Performance Comparison of Decision Fusion Strategies in BMMF-based Image Quality Assessment

Lina Jin*, Seongho Cho#, Tsung-Jung Liu#, 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— **The block-based multi-metric fusion (BMMF) is one of the state-of-the-art perceptual image quality assessment (IQA) schemes. With this scheme, image quality is analyzed in a block-by-block fashion according to the block content type (i.e. smooth, edge and texture blocks) and the distortion type. Then, a suitable IQA metric is adopted to evaluate the quality of each block. Various fusion strategies to combine the QA scores of all blocks are discussed in this work. Specifically, factors such as quality scores distribution and the spatial distribution of each block are examined using statistics methods. Finally, we compare the performance of various fusion strategies based on the popular TID database.**

I. INTRODUCTION

To provide high quality of user experience (QoE) for image processing applications such as image compression, image quality assessment (IQA) plays a key role. IQA aims to predict perceived image quality by human eyes. In general, there are mainly two ways: subjective IQA and objective IQA. Subjective IQA evaluates image quality by human. It is the most widely used and accurate method. However, subjective IQA has many limitations. For example, it is expensive and time-consuming, and it cannot be used in real-time image processing systems such as an image transmission system. Objective IQA is to evaluate image quality that is consistent with human experience automatically.

Research work on objective IQA has been active in the last decade. It was shown in many publications [1-6] that pixel-based methods (*i.e.* PSNR and MSE) have poor correlation with subjective testing. A number of perceptual metrics based on the human visual system (HVS) were proposed to address this shortcoming. Most of them [1-5] attempt to measure image quality using one model. However, image contents and distortions vary from one to another. It is difficult for one model to cover all distorted image situations as demonstrated in [6]. For example, the IQA for image distortion caused by compression (*e.g.* JPEG or JPEG2000) may rely on models different from those for image distortions caused by image mean shift or contrast changes. Following this principle, a block-based multi-metric fusion (BMMF) method was proposed in [7]. In BMMF, image quality is analyzed in a block-by-block fashion based on the block types (i.e. smooth, edge and texture) and the distortion types. Then, a suitable IQA metric is adopted to evaluate the quality of each block. Finally, all block-based quality metrics are fused to result in one final score. It was shown in [7] that BMMF significantly improves image quality performance as compared with other single IQA metrics.

How to fuse quality scores of all blocks to result in one final score is a key problem in BMMF. In [7], the simplest statistical method; namely, the arithmetic mean, was adopted. In this work, different fusion strategies are discussed and compared. Furthermore, to study the influence of quality score in different regions of distorted image contents, the visual attention (VA) model is incorporated in the BMMF framework. Finally, a new multi-dimensional mean fusion is presented. More details will be given in Section 3.

The rest of this paper is organized as follows. The BMMF scheme is reviewed in Section 2. Various fusion strategies are examined in Section 3. Experimental results are presented in Section 4. Finally, concluding remarks are given in Section 5.

II. REVIEW OF BMMF

The block-based multi-metric fusion (BMMF) metric was developed to exploit the observation that image QA metrics are influenced by the content and distortion types of a local image region, since the HVS selects parts of visual contents for analysis and responds [7]. BMMF 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.

Decomposed processing units are first classified into smooth, edge and texture three types using the CART (Classification And Regression Trees) supervised learning method. Block features are extracted using variances of two histograms: a) the gray level and b) the first order derivative. Furthermore, image distortion types are automatically detected and classified into five types by following the method described in [7][8]. We can select the most suitable IQA metric with respect to a block type and a distortion type. This is known as the local perceptual quality metrics.

To derive the global perceptual quality metric for the whole image, all block-based quality metrics are combined to yield one score, called the BMMF metric, as

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

where i is the distortion type index, j is the image block type index (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 distortion type i and block type j, and $\omega_{i,j}$ is a weighting factor determined by MOS using a small training dataset. Here, $\widehat{Q}_{i,j}$ is fusion method of $q_{i,j}$. In [7],$\widehat{Q}_{i,j}$ is calculated as the arithmetic mean. In this work, different fusion strategies $\widehat{Q}_{i,j}$ are proposed in Section III, and its corresponding performances are compared in Section IV. Since most IQA metrics fail in images of distortion type V as reported in [6], corrected $BMMF_c$ specified for distortion group 5 was proposed in [7].

III. Proposed Fusion Strategies for BMMF

Any process to an original image may change image information and results in image noise or distortion. Examples include image compression, watermarking, transmission. However, not every change is noticeable to HVS. Assume one image is divided into different parts, human attention is not allocated equally in the field of view [9].

Since BMMF calculates image quality in blocks, each block quality value can be obtained. Fig. 1 shows one example of comparing one original image in Fig. (a) and four distorted images in Fig. 1(b)-(e). Fig. 1(b) shows the white Gaussian noise from distortion group 1 as defined in [7][8]. Fig.1(c) shows the JPEG compression from distortion group 3, Fig.1(d) shows the JPEG transmission from distortion group 4, and Fig.1(e) shows the local blockwise from distortion group 5. Block classification and the corresponding block quality values for distorted images are shown in the second row of Fig. 1. Block quality values are mapped to the range of [0-8], where higher scores indicate better quality. We see from Fig. 1 that block quality values vary in different image regions.

A. Mean

First, statistics methods of calculating various mean values are proposed in this section. It includes the arithmetic mean, the median, the truncated mean, the weighted mean, and the geometric mean.

- Arithmetic Mean (AM)

The *arithmetic mean* (AM) is defined as the arithmetic average of a set of values, or distribution, often simply called the "mean". The AM is the most widely used and simplest method to calculate one set of data,

$$\hat{Q} = \frac{1}{n}\cdot\sum_{i=1}^{n} q_i, \tag{2}$$

where i indicates the i-th block, n is the total number of one block type, *e.g.* smooth, edge or texture, q_i is quality value calculated by corresponding IQAs as given in [7]. The AM is used to fuse all block quality scores in [7].

- Median Value (MD)

Instead of obtaining the mean to describe overall quality, a *median* value (MD) can be used in the analysis. In statistics and probability theory, the *median* is described as the numerical value separating the higher half from the lower half. If there is an even number of observations, there is no single middle value and the median is then chosen to be the mean of the two middle values. For a block quality score set, the MD can be calculated as

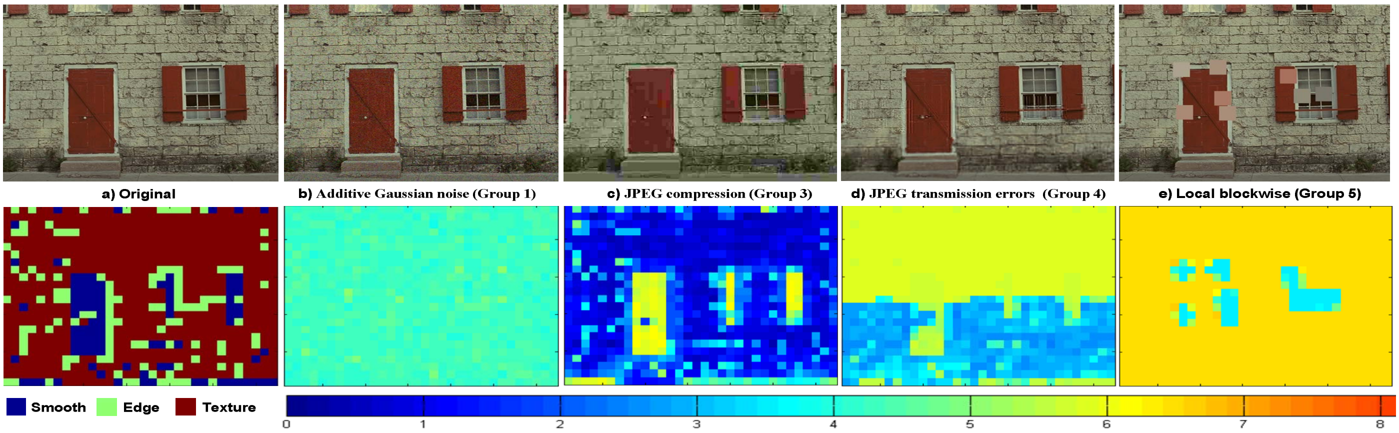
$$\hat{Q} = median\{\widetilde{q_1}, \widetilde{q_2}, \widetilde{q_3}, \dots, \widetilde{q_n}, \}, \tag{3}$$

where $\{\widetilde{q_1}, \widetilde{q_2}, \widetilde{q_3}, \dots, \widetilde{q_n}, \}$ are sorted in an ascending order.

- Truncated Mean (TM)

Sometimes, one set of block quality scores might contain outliers, i.e., a datum which is much lower or much higher than the others. Often, outliers are erroneous data caused by artifacts, and they may have a significant impact on the final arithmetic mean. In this case, one can use a *truncated mean* (TM) or a *trimmed mean* which is a statistical measure of central tendency. The TM involves discarding given parts of the data at the top and/or the bottom ends, typically an equal amount at each end, and then taking the arithmetic mean of the remaining data. The number of values removed is indicated as a percentage of the total number of values. For a vector input of block quality scores, i.e. $q_i, i = 1,2,..n,$ $\hat{Q}$ is calculated as the mean of q, excluding the highest and lowest k data values, where $k = (n * percent)/2$. The TM is a useful estimator since it is less sensitive to outliers than the AM but will still give a reasonable estimate of central tendency.

The TM uses more information from the distribution or the samples than that of the median. However, unless the underlying distribution is a symmetric one, the TM of a sample set is unlikely to produce an unbiased estimator for either the mean or the median. In addition, the AM can be considered as one special set of the TM, in which the number



a) Original

b) Additive Gaussian noise (Group 1)

c) JPEG compression (Group 3)

d) JPEG transmission errors (Group 4)

e) Local blockwise (Group 5)

Fig. 1 Example of one original image, four distorted images (top) and the corresponding BMMF block quality scores (bottom).

of excluding data is *k*=0.

- Weighted Mean (WM)

As mentioned in the introduction section, pixel-based quality metrics do not always success to predict human perception. It has been demonstrated that not every change in image pixels is always noticeable according to many psychological and physiological experiments [9]. Here, it is assumed that some data points of block quality q_i contribute more than others.

The weighted arithmetic mean (WM) (or weighted average) is calculated by giving different weights to different individual values. Since there are lots of blocks in each block type, there will be too many weighting factors if $\hat{Q}$ is calculated based on each q_i. Therefore, block quality scores q_i are classified into M bins. The WM of $\hat{Q}$ is calculated as

$$\hat{Q} = \frac{\sum_{i=1}^{M} \varphi_i \cdot \widetilde{Q_i}}{\sum_{i=1}^{M} \varphi_i}, \tag{4}$$

where $\widetilde{Q_i}$ is the arithmetic mean of quality scores in the i-th bin, M is set to 10 in this work, and φ_i is the weighting factor for the i-th bin and defined as the total number of i-th bin. The weight φ_i represents a measure of the reliability of the influence upon the mean by respective values. Clearly, AM is also a special case of the weighted mean where all data have equal weights, i.e., $\varphi_i = \varphi$.

- Geometric mean (GM)

To study the central tendency or typical value of a set of quality scores, the *geometric mean*(GM), can also be considered. A geometric mean is often used when comparing different items; namely, finding a single "figure of merit" for these items when each item has multiple properties that have different numeric ranges. The geometric mean $\hat{Q}$ is calculated as

$$\hat{Q} = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots q_n}. \tag{5}$$

B. *Visual Attention Model*

Recent research work on psychological and physiological phenomena has demonstrated that not every difference of an image receives the same attention level [10]. To extract image regions of the visual signal for detailed analysis and to study its influences on the final quality level, the visual attention (VA) model is incorporated in the fusion process of BMMF in this section.

There are two categories of VA approaches: the bottom-up and the top-down approaches. The bottom-up approach detects areas of image or video those are salient by low-level features of visual signals. The top-down approach is task-oriented, which is driven by a certain task, and a model is built based on visual features correlated with such a task [9]. Two bottom-up approaches are applied in association with BMMF in this work. One is the ITTI model [11] while the other is the GBVS model [12]. Fig. 2 shows an example of saliency maps from the ITTI model over four distorted images as shown in Fig.1. Saliency maps are in the ranges of [0~1], where the corresponding image block with a higher value indicates more attention than others. These scores are marked gradual color from the light blue to the dark red in the figure.

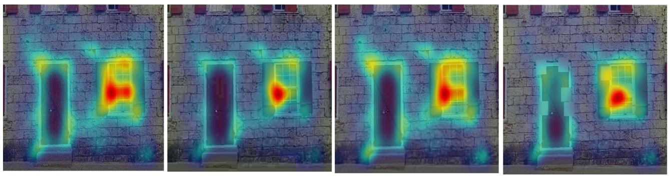


Fig. 2 The ITTI saliency maps over four distorted images of Fig.1b)-e)

We divide image regions into three parts: the non-VA region, the normal region and the VA region by setting two thresholds in saliency maps. The following three methods are employed.

- VA Method 1:

Distorted images are divided into the above three parts, but the block types in each region are not considered, *i.e.* j is the image region index in Eq.(1). IQAs in each distortion group and image region are as the same as in [7]. $\widehat{\mathrm{Q}}$ is calculated as AM.

- VA Method 2:

Distorted images are divided into three parts and block types in each region are considered. Hence, to each image region, there will be $m_b$block types, where. m_b takes the value of 1, 2 or 3. Eq.(1) is applied to each image region.

- VA Method 3:

The final fused quality value of each block type is decided by the ones in the image region, which contains the maximum number of correspond block type.

C. *Multi-dimensional Mean*

Finally, the *Multi-dimensional Mean* (MM) is proposed. In the MM, more than one statistical methods are adopted in each block type and applied to Eq.(1). Two dimensions are studied here. The first dimensional mean is the AM. Two different fusion methods are proposed to include the second dimension metric:

$$\hat{Q} = AM(q_i) + MM(q_i) \tag{6}$$

- MM Method 1: AM and $p\%$ worst quality values

For a distorted image, human tends to pay more attention to image blocks that have lower quality than its neighbor blocks as shown in Fig.1 (e).Therefore, $p\%$ worst quality values are obtained from the sorted block quality set and they are fused with the AM of these data.

- MM Method 2: AM and variance/standard deviation

To measure of how far a set of block quality values is spreat out, the ratio of the variance (var) and the standard deviation (std) is used as the second dimension.

IV. Experimental Results

The TID [13] image database is used to test the proposed fusion strategies. TID contains 25 reference images and 1700 distorted images with 17 different distortion types, which are classified into five groups [7][8]. The Spearman rank-order correlation coefficient (SROCC) between the objective IQA metric and the Mean Opinion Score (MOS) is measured for performance comparison.

TABLE I: COMPARISON OF SROCC VALUES OF PROPOSED FUSION STRATEGIES FOR BMMF AND IQAS ON THE TID DATABASE.

		Group 1	Group 2	Group 3	Group 4	Group 5	ALL
Mean	1 AM	0.9517	0.9254	0.9575	0.9139	0.9025	0.9471
	2. MD	0.9362	0.9102	0.9523	0.8991	0.3629	0.7829
	3. TM	**0.9519**	**0.9259**	**0.9594**	**0.9240**	**0.9027**	**0.9488**
	4. WM	0.9469	0.9257	0.9567	0.8954	0.8694	0.9374
	5. GM	0.8632	0.9052	0.9449	0.8457	0.8285	0.9018
Visual Attention	6. VA1	0.9413	0.9174	0.9426	0.8985	0.8944	0.9383
	7. VA2	**0.9528**	**0.9295**	**0.9582**	**0.9180**	**0.8975**	**0.9454**
	8. VA3	0.9502	0.9227	0.9569	0.9101	0.8364	0.9289
Multidimensional Mean	9. MM1	0.9564	**0.9274**	**0.9621**	**0.9292**	0.9030	**0.9506**
	10. MM2(var)	**0.9577**	0.9257	0.9610	0.9174	0.9046	0.9480
	11 .MM2(std)	0.9559	0.9273	0.9610	0.9156	**0.9053**	0.9475
Other IQMs	FSIM	0.9045	0.8550	0.9514	0.8845	0.7053	0.8805
	PSNR-HA	0.9420	0.8958	0.9300	0.8252	0.8007	0.8680
	IW-SSIM	0.8721	0.8164	0.9361	0.8592	0.7652	0.8559
	SSIM	0.8687	0.8661	0.9387	0.8811	0.5095	0.8087
	PSNR-HVS	0.9402	0.8930	0.9287	0.8292	0.2755	0.5942
	PSNR	0.6583	0.8689	0.8823	0.7246	0.2483	0.5245

Experimental results of above mentioned fusion strategies are listed in Table I. The first five rows show different fusion methods for calculating 'Mean'. The two thresholds used to divide the VA region are chosen to be 0.1 and 0.3, respectively. Two VA models, i.e. the ITTI and the GBVS models, are tested. Since their results are similar, just the ones of the ITTI model are listed in Table I. For MM method 1, $p\%$ is set to 30%. Several state-of-the-art image quality metrics are also compared. They include: SSIM[1], PSNR-HVS[2], PSNR-HA[3], FSIM[4], IW-SSIM[5], and PSNR.

The best results in the fusion strategies under the categories of 'Mean', 'Visual Attention' and 'Multidimensional Mean' are marked in bold. We see from the table that TM performs better than AM in BMMF, where SROCC is equal to 0.9240 for distortion Group 4. Note also that the Visual Attention model does not improve the performance BMMF much. The MM method with $p\% = 30\%$ performs the best as compared with other fusion methods. The SROCC values in distortion Group 3 and 4 have increased to 0.9621 and 0.9230, respectively. Overall, with respect to the whole TID database, the BMMF with various fusion strategies offer better SROCC performance when being compared with other state-of-the-art single IQMs, especially, in distortion groups 2, 4 and 5.

V. CONCLUSION AND FUTURE WORK

BMMF was carefully studied in this work by considering different fusion strategies of quality scores in various blocks. Both the block quality distribution and the block spatial distribution were considered and analyzed using various statistical methods. These fusion strategies were tested and compared with each other with the TID database. We did observe better results using the MM fusion strategy.

The current results are nevertheless preliminary. We would like to find theoretical justification of the improvement due to the MM fusion strategy. We will also continue to seek better fusion strategies for BMMF in the near future.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] L. Jin, K. Egiazarian, C.-C. J. Kuo, "Perceptual image quality assessment using block-based multi-metric fusion (BMMF)," *37th IEEE Intl. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP),* Kyoto, Japan, March 25-30, 2012.

[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] L. Itti, and C. Koch, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience,* vol. 2, no. 3, pp. 194-203, Mar. 2001.

[11] L. Itti, C. Koch, and E. Niebur , "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254-1259 (1998) http://ilab.usc.edu/bu/

[12] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency", *Proceedings of Neural Information Processing Systems (NIPS)*,Vancouver, Canada, 4-9 Dec., 2006.

[13] 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.