VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment

Lin Zhang, *Member, IEEE*, Ying Shen, *Member, IEEE*, and Hongyu Li

***Abstract*— Perceptual image quality assessment (IQA) aims to use computational models to measure the image quality in consistent with subjective evaluations. Visual saliency (VS) has been widely studied by psychologists, neurobiologists, and computer scientists during the last decade to investigate, which areas of an image will attract the most attention of the human visual system. Intuitively, VS is closely related to IQA in that suprathreshold distortions can largely affect VS maps of images. With this consideration, we propose a simple but very effective full reference IQA method using VS. In our proposed IQA model, the role of VS is twofold. First, VS is used as a feature when computing the local quality map of the distorted image. Second, when pooling the quality score, VS is employed as a weighting function to reflect the importance of a local region. The proposed IQA index is called visual saliency-based index (VSI). Several prominent computational VS models have been investigated in the context of IQA and the best one is chosen for VSI. Extensive experiments performed on four large-scale benchmark databases demonstrate that the proposed IQA index VSI works better in terms of the prediction accuracy than all state-of-the-art IQA indices we can find while maintaining a moderate computational complexity. The MATLAB source code of VSI and the evaluation results are publicly available online at http://sse.tongji.edu.cn/linzhang/IQA/VSI/VSI.htm.**

***Index Terms*— Perceptual image quality assessment, visual saliency.**

I. Introduction

QUANTITATIVE evaluation of an image's perceptual quality is one of the most fundamental yet challenging problems in image processing and vision research. Image quality assessment (IQA) methods fall into two categories: subjective assessment by humans and objective assessment by algorithms designed to mimic the subjective judgments. Though subjective assessment is the ultimate criterion of an image's quality, it is time-consuming, cumbersome, and cannot be implemented in systems where a real-time quality score for an image or video sequence is needed. Recently, there has been an increasing interest in developing objective IQA methods. According to the availability of a reference image, objective IQA indices can be classified as full reference (FR), no-reference (NR) and reduced-reference (RR) methods [1]. In this paper, the discussion is confined to FR methods, where the pristine "distortion free" image is known as the reference image.

As a conventional fidelity metric, the peak-to-noise ratio (PSNR) or the mean squared error (MSE), works well for evaluating the quality of images sharing the same content and the same distortion type. However, quality scores predicted by PSNR or MSE do not correlate well with human beings' subjective fidelity ratings when multiple images or multiple distortion types are involved [2]. In the past decade, several sophisticated IQA models have been proposed and some representative ones will be briefly reviewed here.

The noise quality measure index (NQM) [3] and the visual signal-to-noise ratio index (VSNR) [4] emphasize the importance of human visual system (HVS)'s sensitivity to different visual signals, such as the luminance, the contrast, the frequency content, and the interaction between them.

The structural similarity index (SSIM) proposed by Wang *et al*. [5] can be considered as a milestone of the development of IQA models. SSIM is based on the hypothesis that HVS is highly adapted to extract the structural information from the visual scene and therefore a measurement of structural similarity can provide a good approximation of the perceived image quality. In their later work, Wang *et al*. proposed a multi-scale extension of SSIM, namely MS-SSIM [6] and it has been corroborated that MS-SSIM could produce better results than its single scale counterpart. In [7], Wang and Li improved the original MS-SSIM to the information content weighted SSIM index (IW-SSIM) by introducing a new information content weighting (IW)-based quality score pooling strategy.

In [8], Sheikh *et al*. proposed the visual information fidelity index (VIF), which was an extension of its former version, i.e. the information fidelity criterion index (IFC) [9]. In VIF, Sheikh *et al*. treated the FR IQA problem as an information fidelity problem and the fidelity were quantified by the amount of information shared between the reference image and the distorted image. In [10], Zhang *et al*. proposed a Riesz transforms based feature similarity index (RFSIM). In RFSIM, 1st-order and 2nd-order Riesz transforms are used to characterize image's local structures and the Canny edge detector is employed to generate the mask for quality score pooling. Larson and Chandler argued that the HVS performs two distinct strategies when assessing the image quality for

Manuscript received December 16, 2013; revised April 27, 2014 and July 8, 2014; accepted August 2, 2014. Date of publication August 7, 2014; date of current version August 29, 2014. This work was supported in part by the Natural Science Foundation of China under Grant 61201394 and Grant 61303112, in part by the Shanghai Pujiang Program under Grant 13PJ1408700, and in part by the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing, China, under Grant 30920140122007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Damon M. Chandler.

L. Zhang is with the School of Software Engineering, Tongji University, Shanghai 201804, China, and also with the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: cslinzhang@tongji.edu.cn).

Y. Shen and H. Li are with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: yingshen@tongji.edu.cn; hyli@tongji.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2014.2346028

high-quality images and for low-quality images, and accordingly they proposed a most apparent distortion (MAD) based IQA index [11]. The feature similarity index (FSIM) proposed in [12] employs two features to compute the local similarity map, the phase congruency and the gradient magnitude. The authors claimed that the phase congruency and the gradient magnitude play complementary roles in characterizing the local image quality. At the quality score pooling stage of FSIM, phase congruency map is utilized again as a weighting function since it can roughly reflect how perceptually important a local patch is to the HVS. By considering that structural and contrast changes can be effectively captured by gradients, Liu *et al.* proposed a gradient similarity based metric (GSM) [13] for FR IQA. For a thorough survey of modern IQA development, please refer to [14] and [15].

On the other hand, in recent years how to build effective computational visual saliency (VS) models has been attracting tremendous attention [16]–[18]. Given an image, its VS map computed by an appropriate VS model can reflect how "salient" a local region is to the HVS. Intuitively, VS and IQA are intrinsically related because both of them depend on how HVS perceives an image and suprathreshold distortions can be a strong attractor of visual attention [19]. Thus recently, researchers have been trying to incorporate VS information to IQA models to improve their performance. Meanwhile, there are also some other studies focusing on the relationship between visual attention and the perceptual quality of videos [30]–[33].

The relationship between VS and IQA has been investigated by some researchers in previous studies and it is widely accepted that incorporating VS information appropriately can benefit IQA metrics. However, a practical VS-based computational IQA model that could achieve better prediction performance than the other state-of-the-art methods, such as IW-SSIM [7], FSIM_C [12], GSM [13], has not come out yet. In this paper, we expect to fill this research gap to some extent. By analyzing the relationship between the changes of an image's VS map and its perceived quality degradation, we propose a simple yet very effective VS-based index (VSI) for the IQA task. We claim that the VS map cannot be only used as a weighting function at the score-pooling stage, but can also be used as a feature map to characterize the quality of local image regions. The underlying reason is that perceptible quality distortions can lead to measurable changes of images' visual saliency maps. Consequently, in our proposed VSI metric, the role of an image's VS map is twofold: a feature map characterizing the image's local quality, as well as a weighting function indicating the importance of a local region to the HVS when pooling the final quality score. In our work, several eminent VS models have been explored in the context of IQA and the most suitable one is selected for VSI. VSI is thoroughly examined by extensive experiments conducted on four large scale databases. The results show that our proposed VSI works consistently better than all the other state-of-the-art IQA metrics. In addition, the computational complexity of VSI is quite low. The Matlab source code of VSI and the associated evaluation results have been made publicly available online at http://sse.tongji.edu.cn/linzhang/IQA/VSI/VSI.htm.

The remainder of this paper is organized as follows. Section II introduces the works relevant to this paper. Section III presents in detail the proposed VSI metric for IQA. Section IV presents the experimental results and associated discussions. Finally, Section V concludes the paper.

II. Related Works

This section presents works most related to our paper, which covers a brief review of modern VS models and their existing applications in IQA.

A. Computational Visual Saliency Models

As a consequence of evolution, most vertebrates, including humans, have a remarkable ability to automatically pay more attention to salient regions of the visual scene. Building effective computational models to simulate human visual attention has been studied by scholars in psychology, neurobiology, and computer vision for a long time, and some powerful models have been proposed. Although both bottom-up (scene dependant) and top-down (task dependant) factors will affect the visual attention, most of the existing computational VS models are bottom-up since bottom-up attention mechanisms are more thoroughly studied than top-down mechanisms. In bottom-up VS models, it is supposed that visual attention is driven by low-level visual stimulus in the scene, such as intensity, color, orientation, etc.

The first influential and best known VS model was proposed by Itti *et al*. [34] for still images. Itti *et al*.'s model was based on the VS computational architecture introduced by Koch and Ullman [35]. Itti *et al*.'s contribution mainly lies in two aspects. First, they introduced image pyramids for feature extraction, which makes the VS computation efficient. Second, they proposed the biologically inspired "center-surround difference" operation to compute feature dependant saliency maps across scales. In their later work, Itti and Baldi introduced a Bayesian model of surprise aiming to predict eye movements [36]. In [37], following Itti *et al*.'s architecture, Harel *et al*. proposed the graph-based visual saliency (GBVS) model by introducing a novel graph-based normalization and combination strategy. In another work following Itti *et al.*'s framework, Klein and Frintrop [38] modeled the center-surround contrast in an information-theoretic way, in which two distributions of visual feature occurrences are determined for a center and for a surround region, respectively. By analyzing the log-spectrum of the input image, Hou and Zhang [39] proposed a Fourier transform based method to extract the spectral residual (SR) of an image in the spectral domain and to construct the corresponding saliency map in the spatial domain; one prominent advantage of this method is its low computational complexity. In Hou's recent work [40], he developed a saliency algorithm based on the image signature (IS), which can approximate the foreground of an image and can be simply computed as the sign map of the image's DCT (discrete cosine transform) coefficients. In [41], Bruce and Tsotsos proposed the model of attention based on information maximization (AIM), in which an image's saliency is modeled as the maximum information that can be sampled

from it. In [42], Seo and Milanfar used local regression kernels (LRK) as features and used a local "self-resemblance" measure, which indicates the likelihood of saliency, to build an image's saliency map. In [43], Achanta *et al.* proposed a conceptually simple approach for detecting saliency by combining image's responses to band-pass filters from three *CIEL*a*b** channels. In [44], Judd *et al.* extracted various types of features and fed them into a SVM to train a model to predict the visual saliency of a given test image. In [45], Shen and Wu represent an image as a low-rank matrix plus sparse noises, where the low-rank matrix explains the non-salient regions while the sparse noises indicate the salient regions. By integrating prior knowledge from three aspects, frequency prior, color prior, and location prior, Zhang *et al.* [46] proposed an efficient saliency algorithm, namely SDSP (Saliency Detection by combining Simple Priors).

For a complete recent survey of modern VS models, please refer to [16]–[18].

B. Existing Investigations of Visual Saliency in IQA

Recently, increased awareness to the close relationship between VS and quality perception has led to a number of approaches that try to integrate VS into IQA metrics to potentially improve their prediction performance.

In [20], Vu *et al*. designed two experiments to examine visual fixation patterns when judging image quality. Their results revealed that regions where people fixate while judging image quality can be different from those obtained under task-free condition. In their another work [21], five common fidelity metrics were augmented using two sets of fixation data, one set obtained under task-free viewing conditions and another set obtained when viewers were asked to judge image quality. The results show that most metrics could be improved using fixation data and a greater improvement was found using fixations obtained in the task-free condition. Similar results have also been obtained by a recent study [22]. In [23], Larson *et al*. revealed that common metrics (such as SSIM, PSNR, VIF, etc.) could be improved by using spatially varying weights for pooling. In [24], a framework was introduced to extend existing quality metrics by segmenting an image into ROI (region of interest) and background regions. With such a method, the metrics are computed independently on ROI and background regions and then a pooling function is used to derive the final quality score. The abovementioned works demonstrate that if being incorporated appropriately, visual attention data can benefit the design of IQA metrics. However, it should be noted that the visual attention data or the ROI data used in these works are either obtained by eye tracking or are hand-labeled. Thus, these approaches cannot be used in applications where a fully automatic IQA metric is needed.

Rather than using eye-tracking or subjective ROI data, some researchers attempted to incorporate VS information computed by using computational VS models into IQA models. Representative works belonging to this category include [25]–[29] and they share some common characteristics.

At first, these studies are based on the assumption that a distortion occurring in an area that attracts the viewer's attention is more annoying than in any other area, and they attempt to weight local distortions with a local saliency map. Consequently, pooling strategies adopted in these methods share a general form as:

$$S = \sum\nolimits_i^K w_i s_i \Big/ \sum\nolimits_i^K w_i \tag{1}$$

where s_i is the local quality value at the i-th location in the local quality map, w_i is the VS value at the location i, K is the number of points in the image, and S is the final quality score of the examined image.

Secondly, for these methods, the motivation is actually not to design a new IQA index but to demonstrate that a VS-weighted pooling strategy could perform better than the simple "mean" scheme. Thus, for computing the local quality map, they all adopt some existing methods, such as PSNR, SSIM, and VIF, without discussing whether there could be more effective methods to characterize the local image quality.

Thirdly, these works lack extensive evaluations to verify the effectiveness of the proposed IQA indices. Usually, the experiments were performed only on a specific dataset and only some classic IQA indices (e.g., SSIM, VIF, PSNR) were used for comparison. Some recently developed high performance IQA metrics, such as IW-SSIM [7], FSIM/FSIM$_C$ [12], and GSM [13], were not compared with, which makes the elicited conclusions less convincing.

III. VS-Based Index (VSI) for IQA

As stated in Section II, VS has already been used as a weighting function for quality score pooling in some previous studies [25]–[29]. In [12], Zhang *et al*. have shown that perceptible image quality degradation can lead to perceptible changes in image's low-level features. Since the bottom-up VS models are basically based on image's low-level features, the VS values themselves actually vary with the change of image quality. Therefore, why don't we use VS as a feature to compute the local similarity map between the reference image and the distorted image?

We find that quality distortions could give rise to changes in images' VS maps and the intensities of such measurable changes correlate well with the degrees of perceptible quality distortions. To support our claim, we have conducted a statistical analysis on VS maps of images in TID2013 [47], the most comprehensive dataset available for IQA research. In TID2013, there are 25 reference images, 24 distortion types and 5 distortion levels. Hence, for one distortion type at a particular distortion level, there are 25 samples. To perform such an analysis, we at first computed VS maps for all the images by using GBVS model [37] and then for each reference-distortion image pair, we computed the MSE between their VS maps. After that, we averaged MSEs belonging to the same distortion type and the same distortion level. The results are listed in Table I.[1] In each field of Table I, we also present the average subjective score for the corresponding distortion type and distortion level in a bracket. In TID2013, a higher

[1]For better observation, MSE values listed here are the actual MSE values multiplied by 10^6.

TABLE I
AVERAGE MSES OF VS AND AVERAGE SUBJECTIVE SCORES FOR DISTORTIONS AT DIFFERENT LEVELS

Dis. Type	Level 1	Level 2	Level 3	Level 4	Level 5
AGN	2.68 (5.67)	4.10 (5.23)	7.26 (4.85)	16.71 (4.25)	33.92 (3.77)
ANC	0.55 (5.93)	1.17 (5.85)	2.68 (5.53)	5.47 (5.06)	9.36 (4.48)
SCN	15.41 (4.76)	25.54 (4.24)	53.11 (3.70)	111.87 (3.20)	244.19 (2.69)
MN	1.04 (5.92)	2.15 (5.79)	5.12 (5.50)	12.09 (5.08)	32.30 (4.47)
HFN	0.58 (5.89)	1.18 (5.64)	3.57 (5.11)	11.12 (4.15)	39.36 (3.22)
IN	2.02 (4.71)	5.38 (4.36)	10.18 (3.98)	19.34 (3.64)	44.25 (3.20)
QN	10.21 (5.25)	25.17 (4.68)	49.95 (4.11)	107.71 (3.46)	311.32 (2.83)
GB	0.16 (5.65)	0.41 (5.03)	1.20 (4.15)	5.54 (3.31)	40.94 (2.45)
DEN	4.17 (6.06)	9.60 (5.75)	20.72 (4.82)	77.02 (3.44)	289.00 (2.02)
JPEG	0.97 (5.94)	2.73 (5.70)	10.71 (4.91)	68.89 (3.31)	205.86 (2.07)
JP2K	1.64 (5.65)	10.70 (5.02)	67.21 (3.91)	323.61 (2.79)	1350.75 (1.28)
JGTE	0.86 (5.74)	1.90 (5.40)	35.00 (4.61)	133.98 (3.87)	408.88 (2.97)
J2TE	15.27 (4.88)	43.04 (4.18)	48.60 (3.62)	200.03 (2.96)	269.04 (2.59)
NEPN	26.88 (5.72)	38.81 (5.89)	131.92 (5.33)	201.36 (4.71)	361.42 (3.85)
Block	604.50 (3.28)	466.45 (3.34)	459.34 (3.48)	385.14 (3.73)	307.11 (4.06)
MS	0.20 (6.07)	2.78 (6.06)	2.86 (5.63)	26.60 (5.34)	26.95 (4.65)
CTC	0.13 (5.62)	19.79 (6.48)	0.13 (4.48)	83.29 (6.29)	0.19 (3.45)
CCS	0.16 (5.06)	0.11 (4.66)	0.13 (4.23)	0.19 (3.92)	0.30 (3.69)
MGN	3.12 (5.52)	5.47 (5.13)	10.53 (4.69)	24.34 (4.13)	50.43 (3.68)
CN	1.21 (5.88)	3.68 (5.55)	13.93 (5.01)	66.72 (4.17)	200.58 (3.31)
LCNI	6.41 (5.51)	16.55 (5.00)	32.59 (4.32)	91.33 (3.53)	255.46 (2.49)
ICQD	2.86 (5.69)	7.87 (5.26)	36.06 (4.58)	89.54 (3.79)	350.14 (2.91)
CHA	0.71 (6.11)	3.53 (5.87)	28.09 (5.18)	104.38 (4.40)	436.12 (3.00)
SSR	2.67 (5.77)	9.63 (5.10)	68.65 (3.99)	498.20 (2.61)	1487.23 (0.95)

subjective score indicates a better image quality. Based on the results listed in Table I, we could have the following findings.

At first, for nearly all types of distortions, a lower average subjective score corresponds to a severer average VS changes measured by MSE. Secondly, even for most cross-distortion cases, VS changes can also be a good indicator of perceptual image quality. For example, the average subjective score for the distortion "SCN Level 5" (2.69) is much poorer than the one for the distortion "AGN Level 1" (5.67); as expected, the average VS changes for the distortion "SCN Level 5" (244.19) is much severer than the one for the distortion "AGN Level 1" (2.68). The relationship between the quality distortions and VS changes is illustrated using an example in Fig. 1. Images used in Fig. 1 are from TID2013. Fig. 1(a) is a reference image and

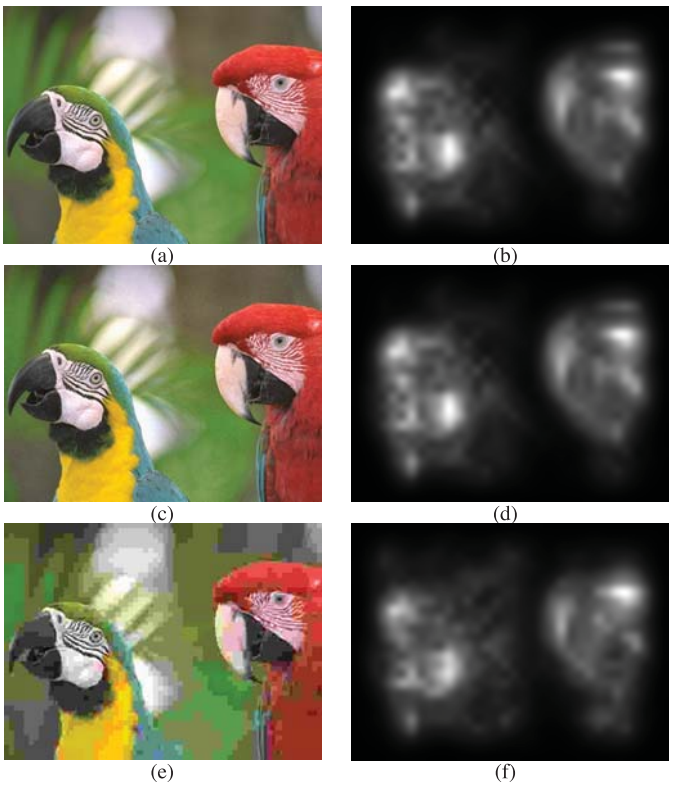

Fig. 1. (a) is a reference image while (c) and (e) are its two distorted versions; subjective scores for (c) and (e) are 5.05 and 2.40, respectively; (b), (d) and (f) are the VS maps of (a), (c) and (e), respectively; the MSE between (d) and (b) is 4.49 while the MSE between (f) and (b) is 198.36.

Figs. 1(c) and (e) are its two distorted versions. The subjective scores for Figs. 1(c) and (e) are 5.05 and 2.40, respectively. Figs. 1(b), 1(d) and 1(f) are the VS maps of Figs. 1(a), 1(c) and 1(e), respectively. The MSE between Figs. 1(d) and 1(b) is 4.49 while the MSE between Figs. 1(f) and 1(b) is 198.36. Fig. 1(e) has a poorer quality than Fig. 1(c) and as expected, its VS map Fig. 1(f) alters much more than Fig. 1(c)'s VS map Fig. 1(d) when being compared with the reference VS map Fig. 1(b).

Based on the above analysis, it can be seen that in most cases, changes of VS maps can be a good indicator of distortion degrees and thus, in this paper we propose to use the VS map as a feature to characterize the image's local quality.

However, from Table I, it can be seen that as a quality distortion indicator, VS map does not work quite well for the distortion type CTC (Contrast Change). The root reason is that due to the normalization operations involved in VS computational models, the VS value at a pixel is a measure to reflect its relative distinctiveness to its surroundings, which makes VS weak to characterize image's absolute contrast. Nonetheless, image's local contrast does affect much HVS' perception of the image quality. We use an example in Fig. 2 to illustrate this fact. Fig. 2(a) is a reference image while 2(b) is a distorted version of it and the distortion type is contrast reduction. Fig. 2(c) and Fig. 2(d) are the VS maps of Fig. 2(a) and Fig. 2(b), respectively. It can be clearly seen that Fig. 2(b) has lower quality than Fig. 2(a). However, such a quality degradation caused by contrast reduction cannot be

Fig. 2. (a) is a reference image while (b) is a distorted version of it (the distortion type is contrast reduction); (c) and (d) are the VS maps computed from (a) and (b) respectively using the GBVS model [37]; (e) and (f) are the GM maps computed from (a) and (b) respectively using the Scharr gradient operator. No significant difference can be observed between (c) and (d), which indicates that VS map behaves poorly in characterizing the contrast loss of the images. By contrast, apparent differences can be observed between (e) and (f), indicating that GM map has a good capability in reflecting the contrast loss of images.

reflected in their VS maps, since no significant difference can be observed between Fig. 2(c) and Fig. 2(d).

Fortunately, we can use an additional feature to compensate for the lack of contrast sensitivity of VS. The simplest feature of this kind may be the gradient modulus (GM). There are several different operators to compute the image gradient, such as the Prewitt operator, the Sobel operator, the Roberts operator [48] and the Scharr operator [49], and here we adopt the Scharr gradient operator, which has been proved very powerful in our previous work [12]. With Scharr gradient operator, the partial derivatives $G_x(\mathbf{x})$ and $G_y(\mathbf{x})$ of an image $f(\mathbf{x})$ are calculated as:

$$G_x(\mathbf{x}) = \frac{1}{16}\begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * f(\mathbf{x})$$
$$G_y(\mathbf{x}) = \frac{1}{16}\begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} * f(\mathbf{x}) \quad (2)$$

The GM of $f(\mathbf{x})$ is then computed as $G(\mathbf{x}) = \sqrt{G_x^2(\mathbf{x}) + G_y^2(\mathbf{x})}$. GM maps of Figs. 2(a) and 2(b) are visualized in Figs. 2(e) and 2(f), respectively. Apparent differences can be observed between Figs. 2(e) and 2(f), indicating that GM map has a good potential capability in reflecting the local contrast loss of images. Therefore, VS and GM are complementary and they reflect different aspects of the HVS in assessing the local quality of the input image.

From Table I, it can also be seen that as a quality distortion indicator, VS map does not work quite well for the distortion type CCS (Change of Color Saturation) either. Actually, color distortion cannot be well characterized by gradient either since usually gradient is computed from the luminance channel of images. Hence, to make the IQA metric possess the capability to deal with color distortions, chrominance information should be given special considerations. Consequently, for RGB color images, we first transform them into an opponent color space [50]:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

The weights in the above conversion are optimized for the HVS [51]. Then, the gradients are computed from L channels. M and N, two chrominance channels, will be used as features to characterize the quality degradation caused by color distortions.

With the extracted VS, GM, and chrominance features, we can define a VS-based index (VSI) for IQA tasks. Suppose that we are going to calculate the similarity between images f_1 and f_2. Denote by VS_1 and VS_2 the two VS maps extracted from images f_1 and f_2 using a specific VS model; denote by G_1 and G_2 the two GM maps; denote by M_1 and M_2 the two M channels; and denote by N_1 and N_2 the two N channels. The computation of VSI consists of two stages. In the first stage, the local similarity map is computed, and in the second stage, we pool the similarity map into a single quality score.

We separate the similarity measurement between $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ into three components, one for VS, one for GM, and the other for chrominance. First, the similarity between $VS_1(\mathbf{x})$ and $VS_2(\mathbf{x})$ is defined as:

$$S_{VS}(\mathbf{x}) = \frac{2VS_1(\mathbf{x}) \cdot VS_2(\mathbf{x}) + C_1}{VS_1^2(\mathbf{x}) + VS_2^2(\mathbf{x}) + C_1} \quad (4)$$

where C_1 is a positive constant to increase the stability of S_{VS}. Similarly, the GM values $G_1(\mathbf{x})$ and $G_2(\mathbf{x})$ are compared as:

$$S_G(\mathbf{x}) = \frac{2G_1(\mathbf{x}) \cdot G_2(\mathbf{x}) + C_2}{G_1^2(\mathbf{x}) + G_2^2(\mathbf{x}) + C_2} \quad (5)$$

where C_2 is another positive constant. The similarity between the chrominance components is simply defined as:

$$S_C(\mathbf{x}) = \frac{2M_1(\mathbf{x}) \cdot M_2(\mathbf{x}) + C_3}{M_1^2(\mathbf{x}) + M_2^2(\mathbf{x}) + C_3} \cdot \frac{2N_1(\mathbf{x}) \cdot N_2(\mathbf{x}) + C_3}{N_1^2(\mathbf{x}) + N_2^2(\mathbf{x}) + C_3} \quad (6)$$

In our experiments, C_1, C_2 and C_3 are all fixed so that the proposed VSI can be conveniently applied to all datasets. Then, $S_{VS}(\mathbf{x})$, $S_G(\mathbf{x})$ and $S_C(\mathbf{x})$ are combined to get the local similarity $S(\mathbf{x})$ of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$. We define $S(\mathbf{x})$ as follows:

$$S(\mathbf{x}) = S_{VS}(\mathbf{x}) \cdot [S_G(\mathbf{x})]^{\alpha} \cdot [S_C(\mathbf{x})]^{\beta} \quad (7)$$

where α and β are two parameters used to adjust the relative importance of VS, GM, and chrominance features.

Having obtained the local similarity $S(\mathbf{x})$ at each location $\mathbf{x}$, the overall similarity between f_1 and f_2 can be calculated.

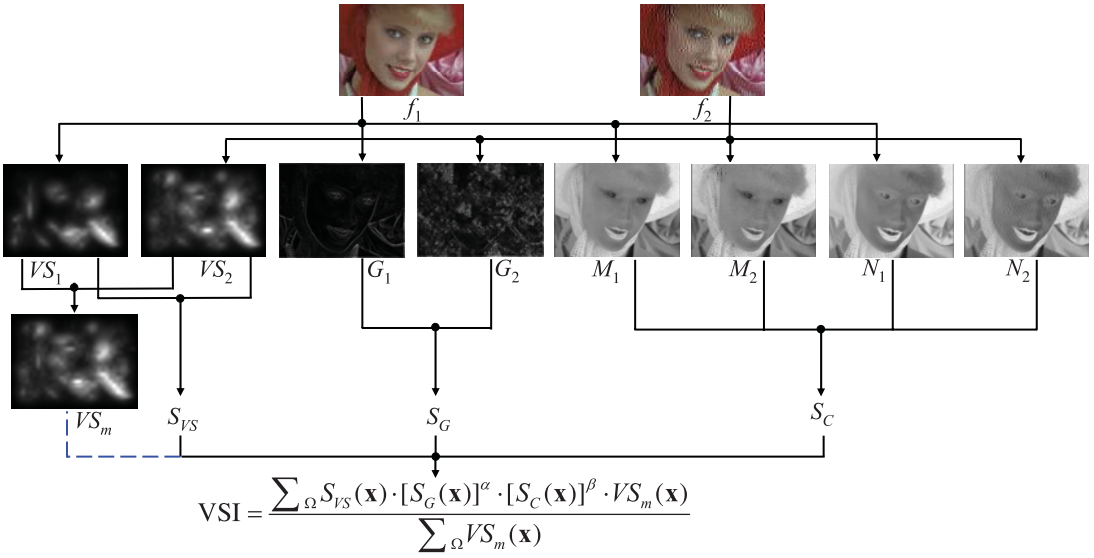

Fig. 3. Illustration for the computational process of the proposed IQA index VSI. f_1 is a reference image and f_2 is a distorted version of f_1.

It has been widely accepted that different locations can have different contributions to the HVS' perception of the image quality and it would be better if the score pooling strategy could be correlated with human visual fixation. Consequently, in our VSI framework, it is natural to choose the VS map to characterize the visual importance of a local region. Intuitively, for a given position $\mathbf{x}$, if anyone of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ has a high VS value, it implies that this position $\mathbf{x}$ will have a high impact on HVS when it evaluates the similarity between f_1 and f_2. Therefore, we use $VS_m(\mathbf{x}) = \max(VS_1(\mathbf{x}), VS_2(\mathbf{x}))$ to weight the importance of $S(\mathbf{x})$ in the overall similarity. Actually, a similar form was used in [12]. Finally, the VSI metric between f_1 and f_2 is defined as:

$$\mathrm{VSI} = \frac{\sum_{\mathbf{x}\in\Omega} S(\mathbf{x}) \cdot VS_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} VS_m(\mathbf{x})} \tag{8}$$

where Ω means the whole spatial domain. It can be easily verified that as a metric function VSI satisfies the symmetry property mentioned in [5]. The procedures to compute VSI are illustrated by an example in Fig. 3.

IV. Experimental Results and Discussions

A. Experimental Protocol

Experiments were conducted on four large-scale image datasets constructed for evaluating IQA indices, including TID2013 [47], TID2008 [52], CSIQ [11] and LIVE [53]. The important information of these four datasets, in terms of the number of reference images, the number of distorted images, the number of quality distortion types, and the number of subjects performing the subjective evaluations, is summarized in Table II. Totally, there are 6345 distorted images contained in these datasets.

Four commonly used performance metrics are employed to evaluate the IQA indices. The first two are the Spearman rank-order correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC), which can measure the prediction monotonicity of an IQA index. These two metrics operate only on the rank of the data points and ignore the relative distance between data points. To compute the other two metrics we need to apply a regression analysis to provide a nonlinear mapping between the objective scores and the subjective mean opinion scores (MOS). The third metric is the Pearson linear correlation coefficient (PLCC) between MOS and the objective scores after nonlinear regression. The fourth metric is the root mean squared error (RMSE) between MOS and the objective scores after nonlinear regression. For the nonlinear regression, we used the following mapping function as suggested by Sheikh *et al.* [53]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5 \tag{9}$$

TABLE II
Benchmark Datasets for Evaluating IQA Indices

Dataset	Reference Images No.	Distorted Images No.	Distortion Types No.	Subjects No.
TID2013	25	3000	25	971
TID2008	25	1700	17	838
CSIQ	30	866	6	35
LIVE	29	779	5	161

TABLE III
SROCC Values Obtained by Using VSI With Different VS Models on the Sub-Dataset

VS Model	SROCC
Itti [34]	0.9035
GBVS [37]	0.9017
AIM [41]	0.8791
LRK [42]	0.9031
SR [39]	0.8892
FT [43]	0.8780
IS [40]	0.9008
SDSP [46]	**0.9061**

TABLE IV
PERFORMANCE COMPARISON OF 13 IQA INDICES ON FOUR BENCHMARK DATASETS

		SSIM_I	SSIM	MS-SSIM	IFC	VIF	VSNR	MAD	GSM	IW-SSIM	RFSIM	FSIM	$FSIM_C$	VSI
TID 2013	SROC	0.7650	0.7417	0.7859	0.5389	0.6769	0.6812	0.7807	0.7946	0.7779	0.7744	0.8015	**0.8510**	**0.8965**
	KROC	0.5864	0.5588	0.6047	0.3939	0.5147	0.5084	0.6035	0.6255	0.5977	0.5951	0.6289	**0.6665**	**0.7183**
	PLCC	0.8195	0.7895	0.8329	0.5538	0.7720	0.7402	0.8267	0.8464	0.8319	0.8333	0.8589	**0.8769**	**0.9000**
	RMSE	0.7105	0.7608	0.6861	1.0322	0.7880	0.8392	0.6975	0.6603	0.6880	0.6852	0.6349	**0.5959**	**0.5404**
TID 2008	SROC	0.8117	0.7749	0.8542	0.5675	0.7491	0.7046	0.8340	0.8504	0.8559	0.8680	0.8805	**0.8840**	**0.8979**
	KROC	0.6200	0.5768	0.6568	0.4236	0.5860	0.5340	0.6445	0.6596	0.6636	0.6780	0.6946	**0.6991**	**0.7123**
	PLCC	0.8173	0.7732	0.8451	0.7340	0.8084	0.6820	0.8308	0.8422	0.8579	0.8645	0.8738	**0.8762**	**0.8762**
	RMSE	0.7732	0.8511	0.7173	0.9113	0.7899	0.9815	0.7468	0.7235	0.6895	0.6746	0.6525	**0.6468**	**0.6466**
CSIQ	SROC	0.8653	0.8756	0.9133	0.7671	0.9195	0.8106	**0.9466**	0.9108	0.9213	0.9295	0.9242	0.9310	**0.9423**
	KROC	0.6866	0.6907	0.7393	0.5897	0.7537	0.6247	**0.7970**	0.7374	0.7529	0.7645	0.7567	0.7690	**0.7857**
	PLCC	0.8566	0.8613	0.8991	0.8384	0.9277	0.8002	**0.9502**	0.8964	0.9144	0.9179	0.9120	0.9192	**0.9279**
	RMSE	0.1355	0.1334	0.1149	0.1431	0.0980	0.1575	**0.0818**	0.1164	0.1063	0.1042	0.1077	0.1034	**0.0979**
LIVE	SROC	0.9571	0.9479	0.9513	0.9259	0.9636	0.9274	**0.9669**	0.9561	0.9567	0.9401	0.9634	**0.9645**	0.9524
	KROC	0.8163	0.7963	0.8045	0.7579	0.8282	0.7616	**0.8421**	0.8150	0.8175	0.7816	0.8337	**0.8363**	0.8058
	PLCC	0.9529	0.9449	0.9489	0.9268	0.9604	0.9231	**0.9675**	0.9512	0.9522	0.9354	0.9597	**0.9613**	0.9482
	RMSE	8.2881	8.9455	8.6188	10.264	7.6137	10.506	**6.9073**	8.4327	8.3473	9.6642	7.6780	**7.5296**	8.6816

TABLE V
OVERALL PERFORMANCES OF IQA INDICES OVER 4 DATASETS

IQA Index	SROCC	KROCC	PLCC
SSIM_I [29]	0.8148	0.6373	0.8404
SSIM [5]	0.7942	0.6108	0.8140
MS-SSIM [6]	0.8419	0.6616	0.8594
IFC [9]	0.6252	0.4733	0.6867
VIF [8]	0.7646	0.6049	0.8261
VSNR [4]	0.7354	0.5622	0.7553
MAD [11]	0.8405	0.6702	0.8619
GSM [13]	0.8452	0.6732	0.8650
IW-SSIM [7]	0.8403	0.6635	0.8649
RFSIM [10]	0.8410	0.6633	0.8657
FSIM [12]	0.8593	0.6891	0.8825
$FSIM_C$ [12]	0.8847	0.7101	0.8928
VSI	**0.9100**	**0.7366**	**0.9033**

TABLE VI
RANKING OF OVERALL PERFORMANCES OF IQA INDICES

IQA Index	SROCC	KROCC	PLCC
SSIM_I [29]	9	9	9
SSIM [5]	10	10	11
MS-SSIM [6]	5	8	8
IFC [9]	13	13	13
VIF [8]	11	11	10
VSNR [4]	12	12	12
MAD [11]	7	5	7
GSM [13]	4	4	5
IW-SSIM [7]	8	6	6
RFSIM [10]	6	7	4
FSIM [12]	3	3	3
$FSIM_C$ [12]	2	2	2
VSI	**1**	**1**	**1**

where β_i, $i = 1, 2, \ldots, 5$, are parameters to be fitted. More details about the definitions and explanations of these four performance metrics can be found in [7].

VSI was compared with the other 12 state-of-the-art or representative IQA indices, including SSIM_I [29], SSIM [5], MS-SSIM [6], IFC [9], VIF [8], VSNR [4], MAD [11], GSM [13], IW-SSIM [7], RFSIM [10], FSIM [12], and $FSIM_C$ [12]. It needs to be pointed out that SSIM_I [29] is a representative IQA model which adopts a visual saliency map as a weighting function for score pooling. Specifically, SSIM_I uses SSIM [5] to compute the local quality map and uses Itti's model [34] to compute the visual saliency map.

B. Evaluation of VS Models, Determination of Parameters, and Examination of Two Roles of VS

In our VSI scheme, the VS map could be computed using various VS models. In order to find an appropriate candidate for VSI, eight eminent computational VS models, including Itti's model [34], GBVS [37], AIM [41], LRK [42], SR [39], FT [43], IS [40] and SDSP [46] were tested for VSI. In order to reduce the burden of parameter adjustment, in this experiment only a sub-dataset of TID2008 was used, which contained the first 8 reference images and the associated 544 distorted images. For each VS model evaluated, the related parameters were tuned experimentally and the tuning criterion was that the parameter value leading to a higher SROCC would be chosen. The SROCC values obtained by VSI with eight different VS models on the tuning dataset are listed in Table III, from which we can see that the SDSP model could achieve better results than the others. Thus, in all of the following experiments, SDSP was used to compute the VS map for VSI. Key parameters α and β are set as 0.40 and 0.02, respectively.

In VSI, the role of the VS map is twofold: a feature map characterizing the image's local quality and a weighting function indicating the importance of a local region for quality score pooling. In this experiment, we will show the benefits brought by these two roles of VS map. The experiment was conducted on TID2013 dataset and we use SROCC as the performance measure. If VS map was used only as a weighting function for quality score pooling (in this case, the local quality map was computed based on the gradient modulus map and two chrominance channels), the SROCC obtained

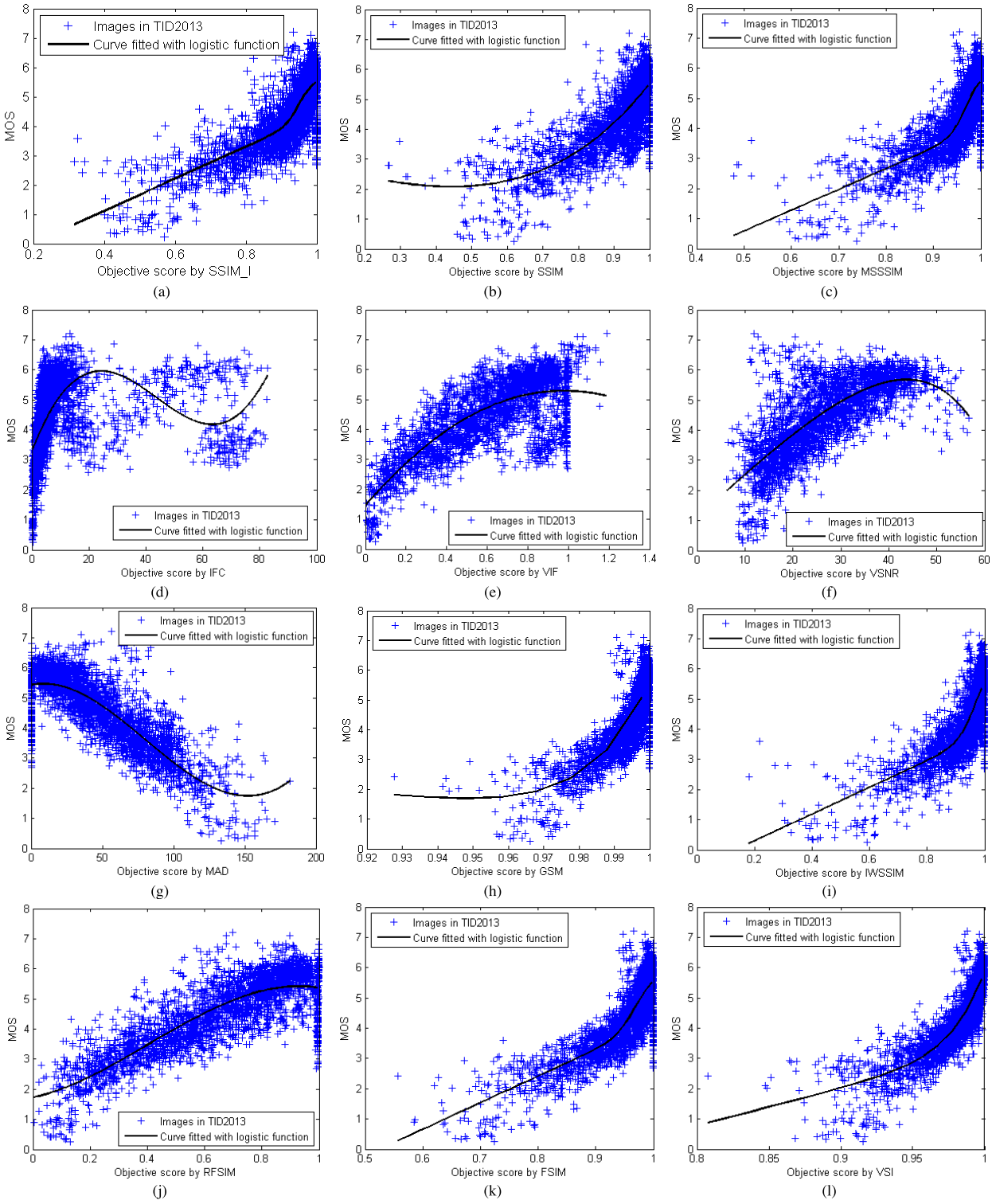

Fig. 4. Scatter plots of subjective MOS against scores obtained by model prediction on the TID2013 database. (a) SSIM_I, (b) SSIM, (c) MS-SSIM, (d) IFC, (e) VIF, (f) VSNR, (g) MAD, (h) GSM, (i) IWSSIM, (j) RFSIM, (k) FSIM, and (l) VSI.

was 0.8802. If VS map was used only as a feature map and we used a simple averaging strategy for quality score pooling, the SROCC was 0.8704. If VS map was used both as a feature map and a weighting function, the SROCC was 0.8965. From this experiment, it can be seen that to better explore the power of VS map, it should be used both as a feature map and a weighting function for quality score pooling.

C. Performance Evaluation

In this section, the prediction performance measured by SROCC, KROCC, PLCC, and RMSE of each competing IQA

TABLE VII
SROCC VALUES OF IQA INDICES FOR EACH TYPE OF DISTORTIONS

	Dis. Type	SSIM_I	SSIM	MS-SSIM	IFC	VIF	VSNR	MAD	GSM	IW-SSIM	RFSIM	FSIM	$FSIM_C$	VSI
TID 2013	AGN	0.8627	0.8671	0.8646	0.6612	0.8994	0.8271	0.8843	**0.9064**	0.8438	0.8878	0.8973	**0.9101**	**0.9460**
	ANC	0.7763	0.7726	0.7730	0.5352	0.8299	0.7305	0.8019	0.8175	0.7515	**0.8476**	0.8208	**0.8537**	**0.8705**
	SCN	0.8505	0.8515	0.8544	0.6601	0.8835	0.8013	**0.8911**	**0.9158**	0.8167	0.8825	0.8750	0.8900	**0.9367**
	MN	0.7895	0.7767	0.8073	0.6932	**0.8450**	0.7072	0.7380	0.7293	0.8020	**0.8368**	0.7944	**0.8094**	0.7697
	HFN	0.8688	0.8634	0.8604	0.7406	0.8972	0.8455	0.8876	0.8869	0.8553	**0.9145**	0.8984	**0.9040**	**0.9200**
	IN	0.7896	0.7503	0.7629	0.6408	**0.8537**	0.7363	0.2769	0.7965	0.7281	**0.9062**	0.8072	0.8251	**0.8741**
	QN	0.8411	0.8657	0.8706	0.6282	0.7854	0.8357	0.8514	**0.8841**	0.8468	**0.8968**	0.8719	**0.8807**	0.8748
	GB	**0.9724**	0.9668	0.9673	0.8907	0.9650	0.9470	0.9319	0.9689	**0.9701**	**0.9698**	0.9551	0.9551	0.9612
	DEN	0.9296	0.9254	0.9268	0.7779	0.8911	0.9081	0.9252	**0.9432**	0.9152	**0.9359**	0.9302	0.9330	**0.9484**
	JPEG	0.9227	0.9200	0.9265	0.8357	0.9192	0.9008	0.9217	0.9284	0.9187	**0.9398**	0.9324	**0.9339**	**0.9541**
	JP2K	0.9575	0.9468	0.9504	0.9078	0.9516	0.9273	0.9511	**0.9602**	0.9506	0.9518	0.9577	**0.9589**	**0.9706**
	JGTE	**0.8581**	0.8493	0.8475	0.7425	0.8409	0.7908	0.8283	0.8512	0.8388	0.8312	0.8464	**0.8610**	**0.9216**
	J2TE	0.8856	0.8828	0.8889	0.7769	0.8761	0.8407	0.8788	**0.9182**	0.8656	**0.9061**	0.8913	0.8919	**0.9228**
	NEPN	0.7885	0.7821	0.7968	0.5737	0.7720	0.6653	**0.8315**	**0.8130**	0.8011	0.7705	0.7917	0.7937	**0.8060**
	Block	0.4563	**0.5720**	0.4801	0.2414	0.5306	0.1771	0.2812	**0.6418**	0.3717	0.0339	0.5489	**0.5532**	0.1713
	MS	**0.7845**	0.7752	**0.7906**	0.5522	0.6276	0.4871	0.6450	**0.7875**	0.7833	0.5547	0.7531	0.7487	0.7700
	CTC	0.3800	0.3775	0.4634	0.1798	**0.8386**	0.3320	0.1972	**0.4857**	0.4593	0.3989	0.4686	0.4679	**0.4754**
	CCS	0.4208	0.4141	0.4099	0.4029	0.3099	0.3677	0.0575	0.3578	**0.4196**	0.0204	0.2748	**0.8359**	**0.8100**
	MGN	0.8092	0.7803	0.7786	0.6143	0.8468	0.7644	0.8409	0.8348	0.7728	0.8464	**0.8469**	**0.8569**	**0.9117**
	CN	0.8711	0.8566	0.8528	0.8160	0.8946	0.8683	0.9064	**0.9124**	0.8762	0.8917	0.9121	**0.9135**	**0.9243**
	LCNI	0.9173	0.9057	0.9068	0.8180	0.9204	0.8821	0.9443	**0.9563**	0.9037	0.9010	0.9466	**0.9485**	**0.9564**
	ICQD	0.8351	0.8542	0.8555	0.6006	0.8414	0.8667	0.8745	**0.8973**	0.8401	**0.8959**	0.8760	0.8815	**0.8839**
	CHA	0.8771	0.8775	0.8784	0.8210	0.8848	0.8645	0.8310	0.8823	0.8682	**0.8990**	0.8715	**0.8925**	**0.8906**
	SSR	0.9488	0.9461	0.9483	0.8885	0.9353	0.9339	0.9567	**0.9668**	0.9474	0.9326	0.9565	**0.9576**	**0.9628**
TID 2008	AGN	0.8169	0.8107	0.8086	0.5806	0.8797	0.7728	0.8386	**0.8606**	0.7869	0.8415	0.8566	**0.8758**	**0.9229**
	ANC	0.8192	0.8029	0.8054	0.5460	**0.8757**	0.7793	0.8255	0.8091	0.7920	0.8613	0.8527	**0.8931**	**0.9118**
	SCN	0.8264	0.8144	0.8209	0.5958	0.8698	0.7665	0.8678	**0.8941**	0.7714	0.8468	0.8483	**0.8711**	**0.9296**
	MN	0.7982	0.7795	0.8107	0.6732	**0.8683**	0.7295	0.7336	0.7452	0.8087	**0.8534**	0.8021	**0.8264**	0.7734
	HFN	0.8823	0.8729	0.8694	0.7318	0.9075	0.8811	0.8864	0.8945	0.8662	**0.9182**	0.9093	**0.9156**	**0.9253**
	IN	0.7232	0.6732	0.6907	0.5345	**0.8327**	0.6471	0.0650	0.7235	0.6465	**0.8806**	0.7452	0.7719	**0.8298**
	QN	0.8221	0.8531	0.8589	0.5857	0.7970	0.8270	0.8160	**0.8800**	0.8177	**0.8880**	0.8564	0.8726	**0.8731**
	GB	**0.9653**	0.9544	0.9563	0.8559	0.9540	0.9330	0.9196	**0.9600**	**0.9636**	0.9409	0.9472	0.9472	0.9529
	DEN	0.9600	0.9530	0.9582	0.7973	0.9161	0.9286	0.9433	**0.9725**	0.9473	0.9400	0.9603	**0.9618**	**0.9693**
	JPEG	0.9234	0.9252	0.9322	0.8180	0.9168	0.9174	0.9275	**0.9393**	0.9184	**0.9385**	0.9279	0.9294	**0.9616**
	JP2K	**0.9774**	0.9625	0.9700	0.9437	0.9709	0.9515	0.9707	0.9758	0.9738	0.9488	0.9773	**0.9780**	**0.9848**
	JGTE	**0.8868**	0.8678	0.8681	0.7909	0.8585	0.8055	0.8661	**0.8790**	0.8588	0.8503	0.8708	0.8756	**0.9160**
	J2TE	**0.8654**	0.8577	0.8606	0.7301	0.8501	0.7909	0.8394	**0.8936**	0.8203	0.8592	0.8544	0.8555	**0.8942**
	NEPN	0.7468	0.7107	0.7377	**0.8418**	0.7619	0.5716	**0.8287**	0.7386	**0.7724**	0.7274	0.7491	0.7514	0.7699
	Block	0.8062	0.8462	0.7546	0.6770	0.8324	0.1926	0.7970	**0.8862**	0.7623	0.6258	**0.8492**	**0.8464**	0.6295
	MS	**0.7357**	**0.7231**	**0.7336**	0.4250	0.5096	0.3715	0.5163	0.7190	0.7067	0.4178	0.6720	0.6554	0.6714
	CTC	0.5391	0.5246	0.6381	0.1713	**0.8188**	0.4239	0.2723	**0.6691**	0.6301	0.5823	0.6481	0.6510	**0.6557**
CSIQ	AGWN	0.8726	0.8974	0.9471	0.8431	**0.9575**	0.9241	**0.9541**	0.9440	0.9380	0.9441	0.9262	0.9359	**0.9636**
	JPEG	0.9619	0.9546	0.9634	0.9412	**0.9705**	0.9036	0.9615	0.9632	**0.9662**	0.9502	0.9654	**0.9664**	0.9618
	JP2K	**0.9784**	0.9606	0.9683	0.9252	0.9672	0.9480	**0.9752**	0.9648	0.9683	0.9643	0.9685	**0.9704**	0.9694
	AGPN	0.8783	0.8922	0.9331	0.8261	**0.9511**	0.9084	**0.9570**	0.9387	0.9059	0.9357	0.9234	0.9370	**0.9638**
	GB	**0.9773**	0.9609	0.9711	0.9527	**0.9745**	0.9446	0.9602	0.9589	**0.9782**	0.9634	0.9729	0.9729	0.9679
	GCD	0.8162	0.7922	**0.9526**	0.4873	0.9345	0.8700	0.9207	0.9354	**0.9539**	**0.9527**	0.9420	0.9438	0.9504
LIVE	JP2K	0.9675	0.9614	0.9627	0.9113	0.9696	0.9551	0.9676	**0.9700**	0.9649	0.9323	**0.9717**	**0.9724**	0.9604
	JPEG	0.9799	0.9764	0.9815	0.9468	**0.9846**	0.9657	0.9764	0.9778	0.9808	0.9584	**0.9834**	**0.9840**	0.9761
	AWGN	0.9749	0.9694	0.9733	0.9382	**0.9858**	0.9785	**0.9844**	0.9774	0.9667	0.9799	0.9652	0.9716	**0.9835**
	GB	0.9660	0.9517	0.9542	0.9584	**0.9728**	0.9413	0.9465	0.9518	**0.9720**	0.9066	**0.9708**	0.9708	0.9527
	FF	**0.9610**	0.9556	0.9471	**0.9629**	**0.9650**	0.9027	0.9569	0.9402	0.9442	0.9237	0.9499	0.9519	0.9430

index on each benchmark dataset is given. The results are listed in Table IV. For each performance measure, the two IQA indices producing the best results are highlighted in boldface. In addition, as suggested by Wang and Li [7], in order to provide an evaluation of the overall performance of the evaluated IQA indices, in Table V we present their weighted-average SROCC, KROCC and PLCC results over four datasets and the weight assigned to each dataset linearly depends on the number of distorted images contained in that dataset. The ranking of the weighted-average performances of the evaluated IQA indices based on three different performance metrics, SROCC, KROCC, and PLCC, is presented in Table VI.

From Table IV, it can be seen that VSI performs consistently well on all the benchmark databases. Particularly, it performs greatly better than all the other competitors on the two largest databases, TID2013 and TID2008. On CSIQ and LIVE, even though it is not the best, VSI performs only slightly worse than the best results. By contrast, for the other methods, they may work well on some database but fail to provide good results on other databases. For example, though VIF and MAD can get pleasing results on LIVE, they perform quite poor on TID2013 and TID2008. In Tables V and VI, the statistical superiority of VSI to the other competing IQA indices is clearly exhibited since no matter which performance measure is used, VSI always achieves the best overall results. Hence, we can conclude that objective scores predicted by VSI correlate much more consistently with subjective evaluations than all the other IQA indices evaluated.

In addition, in Fig. 4 we show the scatter plots of subjective scores against objective scores predicted by some representative IQA indices (SSIM_I, SSIM, MS-SSIM, IFC, VIF, VSNR, MAD, GSM, IW-SSIM, RFSIM, FSIM, and VSI) on TID2013, at present the largest benchmark database for evaluating IQA indices. The curves shown in Fig. 4 were obtained by a nonlinear fitting using Eq. (9). From Fig. 4, it can also be seen that objective scores predicted by VSI is more correlated with subjective ratings than the other competitors.

D. Performance Comparison on Individual Distortion Types

To more comprehensively evaluate an IQA index's ability to predict image quality degradations caused by specific types of distortions, in this experiment, we examined the performance of the competing methods on each type of distortions. We use SROCC as the performance measure. In fact, by using the other measures, such as KROCC, PLCC, and RMSE, similar conclusions could be drawn. The results are summarized in Table VII. There are a total of 52 groups of distorted images in the four databases.

For each database and each distortion type, the first three IQA indices producing the highest SROCC values are highlighted in boldface. It can be seen that VSI is among the top 3 indices 34 times, followed by FSIM_C (27 times) and GSM (25 times). Thus, we can have the following conclusions. In general, when the distortion is of a specific type, VSI performs the best, while FSIM_C and GSM can have comparable performance. Moreover, in this case, VSI, FSIM_C and GSM perform much better than the other IQA indices.

E. Computational Cost

The running speed of each selected IQA index was also evaluated. Experiments were performed on a standard HP Z620 workstation with a 3.2GHZ Intel Xeon E5-1650 CPU and an 8G RAM. The software platform was Matlab R2012a. The time cost consumed by each IQA index for measuring the similarity of a pair of 384×512 color images (taken from TID2013) is listed in Table VIII. From Table VIII we can see that VSI has a moderate computational complexity. Particularly, it runs much faster than the other modern IQA indices which could achieve state-of-the-art prediction performances, such as FSIM_C, FSIM, IW-SSIM, and MAD.

TABLE VIII
TIME COST OF EACH IQA INDEX

IQA Index	Time cost (milliseconds)
SSIM_I [29]	169.8
SSIM [5]	16.8
MS-SSIM [6]	69.2
IFC [9]	523.0
VIF [8]	530.6
VSNR [4]	23.2
MAD [11]	682.0
GSM [13]	17.2
IW-SSIM [7]	258.0
RFSIM [10]	48.4
FSIM [12]	127.2
FSIM_C [12]	130.4
VSI	95.4

V. CONCLUSIONS

In this paper, we proposed a novel metric for IQA, namely visual saliency based index (VSI). It is based on the assumption that an image's VS map has a close relationship with its perceptual quality. In VSI, the visual saliency (VS) map is explored at two stages. At the stage of local quality map computation, the VS map is taken as an image feature; while at the quality score pooling stage, it is used as a weighting function to characterize the importance of a local image region. Several representative VS models were examined under our framework of VSI for IQA tasks, and among them the SDSP model performs the best. The proposed VSI was thoroughly tested and compared with the other 12 state-of-the-art or widely cited IQA indices on four large-scale benchmark IQA databases. The results demonstrated that the proposed IQA index VSI could yield statistically much better results in terms of the prediction accuracy than all the other competing methods evaluated while still maintaining a moderate computational complexity. To some extent, VSI is actually an open framework; therefore, with the advent of even more powerful VS models, VSI of course can be improved accordingly.

REFERENCES

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA, USA: Morgan & Claypool, 2006.
[2] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.
[4] 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.
[5] 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.
[6] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.
[7] 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.

[8] 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.

[9] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[10] L. Zhang, D. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 321–324.

[11] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 001006:1–001006:21, Jan. 2010.

[12] L. Zhang, D. 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.

[13] 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.

[14] 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, May 2011.

[15] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1477–1480.

[16] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2131–2146, Nov. 2011.

[17] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, Jan. 2013.

[18] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 55–69, Jan. 2013.

[19] U. Engelke, H. Kaprykowsky, H. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 50–59, Nov. 2011.

[20] E. C. L. Vu and D. M. Chandler, "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Mar. 2008, pp. 73–76.

[21] E. C. Larson, C. Vu, and D. M. Chandler, "Can visual fixation patterns improve image fidelity assessment?" in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2572–2575.

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

[23] E. C. Larson and D. M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," *Proc. SPIE, Vis. Commun. Image Process.*, vol. 6822, pp. 68222A:1–68222A:16, Jan. 2008.

[24] U. Engelke and H.-J. Zepernick, "Framework for optimal region of interest-based quality assessment in wireless imaging," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011005-1–011005-13, 2010.

[25] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[26] I. Gkioulekas, G. Evangelopoulos, and P. Maragos, "Spatial Bayesian surprise for image saliency and quality assessment," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1081–1084.

[27] Y. Tong, H. Konik, F. Cheikh, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *J. Imag. Sci. Technol.*, vol. 54, no. 3, pp. 30503:1–30503:14, May 2010.

[28] Q. Ma, L. Zhang, and B. Wang, "New strategy for image and video quality assessment," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011019:1–011019:14, Jan. 2010.

[29] M. C. Q. Farias and W. Y. L. Akamine, "On performance of image quality metrics enhanced with visual attention computational models," *Electron. Lett.*, vol. 48, no. 11, pp. 631–633, May 2012.

[30] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 547–558, Aug. 2010.

[31] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Do video coding impairments disturb the visual attention deployment," *Signal Process., Image Commun.*, vol. 25, no. 8, pp. 597–609, Sep. 2010.

[32] J. You, J. Korhonen, A. Perkis, and T. Ebrahimi, "Balancing attended and global stimuli in perceived video quality assessment," *IEEE Trans. Multimedia*, vol. 13, no. 6, pp. 1269–1285, Dec. 2011.

[33] J. You, A. Perkis, M. M. Hannuksela, and M. Gabbouj, "Perceptual quality assessment based on visual attention analysis," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 561–564.

[34] 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.

[35] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiol.*, vol. 4, no. 4, pp. 219–227, Apr. 1985.

[36] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vis. Res.*, vol. 49, no. 10, pp. 1295–1306, Jun. 2009.

[37] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, vol. 19. Cambridge, MA, USA: MIT Press, 2007, pp. 545–552.

[38] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2214–2219.

[39] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[40] 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.

[41] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *J. Vis.*, vol. 9, no. 3, pp. 1–24, Mar. 2009.

[42] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, pp. 15.1–15.27, Nov. 2009.

[43] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[44] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2106–2113.

[45] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 853–860.

[46] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 171–175.

[47] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process.*, Jun. 2013, pp. 106–111.

[48] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Stanford, CT, USA: Cengage Learning, 2008.

[49] B. Jähne, H. Haubecker, and P. Geibler, *Handbook of Computer Vision and Applications*. New York, NY, USA: Academic, 1999.

[50] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 12, pp. 1338–1350, Dec. 2001.

[51] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, and A. Dev, "Color and scale: The spatial structure of color images," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 331–341.

[52] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30–45, 2009.

[53] 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.

Lin Zhang (M'11) received the B.S. and M.S. degrees from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Department of Computing, Hong Kong Polytechnic University, Hong Kong, in 2011. In 2011, he was a Research Assistant with the Department of Computing, Hong Kong Polytechnic University. He is currently an Assistant Professor with the School of Software Engineering, Tongji University, Shanghai. His current research interests include biometrics, pattern recognition, computer vision, and perceptual image/video quality assessment.

Ying Shen (M'13) received the B.S. and M.S. degrees from Software School, Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2009, respectively, and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2012. In 2013, she joined the School of Software Engineering, Tongji University, Shanghai, where she is currently an Assistant Professor. Her research interests include bioinformatics and pattern recognition.

Hongyu Li received the B.E. degree from Tongji University, Shanghai, China, in 2000, the Ph.D. degree from the Department of Computer Science, Fudan University, Shanghai, in 2008, and the Ph.D. degree from Department of Computer Science, University of Eastern Finland, Joensuu, Finland. In 2008, he joined the School of Software Engineering at Tongji University, where he is currently an Associate Professor. His research interests include computer vision, pattern recognition, and motion analysis.