Analysis of Distortion Distribution for Pooling in Image Quality Prediction

Ke Gu, Shiqi Wang, *Member, IEEE,* Guangtao Zhai, *Member, IEEE,* Weisi Lin, *Fellow, IEEE*
Xiaokang Yang, *Senior Member, IEEE,* and Wenjun Zhang, *Fellow, IEEE*

***Abstract*—Image quality assessment (IQA) has been an active research area during last decades. Many existing objective IQA models share a similar two-step structure with measuring local distortion before pooling. Compared with the rapid development for local distortion measurement, seldom effort has been made dedicated to effective pooling schemes. In this paper, we design a new pooling model via the analysis of distortion distribution affected by image content and distortion. That is, distributions of distortion position, distortion intensity, frequency changes, and histogram changes are comprehensively considered to infer an overall quality score. Experimental results conducted on four large-scale image quality databases (LIVE, TID2008, CSIQ and CCID2014) concluded with three valuable findings. First, the proposed technique leads to consistent improvement in the IQA performance for studied local distortion measures. Second, relative to the traditional pooling, the performance gain of our algorithm is beyond 15% on average. Third, the best overall performance made by the proposed strategy outperforms state-of-the-art competitors.**

***Index Terms*—Image quality assessment (IQA), pooling, distortion distribution, multi-scale (MS), ranking-based weighting (RW), frequency variation-induced adjuster (FVA), entropy gain multiplier (EGM)**

I. Introduction

NOWADAYS, how to precisely predict image quality has aroused broad attention from the digital image/video processing community, because on one hand image quality assessment (IQA) can facilitate the research of technologies for 3D imaging [1-3] and image retargeting [4], and on the other hand it has numerous applications in the development and optimization of image/video processing algorithms, e.g. compression [5-7], transmission [8-9] and enhancement [10]. We can divide IQA into subjective assessment and objective assessment. The role of the former one is important, since in most situations, human beings are the ultimate users of the processed images and thus the judges of visual quality. But subjective IQA metrics are labor intensive and not efficient and affordable, and thereby do not work possibly under the condition that hundreds of thousands of images are acquired, compressed and transmitted every moment. So a rising number of studies have been devoted to the exploration of objective IQA models, which automatically predict the visual quality for simulating subjective ratings via mathematical models.

Depending on the availability of *perfect quality* images as original references, objective quality metrics may be classified into three types: 1) *full-reference* (FR) IQA, where reference images are fully accessible on assessing distorted images; 2) *reduced-reference* (RR) IQA, where only partial information or some extracted features in reference images are permitted to help IQA; 3) *no-reference* (NR) IQA, where the access to reference images is wholy forbidden. The earliest FR metrics are perhaps the mean-squared error (MSE) and its equivalent peak signal-to-noise ratio (PSNR), which measure the energy preservation of a distorted image relative to the corresponding clear one. These two metrics have three remarkable merits of low complexity, high portability and clear physical meaning. However they were found to not always correlate well with human judgements of image quality, i.e. mean opinion score (MOS), which leads to a large body of IQA metrics [11].

Existing quality metrics (especially FR-IQA approaches) adopted a common two-stage structure. The first stage targets to estimate the local distortions, which most methods were mainly proposed for. For example, structural similarity index (SSIM) [12] was developed in light of luminance, contrast and structural similarities of the original and distorted images. Visual information fidelity (VIF) [13] is defined as the ratio of the mutual information between the reference and distorted images to the information content of the original image itself. Most apparent distortion (MAD) employs the detection- and appearance-based model for quality prediction [14]. Feature similarity index (FSIM) [15] and gradient similarity measure (GSM) [16] consider the fact that the human visual system (HVS) mainly depends on low-level visual features, such as gradient magnitude, for perceiving image quality. Under the assumption that IQA is highly connected to brain theory and neuroscience, internal generative model (IGM) [17] first partitions an image into the orderly (predicted) and disorderly (unpredicted) regions, and then evaluates them respectively via modified PSNR and SSIM on the above two areas with psychophysical parameters derived from [18].

The second (pooling) stage is to convert the local distortion map into a single visual quality score. Except the traditional average pooling, the multi-scale (MS) model is an alternative pooling strategy. Based on a reasonable assumption that the perceived quality of an image substantially depends upon the

This work was supported in part by the National Science Foundation of China under Grants 61422112, 61371146, 61221001, the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339.

K. Gu and W. Lin are with School of Computer Engineering, Nanyang Technological University, Singapore, 639798 (email: gukesjtuee@gmail.com; wslin@ntu.edu.sg).

S. Wang is with the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China. (email: sqwang1986@gmail.com).

G. Zhai, X. Yang, and W. Zhang are with Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, 200240, China (email: zhaiguangtao/xkyang/zhangwenjun@sjtu.edu.cn).

scale at which the image is analyzed, multi-scale SSIM (MS-SSIM) [18] computes SSIM in each scale level and combines with different weights obtained from a psychophysical test. A limitation of MS-SSIM lies in that it only considers the effect of scale while ignores the influence of visual saliency that usually has a high impact on the HVS when evaluating quality. From this point of view, W-SSIM [19] was yielded by weighting SSIM with a saliency map obtained from an eye-tracking experiment. However, W-SSIM heavily relies on subjective data and just takes the saliency map of the original image into account and thereby is impractical and ineffective for most scenarios. To tackle the dilemma, S_NW-SSIM [20] weights SSIM with a saliency map formed by properly using saliency features from both original and distorted images via the benchmark saliency detection model [21]. Similarly, PF-SSIM [22] combines SSIM with the percentile-and-fixation strategy towards high performance.

Unfortunately, the pooling strategies stated above brought limited performance gain to IQA metrics. Recently, another pooling technique, stemming from the statistical information theory and the NSS model, has introduced the high-accuracy information content weighted SSIM (IW-SSIM) [23]. It is interesting and intriguing to note that the root of IW-SSIM is originated from the VIF technique [13]; that is to improve the IQA performance with IQA methods. With this concern, the authors in [24] designed the structural similarity weighted SSIM (SW-SSIM) by using SSIM to estimate the weights in a block-based manner to advance SSIM itself. These two models, however, just exploit image content but overlook the impact of distortion types and intensities.

Despite the existence of numerous pooling strategies, they all bypassed the joint effects of image content and distortion, for instance, as displayed in Fig. 2 (a), the texture masking effect to noise on the stone wall compared to that on the red door, and the luminance masking effect to noise on the bright stone compared to that on the dark stone. To solve the issue, we devise a new pooling model via the analysis of distortion distribution decided by image content and introduced degradation. Particularly, four models are applied here: 1) the MS model for measuring the distribution of distortion position; 2) the ranking-based weighting (RW) model quantifying the distribution of distortion intensity; 3) the frequency variation-induced adjuster (FVA) gauging the distribution of frequency change; 4) the entropy gain multiplier (EGM) to estimate the distribution of histogram alteration.

The structure of the rest of this paper is arranged below. Section II firstly illustrates the distortion distribution based pooling model. In Section III, a comparison of our approach with existing pooling strategies and popular IQA metrics is conducted on four large-scale image quality databases, which include LIVE [25], TID2008 [26], CSIQ [14] and CCID2014 [27]. The paper is concluded in Section IV.

II. Methodology

In this section, we first briefly review the popular SSIM that has a good ability to detect local distortions. Then how to infer an overall image quality score by taking advantage of MS, RW, FVA and EGM models for pooling is described. We lastly sum up the analysis of distortion distribution-based (ADD) IQA algorithm using the proposed pooling scheme. A primary flowchart of our approach is illustrated in Fig. 1 for readers' conveniences.

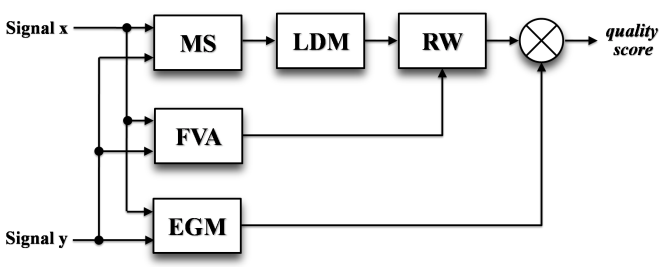


Fig. 1: A basic flowchart of our pooling model. Signal **x**: original image; Signal **y**: distorted image; MS: multi-scale; LDM: local distortion measurement; RW: ranking-based weighting; FVA: frequency variation-induced adjuster; EGM: entropy gain multiplier.

A. Local Distortion Measurement

The most popular IQA method is possibly SSIM which has been embedded into a wide scope of applications. Given the original image signal **x** and its distorted one **y**, the SSIM score is computed by

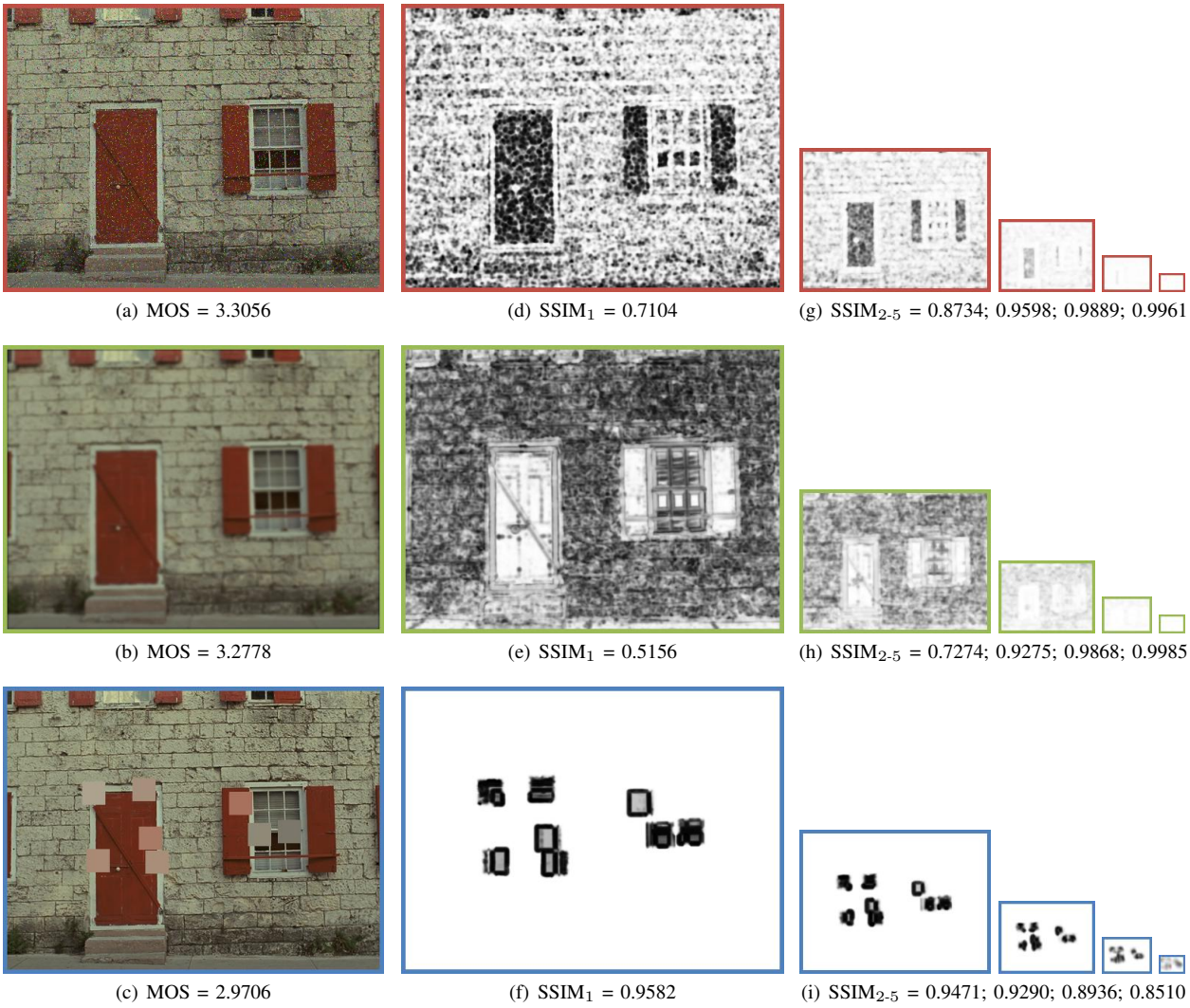
$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M} l(x_i,y_i)\cdot c(x_i,y_i)\cdot s(x_i,y_i) \quad (1)$$

where l, c and s respectively indicate local luminance, contrast and structural similarities; M is the number of local windows in the image. More details can refer to [12].

B. Multi-Scale Model

To provide a straightforward observation, we present three images from the TID2008 database in Figs. 2(a)-(c). The images are of the same original image "stone building" and similar MOS values. Next, we illustrate the distortion maps between the original image and each of distorted ones using SSIM in Figs. 2(d)-(f) and corresponding SSIM scores. In comparison, the three images, though of similar subjective scores, show remarkably distinct SSIM values, indicating the average pooling used in SSIM is not good. We view that the image corrupted by "local block-wise distortions of different intensity" in Fig. 2(c) has a much higher objective prediction relative to other twos, which is mainly caused by the various distributions of distortion position. In this case, the artifacts located in centralized positions in the SSIM distortion map exert a large amount of visual quality degradation, since those artifacts may destroy the understanding of the image content, just as we hardly recognize who it is from a portrait picture in which we dig a big hole on the position of human face.

In fact, this problem can be addressed by the multi-scale mechanism [28], with which the quality of the input image is perceived at different scales before combined with different weights to judge the overall quality score. Hence, we in this work consider using the classical MS model [18]. To specify,



(a) MOS = 3.3056 (d) $SSIM_1$ = 0.7104 (g) $SSIM_{2\text{-}5}$ = 0.8734; 0.9598; 0.9889; 0.9961

(b) MOS = 3.2778 (e) $SSIM_1$ = 0.5156 (h) $SSIM_{2\text{-}5}$ = 0.7274; 0.9275; 0.9868; 0.9985

(c) MOS = 2.9706 (f) $SSIM_1$ = 0.9582 (i) $SSIM_{2\text{-}5}$ = 0.9471; 0.9290; 0.8936; 0.8510

Fig. 2: An example of the natural image "stone building" and associated distorted versions: (a) Impulse noise; (b) Gaussian blur; (c) Local block-wise distortions of different intensity; (d)-(f) Distortion maps and corresponding quality scores of "SSIM" for (a)-(c) at Scale 1; (g)-(i) Distortion maps and corresponding quality scores of "SSIM" for (a)-(c) at Scale 2-5.

as exhibited in Fig. 3, we input the original and distorted images into a system, which iteratively uses a low-pass filter and downsamples the filtered image by a factor of 2. The original image is denoted as Scale 1, and the highest scale as Scale N, which is obtained after $N-1$ iterations. At the j-th scale, the contrast comparison, $c(\mathbf{x},\mathbf{y})$, and the structure comparison, $s(\mathbf{x},\mathbf{y})$, are calculated and denoted as $c_j(\mathbf{x},\mathbf{y})$ and $s_j(\mathbf{x},\mathbf{y})$, respectively. The luminance comparison, $l(\mathbf{x},\mathbf{y})$, is computed only at Scale N and is denoted as $l_N(\mathbf{x},\mathbf{y})$. The overall quality estimation is obtained by integrating the measurements at different scales using

$$S(\mathbf{x},\mathbf{y}) = [l_N(\mathbf{x},\mathbf{y})]^{\alpha_N} \prod_{j=1}^{N} [c_j(\mathbf{x},\mathbf{y})]^{\beta_j} [s_j(\mathbf{x},\mathbf{y})]^{\gamma_j} \qquad (2)$$

where model parameters are assigned as $\alpha_j = \beta_j = \gamma_j$ with $\sum_{j=1}^{N} \gamma_j = 1$ for all j's to control the relative importance of each component. By a psychophysical test [18], $\{\gamma_1, ..., \gamma_5\}$ are set to be $\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$.

We show some results in Fig. 2 to account for how the MS

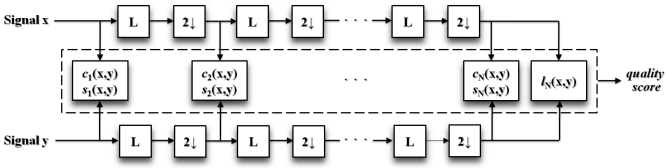


Fig. 3: Multi-scale structural similarity measurement [18]. L: low-pass filtering; 2 ↓: downsampling by 2.

model discriminates distinct distributions of distortion position well; that is how to effectively pool various types of artifacts located in centralized or sparse positions. We see that, despite quite similar subjective ratings for Figs. 2(a)-(c), at the Scale 1 the SSIM score in (f) is noticeably higher than those in (d)-(e). With the scale level increased, however, the SSIM scores for "impulse noise" and "Gaussian blur" become higher, as provided in Figs. 2(g)-(h), since at the large scale the artifacts of sparse distribution will be filtered and smoothed to a great extent. On the contrary, the SSIM score for "local block-wise

distortions of different intensity" become smaller, as shown in Fig. 2(i), since at the large scale the artifacts of centralized distribution will be preserved or even highlighted. Weighting SSIM scores at each scale with the aforesaid psychophysical numbers, we can get the closer scores of 0.9334, 0.8638 and 0.9173 for Figs. 2(a)-(c), and this confirms the effectiveness of the MS model for characterizing different distributions of distortion position.

C. Ranking-Based Weighting

Besides the influence of distributions of distortion position, we find from Figs. 2(b)-(c) that the distortions with intensity distributed non-uniformly (e.g. impulse noise) usually makes a larger punishment to visual quality than those with intensity distributed uniformly (e.g. Gaussian blur). From the viewpoint of low-level vision, this phenomenon is majorly because, as compared to the uniformly distributed distortion intensity, the non-uniformly distributed distortion intensity easily generates higher local contrast and attracts much more attention, and thus degrades the visual quality to a large degree. Hence we insert a simple RW technique into the MS model to highlight the highly distorted image pixels. Note that, by distinguishing various types of artifacts with similar position distributions but different intensity distributions, the RW model is able to improve the MS model. We first rewrite Eq. (2) to be

$$\begin{aligned} S(\mathbf{x},\mathbf{y}) &= [l_N(\mathbf{x},\mathbf{y})c_N(\mathbf{x},\mathbf{y})s_N(\mathbf{x},\mathbf{y})]^{\gamma_N}\prod_{j=1}^{N-1}[c_j(\mathbf{x},\mathbf{y})s_j(\mathbf{x},\mathbf{y})]^{\gamma_j} \\ &= \prod_{j=1}^{N}[t_j(\mathbf{x},\mathbf{y})]^{\gamma_j}. \end{aligned} \tag{3}$$

We then use RW to modify each of t_j $(j = 1, ..., 5)$ as

$$t'_j(\mathbf{x},\mathbf{y}) = \frac{\sum_{u\in\Omega_1} w t_j^r(x_u,y_u) + \sum_{v\in\Omega_2} t_j(x_v,y_v)}{\sum_{u\in\Omega_1} w + \sum_{v\in\Omega_2} 1} \tag{4}$$

where constant parameters w and r are applied to stress the significance of lowest values in t_j (i.e. the most distorted image pixels). Ω_1 includes the smallest $k\%$ values while Ω_2 consists of the rest ones.

Combining MS and RW techniques to advance SSIM, we can get the scores of 0.9923, 0.9816 and 0.9792 for Figs. 2(a)-(c), which indicates the effectiveness of the RW model for discriminating distinct distributions of distortion intensity and the integration of the above two techniques for quality prediction.

D. Frequency Variation-Induced Adjuster

With the adoption of MS and RW models, we also take the visual masking into account; for example, the bright plane regions (the luminance masking) or the texture and edge areas (the texture and edge masking) exert larger masking effects on noise compared to blurriness. In fact, from the perspective of frequency domain, the classical contrast sensitive function (CSF) [29] reveals that the HVS has a stronger depress on the low-frequency domain than the middle-frequency one, but weaker depress than the high-frequency one. The FVA based on the frequency variation is inspired by this.

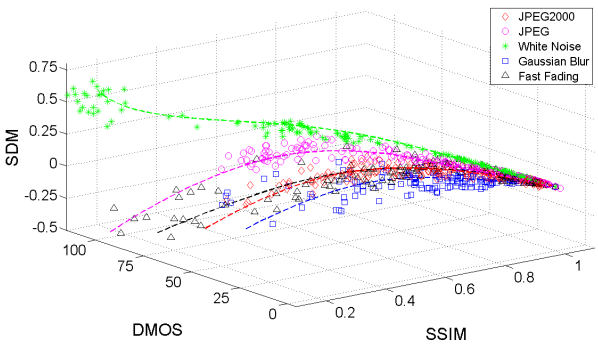


Fig. 4: Scatter plots of SDM vs. SSIM and DMOS on LIVE for five distortion types and associated nonlinear fitting curves (dash lines).

Firstly, we use the difference of the original and distorted images in terms of the structural degradation measurement (SDM), which reflects the similarity of an input image signal and its associated low-pass filtered version, to quantify this frequency variation-induced perceptual difference [30]:

$$\text{SDM}(\mathbf{x},\mathbf{y}) = \text{SSIM}(\mathbf{x},\mathbf{x}_f) - \text{SSIM}(\mathbf{y},\mathbf{y}_f) \tag{5}$$

where $\mathbf{x}_f$ and $\mathbf{y}_f$ are generated applying a low-pass filter to $\mathbf{x}$ and $\mathbf{y}$. In fact, this simple model has been widely employed with the Gaussian kernel or the autoregressive (AR) model for IQA designs [17, 31]. In this implementation, the commonly used and efficient Gaussian kernel is adopted as the low-pass filter. In Fig. 4, we show the scatter plots of SDM versus SSIM and DMOS on the whole LIVE database [25] and five fitted curves for various distortion types, which confirms the good ability of SDM for distinguishing different perceptions induced by frequency variation. Second, we use the FVA to improve the parameter r in Eq. (4) as follows:

$$\begin{aligned} r' &= r + \text{FVA}(\mathbf{x},\mathbf{y}) \\ &= r + h(\text{SDM}(\mathbf{x},\mathbf{y}),\epsilon,\theta) \\ &= r + h(\text{SSIM}(\mathbf{x},\mathbf{x}_f) - \text{SSIM}(\mathbf{y},\mathbf{y}_f),\epsilon,\theta) \end{aligned} \tag{6}$$

where ϵ and θ are fixed parameters with empirical values of 200 and 1e-3. $h(z,a,b) = \frac{\text{sign}(z)\cdot|z|^b}{a}$ is used to adjust the magnitude of z, where $\text{sign}(\cdot)$ obtains the sign of the input variable.

E. Entropy Gain Multiplier

We also find that many typical distortion types, e.g. noise and blur, seldom alter the histogram distribution, while some other distortion types, e.g. contrast adjustment, might reshape the histogram to a large extent, even enhancing the quality of original images. Entropy, Kullback-Leibler (KL) divergence, and Jensen-Shannon (JS) divergence [32] are good candidates. Because KL and JS divergences do not contribute remarkable performance gain but introduce much computational cost, in this paper we therefore choose entropy. We further consider downsampling the original and distorted images to a small

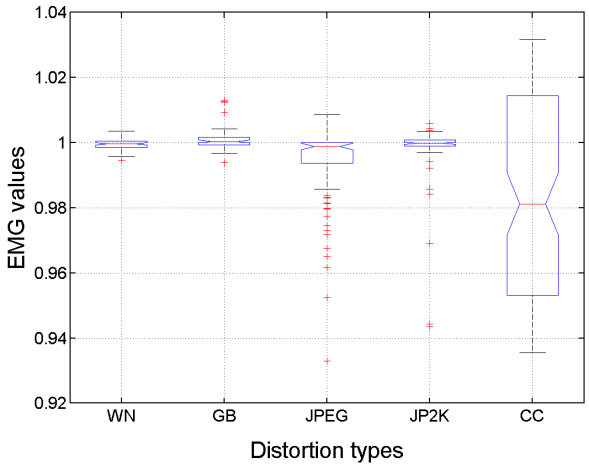


Fig. 5: Box plot of EGM distributions across white noise (WN), Gaussian blur (GB), JPEG and JPEG2000 (JP2K) compressions, and contrast change (CC) in the TID2008 database.

size first for decreasing the implementation time. The EGM multiplier is consequently defined by

$$\mathrm{EGM}(\mathbf{x},\mathbf{y}) = \left(\frac{H(\mathbf{x}_d)+C_4}{H(\mathbf{y}_d)+C_4}\right)^{\phi} \tag{7}$$

where $H(\cdot)$ estimates the entropy of the image. $\mathbf{x}_d$ and $\mathbf{y}_d$ are downsampled images of $\mathbf{x}$ and $\mathbf{y}$. C_4 is a positive constant. ϕ together with C_4 is used to adjust the magnitude of EGM and the relative importance among EGM and other three models of MS, RW and FVA. We empirically assign $\phi = 0.05$ and $C_4 = 9$.

To show the effect of EGM, we compute its values on the image from distinct distortion types in the TID2008 database. Those types consist of white noise, Gaussian blur, JPEG and JPEG2000 compressions, and contrast change. As presented in Fig. 5, the EGM values are around one for the first four distortion types, and thereby are hardly influenced by them. Conversely, the EGM is very sensitive to the contrast change, which can play a supplementary role for the aforementioned MS, RW and FVA models.

F. Proposed Pooling Scheme

Given the original image $\mathbf{x}$ and its distorted version $\mathbf{y}$, we finally show the complete form of the analysis of distortion distribution-based (ADD) ADD-SSIM, in order to make the proposed pooling model clearer:

$$\text{ADD-SSIM}(\mathbf{x},\mathbf{y}) = \mathrm{EGM}(\mathbf{x},\mathbf{y}) \prod_{j=1}^{N} [t'_j(\mathbf{x},\mathbf{y})]^{\gamma_j}. \tag{8}$$

For a cross-validation, another basic IQA model, the gradient similarity index (GSIM)[1] established upon the hypothesis that the HVS understands an image strongly relying on low-level visual features, is used to define ADD-GSIM with entirely the same parameter settings as used in ADD-SSIM.

[1]More details about GSIM can be found in the Appendix A.

III. Experimental Results and Analysis

A. Testing Metrics and Databases

In this section, we will validate the designed ADD pooling model and compare with classical and state-of-the-art metrics, and relevant pooling scheme based IQA models: 1) two classical FR IQA techniques, which include SSIM [12] and GSIM [15]; 2) two state-of-the-art FR IQA metrics, consisting of GSM [16] and IGM [17]; 3) seven pooling models: W-SSIM/GSIM [33], $\mathrm{S_N}$W-SSIM/GSIM [20], IW-SSIM/GSIM [23], SW-SSIM/GSIM [24], F-SSIM/GSIM [15], spectral residual (SR) SR-SSIM/GSIM [34], and standard deviation (SD) SD-SSIM/GSIM [35][2].

Four large-scale databases, including the most three popular LIVE, TID2008, CSIQ and one newly proposed CCID2014, are used as testing beds[3]. The first LIVE database [25] was developed at the University of Texas at Austin. It contains five image data sets, and totally 779 distorted images from 29 sources. The distortion types include JPEG2000 compression, JPEG compression, white noise, Gaussian blur, and fast fading channel distortion of JPEG2000 compressed bitstream. The subjective test for each data set was separately carried out. A cross-comparison set mixing images from all distortion types is then used to help align the subject scores across data sets, and thus adjust the subjective scores of all images.

The second TID2008 database [26] was developed with a joint international effort across Finland, Italy and Ukraine. It is composed of 1,700 distorted images produced from 25 references with 17 distortion types at four degradation levels. These types are: a) additive Gaussian noise; b) additive noise in color components is more intensive than additive noise in the luminance component; c) spatially correlated noise; d) masked noise; e) high frequency noise; f) impulse noise; g) quantization noise; h) Gaussian blur; i) image denoising; j) JPEG compression; k) JPEG2000 compression; l) JPEG transmission errors; m) JPEG2000 transmission errors; n) non eccentricity pattern noise; o) local block-wise distortions of different intensity; p) mean shift (i.e. intensity shift); and q) contrast change.

The third CSIQ database [14] was built at the Oklahoma State University. It consists of 866 images created from 30 pristine images with six types at four to five intensities of distortions. The six distortion types are JPEG compression, JPEG2000 compression, Gaussian blurring, global contrast decrements, additive white Gaussian noise, and additive pink Gaussian noise.

The fourth CCID2014 database [27] was recently completed at Shanghai Jiaotong University. It includes 655 images via eight kinds of transfer mappings to process 15 source ones. The eight types are composed of negative and positive gamma transfers, convex and concave arcs, cubic and logistic functions, mean shifting, and compound functions with mean shifting before logistic transfer.

[2]W-, $\mathrm{S_N}$W-, IW- and SW-GSIM were generated by separately using W, $\mathrm{S_N}$W, IW and SW models on GSIM. F-GSIM, SR-GSIM and SD-GSIM are FSIM, SR-SIM and GMSD built upon the GSIM. Using the same strategy on SSIM, we yield F-SSIM, SR-SSIM and SD-SSIM.

[3]Interested readers can be directed to [11] for more information about subjective image quality databases.

TABLE I: Correlation performance measures on four databases. We bold the best performed method in each group.

LIVE Database (779 images) [25]				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.9104	0.7311	0.9042	11.669
W-SSIM [33]	0.9556	0.8123	0.9512	8.4278
S_NW-SSIM [20]	0.9519	0.8027	0.9484	8.6671
IW-SSIM [23]	0.9567	0.8175	0.9522	8.3468
SW-SSIM [24]	0.9610	0.8241	0.9559	8.0204
F-SSIM [15]	0.9573	0.8165	0.9531	8.2703
SR-SSIM [34]	0.9589	0.8207	0.9533	8.2488
SD-SSIM [35]	0.9289	0.7616	0.9191	10.763
ADD-SSIM (Pro)	**0.9646**	**0.8358**	**0.9587**	**7.7744**
GSIM [15]	0.9206	0.7509	0.9148	11.039
W-GSIM [33]	0.9618	0.8283	0.9590	7.7395
S_NW-SSIM [20]	0.9590	0.8213	0.9564	7.9821
IW-GSIM [23]	0.9520	0.8083	0.9469	8.7825
SW-GSIM [24]	0.9660	0.8382	0.9632	7.3476
F-GSIM [15]	0.9634	0.8335	0.9597	7.6739
SR-GSIM [34]	0.9619	0.8301	0.9555	8.0633
SD-GSIM [35]	0.9603	0.8268	0.9603	7.6214
ADD-GSIM (Pro)	**0.9681**	**0.8474**	**0.9657**	**7.0924**
GSM [16]	0.9561	0.8150	0.9512	8.4323
IGM [17]	**0.9581**	**0.8250**	**0.9570**	**7.9242**

TID2008 Database (1,700 images) [26]				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.6272	0.4562	0.6413	1.0297
W-SSIM [33]	0.8193	0.6257	0.8236	0.7612
S_NW-SSIM [20]	0.8085	0.6146	0.8031	0.7995
IW-SSIM [23]	0.8559	0.6636	0.8579	0.6895
SW-SSIM [24]	0.7907	0.5986	0.8092	0.7884
F-SSIM [15]	0.7935	0.6002	0.8019	0.8017
SR-SSIM [34]	0.8239	0.6343	0.8292	0.7501
SD-SSIM [35]	0.7599	0.5725	0.7407	0.9016
ADD-SSIM (Pro)	**0.8805**	**0.6994**	**0.8860**	**0.6222**
GSIM [15]	0.6952	0.5138	0.7097	0.9454
W-GSIM [33]	0.8827	0.6985	0.8736	0.6530
S_NW-SSIM [20]	0.8741	0.6851	0.8521	0.7023
IW-GSIM [23]	0.8506	0.6600	0.8501	0.7066
SW-GSIM [24]	0.8730	0.6883	0.8738	0.6526
F-GSIM [15]	0.8804	0.6945	0.8738	0.6526
SR-GSIM [34]	0.8913	0.7149	0.8867	0.6205
SD-GSIM [35]	0.8907	0.7092	0.8789	0.6402
ADD-GSIM (Pro)	**0.9094**	**0.7389**	**0.9120**	**0.5504**
GSM [16]	0.8504	0.6596	0.8422	0.7234
IGM [17]	**0.8901**	**0.7103**	**0.8857**	**0.6231**

CSIQ Database (866 images) [14]				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.8378	0.6343	0.8154	0.1520
W-SSIM [33]	0.8972	0.7214	0.8888	0.1203
S_NW-SSIM [20]	0.8892	0.7120	0.8722	0.1284
IW-SSIM [23]	0.9213	0.7529	0.9144	0.1063
SW-SSIM [24]	0.8965	0.7199	0.8905	0.1195
F-SSIM [15]	0.8735	0.6889	0.8642	0.1321
SR-SSIM [34]	0.8834	0.7065	0.8584	0.1347
SD-SSIM [35]	0.8163	0.6424	0.8074	0.1549
ADD-SSIM (Pro)	**0.9330**	**0.7697**	**0.9311**	**0.0958**
GSIM [15]	0.8696	0.6714	0.8462	0.1399
W-GSIM [33]	0.9311	0.7707	0.9165	0.1050
S_NW-SSIM [20]	0.9359	0.7772	0.9239	0.1005
IW-GSIM [23]	0.9090	0.7388	0.8963	0.1164
SW-GSIM [24]	0.9298	0.7689	0.9071	0.1105
F-GSIM [15]	0.9223	0.7539	0.9026	0.1130
SR-GSIM [34]	0.9286	0.7653	0.8854	0.1220
SD-GSIM [35]	**0.9570**	**0.8129**	**0.9541**	**0.0786**
ADD-GSIM (Pro)	0.9422	0.7894	0.9342	0.0937
GSM [16]	0.9108	0.7374	0.8964	0.1163
IGM [17]	**0.9403**	**0.7881**	**0.9281**	**0.0977**

CCID2014 Database (655 images) [27]				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.8136	0.6063	0.8227	0.3717
W-SSIM [33]	0.8239	0.6197	0.8341	0.3607
S_NW-SSIM [20]	0.8013	0.5988	0.8004	0.3920
IW-SSIM [23]	0.7811	0.5898	0.8342	0.3606
SW-SSIM [24]	0.8260	0.6206	0.8414	0.3533
F-SSIM [15]	0.8236	0.6181	0.8375	0.3573
SR-SSIM [34]	0.8184	0.6126	0.8357	0.3591
SD-SSIM [35]	0.7691	0.5550	0.7705	0.4168
ADD-SSIM (Pro)	**0.8767**	**0.6924**	**0.8980**	**0.2878**
GSIM [15]	0.7045	0.5151	0.7695	0.4176
W-GSIM [33]	0.7137	0.5216	0.7841	0.4058
S_NW-SSIM [20]	0.7126	0.5086	0.7472	0.4346
IW-GSIM [23]	0.7231	0.5345	0.7988	0.3934
SW-GSIM [24]	0.7267	0.5356	0.7972	0.3947
F-GSIM [15]	0.7655	0.5704	0.8202	0.3741
SR-GSIM [34]	0.7360	0.5370	0.7833	0.4065
SD-GSIM [35]	0.8077	0.6100	0.8521	0.3422
ADD-GSIM (Pro)	**0.8698**	**0.6840**	**0.8935**	**0.2936**
GSM [16]	**0.7768**	**0.5711**	**0.8073**	**0.3859**
IGM [17]	0.7245	0.5355	0.7992	0.3930

B. Evaluation Protocols

In order to remove the nonlinearity, the objective prediction scores are mapped to subjective human ratings through the nonlinear regression. To specify, on each database, this paper first applies the five-parameter logistic function to preprocess each IQA metric:

$$q(\xi) = \pi_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\pi_2(\xi - \pi_3)}} \right) + \pi_4 \xi + \pi_5 \qquad (9)$$

where ξ and $q(\xi)$ respectively indicate the input and mapped scores, and π_j $(j = 1, 2, 3, 4, 5)$ are five free parameters to be determined during the curve fitting process. We then apply four frequently used performance measures, as suggested by the video quality experts group (VQEG) [36], for evaluating and comparing the proposed algorithm with the IQA models tested. The first and second performance measures are the Spearman rank-order correlation coefficient (SRCC) and the Kendall's rank-order correlation coefficient (KRCC) that are significant non-parametric rank correlation metrics. The third one is the Pearson linear correlation coefficient (PLCC) of subjective human ratings and the converted objective scores for measuring the prediction accuracy. And the last performance index is the root mean-squared error (RMSE) which is quantified between subjective and objective quality scores after the nonlinear regression of Eq. (9), in order to measure the prediction consistency. In the aforesaid four performance evaluations, a value close to 1 for SRCC, KRCC and PLCC, yet close to 0 for RMSE represents superior correlation in accordance with subjective ratings.

C. Performance Measures

We report the correlation performance measures of testing IQA metrics on four databases in Table I. We bold the top two metrics in each group. For a comprehensive comparison, in Table II, we also list two averaged performance evaluations defined by

$$\bar{\delta} = \frac{\sum_i \delta_i \cdot \omega_i}{\sum_i \omega_i} \qquad (10)$$

TABLE II: Average correlation performance indices on four databases. We highlight the top method in each group.

Direct Average				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.7972	0.6070	0.7959	3.3057
W-SSIM [33]	0.8740	0.6948	0.8744	2.4175
S_NW-SSIM [20]	0.8627	0.6820	0.8560	2.4968
IW-SSIM [23]	0.8787	0.7060	0.8897	2.3758
SW-SSIM [24]	0.8685	0.6908	0.8743	2.3204
F-SSIM [15]	0.8620	0.6810	0.8642	2.3904
SR-SSIM [34]	0.8712	0.6935	0.8692	2.3732
SD-SSIM [35]	0.8185	0.6329	0.8094	3.0590
ADD-SSIM (Pro)	**0.9137**	**0.7493**	**0.9184**	**2.1950**
GSIM [15]	0.7975	0.6128	0.8100	3.1354
W-GSIM [33]	0.8723	0.7048	0.8833	2.2258
S_NW-SSIM [20]	0.8704	0.6980	0.8699	2.3049
IW-GSIM [23]	0.8587	0.6854	0.8730	2.4997
SW-GSIM [24]	0.8739	0.7078	0.8853	2.1264
F-GSIM [15]	0.8829	0.7131	0.8891	2.2034
SR-GSIM [34]	0.8795	0.7118	0.8777	2.3031
SD-GSIM [35]	0.9039	0.7397	0.9114	2.1706
ADD-GSIM (Pro)	**0.9224**	**0.7649**	**0.9263**	**2.0075**
GSM [16]	0.8735	0.6958	0.8743	2.4145
IGM [17]	**0.8783**	**0.7147**	**0.8925**	**2.2595**

Database Size-Weighted Average				
Metrics	SRCC	KRCC	PLCC	RMSE
SSIM [12]	0.7585	0.5729	0.7599	2.8040
W-SSIM [33]	0.8635	0.6818	0.8643	2.0499
S_NW-SSIM [20]	0.8527	0.6697	0.8459	2.1197
IW-SSIM [23]	0.8774	0.7008	0.8846	2.0006
SW-SSIM [24]	0.8526	0.6724	0.8607	1.9808
F-SSIM [15]	0.8477	0.6645	0.8507	2.0385
SR-SSIM [34]	0.8622	0.6827	0.8608	2.0132
SD-SSIM [35]	0.8065	0.6216	0.7948	2.5810
ADD-SSIM (Pro)	**0.9076**	**0.7400**	**0.9119**	**1.8464**
GSIM [15]	0.7784	0.5943	0.7890	2.6503
W-GSIM [33]	0.8809	0.7105	0.8849	1.8740
S_NW-SSIM [20]	0.8776	0.7027	0.8708	1.9459
IW-GSIM [23]	0.8621	0.6854	0.8706	2.1003
SW-GSIM [24]	0.8794	0.7099	0.8859	1.7969
F-GSIM [15]	0.8868	0.7141	0.8880	1.8576
SR-GSIM [34]	0.8877	0.7191	0.8829	1.9270
SD-GSIM [35]	0.9050	0.7383	0.9066	1.8294
ADD-GSIM (Pro)	**0.9214**	**0.7620**	**0.9242**	**1.6835**
GSM [16]	0.8720	0.6922	0.8695	2.0380
IGM [17]	**0.8871**	**0.7209**	**0.8946**	**1.8936**

where δ_i $(i = 1, 2, 3, 4)$ indicates the performance index for each database. The first direct average is calculated by setting all of ω_i as one, while ω_i are assigned as the number of images in each database, i.e. 779 for LIVE, 1,700 for TID2008, 866 for CSIQ, and 655 for CCID2014, for the computation of the second database size-weighted average. From Tables I and II, we have derived three important conclusions.

First, popular IQA models have been advanced consistently with the proposed pooling strategy on each testing database, especially on the TID2008 database. As bolded in the above-mentioned two tables, our approach has also obtained higher performance than other testing IQA measures on four image databases, except that our ADD-GSIM is a little less than SD-GSIM on the CSIQ database. In terms of two average results, our scheme entirely and constantly defeats the second-place pooling model, with the performance gain of around 0.025 on SSIM and 0.015 on GSIM for SRCC and PLCC.

Second, it is found that the modified SSIM and/or GSIM using the proposed ADD model are/is even superior to existing classical and state-of-the-art IQA algorithms, particularly on TID2008 and CCID2014 databases as provided in Table I, and two average performance evaluations as listed in Table II. Note that for the newly developed CCID2014 database, there are only our ADD-SSIM/GSIM approach having acquired fairly good results of over 0.85 for both SRCC and PLCC.

Third, the devised pooling model is shown to be constantly valid as compared to other related technologies, which have, more or less, limitations for the pooling design. To be more concretely, the popular IW strategy shows remarkable performance gains for SSIM but not for GSIM, state-of-the-art F, SR and SD schemes work effectively for GSIM only, and the lately designed SW model is just extremely good at improving SSIM and GSIM on the LIVE database.

In order to provide a straightforward comparison, the ratio of the performance gain of each pooling based SSIM/GSIM metric relative to the original SSIM/GSIM in terms of SRCC is computed as follows:

$$R_{xy} = \frac{\text{SRCC}_y - \text{SRCC}_x}{\text{SRCC}_x} \times 100\% \tag{11}$$

where SRCC_x means the SRCC value of SSIM, and SRCC_y means the SRCC value of each testing pooling based SSIM metric. For GSIM, SRCC_x and SRCC_y are similarly defined to be the SRCC values of GSIM and the associated pooling model. We show the ratios on four databases and the direct and database size-weighted averages in Table III. The model on each database with the highest performance improvement is emphasized with boldface. First, the performance gain of the proposed pooling model in most cases has obtained much better results than those of competing quality methods. Next, the gain ratio of our ADD-SSIM and ADD-GSIM techniques compared to the original SSIM and GSIM metrics are respectively about 6% and 5% on LIVE, 40% and 30% on TID2008, 11% and 8% on CSIQ, 7.5% and 23% on CCID2014, 14% and 15% on the direct average, as well as 19% and 18% on the database size-weighted average. It deserves much attention that, relative to SSIM and GSIM, the performance gains of our ADD-SSIM and ADD-GSIM are up to 30% on the large-scale TID2008 database on one hand, and higher than 15% on the weighted average on the other hand.

D. Statistical Significance

Furthermore, we utilize the f-test to measure the statistical significance of the proposed pooling technique. The f-test compares the prediction residuals of each quality metric tested. Defining f as the ratio between two residual variances and f_c (determined by the number of residuals and the confidence level) be the judgement threshold. The difference of performance of each pair of testing IQA models is considered to be significant when $f > f_c$. The confidence is set to be 95% in this work. By experiment, we found that our ADD-SSIM and ADD-GSIM metrics perform very well. To be more specific, on four databases, our pooling technique based ADD-SSIM

TABLE III: The ratio of the performance gain of each pooling based SSIM/GSIM metric relative to the original SSIM/GSIM in terms of SRCC on four databases as well as the direct and database size-weighted means. We bold the metric with the highest performance gain.

Database	W-SSIM	S_NW-SSIM	IW-SSIM	SW-SSIM	F-SSIM	SR-SSIM	SD-SSIM	ADD-SSIM
LIVE [25]	4.966%	4.563%	5.082%	5.556%	5.154%	5.330%	2.029%	**5.954**%
TID2008 [26]	30.63%	28.90%	36.47%	26.08%	26.52%	31.36%	21.15%	**40.39**%
CSIQ [14]	7.081%	6.126%	9.961%	6.999%	4.252%	5.439%	-2.564%	**11.36**%
CCID2014 [27]	1.273%	-1.505%	-3.994%	1.528%	1.232%	0.598%	-5.468%	**7.760**%
Direct Average	9.627%	8.213%	10.22%	8.944%	8.119%	9.271%	2.671%	**14.64**%
Weighted Average	13.84%	12.43%	15.69%	12.41%	11.76%	13.67%	6.335%	**19.67**%

Database	W-GSIM	S_NW-GSIM	IW-GSIM	SW-GSIM	F-GSIM	SR-GSIM	SD-GSIM	ADD-GSIM
LIVE [25]	4.470%	4.164%	3.411%	4.932%	4.642%	4.482%	4.304%	**5.154**%
TID2008 [26]	26.96%	25.73%	22.34%	25.56%	26.63%	28.20%	28.12%	**30.81**%
CSIQ [14]	7.069%	7.629%	4.531%	6.927%	6.060%	6.789%	**10.05**%	8.348%
CCID2014 [27]	1.314%	1.155%	2.648%	3.155%	8.660%	4.468%	14.65%	**23.47**%
Direct Average	9.382%	9.144%	7.674%	9.579%	10.71%	10.28%	13.34%	**15.66**%
Weighted Average	13.17%	12.74%	10.75%	12.98%	13.93%	14.04%	16.27%	**18.38**%

TABLE IV: Comparison of mean computational time (in millisecond/image) and the time ratio of each pooling based SSIM/GSIM technique compared to the original SSIM/GSIM metric on the whole 1,700 images in the TID2008 database.

Metric	SSIM	W-SSIM	S_NW-SSIM	IW-SSIM	SW-SSIM	F-SSIM	SR-SSIM	SD-SSIM	ADD-SSIM
Time (ms)	34.597	16.232	643.75	298.56	8807.0	294.30	19.744	11.682	127.70
Ratio	1	0.47	18.6	8.63	255	8.51	0.57	0.34	3.69

Metric	GSIM	W-GSIM	S_NW-GSIM	IW-GSIM	SW-GSIM	F-GSIM	SR-GSIM	SD-GSIM	ADD-GSIM
Time (ms)	7.9334	7.1212	642.53	310.18	1207.3	286.86	13.745	5.295	63.813
Ratio	1	0.90	80.1	39.1	152	36.2	1.73	0.67	8.04

is statistical better than all testing pooling schemes except that it is statistical equivalent to the SW-SSIM on the LIVE database. Our ADD-GSIM metric is superior or comparable to the overall pooling models based GSIM methods on the popular LIVE database, while it outperforms most pooling models on TID2008, CSIQ and CID2014 databases but it is worse than the SD-GSIM (i.e. the state-of-the-art GMSD [35]) on the CSIQ database. All in all, our approach is proved to be a substantially good pooling strategy using the statistical significance comparison of f-test.

E. Computational Cost

A good quality measure should be simultaneously effective and efficient. Hence we compute the implementation time of each pooling model on all the 1,700 images in the TID2008 database. This experiment was conducted by MATLAB 7.10.0 (R2012a) on a computer with 3.40GHz CPU processor and 4.00GB memory. Table IV reports the average running time and the associated ratio of each pooling based SSIM/GSIM method compared to the original SSIM/GSIM method. With extensive tests, the proposed pooling model takes much less than one second for an image. Since the computation in each scale used in our model is independent of others, the speed-up parallel computing is possibly introduced to reduce the computational time to some extent.

F. Visualized Comparison

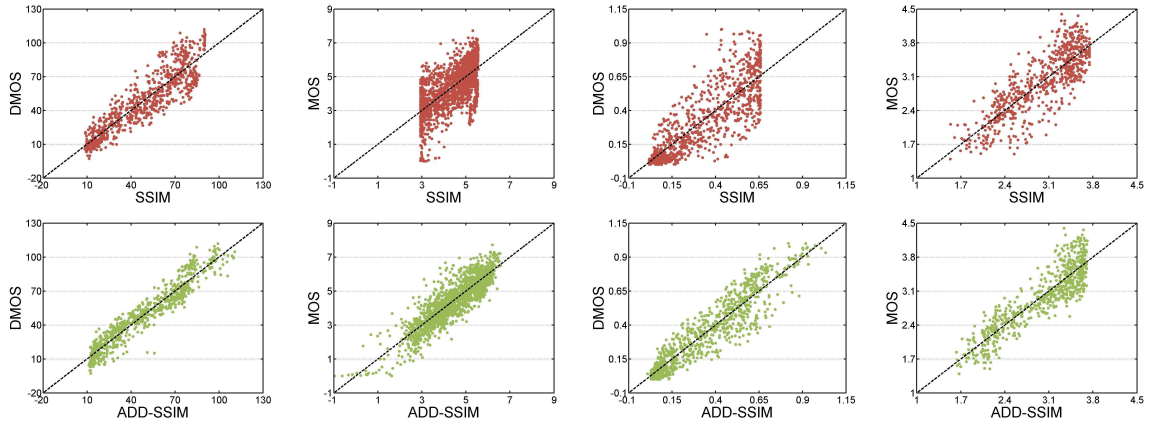
To give a visualized validation, we lastly exhibit the scatter plots of MOS/DMOS versus objective quality predictions of the proposed ADD-SSIM/GSIM (after the nonlinear mapping) on LIVE, TID2008, CSIQ and CCID2014 databases in Figs. 6-7. The original SSIM and GSIM models are used for comparison. Our method always gives reasonable quality scores, where the sample points tend to be clustered closer to the black diagonal lines (meaning perfect prediction) than testing metrics under comparison.

G. Discussions

In the future, we will mainly concentrate on the following three aspects. Besides the luminance information, the first one is to further promote the correlation performance by taking the chromatic information into account, since it conveys crucial semantic structure information [37]. Second, the joint effects of different distortion sources [38] will be employed to make the proposed model robust across multiply distorted image databases [39-40]. The last but not the least consideration lies in that the viewing distance-changed IQA based on adaptive technologies in the proper transform domains [41].

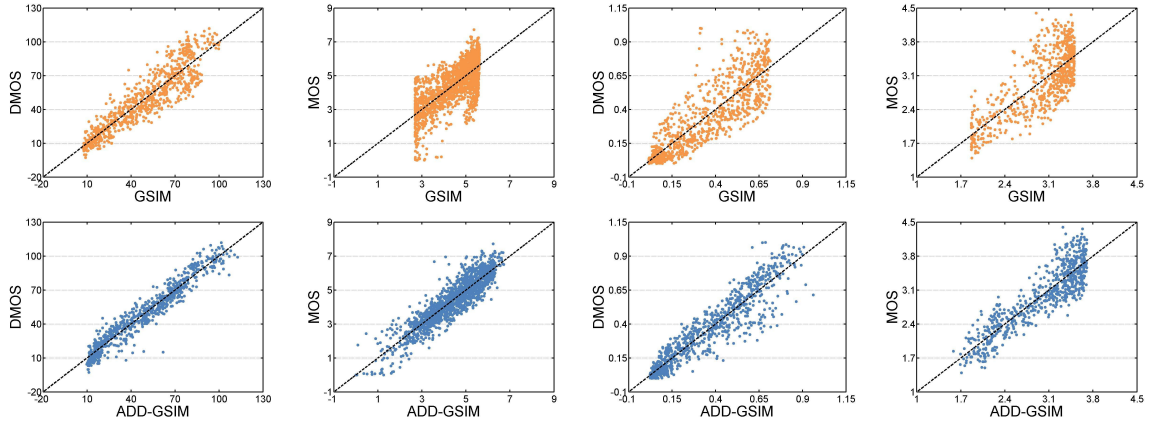
IV. Conclusion

Existing research effort has emphasized more on the local distortion measurement so far, and this research is devoted to the pooling stage which has been less investigated. With the comprehensive analysis, in this paper we have proposed a new pooling strategy with distortion distribution affected by visual content and image distortion. Our technique takes the distributions of distortion position, distortion intensity, frequency alterations, and histogram alterations into account and applies four models − multi-scale, ranking-based weighting, frequency variation-induced adjuster, and entropy gain multiplier − for pooling, to yield an overall quality score. We conduct the extensive experiments on four large-scale image databases (i.e. LIVE, TID2008, CSIQ and CCID2014) to evaluate the proposed quality measure and compare with classical and state-of-the-art IQA methods in terms of four typical correlation performance indices. Results show four important conclusions.



LIVE Database TID2008 Database CSIQ Database CCID2014 Database

Fig. 6: Scatter plots of MOS/DMOS vs. SSIM and ADD-SSIM (after the nonlinear mapping) on four databases.



LIVE Database TID2008 Database CSIQ Database CCID2014 Database

Fig. 7: Scatter plots of MOS/DMOS vs. GSIM and ADD-GSIM (after the nonlinear mapping) on four databases.

First, our proposed pooling scheme leads to consistent IQA performance improvement. Second, relative to the traditional average pooling, our strategy has obtained the performance gain of more than 15% on average. Third, the best overall performance made by the proposed model generally outperforms state-of-the-art competitors in numerical and statistical significance comparisons. Fourth, our pooling scheme shows high performance gains on the TID2008 database that involves various kinds of distortion types. Our source code will be released at https://sites.google.com/site/guke198701/publications.

V. Appendix A

The GSIM is another FR IQA measure which is developed based on the supposition that the HVS understands an image signal heavily depending on low-level visual features. When computing GSIM, we firstly compute the gradient magnitude (GM) for the original image signal $\mathbf{x}$ by

$$G_x = \sqrt{G_h^2 + G_v^2} \tag{12}$$

where

$$G_h = S_h \otimes \mathbf{x} = \frac{1}{16}\begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix} \tag{13}$$

$$G_v = S_v \otimes \mathbf{x} = \frac{1}{16}\begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ +3 & +10 & +3 \end{bmatrix} \tag{14}$$

where S_h and S_v are Sobel convolution masks in horizontal and vertical directions, and the symbol "$\otimes$" represents the convolution operation. Similarly, we denote by G_y the GM of the distorted image signal $\mathbf{y}$. Following the similarity measure that has three merits of symmetry, boundedness and unique maximum [12], GSIM is defined as

$$\text{GSIM}(\mathbf{x}, \mathbf{y}) = \frac{2G_x \cdot G_y + C_5}{G_x^2 + G_y^2 + C_5} \tag{15}$$

where C_5 is a constant with the similar function to C_4.

References

[1] M. J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379-3391, Sept. 2013.

[2] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu, "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Trans. Broadcasting*, vol. 61, no. 2, pp. 154-165, Jun. 2015.

[3] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400-3414, Nov. 2015.

[4] Y. Fang, K. Zeng, Z. Wang, W. Lin, Z. Fang, C.-W. Lin, "Objective quality assessment for image retargeting based on structural similarity," *IEEE J. Emerg. Sel. T. Circuits Syst.*, vol. 4, no. 1, pp. 95-105, Mar. 2014.

[5] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516-529, Apr. 2012.

[6] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-inspired divisive normalization for perceptual video coding," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418-1429, Apr. 2013.

[7] T. Zhao, J. Wang, Z. Wang, and C. W. Chen, "SSIM-based coarse-grain scalable video coding", *IEEE Trans. Broadcasting*, vol. 61, no. 2, pp. 210-221, Jun. 2015.

[8] H. R. Wu, A. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025-2043, Sept. 2013.

[9] M. H. Pinsonm, L.-K. Choi, and A. C. Bovik, "Temporal video quality model accounting for variable frame delay distortions," *IEEE Trans. Broadcasting*, vol. 60, no. 4, pp. 637-649, Dec. 2014.

[10] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480-1494, Sept. 2015.

[11] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297-312, May 2011.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[13] 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.

[14] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, Mar. 2010. [Online], Available: http://vision.okstate.edu/csiq

[15] 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.

[16] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512, Apr. 2012.

[17] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43-54, Jan. 2013.

[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *in Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1398-1402, Nov. 2003.

[19] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971-982, Apr. 2011.

[20] K. Gu, G. Zhai, X. Yang, L. Chen, and W. Zhang, "Nonlinear additive model based saliency map weighting strategy for image quality assessment," *in Proc. IEEE Workshop on Multimedia Signal Process.*, pp. 313-318, Sept. 2012.

[21] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[22] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. T. Sig. Proc.*, vol. 3, no. 2, pp. 193-201, Apr. 2009.

[23] 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.

[24] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Structural similarity weighting for image quality assessment," *in Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, pp. 1-6, Jul. 2013.

[25] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality

[26] N. Ponomarenko *et al.*, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30-45, 2009.

[27] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybernetics*, vol. 46, no. 1, pp. 284-297, Jan. 2016.

[28] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587-607, Mar. 1992.

[29] C. F. Hall and E. L. Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Trans. Syst., Man, Cybern.*, vol. 7, no. 3, pp. 161-170, Mar. 1977.

[30] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An improved full-reference image quality metric based on structure compensation," *in Asia-Pacific Signal and Information Processing Association ASC*, Dec. 2012.

[31] K. Gu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50-63, Jan. 2015.

[32] D. H. Johnson and S. Sinanović, "Symmetrizing the Kullback-Leibler distance," *IEEE Trans. Inf. Theory*, 2001.

[33] X. Min, G. Zhai, Z. Gao, and K. Gu, "Visual attention data for image quality assessment databases," *in Proc. IEEE Int. Symp. Circuits and Systems*, pp. 894-897, Jun. 2014.

[34] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," *in Proc. IEEE Int. Conf. Image Process.*, pp. 1473-1476, Sept. 2012.

[35] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684-695, Feb. 2014.

[36] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000, http://www.vqeg.org/.

[37] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218-3231, Oct. 2015.

[38] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Processing*, Volume 2013, Article ID 905685, 2013.

[39] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," *in Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1693-1697, Nov. 2012.

[40] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcasting*, vol. 60, no. 3, pp. 555-567, Sept. 2014.

[41] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcasting*, vol. 61, no. 3, pp. 520-531, Sept. 2015.

Ke Gu received the B.S. and PhD degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is the reviewer for IEEE T-IP, T-CSVT, T-MM, T-CYB, SPL, Neurocomputing, SPIC, JVCI, SIViP, IET-IP, etc. He has reviewed more than 30 journal papers each year. From Jul. to Nov. 2014, he was a visiting student at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. From Dec. 2014 to Jan. 2015, he was a visiting student at the School of Computer Engineering, Nanyang Technological University, Singapore. From Feb. to Mar. and from Nov. to Dec. 2015, he was a visiting student at the Department of Computer Science and Technology, Peking University, Beijing, China. His research interests include quality assessment, contrast enhancement and visual saliency detection.

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008, and the Ph.D. degree in computer application technology from the Peking University, in 2014. He is currently a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From Apr. 2011 to Aug. 2011, he was with Microsoft Research Asia, Beijing, as an Intern. He has proposed more than 20 technical proposals to ISO/MPEG, ITU-T and AVS video coding standards. His current research interests include video compression and image/video quality assessment.

Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Germany. He received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012. His research interests include multimedia signal processing and perceptual signal processing.

Weisi Lin (F'16) received the Ph.D. degree from Kings College, London University, London, U.K., in 1993. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, and served as a Lab Head of Visual Processing, Institute for Infocomm Research. He authors over 300 scholarly publications, holds 7 patents, and receives over S$ 4 million in research grant funding. He has maintained active long-term working relationship with a number of companies. His research interests include image processing, video compression, perceptual visual and audio modeling, computer vision, and multimedia communication. He served as an Associate Editor of IEEE Transactions on Multimedia, IEEE Signal Processing Letters, and Journal of Visual Communication and Image Representation. He is also on six IEEE Technical Committees and Technical Program Committees of a number of international conferences. He was the Lead Guest Editor for a special issue on perceptual signal processing of the IEEE Journal OF Selected Topics in Signal Processing in 2012. He is a Chartered Engineer in the U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He Co-Chaired the IEEE MMTC special interest group on quality of experience. He was an Elected Distinguished Lecturer of APSIPA in 2012/3.

Xiaokang Yang (SM'04) received the B. S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2000. He is currently a Full Professor and Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From September 2000 to March 2002, he was a Research Fellow in Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has published over 80 refereed papers, and has filed six patents. His current research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition. He actively participates in the International Standards such as MPEG-4, JVT, and MPEG-21. He received the Microsoft Young Professorship Award 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003), and awards from A-STAR and Tan Kah Kee foundations. He is a member of Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society. He was the Special Session Chair of Perceptual Visual Processing of IEEE ICME2006. He is the local co-chair of ChinaCom2007 and the technical program co-chair of IEEE SiPS2007.

Wenjun Zhang (F'11) received the B.S., M.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987 and 1989, respectively. From 1990 to 1993, He worked as a post-doctoral fellow at Philips Kommunikation Industrie AG in Nuremberg, Germany, where he was actively involved in developing HD-MAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a full professor in the Department of Electronic Engineering in 1995. As the national HDTV TEEG project leader, he successfully developed the first Chinese HDTV prototype system in 1998. He was one of the main contributors to the Chinese Digital Television Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He holds more than 40 patents and published more than 90 papers in international journals and conferences. Prof. Zhang's main research interests include digital video coding and transmission, multimedia semantic processing and intelligent video surveillance. He is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television (NERC-DTV), an industry/government consortium in DTV technology research and standardization and the Chair of Future of Broadcast Television Initiative (FOBTV) Technical Committee.