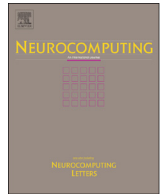




ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Learning a blind quality evaluation engine of screen content images

Ke Gu^{a,*}, Guangtao Zhai^b, Weisi Lin^a, Xiaokang Yang^b, Wenjun Zhang^b^a School of Computer Engineering, Nanyang Technological University, Singapore^b Shanghai Key Laboratory of Digital Media Processing and Transmissions, Institute of Image Communication & Information Processing, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:

Received 22 August 2015

Received in revised form

17 October 2015

Accepted 1 November 2015

Keywords:

Screen content images (SCIs)

Image quality assessment (IQA)

Blind/no-reference (NR)

Statistical model

Machine learning

ABSTRACT

We in this paper investigate how to blindly predict the visual quality of a screen content image (SCI). With the popularity of multi-client and remote-controlling systems, SCIs and the relevant applications have been a hot research topic. In general, SCIs contain texts or graphics in cartoons, ebooks or captures of computer screens. As for blind quality assessment (QA) of natural scene images (NSIs), it has been well established since NSIs possess certain statistical properties. SCIs however do not have reliable statistic models so far and thus the associated blind QA task is hard to be addressed. Aiming at solving this problem, we first extract 13 perceptual-inspired features with the free energy based brain theory and structural degradation model. In order to avoid the overfitting and guarantee the independence of training and testing samples, we then collect 100,000 images and use their objective quality scores computed via a high-accuracy full-reference QA method for SCIs as labels, before learning a new blind quality measure from aforementioned 13 features to the objective quality score. Experimental results performed on a large-scale screen image quality assessment database (SIQAD) demonstrate that the proposed blind quality metric has a good correlation with human perception of quality, even superior to state-of-the-art full-, reduced- and no-reference QA algorithms.

© 2016 Published by Elsevier B.V.

1. Introduction

Screen content images (SCIs), a mixture of graphics, texts and natural scene images, typically include ebooks, emaps, cartoons and captures of computer screens. As the multi-client collaboration and communication systems, such as remote computing [1], cloud gaming [2] and cloud-guided enhancement [3], are increasingly widespread, SCIs have recently aroused broad interests from multimedia and image processing communities. During the process of acquisition, compression and transmission, SCIs are inevitably degraded by various distortions. For instance, contrast altered SCIs are produced for different settings of screen luminance and contrast. Compressed SCIs will generate blockiness and quantization. Noise often appears when SCIs are transmitted. Hence we in this paper propose an objective quality metric for the degradation measurement of SCIs.

Objective image quality assessment (IQA) can be divided into three types: *full reference (FR)*, *reduced reference (RR)* and *no reference (NR)*. With the known reference image, FR-IQA models compute the distance between the reference and distorted images. The variation in image structures is a good distance metric, and on

this basis, some modern FR-IQA models have been devised in [4–12]. Nonetheless, these methods do not perform nicely for the quality evaluation of SCIs. Therefore we have recently proposed an effective structure-induced quality metric (SIQM) [13], which combines the structural similarity measurement with the structural degradation model. Using the new large-scale screen image quality assessment database (SIQAD) [14,15], the SIQM metric was shown to outperform state-of-the-art FR-IQA methods.

Using partial reference image or some extracted features, RR-IQA models work with various kinds of strategies. For example, using the advanced multi-resolution analysis tool - wavelet-based contourlet transform (WBCT) – for valid feature extraction, Gao et al. proposed the WBCT-based RR (RR-WBCT) metric [16]. The free energy based distortion metric (FEDM) [17] was devised to simulate the working of the brain's perception to image quality for RR scenarios. Taking the low-level vision into account, Narwaria et al. developed the Fourier transform based quality measure (FTQM) [18] based on the phase and magnitude of the Fourier transform. By largely reducing the amount of reference information used in SSIM, the structural degradation model (SDM) [19] succeeded in converting SSIM into the effective RR-IQA technique with only few features.

In most cases the (complete or partial) reference image is not available, and thus NR-IQA models become the only choice which has been intensively studied during the past few years. One type of

* Corresponding author. Tel.: +86 13764402720

E-mail addresses: guke.doctor@gmail.com, gukesjtuee@gmail.com (K. Gu).

classical blind IQA metrics is to predict the visual quality of specific distortion types, e.g. blockiness [20,21], image sharpness/blurriness [22–24], and multiple distortions [25]. Another type of popular and mature strategy for the exploration of NR-IQA performs by extracting natural scene statistics (NSS) based features to be mapped to the visual quality score with machine learning. In [26], Moorthy et al. have found that the subband coefficients after discrete wavelet transform (DWT) can be well-fitted using a generalized Gaussian distribution (GGD), and thus developed a two-stage system by distortion classification before quality prediction. In [27], Saad et al. observed that the coefficients after discrete cosine transform (DCT) have better regulations, and without identifying distortion types, they can directly predict the visual quality from local DCT coefficients. In [28], Mittal et al. discovered the NSS in the spatial domain, and thereby proposed an effective and efficient blind quality metric that is usually considered to be the benchmark in existing researches of NR-IQA methods.

As compared to the opinion-aware (OA) NR-IQA metrics stated above, there has been recently a trend towards constructing opinion-unaware (OU) blind IQA techniques without using human scored images. In [29], natural image quality evaluator (NIQE) was explored, without any prior knowledge of image contents or distortion categories, by quantifying the deviation between the model statistics of the distorted image and those extracted from a collection of pristine natural images. To make the NIQE more complete, Zhang et al. in [30] recently came up with the integrated-local NIQE (IL-NIQE) with three new statistical features (i.e. gradient statistics, statistics of Log-Gabor filter responses and statistics of colors) included.

Yet the NSS model is violated when the images do not come from a natural source, such as computer graphics and document images, or when natural images are corrupted [31], making the aforementioned NR-IQA models work not validly for SCIs. In our recent work, a new statistical model was built on natural scene images, which derives a high-accuracy blind quality metric [32]. It is surprising that the statistical regulation approximately exists in lossless screen content images. Hence we in this paper propose a novel blind quality measure for SCIs (BQMS), by first extracting features based on a new screen content statistics model before learning the module from the feature space to the quality score using a large set of training data.

Compared with the previous works, the main contributions of this paper are summarized as follows: (1) to the best of our knowledge, this work is the first one blindly evaluating the quality of SCIs; (2) we construct the first screen content statistics (SCS) model; (3) in contrast to previous works that only use a few hundreds of training samples, a vast number of 100,000 training data are applied for building our BQMS method, in order to avoid the overfitting in the learning; (4) the proposed blind quality metric outperforms state-of-the-art FR-, RR- and NR-IQA approaches using the new SIQAD database dedicated to SCIs [14,15].

The remainder of this paper are arranged as follows: Section 2 presents the statistic model of screen content images before constructs our blind IQA method. In Section 3, thorough experiments are conducted using the large-scale SIQAD database to verify the effectiveness of our BQMS with recently developed IQA metrics. We conclude this paper in Section 4.

2. Methodology

2.1. Screen content statistics

We establish the SCS model using two sets of features of free energy measure and structural degradation information. The free energy theory, from the perspective of human action, perception

and learning, explains and unifies most existing brain principles [33]. Its core concept lies in that, in the brain, the so-called internal generative model controls the cognitive process, with a constructive way to predict meaningful parts and abandon uncertain residuals from an input visual signal. This residual “gap” always exists and it has been found to be highly related to the image quality [17].

To specify, the autoregressive (AR) model, due to its good descriptive ability [34,35], is used to predict meaningful parts of an input image \mathbf{x} , defined as follows:

$$x_n = \mathcal{X}^c(x_n)\mathbf{p} + q_n \quad (1)$$

where n is the pixel index. $\mathcal{X}^c(x_n)$ is a vector of the c nearest pixel indices of x_n . $\mathbf{p} = (p_1, \dots, p_c)^T$ is a vector of AR model coefficients. The superscript “ T ” indicates transpose. q_n is the error term. To solve \mathbf{p} , we need to rewrite the linear system to be the matrix form:

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \|\mathbf{x} - \mathbf{X}\mathbf{p}\|_2 \quad (2)$$

where $\mathbf{x} = (x_1, \dots, x_c)^T$; $\mathbf{X}(t, :) = \mathcal{X}^c(x_t)$. We use the least squares method to find the solution to be $\hat{\mathbf{p}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}$. In real application, an 8-th order AR model is trained in a local 7×7 neighborhood before the pseudo-inverse of the 48×8 matrix are solved via Gaussian eliminations. The predicted $\hat{\mathbf{x}}$ can be estimated as

$$\hat{x}_n = \mathcal{X}^c(x_n)\hat{\mathbf{p}}. \quad (3)$$

Then, following the analysis in [17], the free energy of the image \mathbf{x} can be approximated as the entropy of residual “gap”:

$$F_e(\mathbf{x}) = - \sum_i P_i(\bar{\mathbf{x}}) \log P_i(\bar{\mathbf{x}}) \quad (4)$$

where $\bar{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ being the prediction error between the input image and its predicted one; $P_i(\bar{\mathbf{x}})$ is the probability density of grayscale i in $\bar{\mathbf{x}}$.

Structural degradation model derives from the observation that, after the low-pass filtering, images corrupted by various distortion types and intensities will appear as distinct degrees of spatial frequency reduction. For example, with the same low-pass filter, noisy images in general have higher degrees of frequency decrement than blurred images. As thus, we define the structural degradation information by

$$S_\mu(\mathbf{x}) = E \left(\frac{\sigma(\hat{\mu}_x \hat{\mu}_x) + \epsilon}{\sigma(\hat{\mu}_x) \sigma(\hat{\mu}_x) + \epsilon} \right) \quad (5)$$

$$S_\sigma(\mathbf{x}) = E \left(\frac{\sigma(\hat{\sigma}_x \hat{\sigma}_x) + \epsilon}{\sigma(\hat{\sigma}_x) \sigma(\hat{\sigma}_x) + \epsilon} \right) \quad (6)$$

where $\hat{\mu}_x$ and $\hat{\sigma}_x$ represent the local mean and variance of \mathbf{x} , computed by a normalized Gaussian kernel $\mathbf{w} = \{w(u, v) | u = -U, \dots, U, v = -V, \dots, V\}$ with the standard deviation of 1.5; $\hat{\mu}_x$ and $\hat{\sigma}_x$ are similarly defined using the impulse function instead of the Gaussian kernel; $E(\cdot)$ is to compute the global average; $\sigma_{(AB)}$ indicates the local covariance between A and B ; ϵ is a small positive constant to avoid division-by-zero.

Table 1

Definitions of the 12 structural degradation features for interior and exterior parts and three pairs of (U, V) .

Definitions of structural degradation features	Interior parts		Exterior parts	
	S_μ	S_σ	S_μ	S_σ
$(U, V) = (1, 1)$	S_μ^i	S_σ^i	S_μ^e	S_σ^e
$(U, V) = (3, 3)$	S_μ^s	S_σ^s	S_μ^{es}	S_σ^{es}
$(U, V) = (5, 5)$	S_μ^l	S_σ^l	S_μ^{el}	S_σ^{el}

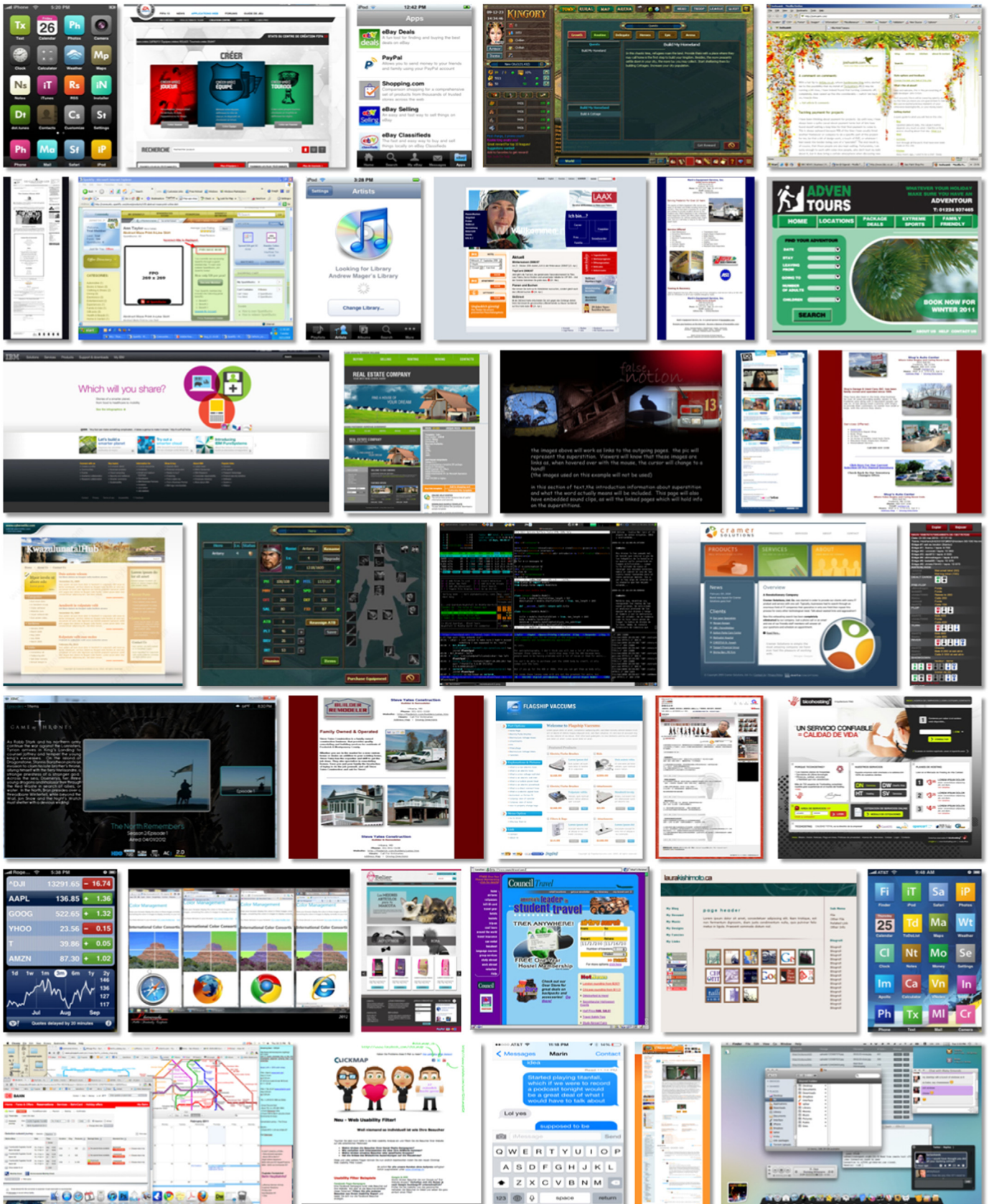


Fig. 1. Representative 37 screen content (“webpage” and “screen snap”) images downloaded from the website of “Google Images”.

Further, we modify the structural degradation information with three considerations. First, we choose three couples of (U, V) as $(1, 1)$, $(3, 3)$ and $(5, 5)$ in the kernels, to introduce different

amounts of neighboring information. Second, $S_{\mu}(\mathbf{x})$ and $S_{\sigma}(\mathbf{x})$ are reversed when $F_e(\mathbf{x})$ is bigger than T , making the structural degradation information of images, which are corrupted by

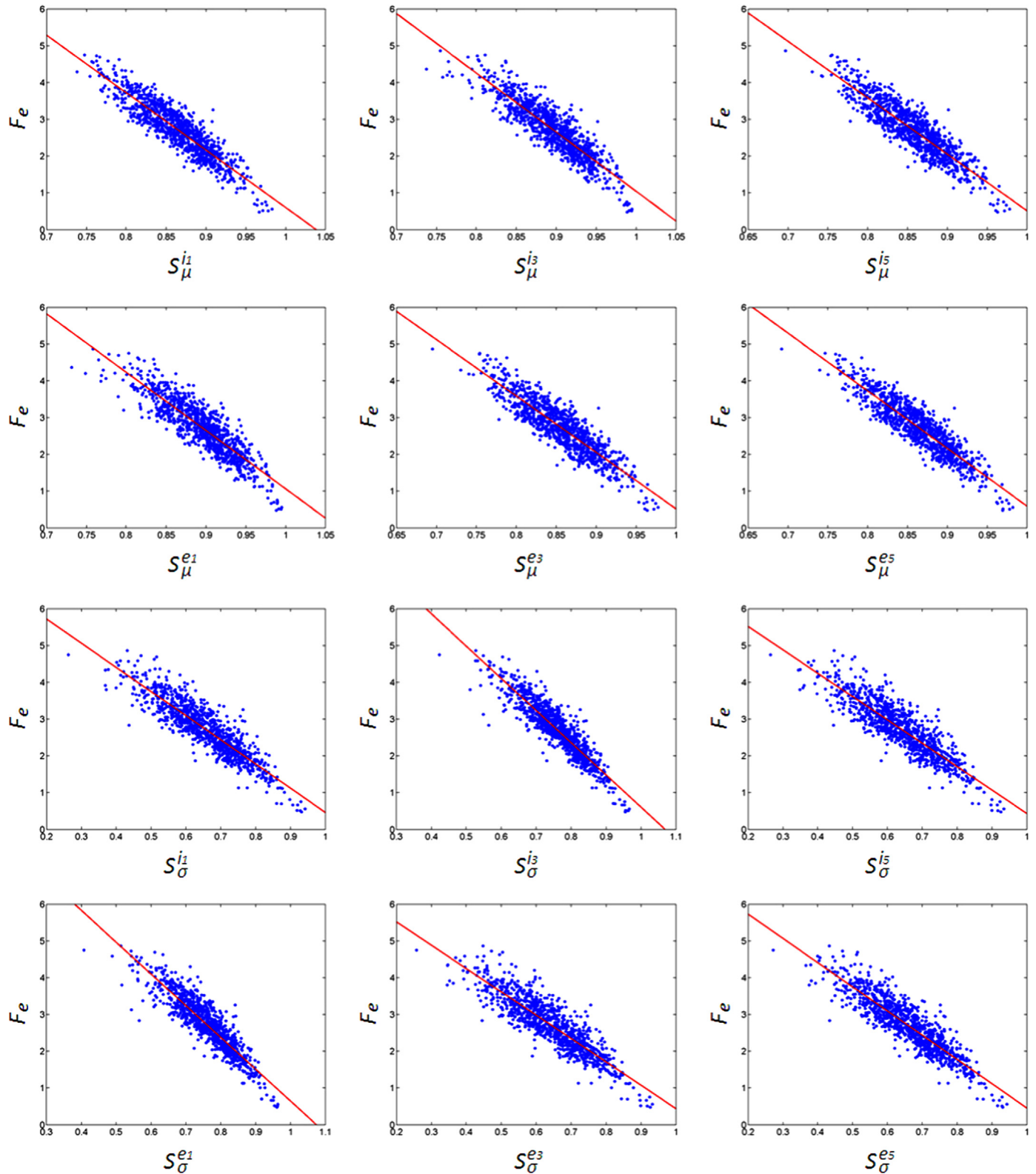


Fig. 2. Scatter plots of the structural degradation features $S_{\mu}^{\tau}(\mathbf{x})$ and $S_{\sigma}^{\tau}(\mathbf{x})$ ($\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$) versus the free energy feature $F_e(\mathbf{x})$ on 1000 original (high-quality) screen content images. The straight lines are fitted with the least square method. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

various distortion types such as blur and noise, have consistent changes with the visual quality. Third, $S_{\mu}(\mathbf{x})$ and $S_{\sigma}(\mathbf{x})$ are computed in the interior 6×6 part and exterior block-edge part, to

discriminate different frequency decrease in interior and exterior areas caused by JPEG compression. Table 1 presents all the 12 structural degradation features.

In order to validate the effectiveness of the screen content statistics, we have first downloaded thousands of “webpage” and “screen snap” images from the “Google Images” website, then used human eyes to observe the overall images one by one, and finally picked up 1000 high-quality images. Those chosen images are of size from 500×221 to 4218×1265 , part of which are exhibited in Fig. 1. It deserves to emphasize that there is not a bit overlapping between the chosen 1000 high-quality images and the original images in the testing SIQAD database. We then compare the structural degradation features $S_\mu^r(\mathbf{x}_0)$ and $S_\sigma^r(\mathbf{x}_0)$, where \mathbf{x}_0 means the original SCI and $\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$, with the free energy feature $F_e(\mathbf{x}_0)$ on the selected 1000 original screen content images and draw the 12 scatter plots in Fig. 2. As shown, the linear dependence of the free energy feature and structural degradation features provides an opportunity to characterize distorted SCIs without original references. We fit the linear regression model:

$$F_e(\mathbf{x}_0) = a_\tau \cdot S_\mu^r(\mathbf{x}_0) + b_\tau \quad (7)$$

$$F_e(\mathbf{x}_0) = c_\tau \cdot S_\sigma^r(\mathbf{x}_0) + d_\tau \quad (8)$$

where $\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$; the parameters a_τ , b_τ , c_τ and d_τ are obtained using the least square method, whose values are reported in Table 2.

Table 2

The estimates of model parameters a_τ , b_τ , c_τ and d_τ for S_μ^r and S_σ^r ($\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$) using the least square method.

Variables	a_τ	b_τ	c_τ	d_τ	
$S_\mu^{i_1}$	-15.690	16.281	$S_\sigma^{i_1}$	-8.7722	9.3716
$S_\mu^{i_3}$	-6.3634	6.7923	$S_\sigma^{i_3}$	-16.115	17.153
$S_\mu^{i_5}$	-15.358	15.872	$S_\sigma^{i_5}$	-6.5959	7.0473
$S_\mu^{e_1}$	-15.630	16.229	$S_\sigma^{e_1}$	-8.6480	9.2890
$S_\mu^{e_3}$	-6.3552	6.7896	$S_\sigma^{e_3}$	-15.873	16.934
$S_\mu^{e_5}$	-15.353	15.869	$S_\sigma^{e_5}$	-6.5785	7.0370

Table 3

Summarization of the used thirteen features. F_s indicates the features of structural degradation information including S_μ^r and S_σ^r .

Feature ID	Feature notation	Feature description
f_{01}	F_e	Free energy entropy based on AR model
$f_{02} - f_{07}$	N_μ^r and N_σ^r , $\tau = \{i_1, i_3, i_5\}$	Difference of F_s and F_e in interior parts
$f_{08} - f_{13}$	N_μ^r and N_σ^r , $\tau = \{e_1, e_3, e_5\}$	Difference of F_s and F_e in exterior parts

2.2. Image quality evaluation

Based on the linear relationship of the free energy feature and the structural degradation features on original SCIs, we further define $N_\mu^r(\mathbf{x}) = F_e(\mathbf{x}) - (a_\tau \cdot S_\mu^r(\mathbf{x}) + b_\tau)$ and $N_\sigma^r(\mathbf{x}) = F_e(\mathbf{x}) - (c_\tau \cdot S_\sigma^r(\mathbf{x}) + d_\tau)$, since N_μ^r and N_σ^r are close to zero for high-quality images (i.e. without distortions or with very few distortions), while they will be far from zero when distortions appear and become larger. In addition, we also consider the free energy feature which has a good correlation with human ratings on noisy and blurred images [17]. Lastly, thirteen features used in our BQMS metric are summarized in Table 3.

The reason for selecting the adopted 13 features might be explained from the following two perspectives. The first one is that, for high-quality screen content images, there exist high correlations between structural degradation information and free energy feature, as given in Fig. 2, since the Gaussian convolution kernel used in the computation of structural degradation information is a uniform low-pass filter, and conversely, the AR model well preserves image textures but is not good at processing image edges. The second perspective is that, e.g. for white noise, the AR model is generally of the better denoising ability than the Gaussian filter, which gives rise to the deviations from the fitted red lines shown in Fig. 2. Based on these two considerations, we have a reason to believe the extracted features can be learned to infer the visual quality scores close to the real subjective opinions.

After the feature extraction, it needs to learn a regression module from features to the image quality score. In order to avoid the problem of overfitting, this paper adopts a total of 100,000 images as training samples. To specify, we apply the aforesaid 1000 high-quality SCIs as original references, which are exactly content-independent of the pristine images used in the testing SIQAD database. We then simulated the distorted SCIs via six typical distortion types, i.e. Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), and JPEG and JPEG2000 (JP2K) compressions. Note that these six distortion types are also included in the SIQAD database. For each image, we create its distorted ones of each type on more than 15 distortion levels with Matlab embedded functions. We finally generate a training dataset consisting of 100,000 distorted screen content images.

But how to label these distorted SCIs? In [36], the authors, instead of training on human opinion scores, used synthetic scores yielded from FR-IQA approaches as training labels. Likewise, in this paper we apply the recently designed SIQM [13], which improves the SSIM weighted by the structural degradation measurement [19]. The SIQM method was proven to work validly for the screen content IQA, and thus the associated quality scores of distorted SCIs are used for approximating subjective ratings.

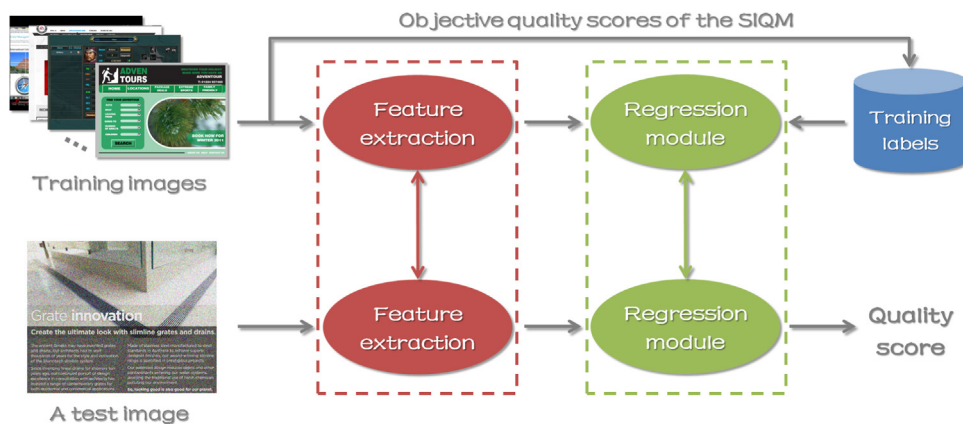


Fig. 3. The primary framework of the blind BQMS method based on statistical model of screen content images.

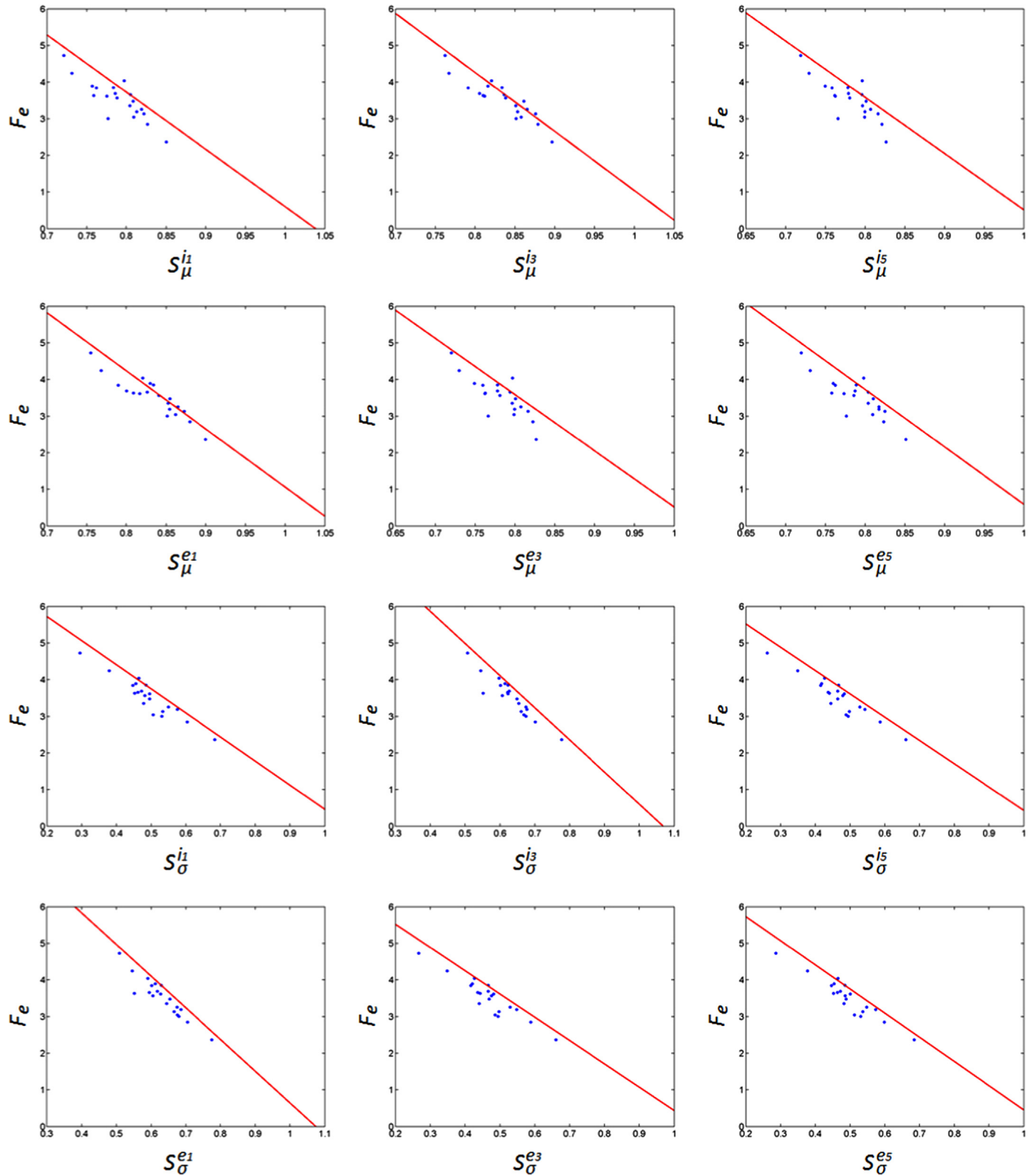


Fig. 4. Scatter plots of the structural degradation features $S_{\mu}^{\tau}(\mathbf{x})$ and $S_{\sigma}^{\tau}(\mathbf{x})$ ($\tau = \{i_1, i_3, i_5, e_1, e_3, e_5\}$) versus the free energy feature $F_e(\mathbf{x})$ on 20 original screen content images in the SIQAD database. The red lines are based on model parameters in Table 2. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Having prepared the training set, we use the support vector regression (SVR) [37] to learn the module in the proposed BQMS. In reality, the SVR-based machine learning technology has penetrated into various kinds of research fields, such as used for image

retrieval [38] and quality evaluation [39]. Given a training dataset $D = \{(x_1, y_1), \dots, (x_r, y_r)\}$, where x_i and y_i are respectively a feature vector of $f_{01} - f_{13}$ in Table 3 and the target output of the SIQM score derived from the i -th training image. Let parameters $\alpha > 0$

and $\varepsilon > 0$, the standard form of SVR can be expressed as:

$$\begin{aligned} \min_{\delta, \mathbf{v}, \mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \alpha \left(\sum_{i=1}^r v_i + \sum_{i=1}^r v_i' \right) \\ \text{s.t.} \quad & \mathbf{w}^T \phi(x_i) + \delta - y_i \leq \varepsilon + v_i, \\ & y_i - \mathbf{w}^T \phi(x_i) - \delta \leq \varepsilon + v_i', \\ & v_i, v_i' \geq 0, \quad i = 1, \dots, r. \end{aligned}$$

where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel function, which is assigned as the Radial Basis Function (RBF) kernel having $K(x_i, x_j) = \exp(-k \|x_i - x_j\|^2)$. The parameters α , ε and k are to be determined using the training samples.

Finally, we illustrate the primary framework of the proposed blind BQMS technique in Fig. 3, in order to help readers readily understand how to apply the metric to predict the visual quality of a distorted screen content image.

3. Experimental results

3.1. Evaluation metrics, databases and protocols

Using the newly released SIQAD database [14,15], the performance of our BQMS model is computed and compared with state-of-the-art IQA techniques, consisting of: (1) FR FSIM [6], GSI [7], IGM [9], VSI [10] and SIQM [13]; (2) RR FEDM [17], QFTB [18] and SDM [19]; (3) NR DIIVINE [26], BLIINDS-II [27], BRISQUE [28], NIQE [29], IL-NIQE [30] and NFERM [32]. The SIQAD database was constructed dedicated to the screen content IQA, which is made up of 980 distorted SCIs created by using seven distortion types at seven degradation levels to corrupt 20 pristine SCIs. Apart from the aforementioned six types of distortions (i.e. GN, GB, MB, CC, JPEG and JP2K), the SIQAD database also includes the Layer Segmentation based Coding (LSC).

As the video quality experts group (VQEG) suggested, we first remove the nonlinearity of objective quality predictions through a five-parameter nonlinear regression function [40]:

$$\mathbf{v}_c = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(\mathbf{v}_o - \beta_3)}} \right) + \beta_4 \mathbf{v}_o + \beta_5 \quad (9)$$

where \mathbf{v}_o , \mathbf{v}_c and \mathbf{v}_s respectively stand for the vectors of original IQA scores, converted IQA scores after the regression of Eq. (9) and subjective MOS values; model parameters β_1, \dots, β_5 are decided during the curve fitting process. After the regression, we calculate four important performance indices, Spearman rank order correlation coefficient (SRC), Kendall's rank-order correlation coefficient (KRC), Pearson linear correlation coefficient (PLC), and root mean-squared error (RMS) to testify and compare the proposed BQMS with other testing IQA metrics on the SIQAD database. The former two evaluations are used for prediction monotonicity, the third one for prediction accuracy, and the last one for prediction consistency. A value close to one for PLC, SRC and KRC while close to zero for RMS indicates the superior correlation performance in line with subjective human ratings. Interested readers can refer to [40] for a more detailed description of these measures.

3.2. Validation of screen content statistics model

Additionally, we also check the robustness of the proposed screen content statistic model and justify the model parameters shown in Table 2. The overall 20 original screen content images contained in the SIQAD database were chosen for testing, since they are extremely different from the 1000 high-quality “webpage” and “screen snap” images downloaded from the “Google Images” website. As illustrated in Fig. 4, we present the scatter plots and the fitted lines using the model parameters listed in Table 2. One can

see that the used statistic model can also be applied to the new 20 screen content images with a small deviation.

3.3. Performance measures and comparisons

We adopt two manners in performance evaluations. The first evaluation is to compare our BQMS model and OA-NR-IQA algorithms (including DIIVINE, BLIINDS-II, BRISQUE and NFERM), coming from the typical training process. Based on the image scene, we randomly divide all the 980 SCIs in the SIQAD database into two groups, one of which covers 784 distorted SCIs associated to 16 original SCIs (i.e. 80% data for training) and the other of which covers 196 distorted SCIs associated to the rest 4 original SCIs (i.e. 20% data for testing). We repeat the aforesaid random 80% train–20% test procedure 1000 times and report the median result across the 1000 iterations, to confirm that the testing method is robust across image scenes and not biased by some specific train-test splits. Table 4 lists the results of our BQMS approach and blind DIIVINE, BLIINDS-II, BRISQUE and NFERM metrics. As expected, the proposed BQMS model always obtains the highest scores across all.

Table 4

Performance indices of our BQMS and two NR-IQA algorithms on the SIQAD database. We highlight the best performed metric with boldface.

80–20%	Type	PLC	SRC	KRC
DIIVINE [26]	NR	0.6912	0.6587	0.4805
BLIINDS-II [27]	NR	0.7255	0.6813	0.5020
BRISQUE [28]	NR	0.7708	0.7237	0.5382
NFERM [32]	NR	0.8007	0.7717	0.5815
BQMS (Proposed)	NR	0.8115	0.8005	0.6056
60–40%	Type	PLC	SRC	KRC
DIIVINE [26]	NR	0.6338	0.6131	0.4440
BLIINDS-II [27]	NR	0.6666	0.6468	0.4719
BRISQUE [28]	NR	0.7525	0.7115	0.5259
NFERM [32]	NR	0.7670	0.7369	0.5458
BQMS (Proposed)	NR	0.8023	0.7892	0.5924
40–60%	Type	PLC	SRC	KRC
DIIVINE [26]	NR	0.5768	0.5644	0.4035
BLIINDS-II [27]	NR	0.6414	0.6243	0.4511
BRISQUE [28]	NR	0.7297	0.6915	0.5058
NFERM [32]	NR	0.7278	0.7027	0.5130
BQMS (Proposed)	NR	0.7904	0.7759	0.5782
20–80%	Type	PLC	SRC	KRC
DIIVINE [26]	NR	0.4910	0.4854	0.3410
BLIINDS-II [27]	NR	0.5471	0.5405	0.3825
BRISQUE [28]	NR	0.6814	0.6495	0.4685
NFERM [32]	NR	0.6667	0.6459	0.4657
BQMS (Proposed)	NR	0.7558	0.7366	0.5406

Table 5

Performance indices of our BQMS and FR- and RR-IQA models on the SIQAD database. We emphasize the top two quality metrics with boldface.

Algorithms	Type	PLC	SRC	KRC	RMS
FSIM [6]	FR	0.5746	0.5652	0.4092	11.612
GSI [7]	FR	0.5515	0.5311	0.3894	11.835
IGM [9]	FR	0.6287	0.6245	0.4594	11.033
VSI [10]	FR	0.5403	0.5199	0.3712	11.938
SIQM [13]	FR	0.8522	0.8455	0.6529	7.4222
FEDM [17]	RR	0.6670	0.5930	0.4097	10.571
QFTB [18]	RR	0.5798	0.5430	0.4165	11.559
SDM [19]	RR	0.5947	0.5891	0.4221	11.406
NIQE [29]	NR	0.3749	0.3568	0.2462	13.152
IL-NIQE [30]	NR	0.1497	0.0994	0.0706	14.027
BQMS (Proposed)	NR	0.7549	0.7223	0.5299	9.3042

Apart from the 80% train–20% test procedure, other train-test procedures with 60–40%, 40–60%, 20–80% are used to validate and compare BQMS with other four NR-IQA models, as reported in Table 4. Likewise, our BQMS algorithm has constantly achieved the optimal results. And moreover, one can easily see that, with the training data decreased and testing data increased, the performance of our model is slightly fluctuated whereas other four blind metrics show a large degree of performance reductions. The above phenomena might be explained by the fact that the NSS which two testing blind IQA metrics rely on does not come into existence

	M_{01}	M_{02}	M_{03}	M_{04}	M_{05}	M_{06}	M_{07}	M_{08}	M_{09}	M_{10}	M_{11}	
M_{01}	0	0	0	0	-1	-1	0	0	0	1	-1	M_{01} : FSIM
M_{02}	0	0	0	0	-1	-1	0	0	0	1	-1	M_{02} : GSI
M_{03}	0	0	0	0	-1	-1	1	0	1	1	-1	M_{03} : IGM
M_{04}	0	0	0	0	-1	-1	0	0	0	1	-1	M_{04} : VSI
M_{05}	1	1	1	1	0	1	1	1	1	1	1	M_{05} : SIQM
M_{06}	1	1	1	1	-1	0	1	1	1	1	-1	M_{06} : FEDM
M_{07}	0	0	-1	0	-1	-1	0	-1	0	0	-1	M_{07} : FTQM
M_{08}	0	0	0	0	-1	-1	1	0	1	1	-1	M_{08} : SDM
M_{09}	0	0	-1	0	-1	-1	0	-1	0	0	-1	M_{09} : NIQE
M_{10}	-1	-1	-1	-1	-1	-1	0	-1	0	0	-1	M_{10} : IL-NIQE
M_{11}	1	1	1	1	-1	1	1	1	1	1	0	M_{11} : BQMS

Fig. 5. Statistical significance comparison of the proposed blind BQMS algorithm and other testing IQA methods using the *F*-test.

in original screen content images, but conversely, our BQMS metric is established on a good screen content statistics model.

The second manner is to measure correlation performance of our BQMS and FR-, RR- and OU-NR-IQA models (including NIQE and IL-NIQE). As mentioned previously, rather than training on only hundreds of images, we utilize a large set of 100,000 distorted SCIs and the corresponding SIQM scores as training samples to learn a module for our (opinion-unaware) BQMS method, which is subsequently used for comparing with FR-, RR- and OU-NR-IQA techniques. From Table 5, we can easily find that the proposed BQMS algorithm, despite the use of SIQM scores (instead of human opinions) as training labels, has obtained very encouraging performance results, noticeably better than OU-NR NIQE and IL-NIQE models. Due to the presence of the (complete or partial) original references for assistance in the FR- and RR-IQA techniques, they are considered hardly matchable with NR-IQA metrics. Despite this, our BQMS model is still superior to most state-of-the-art FR- and RR-IQA algorithms as well as just inferior to the SIQM method which was recently proposed and specific to the IQA of screen content images.

The statistical significance comparison is further conducted by comparing the prediction errors between v_c (converted scores) and v_s (subjective scores) of each IQA model tested. Supposing that the prediction errors of each metric conform to the Gaussian distribution, we compute the *F*-test on the errors of our BQMS technique and other IQA methods. A value of “1” indicates that the metric in the row is significantly superior to the one in the column with a confidence greater than 95%, a value of “0” indicates that the metric in the row is significantly comparable to the one in the

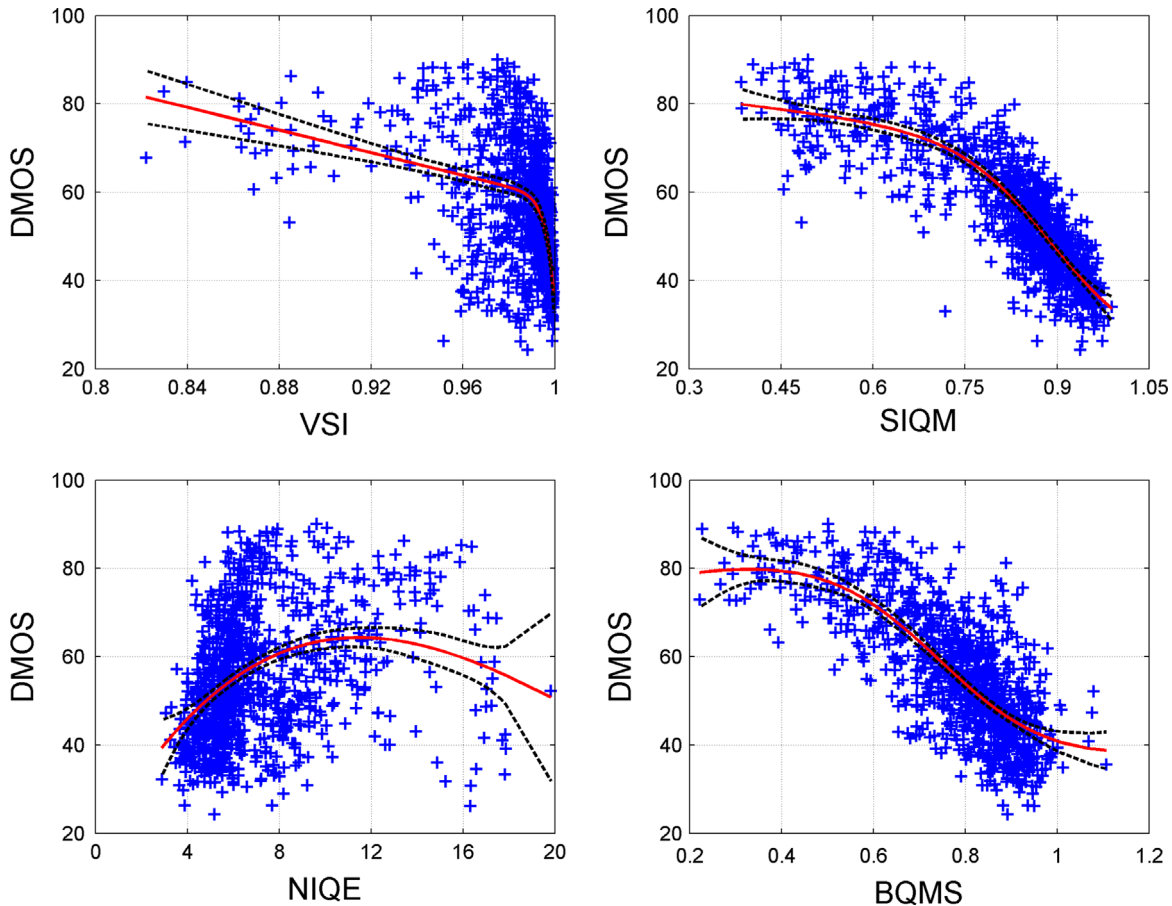


Fig. 6. Scatter plots of DMOS versus lately developed FR VSI, FR SIQM, NR NIQE and our blind BQMS metric on the SIQAD database. The (red) lines are curves fitted with the logistic function and the (black) dash lines are 95% confidence intervals. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

column, and a value of “–1” indicates that the metric in the row is statistically inferior to the one in the column. We list the results of statistical significance in Fig. 5. It can be viewed that the high performance has been obtained by our blind IQA metric, whose results are “1” in most conditions. Specifically, except the SIQM metric dedicated to the IQA of screen content, our blind algorithm is completely statistically better than the FR-, RR- and NR-IQA metrics tested in this work.

At last, we straightforwardly compare the correlation performance in light of visualized scatter plots. As shown in Fig. 6, we display the scatter plots of subjective DMOS values and objective quality predictions of state-of-the-art FR VSI, FR SIQM, NR NIQE and our blind BQMS technique. Obviously, our method has acquired the impressive convergency and monotonicity, much better than recently designed VSI and NIQE metrics while slightly inferior to the FR SIQM metric that is specific to the screen content IQA.

The quality assessment of screen content images is still in a preliminary stage, and there is some room for performance gain of the blind IQA model. The first suggestion about the future work is to improve the proposed screen content statistics model, e.g. to make the statistic model more robust across different contents and thus to reduce the deviations given in Fig. 4. And the second one is to devise higher-accuracy FR-IQA metrics for screen content, with whose scores as training labels to learn a more reliable blind quality method.

4. Conclusion

In this paper we have put forward a new blind quality metric for screen content images. Based on the new screen content statistics model, we have proposed 13 features and proven their effectiveness on the SIQAD database. In order to avoid the overfitting in training, we have used a large set of 100,000 SCLs and the associated SIQM scores to construct the BQMS technique. With respect to state-of-the-art RR- and FR-IQA methods, our BQMS algorithm has acquired better correlation performance in line with subjective quality scores. In the future, two plans will be carried out. One is to improve the screen content statistics model and the other is to employ more reliable training samples by advancing FR quality metrics of screen content images. Our source code will be available soon at <http://www.ntu.edu.sg/home/wslin/Publications.htm> and <https://sites.google.com/site/guke198701/publications>.

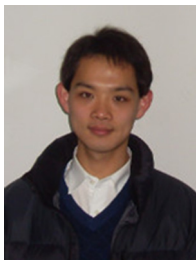
Acknowledgments

This work was supported in part by the National Science Foundation of China under Grants (61025005, 61371146, 61221001, and 61390514), the Foundation for the Author of National Excellent Doctoral Dissertation of PR China under Grant 201339, and the Shanghai Municipal Commission of Economy and Informatization under Grant 140310.

References

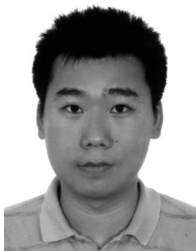
- [1] H. Shen, Y. Lu, F. Wu, S. Li, A high-performance remote computing platform, In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications, March 2009, pp. 1–6.
- [2] C.-Y. Huang, K.-T. Chen, D.-Y. Chen, H.-J. Hsu, C.-H. Hsu, GamingAnywhere: the first open source cloud gaming system, *ACM Trans. Multimed. Comput.* 10 (January (1)) (2014) 173–185.
- [3] S. Wang, K. Gu, S. Ma, W. Lin, X. Zhang, W. Gao, Cloud based image contrast enhancement, In: Proceedings of the IEEE International Conference on Multimedia Big Data, April 2015, pp. 148–155.
- [4] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (April (4)) (2004) 600–612.
- [5] H. Liu, I. Heynderickx, Visual attention in objective image quality assessment: based on eye-tracking data, *IEEE Trans. Circuits Syst. Video Technol.* 21 (April (7)) (2011) 971–982.
- [6] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, *IEEE Trans. Image Process.* 20 (August (8)) (2011) 2378–2386.
- [7] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, *IEEE Trans. Image Process.* 21 (April (4)) (2012) 1500–1512.
- [8] K. Gu, G. Zhai, X. Yang, W. Zhang, An improved full-reference image quality metric based on structure compensation, In: Asia-Pacific Signal and Information Processing Association ASC, December 2012.
- [9] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, *IEEE Trans. Image Process.* 22 (January (1)) (2013) 43–54.
- [10] L. Zhang, Y. Shen, H. Li, VSI: a visual saliency induced index for perceptual image quality assessment, *IEEE Trans. Image Process.* 23 (October (10)) (2014) 4270–4281.
- [11] W. Zhang, A. Borji, Z. Wang, P. Le Callet, H. Liu, The application of visual saliency models in objective image quality assessment: a statistical evaluation, *IEEE Trans. Neural Netw. Learn. Syst.* (2016), in press, <http://dx.doi.org/10.1109/TNNLS.2015.2461603>.
- [12] H. Chang, Q. Zhang, Q. Wu, Y. Gan, Perceptual image quality assessment by independent feature detector, *Neurocomputing* 151 (March (3)) (2015) 1142–1152.
- [13] K. Gu, S. Wang, G. Zhai, S. Ma, W. Lin, Screen image quality assessment incorporating structural degradation measurement, In: Proceedings of the IEEE International Symposium on Circuits and Systems, May 2015, pp. 125–128.
- [14] H. Yang, Y. Fang, W. Lin, Z. Wang, Subjective quality assessment of screen content images, in: Proceedings IEEE International Workshop on Quality of Multimedia Experience, September 2014, pp. 257–262.
- [15] H. Yang, Y. Fang, W. Lin, Perceptual quality assessment of screen content images, *IEEE Trans. Image Process.* 24 (November (11)) (2015) 4408–4421.
- [16] X. Gao, W. Lu, X. Li, D. Tao, Wavelet-based contourlet in quality evaluation of digital images, *Neurocomputing* 72 (December (1)) (2008) 378–385.
- [17] G. Zhai, X. Wu, X. Yang, W. Lin, W. Zhang, A psychovisual quality metric in free-energy principle, *IEEE Trans. Image Process.* 21 (January (1)) (2012) 41–52.
- [18] M. Narwaria, W. Lin, I.V. McLoughlin, S. Emmanuel, L.T. Chia, Fourier transform-based scalable image quality measure, *IEEE Trans. Image Process.* 21 (August (8)) (2012) 3364–3377.
- [19] K. Gu, G. Zhai, X. Yang, W. Zhang, A new reduced-reference image quality assessment using structural degradation model, In: Proceedings of the IEEE International Symposium on Circuits and Systems, May 2013, pp. 1095–1098.
- [20] S.A. Golestaneh, D.M. Chandler, No-reference quality assessment of JPEG images via a quality relevance map, *IEEE Signal Process. Lett.* 21 (February (2)) (2014) 155–158.
- [21] L. Li, W. Lin, H. Zhu, Learning structural regularity for evaluating blocking artifacts in JPEG images, *IEEE Signal Process. Lett.* 21 (August (8)) (2014) 918–922.
- [22] H. Liu, N. Klomp, I. Heynderickx, A no-reference metric for perceived ringing artifacts in images, *IEEE Trans. Circuits Syst. Video Technol.* 20 (April (4)) (2010) 529–539.
- [23] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, *IEEE Trans. Image Process.* 24 (October (10)) (2015) 3218–3231.
- [24] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, *IEEE Trans. Cybern.* 46 (1) (2016) 39–50.
- [25] K. Gu, G. Zhai, X. Yang, W. Zhang, Hybrid no-reference quality metric for singly and multiply distorted images, *IEEE Trans. Broadcast.* 60 (September (3)) (2014) 555–567.
- [26] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from scene statistics to perceptual quality, *IEEE Trans. Image Process.* 20 (December (12)) (2011) 3350–3364.
- [27] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, *IEEE Trans. Image Process.* 21 (August (8)) (2012) 3339–3352.
- [28] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (December (12)) (2012) 4695–4708.
- [29] A. Mittal, R. Soundararajan, A.C. Bovik, Making a ‘completely blind’ image quality analyzer, *IEEE Signal Process. Lett.* 22 (March (3)) (2013) 209–212.
- [30] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, *IEEE Trans. Image Process.* 24 (August (8)) (2015) 2579–2591.
- [31] D.L. Ruderman, The statistics of natural images, *Netw. Comput. Neural Syst.* 5 (4) (1994) 517–548.
- [32] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, *IEEE Trans. Multimed.* 17 (January (1)) (2015) 50–63.
- [33] K. Friston, The free-energy principle: A unified brain theory?, *Nat. Rev. Neurosci.* 11 (February) (2010) 127–138.
- [34] X. Wu, G. Zhai, X. Yang, W. Zhang, Adaptive sequential prediction of multi-dimensional signals with applications to lossless image coding, *IEEE Trans. Image Process.* 20 (January (1)) (2011) 36–42.

- [35] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, Visual saliency detection with free energy theory, *IEEE Signal Process. Lett.* 22 (October (10)) (2015) 1552–1555.
- [36] P. Ye, J. Kumar, D. Doermann, Beyond human opinion scores blind image quality assessment based on synthetic scores, In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 4241–4248.
- [37] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) [Online]. Available: <http://www.csie.ntu.edu.tw/~simjlin/libsvm>.
- [38] T. Kanimozhi, K. Latha, An integrated approach to region based image retrieval using firefly algorithm and support vector machine, *Neurocomputing* 151 (March (3)) (2015) 1099–1111.
- [39] S. Wang, C. Deng, B. Zhao, G.-B. Huang, B. Wang, Gradient-based no-reference image blur assessment using extreme learning machine, *Neurocomputing* (2016), to appear.
- [40] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, March 2000 (<http://www.vqeg.org/>).



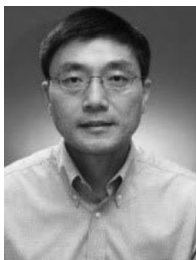
Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is the reviewer for IEEE T-IP, T-CSVT, T-MM, T-CYB, SPL, Neurocomputing, SPIC, JVCI, SIViP, IET-IP, etc. From July to November 2014, he was a visiting student at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. From December 2014 to January 2015, he was a visiting student at the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include quality assessment, visual saliency detection and contrast enhancement. More information can be found in his

homepage <https://sites.google.com/site/guke198701/home>.



Guangtao Zhai received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-

Nuremberg, Germany. He received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012. His research interests include multimedia signal processing and perceptual signal processing.



Weisi Lin received the Ph.D. degree from Kings College, London University, London, U.K., in 1993. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, and served as a Lab Head of Visual Processing, Institute for Infocomm Research. He authors over 300 scholarly publications, holds 7 patents, and receives over 4 million dollars in research grant funding. He has maintained active long-term working relationship with a number of companies. His research interests include image processing, video compression, perceptual visual and audio modeling, computer vision, and multimedia communication. He served as an Associate Editor of IEEE

Transactions on Image Processing, *IEEE Transactions on Multimedia*, *IEEE Signal Processing Letters*, and *Journal of Visual Communication and Image Representation*. He is also on six IEEE Technical Committees and Technical Program Committees of a number of international conferences. He was the Lead Guest Editor for a special issue on perceptual signal processing of the *IEEE Journal Of Selected Topics in Signal Processing* in 2012. He is a Chartered Engineer in the U.K., a fellow of the Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He Co-Chaired the IEEE Multimedia Communications Technical Committee (MMTC) special interest group on quality of experience. He was an Elected Distinguished Lecturer of APSIPA in 2012/3. More information can be found in his homepage <http://www.ntu.edu.sg/home/wslin/index.htm>.



Xiaokang Yang received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2000. He is currently a Full Professor and Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From September 2000 to March 2002, he was a Research Fellow in Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has published over 80 refereed papers, and has filed six patents. His current research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition. He actively participates in the International Standards such as MPEG-4, JVT, and MPEG-21. He received the Microsoft Young Professorship Award 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003), and awards from A-STAR and Tan Kah Kee foundations. He is a member of Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society. He was the Special Session Chair of Perceptual Visual Processing of IEEE ICME2006. He is the local co-chair of ChinaCom2007 and the technical program co-chair of IEEE SiPS2007.



Wenjun Zhang received the B.S., M.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987 and 1989, respectively. From 1990 to 1993, He worked as a post-doctoral fellow at Philips Kommunikation Industrie AG in Nuremberg, Germany, where he was actively involved in developing HD-MAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a full professor in the Department of Electronic Engineering in 1995. As the national HDTV TEEG project leader, he successfully developed the first Chinese HDTV prototype system in 1998. He was one of the main contributors to the Chinese Digital Television

Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He holds more than 40 patents and published more than 90 papers in international journals and conferences. Prof. Zhang's main research interests include digital video coding and transmission, multimedia semantic processing and intelligent video surveillance. He is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television (NERC-DTV), an industry/government consortium in DTV technology research and standardization and the Chair of Future of Broadcast Television Initiative (FOBTV) Technical Committee.