A Fast Reliable Image Quality Predictor by Fusing Micro- and Macro-Structures

Ke Gu, Leida Li, *Member, IEEE*, Hong Lu, Xiongkuo Min, and Weisi Lin, *Fellow, IEEE*

***Abstract*—A fast reliable computational quality predictor is eagerly desired in practical image/video applications, such as serving for the quality monitoring of real-time coding and transcoding. In this paper, we propose a new perceptual image quality assessment (IQA) metric based on the human visual system (HVS). The proposed IQA model performs efficiently with convolution operations at multiscales, gradient magnitude, and color information similarity, and a perceptual-based pooling. Extensive experiments are conducted using four popular large-size image databases and two multiply distorted image databases, and results validate the superiority of our approach over modern IQA measures in efficiency and efficacy. Our metric is built on the theoretical support of the HVS with lately designed IQA methods as special cases.**

***Index Terms*—Color information, gradient operator, perceptual image quality assessment (IQA), pooling, structure.**

I. Introduction

QUALITY assessment and monitoring play a crucial role in various respects of scientific research and applicational development. For the sake of the continuous increase in complexity and expense of systems and softwares, quality degradation is more easily introduced than ever before. On one hand, the issue of fault diagnosis and detection, particularly quality related, have aroused much attention very recently [1]–[6], and on the other hand, some significant progress has been made in the study of quality-based optimization and visual enhancement [7]–[10].

In this paper, we investigate the problem of full-reference (FR) image quality assessment (IQA), which can be used for the quality monitoring in real-time coding and transcoding systems replacing the traditional manual labor. It has come to a broad agreement that subjective assessment is the decisive criterion for a given image/video signal, since human beings are usually the ultimate consumers. Subjective assessment is however costly, time consuming, and labor intensive, and it turns out to be hardly incorporated into practical applications. Therefore, a vast majority of researches has been devoted to the exploration of objective IQA metrics toward simulating subjective opinion scores via mathematical models [11].

One type of classical solution to modeling FR-IQA models resort to the measurement of structural similarity between the original and distorted images [12] and its variants improved from different perspectives [13]–[16]. The second type of FR-IQA models were devised using other advanced tactics, e.g., assuming that the human visual system (HVS) implements two distinct manners when predicting the image quality [17]–[19]. It was found lately that low-level visual features, especially the image gradient magnitude (GM), perform effectively in image quality evaluation. On this basis, several prevailing IQA methods were proposed by combining the GM with visual saliency, masking effects, and free energy theory systematically [20]–[24].

Despite the successfulness of numerous IQA methods, the pursuit of more effective and efficient algorithms is never changed. To this end, we in this paper come up with a novel image quality metric based on Perceptual SIMilarity (PSIM) measure. To specify, considering some prior knowledge about the human perception to visual quality, our PSIM algorithm works with four steps. First, we extract the GM maps of the input original and distorted images inspired from the fact that the center-surround cell in eyes conducts a local comparison exhibiting a lateral inhibition [25] and moreover the GM implements better in IQA tasks than other frequently used local operators such as Laplacian of Gaussian (LOG). The second and third steps in the proposed IQA method are to compute the similarity of GM maps at multiscales (MS) since the HVS usually employs MS decomposition [26], [27], and to measure the chromatic channel degradation caused by artifact injection due to the important influence of chromatic information in visual perception [28], [29]. In the fourth step, a reliable perceptual-based pooling groups visual degradation measures above to predict the objective quality.

Manuscript received December 31, 2015; revised September 21, 2016 and November 2, 2016; accepted November 12, 2016. Date of publication January 16, 2017; date of current version April 10, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61305011 and Grant 61379143, and in part by the Qing Lan Project of Jiangsu Province. *(Corresponding author: Hong Lu.)*

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, BJUT Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

L. Li is with the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: lileida@cumt.edu.cn).

H. Lu is with the School of Automation, Nanjing Institute of Technology, Nanjing 211167, China (e-mail: luhong2006@yahoo.com).

X. Min is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: minxiongkuo@sjtu.edu.cn).

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

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

Digital Object Identifier 10.1109/TIE.2017.2652339

0278-0046 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The remainder of this paper is organized as follows. In Section II, our IQA metric is introduced in detail. Section III provides results of experiments on six image databases (LIVE [30], CSIQ [18], TID2008 [31], TID2013 [32], LIVEMD [33], and MDID2013 [34]) for demonstrating the superiority and efficiency of the proposed PSIM model. We summarize the contributions and draw the conclusions in Sections IV and V.

II. Proposed PSIM Method

A. Gradient Magnitude Extractor

During the last decades, it has been widely accepted that structural information, particularly the image gradient, serves as one of the most significant function in evaluating perceived quality [20]–[24]. To be more concretely, given an image signal $\mathbf{s}$, the famous Prewitt operator [35] is applied to the gradient extraction as

$$\mathbf{g} = \sqrt{\mathbf{g}_h^2 + \mathbf{g}_v^2} \tag{1}$$

where $\mathbf{g}_h = \mathbf{p}_h \star \mathbf{s}$ and $\mathbf{g}_v = \mathbf{p}_v \star \mathbf{s}$ with the symbol "$\star$" being the convolution operation. $\mathbf{p}_h = [1, 0, -1; 1, 0, -1; 1, 0, -1]$ and $\mathbf{p}_v = \mathbf{p}_h^T$, separately indicating Prewitt convolution masks in the horizontal and vertical directions. In Fig. 1, we display an intuitive example including five images chosen from the TID2013 database [32] and their corresponding GM maps.

As a matter of fact, given a visual signal input, the center-surround cell conducts a local comparison, which exhibits the lateral inhibition; that is to excite light in the center while inhibiting light in the surround. The used gradient operator can be treated as a fast and effective manner to perform the local comparison. Other similar operators, such as LOG, Sobel, and Scharr operators, can be also applied here but instead, by tests, bring at least 2% performance reduction. So the Prewitt operator is used in our PSIM metric.

B. Gradient Magnitude Similarity

Similarity measure is conducted on the GM maps between the original and distorted images. According to suggestions given in psychophysical masking studies [26] and neuropsychological recordings in [27], there exist mechanisms selective to narrow ranges of spatial frequencies and orientations. To mimic the above-mentioned function which possibly happens in the visual cortex, MS decompositions are commonly used in building visual models, e.g., in the IQA study [13]. But that MS strategy is not efficient enough for real-time applications, which encourages us to develop a concise and practical MS model instead.

1) Microstructural Similarity: In the design of our approach, a two-scale model, due to its efficacy and efficiency, is taken into account. First, we use the Prewitt operator to convolve the original and distorted images twice at a large scale where an image is processed by a low-pass filter and then down sampled by a proper factor of 2. We denote the original image and its distorted one as $\mathbf{o}$ and $\mathbf{d}$, and the associated GM maps as $\dot{\mathbf{g}}_o$ and $\dot{\mathbf{g}}_d$. Five sample images are exhibited in Fig. 1(a)–(e), which are chosen from the TID2013 database and composed of the source image and four distorted images corrupted by "white noise," "Gaussian blur," "JPEG compression" and "local block-wise distortions of different intensity," and their GM maps in Fig. 1(f)–(j). As seen, the features $\dot{\mathbf{g}}_o$ and $\dot{\mathbf{g}}_d$ succeed in conducting local comparisons.

In practice, it was found that most existing IQA methods implement this downsampling procedure before predicting visual quality [12], [20], [23], [24]. In our recent work [15], we managed to search for the optimal scale according to the image size and the given viewing distance. By analysis in the spatial domain, an appropriate scaling coefficient z is approximated to be the square root of the ratio of the image size and the focused visual scope

$$\begin{aligned} z &= \sqrt{\frac{\mathcal{A}}{\tilde{\mathcal{A}}}} = \sqrt{\frac{h * w}{\tilde{h} * \tilde{w}}} \\ &= \sqrt{\frac{1}{4\tan\left(\frac{\theta_h}{2}\right) * \tan\left(\frac{\theta_w}{2}\right)} * \left(\frac{h}{d}\right)^2 * r} \end{aligned} \tag{2}$$

where d is the viewing distance; $\mathcal{A}$ and $\tilde{\mathcal{A}}$ stand for the image size and the visual scope at the given d; h and w are the image height and width; $\tilde{h}$ and $\tilde{w}$ are the visual height and width; θ_h and θ_w separately indicate horizontal and vertical visual angles; r represents the aspect ratio of image width to height, i.e., $r = \frac{w}{h}$. With empirical parameters about visual angles[1] and reasonable hypothesis about image sizes and the viewing distance,[2] the scaling coefficient z is found to be 0.4955 ($\approx$0.5). So the used downsampling factor "2" can be considered as approximately optimal in typical viewing conditions.

Next, we employ the frequently used similarity measure to detect the difference of GM maps between the original image and its contaminated version at a large scale stated above

$$\dot{\mathbf{g}}_m(\mathbf{o}, \mathbf{d}) = \frac{2\,(\dot{\mathbf{g}}_o) * (\dot{\mathbf{g}}_d) + c_1}{(\dot{\mathbf{g}}_o)^2 + (\dot{\mathbf{g}}_d)^2 + c_1} \tag{3}$$

where c_1 is a positive fixed number to avoid the instability. This similarity metric has the following three advantages:

1) symmetry: $\dot{\mathbf{g}}_m(\mathbf{o}, \mathbf{d}) = \dot{\mathbf{g}}_m(\mathbf{d}, \mathbf{o})$;
2) boundedness: $\dot{\mathbf{g}}_m(\mathbf{o}, \mathbf{d}) \leq 1$;
3) unique maximum: $\dot{\mathbf{g}}_m(\mathbf{o}, \mathbf{d}) = 1$ if and only if $\mathbf{o} = \mathbf{d}$.

In Fig. 2, we display four $\dot{\mathbf{g}}_m$ maps associated to Fig. 1(b)–(e) with respect to Fig. 1(a), where the brighter gray level means the higher similarity and thus the lower-level distortion. It can be viewed that $\dot{\mathbf{g}}_m$, the GM maps' difference at a large scale, is good at reflecting the variations in microstructure (e.g., details) of the original and distorted images.

2) Macrostructural Similarity: Visual saliency has long been applied to numerous valuable applications, for instance, tone mapping operation [37], contrast enhancement [7], [8], fixation prediction [38], visual tracking [39]–[41], quality evaluation [10], [24], etc. However, in [24], it was viewed that

[1] Horizontal and vertical visual angles are generally assigned to be about 120° and 150° [36]. Since the real view angle becomes narrower to about one third of the common value when one concentrates on the details of an image and scores it. The horizontal and vertical visual angles can be reasonably considered to be 40° and 50°, respectively.

[2] In this paper, we suppose that typical images are of the aspect ratio 3:2 and the general viewing distance is about three times the image height, as used in most image databases.

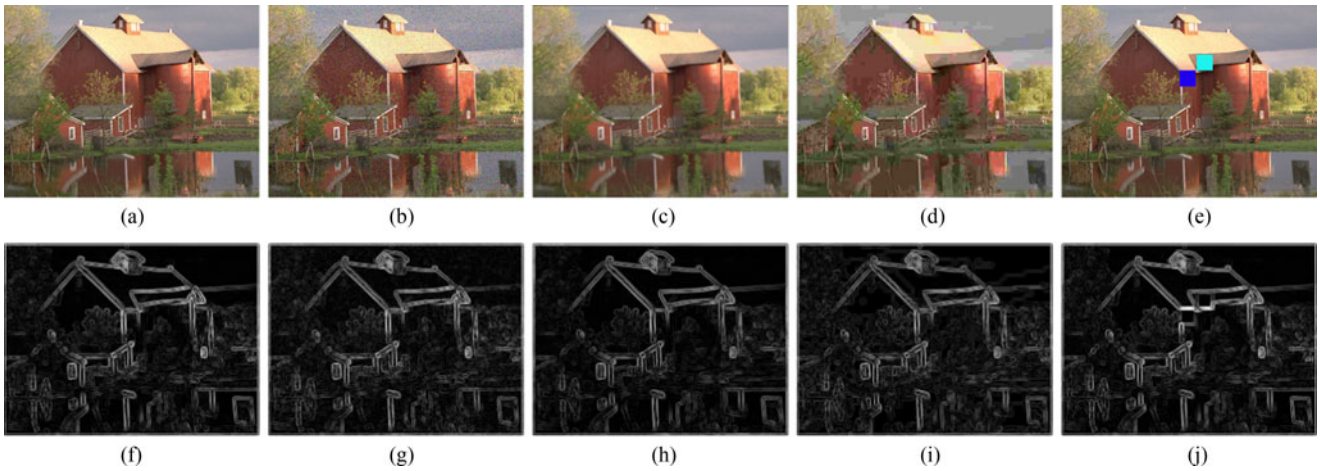

Fig. 1. Sample images from the TID2013 database [32] and associated GM maps: (a) original image; (b) white noise; (c) Gaussian blur; (d) JPEG compression; (e) local block-wise distortions of different intensity; (f)–(j) GM maps of (a)–(e).

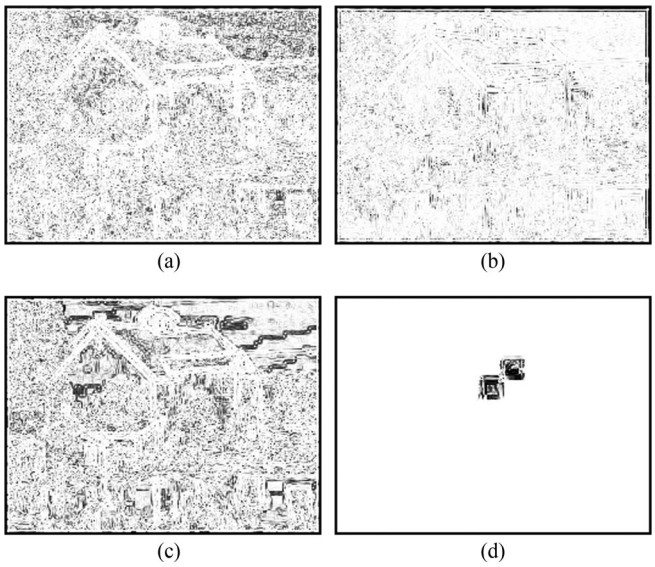

Fig. 2. Illustration of $\dot{\mathrm{g}}_m$ maps corresponding to Fig. 1(b)–(e).

state-of-the-art saliency detection techniques that predict fixations better, e.g., image signature (IS) model [42], perform worse in assessing image quality as compared to saliency detection models which have inferior fixation predictions, e.g., Itti model [44]. In general, visual saliency theory tells that humans pay more attention to salient regions in an image, but from the images shown in Fig. 1(b) and (d), we can see that the artifacts which play a significant role in degrading the visual quality are located in the smooth "sky" regions at the top of images instead of salient "chalet" area in the middle. So we have a reason to suppose there may exist a mechanism which takes effect and is related to visual saliency. We hypothesize that this mechanism mainly concerns the variations in macrostructure (e.g., contours) of images.

To measure such variations in macrostructure, we compute the GM maps at a small scale to compare the difference between the original and distorted images, and denote the associated GM map of the original image as $\ddot{\mathrm{g}}_o$ and that of the distorted one as $\ddot{\mathrm{g}}_d$. Afterward, we exert a distance metric, which is a similarity measure typically used in the computer vision field, on the GM maps at a small scale

$$\ddot{\mathrm{g}}_m(\mathbf{o},\mathbf{d}) = \frac{\min(\ddot{\mathrm{g}}_o,\ddot{\mathrm{g}}_d)+c_2}{\max(\ddot{\mathrm{g}}_o,\ddot{\mathrm{g}}_d)+c_2} \tag{4}$$

where c_2 is for stability when the denominator is extremely small. Similarly, this distance metric also has three advantages of symmetry, boundedness, and unique maximum.

The GM map extracted at a small scale characterizes the macrostructure of an image.[3] We normalize the GM map using a Gaussian kernel to smooth it. It was found that the smoothed GM map basically detects the salient regions in an image, and the result is highly similar to the saliency map computed by the state-of-the-art IS model [42]. We take an example to illustrate this. Five sample images of the same size 681×511 were chosen from the classical Toronto dataset [45]. These five images include various scenes, as given in the leftmost column in Fig. 3. The smoothed GM map at a small scale[4] are given in the middle column in Fig. 3, and the saliency maps computed using the popular model [42] are used for comparison, as given in the rightmost column in Fig. 3. The phenomenon is surely not a coincidence. In [46], the authors have revealed that, for natural images, a gradient operator combined with Gaussian postprocessing at a small scale is capable of searching for salient objects, just as the examples provided in Fig. 3. Experimental results demonstrate that our proposed metric is superior to the recently developed visual saliency-based quality index (VSI) [24]. Based on the analysis and results stated above, we believe that the reason why using visual saliency is benefit for raising IQA performance is that the macrostructure is included in the detected salient areas.

[3]In this paragraph, "the GM map" means "the GM map extracted at a small scale." We omit "extracted at a small scale" for simplicity.

[4]In this implementation, we follow a popular transforming coefficient used in the popular model [42], [43], and resize the image to a coarse 64×48 pixel representation.

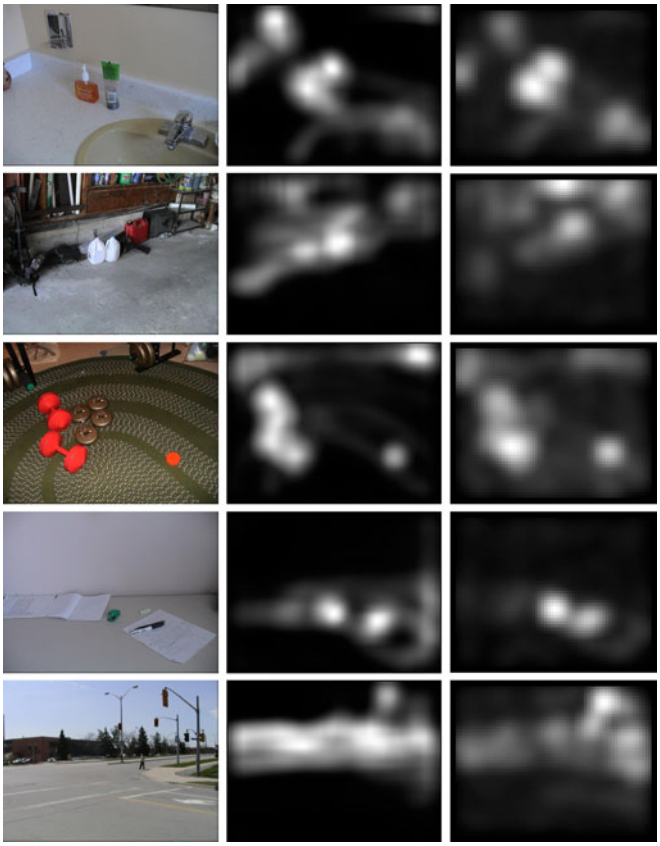

Fig. 3. Comparison of saliency detection algorithms: The leftmost column shows the representative images in the Toronto dataset [45], the middle shows the GM maps computed at a small scale, and the rightmost shows the saliency maps detected by using [42].

The design principle behind our proposed PSIM metric is an improved version of the concept of the MS structure. It is reasonable that humans implement two typical perceptions when understanding a visual input. One is the basic perception, which is used to estimate the macrostructure, e.g., the completeness of objects; the other is the detailed perception, which is used to measure the microstructure, e.g., the sharpness of edges. Toward keeping the design principle consistent with the typical "multiscale," we associate structures in "large scale" and "small scale" with basic perception and detailed perception of humans.

C. Color Information Similarity

Since color information plays a significant part in visual perception [28], [29], we also take account of the variations in chrominance channels of the distorted image compared to its original one. Before computing the GM maps, the simple and widely used YIQ color space [47] is applied to transfer an input RGB color image

$$\begin{cases} \mathbf{y} = 0.299\,\mathbf{r} + 0.587\,\mathbf{g} + 0.114\,\mathbf{b} \\ \mathbf{i} = 0.596\,\mathbf{r} - 0.274\,\mathbf{g} - 0.322\,\mathbf{b} \\ \mathbf{q} = 0.211\,\mathbf{r} + 0.523\,\mathbf{g} + 0.312\,\mathbf{b} \end{cases} \tag{5}$$

where $\mathbf{y}$ conveys the luminance information and $\mathbf{i}$ and $\mathbf{q}$ the chrominance information. In this study, we use $\mathbf{y}$ to compute the GM maps at two (large and small) scales, and use $\mathbf{i}$ and $\mathbf{q}$ to measure the chromatic distinction between the original and distorted images as follows:

$$\mathbf{i}_m(\mathbf{o},\mathbf{d}) = \frac{2\,(\mathbf{i}_o) * (\mathbf{i}_d) + c_3}{(\mathbf{i}_o)^2 + (\mathbf{i}_d)^2 + c_3} \tag{6}$$

$$\mathbf{q}_m(\mathbf{o},\mathbf{d}) = \frac{2\,(\mathbf{q}_o) * (\mathbf{q}_d) + c_3}{(\mathbf{q}_o)^2 + (\mathbf{q}_d)^2 + c_3} \tag{7}$$

where $\mathbf{i}_o$ and $\mathbf{i}_d$ ($\mathbf{q}_o$ and $\mathbf{q}_d$) indicate $\mathbf{i}$ ($\mathbf{q}$) chromatic channels of images $\mathbf{o}$ and $\mathbf{d}$. c_3 is of similar function to c_1 and c_2.

D. Perceptual-Based Pooling

After implementing GM similarity measures at two scales, how to pool them validly is still a tough problem. A natural strategy is the simplest global average, defined as follows:

1) for GM similarity at a large scale:

$$L^{\alpha} = \frac{1}{U}\sum_{u=1}^{U}(\dot{\mathbf{g}}_m(o_u, d_u))^{\alpha}; \tag{8}$$

2) for GM similarity at a small scale:

$$S^{\beta} = \frac{1}{V}\sum_{v=1}^{V}(\ddot{\mathbf{g}}_m(o_v, d_v))^{\beta}; \tag{9}$$

3) for color information similarity:

$$C^{\theta} = \frac{1}{W}\sum_{w=1}^{W}(\mathbf{i}_m(o_w, d_w)\,\mathbf{q}_m(o_w, d_w))^{\theta} \tag{10}$$

where α, β, and θ are three positive fixed numbers; U, V, and W represent the total number of elements in the vector.

The pooling schemes were however found to work poorly, since they treat the distortions of different types and locations equally. To address this problem, it is natural to also consider the high-distortion-based pooling (i.e., quality-based pooling), whose intuitive idea is to emphasize high-distortion regions. From the perspective of saliency, locations of high-distortion levels attract more viewers' attentions, resulting in a serious degradation on the perceived visual quality. As a result, we further make use of the high-distortion-based pooling which is defined by

$$H^{\tau} = \frac{1}{X}\sum_{x=1}^{X}(\dot{\mathbf{h}}_m(o_x, d_x))^{\tau} \tag{11}$$

where $\dot{\mathbf{h}}_m$ is the vector consisting of the smallest $\gamma\%$ values in $\dot{\mathbf{g}}_m$; τ is a positive constant of similar function to α, β, and θ; X is the number of elements in the vector $\dot{\mathbf{h}}_m$.

We finally fuse the aforementioned four features together and derive the PSIM metric (as seen in Fig. 4) to be

$$\mathrm{PSIM} = \frac{H^{\tau}}{L^{\alpha}} * S^{\beta} * C^{\theta} \tag{12}$$

where α, β, θ, and τ are fixed parameters. Toward reliably determining the parameters used in our PSIM algorithm, we deploy a two-phase strategy. In the first phase, we initialize those parameters using the state-of-the-art VSI metric and big-data

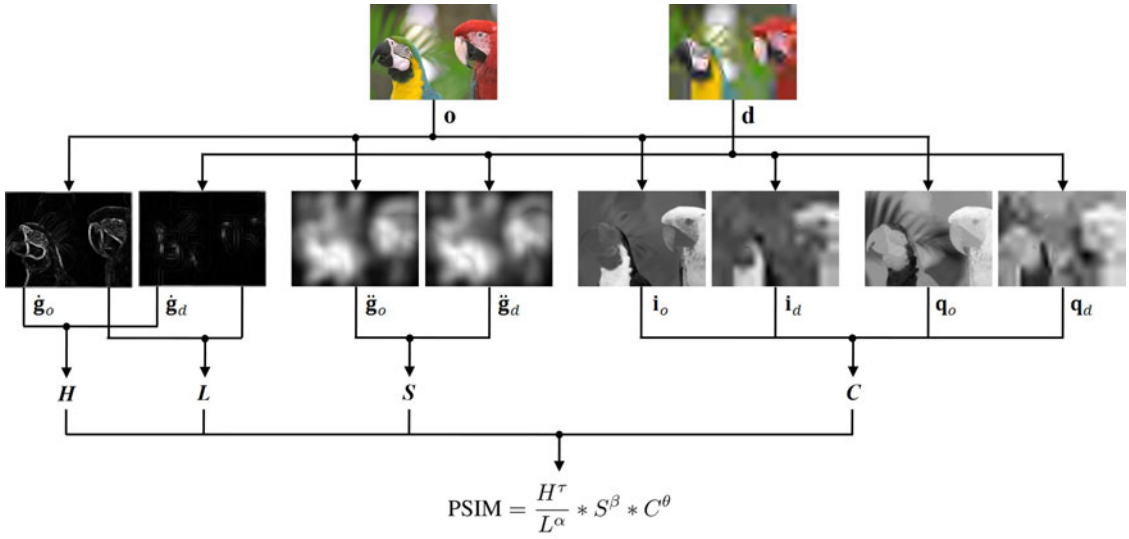

Fig. 4. Illustration for the computing scheme of our PSIM metric. $\mathbf{o}$ and $\mathbf{d}$ represent the original and distorted images.

images. To specify, we use 20 distortion types and 10 distortion intensities to corrupt the 500 pristine images in the Berkeley database [48] and thus generate 100 000 distorted images. We label those images with the quality scores of the VSI metric for determining the parameters used in our PSIM model. We adopt a robust and highly generalized regression method, maximum-rank order correlation coefficient estimator (MROCCE) [49], for validly estimating the model parameters. The MROCCE targets to make the Kendall's rank order correlation coefficient of the quality scores between our model and VSI metric maximized. More information can be directed to [49]. Having initialized the used parameters, in the second phase, we follow the method applied in [50] to fine tune the parameters used in our PSIM metric. A leave-one-out cross-validation procedure is implemented to ensure the robustness of the proposed PSIM approach. In terms of image contents, we divide the TID2013 database into 24 training sets, which includes 2880 images, and one testing set, which includes 120 images. We repeat the above division 25 times and use the MROCCE method to yield 25 groups of values of parameters. It was observed that the 25 groups of parameters are different but fairly close to each other, which indicates the robustness of the estimated parameters, and thus we eventually selected one group of parameters.

The design philosophy behind the above pooling scheme is not just intuitive. The values of α, β, θ, and τ are very small numbers. Hence we can approximate (15) based on the Chebyshev's inequality and derive

$$\text{PSIM} \approx \frac{H^\tau}{L^\tau} * L^\phi * S^\beta * C^\theta \tag{13}$$

where $\phi = \tau - \alpha > 0$. The first term indicates that for diverse images with the same L^τ, the more uneven distribution of distortion levels (gauged by H^τ) will give rise to the worse quality. The second and third terms reflect the variations of microstructure and macrostructure between the original and distorted images. And the last term estimates the difference in light of the color information. Moreover, via extensive tests, the proposed pooling scheme leads to 3%∼7% performance gain beyond popular existing pooling strategies [51]. Note that, as compared with a previous conference paper [52], this work supplements more features toward better performance, and meanwhile, provides detailed discussions about why choose these features, how to determine parameters, etc.

III. Experimental Results

A. Evaluation Protocols

In this paper, four large-scale databases with totally 6345 distorted images are used as testing beds for IQA evaluations. The first is the LIVE database [30], which was explored at the University of Texas at Austin in the year of 2006 and consists of 779 distorted images associated with 29 original versions and five typical distortion types. The second one is the CSIQ database [18], which was completed at Oklahoma State University in the year of 2009 and contains 866 images created by exerting six distortion types at four to five distortion intensities on 30 sources. The third and fourth databases are TID2008 [31] and TID2013 [32], built by a joint international effort across Finland, Italy, and Ukraine in 2008 and 2013. In TID2008, there are 1700 images in sum, by corrupting 25 references with 17 degradation types at four distortion levels. TID2013 covers up to 3000 images from the same 25 sources with 24 distortion types at 5 various intensities. We summarize the main information in Table I.

The performance of an IQA approach is typically evaluated from three aspects from the angle of prediction power [53]:

1) prediction accuracy;
2) prediction monotonicity;
3) prediction consistency.

The computation of the correlation performance requires a regression procedure to decrease the nonlinearity of predicted scores. Denoting $\mathbf{s}_o$, $\mathbf{s}_p$, and $\mathbf{s}_m$ as the vectors of the original

TABLE I
SUMMARIZATION OF MAIN INFORMATION ABOUT IMAGE QUALITY DATABASES TESTED

Databases	LIVE [30]	CSIQ [18]	TID2008 [31]	TID2013 [32]	LIVEMD [33]	MDID2013 [34]
Number of reference images	29	30	25	25	15	12
Number of distorted images	779	886	1700	3000	450	324
Number of distortion types	5	6	17	24	5	1
Number of observers	29	35	838	971	37	25
Image format	BMP	PNG	BMP	BMP	PNG	PNG
Establishing time	2007	2009	2008	2013	2012	2013
Establishing place	USA	USA	Finland	Finland	USA	China

IQA scores, the IQA scores after regression and the subjective MOS values, the logistic regression function is deployed for the nonlinear regression [53]

$$\mathbf{s}_p = \pi_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\pi_2 (\mathbf{s}_o - \pi_3)}} \right) + \pi_4 \mathbf{s}_o + \pi_5 \qquad (14)$$

where π_1 to π_5 are regression model parameters that are determined during the curve fitting process.

After the regression, five indices are used for performance measures [53]. The first and second indices are the Spearman rank order correlation coefficient (ξ_S) and the Kendall's rank order correlation coefficient (ξ_K) between $\mathbf{s}_p$ and $\mathbf{s}_m$, for evaluating the prediction monotonicity. The third one is the Pearson linear correlation coefficient (ξ_P) between $\mathbf{s}_p$ and $\mathbf{s}_m$, in order to estimate the prediction accuracy. The last two are the root mean square error (ξ_R) and the mean absolute error (ξ_M) between $\mathbf{s}_p$ and $\mathbf{s}_m$, to judge the prediction consistency. Of the above five indices, a good IQA model is expected to attain high values in ξ_S, ξ_K, and ξ_P, as well as low values in ξ_R and ξ_M.

B. Performance Measures

Two types of seven IQA models are used in our tests for comparison. The first type of metrics includes three classical PSNR, SSIM [12], and VSNR [17], which are extremely good at the quality prediction of images corrupted by commonly encountered distortion types, e.g., noise, blur, and compression. The second type includes four recently developed GSM [21], IGM [22], GMSD [23], and VSI [24].

As tabulated in Table II, the proposed PSIM algorithm has achieved significantly higher correlation performance on the popular four large-scale image databases. To specify, on the LIVE database, the accuracy of our method is beyond 95% in ξ_S and ξ_P. In terms of prediction monotonicity (ξ_S and ξ_K), the proposed PSIM metric is the best across the IQA metrics tested. In the comparison about prediction accuracy (ξ_P) and consistency (ξ_R), our approach has also attained the second-place performance results.

The obvious superiority of our PSIM method is validated on the TID2008 and CSIQ databases as well, beating other competing quality measures. The proposed model has gained very high prediction monotonicity and accuracy (ξ_S and ξ_P) of larger than 96% on the CSIQ database, and of larger than 90% on the TID2008 database. Our technique is clearly better than other competing IQA models.

On the TID2013 database, there is only our PSIM metric acquiring the prediction accuracy of beyond 90%. Apart from this, in the comparison of prediction consistency (ξ_R and ξ_M), our IQA metric has attained the top performance among all the tested IQA models. In the comparison of prediction monotonicity (ξ_S and ξ_K), our measure is just a little lower than the lately designed VSI metric, which combines various kinds of complicated models.

We further conduct comprehensive comparisons using two commonly seen average measures, as defined by

$$\bar{\delta} = \frac{\sum_i c_i \cdot f_i}{\sum_i f_i} \qquad (15)$$

where f_i and c_i $(i = 1, 2, 3, 4)$ indicate weighting parameters and correlation measures for four testing databases. For the first direct average, all of d_i are set as one. For the second database size-weighted average, we assign d_i as the number of images in each database, i.e., 779 for LIVE, 1700 for TID2008, 866 for CSIQ, and 3000 for TID2013. Table II lists the aforesaid two average results across the entire eight IQA measures. Note that there exist remarkable differences in the ranges of MOS/DMOS in distinct databases, which makes the comparison of prediction consistency (ξ_R and ξ_M) not fair, so we divide ξ_R and ξ_M by the maximum dynamic range of subjective scores in each database before computing the average results using (16). It is apparent that our PSIM method has constantly obtained better prediction accuracy on average performance comparison, largely superior to the second- and third-place VSI and GMSD algorithms.

C. Computational Cost

Table III provides the average run time of the eight IQA methods on the overall 3000 images of size 512×384 in the TID2013 database. The entire quality measures were run using the software platform of MATLAB R2010a (7.10.0) on a computer with 3.40 GHz CPU processor and 4.00 GB RAM. All the source codes of competing IQA methods were obtained from their authors or websites. It can be easily found that our PSIM model merely needs less than 40 ms, about three times faster than the recently explored VSI metric, which has achieved the second best correlation performance. The implementation time of each IQA algorithm tested on LIVE, CSIQ, and TID2008 databases are also reported in Table III for comparison. From the results, we can derive the similar conclusions.

TABLE II
PERFORMANCE MEASURES ON FOUR TESTING DATABASES AND TWO AVERAGES

Metrics	LIVE Database (779 images) [30]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.8756	0.6865	0.8723	13.360	10.509
SSIM [12]	0.9104	0.7311	0.9042	11.669	9.2279
VSNR [17]	0.9279	0.7624	0.9236	10.476	8.0446
GSM [21]	0.9561	0.8150	0.9512	8.4323	6.5787
IGM [22]	0.9581	0.8250	0.9570	7.9242	6.0101
GMSD [23]	**0.9603**	**0.8268**	**0.9603**	**7.6214**	**5.7892**
VSI [24]	0.9524	0.8058	0.9482	8.6812	6.8593
PSIM (Pro.)	**0.9622**	**0.8294**	**0.9584**	**7.7996**	**6.1452**

Metrics	CSIQ Database (866 images) [18]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.8033	0.6022	0.8000	0.1575	0.1195
SSIM [12]	0.8378	0.6343	0.8154	0.1520	0.1158
VSNR [17]	0.8109	0.6247	0.8005	0.1573	0.1161
GSM [21]	0.9108	0.7374	0.8964	0.1163	0.0847
IGM [22]	0.9403	0.7881	0.9281	0.0977	0.0715
GMSD [23]	**0.9570**	**0.8129**	**0.9541**	**0.0786**	**0.0586**
VSI [24]	0.9423	0.7857	0.9279	0.0979	0.0706
PSIM (Pro.)	**0.9621**	**0.8273**	**0.9642**	**0.0696**	**0.0528**

Metrics	TID2008 Database (1,700 images) [31]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.5834	0.4256	0.5734	1.0994	0.8327
SSIM [12]	0.6272	0.4562	0.6413	1.0297	0.8207
VSNR [17]	0.7044	0.5340	0.6818	0.9813	0.6904
GSM [21]	0.8504	0.6596	0.8422	0.7234	0.5438
IGM [22]	0.8901	0.7103	**0.8857**	**0.6231**	**0.4697**
GMSD [23]	0.8907	0.7092	0.8789	0.6402	0.4833
VSI [24]	**0.8979**	**0.7123**	0.8763	0.6466	0.4904
PSIM (Pro.)	**0.9120**	**0.7395**	**0.9078**	**0.5628**	**0.4195**

Metrics	TID2013 Database (3,000 images) [32]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.7027	0.5193	0.7017	0.8833	0.6567
SSIM [12]	0.6274	0.4563	0.6860	0.9020	0.7264
VSNR [17]	0.6971	0.5189	0.7136	0.8683	0.6506
GSM [21]	0.7946	0.6255	0.8463	0.6603	0.4941
IGM [22]	0.8097	0.6395	0.8561	0.6406	0.4811
GMSD [23]	0.8044	0.6339	0.8590	0.6346	0.4810
VSI [24]	**0.8964**	**0.7182**	**0.8999**	**0.5406**	**0.4052**
PSIM (Pro.)	**0.8925**	**0.7161**	**0.9080**	**0.5193**	**0.3926**

Metrics	Direct Average (6,345 images)				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.7412	0.5584	0.7368	0.1358	0.1033
SSIM [12]	0.7507	0.5695	0.7617	0.1291	0.1017
VSNR [17]	0.7851	0.6100	0.7799	0.1251	0.0922
GSM [21]	0.8780	0.7094	0.8841	0.0946	0.0709
IGM [22]	0.8996	0.7407	0.9067	0.0849	0.0634
GMSD [23]	0.9031	0.7457	0.9131	0.0798	0.0602
VSI [24]	**0.9223**	**0.7555**	**0.9131**	**0.0837**	**0.0630**
PSIM (Pro.)	**0.9322**	**0.7780**	**0.9346**	**0.0713**	**0.0543**

Metrics	Database Size-Weighted Average				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.7057	0.5260	0.7017	0.1338	0.1009
SSIM [12]	0.6908	0.5143	0.7185	0.1300	0.1033
VSNR [17]	0.7430	0.5673	0.7427	0.1255	0.0924
GSM [21]	0.8452	0.6732	0.8650	0.0947	0.0709
IGM [22]	0.8673	0.7015	0.8863	0.0868	0.0651
GMSD [23]	0.8675	0.7022	0.8898	0.0840	0.0635
VSI [24]	**0.9100**	**0.7366**	**0.9033**	**0.0817**	**0.0614**
PSIM (Pro.)	**0.9158**	**0.7514**	**0.9218**	**0.0725**	**0.0549**

We bold the best two performed algorithms.

TABLE III
COMPARISON OF COMPUTATIONAL COST (MILLISECONDS/IMAGE) ON THE FOUR TESTING IMAGE DATABASES

Metrics	PSNR	SSIM	VSNR	GSM	IGM	GMSD	VSI	PSIM
TID2013	1.60	19.9	175	11.3	10 784	3.95	116	38.1
LIVE	3.76	34.5	269	19.1	17 641	6.84	154	62.8
CSIQ	2.07	24.7	214	13.8	13 163	5.15	131	47.5
TID2008	1.51	19.7	179	11.3	9512	3.95	114	38.1

D. Intuitive Comparison

We also show the scatter plots of subjective DMOS values versus objective quality predictions of the eight testing IQA algorithms on the CSIQ database, as illustrated in Fig. 5. For a comprehensive comparison, in each scatter plot the sample points corresponding to different distortion types are discriminated by distinct colors: red for additive white Gaussian noise (AWGN); green for JPEG compression; blue for JPEG2000 compression (JP2K); cyan for additive pink Gaussian noise (APGN); magenta for Gaussian blur; yellow for global contrast decrements (GCD). Notice that a good IQA method should predict the visual quality consistently across distinct types of distortions. Referring to Fig. 5, it can be seen that the scatter plot of the PSIM is more concentrated across various groups of distortion categories, and thus has better consistency. For instance, its sample points associated to the GCD distortion are near to other five distortion types. Nonetheless, the sample points associated to the GCD distortion for other testing IQA models are relatively far from other five categories. The proposed IQA model is therefore able to acquire such high performance accuracy.

E. Validation on Multiple Distortions

During the last few years, much attention has been shifted to the quality evaluation of multiply distorted images. Two databases (LIVEMD [33] and MDID2013 [34]) are further applied for performance comparison. LIVEMD was the first database for this issue, completed at the University of Texas at Austin in the year of 2012. A pair of image subsets are separately produced adding white noise or JPEG compression to blurred images, making up 225 images in each subset and a total of 450 images in the database. MDID2013 was explored at Shanghai Jiaotong University in the year of 2013. It consists of 324 images produced by successively using Gaussian blur, JPEG compression, and additive white noise to corrupt ten original images of size 768×512 for one half and of size 1280×720 for the rest. For convenience, interested readers can be directed to Table I for main information of the above-mentioned two databases.

We list the results of correlation performance in Table IV. On the LIVEMD database, the scores of prediction accuracy (ξ_P) and consistency (ξ_R) for the proposed PSIM model are higher than 0.85, better than or equivalent to other IQA measures tested. In comparison, on the MDID2013 database, our PSIM method

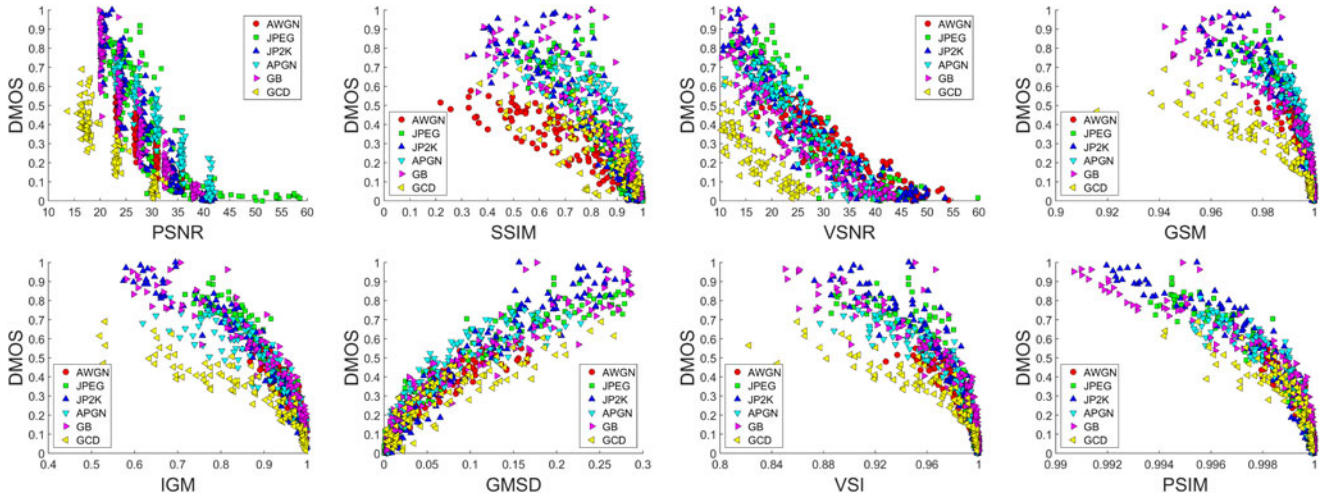

Fig. 5. Scatter plots of DMOS versus PSNR, SSIM, VSNR, GSM, IGM, GMSD, VSI and our PSIM metrics on the CSIQ database. AWGN: additive white Gaussian noise (red); JPEG: JPEG compression (green); JP2K: JPEG2000 compression (blur); APGN: additive pink Gaussian noise (cyan); GB: Gaussian blur (magenta); GCD: global contrast decrements (yellow).

TABLE IV
PERFORMANCE INDICES OF EIGHT TESTING IQA METRICS ON MULTIPLY DISTORTED LIVEMD AND MDID2013 DATABASES

Metrics	LIVEMD Database (450 images) [33]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.6771	0.5003	0.7418	12.682	10.218
SSIM [12]	0.6534	0.4703	0.7384	12.755	10.408
VSNR [17]	0.7727	0.5775	0.8117	11.045	8.8326
GSM [21]	0.8454	0.6550	0.8808	8.9562	7.1738
IGM [22]	0.8562	0.6681	0.8859	8.7744	7.1120
GMSD [23]	0.8448	0.6547	0.8810	8.9492	7.0778
VSI [24]	0.8121	0.6189	0.8577	9.7221	7.8094
PSIM (Pro.)	0.8506	0.6598	0.8847	8.8160	6.9647

Metrics	MDID2014 Database (324 images) [34]				
	ξ_S	ξ_K	ξ_P	ξ_R	ξ_M
PSNR	0.5604	0.3935	0.5647	0.0419	0.0343
SSIM [12]	0.5518	0.3864	0.5567	0.0422	0.0345
VSNR [17]	0.3923	0.2610	0.4176	0.0462	0.0380
GSM [21]	0.6637	0.4600	0.6647	0.0380	0.0307
IGM [22]	0.8235	0.6244	0.8211	0.0290	0.0230
GMSD [23]	0.8310	0.6283	0.8310	0.0283	0.0227
VSI [24]	0.6932	0.4919	0.7039	0.0361	0.0294
PSIM (Pro.)	0.8618	0.6630	0.8588	0.0260	0.0208

works the best using correlations against reported subjective opinion scores, beyond 0.85 for prediction accuracy (ξ_P) and consistency (ξ_R).

IV. SUMMARIZATION

In this section, we will compare our proposed PSIM model with recently proposed GMSD and VSI metrics, and reveal their relationships and differences. We denote the vector $\hat{\mathbf{g}}$ as the GM similarity map [23], in which the elements belong to $[0, 1]$ with most of them close to one, and denote $\mathcal{S}$ and $\mathcal{E}$ as the standard deviation and global mean. The sign of "$\doteq$" will be used below to denote that both sides have the equal prediction monotonicity (ξ_S and ξ_K). Then we set

$$A = \mathcal{E}(\hat{\mathbf{g}}^2) \approx 1 \qquad B = \mathcal{E}(\hat{\mathbf{g}})^2 \approx 1 \tag{16}$$

$$\dot{A} = 1 - A \approx 0 \qquad \dot{B} = 1 - B \approx 0. \tag{17}$$

According to the Taylor's formula, when $(-\dot{A})$ near to zero, we have

$$\begin{aligned} \ln(1-\dot{A}) &= \ln(1+(-\dot{A})) \\ &= (-\dot{A}) - \sum_{n=2}^{+\infty}(-1)^n \frac{(-\dot{A})^n}{n} \\ &\approx -\dot{A}. \end{aligned} \tag{18}$$

Hence,

$$\begin{aligned} \frac{A}{B} &\doteq \ln\frac{A}{B} \\ &= \ln A - \ln B \\ &= \ln(1-\dot{A}) - \ln(1-\dot{B}) \\ &\approx (-\dot{A}) - (-\dot{B}) \\ &= -(1-A) - (-(1-B)) \\ &= A - B \end{aligned} \tag{19}$$

i.e.,

$$\mathcal{E}(\hat{\mathbf{g}}^2) - \mathcal{E}(\hat{\mathbf{g}})^2 \doteq \frac{\mathcal{E}(\hat{\mathbf{g}}^2)}{\mathcal{E}(\hat{\mathbf{g}})^2}. \tag{20}$$

Fusing the above formulas, we derive from the GMSD [23]

$$\text{GMSD} = \mathcal{S}(\hat{\mathbf{g}}) = \mathcal{E}(\hat{\mathbf{g}}^2) - \mathcal{E}(\hat{\mathbf{g}})^2 \doteq \frac{\mathcal{E}(\hat{\mathbf{g}}^2)}{\mathcal{E}(\hat{\mathbf{g}})^2}. \tag{21}$$

That is, GMSD can be approximated as the ratio of the local distortion-based mean to the global mean, i.e., the component $\frac{H^\tau}{L^\alpha}$ in (15). Next, as compared to the VSI metric [24], we suppose the first term $\frac{H^\tau}{L^\tau}$ in (13) has a similar function of $\max(VS_o, VS_d)$, where VS_o and VS_d stand for the saliency maps of original and distorted images, and thus the proposed PSIM model is a variant of the VSI metric. To sum up, our

PSIM works as a general framework which basically reduces to existing state-of-the-art IQA methods as particular cases.

Furthermore, in comparison to the recent GMSD and VSI models, this paper has made several main contributions:

1) Using simple operators, PSIM has obtained the optimal performance on average and promising computational efficiency.
2) we have illustrated why, using the image GM (for approximating the function of the center-surround cell) and downsampling (for approximating the influence of viewing distance and image resolution), which were applied in GMSD and VSI without explanations.
3) we have presented a reliable supposition to illustrate why visual saliency takes effect in IQA because the macrostructure is included in the detected salient regions.
4) a perceptual-based pooling scheme, where a new strategy is included for coarse and fine tunings of the parameters used, has been proposed to fuse the difference of microstructure, macrostructure, and color information between the original and distorted images.

V. Conclusion

In this paper, we introduced a novel fast perceptual-based IQA metric. The proposed PSIM metric combines the GM similarities at two scales, the color information similarity, and a reliable perceptual-based pooling. We performed the proposed IQA algorithm on four large-size singly distorted image databases (LIVE, TID2008, CSIQ, and TID2013) and two multiply distorted image databases (LIVEMD and MDID2013). Results of experiments confirm that our PSIM approach clearly performs better than classical and recent visual quality evaluators in efficiency and efficacy comparisons. It is worthy to stress that the proposed model is a general framework including some recently proposed IQA methods as its special cases. The code will be released at https://sites.google.com/site/guke198701/publications.

References

[1] S. Yin, H. J. Gao, and O. Kaynak, "Data-driven control and process monitoring for industrial applications—Part I," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6356–6359, Nov. 2014.

[2] S. Yin, H. J. Gao, and O. Kaynak, "Data-driven control and process monitoring for industrial applications—Part II," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 583–586, Jan. 2015.

[3] J. Jiao, H. Yu, and G. Wang, "A quality-related fault detection approach based on dynamic least squares for process monitoring," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2625–2632, Apr. 2016.

[4] Z. W. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques - Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[5] Z. W. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques - Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

[6] Y. Zhang, Y. Fan, and N. Yang, "Fault diagnosis of multimode processes based on similarities," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2606–2614, Apr. 2016.

[7] 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, Sep. 2015.

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

[9] K. Gu *et al.*, "Blind quality assessment of tone-mapped images via analysis of information, naturalness and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.

[10] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2575544.

[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] 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.*, Nov. 2003, pp. 1398–1402.

[14] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1–13, Jun. 2016.

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

[16] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Objective quality assessment and perceptual compression of screen content images," *IEEE Comput. Graphics Appl.*, to be published, doi:10.1109/MCG.2016.46.

[17] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[18] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Mar. 2010, Art. no. 011006. [Online]. Available: http://vision.okstate.edu/csiq

[19] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new psychovisual paradigm for image quality assessment: From differentiating distortion types to discriminating quality conditions," *Signal, Image Video Process.*, vol. 7, no. 3, pp. 423–436, May 2013.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[25] E. H. Adelson, "Lightness perception and lightness illusions," *in The New Cognitive Neurosciences*, 2nd ed. Cambridge, MA, USA: MIT Press, 2000, pp. 339–351.

[26] C. F. Stromeyer and B. Julesz, "Spatial-frequency masking in vision: Critical bands and spread of masking," *J. Opt. Soc. Amer.*, vol. 62, no. 10, pp. 1221–1232, Oct. 1972.

[27] R. De Valois, D. Albrecht, and L. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vis. Res.*, vol. 22, no. 5, pp. 545–559, 1982.

[28] J. Tanaka, D. Weiskopf, and P. Williams, "The role of color in high-level vision," *Trends Cognitive Sci.*, vol. 5, no. 5, pp. 211–215, May 2001.

[29] S. G. Solomon and P. Lennie, "The machinery of colour vision," *Nature Rev. Neurosci.*, vol. 8, pp. 276–286, Apr. 2007.

[30] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," (2006). [Online]. Available: http://live.ece.utexas.edu/research/quality

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

[32] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Sig. Process.: Image Commun.*, vol. 30, pp. 57–55, Jan. 2015.

[33] 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.*, Nov. 2012, pp. 1693–1697.
[34] 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.
[35] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision.* New York, NY, USA: McGraw-Hill, 1995.
[36] E. Bruce Goldstein, *Sensation and Perception*, 1979.
[37] Z. Li and J. Zheng, "Visual-salience-based tone mapping for high dynamic range images," *IEEE Trans. Ind. Electron.*, vol. 61, no. 12, pp. 7076–7082, Dec. 2014.
[38] L. Zhang, Y. Xia, R. Ji, and X. Li, "Spatial-aware object-level saliency prediction by learning graphlet hierarchies," *IEEE Trans. Ind. Electron.*, vol. 62, no. 2, pp. 1301–1308, Feb. 2015.
[39] J. Yu, F. Sun, D. Xu, and M. Tan, "Embedded vision-guided 3-D tracking control for robotic fish," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 355–363, Jan. 2016.
[40] X. Li, Q. Ren, and J. Xu, "Precise speed tracking control of a robotic fish via iterative learning control," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2221–2228, Apr. 2016.
[41] S.-L. Dai, M. Wang, and C. Wang, "Neural learning control of marine surface vessels with guaranteed transient tracking performance," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1717–1727, Mar. 2016.
[42] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, Jan. 2012.
[43] 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.
[44] 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.
[45] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *J. Vis.*, vol. 9, no. 3, pp. 1–24, Mar. 2009.
[46] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.
[47] C. Yang and S. H. Kwok, "Efficient gamut clipping for color image processing using LHS and YIQ," *Opt. Eng.*, vol. 42, no. 3, pp. 701–711, Mar. 2003.
[48] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jul. 2001, pp. 416–423.
[49] H. Wang, "A note on iterative marginal optimization: A simple algorithm for maximum rank correlation estimation," *Comput. Statist. Data Anal.*, vol. 51, no. 6, pp. 2803–2812, 2007, Mar. 2007.
[50] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
[51] 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.
[52] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 506–510.
[53] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," (2000, Mar.). [Online]. Available: http://www.vqeg.org/

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively.

He is currently with Beijing University of Technology, Beijing, China. His research interests include quality assessment, contrast enhancement, and visual saliency detection.

Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo 2016. He is the leading special session organizer for VCIP2016 and ICIP2017. He is a Reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON BROADCASTING, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, IEEE ACCESS, *Information Sciences*, *Neurocomputing*, *Multimedia Tools and Applications*, *SPIC*, the *Journal of Visual Communication and Image Representation*, *Digital Signal Processing*, etc.

Leida Li (M'14) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively.

He is currently an Associate Professor with the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, China. His current research interests include multimedia quality assessment, information hiding, and image forensics.

Hong Lu received the M.S. degree from the School of Electrical Engineering, Southeast University, Nanjing, China, in 2003, and the Ph.D. degree in control theory and control engineering from the School of Automation, Southeast University, in 2009.

She is currently an Associate Professor with the School of Automation, Nanjing Institute of Technology, Nanjing, China. Her current research interests include object detection, object tracking, and image and video data mining.

Xiongkuo Min received the B.E. degree in electronic engineering from Wuhan University, Wuhan, China, in 2013. He is currently working toward the Ph.D. degree at the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include image/video quality assessment, visual attention modeling, and perceptual signal processing.

Mr. Min was a recipient of the Best Student Paper Award at the 2016 IEEE International Conference on Multimedia and Expo.

Weisi Lin (F'16) received the Ph.D. degree from Kings College London, London, U.K.

He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with more than 340 refereed papers published in international journals and conference proceedings.

Dr. Lin is a Fellow of the Institution of Engineering and Technology, U.K., and an Honorary Fellow of the Singapore Institute of Engineering Technologists.