Image Quality Assessment Using Multi-Method Fusion

Tsung-Jung Liu, *Student Member, IEEE*, Weisi Lin, *Senior Member, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*— A new methodology for objective image quality assessment (IQA) with multi-method fusion (MMF) is presented in this paper. The research is motivated by the observation that there is no single method that can give the best performance in all situations. To achieve MMF, we adopt a regression approach. The new MMF score is set to be the nonlinear combination of scores from multiple methods with suitable weights obtained by a training process. In order to improve the regression results further, we divide distorted images into three to five groups based on the distortion types and perform regression within each group, which is called "context-dependent MMF" (CD-MMF). One task in CD-MMF is to determine the context automatically, which is achieved by a machine learning approach. To further reduce the complexity of MMF, we perform algorithms to select a small subset from the candidate method set. The result is very good even if only three quality assessment methods are included in the fusion process. The proposed MMF method using support vector regression is shown to outperform a large number of existing IQA methods by a significant margin when being tested in six representative databases.**

***Index Terms*— Context-dependent MMF (CD-MMF), context-free MMF (CF-MMF), image quality assessment (IQA), machine learning, multi-method fusion (MMF).**

I. Introduction

VARIOUS image quality assessment methods have been developed to reflect human visual quality experience. The mean-square error (MSE) and the peak-signal-to-noise-ratio (PSNR) are two widely used ones. However, they may not correlate with human perception well [1], [2]. During the last decade, a number of new quality indices have been proposed as better alternatives. Examples are summarized in Table I, including MS-SSIM [3],SSIM [4], VIF [5], VSNR [6], NQM [7], PSNR-HVS [8], IFC [9], PSNR, FSIM [10], and MAD [11]. So far, there is not a single quality index that significantly outperforms others. Some method may be superior for one image distortion type but inferior for others. Thus, the idea of multi-method fusion (MMF) arises naturally.

The major contribution of this research is to provide a new perspective in visual quality assessment to complement existing efforts that target at developing a single method suitable for certain types of images. Given the complex and diversifying nature of general visual content and distortion types, it would be challenging to solely rely on a single method. On the other hand, we demonstrate that it is possible to achieve significantly better performance by fusing multiple methods with proper means (e.g., machine learning) at the cost of higher complexity. The performance of the proposed MMF scheme will be improved continuously when new methods invented by the research community are incorporated.

TABLE I
TEN BETTER-RECOGNIZED IQA METHODS

IQA Method Index	Abbreviation	Full Name
m1	MS-SSIM	Multi-Scale Structural Similarity
m2	SSIM	Structural Similarity
m3	VIF	Visual Information Fidelity
m4	VSNR	Visual Signal-to-Noise Ratio
m5	NQM	Noise Quality Measure
m6	PSNR-HVS	Peak Signal-to-Noise Ratio Human Visual System
m7	IFC	Information Fidelity Criterion
m8	PSNR	Peak Signal-to-Noise Ratio
m9	FSIM	Feature Similarity
m10	MAD	Most Apparent Distortion

To achieve MMF, we adopt a regression approach (e.g., support vector regression (SVR)). First, we collect a number of image quality evaluation methods. Then, we set the new MMF score to be the nonlinear combination of scores from multiple methods with suitable weighting coefficients. Clearly, these weights can be obtained by the regression approach.

Although it is possible to perform MMF independently of image distortion types, the assessment result can be improved if we take image distortion types into account. To be more specific, we may divide image distortion types into several major groups (i.e., contexts) and perform regression within each group. In other words, the term context in this work means an image group consisting of similar distortion types We call the one independent of distortion type as "context-free MMF" (CF-MMF) and the one depending on distortion types as "context-dependent MMF" (CD-MMF), respectively. For CD-MMF, one important task is to determine the context automatically. Here, we use a machine learning approach for context determination. As a result, the proposed CD-MMF system consists of two steps: 1) context determination and 2) MMF for a given context.

This paper is an extension of our previous work in [12] with a substantial amount of new material. First, we provide a detailed discussion on support vector regression (SVR)

Manuscript received March 29, 2012; revised November 25, 2012; accepted December 2, 2012. Date of publication December 24, 2012; date of current version March 11, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhou Wang.

T.-J. Liu and C.-C. Jay Kuo are with the Ming Hsieh Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail: liut@usc.edu; cckuo@sipi.usc.edu).

W. Lin is with the School of Computer Engineering, Nanyang Technological University, 639798 Singapore (e-mail: wslin@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2236343

1057–7149/$31.00 © 2012 IEEE

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

theory, which is the machine learning tool used for fusion of multiple methods in Section III.B. Second, we offer an elaborated study on the fusion rules in Section V. All material in Section V is new. Specifically, we develop a new fused IQA methods selection algorithm called the Biggest Index Ranking Difference (BIRD) that is used to select the most appropriate method for fusion so as to reduce the complexity of the MMF method proposed in [12]. Furthermore, we compare BIRD with another fused IQA methods selection algorithm called the Sequential Forward Method Selection (SFMS) in terms of performance accuracy and complexity. Third, we conduct a more thorough experimental evaluation. Only the TID database was tested in [12]. Here, we test the performance of the MMF method against six publicly available image quality databases in Section VI. Finally, we replace two poor-performed methods in [12] with two more recently developed approaches (i.e., FSIM [10] and MAD [11]) in the MMF process so as to achieve a better correlation between predicted objective quality scores and human subjective scores. This also demonstrates that the proposed MMF approach is able to accommodate new methods to produce better results. The Pearson Linear Correlation Coefficient (PLCC) performance of the proposed MMF method ranges from 94% to 98% with respect to various image quality databases.

The rest of this paper is organized as follows. Some prior related works are reviewed in Section II. The MMF process based on regression is described in Section III. The MMF types, context definition and determination are investigated in Section IV. In Section V, we try to reduce the complexity of the proposed MMF approach via two fused IQA methods selection algorithms. Experimental results are reported in Section VI, where extensive performance comparisons are made across multiple image quality databases. Finally, concluding remarks and future works are given in Section VII.

II. Review of Previous Work

Machine learning and multiple-metric based image quality assessment methods were reported in the literature before. Luo [13] proposed a two-step algorithm to assess image quality. First, a face detection algorithm is used to detect human faces from the image. Second, the spectrum distribution of the detected region is compared with a trained model to determine its quality score. The restriction is that it primarily applies to images that contain human faces. Although the authors claimed it's not difficult to generalize faces to other objects, they still only provided the results which used the images containing human faces to prove the feasibility of their algorithm.

Suresh *et al.* [14], [15] proposed the use of a machine learning method to measure the visual quality of JPEG-coded images. Features are extracted by considering factors related with the human visual sensitivity, such as edge length, edge amplitude, background luminance, and background activity. The visual quality of an image is then computed using the predicted class number and their estimated posterior probability. It was shown by experimental results that the approach performs better than other metrics. However, it is only applicable to JPEG-coded images since the above features are calculated based on the DCT blocks.

The machine learning tool has been used in developing an objective image quality metric. For example, Narwaria and Lin [16] proposed to use singular vectors out of singular value decomposition (SVD) as features to quantify the major structural information in images. Then, they applied support vector regression (SVR) for image quality prediction, where the SVR method has the ability to learn complex data patterns and maps complicated features into a proper score.

Moreover, Leontaris *et al.* [17] collected 15 metrics, and evaluated each one of them to see if they satisfy the expectation of a good video quality metric. In the end, they linearly combined two metrics (MCEAM and GBIM) to get a hybrid metric by using simple coefficients as the weights.

In summary, the works in [13]–[16] used extracted features for model training and test. They are related to machine learning. Another work [17] showed the advantage of integrating results from two methods (although it did not use the machine learning approach).

As compared with the previous work, the proposed multi-method fusion (MMF) idea is new since it offers a generic framework that enables better performance than existing methods and serves as a reference for future research. The MMF is developed in a systematic manner to complement the existing (and even future) approaches.

III. Multi-method Fusion (MMF)

A. Motivation

Many objective quality indices have been developed during the last decade. We consider the ten existing better-recognized methods given in Table I. Furthermore, we consider 17 image distortion types as given in the TID2008 database [18]. They are listed in Table II.

In Table III, we list the top three quality indices for each distortion type in terms of Pearson linear correlation coefficient (PLCC). We observe that different quality indices work well with respect to different image distortion types. For example, the PSNR may not accurately predict quality scores for most distortion types, but it does work well for images corrupted by additive noise as well as quantization noise. Generally speaking, the PSNR and its variant PSNR-HVS work well for image distortion types #1-7 while FSIM works well for image distortion types #8-17.

B. Support Vector Regression (SVR)

In order to develop the MMF method that handles all image distortion types and extents, we would like to integrate the scores obtained from multiple quality indices into one score. Although there exist many different fusion tools, we adopt a support vector regression approach here due to its relatively superior performance.

Suppose we have a set of training data $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$, where $\boldsymbol{x}_i \in R^n$ is a feature vector, and $y_i \in R$ is the target output. In ε support vector regression (ε-SVR) [19], we want to find a linear function,

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b = \boldsymbol{w}^T \boldsymbol{x} + b \tag{1}$$

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE II
IMAGE DISTORTION TYPES IN TID2008 DATABASE

Type	Type of Distortion
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
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change

TABLE III
TOP THREE QUALITY INDICES FOR IMAGE DISTORTION TYPES IN TABLE II (IN TERMS OF THE PLCC PERFORMANCE

Type \ Method	Best method (PLCC)	2nd best method (PLCC)	3rd best method (PLCC)
1	m6 (0.9366)	m8 (0.9333)	m3 (0.8717)
2	m8 (0.9285)	m6 (0.9137)	m3 (0.9004)
3	m8 (0.9524)	m6 (0.9510)	m10 (0.8745)
4	m3 (0.8928)	m8 (0.8737)	m6 (0.8240)
5	m6 (0.9730)	m8 (0.9708)	m3 (0.9464)
6	m8 (0.9084)	m6 (0.8651)	m3 (0.8263)
7	m6 (0.8965)	m8 (0.8911)	m2 (0.8745)
8	m1 (0.9506)	m2 (0.9452)	m9 (0.9414)
9	m9 (0.9680)	m2 (0.9664)	m1 (0.9638)
10	m6 (0.9720)	m9 (0.9710)	m2 (0.9608)
11	m9 (0.9801)	m10 (0.9789)	m1 (0.9751)
12	m1 (0.8844)	m9 (0.8823)	m10 (0.8784)
13	m6 (0.9256)	m2 (0.8574)	m9 (0.8491)
14	m7 (0.8394)	m10 (0.8315)	m3 (0.7460)
15	m2 (0.8768)	m9 (0.8531)	m3 (0.8434)
16	m2 (0.7547)	m6 (0.7099)	m8 (0.7076)
17	m3 (0.9047)	m9 (0.7706)	m1 (0.7689)

which has at most deviation ε from the actually obtained target outputs y_i for all the training data and at the same time as flat as possible, where $\boldsymbol{w} \in R^n$, $b \in R$. In other words, we want to find $\boldsymbol{w}$ and b such that

$$||f(\boldsymbol{x}_i) - y_i||_1 \leq \varepsilon \forall i = 1, \ldots, m. \tag{2}$$

where $||\cdot||_1$ is the l_1 norm.

Flatness in (1) means we have to seek a smaller $\boldsymbol{w}$ [20]. For this reason, it is required to minimize $||\boldsymbol{w}||_2^2$, where $||\cdot||_2$ is the Euclidean (l_2) norm. Generally, this can be written as a convex optimization problem by requiring

$$\begin{aligned} &\min \frac{1}{2}||\boldsymbol{w}||_2^2 \\ &\text{subject to} \quad ||f(\boldsymbol{x}_i) - y_i||_1 \leq \varepsilon, i = 1, \ldots, m. \end{aligned} \tag{3}$$

Introducing two slack variables $\varepsilon_i \geq 0$, $\hat{\varepsilon}_i \geq 0$, to cope with infeasible constraints of (3), (3) becomes

$$\begin{aligned} &\min \frac{1}{2}||\boldsymbol{w}||_2^2 + C\Sigma_{i=1}^m(\varepsilon_i + \hat{\varepsilon}_i) \\ &\text{subject to} \quad y_i \leq f(\boldsymbol{x}_i) + \varepsilon + \varepsilon_i \\ &\qquad y_i \geq f(\boldsymbol{x}_i) - \varepsilon - \hat{\varepsilon}_i \\ &\qquad \varepsilon_i \geq 0, \hat{\varepsilon}_i \geq 0, i = 1, \ldots, m. \end{aligned} \tag{4}$$

where C is a penalty parameter for the error term. The optimization problem (4) can be solved through its Lagrangian dual problem

$$\begin{aligned} \max \; &-\frac{1}{2}\Sigma_{i=1}^m\Sigma_{j=1}^m(a_i - \hat{a}_i)(a_j - \hat{a}_j)\boldsymbol{x}_i^T\boldsymbol{x}_j \\ &- \varepsilon\Sigma_{i=1}^m(a_i + \hat{a}_i) + \Sigma_{i=1}^m(a_i - \hat{a}_i)y_i \\ \text{subject to} \quad &\Sigma_{i=1}^m(a_i - \hat{a}_i) = 0 \\ &0 \leq a_i, \hat{a}_i \leq C, i = 1, \ldots, m. \end{aligned} \tag{5}$$

where $a_i \geq 0$ and $\hat{a}_i \geq 0$, being Lagrange multipliers. After solving (5), we can obtain

$$\boldsymbol{w} = \Sigma_{i=1}^m(a_i - \hat{a}_i)\boldsymbol{x}_i \tag{6}$$

$$f(x) = \Sigma_{i=1}^m(\hat{a}_i - a_i)\boldsymbol{x}_i^T\boldsymbol{x} + b \tag{7}$$

By using Karush-Kuhn-Tucker (KKT) conditions as below

$$\begin{aligned} a_i(\varepsilon + \varepsilon_i + \boldsymbol{w}^T\boldsymbol{x}_i + b - y_i) &= 0 \\ \hat{a}_i(\varepsilon + \hat{\varepsilon}_i - \boldsymbol{w}^T\boldsymbol{x}_i - b + y_i) &= 0 \\ (C - a_i)\varepsilon_i &= 0 \\ (C - \hat{a}_i)\hat{\varepsilon}_i &= 0 \end{aligned} \tag{8}$$

where b can be computed as follows:

$$b = \begin{cases} y_i - \varepsilon - \boldsymbol{w}^T\boldsymbol{x}_i, 0 < a_i < C \\ y_i + \varepsilon - \boldsymbol{w}^T\boldsymbol{x}_i, 0 < \hat{a}_i < C \end{cases} \tag{9}$$

The support vectors are defined as those data points that contribute to predictions given by (7), and are $\boldsymbol{x}_i$'s where $a_i - \hat{a}_i \neq 0$. The complexity of $f(\boldsymbol{x})$ is related to the number of support vectors.

Similarly, for nonlinear regression, we just define $f(\boldsymbol{x}) = \boldsymbol{w}^T\varphi(\boldsymbol{x}) + b$, and $\varphi(\boldsymbol{x})$ denotes a fixed feature-space transformation. Then

$$w = \Sigma_{i=1}^m(a_i - \hat{a}_i)\varphi(\mathbf{x}_i) \tag{10}$$

and

$$\begin{aligned} f(\boldsymbol{x}) &= \Sigma_{i=1}^m(\hat{a}_i - a_i)\varphi^T(\boldsymbol{x}_i)\varphi(\boldsymbol{x}) + b \\ &= \Sigma_{i=1}^m(\hat{a}_i - a_i)K(\boldsymbol{x}_i, \boldsymbol{x}) + b \end{aligned} \tag{11}$$

where $K(\boldsymbol{x}_i, \boldsymbol{x})$ is a kernal function. The kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ can be defined as

$$K(\boldsymbol{x}_i\boldsymbol{x}_j) = \varphi^T(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j) \tag{12}$$

There are four basic kernels [21]. We list two commonly used ones:

1) Linear:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T\boldsymbol{x}_j \tag{13}$$

2) Radial basis function (RBF):

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2), \gamma > 0 \tag{14}$$

where γ is a kernel parameter.

Since the proper value for the parameter ε is difficult to determine, we try to resolve this problem by using a different version of the regression algorithm, ν support vector regression (ν-SVR) [19], in which ε itself is a variable in the optimization process and is controlled by another new parameter $\nu \in (0, 1)$. In fact, ν is a parameter that can be used to control the number

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

of support vectors and the upper bound on the fraction of error points. Hence, this makes ν a more convenient parameter than ε in adjusting the accuracy level to the data. Therefore, ν-SVR is to solve

$$\begin{aligned} &\min \frac{1}{2}||\boldsymbol{w}||_2^2 + C\left(\upsilon\varepsilon + \frac{1}{m}\Sigma_{i=1}^{m}(\varepsilon_i + \hat{\varepsilon}_i)\right) \\ &\text{subject to } \; y_i \le f(\boldsymbol{x}_i) + \varepsilon + \varepsilon_i \\ &\qquad\qquad\; y_i \ge f(\boldsymbol{x}_i) - \varepsilon - \hat{\varepsilon}_i \\ &\qquad\qquad\; \varepsilon_i \ge 0, \hat{\varepsilon}_i \ge 0, i = 1, \ldots, m, \varepsilon \ge 0. \end{aligned} \tag{15}$$

The dual problem is

$$\begin{aligned} &\max -\frac{1}{2}\Sigma_{i=1}^{m}\Sigma_{j=1}^{m}(a_i - \hat{a}_i)(a_j - \hat{a}_j)K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &\quad + \Sigma_{i=1}^{m}(a_i - \hat{a}_i)y_i \\ &\text{subject to } \; \Sigma_{i=1}^{m}(a_i - \hat{a}_i) = 0, \\ &\qquad\qquad\; \Sigma_{i=1}^{m}(a_i + \hat{a}_i) \le C\upsilon, \\ &\qquad\qquad\; 0 \le a_i, \hat{a}_i \le C/m, i = 1, \ldots, m. \end{aligned} \tag{16}$$

Following the same procedure, we can obtain the same expressions for $\boldsymbol{w}$ and $f(\boldsymbol{x})$ as in (10) and (11). In this work, we choose ν-SVR as our tool to do all the regressions because of its convenience on parameter selection.

C. MMF Scores

Consider the fusion of n image quality assessment methods with m training images. For the i-th training image, we can compute its quality score individually, which is denoted by x_{ij}, where $i = 1, 2, \ldots, m$, denoting the image index and $j = 1, 2, \ldots, n$, denoting the method index. Also, we define the quality score vector $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,n})^T$ for the i-th image. The new MMF quality score is defined as

$$mmf(\boldsymbol{x}_i) = \boldsymbol{w}^T\varphi(\boldsymbol{x}_i) + b \tag{17}$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ is the weight vector, and b is the bias.

D. Data Scaling and Cross-Validation

Before applying SVR, we need to linearly scale the scores obtained from each quality index to the same range [0, 1] to avoid the quality index in larger numerical ranges (e.g., PSNR) dominating those in smaller numerical ranges (e.g., SSIM). Another advantage is to avoid numerical difficulties during the calculation. The linear scaling operation is performed for both training and testing data [21].

In all experiments, we use the n-fold (e.g., n equals to 5) cross-validation strategy (which is widely used in the machine learning [22]) to select our training and testing sets. First, we divide the image set into n sets. One set is used for testing, and the remaining n-1 sets are used for training. Then, we do this n times, where we make sure each set is only used as the testing set once. The testing results from n folds are then combined and averaged to compute the overall correlation coefficients and error. This procedure can prevent the over-fitting problem.

E. Training Stage

In the training stage, we would like to determine the weight vector $\boldsymbol{w}$ and the bias b from the training data that minimize the difference between $mmf(\boldsymbol{x}_i)$ and the (differential) mean opinion score ($(D)MOS_i$) obtained by human observers; namely,

$$\left|\left|mmf(\boldsymbol{x}_i) - (D)MOS_i\right|\right|, i = 1, \ldots, m. \tag{18}$$

where $||\cdot||$ denotes a certain norm. Several commonly used difference measures include the Euclidean (l_2) norm, the l_1 norm and the l_∞ norm. The Euclidean norm will lead to the standard least square curve fitting problem. However, this choice will penalize the quality metrics that have a few outliers severely.

Similar as (2), we demand that the maximum absolute difference in (18) is bounded by a certain level (denoted by ε), and adopt the support vector regression (SVR) [23] for its solution, (i.e., to determine the weight vector and the bias). In conducting SVR, we choose (14) as the kernel function. The main advantage of RBF kernel is to be able to handle the case when the relation between $(D)MOS_i$ and quality score vector $\boldsymbol{x}_i$ is nonlinear. Besides, the number of hyperparameters influences the complexity of model selection. The RBF kernel also has less hyperparameters than the polynomial kernel. That's the reason why the RBF kernel becomes our first choice.

In addition, after a series of experiments, we find out using RBF kernel always gives us better performance than other kernels (linear or polynomial) in all cases. Hence, we choose the RBF kernel in all the experiments we conducted. The choice of RBF kernel is also corroborated in [21].

F. Testing Stage

In the test stage, we use the quality score vector $\boldsymbol{x}_k$ of the kth test image, where k= 1, 2, …, l, with l being the number of test images, and (17) to determine the quality score of the MMF method, mmf $(\boldsymbol{x}_k)$. Clearly, the test can be done very fast as long as we have the trained model.

IV. Contexts for MMF Scheme

To achieve better quality scores for the MMF method, we may cluster image distortion types into several distinct groups and then determine the regression rule for each group individually. We call each group as a context, which consists of similar image distortion types. Then, the resulting scheme is called the context-dependent MMF (CD-MMF) while the scheme without the context classification stage is called context-free MMF (CF-MMF).

This involves two issues: 1) the definition of contexts and 2) automatic determination of contexts. They will be discussed in Sections IV.A and IV.B, respectively.

A. Context Definition in CD-MMF

To define the contexts for six image quality databases, including A57 [24], CSIQ [25], IVC [26], LIVE [27], TID2008 [18], and Toyoma [28], we combine similar image distortion types into one group (i.e., context). The detailed context

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
CONTEXT DEFINITION FOR EACH DATABASE

Database \ Context	I	II	III	IV	V
A57	Additive White Gaussian Noise	Gaussian Blur	JPEG, JPEG2000, JPEG2000 w/ DCQ	Quantization of Subband of DWT	
CSIQ	White Noise, Pink Noise	Gaussian Blur	JPEG, JPEG2000	Contrast Decrease	
IVC	LAR Coding	Blur	JPEG, JPEG2000		
LIVE	White Noise	Gaussian Blur	JPEG, JPEG2000	Fast Fading Rayleigh	
TID2008	1~7	8~9	10~11	12~13	14~17
Toyoma	JPEG, JPEG2000				

definition for each database is in Table IV. However, there are 17 types of distortions for TID2008 database [18]. Thus, in order to make the classification and cross-database comparison easy, we have to classify distortion types into 5 contexts according to the distortion characteristics described below:

- Group I: Collection of all kinds of additive noise
- Group II: Blurring
- Group III: JPEG + JPEG2000
- Group IV: Error caused by transmission
- Group V: Intensity deviation

B. Automatic Context Determination in CD-MMF

To perform CD-MMF, the system should be able to determine the context automatically. We extract the following five features and apply a machine learning approach to achieve this task. The first three features [29] are calculated horizontally and vertically, and combined into a single value by averaging.

(i) *Blockiness (along the Horizontal Direction):* It is defined as the average differences across block boundaries

$$B_h = \frac{1}{M(\lfloor N/8 \rfloor - 1)} \Sigma_{i=1}^{M} \Sigma_{j=1}^{\lfloor N/8 \rfloor - 1} |d_h(i, 8j)| \tag{19}$$

where $d_h(i, j) = x(i, j+1) - x(i, j), j \in [1, N-1]$, being the difference signal along horizontal line, and $x(i, j), i \in [1, M], j \in [1, N]$ for an image of size $M \times N$.

(ii) *Average Absolute Difference Between In-Block Image Samples (along the Horizontal Direction)*

$$A_h = \frac{1}{7}\left[\frac{8}{M(N-1)} \Sigma_{i=1}^{M} \Sigma_{j=1}^{N-1} |d_h(i, j)| - B_h\right] \tag{20}$$

(iii) *Zero-Crossing (ZC) Rate*

Define

$$z_h(m, n) = \begin{cases} 1 & \text{ZC happens at } d_h(m, n) \\ 0 & \text{otherwize} \end{cases}$$

Then, the horizontal ZC rate can be estimated as:

$$Z_h = \frac{1}{M(N-2)} \Sigma_{i=1}^{M} \Sigma_{j=1}^{N-2} z_h(i, j) \tag{21}$$

We only define the three horizontal components above; one can calculate vertical components B_v, A_v, Z_v in a similar fashion. Finally, the desired features are given by

$$B = \frac{B_h + B_v}{2}, A = \frac{A_h + A_v}{2}, Z = \frac{z_h + z_v}{2} \tag{22}$$

For the justification of these 3 features, the reader is referred to [29]. In addition, we introduce two more features below.

(iv) *Average Edge-Spread*

First, edge detection is applied to the image. For each edge pixel, we search the gradient direction to count the number of pixels with an increasing grey level value in the "+" direction, a decreasing grey level value in the "−" direction, and stop when significant gradient does not exist. Then, the sum of these two pixel counts is the edge-spread. The average edge-spread is computed by dividing the total amount of edge-spread by the number of edge pixels in the image. The details about this feature are given in [30].

(v) *Average Block Variance in the Image*

First, we divide the whole image into blocks of size 4×4 and classify them into "smooth" or "non-smooth" blocks based on the existence of edges. Then, we collect a set of smooth blocks of size 4×4 and make sure that they do not across the boundary of 8×8 DCT blocks. Finally, we compute the variance of each block and obtain the average.

For each database, we can use the support vector machine (SVM) algorithm and the cross-validation method to classify images into different contexts with these five features. The correct context classification rate for the selected database is listed in Table V by using SVM and the 5-fold cross-validation, where the average accuracy of the tests over 5 sets is taken as the performance measure.

Although the classification rate is reasonable, it is still lower than 90% for most databases. However, the classification rate of the context determination is not that critical since it is simply an intermediate step. The overall performance of visual quality evaluation (as to be discussed in the next section) is what we care about. Actually, as long as we use the same context classification rule in the training and the testing stages, the fusion rule will be properly selected with respect to the classified group.

Once the context of an image is given, we can apply a different fusion rule (i.e., the combination of multiple IQA methods, called the fused IQA methods) to a different context so as to optimize the assessment performance of the proposed MMF method. The block diagram of the MMF quality assessment system is given in Fig. 1.

V. FUSED IQA METHODS SELECTION

In either CF-MMF or CD-MMF, we need to find out what is the best combination for fused IQA methods, which should not

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Sequential Forward Method Selection (SFMS)

1) Start with the empty method set $M_0=\{\Phi\}$.
2) Select the next best method.

$$m^* = \arg\max_{m \in M - M_k} J(M_k + m)$$

3) Update $M_{k+1}= M_k + m^*$; k=k+1.
4) Go to 2).

TABLE V
CONTEXT CLASSIFICATION RATE FOR EACH DATABASE

Database	A57	CSIQ	IVC	LIVE	TID2008	Toyoma
Context Classification Accuracy	88.9%	78.1%	81.6%	80.7%	83.6%	100%

only achieve higher correlation with MOS (DMOS) but also have lower complexity. Given above requirements, the fused IQA methods can be selected by the algorithms as follows.

A. Sequential Forward Method Selection (SFMS)

First, given a method set $M = \{m_j | j=1, \ldots, 10\}$, we want to find a subset $M_N = \{m_{i_1}, m_{i_2}, \ldots, m_{i_N}\}$, with N<10, to optimize an objective function $J(M_N)$, which can be one of the following three forms:

$$J(M_N) = PLCC(mmf(M_N), (D)MOS) \quad (23)$$

$$J(M_N) = SROCC(mmf(M_N), (D)MOS) \quad (24)$$

$$J(M_N) = RMSE(mmf(M_N), (D)MOS) \quad (25)$$

where PLCC, SROCC, and RMSE are the Pearson linear correlation coefficient, Spearman rank order correlation coefficient, and root-mean-squared error between predicted objective scores and subjective scores, respectively. If (23) or (24) is used, then maximization needs to be applied. Otherwise, we manage to minimize (25) instead. Here, we choose (23) since PLCC represents the prediction accuracy of evaluation [31].

Sequential Forward Method Selection (SFMS) is the simplest greedy search algorithm to achieve the above goal. Starting from a method set M_k (being empty at the start), we sequentially add one method m^* that results in the highest objective function $J(M_k + m^*)$ between MOS (DMOS) and the SVR output *mmf* $(\boldsymbol{x}_i)$ to the set when combined with the method set M_k that have already been selected. The algorithm can be stated below for clarity.

B. Biggest Index Ranking Difference (BIRD)

Since we use n-fold cross-validation, we can obtain n different sets of training data for each of the ten candidate methods. Then, we define a characteristic index I_j for the jth method in the candidate method set as

$$I_j = \frac{1}{n} \Sigma_{i=1}^{n} \frac{Var[F_{j,i}]}{Mean[F_{j,i}]}, i = 1, \ldots, n; j = 1, \ldots, 10, \quad (26)$$

where $F_{j,i}$ is the ith fold training data of the jth method. The more diverse between two methods is for the trained model, the higher their characteristic index difference is. Using the index I_j, we develop the following algorithm to reduce the number of fused IQA methods.

Algorithm 2 Biggest Index Ranking Difference (BIRD)

1) Find the index of the most correlated method, and denoted as k.

$$k = \arg\max_j J(m_j)$$

2) Set the threshold value $n_{th,db\ s}$ for database s as follows:
For CF–MMF,

$$n_{th,db\ s} = \left\lceil \frac{1 - PLCC_{PSNR,db\ s}}{0.1} \right\rceil \quad (27)$$

For CD–MMF,

$$n_{th,db\ s} = \max\left(2, \left\lceil \frac{1 - PLCC_{PSNR,db\ s}}{0.1} \right\rceil - 2\right) \quad (28)$$

where $\lceil \cdot \rceil$ denotes the ceiling function, and $PLCC_{PSNR,db\ s}$ represents the Pearson linear correlation coefficient between PSNR and MOS (DMOS) for database s.
3) Compute index I_j, $j = 1,\ldots,10$.
4) Sort the methods from the smallest to the largest by index I_j and denote the ranking of method j as $r(m_j)$.
$i_1 = k$;
for $n = 1$ to $n_{th,db\ s}$
$G_n = \{i_1, \ldots, i_n\}$;
$i_{n+1} = \arg\max_{i \in \{1,\ldots,10\} - G_n} |r(m_i) - r(m_{i_n})|$;
end
5) Choose the method set $\{m_{i_1}, m_{i_2}, \ldots, m_{i_{n_{th,db\ s}+1}}\}$ as the fused IQA methods.

TABLE VI
COMPLEXITY COMPARISON OF THE FUSED IQA METHODS SELECTION ALGORITHMS IN TERMS OF REQUIRED ARITHMETIC OPERATIONS

Method	Exhaustive Search	SFMS	BIRD
N = 3	122	33	25
N = 4	212	42	26
N = 5	254	50	27
N = 6	212	57	28

Basically, this idea of finding the biggest index ranking difference (BIRD) is analogous to picking up the most dissimilar feature for the current feature. After choosing the first fused IQA method, we choose the one which has the biggest index ranking difference with the first one as the second fused IQA method since it has the most different characteristics comparing to the first chosen fused IQA method. Following the same idea, we can decide the third, the fourth, and the fifth fused IQA methods and so on.

Actually, we may think of (26) as a representation of the characteristics of the method j. Intuitively, we do not want to combine two methods having similar characteristic index value together since this way of fusion cannot give us extra advantage comparing to the original single method. That's because these two methods may share a lot of common characteristics. We may only need one of them. For instance, PSNR (m8) and PSNR-HVS (m6) both can predict image

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

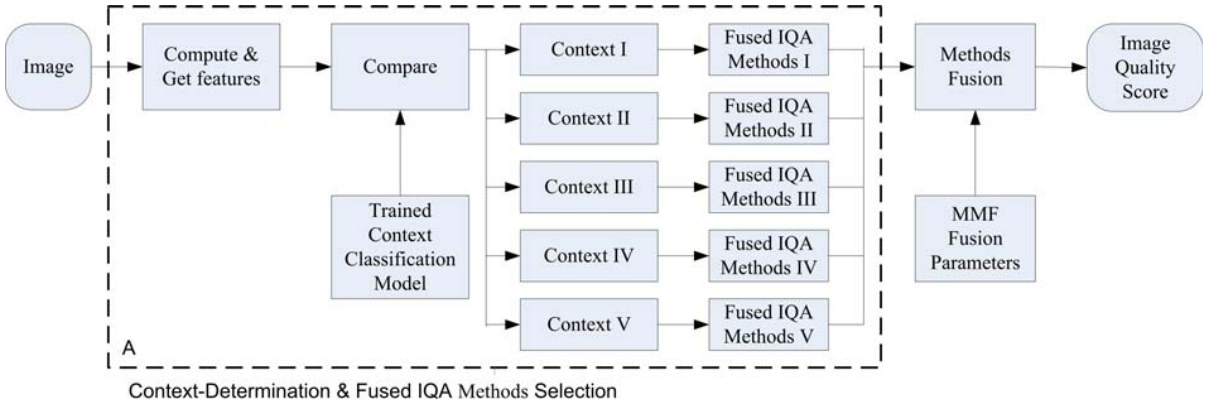


Fig. 1. Block diagram of the proposed CF-MMF (without block A) and CD-MMF (with block A) quality assessment system.

quality well on additive noise distortion types. But in order to achieve better prediction results, we never combine them together (See the BIRD case in Tables VIII and XI). Therefore, instead of doing fusion with two similar methods, we choose the next fused IQA method by picking up the method which has the most complementary characteristics to obtain extra advantage over original one.

C. Complexity Analysis

Since we have ten methods in the candidate method set, exhaustive evaluation of method subsets involves $\binom{10}{N}$ combinations for a fixed value of N, and 2^{10}combinations if the selection of N is to be optimized as well. This scale of combinations is unfeasible, even for moderate values of N. So a search procedure must be used in practice. We compare the complexity of two fused IQA methods selection approaches mentioned in Table VI. The complexity of the exhaustive evaluation is also listed together in Table V for easy comparison. Here, the algorithm complexity is measured by the total number of required arithmetic operations for 1) PLCC computation, 2) value sorting in descending order, and 3) the number of methods (features) selected for the fusion. It appears that SFMS and BIRD all have relatively low complexity comparing to the exhaustive evaluation, especially BIRD. In summary, both method selection algorithms can save us more time on finding the best combination of fused IQA methods in a systematic way for the training process. We will further demonstrate their performances in Section VI.

D. Discussion

Besides the algorithms we mentioned above, principal components analysis (PCA) [32] is a well-known method for feature reduction. It transforms the existing features into a lower dimensional space. The reason we did not use it is because we only have ten methods (features). It seems no need to reduce the dimension a lot. After all, we still need 3 to 6 methods (features) for the fusion. Moreover, we do not want to lose the score information of original methods (features). That's why we use the methods which can select the subset of features instead of transforming the features into a lower dimensional space with different units.

VI. Performance Evaluation

A. Databases

For performance evaluation, six image quality databases (A57, CSIQ, IVC, LIVE, TID2008, and Toyoma) are used. We briefly introduce each database as below.

The A57 Database [24] has 3 original images, and 54 distorted images, including six distortion types - quantization of the LH subbands of a 5-level DWT of the image using the 9/7 filters, additive Gaussian white noise, JPEG compression, JPEG2000 compression, JPEG2000 compression with the Dynamic Contrast-Based Quantization (DCQ), and Gaussian blurring. The subjective quality scores used for this database are DOMS, ranging from 0 to 1.

The Categorical Image Quality (CSIQ) Database [25] contains 30 reference images, and each image is distorted using 6 types of distortions - JPEG compression, JPEG2000 compression, global contrast decrements, additive Gaussian white noise, additive Gaussian pink noise, and Gaussian blurring - at 4 to 5 different levels, resulting in 866 distorted images. The score ratings (0 to 1) are reported in the form of DMOS.

The IVC Database [26] has 10 original images, and 185 distorted images, including 4 types of distortions – JPEG, JPEG2000, locally adaptive resolution (LAR) coding, and blurring. The subjective quality scores provided in this database are MOS, ranging from 1 to 5.

The LIVE Image Quality Database [27] has 29 reference images and 779 test images, including five distortion types - JPEG2000, JPEG, white noise in the RGB components, Gaussian blur, and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model. The subjective quality scores provided in this database are DMOS, ranging from 0 to 100.

The Tampere Image Database (TID2008) [18] includes 25 reference images, 17 types of distortions for each reference image (Table II), and 4 different levels for each type of distortion. The whole database contains 1700 distortion images. MOS is provided in this database, and the scores range from 0 to 9.

The Toyoma Database [28] has 14 original images, and 168 distorted images, including two types of distortions – JPEG, and JPEG2000. The subjective scores in this database are MOS, ranging from 1 to 5.

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

From the above introduction, we know that these six databases contain different visual contents, an abundant amount of distorted images, and diversifying image distortion types. Therefore, they can provide an excellent ground for performance evaluation.

B. Test Methodology and Performance Measure

We use the following three indices to measure IQA model performance [31], [33]. The first index is the Pearson linear correlation coefficient (PLCC) between objective/subjective scores after non-linear regression analysis. It provides an evaluation of prediction accuracy. The second index is the Spearman rank order correlation coefficient (SROCC) between the objective/subjective scores. It is considered as a measure of prediction monotonicity. The third index is the root-mean-squared error (RMSE). Before computing the first and second indices, we need to use the logistic function and the procedure outlined in [31] to fit the objective model scores to the MOS (or DMOS). The monotonic logistic function used to fit the objective prediction scores to the subjective quality scores [31] is:

$$f(x) = \frac{\beta_1 - \beta_2}{1 + e^{-(x-\beta_3)/|\beta_4|}} + \beta_2 \tag{29}$$

where x is the objective prediction score, $f(x)$ is the fitted objective score, and the parameters β_j $(j = 1,2,3,4)$ are chosen to minimize the least squares error between the subjective score and the fitted objective score. For an ideal match between the objective prediction scores and the subjective quality scores, we will have PLCC=1, SROCC=1 and RMSE=0.

One thing needs to be paid attention here. Without the logistic function, the method is still able to obtain reasonable results because RBF does a similar job. However, we observe that the logistic function enhances the performance since it offers better quality rating at the extremes of the test range. Moreover, to have a fair comparison among all IQA methods, it is still better to include the logistic function in the MMF approach.

C. CF-MMF

First, we list the performance results of CF-MMF in Table VII and plot the PLCC results in Fig. 2 to see the trend better. The methods in Table VII are selected by SFMS algorithm, as described in Section V. This selection algorithm can achieve better performance with a smaller number of methods reliably. Besides, the change of their application order does not affect the results. For example, {m9, m3, m7}, {m7, m9, m3} and {m3, m7, m9} all give the same performance. As shown in Fig. 2 (the method used in each case has been listed in Table VII), when the number of methods goes from 1 to 3, the performance improves drastically. There is still a little improvement when the number of methods increases from 3 to 8. Once the method number is over 8, the performance stays the same and sometimes even worse. Hence, we may have at least required 3 methods and up to 8 for the fusion. This phenomenon is well known as "the curse of the dimensionality" [34], [35]. The main reasons why the performance of high-dimensional data degrades rather than improve are 1) the increased noise and error when adding more features, and 2) the amount of data is not enough to obtain statistically sound and reliable estimates. This phenomenon is also called Hughes effect [36] in the field of machine learning. Therefore, we need the feature reduction (or selection) techniques introduced in Section V to reduce the dimension and achieve better results.

TABLE VII
PERFORMANCE MEASURE OF CF-MMF WITH SFMS IN TID2008 DATABASE

No. of Fused IQA Methods	Selected Fused IQA Methods	PLCC	SROCC	RMSE
1	m9	0.8726	0.8801	0.6556
2	m9, m3	0.9053	0.8953	0.5700
3	m9, m3, m7	0.9307	0.9275	0.4907
4	m9, m3, m7, m10	0.9432	0.9403	0.4460
5	m9, m3, m7, m10, m6	0.9502	0.9466	0.4183
6	m9, m3, m7, m10, m6, m1	0.9525	0.9487	0.4087
7	m9, m3, m7, m10, m6, m1, m8	0.9543	0.9495	0.4012
8	m9, m3, m7, m10, m6, m1, m8, m5	0.9546	0.9494	0.3999
9	m9, m3, m7, m10, m6, m1, m8, m5, m4	0.9543	0.9493	0.4008
10	m9, m3, m7, m10, m6, m1, m8, m5, m4, m2	0.9532	0.9483	0.4056

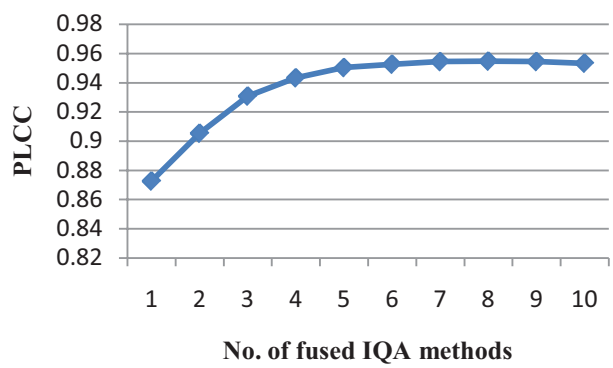


Fig. 2. PLCC performance of CF-MMF in the TID2008 database.

To balance the performance and complexity, the number of fused IQA methods needed for each database can be computed via (27), which is equal to $n_{th,db\,s}$+1. Hence, 3 methods are used for CSIQ and LIVE databases, 4 methods for IVC database, 5 methods for A57 and Toyoma databases, and 6 methods for TID2008 database, respectively. Here, more than 3 methods are determined for IVC, A57, Toyoma and TID2008 databases because they are the databases which are less correlated with PSNR or other image quality indices (See Table XIII). We need to fuse more methods to achieve the same performance as in the other two databases. The selected fused IQA methods by using two algorithms (SFMS and BIRD) are listed in Table VIII. The corresponding

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII
SELECTED FUSED IQA METHODS FOR CF-MMF IN SIX DATABASES

Selection Algorithm / Database	SFMS	BIRD
A57	m4, m9, m10, m7, m1	m4, m8, m5, m10, m2
CSIQ	m10,m3, m6	m10, m1, m7
IVC	m9, m10, m5, m4	m9, m10, m1, m3
LIVE	m10, m3, m6	m10, m1, m3
TID2008	m9, m3, m7, m10, m6, m1	m9, m7, m1, m10, m2, m3
Toyoma	m10, m5, m6, m3, m2	m10, m1, m7, m2, m3

TABLE IX
PERFORMANCE MEASURE FOR CF-MMF IN SIX DATABASES

Database	Selection Algorithm	PLCC	SROCC	RMSE
A57	SFMS	0.9604	0.9590	0.0685
(5 methods)	BIRD	0.9465	0.9475	0.0793
CSIQ	SFMS	0.9797	0.9755	0.0527
(3 methods)	BIRD	0.9698	0.9657	0.0641
IVC	SFMS	0.9352	0.9226	0.4313
(4 methods)	BIRD	0.9205	0.9096	0.4760
LIVE	SFMS	0.9734	0.9732	6.2612
(3 methods)	BIRD	0.9712	0.9710	6.5131
TID2008	SFMS	0.9525	0.9487	0.4087
(6 methods)	BIRD	0.9482	0.9434	0.4261
Toyoma	SFMS	0.9477	0.9419	0.3995
(5 methods)	BIRD	0.9451	0.9402	0.4091

performances for both algorithms are also compared in Table IX. We can see that the performance is better by using the methods selected by SFMS. However, although the performance degrades a little when using BIRD, the complexity of BIRD is lower than SFMS, as demonstrated in Table VI.

D. CD-MMF

For CD-MMF, we increase the number of fused IQA methods from 1 to 5, and the performance stays almost the same (i.e., stop improving) when the number of fused methods is over 4 (Table X). Note that the methods in Table X are selected by SFMS algorithm, which can help us reliably achieve better performance with a smaller number of methods. The PLCC results are also plotted in Fig. 3 for comparison. We see that the quality prediction reaches the best correlation when the number of fused methods increases to four. Thus, we have similar conclusions as in CF-MMF, except for the upper limit of fused IQA methods decreasing from 8 to 4, and this allows us to lower the complexity of the proposed system for CD-MMF. The reason why CD-MMF works as well as CF-MMF and needs less fused IQA methods is because of the pre-classification of distortions in the first stage. Then for these specific known distortions, we can use smaller number of fused IQA methods to achieve the same level of performance.

As already mentioned earlier in Section V, the optimal combination of fused IQA methods for each context is selected by two algorithms (SFMS and BIRD). The fused IQA methods selection rules for six databases are summarized in Table XI for CD-MMF. And the performance measures for both algorithms are shown in Table XII. We can observe that the

TABLE X
PERFORMANCE MEASURE OF CD-MMF WITH SFMS IN TID2008 DATABASE

No. of fused IQA Methods / Measure	1	2	3	4	5
PLCC	0.9139	0.9265	0.9517	0.9538	0.9539
SROCC	0.9046	0.9174	0.9439	0.9463	0.9464
RMSE	0.5588	0.5049	0.4120	0.4032	0.4026

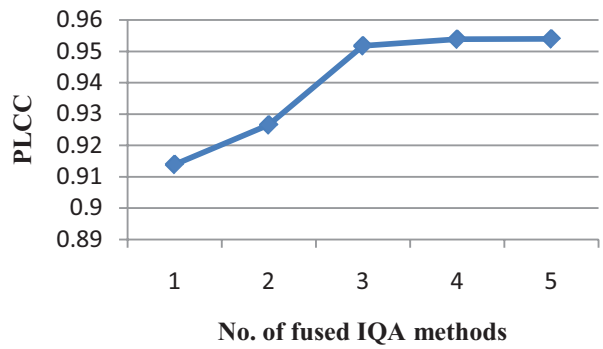


Fig. 3. PLCC comparison of CD-MMF in the TID2008 database.

performance difference between SFMS and BIRD becomes smaller in CD-MMF. Especially, BIRD works a little better than SFMS in Toyoma database for CD-MMF setting. For the same reason as in CF-MMF, the number of fused IQA methods for each database can be decided via (28), which is still equal to $n_{th,db\,s}+1$. Thus, 3 methods are adopted for all databases, except TID2008. Here, 4 methods are used for TID2008 database instead since it is the most difficult (in term of correlation) database among all.

One more thing needs attention here. As shown in Table XI, the most frequent combinations of methods are {m10, m1, m3} and {m10, m1, m7} for BIRD algorithm, which are highlighted in bold in Table XI. These two combinations work well in 9 cases, especially for Blur and JPEG+JPEG2000 distortions. However, there is no combination of methods that works consistently well for SFMS algorithm among all databases. Therefore, the BIRD algorithm is more general than the SFMS algorithm.

E. Performance Comparison Between MMF and the Existing Methods

Finally, we compare the two proposed methods (i.e., CF-MMF, CD-MMF) and other mentioned quality indices with respect to the databases A57, CSIQ, IVC, LIVE, TID2008, and Toyoma. Here, we include one more quality index, information content weighted SSIM (IW-SSIM) [37] into the comparison since it has been proved to work better than other variants of SSIM. The results are shown in Tables XIII, where the top three ranked methods are highlighted in bold. Apparently, top three ranked methods are all our proposed approaches, except VSNR in A57, and FSIM in IVC database.

We observe that CD-MMF achieves the same performance as CF-MMF but with a smaller number of methods as shown in Table XIII. In other words, CD-MMF will perform better

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE XI
SELECTED FUSED IQA METHODS FOR CD-MMF IN SIX DATABASES

Database	Context \ Algorithm	SFMS	BIRD
A57	I	m10, m6, m7	m10, m7, m8
	II	m4, m2, m1	m4, m8, m1
	III	m1, m10, m4	m1, m10, m5
	IV	m9, m10, m4	m9, m7, m8
CSIQ	I	m10, m6, m2	**m10, m1, m7**
	II	m10, m8, m5	**m10, m1, m3**
	III	m10, m2, m8	**m10, m1, m3**
	IV	m3, m6, m7	m3, m1, m4
IVC	I	m10, m6, m1	**m10, m1, m3**
	II	m10, m3, m7	**m10, m7, m1**
	III	m7, m2, m10	m7, m10, m1
LIVE	I	m3, m5, m10	m3, m9, m10
	II	m3, m1, m6	m3, m1, m7
	III	m10, m2, m4	**m10, m1, m3**
	IV	m3, m6, m9	**m3, m1, m10**
TID2008	I	m6, m4, m5, m1	m6, m4, m10, m1
	II	m10, m6, m1, m5	m10, m4, m3, m7
	III	m9, m8, m6, m10	m9, m4, m3, m7
	IV	m3, m8, m9, m4	m3, m4, m5, m8
	V	m10, m7, m3, m1	m10, m3, m7, m9
Toyoma	I	m10, m5, m6	**m10, m1, m7**

TABLE XII
PERFORMANCE MEASURE FOR CD-MMF IN SIX DATABASES

Database	Selection Algorithm	PLCC	SROCC	RMSE
A57	SFMS	0.9411	0.9498	0.0831
(3 methods)	BIRD	0.9347	0.9354	0.0874
CSIQ	SFMS	0.9675	0.9668	0.0664
(3 methods)	BIRD	0.9630	0.9609	0.0707
IVC	SFMS	0.9453	0.9382	0.3976
(3 methods)	BIRD	0.9374	0.9285	0.4244
LIVE	SFMS	0.9802	0.9805	5.4134
(3 methods)	BIRD	0.9801	0.9798	5.4239
TID2008	SFMS	0.9538	0.9463	0.4032
(4 methods)	BIRD	0.9476	0.9422	0.4289
Toyoma	SFMS	0.9456	0.9411	0.4071
(3 methods)	BIRD	0.9462	0.9421	0.4051

than CF-MMF when using the same number of methods (see Table XIII (c), (d), (e), (f)). The A57 and CSIQ databases are the only exceptions among these six databases. There is no obvious improvement by using CD-MMF in CSIQ database as shown in Table XIII (b). We think that it may be caused by the lower context classification rate (see Table V). In addition, the performance of CD-MMF is lower than CF-MMF in A57 database (Table XIII (a)). This probably results from the small number of training images (less than 10 for each context) we can use in A57 after the context classification. In order to see the trend better, we plot the corresponding bar chart of the PLCC for the TID2008 database in Fig. 4. Clearly, CD-MMF ranks the first (with the highest PLCC, SROCC and the smallest RMSE), and CF-MMF the second among all 15 IQA models in comparison.

The scatter plots of predicted objective scores vs. MOS for all the IQA models along with the best fitting logistic functions are shown in Fig. 5. Each point on the plot represents

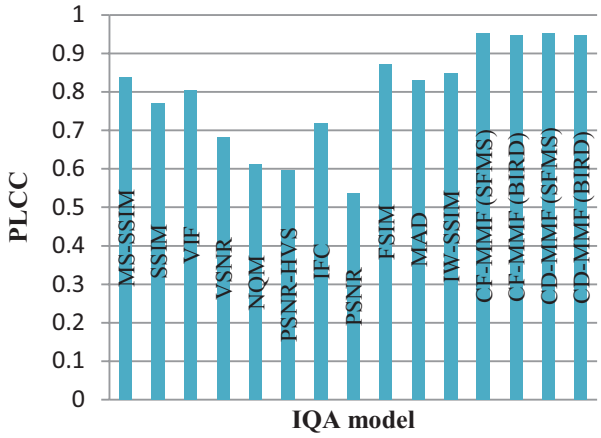


Fig. 4. Comparison of the PLCC measure of 15 IQA models in the TID2008 database.

one image in the database, the horizontal axis corresponds to the (scaled) IQA score for that image and the vertical axis corresponds to the subjective MOS for that image. The scatter plots confirm the results shown in Table XIII (e). Obviously, CF-MMF and CD-MMF have the best correlation with human judgments (i.e., MOS).

F. Cross-Database Evaluation

To test the generality of our proposed approach, we perform the cross-database evaluation since MMF is a learning-oriented method [38]. As we know, each database contains different image distortion types. In order to have a fair performance evaluation, we divide the images into several distortion groups (i.e., contexts) according to Table IV. There are four major distortion groups across all databases, namely:

1) Additive Noise: Context I in A57, CSIQ, LIVE and TID2008.
2) Blur: Context II in A57, CSIQ, IVC, LIVE and TID2008.
3) JPEG+JPEG2000: Context III in A57, CSIQ, IVC, LIVE, TID2008, and context I in Toyoma.
4) Transmission Error: Context IV in LIVE and TID2008.

We use all the images from one distortion group in one database for the training, and test the images from the same distortion group in the other remaining databases. For example, we build a trained model (for additive noise) from LIVE database. Then, we use this trained model to do the testing on A57, CSIQ, and TID2008. Here, we use LIVE and TID2008 as our training databases since they cover most distortion groups and also have larger sizes of training images. Moreover, we only choose 3 methods for fusion in each case to reduce the complexity of MMF. The performance results are summarized in Table XIV.

For additive noise distortions, we see that using LIVE database as the trained model would give us slightly better performance than TID2008 in A57 and CSIQ. The performance with TID2008 is acceptable, but not as good as that with these aforementioned two databases. This is probably because

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE XIII
PERFORMANCE COMPARISON AMONG 15 IQA MODELS IN SIX DATABASES

(a) A57 Database (54 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.8737	0.8578	0.1196
SSIM	0.8019	0.8067	0.1469
VIF	0.6160	0.6223	0.1936
VSNR	**0.9502**	**0.9359**	**0.0766**
NQM	0.8027	0.7978	0.1466
PSNR-HVS	0.8832	0.8502	0.1153
IFC	0.4549	0.3187	0.2189
PSNR	0.6347	0.6189	0.1899
FSIM	0.9253	0.9181	0.2458
MAD	0.9059	0.9014	0.1041
IW-SSIM	0.9024	0.8713	0.1059
CF-MMF (SFMS) (5 methods)	**0.9604**	**0.9590**	**0.0685**
CF-MMF (BIRD) (5 methods)	**0.9465**	**0.9475**	**0.0793**
CD-MMF (SFMS) (3 methods)	0.9411	0.9498	0.0831
CD-MMF (BIRD) (3 methods)	0.9347	0.9354	0.0874

(b) CSIQ Database (866 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.8666	0.8774	0.1310
SSIM	0.8594	0.8755	0.1342
VIF	0.9253	0.9194	0.0996
VSNR	0.8005	0.8108	0.1573
NQM	0.7422	0.7411	0.1759
PSNR-HVS	0.8231	0.8294	0.1491
IFC	0.8358	0.7671	0.1441
PSNR	0.8001	0.8057	0.1576
FSIM	0.9095	0.9242	0.1091
MAD	0.9502	0.9466	0.0818
IW-SSIM	0.9025	0.9212	0.1131
CF-MMF (SFMS) (3 methods)	**0.9797**	**0.9755**	**0.0527**
CF-MMF (BIRD) (3 methods)	**0.9698**	**0.9657**	**0.0641**
CD-MMF (SFMS) (3 methods)	**0.9675**	**0.9668**	**0.0664**
CD-MMF (BIRD) (3 methods)	0.9630	0.9609	0.0707

(c) IVC Database (185 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.9108	0.8971	0.5031
SSIM	0.9117	0.9018	0.5007
VIF	0.9026	0.8964	0.5244
VSNR	0.8027	0.7993	0.7265
NQM	0.8489	0.8343	0.6440
PSNR-HVS	0.8648	0.8590	0.6118
IFC	0.9093	0.8993	0.5069
PSNR	0.7192	0.6885	0.8465
FSIM	**0.9376**	**0.9262**	**0.4236**
MAD	0.9210	0.9146	0.4748
IW-SSIM	0.9228	0.9125	0.4693
CF-MMF (SFMS) (4 methods)	0.9352	0.9226	0.4313
CF-MMF (BIRD) (4 methods)	0.9205	0.9096	0.4760
CD-MMF (SFMS) (3 methods)	**0.9453**	**0.9382**	**0.3976**
CD-MMF (BIRD) (3 methods)	**0.9374**	**0.9285**	**0.4244**

(d) LIVE Database (779 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.9402	0.9521	9.3038
SSIM	0.9384	0.9479	9.4439
VIF	0.9597	0.9636	7.6737
VSNR	0.9235	0.9279	10.4816
NQM	0.9128	0.9093	11.1570
PSNR-HVS	0.9134	0.9186	11.1228
IFC	0.9261	0.9259	10.3052
PSNR	0.8701	0.8756	13.4685
FSIM	0.9540	0.9634	8.1938
MAD	0.9672	0.9669	6.9419
IW-SSIM	0.9425	0.9567	9.1301
CF-MMF (SFMS) (3 methods)	**0.9734**	**0.9732**	**6.2612**
CF-MMF (BIRD) (3 methods)	0.9712	0.9710	6.5131
CD-MMF (SFMS) (3 methods)	**0.9802**	**0.9805**	**5.4134**
CD-MMF (BIRD) (3 methods)	**0.9801**	**0.9798**	**5.4239**

(e) TID2008 Database (1700 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.8389	0.8528	0.7303
SSIM	0.7715	0.7749	0.8537
VIF	0.8055	0.7496	0.7953
VSNR	0.6820	0.7046	0.9815
NQM	0.6103	0.6243	1.0631
PSNR-HVS	0.5977	0.5943	1.0759
IFC	0.7186	0.5707	0.9332
PSNR	0.5355	0.5245	1.1333
FSIM	0.8710	0.8805	0.6592
MAD	0.8306	0.8340	0.7474
IW-SSIM	0.8488	0.8559	0.7094
CF-MMF (SFMS) (6 methods)	**0.9525**	**0.9487**	**0.4087**
CF-MMF (BIRD) (6 methods)	**0.9482**	**0.9434**	**0.4261**
CD-MMF (SFMS) (4 methods)	**0.9538**	**0.9463**	**0.4032**
CD-MMF (BIRD) (4 methods)	0.9476	0.9422	0.4289

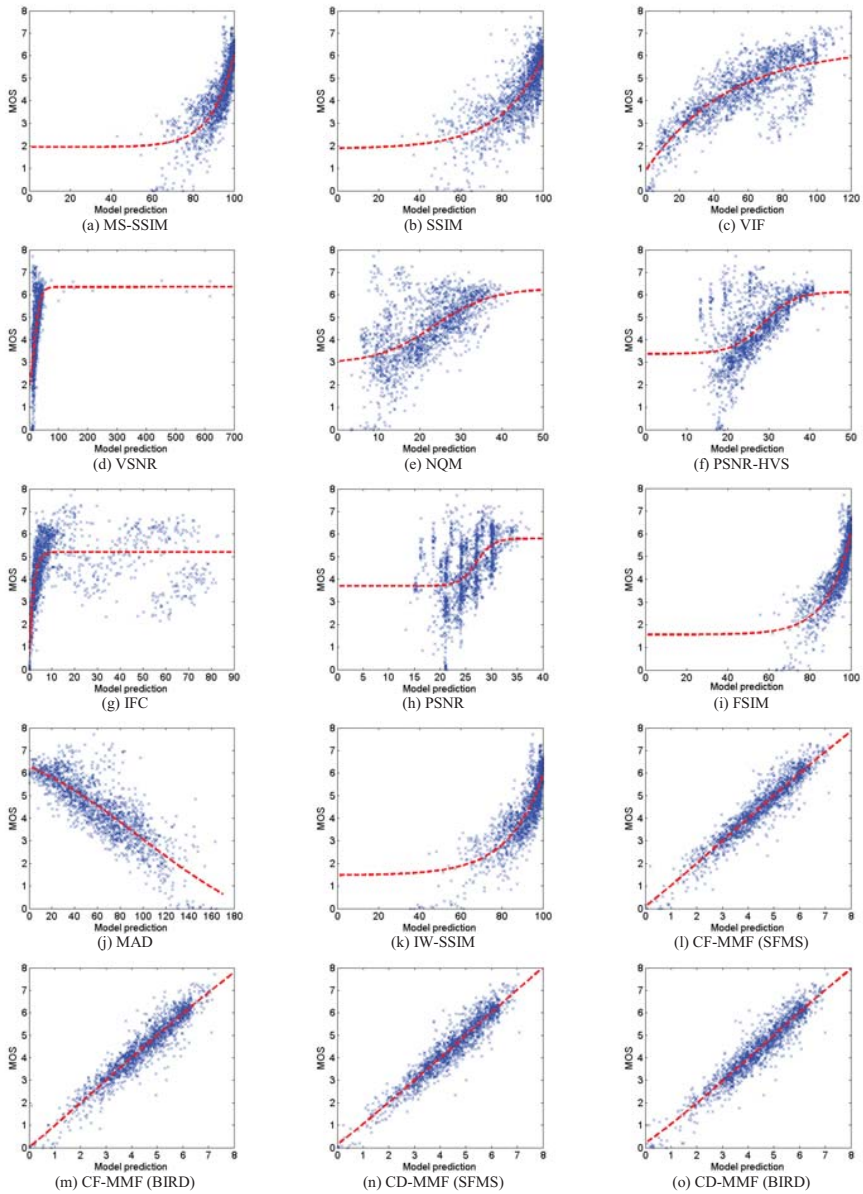
(f) Toyoma Database (168 images)			
IQA Model \ Measure	PLCC	SROCC	RMSE
MS-SSIM	0.8948	0.8911	0.5588
SSIM	0.8877	0.8794	0.5762
VIF	0.9137	0.9077	0.5087
VSNR	0.8705	0.8609	0.6160
NQM	0.8893	0.8871	0.5724
PSNR-HVS	0.7884	0.7817	0.7700
IFC	0.8404	0.8355	0.6784
PSNR	0.6355	0.6133	0.9663
FSIM	0.9077	0.9059	0.5253
MAD	0.9406	0.9362	0.4248
IW-SSIM	0.9244	0.9203	0.4774
CF-MMF (SFMS) (5 methods)	**0.9477**	**0.9419**	**0.3995**
CF-MMF (BIRD) (5 methods)	0.9451	0.9402	0.4091
CD-MMF (SFMS) (3 methods)	**0.9456**	**0.9411**	**0.4071**
CD-MMF (BIRD) (3 methods)	**0.9462**	**0.9421**	**0.4051**

TID2008 contains 7 kinds of additive noises, and LIVE only contains 1 type of additive noise.

For blur distortions, the TID2008 trained model has better prediction performance than LIVE in A57 and CSIQ. Regarding JPEG and JPEG2000 distortions, the LIVE trained model predicts more accurately than the TID2008 trained one in A57, IVC and Toyoma. For transmission errors, the TID2008 trained model can provide a better performance for LIVE than the LIVE trained model can provide for TID2008.

Moreover, we use all the images from one database for the training, and test the images from the other remaining databases. For instance, we build a trained model from LIVE database. Then, we use this trained model to do the testing on A57, CSIQ, IVC, TID2008 and Toyoma. Here, we also use LIVE and TID2008 as our training databases since they cover more distortion groups and also have larger sizes of

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.



(a) MS-SSIM (b) SSIM (c) VIF
(d) VSNR (e) NQM (f) PSNR-HVS
(g) IFC (h) PSNR (i) FSIM
(j) MAD (k) IW-SSIM (l) CF-MMF (SFMS)
(m) CF-MMF (BIRD) (n) CD-MMF (SFMS) (o) CD-MMF (BIRD)

Fig. 5. Scatter plots and best fitting logistic function of objective IQA scores versus MOS for the TID2008 database.

TABLE XIV
CROSS-DATABASE PLCC PERFORMANCE OF MMF (IN TERMS OF DISTORTION GROUPS)

Distortion Group	Test Database / Model	A57	CSIQ	IVC	LIVE	TID2008	Toyoma
Additive Noise	LIVE	0.9827	0.9662	-	-	0.9195	-
	TID2008	0.9782	0.9586	-	0.9863	-	-
Blur	LIVE	0.8754	0.9467	0.9841	-	0.9255	-
	TID2008	0.9449	0.9652	0.9841	0.9455	-	-
JPEG+JPEG2000	LIVE	0.9725	0.9687	0.9191	-	0.9755	0.9361
	TID2008	0.9289	0.9685	0.9055	0.9710	-	0.8651
Transmission Error	LIVE	-	-	-	-	0.8898	-
	TID2008	-	-	-	0.9607	-	-

training images. Similarly, we only choose 3 IQA methods for fusion in each case to reduce the complexity of MMF. The performance results are summarized in Table XV. We can observe the PLCC for TID2008 is below 0.9 since only 3 IQA

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE XV
CROSS-DATABASE PLCC PERFORMANCE OF MMF (IN TERMS OF WHOLE DATABASE)

Model \ Test Database	A57	CSIQ	IVC	LIVE	TID2008	Toyoma
LIVE	0.8839	0.9710	0.9300	-	0.8984	0.9374
TID2008	0.9454	0.9127	0.9277	0.9644	-	0.9132

TABLE XVI
N-FOLD CROSS-VALIDATION PERFORMANCE OF CF-MMF WITH SFMS IN TID2008 DATABASE

n	2	3	4	5
PLCC	0.9501	0.9515	0.9531	0.9525
SROCC	0.9465	0.9477	0.9497	0.9487
RMSE	0.4184	0.4130	0.4063	0.4087

TABLE XVII
PERFORMANCE COMPARISON AMONG FUSED IQA METHODS SELECTION ALGORITHMS FOR CF-MMF IN LIVE DATABASE

Algorithm	Fused IQA Methods	PLCC	SROCC	RMSE
Exhaustive Search	m10, m3, m6	0.9734	0.9732	6.2612
SFMS	m10, m3, m6	0.9734	0.9732	6.2612
BIRD	m10, m1, m3	0.9712	0.9710	6.5131
Random Selection 1	m1, m3, m5	0.9541	0.9556	8.1821
Random Selection 2	m2, m4, m6	0.9437	0.9473	9.0346

methods are fused for this case. However, the PLCC is over 0.91 for most cases. This further verifies the generality and robustness of our proposed system.

G. Further Discussion

First, we want to investigate how the performance changes when we use different selections of n for n-fold cross-validation. We perform n-fold ($n = 2\sim5$) cross-validation in TID2008 database. Table XVI shows the performance results. There is little difference for the performance among different n values. As long as there are enough training data, the results will be similar for different choices of n values.

Second, we would like to compare the performances among several fused IQA methods selection algorithms, which include exhaustive search, SFMS, BIRD, and random selection. The results are listed in Table XVII.

Obviously, both SFMS and BIRD have better performance than the random selection of fused IQA methods. Particularly, SFMS has the same performance as the exhaustive search of fused IQA methods since they all select the same methods for fusion. However, the computation complexity is much lower for SFMS and BIRD comparing to the exhaustive search (see N = 3 case in Table VI). Hence, we can conclude that a good fused IQA methods selection algorithm (e.g., SFMS or BIRD) is indeed necessary for both performance improvement and complexity reduction.

Next, we want to test if different objective functions have different impacts on the performance. Three objective functions (23), (24), and (25) are applied separately on the fused IQA methods selection algorithm. From Table XVIII, we know that they all lead to choose the same methods for fusion. This result is quite reasonable since PLCC becomes higher when SROCC value increases in most of the cases. In the meantime, the RMSE value also decreases accordingly. That's why we obtain the same performance no matter which one is applied.

TABLE XVIII
PERFORMANCE COMPARISON AMONG DIFFERENT OBJECTIVE FUNCTIONS $J(Y_N)$ FOR CF-MMF (WITH SFMS) IN LIVE DATABASE

$J(Y_N)$	Fused IQA Methods	PLCC	SROCC	RMSE
(23)	m10, m3, m6	0.9734	0.9732	6.2612
(24)	m10, m3, m6	0.9734	0.9732	6.2612
(25)	m10, m3, m6	0.9734	0.9732	6.2612

TABLE XIX
PERFORMANCE COMPARISON IN TID2008 DATABASE

Model	IQA Methods	Weights	PLCC	SROCC	RMSE
Linear combination [17]	m9, m3	[0.95, 0.05]	0.8751	0.8820	0.6495
		[0.85, 0.15]	0.8752	0.8753	0.6491
		[0.5, 0.5]	0.8476	0.8196	0.7120
		[0.15, 0.85]	0.8197	0.7681	0.7687
		[0.05, 0.95]	0.8106	0.7554	0.7859
SVR	m9, m3	[6.2830, 7.5598]	0.9053	0.8953	0.5700

In Table XIX we compare the performance between IQA methods combined by linear weights and SVR. We try several linear combinations of weights mentioned in [17] to combine m9 and m3. The best PLCC result is 0.8751, which is still lower than 0.9053, obtained by SVR. The comparison in Table XIX clearly shows the advantage of SVR over linear combination. Here, we only choose two methods for fusion in SVR to have a fair comparison with linear combination approach.

Finally, the computational complexity among 15 IQA models is compared in Table XX. The measurement is in terms of the computation time required to evaluate an image of size 512 x 512 by using a computer with Intel core i7 processor @1.73 GHz. It can be seen that 10 of the other 11 IQA models require less than 5 seconds finishing the assessment, except MAD. MAD still needs about 54 seconds to complete the job. Since MAD is also one of the IQA methods used in CF-MMF and CD-MMF, the computation time for both MMF-based approaches is definitely greater than 54 seconds. As we can see in Table XX, the computation time for MMF methods is approximately 1 minute, while CD-MMF also requires 3 to 4 seconds more than CF-MMF for the assessment. However, it would be 5 to 6 seconds shorter when we select methods by using BIRD instead of SFMS.

VII. CONCLUSION

As far as we know, there is no single existing image quality index gives the best performance in all situations. Thus, in this work, we have proposed an open, inclusive framework for better performance with the current level of technology and for easy extension when new technology emerges. To be more

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

TABLE XX
COMPARISON OF COMPUTATION TIME AMONG 15 IQA MODELS IN CSIQ DATABASE

IQA Model	Time (sec/image)
MS-SSIM	0.21
SSIM	0.10
VIF	3.22
VSNR	0.63
NQM	0.49
PSNR-HVS	4.95
IFC	2.00
PSNR	0.07
FSIM	0.79
MAD	53.57
IW-SSIM	0.94
CF-MMF (SFMS) (3 methods fused)	61.74
CF-MMF (BIRD) (3 methods fused)	55.78
CD-MMF (SFMS) (3 methods fused)	64.40
CD-MMF (BIRD) (3 methods fused)	59.63

specific, we have presented a multi-method fusion (MMF) approach for image quality assessment and proposed two MMF-based quality indices, based upon machine learning. It was shown by experiments with six different databases (totally 3752 images) that both of them outperform state-of-the-art quality indices by a significant margin.

As expected, the complexity of the MMF method is higher since it involves the calculation of multiple methods. However, with the help of the algorithms (SFMS and BIRD), we can reduce the number of fused methods and lower the complexity of MMF. As long as we can keep the number of fused IQA methods not more than three, the computational time of MMF is around 1 minute per image. In most cases, we only need 3 methods for the fusion to achieve satisfactory performance (i.e., PLCC over 0.93). Even in the most difficult (in terms of correlation) TID2008 database, MMF can also achieve 0.9307 and 0.9517 on PLCC (Tables VII and X) by using 3 methods for CF-MMF and CD-MMF, respectively. The performance of MMF outperforms other well-known image quality indices.

Another advantage of the proposed MMF methodology is its flexibility in including new methods. For example, in this work, we have added two new methods, which are FSIM [10] and MAD [11], into the candidate method set to replace poor-performed methods (VIFP [5] and UQI [39]) that we used in [12]; It turns out that the PLCC performance of CD-MMF improves from 0.9438 to 0.9538 for TID2008 database. This is a good demonstration of the forward inclusiveness of the proposed methodology.

Although the proposed MMF has excellent performance, one issue concerning context classification for CD-MMF needs to be resolved in the future. Since one image may consist of multiple distortion types, the strict classification of images into one specific context may lead to the wrong context category, and then affect the subsequent quality prediction. One possible and better solution to overcome this shortcoming is to use unsupervised classification for context determination. Another alternative is to attach beliefs to the classification of the context and weight the corresponding regressed predicted quality indices (as in [40]). The two approaches mentioned above should be able to help to further improve the overall performance of MMF.

In the end, we perform the tests across databases. The results are quite promising. Therefore, the generality of our proposed MMF approach is also demonstrated.

REFERENCES

[1] B. Girod, "What's wrong with mean-squared error?," *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[2] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[3] Z. Wang, E. Simoncelli, A. Bovik, and M. Matthews, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[4] 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, Apr. 2004.

[5] 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.

[6] 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.

[7] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.

[8] K. Egiazarian, J. Astola, N. Ponomarenko, and V. Lukin, F. Battisti, and M. Carli, "A new full-reference quality metrics based on HVS," in *Proc. 2nd Int. Workshop Video Process. Quality Metrics*, 2006, pp. 1–4.

[9] 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.

[10] 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, Aug. 2011.

[11] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, Mar. 2010.

[12] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment," in *Proc. IEEE 3rd Int. Workshop Qual. Multimedia Experience*, Sep. 2011, pp. 72–77.

[13] H. Luo, "A training-based no-reference image quality assessment algorithm," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 2973–2976.

[14] S. Suresh, V. Babu, and N. Sundararajan, "Image quality measurement using sparse extreme learning machine classifier," in *Proc. IEEE 9th Int. Conf. Control, Autom., Robot. Vis.*, Dec. 2006, pp. 1–6.

[15] S. Suresh, V. Babu, and H. J. Kim, "No-reference image quality assessment using modified extreme learning machine classifier," *J. Appl. Soft Comput.*, vol. 9, no. 2, pp. 541–552, Mar. 2009.

[16] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.

[17] A. Leontaris, P. C. Cosman, and A. R. Reibman, "Quality evaluation of motion-compensated edge artifacts in compressed video," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 943–956, Apr. 2007.

[18] *Tampere Image Database* [Online]. Available: http://www.ponomarenko.info/tid2008.htm

[19] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Inf. Process., Lett. Rev.*, vol. 11, no. 10, pp. 203–224, Oct. 2007.

[20] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[21] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., National Taiwan Univ., Taipei, Taiwan, Tech. Rep., Apr. 2010.

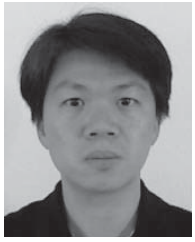
[22] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer-Verlag, 2006.

[23] C.-C. Chang and C.-J. Lin, "LIBSVM : A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

[24] *A57 Database* [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.

[25] *Categorical Image Quality (CSIQ) Database* [Online]. Available: http://vision.okstate.edu/csiq
[26] *IVC Image Quality Database* [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb
[27] *LIVE Image Quality Assessment Database* [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm
[28] *Toyoma Database*. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html
[29] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, pp. 477–480.
[30] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Proc. IEEE 7th Int. Symp. Signal Process. Appl.*, Jul. 2003, pp. 469–472.
[31] *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase I*. (2000, Mar.) [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI
[32] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
[33] *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*. (2003, Aug.) [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII
[34] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ: Princeton Univ. Press, 1961.
[35] J. H. Friedman, "On bias, variance, 0/1—loss, and the curse-of-dimensionality," *Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 55–77, 1997.
[36] T. Oommen, D. Misra, N. K. C. Twarakavi, A. Prakash, B. Sahoo, and S. Bandopadhyay, "An objective analysis of support vector machine based classification for remote sensing," *Math. Geosci.*, vol. 40, no. 4, pp. 409–424, 2008.
[37] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[38] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Recent developments and future trends in visual quality assessment," in *Proc. APSIPA ASC*, Oct. 2011, pp. 1–10.
[39] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[40] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.



Tsung-Jung Liu (S'10) received the B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, and the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 1998 and 2001, respectively. Now, he is pursuing the Ph.D. degree in electrical engineering at University of Southern California (USC), Los Angeles.

He is currently a Research Assistant with the Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, USC. He was also a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore, in summer 2011. His current research interests include machine learning, perceptual image/video processing, and visual quality assessment.

Weisi Lin (M'92–SM'98) received the B.Sc. and M.Sc. degrees from Zhongshan University, Guangzhou, China, and the Ph.D. degree from the King's College, London University, London,U.K.

He was the Head of the Laboratory of Visual Processing and the Acting Manager with Department of Media Processing, Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision. He has authored or co-authored more than 200 refereed papers in international journals and conferences.

Dr. Lin is on the Editorial Board of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS and the *Journal of Visual Communication and Image Representation*. He was the Lead Guest Editor of a special issue on perceptual signal processing, the IEEE Journal of Selected Topics in Signal Processing in 2012. He is the Chair of the IEEE MMTC Special Interest Group on Quality of Experience. He has been elected as a Distinguished Lecturer of APSIPA (2012/3). He is the Lead Technical Program Chair of Pacific-Rim Conference on Multimedia (PCM) in 2012, and the Technical Program Chair of the *IEEE International Conference on Multimedia and Expo (ICME)*, in 2013. He is a Chartered Engineer in U.K., a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

C.-C. Jay Kuo (F'99) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is currently the Director of the Media Communications Laboratory and a Professor of electrical engineering, computer science, and mathematics with the Department of Electrical Engineering and Integrated Media Systems Center, University of Southern California, Los Angeles, and the President of the Asia-Pacific Signal and Information Processing Association. His current research interests include digital image and video analysis and modeling, multimedia data compression, communication, and networking. He has authored or co-authored around 200 journal papers, 850 conference papers, and 10 books.

Dr. Kuo is a fellow of the American Association for the Advancement of Science and the International Society for Optical Engineers.

Authorized licensed use limited to: Manuek Castro. Downloaded on August 12,2025 at 05:13:50 UTC from IEEE Xplore. Restrictions apply.