Quality Assessment Considering Viewing Distance and Image Resolution

Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, *Senior Member, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

***Abstract*—Viewing distance and image resolution have substantial influences on image quality assessment (IQA), but this issue has been highly overlooked in the literature so far. In this paper, we examine the problem of optimal resolution adjustment as a preprocessing step for IQA. In general, the sampling of visual information by human eyes' optics is approximately a low-pass process. For a given visual scene, the amount of the extractable information greatly depends on the viewing distance and image resolution. We first introduce a novel dedicated viewing distance-changed image database (VDID2014) with two groups of typical viewing distances and image resolutions to promote the IQA study for this issue. Then we design a new effective optimal scale selection (OSS) model in dual-transform domains, in which a cascade of adaptive high-frequency clipping in the discrete wavelet transform domain and adaptive resolution scaling in the spatial domain is used. Validation of our technique is conducted on five image databases (LIVE, IVC, Toyama, VDID2014, and TID2008). Experimental results show that the performance of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) can be substantially improved by applying these metrics to OSS model preprocessed images, superior to classical multi-scale-PSNR/SSIM and comparable to the state-of-the-art competitors.**

***Index Terms*—Image quality assessment (IQA), subjective/objective assessment, viewing distance, image resolution, adaptive resolution scaling, adaptive high-frequency clipping.**

I. Introduction

IMAGE quality assessment (IQA) is an important research area due to its possible application in the development and optimization of various image processing algorithms [1]–[5]. Generally, IQA can be divided into both subjective assessment and objective assessment. The former one aims to obtain mean opinion scores (MOSs) through expensive and complicated subjective viewing tests. This usually causes subjective assessment not practical for real-time applications despite the fact that it is always known as the ultimate quality measure. Last decades have witnessed the rise of a large quantity of objective full-reference (FR), reduced-reference (RR) and no-reference (NR) IQA approaches. Since the original image is sometimes not accessible, several researches in recent years have been devoted to the exploration of RR and NR IQA [5]–[11]. Nonetheless, these two types of IQA techniques were mainly effective for specific distortions, such as blur, noise, image compression and contrast change, and their performance results are not often satisfied.

In reality, most existing high-accuracy IQA methods were developed for the FR scenario [12]–[21]. To date, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [13] are perhaps the most popular benchmark models, which have been embedded into many image/video processing systems. Thereafter, Wang *et al*. further proposed multi-scale SSIM (MS-SSIM) [14] by estimating quality in each scale level before integrated with psychophysical weights, and also developed information content weighted PSNR/SSIM (IW-PSNR/SSIM) [16] by fusing the MS model with the natural scene statistics (NSS) inspired visual information fidelity (VIF) [15]. Lately, the MS model was also used widely, e.g., internal generative mechanism (IGM) [19].

However, it has been argued in the survey about IQA [22] that the common subjective image quality databases [23]–[25] separately used distinct selection modes for viewing distances and image resolutions[1] during subjective tests, although some recommendations regarding the viewing distance with respect to monitor resolution have been given for television pictures [26], multimedia [27], 3DTV [28], flat panel displays [29], and mobile devices [30]. In fact, besides these recommendations, the viewing distance is largely determined by other factors, such as the size of house and the location of furniture. Given a fixed monitor size, higher level compression can be made as the viewing distance becomes farther, in order to save the bandwidth without the loss of the quality of experience (QoE), and thus some modifications to existing IQA measures should be done as well. Obviously, the MS model of constant weights for fixed levels tends to work ineffectively in different viewing distances and image resolutions.

To specify, the QoE, on one hand, is seriously influenced by the image/video resolution, which promotes the constant pursuit of high-definition technologies, and on the other hand, it is also strongly affected by the viewing distance. An example in Fig. 1 reveals distinct perceptual quality for the same frame

Manuscript received July 8, 2014; revised March 26, 2015; accepted April 1, 2015. Date of publication August 11, 2015; date of current version September 2, 2015. This work was supported in part by the National Science Foundation of China under Grant 61025005, Grant 61371146, Grant 61221001, and Grant 61390514, in part by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339, and in part by the Shanghai Municipal Commission of Economy and Informatization under Grant 140310.

The authors are with the Institute of Image Communication and Information Processing, Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: guke.doctor@gmail.com).

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/TBC.2015.2459851

[1]Note that the image resolution is different from the image size (in display), which should be the image resolution multiplied by dot pitch of the screen.

Fig. 1. Video displayed at the viewing distances of four and six times the image weight. Three representative regions are highlighted with red rectangles in each video screen for comparison.

resolution and the same environment but at different viewing distances. The left video stream is encoded with 8Mps bitrate, 25f/s framerate and the resolution of 1920×1080, while the other is created using the same parameters but at 2Mps bitrate for comparison. These two pictures were captured by using a digital camera which was configured to simulate the behavior of human eyes as much as possible. We highlighted three representative areas with red rectangles in each video screen. The left subfigure was taken at the distance of four times the video height, and we can easily observe the difference between two video frames. The right subfigure presents an image taken at the distance of six times the video height, but this time, the difference is hardly found. This phenomenon can be explained by the fact that the perception to image details mostly depends on the effective resolution of the human visual system (HVS). As the viewing distance increases, the viewing angel shrinks and less image details can be noticed. As a result, we believe that it is necessary to consider viewing distances and image resolutions in IQA designs for images/videos.

For a comprehensive test and comparison of our technique with existing related models, we first propose a new dedicated viewing distance-changed image database (VDID2014). This database consists of 160 images generated from eight pristine versions of two typical aspect ratios (height/width), and 320 differential MOS (DMOS) values collected from 20 inexperienced observers at two typical viewing distances, i.e., four and six times of the image height in terms of the ratio of the two physical distances.

Traditionally, most studies of objective or subjective assessment were separately conducted, as shown in Fig. 2(a). The former works to compute the local distortion map followed by an effective pooling method, while the latter uses relevant methodologies (e.g., single-stimulus) to collect human ratings of image quality before processing the raw data to remove biased scores. As stated above, objective metrics regardless of viewing conditions are not reasonable. Despite the significance, very limited efforts have been made for this issue [22]. Hence we in this paper devote to deploying the impact of viewing distances and image resolutions on IQA, thus to make quality metrics work better and more practical. Note that many mechanisms of the HVS, e.g., masking effects, have serious influence on visual quality, but they have been widely inserted into existing IQA tasks. We thereby leave these mechanisms aside, and focus on exploring the effect of viewing distance and image resolution in this research.

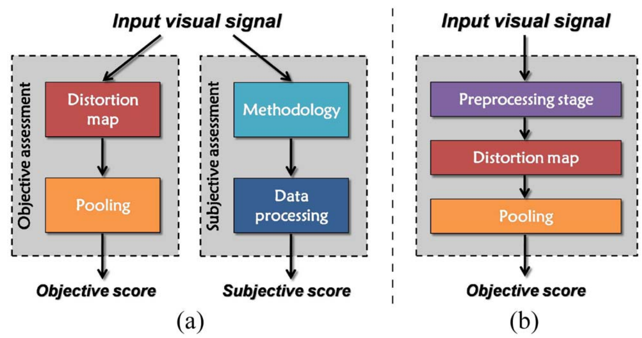

Fig. 2. Flowcharts of: (a) traditional objective and subjective assessments; (b) objective assessment considering viewing distances and image resolutions in an extra preprocessing stage.

A simple self-adaptive scale transform (SAST) model [31] was lately designed to simulate the spatial filtering mechanism of the HVS. Its fundamental idea is to estimate the suitable scaling parameter from the original image resolution and the given viewing distance before resizing input images to boost the performance. Instead of using the spatial domain, another recent work [32] focused on discarding part of image details by adaptive high-frequency clipping (AHC) in the discrete wavelet transform (DWT) domain, and then synthesizing the AHC model filtered subband coefficients back to an image at its original resolution to be used by IQA metrics.

It is natural to combine the above two models to derive a more effective preprocessing method, so as to better remove indiscernible details caused by the varying viewing distance and image resolution in different but complementary domains. We also found that the real viewing field are not identical for the testing images with the same image height and viewing distance but distinct widths. The human eyes' physiological structure indicates that, in this situation, the perceived image resolution can be adjusted by flattering or thickening the crystalline lens [33]. So a new optimal scale selection (OSS) model in spatial and DWT domains is proposed by resizing the AHC model filtered images to the optimal scale estimated using the improved SAST model that considers the human eyes' physiological mechanism.

Fig. 3. Eight lossless natural images with resolutions of 768×512 and 512×512 used in the VDID2014 database.

TABLE I
SUBJECTIVE EXPERIMENTAL CONDITIONS AND PARAMETERS

Method	Single-stimulus (SS)
Evaluation scales	Continuous quality scale from 0 to 1
Color depth	24-bits/pixel color images
Image coder	Portable Network Graphic (PNG)
Distortion type	JPEG2000, JPEG, blur, noise
Subjects	Twenty inexperienced subjects
Image resolution	768×512; 512×512
Viewing distance	Four / six times the image height
Room illuminance	Dark

The remainder of this article is organized as follows. In Sections II and III, we respectively introduce the new VDID2014 database and the proposed OSS model. Section IV compares our OSS model based IQA methods with competitive quality metrics on five databases (LIVE [23], IVC [24], Toyama [25], VDID2014 and TID2008 [34]). We finally summarize the contributions and conclude the paper in Sections V and VI.

II. VDID2014 DATABASE

To thoroughly understand the impact of viewing conditions, we construct a new dedicated VDID2014 database with two classes of typical viewing distances and image resolutions used in classical image databases. It has been mentioned that both above two factors substantially affect human perceptions of image/video quality. But most existing image databases, even those with clear records of viewing conditions, do not take into account the effects of viewing distances and image resolutions on the research of perceptual IQA.

The VDID2014 database includes eight pristine images with resolutions of 768×512 and 512×512, as exhibited in Fig. 3. A total number of 160 images were produced by adding four commonly encountered distortion types: Gaussian blur, white noise, and JPEG2000 and JPEG compressions.

1) Gaussian blur: We utilized Gaussian kernels (standard deviation $\sigma_G = 0.25, 0.5, 1, 1.75, 2.5$) and a 11×11 window with Matlab commands *fspecial* and *imfilter*. Each of R, G and B image planes was blurred by the same kernel.
2) White noise: Noise generated from a standard normal pdf of variance $\sigma_N^2 (= 0.0003, 0.001, 0.003, 0.01, 0.03)$ was added to each of the three channels R, G and B with the *imnoise* Matlab function.
3) JPEG2000: We used the Matlab *imwrite* command to create JPEG2000 compressed images by setting the Q parameter as 15, 30, 60, 120, 240.
4) JPEG: The Matlab *imwrite* command was used to produce JPEG compressed images with five quality levels ($Q = 75, 45, 25, 10, 5$).

The experiment was conducted using a single-stimulus (SS) method according to ITU-R BT.500-13 [26]. We designed an interactive system to automatically display the test images and collect the subjective quality scores using graphical user interface (GUI) in MATLAB, similar to that used in [4]. 20 inexperienced subjects were involved in this study. Most of viewers were college students with various kinds of majors. The entire test was classified into two consecutive sessions, with viewing distances of four and six times the image height. In each session, every subject viewed and graded 160 images. The presentation order was randomized for reducing memory effects on opinion scores. During rating each image, the subjects were asked to provide their overall sensation of quality on a continuous quality scale from 0 to 1. Table I summarizes major information about the test conditions and parameters.

Next, the gathered 320 subjective DMOS values were computed for all testing images. First, we set z_{abc} as the score provided by subject a to the distorted image I_b at the viewing distance of c times the image height, where $a = \{1, \ldots, 20\}$, $b = \{1, \ldots, 160\}$, $c = \{4, 6\}$. And z'_{abc} indicates the original image's score which is defined similarly to z_{abc}. Specifically, we processed the data as follows:

1) Differential scores: The raw rating assigned to an image was subtracted from the rating assigned to its reference image to form the DMOS value $d_{abc} = z_{abc}$-z'_{abc}.
2) Outliers screening: The subjective scores are easy to be contaminated by outliers given by inattentive subjects. To avoid this, we screened the outliers of all viewers' ratings using the method in [35]. Particularly, we adopted a simple outlier detection by treating raw DMOS value for an image to be an outlier if it was outside an interval of the standard deviation width about the mean score for that image.
3) Mean score: The final DMOS for the image I_b is defined as $\frac{1}{N_A}\sum_a d_{abc}$, where N_A is the number of subjects.

To show the value of building VDID2014 database with different viewing distances and image resolutions, we compute and plot the distributions of quality scores gathered from the test in Fig 4: the left plot (overall database) stands for the histogram of all processed opinion scores, while the middle plot (part 1) and right plot (part 2) independently indicate the histograms of DMOS values acquired at the viewing distance of four and six times the image height. We can easily find in the latter two plots that, for the same images, the mean score in part 2 (six times the image height) is clearly smaller than that in part 1 (four times the image height), which is mainly caused by different viewing distances.

III. METHODOLOGY

Presently, the most popular FR IQA algorithms are perhaps PSNR and SSIM, because of their wide adoption in various image processing systems. PSNR is computed from the

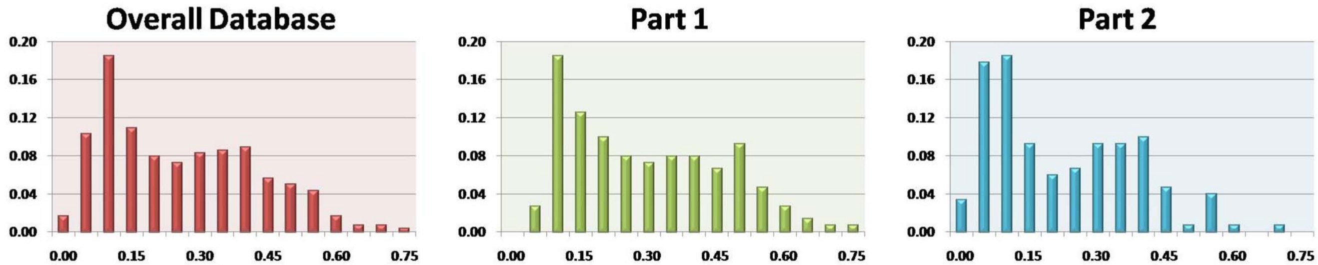

Fig. 4. Histogram of differential mean opinion scores (DMOSs) collected in the subjective viewing test of the VDID2014 database.

mean-squared error (MSE), which quantifies the energy difference between the reference and distorted images. Suppose that r_i and d_i are the i-th pixel values in the reference image $\mathbf{r}$ and its distorted image $\mathbf{d}$. The MSE and PSNR are defined as $\text{MSE} = \frac{1}{M}\sum_{i=1}^{M}(r_i - d_i)^2$ and $\text{PSNR} = 10\log_{10}(\frac{255^2}{\text{MSE}})$, where M is the total number of pixels in the image.

The basic idea of SSIM is to incorporate local luminance, contrast and structural similarities between the reference and distorted images [13]. To be more concretely, the luminance, contrast and structural similarities are defined by

$$l(\mathbf{r}, \mathbf{d}) = \frac{2\mu_r\mu_d + c_1}{\mu_r^2 + \mu_d^2 + c_1} \tag{1}$$

$$c(\mathbf{r}, \mathbf{d}) = \frac{2\sigma_r\sigma_d + c_2}{\sigma_r^2 + \sigma_d^2 + c_2} \tag{2}$$

$$s(\mathbf{r}, \mathbf{d}) = \frac{\sigma_{rd} + c_3}{\sigma_r\sigma_d + c_3} \tag{3}$$

where c_1 to c_3 are three small fixed values to avoid instability when the denominators are close to zero. A 11×11 circular-symmetric Gaussian weighting function $\mathbf{g} = \{g_i | i = 1, 2, \ldots, N\}$ having standard deviation of 1.5 and normalized to unit sum ($\sum_{i=1}^{N} g_i = 1$) is used here. The statistics μ_r, μ_d, σ_r^2, σ_d^2 and σ_{rd} follow the definitions in [13]. The final SSIM score is acquired by

$$\text{SSIM} = \frac{1}{W}\sum_{i=1}^{W} l(r_i, d_i) \cdot c(r_i, d_i) \cdot s(r_i, d_i) \tag{4}$$

where W is the number of local windows in the image.

A. Downsampling Model

It has been realized that considerations of external factors during the viewing session, e.g., the viewing distance and the image resolution, have great influences on IQA performance. A simple and empirical downsampling strategy for preprocessing images before using SSIM was described in [22]:

$$Z_d = \max(1, \text{round}(H_i/256)) \tag{5}$$

where H_i indicates the image height.

B. SAST Model

As the viewing distance grows, the viewing angle shrinks in a gradual way. So the downsampling model with only limited scale parameters is not an ideal solution. For instance,

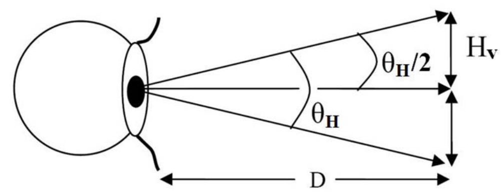

Fig. 5. Illustration of the human visual angle in the horizontal direction. The vertical visual angle has a similar illustration.

at the same distance, images with resolutions of 651×651 and 630×630 have $Z_d = 3$ and $Z_d = 2$, and thereby their downsampled images have resolutions of 217×217 and 315×315. This suggests that the larger image will be scaled down to a smaller resolution in some cases, which violates our common sense. To this end, we in previous work introduced a continually-changing resolution scaling scheme (SAST) [31].

First, we define the visual scope S_v of human eyes for a viewing distance D:

$$S_v = H_v \cdot W_v \tag{6}$$

where H_v and W_v are the visual height and width. According to the illustration in Fig. 5, the two variables can be obtained by

$$H_v = 2\tan\left(\frac{\theta_H}{2}\right) \cdot D \tag{7}$$

$$W_v = 2\tan\left(\frac{\theta_W}{2}\right) \cdot D \tag{8}$$

where θ_H and θ_W separately indicate horizontal and vertical visual angles. θ_H and θ_W are generally assigned to be about 120^o and 150^o [36]. Considering the fact that the real view angle (i.e., angle of gaze) usually becomes narrower to about one third of the common value when one concentrates on the details of an image and scores it. We therefore choose θ_H and θ_W to be 40^o and 50^o in this paper.

Second, it was observed in Fig. 1 that distinguishing tiny artifacts in a visual signal will be more difficult with increasing the viewing distance. On this base, the proper transform scale Z_{sast} can be approximated in the spatial domain by the square root of the ratio between the image resolution and the focused visual scope:

$$\begin{aligned} Z_{\text{sast}} &= \sqrt{\frac{H_i \cdot W_i}{H_v \cdot W_v}} \\ &= \sqrt{\frac{1}{4\tan\left(\frac{\theta_H}{2}\right) \cdot \tan\left(\frac{\theta_W}{2}\right)} \cdot \left(\frac{H_i}{D}\right)^2 \cdot \frac{1}{\gamma}} \end{aligned} \tag{9}$$

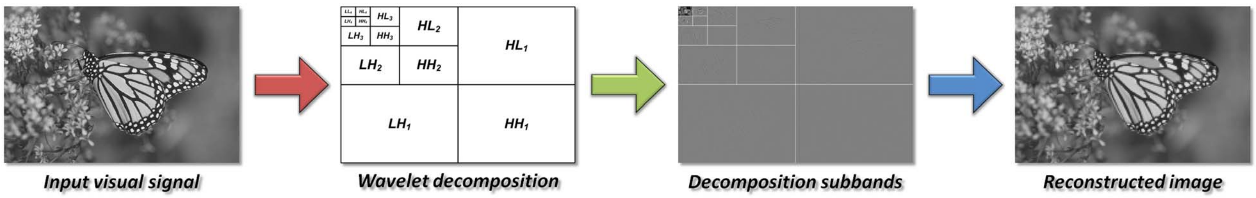

Fig. 6. Sample image and its reconstructed one using wavelet decomposition at the distance of four times the image height.

where H_i/D is a pre-set environment parameter provided by each image database, which will be presented in Table II, and γ represents the aspect ratio of image height to image width defined by

$$\gamma = \frac{H_i}{W_i} \tag{10}$$

and W_i stands for the image width.

C. AHC Model

The image resolution scaling is essentially to discard part of the high resolution information. To specify, image details are hard to be found when the ratio of image height and viewing distance (H_i/D) is small. Hence the AHC model works to remove part of image details in wavelet subbands, thus to be a preprocessing step for IQA [32].

The Haar wavelet decomposition was applied in the AHC, on account of its simple mother wavelet defined as

$$\psi(t) = \begin{cases} 1 & 0 \le t \le 1/2 \\ -1 & 1/2 \le t \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

with the scaling function $\phi(t)$ being

$$\phi(t) = \begin{cases} 1 & 0 \le t \le 2 \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

Alternatively, the Haar wavelet has a good ability and a wide application in many research fields. More details can be found in [12]. An example of the Haar wavelet decomposition is presented in Fig. 6 by using the "monarch" image.

Next, we introduced a weighting function to assign different weights in each of LH, HL and HH subbands:

$$w = \rho \cdot \phi^{(kv_1 + v_2)} \tag{13}$$

where $v_1 = l - L$ with l and L being the current processed layer and the number of decomposition layers. $v_2 = D/D_0$ with D_0 being a virtual baseline distance. ϕ is a fixed bottom number. k is used to adjust the relative importance of v_1 and v_2. In [17], the authors applied three gradient operators for horizontal and vertical directions based on a primary study of V1 cells that human eyes have maximum responses to horizontal and vertical stimuli [37]. Thus this paper uses the coefficient ρ to better retain LH and HL subbands than the HH subband on the same level:

$$\rho = \begin{cases} 1/2 & \text{when LH or HL} \\ 1 & \text{when HH.} \end{cases} \tag{14}$$

Each computed weight is compared with a threshold (thr = 1 in this work). If the weight is greater than the threshold, the associated subband will be clipped out.

Actually, Eq. (13) indicates that we incline to cut off the subband in the smaller layer (for v_1 term) and at the farther viewing distance (for v_2 term). The wavelet reconstruction is used to get the final image, as exemplified in Fig. 6. Note that although the multi-resolution analysis is taken inside the AHC model, the terminal output is a single-scale image with part of high-frequency details properly erased.

D. Proposed OSS Model

Physiological studies of human eyes suggest that light from an object outside the eye is imaged on the retina that lines the inside of the wall's overall posterior portion, when the eye is appropriately focused [33]. Despite the fixed distance between the crystalline lens and the retina (the imaging region), the focal length accomplishing proper focus is gained by varying the shape of crystalline lens. The fibers in the ciliary body can complete this task by flattening or thickening the crystalline lens for distant or near objects. This process is called "accommodation". For long distances, the negative accommodation is to adjust the eyes by relaxation of the ciliary muscles, whereas the positive accommodation works by contraction of the ciliary muscles for short distances.

As a matter of fact, it was found in our subjective viewing test that the adaptive image resolution scaling is also affected by the aspect ratio. When one watches any image/video signal deviated from the optimal aspect ratio, the visual field is not exactly full of the whole image/video picture, and this makes human eyes tend to attenuate the crystalline lens to zoom in to the meaningful part. So we modify the transform coefficient Z_{sast} as follows:

$$Z'_{\text{sast}} = Z_{\text{sast}}^{\left(1 - \frac{|\gamma - \gamma_{\text{o}}|^{\beta}}{\alpha}\right)} \tag{15}$$

where α and β are both selected as 2 to control the speed of the modification process caused by different aspect ratios. γ_{o} means the optimal aspect ratio of human eyes. Considering that the aspect ratio of (H:W) 9:16 has become the most common aspect ratio for televisions and computer monitors and is also the international standard format of digital television and analog widescreen television, it is reasonable to suppose the optimal aspect ratio be what the human eyes are well suited to, and define γ_{o} as 9:16 in this implementation. That is to say, this function implies that the human eyes will further adjust the transform coefficient when the aspect ratio (H:W) is not the optimal aspect ratio. Next, we take into account that the models by adaptive resolution scaling and adaptive high-frequency clipping effectively work in different and complementary domains, and thus use a cascade of aforementioned two methods (the AHC model followed by the modified SAST model) to derive the proposed OSS model.

It needs to stress that we focus on the influence of viewing distance and image resolution on IQA. As shown in Fig. 2(b), with respect to the traditional framework in which the input images are directly used to compute the objective score, in this paper the objective assessment preprocesses input visual signals by considering viewing distance and image resolution before using IQA models, e.g., PSNR and SSIM.

TABLE II
SPECIFICATIONS OF LIVE, IVC, TOYAMA, AND VDID2014

Database name	Image resolution ($W_i \times H_i$)	D / H_i	Number
LIVE	768×512 480×720 640×512 632×505 634×505 618×453 610×488 627×482 634×438	3∼3.75	779
VDID2014	768×512 512×512	4 & 6	320
IVC	512×512	4	185
Toyama	768×512	6	168

IV. EXPERIMENTAL RESULTS AND COMPARISON

In this section, we will testify and compare the performance of the proposed model with 22 classical and state-of-the-art IQA approaches, which are given as follows.

- PSNR and SSIM [13], the benchmark IQA methods with a wide usage in the image processing literature.
- D-PSNR/SSIM, modified by the simple and empirical downsampling strategy [22].
- SAST-PSNR/SSIM [31], performing by the adaptive resolution scaling before PSNR and SSIM are used.
- AHC-PSNR/SSIM [32], computing PSNR and SSIM on the AHC model filtered image signals.
- OSS-PSNR/SSIM, systematically incorporating the AHC model with the modified SAST model.
- MS-PSNR/SSIM [14], using PSNR/SSIM in each scale level followed by fusing the values in different levels with distinct weights obtained via a psychophysical test.
- IW-PSNR/SSIM [16], combining the MS and NSS models to derive the currently optimal pooling scheme.
- FSIM [17], applying complementary phase congruency and gradient magnitude in characterizing the image local quality due to the fact that the HVS understands an image mainly relying on the low-level features.
- GSIM [18], developed by emphasizing on the gradient magnitude similarity because the image gradient conveys visual information and favors scene understanding.
- IGM [19], decomposing an image into the predicted and disorderly parts by the free energy based brain theory [38], before integrating the modified PSNR and SSIM values on those two parts with psychophysical weights [14].
- GMSD [20], predicting visual quality by a new effective and efficient pooling scheme – the standard deviation of the pixel-wise gradient magnitude similarity map.
- X-FSIM/GSIM (X = {D, SAST, AHC, OSS}), similar to X-PSNR/SSIM but applied to FSIM/GSIM.

To measure those above IQA metrics, subject-rated image quality databases are required as testing beds. In addition to our new VDID2014, three related databases (LIVE, IVC and Toyama) are chosen in this work. Detailed information of the three databases is listed below as follows.

- The LIVE database [23] includes five image datasets of 982 images, which are divided into 29 original images of nine resolutions and 779 distorted images corrupted by five distortion types. These types involve: 1) JPEG2000 compression; 2) JPEG compression; 3) White noise; 4) Gaussian blur; 5) Fast fading channel distortion of JPEG2000 compressed bitstream. The subjective test was carried out at the viewing distance of 3∼3.75 times the image height. The entire raw scores are processed by an outlier detection and subject rejection algorithm according to [35].
- The IVC database [24] contains 185 images created from 10 sources. Four distortion types are applied: 1) JPEG compression; 2) JPEG2000 compression; 3) Local adaptive resolution (LAR) coding; 4) Blurring. The resolution of all images is regulated of 512×512, and the viewing distance was fixed at four times the image height.
- The Toyama database [25] consists of 168 images, which generated by exerting JPEG and JPEG2000 compressions on 12 natural images. The testing process was conducted at the fixed distance of six times the image height.

Table II summarizes some relevant information about the used image quality databases[2], in order to straightforwardly show them to the readers.

As the suggestion given by the video quality experts group (VQEG) [39], we first apply the nonlinear regression between the subjective ratings and the prediction scores of each IQA method above with the four-parameter logistic function:

$$q(\varepsilon) = \frac{\tau_1 - \tau_2}{1 + \exp\left(-\frac{\varepsilon - \tau_3}{\tau_4}\right)} + \tau_2 \tag{16}$$

where ε and $q(\varepsilon)$ are the input and mapped scores, and τ_1 to τ_4 are free parameters to be determined. Then we employ five typical performance indices to testify and compare the proposed algorithms with the IQA metrics tested. The five indices include Pearson linear correlation coefficient (PLCC) for measuring prediction accuracy, Spearman rank-order correlation coefficient (SRCC) and Kendall's rank-order correlation coefficient (KRCC) for measuring prediction monotonicity, as well as average absolute prediction error (AAE) and root mean-squared (RMS) error for measuring prediction consistency. Concrete definitions about the above five evaluations can be found in [7]. Note that a value close to 1 for PLCC, SRCC, KRCC, yet close to 0 for AAE, RMS means superior correlation with subjective human ratings.

[2]In those testing databases the information concerning the relationship of sizes between the image and the screen has not been illustrated, so we in this paper suppose the testing image is full of the screen/monitor height or width with the aspect ratio unchanged, and thus obtain the D/H_i. Actually, this number should be derived using the ratio of the two physical distances, just as in our VDID2014 database.

TABLE III

PERFORMANCE EVALUATIONS AND TWO AVERAGES OF THE COMPETING IQA MEASURES ON LIVE, IVC, TOYAMA, AND VDID2014 DATABASES. THE BEST TWO PERFORMED METRICS IN THE PSNR-TYPE AND SSIM-TYPE OF ALGORITHMS ARE HIGHLIGHTED WITH BOLDFACE

IQA Measures	LIVE database (779 images) [23]					IVC database (185 images) [24]				
	PLCC	SRCC	KRCC	AAE	RMS	PLCC	SRCC	KRCC	AAE	RMS
PSNR	0.8701	0.8756	0.6865	10.539	13.469	0.7192	0.6886	0.5220	0.6689	0.8465
D-PSNR [22]	0.8995	0.9031	0.7227	9.4220	11.940	0.8791	0.8721	0.6922	0.4266	0.5808
SAST-PSNR [31]	0.9134	0.9160	0.7450	8.5101	11.121	0.8953	0.8889	0.6992	0.4328	0.5428
AHC-PSNR [32]	0.9295	0.9314	0.7731	7.6244	10.077	**0.9107**	**0.9019**	**0.7188**	**0.3995**	**0.5032**
OSS-PSNR (Pro.)	**0.9304**	**0.9328**	**0.7768**	**7.4987**	**10.012**	**0.9123**	**0.9041**	**0.7226**	**0.3968**	**0.4990**
MS-PSNR [14]	0.9071	0.9110	0.7366	8.9532	11.503	0.8388	0.8340	0.6479	0.4934	0.6634
IW-PSNR [16]	**0.9329**	**0.9328**	**0.7800**	**7.3262**	**9.8394**	0.9055	0.8999	0.7168	0.4100	0.5170
SSIM	0.9014	0.9104	0.7311	9.3341	11.832	0.7924	0.7788	0.5939	0.5547	0.7431
D-SSIM [22]	0.9300	0.9391	0.7768	8.2062	10.044	0.9117	0.9017	0.7221	0.3772	0.5007
SAST-SSIM [31]	0.9305	0.9448	0.7914	8.1933	10.011	0.9042	0.8905	0.7062	0.4109	0.5203
AHC-SSIM [32]	0.9321	0.9477	0.7987	8.1506	9.8967	0.9066	0.8957	0.7170	0.4047	0.5142
OSS-SSIM (Pro.)	0.9316	**0.9499**	**0.8078**	8.1000	9.9310	**0.9144**	**0.9035**	**0.7291**	**0.3846**	**0.4933**
MS-SSIM [14]	**0.9338**	0.9448	0.7927	**7.7605**	**9.7788**	0.8931	0.8846	0.7006	0.4122	0.5480
IW-SSIM [16]	**0.9425**	**0.9567**	**0.8175**	**7.4405**	**9.1317**	**0.9228**	**0.9125**	**0.7339**	**0.3698**	**0.4693**
FSIM [17]	0.9540	0.9634	0.8335	6.4647	8.1907	0.9378	0.9263	0.7566	0.3380	0.4228
GSIM [18]	0.9443	0.9561	0.8150	7.1888	8.9883	0.9390	0.9292	0.7619	0.3279	0.4190
IGM [19]	0.9565	0.9581	0.8250	6.0742	7.9686	0.9128	0.9025	0.7283	0.3783	0.4976
GMSD [20]	0.9568	0.9603	0.8268	6.1990	7.9447	0.9234	0.9148	0.7373	0.3743	0.4678

IQA Measures	Toyama database (168 images) [25]					VDID2014 database (320 images)				
	PLCC	SRCC	KRCC	AAE	RMS	PLCC	SRCC	KRCC	AAE	RMS
PSNR	0.6355	0.6132	0.4443	0.7832	0.9662	0.8494	0.8678	0.6779	0.0696	0.0915
D-PSNR [22]	0.7654	0.7583	0.5605	0.6453	0.8053	0.9061	0.9163	0.7414	0.0562	0.0734
SAST-PSNR [31]	0.8343	0.8272	0.6269	0.5524	0.6898	0.9423	0.9463	0.7968	0.0437	0.0581
AHC-PSNR [32]	**0.8649**	**0.8619**	**0.6711**	**0.4988**	**0.6282**	**0.9567**	**0.9569**	**0.8155**	**0.0390**	**0.0505**
OSS-PSNR (Pro.)	**0.8623**	**0.8621**	**0.6724**	**0.5014**	**0.6338**	**0.9611**	**0.9590**	**0.8219**	**0.0370**	**0.0479**
MS-PSNR [14]	0.7522	0.7411	0.5493	0.6557	0.8246	0.9045	0.9147	0.7401	0.0564	0.0740
IW-PSNR [16]	0.8501	0.8475	0.6508	0.5219	0.6590	0.9398	0.9327	0.7723	0.0458	0.0593
SSIM	0.7978	0.7870	0.5922	0.5891	0.7545	0.8261	0.8422	0.6416	0.0744	0.0978
D-SSIM [22]	0.8877	0.8794	0.6939	0.4451	0.5762	0.8872	0.8958	0.7076	0.0620	0.0800
SAST-SSIM [31]	0.9072	0.9048	0.7289	0.4215	0.5265	0.9144	0.9181	0.7473	0.0531	0.0702
AHC-SSIM [32]	0.9142	0.9117	0.7387	0.3944	0.5071	**0.9188**	**0.9349**	**0.7728**	**0.0533**	**0.0685**
OSS-SSIM (Pro.)	**0.9373**	**0.9371**	**0.7814**	**0.3318**	**0.4363**	**0.9280**	**0.9352**	**0.7787**	**0.0491**	**0.0646**
MS-SSIM [14]	0.8926	0.8870	0.7049	0.4328	0.5641	0.8910	0.8995	0.7131	0.0614	0.0787
IW-SSIM [16]	**0.9243**	**0.9202**	**0.7537**	**0.3696**	**0.4775**	0.9129	0.9179	0.7442	0.0550	0.0708
FSIM [17]	0.9064	0.9050	0.7280	0.4053	0.5287	0.9208	0.9247	0.7568	0.0517	0.0677
GSIM [18]	0.9279	0.9232	0.7535	0.3630	0.4666	0.9170	0.9192	0.7439	0.0544	0.0692
IGM [19]	0.8708	0.8654	0.6735	0.4871	0.6152	0.9322	0.9293	0.7657	0.0479	0.0628
GMSD [20]	0.8579	0.8528	0.6588	0.5014	0.6430	0.9213	0.9274	0.7595	0.0530	0.0675

IQA Measures	Direct average					Database size-weighted average				
	PLCC	SRCC	KRCC	AAE	RMS	PLCC	SRCC	KRCC	AAE	RMS
PSNR	0.7686	0.7613	0.5827	3.0151	3.8432	0.8192	0.8197	0.6356	5.8453	7.4657
D-PSNR [22]	0.8625	0.8624	0.6792	2.6375	3.3498	0.8828	0.8853	0.7041	5.1963	6.5891
SAST-PSNR [31]	0.8963	0.8946	0.7170	2.3848	3.1029	0.9083	0.9090	0.7369	4.6944	6.1281
AHC-PSNR [32]	**0.9155**	**0.9131**	**0.7446**	2.1404	2.8147	**0.9256**	**0.9253**	**0.7637**	4.2077	5.5542
OSS-PSNR (Pro.)	**0.9165**	**0.9145**	**0.7484**	**2.1085**	**2.7981**	**0.9270**	**0.9267**	**0.7678**	**4.1398**	**5.5189**
MS-PSNR [14]	0.8506	0.8502	0.6685	2.5397	3.2663	0.8799	0.8823	0.7044	4.9546	6.3677
IW-PSNR [16]	0.9071	0.9032	0.7300	**2.0760**	**2.7687**	0.9214	0.9187	0.7553	**4.0533**	**5.4341**
SSIM	0.8294	0.8296	0.6397	2.6381	3.3569	0.8589	0.8643	0.6778	5.1630	6.5516
D-SSIM [22]	0.9041	0.9040	0.7251	2.2726	2.8002	0.9133	0.9179	0.7450	4.5159	5.5367
SAST-SSIM [31]	0.9141	0.9146	0.7434	2.2697	2.7820	0.9209	0.9274	0.7636	4.5085	5.5137
AHC-SSIM [32]	0.9179	0.9225	0.7568	2.2508	2.7466	0.9239	0.9341	0.7756	4.4817	5.4489
OSS-SSIM (Pro.)	**0.9276**	**0.9313**	**0.7742**	2.2144	**2.7323**	**0.9292**	**0.9392**	**0.7884**	4.4401	5.4569
MS-SSIM [14]	0.9026	0.9039	0.7279	**2.1667**	2.7424	0.9144	0.9204	0.7533	**4.2796**	**5.3988**
IW-SSIM [16]	**0.9256**	**0.9268**	**0.7623**	**2.0587**	**2.5373**	**0.9314**	**0.9383**	**0.7833**	**4.0938**	**5.0298**
FSIM [17]	0.9298	0.9298	0.7687	1.8149	2.3025	0.9391	0.9434	0.7946	3.5697	4.5243
GSIM [18]	0.9321	0.9319	0.7686	1.9835	2.4858	0.9357	0.9407	0.7854	3.9526	4.9449
IGM [19]	0.9181	0.9138	0.7481	1.7469	2.2861	0.9357	0.9339	0.7821	3.3739	4.4236
GMSD [20]	0.9148	0.9138	0.7456	1.7819	2.2807	0.9333	0.9348	0.7812	3.4432	4.4112

Table III shows the performance results of 18 testing IQA metrics on four related databases. To compare all performance indices of those IQA measures, we further provide the average values of PLCC, SRCC, KRCC, AAE and RMS (after the nonlinear regression) in Table III. In this research, two kinds of average results are reported: 1) the direct average among the correlation scores for each quality metric; 2) the database size-weighted average depending on the number of images in each database, i.e., 779 for LIVE, 185 for IVC, 168 for Toyama, and 320 for VDID2014.

Additionally, we also tabulate and compare in Table IV the correlation performances of FSIM, GSIM, FSIM-type,

TABLE IV
PERFORMANCE MEASURES OF FSIM-, GSIM-TYPES OF METHODS AS WELL AS THE STATE-OF-THE-ART IGM AND GMSD ON THE VDID2014 DATABASE. WE EMPHASIZE THE BEST PERFORMED METRICS IN FSIM-, GSIM-TYPES OF ALGORITHMS

Models	PLCC	SRCC	KRCC	AAE	RMS	Models	PLCC	SRCC	KRCC	AAE	RMS
FSIM	0.9208	0.9247	0.7568	0.0517	0.0677	GSIM	0.9170	0.9192	0.7439	0.0544	0.0692
D-FSIM	0.9208	0.9247	0.7568	0.0517	0.0677	D-GSIM	0.9170	0.9192	0.7439	0.0544	0.0692
SAST-FSIM	0.9359	0.9402	0.7879	0.0465	0.0611	SAST-GSIM	0.9330	0.9270	0.7675	0.0477	0.0624
AHC-FSIM	0.9416	0.9491	0.8037	0.0449	0.0584	AHC-GSIM	0.9354	0.9329	0.7749	0.0481	0.0613
OSS-FSIM	**0.9438**	**0.9528**	**0.8107**	**0.0438**	**0.0573**	OSS-GSIM	**0.9393**	**0.9388**	**0.7842**	**0.0465**	**0.0595**
IGM	0.9322	0.9293	0.7657	0.0479	0.0628	GMSD	0.9213	0.9274	0.7595	0.0530	0.0675

TABLE V
PERFORMANCE COMPARISON OF PSNR/SSIM-TYPE OF METHODS WITH F-TEST (STATISTICAL SIGNIFICANCE). THE SYMBOL "+1", "0," OR "−1" MEANS THAT THE METRIC IS STATISTICALLY (WITH 95% CONFIDENCE) BETTER, UNDISTINGUISHABLE, OR WORSE THAN THE CORRESPONDING METHODS

OSS-PSNR	PSNR	D-PSNR	MS-PSNR	IW-PSNR	***OSS-SSIM***	SSIM	D-SSIM	MS-SSIM	IW-SSIM
LIVE	+1	+1	+1	0	LIVE	+1	0	0	−1
IVC	+1	+1	+1	0	IVC	+1	0	+1	0
Toyama	+1	+1	+1	0	Toyama	+1	+1	+1	0
VDID2014	+1	+1	+1	+1	VDID2014	+1	+1	+1	0

GSIM-type of methods (X-FSIM and X-GSIM with X = {D, SAST, AHC, OSS}) on the VDID2014 database, since this database is composed of frequently used distortion types, and distinct viewing distances and image resolutions. For a convenient comparison across different IQA techniques, we highlight the top two models with boldface. From Tables III and IV, four major observations can be found:

- First, the performance improvement of OSS model based PSNR and SSIM relative to the original versions are more than 6.5% and 4.3% on LIVE, more than 31% and 16% on IVC, more than 40% and 19% on Toyama, more than 10% and 11% on VDID2014, more than 20% and 12% on direct average, and more than 13% and 8.6% on database size-weighted average. It is apparent that the proposed OSS method leads to consistent and considerable performance improvements for modified PSNR and SSIM. Furthermore, Table III presents that, for the average results of correlation performance, our OSS model outperforms MS-PSNR/SSIM and IW-PSNR, and competes well with IW-SSIM and recent FSIM, GSIM, IGM and GMSD.
- Second, despite the high performance, the proposed technique has a low computational cost. In fact, our OSS model runs by performing a pre-filtering based on adaptive high-frequency clipping in the DWT domain, and then resizing the filtered images to the optimal scale estimated using the modified adaptive image resolution scaling.
- Third, it deserves attention that, in contrast to the MS and IW based PSNR and SSIM methods as well as recently proposed FSIM, GSIM, IGM, GMSD metrics, our model is more effective on Toyama and VDID2014 databases. This phenomenon is not just a coincidence, but can be explained by the facts that: 1) the psychophysical testing based MS and IW models were designed under the general viewing distance of roughly 3∼4 times of the image height, which results in their very high performances on LIVE and IVC; 2) the total (or part) of subjective image quality scores in Toyama and VDID2014 were obtained using a far viewing distance of six times the image height.
- Fourth, our technique is robust not only among different databases, but also among distinct single-scale IQA metrics. We have demonstrated the performance increase of the OSS model based PSNR and SSIM on each database. Also, we have validated the effectiveness of the proposed OSS model for improving FSIM, GSIM and GMSD on VDID2014, superior to state-of-the-art FSIM, GSIM, IGM and GMSD. It is easy to find the remarkable performance gains of PSNR over SSIM, FSIM, GSIM and GMSD by using the OSS model. This phenomenon is perhaps because, as compared to the simple PSNR, all of SSIM, FSIM, GSIM and GMSD methods adopt low-pass filtering modules that are very likely to conflict with the proposed OSS model.

Besides, the statistical significance of the proposed OSS model is evaluated by the F-test that computes the prediction residuals between the converted objective scores (after the nonlinear mapping) and the subjective ratings. Let F be the ratio between two residual variances, and $F_{critical}$ (determined by the number of residuals and the confidence level) be the judgement threshold. If $F > F_{critical}$, then the difference of performance between these two metrics is significant. The statistical significance between our technique and the other IQA methods in comparison is listed in Table V, where the symbol "+1", "0" or "−1" means that the proposed metric is statistically (with 95% confidence) better, indistinguishable, or worse than the corresponding metric, respectively. It can be easily viewed that our algorithm is in most cases superior to the classical MS model and comparable to the currently optimal IW model, which proves the effectiveness of the proposed scheme. Furthermore, it should be noted that, compared to the IW model that analyzes visual signals in the brain using the multi-scale strategy and information content, our OSS method works in an alternative stage of the HVS for simulating the images projected on the retina. So some future work might attempt to systematically incorporate OSS and IW schemes to

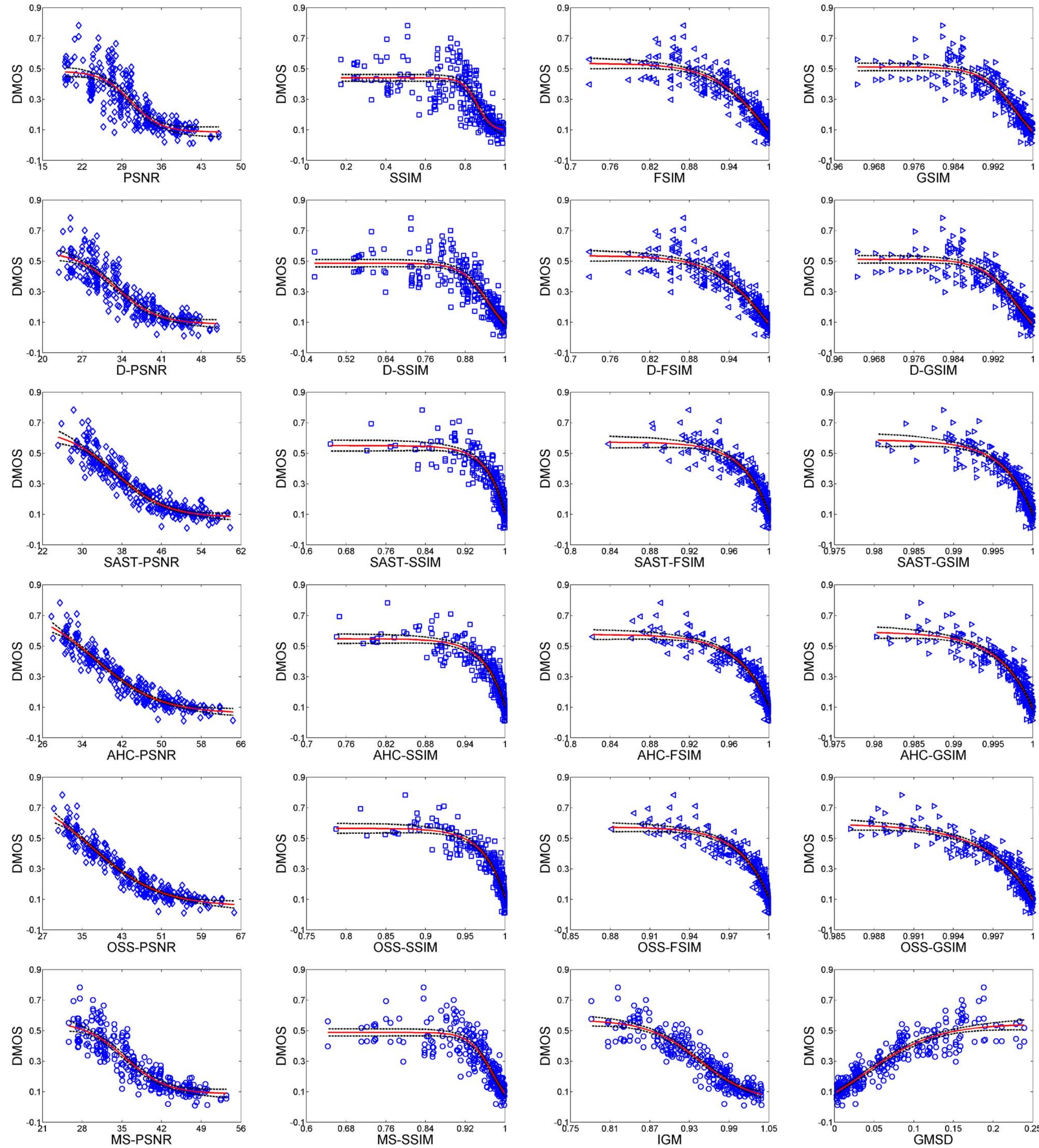

Fig. 7. Scatter plots of DMOS vs. PSNR-, SSIM-, FSIM-, and GSIM-types of methods based on the downsampling model, the SAST model, the AHC model, and the proposed OSS model, as well as MS-PSNR/SSIM and state-of-the-art IGM and GMSD on the VDID2014 database. The (red) lines are curves fitted with the logistic function and the (black) dash lines are 95% confidence intervals.

built a higher-performance model. Combining the results in Table III, we also find that the OSS based PSNR and SSIM are just a little inferior to state-of-the-art FSIM, GSIM, IGM and GSMD on LIVE and IVC, while is equal to or higher than those four IQA approaches on Toyama and VDID2014, which confirms the effectiveness of our model as well.

We display scatter plots of DMOS versus PSNR-, SSIM-FSIM-, GSIM-types of methods using various models on the VDID2014 database in Fig. 7. From top to bottom, the first to fifth rows are: the original IQA approaches, the downsampling model, the SAST model, the AHC model, and the proposed OSS model. The scatter plots of MS-PSNR/SSIM and state-of-the-art IGM and GMSD are also illustrated in the sixth row for comparison. Clearly, the convergence and monotonicity of our OSS algorithm is better than other metrics presented in this figure.

TABLE VI
SENSITIVITY TESTING WITH DIFFERENT PARAMETERS ACCORDING TO SRCC VALUES ON THE VDID2014 DATABASE

ϕ	9	9.5	**10**	10.5	11	k	1.5	1.75	**2**	2.25	2.5
OSS-PSNR	0.9590	0.9590	**0.9590**	0.9590	0.9590	OSS-PSNR	0.9590	0.9590	**0.9590**	0.9590	0.9590
OSS-SSIM	0.9352	0.9352	**0.9352**	0.9352	0.9352	OSS-SSIM	0.9352	0.9352	**0.9352**	0.9352	0.9352

ψ	1.5	1.75	**2**	2.25	2.5	D_0	496	504	**512**	520	528
OSS-PSNR	0.9483	0.9472	**0.9590**	0.9590	0.9590	OSS-PSNR	0.9583	0.9583	**0.9590**	0.9590	0.9590
OSS-SSIM	0.9213	0.9203	**0.9352**	0.9352	0.9352	OSS-SSIM	0.9339	0.9339	**0.9352**	0.9352	0.9352

α	1	1.5	**2**	2.5	3	β	1	1.5	**2**	2.5	3
OSS-PSNR	0.9592	0.9593	**0.9590**	0.9592	0.9588	OSS-PSNR	0.9583	0.9591	**0.9590**	0.9584	0.9586
OSS-SSIM	0.9382	0.9365	**0.9352**	0.9342	0.9334	OSS-SSIM	0.9381	0.9366	**0.9352**	0.9338	0.9323

TABLE VII
PERFORMANCE COMPARISON OF DISTINCT COMBINATIONS OF COMPONENTS EMPLOYED IN THE PROPOSED OSS MODEL

IQA Measures	LIVE	Toyama	IVC	VDID2014	IQA Models	LIVE	Toyama	IVC	VDID2014
PSNR	0.8756	0.6132	0.6886	0.8678	SSIM	0.9104	0.7870	0.7788	0.8422
SAST-PSNR	0.9160	0.8272	0.8889	0.9463	SAST-SSIM	0.9448	0.9048	0.8905	0.9181
AHC-PSNR	**0.9314**	**0.8619**	0.9019	0.9569	AHC-SSIM	**0.9477**	0.9117	0.8957	**0.9349**
MOD-SAST-PSNR	0.9073	0.8231	0.8819	0.9497	MOD-SAST-SSIM	0.9058	0.9010	**0.8981**	0.9326
SAST-AHC-PSNR	0.9298	0.8616	**0.9053**	**0.9574**	SAST-AHC-SSIM	0.9475	**0.9372**	0.8968	0.9278
OSS-PSNR	**0.9328**	**0.8621**	**0.9041**	**0.9590**	OSS-SSIM	**0.9499**	**0.9371**	**0.9031**	**0.9352**

TABLE VIII
PERFORMANCE INDICES OF TESTING IQA APPROACHES ON THE TID2008 DATABASE. WE BOLD THE TOP TWO METRICS

Models	PLCC	SRCC	KRCC	AAE	RMS	Models	PLCC	SRCC	KRCC	AAE	RMS
PSNR	0.7617	0.7718	0.5686	0.6708	0.8566	SSIM	0.7097	0.7270	0.5270	0.7226	0.9314
D-PSNR	0.8934	0.9097	0.7465	0.4209	0.5940	D-SSIM	0.8525	0.8742	0.6763	0.5409	0.6911
SAST-PSNR	0.8978	0.9111	0.7480	0.4130	0.5823	SAST-SSIM	0.8659	0.8836	0.6883	0.5212	0.6613
AHC-PSNR	**0.8988**	**0.9137**	**0.7533**	**0.4067**	**0.5795**	AHC-SSIM	0.8613	0.8872	0.6931	0.5158	0.6717
OSS-PSNR	**0.9033**	**0.9174**	**0.7558**	**0.4039**	**0.5672**	OSS-SSIM	**0.8753**	**0.8967**	**0.7064**	**0.5021**	**0.6393**
MS-PSNR	0.8855	0.9017	0.7308	0.4485	0.6143	MS-SSIM	0.8617	0.8792	0.6842	0.5323	0.6708
IW-PSNR	0.8799	0.8937	0.7079	0.4974	0.6281	IW-SSIM	**0.9101**	**0.9044**	**0.7369**	**0.4086**	**0.5478**
PAMSE [40]	0.8891	0.9162	0.7566	0.4485	0.6050	SMSE [40]	0.8457	0.8659	0.6764	0.5337	0.7054

In this article, we set the parameters used in the proposed OSS model as follows, $\phi = 10$, $k = 2$, $\psi = 2$, $D_0 = 512$, $\alpha = 2$ and $\beta = 2$. Due to the usage of many parameters, we want to further discuss their sensitivity. For each of six parameters, we enumerate four numbers in a proper interval around the assigned value while fixing other five parameters. SRCC is adopted here since it is one of the most popular performance index and has been widely used to find the suitable parameters in quite a few IQA metrics [3], [4]. We report the results of the sensitivity of parameters on the VDID2014 database in Table VI. The values used in our OSS model are emphasized. It is apparent that the proposed model has a stable performance when the used parameters change, and thus it is robust and tolerant to varying values of parameters.

Another comparison is conducted using two other versions of the proposed algorithm. The first version only utilizes the modified SAST model (dubbed as MOD-SAST-PSNR and MOD-SAST-SSIM) for preprocessing input image signals, while the second one is composed of the original SAST model and the AHC model (dubbed as SAST-AHC-PSNR and SAST-AHC-SSIM). We indicate the five performance evaluations of the aforesaid two versions as well as the original PSNR/SSIM, SAST-PNSR/SSIM, AHC-PSNR/SSIM and OSS-PSNR/SSIM on LIVE, IVC, Toyama and VDID2014 databases in Table VII. Two important conclusions can be derived: 1) the two versions tested both contribute to resulting performance of the proposed OSS to some extent on particular database; 2) the second version stated above largely advances the performance of SSIM on the Toyama database.

A cross-validation experiment is also implemented by using the large-scale TID2008 database, which is composed of 1,700 images by corrupting 25 reference images with 17 distortion types at 4 distortion levels [34]. In this work we only validate the IQA performance for the first 13 categories of structural distortions and add a new perceptual fidelity aware MSE [40] for comparison. Since there is not a definite viewing distance for the TID2008 database, we suppose its distance value to be the commonly used three times the image height. Table VIII lists the performance measures of the original PSNR/SSIM, D-PSNR/SSIM, SAST-PSNR/SSIM, AHC-PSNR/SSIM, OSS-PSNR/SSIM, MS-PSNR/SSIM, IW-PSNR/SSIM and SMSE, PAMSE. The results confirm that the proposed OSS model has derived greatly high accuracy. In particularly, our technique improves PSNR to a large extent, outperforming the classical MS and recently designed IW, SMSE and PAMSE models.

We finally want to mention three important points. First, an approximately optimal low-pass filter based on the retina model may be used, but it is very complex and costs quite a lot of time [41]. So we in this paper design the simple and effective OSS model, aiming to find a good tradeoff between the performance accuracy and the computational complexity. Second, the modified SAST model is an isotropic filter for simulating the process that the viewing angle shrinks in a gradual way with the viewing distance increased, while the anisotropic AHC model beforehand preserves the horizontal and vertical directions. Hence the concatenation of the AHC before the modified SAST, first filtering out high-frequency information while preserving the horizontal and vertical directions followed by resizing the image to the optimal scale, can achieve such high performance. Third, a more reasonable prefiltering model should include visual saliency [42], which will be explored in the future work.

V. Summarization

Several contributions have been made in this work. First, this is the first paper that comprehensively analyzes the impact of viewing distance and image resolution on visual quality in light of subjective and objective assessments. Second, the proposed OSS model is simple. Third, this framework is robust because it takes advantage of basic attributes of the HVS in both DWT and spatial domains. Fourth, the proposed metric accurately predicts visual quality of images shown at a far viewing distance. Fifth, our algorithm is robust and tolerant to varying values of parameters used.

VI. Conclusion

This paper has investigated the problem of the influence of viewing distance and image resolution on IQA performance. First, we introduced a new dedicated viewing distance-changed image database (VDID2014), which includes 160 images with two typical image resolutions and associated 320 subjective scores obtained at two typical viewing distances. Next, we developed a novel optimal scale selection (OSS) model to deal with this problem. In order to validly remove the undiscernible details that are caused by the varying viewing distance and image resolution, this model works via a cascade of adaptive high-frequency clipping in the DWT domain and adaptive resolution scaling in the spatial domain. A comparison of the proposed OSS model with a large set of IQA approaches are conducted on LIVE, IVC, Toyama, VDID2014 and TID2008 databases. Experimental results are provided to demonstrate the effectiveness of the OSS method. Furthermore, it is worth emphasizing that our metric is not limited in FR IQA, but also possibly extended to improving RR and NR IQA metrics. Matlab codes of our algorithm and the VDID2014 database will be available at https://sites.google.com/site/guke198701/publications.

References

[1] 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.

[2] 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.

[3] 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, 2015.

[4] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 45, 2015.

[5] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[6] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2013–2026, Dec. 2013.

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

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

[9] S. Wang, X. Zhang, S. Ma, and W. Gao, "Reduced reference image quality assessment using entropy of primitives," in *Proc. Pict. Coding Symp.*, San Jose, CA, USA, Dec. 2013, pp. 193–196.

[10] D. Liu, Y. Xu, Y. Quan, and P. Le Callet, "Reduced reference image quality assessment using regularity of phase congruency," *Signal Process. Image Commun.*, vol. 29, no. 8, pp. 844–855, Sep. 2014.

[11] D. Liu, Y. Xu, Y. Quan, Z. Yu, and P. Le Callet, "Directional regularity for visual quality estimation," *Signal Process.*, vol. 110, pp. 211–221, May 2015.

[12] Y. K. Lai and C.-C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," *J. Vis. Commun. Image Represent.*, vol. 11, no. 1, pp. 17–40, Mar. 2000.

[13] 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.

[14] 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.*, vol. 2. Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

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

[23] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2007). *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality

[24] A. Ninassi, P. Le Callet, and F. Autrusseau. (2006). *Subjective Quality Assessment-IVC Database*. [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb

[25] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad. (2008). *MICT Image Quality Evaluation Database*. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[26] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, Int. Telecommun. Union Recomm., Geneva, Switzerland, 2012.

[27] *Methodology for the Subjective Assessment of Video Quality in Multimedia Applications*, document ITU-R BT.1788, Int. Telecommun. Union Recomm., Geneva, Switzerland, 2007.

[28] *Subjective Methods for the Assessment of Stereoscopic 3DTV Systems*, document ITU-R BT.2021, Int. Telecommun. Union Recomm., Geneva, Switzerland, 2012.

[29] *General Viewing Conditions for Subjective Assessment of Quality of SDTV and HDTV Television Pictures on Flat Panel Displays*, document ITU-R BT.2022, Int. Telecommun. Union Recomm., Geneva, Switzerland, 2012.

[30] A. K. Moorthy, L.-K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[31] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Self-adaptive scale transform for IQA metric," in *Proc. IEEE Int. Symp. Circuits Syst.*, Beijing, China, May 2013, pp. 2365–2368.

[32] K. Gu *et al.*, "Adaptive high-frequency clipping for improved image quality assessment," in *Proc. IEEE Vis. Commun. Image Process.*, Kuching, Malaysia, Nov. 2013, pp. 1–5.

[33] W. A. N. Dorland, *Dorland's Medical Dictionary*. Philadelphia, PA, USA: Saunders Press, 1980.

[34] N. Ponomarenko *et al.*, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.

[35] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[36] E. B. Goldstein, *Sensation and Perception*. Belmont, CA, USA: Wadsworth, 1979.

[37] R. J. W. Mansfield, "Neural basis of the orientation preference in primates," *Science*, vol. 186, no. 4169, pp. 1133–1135, 1974.

[38] K. Friston, "The free-energy principle: A unified brain theory?" *Nat. Rev. Neurosci.*, vol. 11, pp. 127–138, Feb. 2010.

[39] VQEG. (Mar. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: http://www.vqeg.org/

[40] W. Xue, X. Mou, L. Zhang, and X. Feng, "Perceptual fidelity aware mean squared error," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 705–712.

[41] G. Zhai, A. Kaup, J. Wang, and X. Yang, "Retina model inspired image quality assessment," in *Proc. IEEE Vis. Commun. Image Process.*, Kuching, Malaysia, Nov. 2013, pp. 1–6.

[42] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Visual saliency detection with free energy theory," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1552–1555, Oct. 2015.

Ke Gu received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently pursuing the Ph.D. degree.

In 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, for five months and the School of Computer Engineering, Nanyang Technological University, Singapore, from 2014 to 2015. His research interests include quality assessment and contrast enhancement. He is the Reviewer for some IEEE TRANSACTIONS and journals, including the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE SIGNAL PROCESSING LETTERS, *Neurocomputing*, the *Journal of Visual Communication and Image Representation*, and *Signal, Image and Video Processing*.

Min Liu received the B.E. degree in electronic engineering from Xidian University, Xi'an, China, in 2012. She is currently pursuing the Ph.D. degree with Shanghai Jiao Tong University. Her research interests include image/video quality assessment and perceptual signal processing.

Guangtao Zhai 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.

He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. From 2006 to 2007, he was a Student Intern with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore, and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, from 2008 to 2009, 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. His research interests include multimedia signal processing and perceptual signal processing. He was a recipient of the National Excellent Ph.D. Thesis Award from the Ministry of Education of China in 2012.

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 the Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From 2000 to 2002, he was a Research Fellow with the Centre for Signal Processing, Nanyang Technological University, Singapore. From 2002 to 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. His current research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition. He has published over 80 refereed papers and holds six patents. He actively participates in the International Standards such as MPEG-4, JVT, and MPEG-21. He was a recipient of the Microsoft Young Professorship Award in 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing in 2003, and awards from A-STAR and Tan Kah Kee foundations. He is a member of the Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society.

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 with Philips Kommunikation Industrie AG, 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 with 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. His main research interests include digital video coding and transmission, multimedia semantic processing and intelligent video surveillance. He holds over 40 patents and has published over 90 papers in international journals and conferences. He is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television, an industry/government consortium in DTV technology research and standardization. Prof. Zhang is the Chair of Future of Broadcast Television Initiative Technical Committee.