PERCEPTUAL IMAGE QUALITY ASSESSMENT USING BLOCK-BASED MULTI-METRIC FUSION (BMMF)

Lina Jin[*], *Karen Egiazarian*[*] *and C.-C. Jay Kuo*[#]

[*]Department of Signal Processing, Tampere University of Technology, Finland
[#]Ming-Hsieh Department of Electrical Engineering, University of Southern California, USA

ABSTRACT

A new block-based multi-metric fusion (BMMF) approach is proposed for perceptual image quality assessment. The proposed BMMF scheme automatically detects image content and distortion types in a block via machine learning, which is motivated by the observation that the performance of an image quality metric is highly influenced by these factors. Locally, image block content is classified into three types; namely, smooth, edge and texture. Image distortion is detected and grouped into five types. An appropriate image quality metric is adopted for each block by considering its content and distortion types, and then all block-based quality metrics are fused to result in one final score. Furthermore, a corrected version of BMMF is derived for a specific group of distortions based on image complexity analysis. The proposed BMMF scheme is tested on TID database with its Spearman Correlation equal to 0.9471, which outperforms today's state-of-the-art image quality metrics.

***Index Terms*—** Image quality assessment, MMF, BMMF.

1. INTRODUCTION

Perceptual image quality metrics play an important role in many visual applications, *e.g.,* lossy compression of images and video, image denoising, watermarking, etc. The goal of objective image quality assessment (QA) is to automatically evaluate image and video quality that is consistent with human visual perception. In some applications, a reference image (or image sequence) is available in the evaluation of its counterpart under a negative influence of distortion. The derived metrics are called full-reference (FR) quality metrics. PSNR and MSE are two widely used objective quality metrics. However, they do not correlate with subjective human assessment very well.

Research on objective QA metrics has been active in the last decade, and a large number of metrics have been proposed. To compare the performance of different QA metrics, the TID image database [1] was established. It consists of 25 reference images and 1700 distorted images. There are 17 distortion types which often appear in digital image processing applications. The performance analysis of several state-of-the-art image QA metrics was conducted in [2] for the whole TID with a particular subset of distortion types. It was observed in [2] that no single QA metric can perform the best with respect to all possible image contents and distortion types. The Spearman Correlation of the well known metric, MS-SSIM [3], is about 0.85 for the whole TID database. Recently, several other objective QA metrics such as PSNR-H(M)A [4], sPHVS [5], IW-SSIM [6] have been developed. Their Spearman Correlations are around 0.86 for the TID database. Up to now, the best single image QA metric for TID is FSIM [7], which has the Spearman Correlation equal to 0.88.

A block-based multi-metric fusion (BMMF) approach is proposed to develop a new image QA metric in this work. The BMMF metric automatically detects image content and distortion types in a block via machine learning and selects the most suitable QA metric accordingly. Finally, all block-based quality metrics are fused to result in one final score. More details of the BMMF metric will be presented in Sec. 2 and its performance evaluation is conducted in Sec. 3. Finally, concluding remarks are given in Sec. 4.

2. PROPOSED BMMF QA METRIC

Since the Human Visual System (HVS) selects parts of visual contents for analysis and responds, image QA metrics are influenced by the content and distortion types of a local image region. Being motivated by this observation, we decompose images into blocks of a smaller size, classify them into three types (*i.e.* smooth, edge and texture), and select a suitable QA metric for each region accordingly. The block-diagram of the proposed BMMF metric is shown in Fig. 1. Each module is described in one of the following sections.

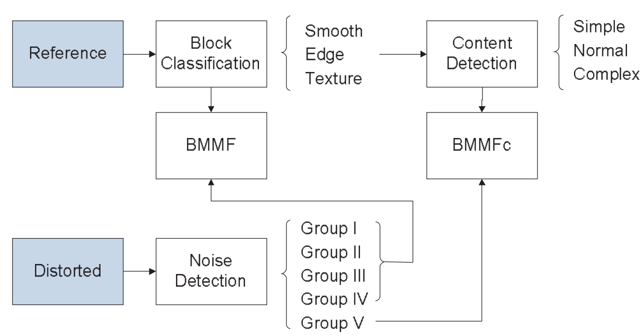
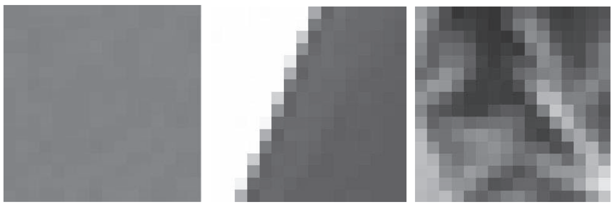


Figure 1. The block-diagram of the proposed BMMF QA metric.

2.1 Image Block Classification

Image blocks in a reference image are classified into three types: smooth, edge and texture blocks. A smooth block is one without obvious intensity variation. An edge block contains two intensity levels with a narrow transition interval. A texture block contains a certain amount of intensity variations. Examples of smooth, edge and texture blocks of size 16x16 are given in Fig. 2, respectively.



a) smooth b) edge c) texture

Figure 2. An example of three block types

There exist many block classification algorithms. Here, we adopt the Classification And Regression Trees (CART) supervised learning method to classify image blocks. The CART is a recursive partitioning method. It builds a directed decision tree (or a regression tree), where by convention the root node is displayed at the top, connected by successive (directional) links or branches to other nodes. Block features are extracted using two histograms: a) the gray level, and b) the first order derivative. The variances of these two histograms are calculated. Then, we get four variances in a block: 1) $\boldsymbol{\sigma_{f1}^2}$-variances of frequency counts in gray levels; 2) $\boldsymbol{\sigma_{f2}^2}$-variances of frequency counts in first order derivatives; 3) $\boldsymbol{\sigma_{p1}^2}$-variances of all pixel values; 4) $\boldsymbol{\sigma_{p2}^2}$-variances of all first order derivative values. Table I gives four variances to each block in Fig. 2. We see clearly that these four variance values can be used to distinguish different block types well.

Table I. Four variances for three blocks as shown in Fig. 2.

	Smooth	**Edge**	**Texture**
$\boldsymbol{\sigma_{f1}^2}$	586.1	104.1	1.6
$\boldsymbol{\sigma_{f2}^2}$	304.9	268.8	3.4
$\boldsymbol{\sigma_{p1}^2}$	7.5	4734.6	2375.9
$\boldsymbol{\sigma_{p2}^2}$	9.2	986.1	1176.2

Two reference images from the TID database, which contain smooth, edge and textures areas obviously, are used to build the training data set. We took 600x2=1200 blocks of size 16x16, tagged them with smooth, edge and texture blocks, respectively, and built the CART with proper threshold values associated with each decision node.

2.2 Image Distortion Classification

Image distortions in TID database were classified into five types as shown in Table II using the machine learning algorithm in [8]. Five features are calculated: 1) blockiness, 2) average absolute difference between in-block image samples, 3) zero-crossing (ZC) rate, 4) average edge-spread and 5) average block variance in the image. For details, we refer to [8]. We follow the framework in [8] and classify image distortion types into five groups as given in Table II.

Table II Distortion Types in TID and Grouping

	Type of distortion
Group I	1 Additive Gaussian noise
	2 Different additive noise in color components
	3 Spatially correlated noise
	4 Masked noise
	5 High frequency noise
	6 Impulse noise
Group II	7 Quantization noise
Group III	8 Gaussian blur
	9 Image denoising
	10 JPEG compression
	11 JPEG2000 compression
Group IV	12 JPEG transmission errors
	13 JPEG2000 transmission errors
Group V	14 Non eccentricity pattern noise
	15 Local blockwise distortions of different intensity
	16 Mean shift (intensity shift)
	17 Contrast change

2.3 BMMF

As demonstrated in [2], although a single QA metric cannot perform well on all distortion types; some of them can perform well with respect to a distortion type in a specific image content. For example, the PSNR metric performs well on 'Different additive noise in color components' and 'Impulse noise' although it is generally perceived as a non-ideal metric. After analyzing the performance of different quality metrics, we choose some of them to calculate the quality score of image blocks as shown in Table III, where the first column and the first row indicate the distortion type and the block type, respectively. Note also that, since our metric is block-based, multi-scale and wavelet based metrics (e.g., MS-SSIM, VIF, VSNR) are excluded from our choice.

Table III. Selected QA metrics for five distortion sets and three image block types

	Smooth	**Edge**	**Texture**
Group I	PSNR-HVS	PSNR-HVS-M	PSNR-HVS
Group II	FSIM	FSIM	PSNR-HVS
Group III	PSNR-HMA	PSNR-HA	FSIM
Group IV	PSNR-HVS	FSIM	FSIM
Group V	$*\text{MSE}_\text{H}^{(\text{S})}$	$*\text{MSE}_\text{H}^{(\text{E})}$	$*\text{MSE}_\text{H}^{(\text{T})}$

Once we assign the quality score to each individual block based on the QA metric given in Table III. The next step is to fuse all quality scores into a single one for the whole image. For image x with a distortion type i, we define its final BMMF score as

$$BMMF_i(x) = \sum_{j=1}^{3} \omega_{i,j} \cdot \hat{Q}_{i,j}(B_j^o, B_j^d), \tag{1}$$

where i is the index of distortion set, j is the image block type (say, 1-smooth, 2-edge, 3-texture), B_j^o and B_j^d are blocks in reference and distorted images, respectively, $Q_{i,j}$ is the selected quality metric for a distortion type i and block type j, $\hat{Q}_{i,j}$ is the mean of $Q_{i,j}$, and $\omega_{i,j}$ is a weighting factor determined by MOS using a small training dataset.

The proposed BMMF works well for distortion types I-IV. However, it does not work very well for distortion type V. To address this problem, we first examine Distortion Type V and then consider an enhanced BMMF solution as detailed in the next section.

3. CORRECTED BMMF QA METRIC

3.1 Analysis of Distortion Type V

Most QA metrics fail in images of distortion type V as reported in [2]. Thus, it is worthwhile to examine this distortion type in depth. Noise may be produced in the process of image coding or watermarking in Distortion #14 (non eccentricity pattern noise). Thus, the visibility of this type of noise is highly dependent on image content. That is, at the same distortion level, the distortion is less visible to HVS if the underlying image contains more textures. Distortions are caused by image acquisition, impainting or gamma correction in Distortion #15, #16 and #17. When images are captured by digital devices such as a camera, their quality can be influenced by a mean shift (or intensity shift) and/or contrast change due to different lighting conditions. This distortion is more obvious in homogenous regions.

According to physiological and psychological study on HVS, human being responds to image differences with an unequal attention level. That is, the HVS selects part of visual signals for detailed analysis and then responds, which is referred to as the Visual Attention (VA) model [9]. As a result, image quality is influenced by its content structure. It may not affect human perceptual image assessment much if the distortion is not in the VA region, but it will have a negative impact on some full reference metrics such as PSNR and MSE.

(a) Simple (b) Normal (c) Complex

Figure 3. Illustration of different image contents.

3.2 Image Complexity and Corrected BMMF Metric

Based on the discussion in Sec. 3.1, we classify image content into three different types using CART according to its structure complexity in this work; namely, simple, normal and complex as shown in Fig. 3. The simple image structure has simple background and a few foreground objects with clear object boundaries and little texture. The normal image structure contains a little bit more texture. The complex image has no dominant edges that can be used in segmentation but consists of multiple texture regions, such as creek, river banks, trees and mountains in the background as shown in Fig. 3(c). To differentiate image content types, we select two gradients G_H and G_V (horizontal and vertical) and a variance σ_e^2(frequency counts of image block types) as the discriminant features.

A full reference image QA metric that accounts for peculiarities of human perception of contrast and mean brightness distortions was proposed in [4]. By following the framework in [4], we propose a modified version of PSNR-HA, which takes into account the influence of non eccentricity pattern noise and intensity deviation distorted images below. First, we have

$$I_b = I^d + \delta, \tag{2}$$

$$I_c = \overline{I_b} + (I_b - \overline{I_b}) \cdot Popr, \tag{3}$$

where $I_b$and I_c denote the degraded brightness and contrast in distorted image I^d, respectively, δ is the mean difference between original image $I^o$and distorted image I^d, and $Popr$ is a factor used to compensate possible contrast change and defined as

$$Popr = \frac{\sum(I^o - \overline{I^o})\sum(I_b - \overline{I_b})}{\sum(I_b - \overline{I_b})^2}, \tag{4}$$

and where $\overline{I^o}$ and $\overline{I_b}$ are the mean values of I^o and I_b, respectively. Then, we define the MSE measure of distorted image I_b as

$$\widetilde{MSE}_b = K \sum_{m=1}^{M-7} \sum_{n=1}^{N-7} \sum_{i=1}^{8} \sum_{j=1}^{8} \Big((X[i,j]_{mn} - X[i,j]_{mn}^{d}) T[i,j] \Big)^2 \tag{5}$$

where M, N denote the image size, $\mathrm{K} = 1/[(\mathrm{M}-7)(\mathrm{N}-7)64]$ is a normalization factor, X_{mn} are DCT coefficients of 8x8 image block in I^o for which the coordinates of its left upper corner are equal to m and n, X_{mn}^d are the DCT coefficients of the corresponding block, and T is the matrix of correcting factors determined by the CSF and obtained by normalizing the quantization table in JPEG [10] and squaring the values. The distortion measure $\widetilde{MSE}_c$ can be derived similarly. After some manipulations, we obtain:

$$\widetilde{MSE}_H = \widetilde{MSE}_c - C\sigma_b^2 + \begin{cases} (\widetilde{MSE}_b - \widetilde{MSE}_c)K1, Popr < 1 \\ (\widetilde{MSE}_b - \widetilde{MSE}_c)K2, Popr \geq 1 \end{cases}, \tag{6}$$

$$MSE_H = \widetilde{MSE}_H + \delta^2 K3 + \begin{cases} \delta K4, & \delta < 0 \ \ and \ \widetilde{MSE}_H > |\delta| \\ 0, & otherwise \end{cases}, \tag{7}$$

$$BMMFc = \begin{cases} MSE_H^{(E)}, & I^o \ is \ Simple \\ MSE_H^{(S+E+T)}, & I^o \ is \ Normal \ or \ Complex \end{cases}, \tag{8}$$

where $\sigma_b^2$is the variance of I_b, $MSE_H^{(E)}$ is the cumulative value of MSE_H from all edge image blocks, $MSE_H^{(S+E+T)}$ is the cumulative value of all smooth, edge and texture blocks, C is a positive constant when $I^o$is simple and normal, and C is a negative constant when I^o is complex, and K1,K2,K3,K4 are correction factors and tuned as suggested in [4]. Finally, $BMMFc$ is the desired corrected QA metric calculated based on the complexity of I^o.

Table IV. Comparison of SROCC values of several image QA metrics on the TID database.

	Group I	Group II	Group III	Group IV	Group V	ALL
BMMF	**0.9517**[1]	**0.9254**[1]	**0.9575**[1]	**0.9139**[1]	**0.9025**[1]	**0.9471**[1]
FSIM	0.9045	0.8550	**0.9514**[2]	0.8845	0.7053	**0.8805**[2]
PSNR-HA	**0.9420**[2]	0.8958	0.9300	0.8252	0.8007	0.8680
IW-SSIM	0.8721	0.8164	0.9361	0.8592	0.7652	0.8559
MS-SSIM	0.8559	0.8527	0.9361	0.8759	0.7279	0.8527
PSNR-HMA	0.9353	**0.9056**[2]	0.9341	0.8224	**0.8218**[2]	0.8460
SSIM	0.8687	0.8661	0.9387	0.8811	0.5095	0.8087
VIF	0.9162	0.7946	0.9243	**0.8879**[2]	0.5310	0.7495
VSNR	0.8978	0.8260	0.9052	0.7711	0.5544	0.7046
PSNR-HVS	0.9402	0.8930	0.9287	0.8292	0.2755	0.5942
PSNR-HVS-M	0.9286	0.8985	0.9297	0.8214	0.2736	0.5594
PSNR	0.6583	0.8689	0.8823	0.7246	0.2483	0.5245

4. EXPERIMENTAL RESULTS

The proposed BMMF metric is validated on the TID [1] image database, which contains 25 reference images and 68 distorted versions of each reference image so that there are 1700 distorted images in total. There are 17 distortion types as shown in Table II. Due to the space limitation, we only use the Spearman rank-order correlation coefficient (SROCC) between the objective QA metric and the Mean Opinion Score (MOS) for performance comparison. Several state-of-the-art image quality metrics are compared with the proposed BMMF metric. They include: FSIM [7], PSNR-HA(-M) [4], IW-SSIM [6], MSSIM [3], SSIM [12], VIF [13], VSNR [14], PSNR-HVS(-M) [10], [11] and PSNR. The averaged results with respect to the five distortion types are shown in Table III.

The two best results are shown in bold and their ranks are marked in the right-up-corner. We see from the table that proposed BMMF QA metric performs the best in all five distortion types. In additional, most image quality metrics fail for images in Group V. The proposed corrected BMMF metric has an SROCC equal to 0.9025. Overall, with respect to the whole TID database, the BMMF metric offers the best SROCC performance with an average of 0.9471, which outperforms the second best metric by a significant margin.

5. CONCLUSION

A new full reference block-based multi-metric fusion (BMMF) approach was proposed for perceptual image quality assessment. The BMMF scheme automatically detects image content and distortion types in a block via machine learning. A corrected version of BMMF was also derived based on image complexity analysis. The superior performance of the BMMF scheme was demonstrated with the TID database.

6. REFERENCES

[1] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics", *in Advances of Modern Radioelectronics,* Vol. 10, pp. 30-45, 2009.

[2] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, V. Lukin "Metrics performance comparison for color image database", *Int. workshop on video processing and quality metrics for consumer electronics(VPQM)*, Scottsdale, Arizona, USA. Jan. 14-16, 2009.

[3] Z. Wang, E. P. Simoncelli, A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. on Signals, Systems and Computers*, pp. 1398-1402, San Francisco, USA, 9-12 Nov., 2003.

[4] N. Ponomarenko, O. Eremeev, V. Lukin, K. Egiazarian, M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting", *Proceedings of CADSM*, pp.305-311, Polyana-Svalyava, Ukraine, 23-25 Feb., 2011.

[5] L. Jin, N. Ponomarenko, K. Egiazarian, "Novel Image Quality Metric Based on Similiarty", *Int. Symp. on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, June 30-July 1, 2011.

[6] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment", *IEEE Trans. on Image Processing*, vol 20, no. 5, pp1185 – 1198, May 2011.

[7] L. Zhang, L. Zhang, X. Mou, D. Zhang, "FSIM: a feature similarity index for image quality assessment*", IEEE Trans. Image Processing,* vol 20, no. 5, pp 2378-2386, August 2011.

[8] T. Liu, W. Lin, C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment," *Intl. Workshop on Quality of Multimedia Experience (QoMEX2011),* pp. 72-77, Mechelen, Belgium, Sept. 7-9, 2011.

[9] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey", *Jounal of Visual Communication and Image Representation*, vol 22, no. 4, pp297-312, 2011

[10] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, "New full-reference quality metrics based on HVS", *Proc. of Int. Workshop on Video Processing and Quality Metrics(VPQM2006)*, Scottsdale, USA, Jan. 2006.

[11] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, "On between-coefficient contrast masking of DCT basis functions", *Proc. of Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM2007),* Scottsdale, Arizona, USA, 25-26 January, 2007.

[12] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. on Image Proc.*, vol. 13, issue 4, , pp. 600-612, April, 2004.

[13] H.R. Sheikh, A.C. Bovik, "Image information and visual quality", *IEEE Trans. on Image Pro*cessing, vol.15, no.2, pp. 430-444, 2006.

[14] D.M. Chandler, S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", *IEEE Trans. on Image Processing*, Vol. 16 (9), pp. 2284-2298, 2007.