better performance compared to other IQA metrics. However, the standard deviation based pooling strategy is only adapted in the local quality map of the gradient magnitude feature. Since different regions and contents may affect the estimation of local quality, various weighting strategies were proposed based on the properties of HVS, these weighting strategies include [8]: information content weighting [4], region type weighting [9], visual fixation and quality based weighting [10], using actual visual attention information [11], and a feature map index (FSIM) [12]. These weighting strategies may achieve better performance but at the expense of higher computational complexity and longer run-time [3].

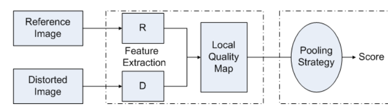
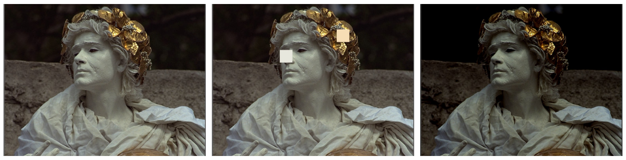


Fig. 1. The frame-work of the two-step FR-IQA models.

The proposed idea is based on two observations: 1) Human vision strongly depends on the seriously distorted regions than non-distorted regions, especially when the distorted regions occupy a very small area in the whole image. The image quality has a non-linear relationship with the distortion of each pixel. In many case, if an image is corrupted or distorted in a very small area and the remaining part of the image is

The project is supported by the National Natural Science Foundation of China (No. 30900328, 61172179, 61103121, 81301278), the Natural Science Foundation of Fujian Province of China (No. 2012J05160), The National Key Technology R&D Program (2012BAI07B06), the Fundamental Research Funds for the Central Universities (No. 2011121051, 2013121023), the NCETFJ.



(a) Refence Image (b) MSE=184.3 (c) MSE=1247.6

Fig. 2. The images are from the public database TID2008 [15]. (a) is a reference image, (b) is distorted with only two small white blocks, (c) is distorted with entirety intensity shift.

perfect, the image quality drastically decreases with the presence of some local seriously distorted pixels or regions in the images; the image quality of such an image is poorer than the image with global unserious distortion, even if the distorted regions are very small. 2) Subjective image quality against decreasing/increasing metrics is also a non-linear procedure. Actually the observation is validated with the definition of popular metrics mean square error (MSE) and peak signal-to-noise ratio (PSNR). PSNR, which can be considered as a non-linear function of MSE, always has better performances than MSE in image quality assessment [13].

Considering the perspective of human vision and computational cost, an efficient and effective pooling strategy with harmonic mean is proposed in this paper. In order to better exploit the advantages of the harmonic mean pooling, separate pooling is used for each feature in the multi-feature map generated in first stage. If there exists only one feature in the first step, the harmonic mean pooling can be directly applied. In this letter, the harmonic mean pooling strategy is applied to some popular methods, they are SSIM [6], gradient SSIM [14], and FSIM [12].

The contributions of the proposed work can be summarized as: 1) a reciprocal transformation is used for a better definition of subjective image quality, which is significantly affected by small but seriously distorted regions in the image; 2) another reciprocal transformation is used to represent the non-linear relationship between the subjective visual quality and the quality scores for different levels of distortion in the images; 3) the computational cost of proposed model is very efficient, since it only contains one average operation and two reciprocal transformations.

2. HARMONIC MEAN POOLING

For IQA metrics, the local quality map is obtained and normalized to (0, 1], where small values represent the poor local quality of the distorted images. The harmonic mean is one of the several types of averaging operators, and the harmonic mean H of the positive real numbers **x**=$[x_1,x_2,..,x_n]$ is defined as:

$$H(X) = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} \tag{1}$$

where N is the total number of values. The calculation of the harmonic mean can be divided into two steps, and the equation (1) can be rewritten as equation (2) and (3):

$$h(X) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{x_i} \tag{2}$$

$$H(X) = \frac{1}{h(X)} \tag{3}$$

In first step as modeled in equation (2), the values of the map are non-linearly transformed with the reciprocal operation and the arithmetic mean h is obtained. The smaller values will have higher impact in the reciprocal transformation resulting in larger contributions in the average operation of the equation (2). As a result, h experiences higher sensitivity against the smaller values of the map. At the same time, many studies [10][11][16] have demonstrated that the worst sections of the images attract more of the visual attention of the observers. In other words, a small region of severely distorted pixels will have significantly high influence on the overall image quality. As shown in Fig.2 where: (a) is a reference image; (b) and (c) are the distorted images with different distortion types; (b) is distorted with two small white blocks, while (c) is distorted with an intensity shift in the entire image. Although (c) is has much higher mean square error (MSE) compared to (b), most people will agree that (c) has better image quality in terms of subjective evaluation. The Fig.2 explains that a small part of poor sections is very important for visual attention. Therefore, the non-linear transformation is very suitable to cover the relationship between the local worst sections and the whole image quality.

In the second step as modeled in equation (3), the second reciprocal transformation is incorporated to obtain the final pooling result H. The second reciprocal transformation incorporates the non-linear relationship between different distorted images. It is used to improve the linear consistency between the objective evaluation and the subjective evaluation. An experiment is conducted to demonstrate that the linear consistency with subjective evaluation of H is better than that of h.

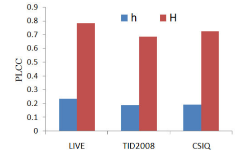


Fig. 3. The PLCC of using h and H to evaluate on LIVE, TID2008, CSIQ databases.

The harmonic mean pooling is used with the most popular method SSIM [6], where *h* and *H* are used to evaluate the respective image qualities. The non-linear regression method, Pearson liner correlation coefficient (PLCC) is used to evaluate the performance on three large databases (LIVE [17], TID2008 [15], and CSIQ [18]), as shown in Fig. 3. It can be seen that *H* is quite better than *h* on all of three databases; it is evidently improved by the second reciprocal transformation that incorporates the non-linear relationship between the images.

3. FR-IQA ALGORITHMS WITH PROPOSED STRATEGY

The final quality score (QS) using harmonic mean pooling is computed after the extraction of the local quality map, simply as:

$$QS = \sum\nolimits_{j=1}^{M} \omega_j H(X_j) \tag{4}$$

Where M is the number of similarity feature maps, X_j is the similarity feature map, ω_j is the weight that represent the significance of the feature map X_j. The higher value of QS represents a better image perceptual quality.

3.1. SSIM with Harmonic Mean Pooling

The most popular IQA algorithm is the structural similarity index (SSIM) [6][7]; the SSIM index uses three separate comparisons of the local luminance (*l*), contrast (*c*), and structure (*s*) between the original and distorted image. The SSIM is defined as:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{5}$$

where x and y represent two local image patches that are extracted from the original and distorted images, respectively. α , β and γ are the parameters used to adjust the importance of the three components. The overall SSIM value of the whole image is obtained simply by averaging (arithmetic mean) the SSIM map. The SSIM index contains three feature maps that are finally grouped together by multiplying them. In order to better exploit the benefits of the harmonic mean pooling, the harmonic mean pooling is applied to each of the three feature maps. Therefore, the new formulated SSIM index using harmonic mean pooling (HM-SSIM) is defined as:

$$HM-SSIM = \omega_1 H(C) + \omega_2 H(L) + \omega_3 H(S) \tag{6}$$

where *C*, *L*, and *S* are the three feature similarity maps; ω_1 ,ω_2 and ω_3 are the weights of the feature importance.

3.2. GSSIM with Harmonic Mean Pooling

The Gradient-based Structural Similarity (GSSIM) [14] is an improved image quality assessment based on edge information. It extracts the edge information using gradient operators to construct the gradient magnitude map; the structural similarity is them calculated using similar technique as in the SSIM. Similar to the HM-SSIM, the new formulated GSSIM index using harmonic mean pooling (HM-GSSIM) is defined as:

$$HM-GSSIM = \omega_1 H(C_g) + \omega_2 H(L_g) + \omega_3 H(S_g) \tag{7}$$

Where C_g, L_g, and S_g are the three feature similarity maps; ω_1, ω_2 and ω_3 are the weights of the feature importance.

Table 1. The comparisons of the original and new formulated merics .

Model	SROCC	KROCC	PLCC	RMSE
	CSIQ	Database		
SSIM	0.876	0.691	0.861	0.133
HM-SSIM	**0.941**	**0.781**	**0.938**	**0.091**
GSSIM	0.872	0.686	0.861	0.133
HM-GSSIM	**0.904**	**0.723**	**0.910**	**0.109**
FSIM	0.924	0.757	0.912	0.108
HM-FSIM	**0.947**	**0.792**	**0.942**	**0.088**
	TID2008	Database		
Model	SROCC	KROCC	PLCC	RMSE
SSIM	0.775	0.577	0.773	0.851
HM-SSIM	**0.832**	**0.632**	**0.820**	**0.769**
GSSIM	0.731	0.569	0.762	0.873
HM-GSSIM	**0.826**	**0.624**	**0.819**	**0.771**
FSIM	0.881	0.695	0.874	**0.653**
HM-FSIM	**0.893**	**0.703**	**0.873**	0.656
	LIVE2	Database		
Model	SROCC	KROCC	PLCC	RMSE
SSIM	0.948	0.796	0.945	8.946
HM-SSIM	**0.953**	**0.805**	**0.949**	**8.636**
GSSIM	0.918	0.765	0.920	10.740
HM-GSSIM	**0.943**	**0.784**	**0.937**	**9.477**
FSIM	**0.963**	**0.834**	**0.960**	**7.674**
HM-FSIM	0.962	0.829	**0.960**	7.696

3.3. FSIM with Harmonic Mean Pooling

Zhang et al. proposed a feature similarity (FSIM) index [12] for full reference IQA based on the fact that human visual system (HVS) understands an image mainly according to its low-level features. Specifically, the phase congruency (PC) is used as the primary feature in FSIM. In addition, the image gradient magnitude (GM) is employed as the secondary feature in the FSIM. The local quality similarity map is computed as:

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{{PC_1}^2(x) + {PC_2}^2(x) + T_1} \tag{8}$$

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{{G_1}^2(x) + {G_2}^2(x) + T_2} \tag{9}$$

After obtaining the local quality map, PC is used again as a weighting function to derive a single quality score. Similar to the HM-SSIM, the harmonic mean pooling is applied separately using PC and GM features. After obtaining the local quality map, the harmonic mean pooling is used with FSIM (HM-FSIM) as:

$$HM\text{−}FSIM = \omega_1 H(S_{PC}) + \omega_2 H(S_G) \tag{10}$$

Where ω_1 and ω_2 are the weight of the feature importance.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In order to verify the proposed IQA model, experiments were conducted on three large scale publicly available databases: LIVE [17], TID2008 [15], CSIQ [18]. Usually four important criteria are used to compare the performance in the evaluation of the IQA metrics. The Spearman rank-order correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC) are used to estimate the test agreement between the DMOS and model predictions. The Pearson liner correlation coefficient (PLCC) is used to evaluate the prediction accuracy, and the root mean square error (RMSE) is used to evaluate the prediction consistency.

For HM-SSIM and HM-GSSIM, the parameters of the first step are not change, in the pooling strategy, ω_1=0, ω_2=ω_3 =0.5. During the experiment, we find that the structural feature is much important than local luminance feature, and the paper [16][19] demonstrated that ignoring the luminance comparison produces no drop in metric performance. For HM-FSIM, we set the parameters as ω_1=ω_2=0.5. The proposed new formulated metrics and their original metrics performance is presented in Table I.

In terms of the prediction performance observed in the Table I, most of the new metrics with harmonic mean pooling perform better than the original metrics. It can be noticed that the harmonic mean pooling improves the performance of the popular IQA metrics. Moreover, the pooling strategy is very easy to apply on the other IQA metrics that involve the two-step procedure shown in Fig.1.

The experiments in Fig.4 are used to demonstrate that the pooling strategy is having a better performance especially for the images with small but serious distortion. As shown in Fig.4, two examples are presented to evaluate the difference of the new formulated and original metrics. The images (b) and (e) have small patches of heavily distorted regions; (c) and (f) are distorted overall with blur and noise. The comparison results of distorted images in Fig.3 are presented in Table II.

It can be observed from Table II and Fig.3 that the harmonic mean pooling offers a better performance in the images with small patches of heavy distortion. The small patches of heavy distortion would have a highly negative impression on human vision system and subjective assessment. From Table II, it can be noticed that SSIM, GSSIM, and FSIM produced a wrong assessment of the image quality, but we can see that HM-SSIM, HM-GSSIM and HM-FSIM get the correct evaluation.

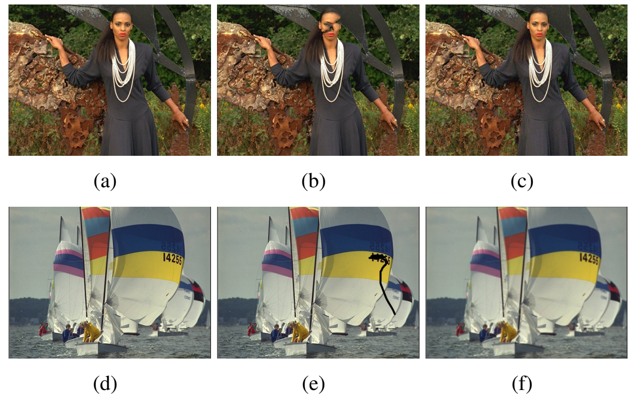


Fig. 4. (a)and (d) are reference images, and the others are distorted image. (b) and (e) are only with small part heavily distorted sections. (c) and (f) are with whole distorted of blur and noise. Most people may agree that (c) is better than (b), and (f) is better than (e).

Table 2. The comparison scores of distorted images in Fig.4.

Model	(b)	(c)	(e)	(f)
SSIM	0.988	0.981	0.980	0.957
HM-SSIM	0.974	0.990	0.861	0.977
GSSIM	0.977	0.941	0.976	0.799
HM-GSSIM	0.928	0.970	0.767	0.932
FSIM	0.993	0.992	0.978	0.962
HM-FSIM	0.995	0.996	0.925	0.978

5. CONCLUSIONS

In this paper, a novel pooling strategy based on harmonic mean is proposed to calculate the final quality score on the local quality map. A reciprocal transformation is used to emphasize the larger weight of the local seriously distorted regions in the averaging. Another reciprocal transformation is applied to describe the non-linear relationship between the visual quality and the quality scores. Experimental results have indicated that the proposed pooling strategy achieves better performances than the original popular metrics used in the subject validation. In addition, the proposed pooling strategy is also very easy and efficient to implement in the metrics definition, since only few simple operations are required.

6. REFERENCES

[1] D.M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *ISRN Signal Processing*, vol. 2013, pp. 53, 2013.

[2] W. Lin and C.C. Jay, Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proceedings of ICIP*. IEEE, 2012.

[4] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.

[5] W. Xue, L. Zhang, X. Mou, and A.C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing,*, vol. 23, no. 5, pp. 684–695, 2014.

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[7] W. Zhou, Alan C.B., Hamid R.S., and P.S. Eero, "The ssim index for image quality assessment," [online] Available: http://www.cns.nyu.edu/ lcv/ssim/.

[8] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *IEEE International Conference on Image Processing*. IEEE, 2006, pp. 2945–2948.

[9] C. Li and A.C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 517–526, 2010.

[10] A.K. Moorthy and A.C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[11] A. Ninassi, M.O. Le, C.P. Le, and D. Barbba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in *IEEE International Conference on Image Processing*. IEEE, 2007, vol. 2, pp. 169–172.

[12] L. Zhang, D. Zhang, and X. Mou, "Fsim: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[13] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[14] G.H. Chen, C.L. Yang, and S.L. Xie, "Gradient-based structural similarity for image quality assessment," in *IEEE International Conference on Image Processing*. IEEE, 2006, pp. 2929–2932.

[15] N. Ponomarenko, V. Lukin, A. Zelensky, E. Karen, Jaakko A., C. Marco, and B. Federica, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[16] B.P. Bondzulic and V.S. Petrovic, "Additive models and separable pooling, a new look at structural similarity," *Signal Processing*, vol. 97, pp. 110–116, 2014.

[17] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 2," 2005, [online] Available: http://live.ece.utexas.edu/research/quality.

[18] E.C. Larson and D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006:1–011006:21, 2010.

[19] D.M. Rouse and S.S. Hemami, "Understanding and simplifying the structural similarity metric," in *IEEE International Conference on Image Processing*. IEEE, 2008, pp. 1188–1191.