Blind Image Quality Assessment based on High Order Statistics Aggregation

Jingtao Xu, *Student Member, IEEE,* Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann, *Fellow, IEEE*

***Abstract*—Blind image quality assessment (BIQA) research aims to develop a perceptual model to evaluate the quality of distorted images automatically and accurately without access to the non-distorted reference images. State-of-the-art general purpose BIQA methods can be classified into two categories according to the types of features used. The first includes handcrafted features which rely on the statistical regularities of natural images. These, however, are not suitable for images containing text and artificial graphics. The second includes learning based features which invariably require large codebook or supervised codebook updating procedures to obtain satisfactory performance. These are time consuming and not applicable in practice.**

In this paper, we propose a novel general purpose BIQA method based on High Order Statistics Aggregation (HOSA), requiring only a small codebook. HOSA consists of three steps. First, local normalized image patches are extracted as local features through a regular grid and a codebook containing 100 codewords is constructed by K-means clustering. In addition to the mean of each cluster, the diagonal covariance and coskewness (i.e., dimension wise variance and skewness) of clusters are also calculated. Second, each local feature is softly assigned to several nearest clusters and the differences of high order statistics (mean, variance and skewness) between local features and corresponding clusters are softly aggregated to build the global quality aware image representation. Finally, support vector regression (SVR) is adopted to learn the mapping between perceptual features and subjective opinion scores. The proposed method has been extensively evaluated on ten image databases with both simulated and realistic image distortions, and shows highly competitive performance to state-of-the-art BIQA methods.

***Index Terms*—Blind image quality assessment, codebook, high order statistics, local feature aggregation, feature normalization, support vector regression.**

I. Introduction

WITH the rapid surge in visual media content and the development of network technologies, the growing amount of digital image processing applications have brought in the requirement of high quality of experience (QoE) for consumers [1]. Since humans are sensitive to visual signal impairments, e.g., blockiness, blurriness, noisiness and transmission loss, it is crucial to evaluate the influences of various distortions on the perceived image quality through a quantitative approach. Subjective quality evaluation is the most accurate and reliable approach for image quality assessment (IQA), but it is time consuming, expensive, non-reproducible, and unable to be implemented in real world systems. Therefore, automatic objective IQA methods which are consistent with human perception are highly desirable.

In general, objective IQA methods can be categorized into three classes according to the availability of reference image [2]. Full-reference (FR) IQA methods require the non-distorted reference images while reduced-reference (RR) IQA methods only need the quality aware features from reference images. Typical FR and RR methods include SSIM [3], FSIM [4], RRED [5] and GMSD [6]. The third class is blind IQA (BIQA) methods which do not require any information from reference images. Since the information regarding the reference image is not available in practice, among all kinds of IQA methods, BIQA methods are extremely valuable and particularly challenging.

Early BIQA research aims to evaluate images corrupted by specific distortion types, e.g., blockiness [7], ringing effectiveness [8] and blurriness [9]–[11]. Though satisfying results have been obtained by these methods, their universality is limited by the given distortion type in practice. By contrast, general purpose BIQA methods do not require any exact prior knowledge of distortion, therefore they are much more practical and can be applied under various scenarios. However, it is very challenging to explore versatile quality aware feature which is sensitive to diverse distortion types and invariable to different image contents. Fortunately, during the last decade, numerous general purpose BIQA methods have been proposed. Existing general purpose BIQA metrics are classified into two categories according to the types of features used.

The first category is based on the well-chosen handcrafted features which are sensitive to the presence of different distortions, such as natural scene statistics (NSS), image gradients and image entropies. These features are generally represented by the parameters of particular probabilistic models. The commonly used models include generalized Gaussian distribution (GGD) [12], Gaussian distribution [13], Weibull distribution [14] and wrapped Cauchy distribution [15]. Methods belong to this category include [16]–[23]. Moorthy *et al.* [16] propose a two-stage framework for BIQA with one classifier and several regressors for each distortion type. A richer set of NSS features are extracted from steerable pyramid wavelet transform coefficients. However, it requires distortion type information before training, and assumes that the test images contain

J. Xu, H. Du and Y. Liu are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876 China (e-mail: xjt678@gmail.com; duhaiqing@bupt.edu.cn; liuyo@bupt.edu.cn).

P. Ye is with Airbnb, San Francisco, CA 94013 USA (e-mail: pengye@umiacs.umd.edu).

Q. Li is with the School of Computer Engineering, Nanyang Technological University, Singapore (e-mail: qli013@e.ntu.edu.sg).

D. Doermann is with the Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 USA (e-mail: doermann@umd.edu).

TABLE I
SUMMARY OF PREVIOUS GENERAL PURPOSE BIQA METHODS CITED IN THIS PAPER (N: IMAGE PATCH NUMBER)

Category	Method	Feature	Regression	Dim.
Handcrafted feature based	BIQI [24]	Wavelet coefficient statistics	SVM for classification, SVR for regression	18
	DIIVINE [16]	Steerable pyramid wavelet coefficient statistics	SVM for classification, SVR for regression	88
	BLIINDS-2 [17]	DCT coefficient statistics	Probabilistic model or SVR	10
	BRISQUE [18]	Spatial normalized image statistics	SVR	36
	GM-LOG [20]	Joint statistics of image GM and LOG coefficients	SVR	40
	NR-GLBP [21]	Rotation-invariant uniform LBP of image LOG coefficients	SVR	72
	DBN [22]	NSS, image texture, noisiness and blurriness based features	Deep belief network	16689
	ILNIQE [23]	NSS, gradient, log-Gabor filter response and color features	MVG	430×N
Feature learning based	CBIQ [25]	Image patch Gabor filter responses and hard pooling	SVR	10000
	CORNIA [26]	Spatial normalized image patches and max + min pooling	SVR	20000
	SFL [27]	CORNIA feature and codebook optimization	SVR	200
	QAF [28]	Image patch log-Gabor filter responses and max pooling	Random forest	10000
	CNN [29]	Spatial normalized image patches	Convolutional neural network	1024×N

the trained distortion types, which makes it not applicable in practice. In [17], Saad *et al.* train a probabilistic model with statistical features extracted from the DCT domain, but the block processing based algorithm is extraordinarily time consuming. Mittal *et al.* [18] utilize GGD to describe the statistics of local normalized images to constitute an efficient BIQA model. Nevertheless, it may lose accuracy when fitting errors of GGD parameter estimation are large [19]. Xue *et al.* [20] propose GM-LOG which extracts competitive histogram-based features from the joint statistics of image gradients magnitude (GM) and Laplacian of Gaussian (LOG) responses. But extracted features cannot deal with all types of distortions, such as multiply distortions. Zhang *et al.* [21] utilize the rotation invariant uniform local binary pattern (LBP) to encode image LOG responses for BIQA. But it is still affected by image contents. In [22], Tang *et al.* propose a model to extract three types of features from the statistics of natural images, distortion textures, and blurriness/noisiness. Then it puts all the features into a deep belief network (DBN) to learn the model. Recently Zhang *et al.* [23] extract abundant perceptual patch level features from pristine images, and measured the distance between distorted image and pristine images via multivariate Gaussian (MVG) model. All these image quality aware features are handcrafted.

To avoid the difficulties coming with handcrafted features, feature learning based BIQA methods try to automatically learn the quality aware features from images. The representative codebook-based framework (a.k.a bag of words model) [30] and several extensions [31], [32] which have gained popularity in image classification and retrieval tasks, have also been employed for BIQA. Ye *et al.* [25] propose a codebook-based framework, CBIQ, to calculate image quality with corresponding codewords occurrence histogram, but the codebook size is extremely large, nearly 120K. Later they extend CBIQ with an unsupervised feature learning method, CORNIA [26]. CORNIA calculates the dot products between normalized image patches and codewords to reveal the image quality. Followed by max and min pooling, the final quality aware feature is generated. With a 10K codeword codebook, it achieves state-of-the-art BIQA performance. But when the codebook size decreases to hundreds, the performance drops significantly. In [27], the authors propose a supervised filter learning (SFL) approach with stochastic gradient descent (SGD) to optimize a 100 codeword codebook and quality evaluation alternatively. The performance is acceptable but still inferior to CORNIA and the supervised codebook updating procedure in the training stage is time consuming. The quality aware filter (QAF) model [28] extracts image log-Gabor responses to formulate codebook using sparse filter learning. Then a random forest is employed to learn the mapping between features and subjective scores. However, higher performance is obtained with a large codebook and complex codebook learning procedure. During each train-test loop, the 10K codeword codebook needs to be reconstructed. With normalized image patches, Kang *et al.* [29] propose a shallow convolutional neural network (CNN) based BIQA method [1] which could even predict local image quality. A summary of BIQA methods cited in this paper can be found in Table I [2].

Previous feature learning based BIQA methods [26]–[28] only incorporate image zero order statistics (word counting) which is insufficient for BIQA. When codebook size decreases, the performance drops significantly, e.g., CORNIA with a 100 codeword codebook has a much worse performance on the LIVE database compared to original CORNIA (0.8572 vs 0.9417). In addition, since most handcrafted features are generally derived from natural scene images, they cannot predict the perceptual quality of all kinds of images, such as screen content images which contain text, graphics and natural pictures simultaneously and document images. In order to solve the aforementioned problems from the two categories of BIQA methods, we propose a novel BIQA framework based on image High Order Statistics Aggregation (HOSA). This is an extension of our previous work [33].

Our high order statistics aggregation strategy investigates to utilize the statistical differences between codebook and images to build an efficient and robust BIQA model to overcome these difficulties. High order statistics have been applied to many computer vision applications, including, for example, image classification [34], image retrieval [35], [36] and image

[1]Actually the CNN based method requires supervised codebook updating. The convolution kernels are codewords, and the convolution procedure is as same as dot product computation in CORNIA, SFL and QAF. The main difference between it and SFL is that it optimizes codebook construction and quality evaluation simultaneously with large amount of neural network parameters in the training stage.

[2]This is not a complete list of existing general purpose BIQA methods.

aesthetic evaluation [37]. The success of these high order statistics based features in the above applications motivated us to introduce them into the task of BIQA.

The main advantages of HOSA are three folds. The first advantage is high order statistics application. Different types of distortions not only change the low order statistics of local patch coefficient distribution, but also its high order statistics. In addition to the mean of each cluster [30], the dimension wise variance and skew of clusters are also calculated to form a detailed quality aware codebook to approximate the distribution of low-level features. Then the soft weighted high order statistics differences between local features and corresponding clusters are calculated. With such a small codebook, it could describe the relationship between codebook and the features from one image more comprehensively and therefore improve the quality prediction performance. The second one is the high generalization ability. The proposed method can be applied in a wide spectrum of image types, including natural images, screen content images and document images. And it can also well reflect the impact of both simulated and realistic distortions on perceptual quality. The third one is fast computation. Since much smaller codebook used, the quality aware representation computation has a faster speed compared to previous feature learning based BIQA methods and has the potential to be applied into real world applications.

The rest of this paper is organized as follows. Section II describes the details of proposed method HOSA. Extensive experiment results on ten databases and discussions are presented in Section III. Section IV concludes the paper.

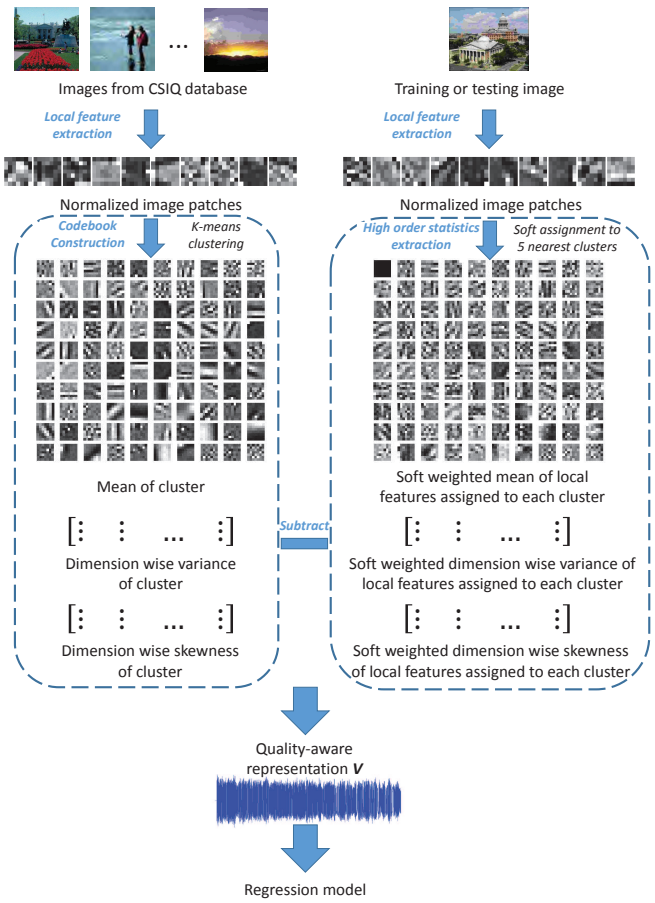


Fig. 1. Pipeline of the proposed method HOSA.

II. Methodology

In this section, we show how the high order statistics can be adapted to the BIQA task. Fig. 1 illustrates the pipeline of proposed HOSA model. The key components in our proposed BIQA framework are the extraction of local features, the construction of comprehensive codebook, high order statistics aggregation, and regression.

A. Local Feature Extraction

In this paper, normalized raw image patches are extracted from images as local features. Given a grayscale image, the local feature $\boldsymbol{x}(i,j)$ is extracted from $B\times B$ patch $\boldsymbol{I}(i,j)$, where (i,j) is the indices sampled on a regular grid over the entire image. The contrast normalization scheme which has been widely used in BIQA domain [18], [26] is applied to each patch as follows:

$$\boldsymbol{x}(i,j)=\frac{\boldsymbol{I}(i,j)-\mu}{\sigma+10}, \tag{1}$$

where μ and σ are the local mean and standard deviation of patch $\boldsymbol{I}(i,j)$, and the constant 10 prevents instability when the denominator approaches zero. For each image, N normalized patches are extracted: $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N]\in\boldsymbol{R}^D$ $(D=B\times B)$, where each column corresponds to one local patch. This normalization can be also regarded as a simpler version of divisive normalization transform (DNT) [38] to mimic the early nonlinear processing in human visual system (HVS) and reduce redundancies in local features. In addition, a zero-phase component analysis (ZCA) whitening process [39] is applied to further remove the linear correlations between local features.

B. Codebook Construction

The essence of codebook framework is based on the idea of partitioning the local feature space into informative regions whose internal structure can be parameterized linearly. The resulted regions are generally called visual codewords and a set of visual codewords form a visual codebook [40]. In this paper, we followed the codebook construction protocol in [26]. In particular, we construct a 100 codeword codebook using K-means clustering [41] from the images in CSIQ database [42].

Given a set $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N]\in\boldsymbol{R}^D$ of N local features, K-means seeks K centers $[\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_K]\in\boldsymbol{R}^D$ and the data hard assignments $[q_1,q_2,\ldots,q_N]\in\{1,2,\ldots,K\}$, such that the cumulative approximation error $\sum_{i=1}^{N}\|\boldsymbol{x}_i-\boldsymbol{\mu}_{q_i}\|^2$ could be minimized. Traditional K-means clustering only provides the mean $(\boldsymbol{\mu})$ of each cluster which can only be utilized to extract zero and first order statistics. In order to extract high order statistics information for quality assessment, we further calculate the covariance $(\boldsymbol{\sigma}^2)$ and coskewness $(\boldsymbol{\gamma})$ for each cluster. These matrices are assumed diagonal as the computational cost of diagonal matrices is much lower than

the cost involved for full matrices. Therefore, the dimension wise variance and skew are provided. For the kth cluster, two D-dimensional vectors, $\boldsymbol{\sigma}_k^2$ and $\boldsymbol{\gamma}_k$, are calculated. Finally the generated codebook is described by a set of parameters $\lambda = \{\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2, \boldsymbol{\gamma}_k, k = 1, 2, \ldots, K\}$. Each cluster represents one codeword. In this paper, we adopt the Pelleg's K-means algorithm [43] which could accelerate the iterative process by utilizing K-D trees. The $\boldsymbol{\mu}$ of 100 codewords from K-means clustering is also shown in the left of Fig. 1.

We observe that the constructed 100 codewords (mean) contain patterns which capture various distortion characteristics, e.g., codewords with horizontal and vertical lines represent blockiness, some with "dense points" account for noise, and some with flat patterns illustrate blurriness. Beyond that, it is interesting that there are some specific codewords including spots, lines, and different angles of edges which may measure image primitive structural information. With these representative quality aware codewords, we determine whether the statistical differences between local features and codebook could represent the image perceptual quality.

C. High Order Statistics Aggregation

Previous feature learning based BIQA methods [26]–[28] generally consider zero order statistics from images which is insufficient for BIQA. Either soft assignment or hard assignment is used in former works, only codeword counting information is extracted to represent the relationship between local features from one image to codebook. By contrast, we aggregate high order statistics differences between local features and clusters to build the global quality aware representation with a much smaller codebook. This image statistics aggregation scheme describes the approximate location of image local features in each cluster (relatively to the mean, variance and skewness). Different clusters represent diverse distortion patterns and levels, therefore this relative relationship will vary with image quality level.

In HOSA, we modify our former BIQA algorithm LFA [33] by adding the soft weighted differences of high order statistics (i.e., variance and skewness). For each single local feature $\boldsymbol{x}_i$, r nearest codewords $rNN(\boldsymbol{x}_i)$ are selected by Euclidean distance. This "soft" assignment attempts to solve the problem of codeword uncertainty and plausibility. Codeword uncertainty is the problem of selecting the correct codeword out of two or more similar candidates, while codeword plausibility refers to the problem of selecting a codeword without a suitable candidate in the entire codebook. Both can be alleviated by soft assignment with kernel similarity weights without introducing large quantization error.

First we calculate the residuals between the soft weighted mean of local features assigned to cluster k and the mean of cluster k:

$$\begin{cases} \boldsymbol{m}_k^d = \hat{\boldsymbol{\mu}}_k^d - \boldsymbol{\mu}_k^d = \sum\limits_{i:k\in rNN(\boldsymbol{x_i})} \left[\omega_{ik}\boldsymbol{x}_i^d\right] - \boldsymbol{\mu}_k^d \\ \omega_{ik} = \dfrac{e^{-\beta\|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2}}{\sum_{j:k\in rNN(\boldsymbol{x_j})} e^{-\beta\|\boldsymbol{x}_j - \boldsymbol{\mu}_k\|^2}} \end{cases} \tag{2}$$

where $\hat{\boldsymbol{\mu}}_k$ is the mean of the local features assigned to codeword k, the superscript d denotes the d-th dimension of a vector and ω_{ik} is the Gaussian kernel similarity weight between local feature $\boldsymbol{x}_i$ and codeword k. The sum of the weights for each codeword is 1.

However, there are some unexpected situations in reality, e.g., two sets of local features have the same mean but different variance will generate exactly the same $\boldsymbol{m}_k$ which cannot distinguish the image quality; the mean of assigned features is equal to the codeword will produce zero $\boldsymbol{m}_k$. In order to resolve these troubles, we propose to extract soft weighted high order statistics differences between local features and codewords to further discriminate different quality-level images.

Similarly, we formulate the second order statistics as follows:

$$\boldsymbol{v}_k^d = \hat{\boldsymbol{\sigma}_k^2}^d - \boldsymbol{\sigma}_k^{2d} = \sum_{i:k\in rNN(\boldsymbol{x_i})} \left[\omega_{ik}(\boldsymbol{x}_i^d - \hat{\boldsymbol{\mu}}_k^d)^2\right] - \boldsymbol{\sigma}_k^{2d}, \tag{3}$$

where $\hat{\boldsymbol{\sigma}_k^2}$ is the dimension wise variance of local features assigned to codeword k. Therefore $\hat{\boldsymbol{\sigma}_k^2}^d$ is the variance of dth dimension in cluster k.

As for standard Gaussian distribution, the first and second order statistics is sufficient to describe the distribution. However, low-level image features are not usually Gaussian distribution in practice [44]. Therefore, we also employ third order statistics to exploit complementary information for quality evaluation. The third order statistics can be written as follows:

$$\boldsymbol{s}_k^d = \hat{\boldsymbol{\gamma}_k}^d - \boldsymbol{\gamma}_k^d = \sum_{i:k\in rNN(\boldsymbol{x_i})} \left[\frac{\omega_{ik}(\boldsymbol{x}_i^d - \hat{\boldsymbol{\mu}}_k^d)^3}{(\hat{\boldsymbol{\sigma}_k^2}^d)^{\frac{3}{2}}}\right] - \boldsymbol{\gamma}_k^d, \tag{4}$$

where $\hat{\boldsymbol{\gamma}}_k$ is the dimension wise skewness of local features assigned to codeword k. And $\hat{\boldsymbol{\gamma}_k}^d$ is the skew of dth dimension for cluster k. Then all the three types of statistics differences are concatenated to a single long quality aware feature $\boldsymbol{V} = [\boldsymbol{m}_k^\top, \boldsymbol{v}_k^\top, \boldsymbol{s}_k^\top], k = 1, 2, \ldots, K$.

The new image quality aware representation contains the first, second and third order statistics (c.f. equations (2)-(4)). With a given codebook of size K, full quality aware representation provides a vector of dimensionality $3DK$. Removing image content influence is a key factor for BIQA [16], [20]–image patch normalization and ZCA can remove some amount of its influence. This representation still suffers the problem of frequent local features (e.g., resulting from a repeated structure in the image such as woodiness desktop, grass and brick wall) which can severely decrease the contribution of other important dimensions and hurt the overall feature effectiveness. In other words, the constructed codewords not only represent image quality information but also image contents. Images with similar contents which have frequent local features will increase image feature similarity, even through they may have different image quality scores. Another situation is that the images have similar quality scores, but the distortion introduces frequency component at different dimensions, such as JPEG distortion produces different types

TABLE II
TOP FIVE SROCC BETWEEN EACH DIMENSION OF AGGREGATED FEATURE AND SUBJECTIVE SCORES IN LIVE, TID2013 AND CSIQ DATABASES

Database	Dist.	First order feature	Second order feature	Third order feature
LIVE	JP2K	**0.8625** 0.8592 0.8575 0.8525 0.8461	**0.8672 0.8658 0.8631 0.8617** 0.8589	0.8182 0.8124 0.8106 0.8057 0.8008
	JPEG	0.8748 0.8617 0.8505 0.8503 0.8409	**0.9017 0.8971 0.8918 0.8915 0.8914**	0.8645 0.8622 0.8612 0.8538 0.8530
	WN	0.9355 0.9186 0.9053 0.9046 0.9020	**0.9653 0.9643 0.9636 0.9621 0.9606**	0.9507 0.9496 0.9463 0.9427 0.9416
	GB	**0.9550 0.9545 0.9543** 0.9540 0.9539	0.9507 0.9503 0.9503 0.9501 0.9500	**0.9579 0.9564** 0.9523 0.9517 0.9513
	FF	0.7798 0.7710 0.7696 0.7675 0.7653	**0.8067 0.7997 0.7988 0.7985** 0.7975	**0.8183** 0.7878 0.7789 0.7767 0.7688
	All	**0.6543 0.6371 0.6343 0.6207** 0.6137	0.6179 0.5712 0.5612 0.5598 0.5591	**0.6432** 0.6052 0.5864 0.5797 0.5771
TID2013	WN	0.8228 0.7941 0.7904 0.7868 0.7854	**0.8898 0.8823 0.8803 0.8793 0.8781**	0.8588 0.8190 0.7787 0.7755 0.7561
	GB	**0.9249 0.9125 0.9105 0.9104 0.9060**	0.8867 0.8637 0.8589 0.8585 0.8572	0.8951 0.8936 0.8932 0.8930 0.8903
	JPEG	**0.9223 0.9168 0.9147 0.9123 0.9103**	0.8986 0.8977 0.8977 0.8961 0.8926	0.8596 0.8524 0.8481 0.8368 0.8346
	JP2K	**0.9090 0.9081** 0.9068 0.9062 0.9058	**0.9080 0.9072** 0.9066 0.9059 0.9053	**0.9073** 0.9069 0.9054 0.8994 0.8981
	All	**0.7294** 0.7190 0.7184 0.7058 0.7012	**0.7235** 0.6848 0.6747 0.6647 0.6517	**0.7524 0.7288 0.7254** 0.7204 0.7186
CSIQ	WN	0.8867 0.8851 0.8845 0.8826 0.8817	**0.9211 0.9203 0.9200 0.9194 0.9189**	0.8570 0.8391 0.8347 0.8183 0.8131
	JPEG	**0.8659** 0.8504 0.8481 0.8474 0.8394	0.8438 0.8404 0.8362 0.8324 0.8291	**0.8729 0.8713 0.8677 0.8645** 0.8619
	JP2K	**0.8604 0.8598 0.8539 0.8539** 0.8537	0.8397 0.8364 0.8354 0.8350 0.8335	**0.8699** 0.8501 0.8420 0.8403 0.8378
	GB	0.8604 0.8598 0.8539 0.8539 0.8537	**0.9303 0.9302 0.9300 0.9299 0.9297**	0.9282 0.9273 0.9221 0.9203 0.9200
	All	**0.7218 0.7153 0.6697** 0.6585 0.6585	**0.6927** 0.6691 0.6680 0.6593 0.6572	**0.7284** 0.6652 0.6548 0.6410 0.6397

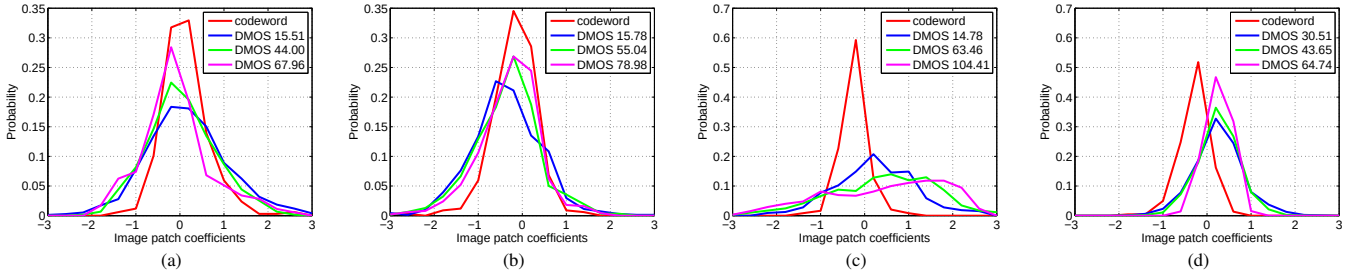


Fig. 2. Examples of image patch coefficient distribution. For each subfigure, the red line is drawn from one dimension of one random selected codeword. The blue, green and pink line represent the distributions drawn from patches assigned to this codeword in different images. Images contain different types and levels of distortions: (a) JP2K distortion, (b) JPEG distortion, (c) WN distortion, (d) GB distortion.

of blocks, and decreases the feature similarity. These kinds of similarity corruption are not expected for quality evaluation.

However, this problem can be alleviated by an element wise signed power normalization [45] on the aggregated feature:

$$f(v) = sign(v)|v|^{\alpha}, \tag{5}$$

where α is the parameter to control the inhibition degree on the frequent components and v is one of the feature value. When α is equal to zero, $\boldsymbol{V}$ converges to a ternary representation. When α is one, there will be no punishment on the frequent generic features. The final feature is also L2-normalized before regression computation.

D. Feature Analysis

To illustrate the effectiveness of the proposed local aggregated feature and the role of high order statistics, we compute the Spearman's rank order correlation coefficient (SROCC) between each dimension of $\boldsymbol{V}$ and subjective scores from the LIVE, TID2013 and CSIQ databases. The top five correlation coefficients for each type of feature are presented in Table II and the top five coefficients for the whole feature are bolded. It shows that after aggregation, the feature generated from local image patches has high correlation with subjective image quality scores. And a number of top five correlation coefficients come from the second and third order feature which demonstrates the contribution of high order statistics.

We further present a few examples for image patch empirical coefficient distributions with different types and levels of distortions in Fig. 2. Discriminating images at different quality levels alter the distribution in various ways, not only the first order statistics are different, but also high order statistics. This could explain why high order features like variance and skewness are helpful for image quality assessment.

In Fig. 3, we provide some concrete image examples to compare features of CORNIA and HOSA. Six pairs of images with different types and levels of distortions are shown. From the first pair to the last pair, the DMOS differences between two images increase gradually. We compared the linear correlation coefficient of features between two images in each pair of images. It is expected that the feature correlation decrease when the two images have larger quality difference. As we can observe from Fig. 4, both features generated by CORNIA-10K and CORNIA-100 [3] have high linear correlations, even though the two images have very different DMOSs. For the regression computation, these images with high correlated features will receive similar quality prediction scores while the subjective scores are quite different. Conversely HOSA extracts distinguishable features for images with different quality scores. And the correlation coefficient decreases gradually when the quality difference becomes larger.

E. Regression Model

Given a database of distorted images and associated subjective quality scores, the extracted features are used to train

[3]Here CORNIA-10K represents original CORNIA [26] while CORNIA-100 represents CORNIA computed on the 100 codeword codebook.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2585880, IEEE Transactions on Image Processing

Fig. 3. Image samples from LIVE database: (a) "sailing4" with JP2K distortion (left) and "house" with FF distortion (right), (b) "woman" with FF distortion (left) and "bikes" with JP2K distortion (right), (c) "lighthouse" with JP2K distortion (left) and "buildings" with JPEG distortion (right), (d) "student sculpture" with GB distortion (left) and "cemetery" with FF distortion (right), (e) "sailing2" with GB distortion (left) and "church and capitol" with JP2K distortion (right), (f) "dancers" with GB distortion (left) and "monarch" with JPEG distortion (right). From image pairs (a) to (f), the absolute DMOS difference between two images increase gradually.

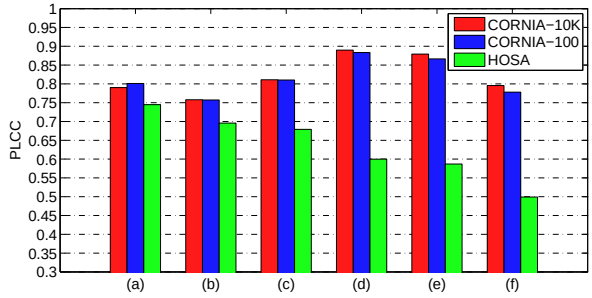


Fig. 4. Linear correlation coefficients of features from different pairs of images.

a linear kernel support vector regressor (SVR) to conduct image perceptual quality prediction. We choose LIBLINEAR package [46] to implement the SVR in this work.

III. Experimental Results and Discussions

A. Experiments on Images with simulated distortions

1) Experimental Protocol: First we conduct experiments to evaluate the competing methods on six image databases with simulated distortions. Except for the commonly-used LIVE database [47], TID2013 database [48], CSIQ database [42], Toyama database [49] and LIVE multiply distorted database (MLIVE) [50], we also tested on one new screen image quality assessment database (SIQAD) [51], to show the generalization ability of the proposed method.

The LIVE database is generated from 29 different reference images. The distortions included in this database are: JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blur (GB) and simulated fast fading Rayleigh channel (FF). It contains 779 distorted images in total. The TID2013 database consists of 25 reference images and generates 3000 distorted images with 24

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2585880, IEEE Transactions on Image Processing

TABLE III
DATABASE INFORMATION

Database	Ref. Images	Dist. Type Num.	Dist. Images	Score Type	Score Range
LIVE	29	5	779	DMOS	[0, 100]
TID2013	25	4	480	MOS	[0, 9]
CSIQ	30	4	600	DMOS	[0, 1]
Toyama	14	2	168	MOS	[0, 5]
MLIVE	15	2	450	DMOS	[0, 100]
SIQAD	20	4	560	DMOS	[0, 100]

various types of distortions. The CSIQ database is a medium size database which consists of 866 distorted images with 6 distortion types. The Toyama database is a smaller database which contains two types of distortions: JPEG and JP2K. The database consists of 168 distorted images from 14 reference images. The MLIVE database includes images distorted by two multiple types of distortion. One is associated with images corrupted by GB followed by JPEG (GB+JPEG), and one is associated with images corrupted by GB followed by WN (GB+WN). The newly proposed SIQAD database is composed of 980 screen content images created by corrupting 20 source images with 7 distortion types. The images in this database include text, graphics and pictures together which is a novel challenge to traditional BIQA methods. In our experiments, for the TID2013, CSIQ and SIQAD databases, we mainly consider four types of distortions which appear in LIVE database, i.e., JP2K, JPEG, WN and GB. The summarized information of six databases are shown in Table III. The subjective scores range and type are different for six databases. In order to unify the subjective scores, human opinion scores from different databases are all mapped to the range of [0,100] as in the LIVE database. It should be also mentioned that both TID2013 and MLIVE contain various types of distortions including several multiple and localized distortions.

Since HOSA belongs to the category of feature learning based BIQA, we compare it with CORNIA-10K [26] and CORNIA-100. Furthermore, we also compare HOSA to other state-of-the-art handcrafted feature based BIQA methods, i.e., BIQI [24], DIIVINE [16], BLIINDS-2 [17], BRISQUE [18], GM-LOG [20] and ILNIQE [23] which is a leading opinion free BIQA method with NSS based features. The source codes of these competing methods are all publicly available.

Three commonly used criteria, SROCC which measures the prediction monotonicity, Pearson's linear correlation coefficient (PLCC) and root mean squared error (RMSE), which measure the prediction accuracy, are employed to evaluate the performance of competing BIQA algorithms. A good BIQA model should have larger value of SROCC and PLCC, and a smaller value of RMSE. According to the report from Video Quality Expert Group (VQEG) [52], the relationship between the subjective scores and the predicted scores may not be linear due to the nonlinear quality rating of human observers. Therefore before calculating PLCC and RMSE, a nonlinear logistic regression processing [47] is applied to map the algorithm scores to subjective opinion scores:

$$f(x) = a_1\left(\frac{1}{2} - \frac{1}{1+e^{a_2(x-a_3)}}\right) + a_4 x + a_5, \qquad (6)$$

where a_1, a_2, a_3, a_4 and a_5 are parameters determined by the nonlinear regression procedure.

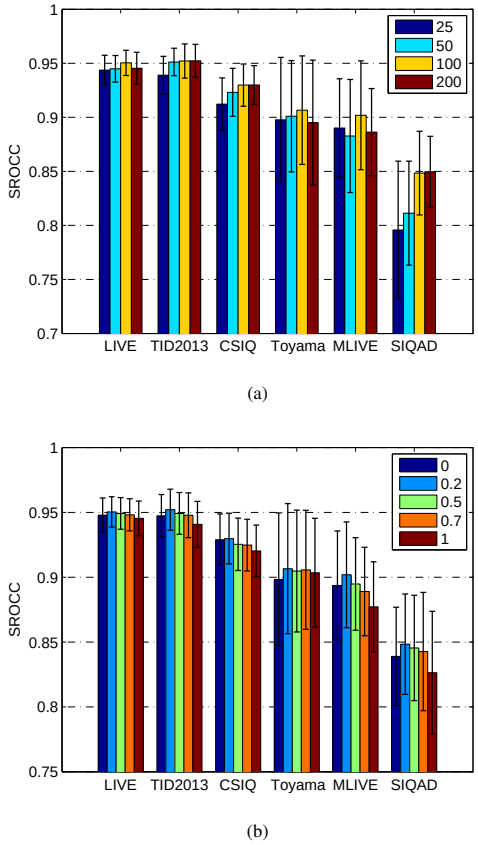


Fig. 5. Performance (SROCC) with different parameters on six databases: (a) different codebook size K, (b) different power normalization parameter α.

2) Implementation Details: There are several parameters in HOSA need to be decided. We tuned the parameters with a five-fold cross validation on the training set from LIVE database. The image patch size B is set to 7. The nearest codewords number r for each local feature is 5 and Gaussian weight function's parameter β is 0.05. In our experiments, we found that the optimal SVR parameters do not vary much with different databases. Thus the cost parameter C is set to 128 and the parameter ε is 0.5 for most experiments.

To investigate the influence of codebook size K and signed power normalization parameter α on the quality estimation performance, we assign $K = \{25, 50, 100, 200\}$, $\alpha = \{0, 0.2, 0.5, 0.7, 1\}$ and compute the SROCC of HOSA on six databases. The results are shown in Fig. 5. We can observe that for $K = 100$ and $\alpha = 0.2$, higher and stable performance is obtained for all databases. Although 200 codeword codebook could provide more promising results than 100 codeword codebook on few databases, considering the overall performance and feature complexity (larger codebook with higher feature dimensionality), the 100 codeword codebook is sufficient for current BIQA tasks. Thus the feature dimensionality for HOSA is 14.7K in total which is also acceptable compared to CORNIA's 20K feature dimensionality.

TABLE IV
OVERALL PERFORMANCE OF THE COMPETING BIQA METHODS ON SIX IMAGE DATABASES WITH SIMULATED DISTORTIONS.

Database		BIQI	DIIVINE	BLIINDS-2	BRISQUE	GM-LOG	ILNIQE	CORNIA-10K	CORNIA-100	HOSA
LIVE 779 images	SROCC	0.8642	0.9162	0.9302	0.9409	**0.9503**	0.9020	0.9417	0.8572	**0.9504**
	PLCC	0.8722	0.9172	0.9357	0.9450	**0.9539**	0.9085	0.9434	0.8579	**0.9527**
	RMSE	13.2852	10.8103	9.6189	8.9048	**8.1723**	11.4007	9.0204	14.0175	**8.2858**
TID2013 480 images	SROCC	0.8191	0.8753	0.8786	0.8917	**0.9282**	0.8871	0.8998	0.8276	**0.9521**
	PLCC	0.8407	0.8859	0.9053	0.9176	**0.9439**	0.9030	0.9277	0.8554	**0.9592**
	RMSE	0.7569	0.6474	0.5921	0.5534	**0.4629**	0.6020	0.5239	0.7280	**0.3941**
CSIQ 600 images	SROCC	0.8115	0.8760	0.9140	0.9099	**0.9228**	0.8885	0.8930	0.8216	**0.9298**
	PLCC	0.8476	0.8983	0.9323	0.9278	**0.9408**	0.9173	0.9175	0.8476	**0.9480**
	RMSE	0.1491	0.1220	0.1010	0.1044	**0.0950**	0.1098	0.1123	0.1492	**0.0887**
Toyama 168 images	SROCC	0.5949	0.8198	0.7995	0.8500	0.8551	0.7772	**0.8565**	0.7160	**0.9066**
	PLCC	0.5948	0.7915	0.7672	0.8269	0.8371	0.7798	**0.8434**	0.7074	**0.8999**
	RMSE	1.0039	0.7720	0.7914	0.7099	0.6897	0.7759	**0.6781**	0.8906	**0.5409**
MLIVE 450 images	SROCC	0.8745	0.8738	0.8872	0.8972	0.8237	**0.9019**	0.9007	0.8403	**0.9019**
	PLCC	0.9008	0.8936	0.9028	**0.9207**	0.8632	0.9142	0.9150	0.8715	**0.9262**
	RMSE	7.9660	8.3843	8.1330	**7.3168**	9.4198	7.6327	7.6737	9.2272	**6.9739**
SIQAD 560 images	SROCC	0.6533	0.7279	0.7561	0.7715	0.7989	0.5429	**0.8352**	0.7321	**0.8484**
	PLCC	0.7304	0.7768	0.7982	0.8210	0.8330	0.5870	**0.8533**	0.7848	**0.8636**
	RMSE	9.4427	8.6903	8.3688	7.9383	7.7005	11.2156	**7.1989**	8.6183	**6.9594**
Weighted average	SROCC	0.7944	0.8555	0.8731	0.8843	0.8894	0.8238	**0.8950**	0.8121	**0.9182**
	PLCC	0.8251	0.8722	0.8907	0.9043	0.9075	0.8438	**0.9094**	0.8357	**0.9295**

TABLE V
RESULTS OF WILCOXON RANK-SUM TEST PERFORMED BETWEEN SROCCS OBTAINED BY COMPETING BIQA METHODS ON SIX DATABASES. 1 (-1) INDICATES HOSA IS STATISTICALLY SUPERIOR (INFERIOR) THAN THE ALGORITHM IN THE COLUMN. 0 INDICATES HOSA IS STATISTICALLY EQUIVALENT TO THE ALGORITHM IN THE COLUMN.

Database	BIQI	DIIVINE	BLIINDS-2	BRISQUE	GM-LOG	ILNIQE	CORNIA-10K	CORNIA-100
LIVE	1	1	1	1	0	1	1	1
TID2013	1	1	1	1	1	1	1	1
CSIQ	1	1	1	1	1	1	1	1
Toyama	1	1	1	1	1	1	1	1
MLIVE	1	1	1	1	1	0	0	1
SIQAD	1	1	1	1	1	1	1	1

TABLE VI
PERFORMANCE EVALUATION (SROCC) OF EACH TYPE OF FEATURE IN HOSA.

Feature type	LIVE	TID2013	CSIQ	Toyama	MLIVE	SIQAD
mean	0.9394	0.9294	0.9082	0.9254	0.8823	0.8137
variance	0.9369	0.9318	0.9266	0.8877	0.8259	0.8340
skewness	0.9368	0.9391	0.8929	0.8853	0.8438	0.8159
m. + v.	0.9480	0.9411	0.9277	0.9053	0.8915	0.8343
m. + v. + s.	0.9504	0.9521	0.9298	0.9066	0.9019	0.8484

3) Performance on Individual Databases: First we evaluated all methods on six databases separately. According to the reference images, each database was partitioned into two non-content overlapping subsets 1000 times. In our experiments, 80% of images were used for training and the remaining for testing. The median value of SROCC, PLCC and RMSE over 1000 train-test loops are reported in Table IV. The top two BIQA methods for each criterion are highlighted in boldface.

Table IV leads us to the following conclusions. First, HOSA always lies in the best two positions for all databases. This demonstrates that the proposed feature aggregation scheme has better ability to represent image quality with various image contents and distortion types. Second, with the same size codebook, HOSA performs continuously better than CORNIA-100 over six databases. More surprisingly, HOSA is also superior to CORNIA-10K which extract the quality aware feature from a much larger 10K codeword codebook. We believe this phenomenon is mainly attributed to redundant information (similar codewords) in large codebook which could interfere with the quality evaluation. Finally, comparing to state-of-the-art handcrafted feature based BIQA methods, e.g., GM-LOG and ILNIQE, HOSA shows quite competitive performance. On TID2013, Toyama and SIQAD, HOSA performs much better than others. It is interesting that the traditional handcrafted features cannot represent the screen content images which contain not only natural scenes but also text and graphics. However, HOSA liked feature learning based methods could catch the specific information varied with distortion degree for text and graphics.

In the bottom two rows of Table IV, we also present the weighted average SROCC and PLCC of competing methods (the weights are based on the numbers of images tested in six databases). Note that the weighted average of RMSE cannot be computed since the subjective scores scale differently in the six databases. It can be seen that HOSA still performs the best among all competing BIQA methods, followed by CORNIA-10K and GM-LOG.

To further prove the superiority of HOSA over the competing BIQA methods, we calculated the statistical significance by the Wilcoxon rank-sum test which measures the equivalence of the median values of two independent samples. It was performed at a significance level of 5% using the 1000 SROCC values of all pairs of BIQA methods. The results are listed in Table V. We can see that on TID2013, CSIQ, Toyama and SIQAD, HOSA performs statistically better than all other BIQA methods. HOSA is on par with CORNIA and ILNIQE on MLIVE, and GM-LOG on LIVE, respectively. Generally, HOSA is more statistically significant.

TABLE VII
PERFORMANCE EVALUATION (SROCC) ON INDIVIDUAL DISTORTIONS.

Database	Dist.	BIQI	DIIVINE	BLIINDS-2	BRISQUE	GM-LOG	ILNIQE	CORNIA-10K	CORNIA-100	HOSA
LIVE	JP2K	0.8464	0.9164	0.9301	0.9169	**0.9262**	0.9017	0.9211	0.8211	**0.9331**
	JPEG	0.8936	0.9028	0.9505	**0.9650**	**0.9631**	0.9312	0.9382	0.8370	0.9549
	WN	0.9640	**0.9813**	0.9471	0.9800	**0.9831**	0.9746	0.9568	0.9241	0.9729
	GB	0.8696	0.9299	0.9146	0.9519	0.9293	0.9111	**0.9573**	0.9217	**0.9524**
	FF	0.7606	0.8627	0.8741	0.8754	0.8994	0.8270	**0.9057**	0.8647	**0.9003**
TID2013	WN	0.8746	0.8845	0.8315	0.8568	**0.9385**	0.9008	0.7546	0.6741	**0.9215**
	GB	0.8969	**0.9369**	0.8731	0.9201	0.9192	0.8623	0.9234	0.8846	**0.9538**
	JPEG	0.8800	0.8685	0.8546	0.8723	**0.9084**	0.8846	0.8654	0.8080	**0.9283**
	JP2K	0.8150	0.8662	0.9016	0.9011	**0.9280**	0.9188	0.9123	0.8592	**0.9453**
CSIQ	WN	0.8750	0.9034	0.9368	**0.9379**	**0.9471**	0.8692	0.8080	0.6196	0.9192
	JPEG	0.7867	0.8802	**0.9254**	0.9248	0.9161	0.9045	0.8888	0.8347	**0.9254**
	JP2K	0.7326	0.8662	0.9052	0.8952	0.9177	**0.9235**	0.9055	0.8509	**0.9244**
	GB	0.8031	0.8754	**0.9164**	0.9123	0.9132	0.8687	0.9066	0.8825	**0.9266**
Toyama	JPEG	0.5689	0.8880	0.8192	0.8906	**0.9199**	0.8106	0.8911	0.7678	**0.9391**
	JP2K	0.7773	0.7192	0.7845	0.8271	0.7996	0.7931	**0.8419**	0.6978	**0.8764**
MLIVE	GB+JPEG	0.8735	0.8773	0.8993	0.9029	0.8237	**0.9111**	0.9006	0.8403	**0.9287**
	GB+WN	0.8825	0.8819	0.8898	**0.9022**	0.8632	**0.9235**	0.8991	0.8715	0.8918
SIQAD	JP2K	0.3016	0.4527	0.6234	0.4466	0.6716	0.4633	**0.7348**	0.4735	**0.7701**
	JPEG	0.3125	0.3519	0.3755	0.5690	0.4442	0.3766	**0.7682**	0.4899	**0.7523**
	WN	0.8199	0.8528	**0.8708**	0.8621	**0.8889**	0.8615	0.8404	0.8207	0.8530
	GB	0.8481	**0.8990**	0.8626	**0.8963**	0.8768	0.5807	0.8736	0.8262	0.8840
Hit count		0	3	3	4	**9**	3	5	0	**15**
Mean		0.7801	0.8380	0.8517	0.8671	0.8751	0.8255	**0.8759**	0.7890	**0.9073**
STD		0.1765	0.1534	0.1306	0.1262	0.1190	0.1577	**0.0634**	0.1276	**0.0564**

In HOSA, three types of features which represent the first order (mean), second order (variance) and third order (skewness) statistics respectively are utilized. In order to understand the relative contribution of each type of feature in HOSA, we separately evaluated the performance of each feature on six databases. SROCC is used as the performance metric. The results are reported in Table VI. From the results shown in Table VI, we can draw the following conclusions. First, except for Toyama, using only a single type of feature, the performance is obviously inferior than the combined feature. Second, skewness based feature is a good complement to the mean and variance based features which further validate the non Gaussian characteristics in low level image features. In addition, we also added the kurtosis statistics (fourth order) based feature, but no obvious improvement is found.

4) Performance on Individual Distortion: For individual distortion, we tested the images belonging to each distortion in the testing set with the model trained on 80% of images including all types of distortions in that database. The results are summarized in Table VII and the best two results are highlighted in bold. For brevity, we only present SROCC results. Similar conclusions were obtained for PLCC and RMSE. In addition, we report the hit count (i.e., the number of times ranked in the top 2 for each distortion type), mean and standard deviation (STD) of the performance for each BIQA model at the bottom of the table. It can be seen that HOSA has the highest hit count (i.e. 15 times), followed by GM-LOG (10 times) and CORNIA-10K (6 times). Furthermore, HOSA has the highest average performance over all types of distortions and the lowest STD.

In order to further validate the generalization ability of HOSA, we conducted the 1000 train-test experiment for all BIQA models on the entire TID2013 and CSIQ databases. The distortion types in TID2013 database include: #01 additive white Gaussian noise, #02 additive noise in color components, #03 additive Gaussian spatially correlated noise, #04 masked noise, #05 high-frequency noise, #06 impulse noise, #07 quantization noise, #08 Gaussian blur, #09 image denoising, #10 JPEG compression, #11 JPEG2000 compression, #12 JPEG transmission errors, #13 JPEG2000 transmission errors, #14 non eccentricity pattern noise, #15 local blockwise distortion of different intensity, #16 mean shift, #17 contrast change, #18 change of color saturation, #19 multiplicative Gaussian noise, #20 comfort noise, #21 lossy compression of noisy images, #22 image color quantization with dither, #23 chromatic aberrations and #24 sparse sampling and reconstruction. The other two distortion types in CSIQ database are additive pink Gaussian noise (PN) and contrast change distortion (CTC). The results are presented in Table VIII and IX and the best two results are highlighted in boldface.

From Table VIII and IX, we can draw several conclusions. First, HOSA performs continuously better than both CORNIA-10K and CORNIA-100 which further validates its effectiveness. Second, almost all BIQA models fail to evaluate several uncommon distortions, such as #14, #15, #16, #17, #18 in TID2013 database and contrast distortion in CSIQ database. #14 and #15 consist of very localized distortion patterns which have limited influence on global image feature. Contrast distortion, #16 and #17 are correlated to image luminance change which is generally overlooked since current algorithms always work on normalized images. And #18 is mainly about color saturation thus most BIQA methods which based on luminance image processing fail to accurately estimate the resultant quality. All these failed examples could lead us to develop more comprehensive BIQA models in the future. Recently [53], [54] made efforts to measure the contrast distortion and [55] tried to evaluate multiply distortions with gradient information. All of them showed some promising results.

Since the training based BIQA methods rely on the trained distortion types, it is not easy to predict image quality with unseen distortion. We perform a leave one distortion type out

TABLE VIII
PERFORMANCE EVALUATION (SROCC) ON ENTIRE TID2013 DATABASE.

Method	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13
BLIINDS-2	0.7142	**0.7282**	**0.8245**	0.3577	0.8523	0.6641	0.7799	0.8523	0.7538	0.8077	0.8615	0.2512	**0.7550**
BRISQUE	0.6300	0.4235	0.7265	0.3210	0.7754	0.6692	0.5915	0.8446	0.5533	0.7417	0.7988	0.3012	0.6715
GM-LOG	0.7808	0.5881	0.8177	**0.5449**	**0.8892**	0.6593	0.8000	0.8485	0.7531	0.7992	0.8431	0.3985	**0.7473**
ILNIQE	**0.8909**	**0.8231**	**0.9292**	**0.6488**	0.8811	**0.8023**	**0.8808**	0.8454	0.7781	**0.8746**	**0.9106**	0.3100	0.6246
CORNIA-10K	0.3408	-0.1962	0.6892	0.1835	0.6071	-0.0138	0.6731	**0.8957**	**0.7866**	0.7854	0.8831	**0.5515**	0.5469
CORNIA-100	0.5591	0.1865	0.4846	0.3038	0.7143	0.2945	0.5874	0.8627	0.7644	0.6059	0.8138	0.3391	0.4677
HOSA	**0.8529**	0.6250	0.7820	0.3677	**0.9046**	**0.7746**	**0.8101**	**0.8924**	**0.8702**	**0.8931**	**0.9323**	**0.7472**	0.7012
Method	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	All	
BLIINDS-2	0.0812	**0.3713**	0.1585	-0.0823	**0.1092**	0.6987	0.2223	0.4505	0.8146	0.5676	0.8562	0.5504	
BRISQUE	0.1751	0.1835	0.1545	0.1246	0.0315	0.5596	0.2823	0.6803	0.8038	0.7145	0.7995	0.5615	
GM-LOG	**0.2054**	0.2419	0.0758	**0.2946**	-0.1831	0.7246	0.2502	0.6419	**0.8565**	0.6582	**0.9031**	**0.6750**	
ILNIQE	-0.1173	-0.0512	**0.2221**	0.0262	-0.1008	**0.7358**	0.3877	**0.8692**	0.7931	**0.7887**	0.8931	0.5165	
CORNIA-10K	0.1605	0.0962	0.0077	**0.4233**	-0.0554	0.2593	**0.6064**	0.5546	0.5919	0.7592	0.9023	0.6509	
CORNIA-100	0.1911	0.0704	0.0585	0.1719	0.0527	0.4785	0.3538	0.6377	0.5688	0.6803	0.8585	0.5167	
HOSA	**0.1989**	**0.3273**	**0.2327**	0.2938	**0.1185**	**0.7819**	**0.5315**	**0.8354**	**0.8554**	**0.8014**	**0.9052**	**0.7280**	

TABLE IX
PERFORMANCE EVALUATION (SROCC) ON ENTIRE CSIQ DATABASE.

Method	WN	CTC	PN	GB	JP2K	JPEG	All
BLIINDS-2	0.7019	0.3358	**0.8120**	**0.8803**	0.8501	0.8464	0.7740
BRISQUE	0.7068	0.3923	0.7935	0.8525	0.8478	0.8608	0.7619
GM-LOG	**0.8022**	**0.5708**	0.7515	0.8707	**0.8981**	**0.8803**	**0.8035**
ILNIQE	**0.8676**	**0.5443**	**0.8836**	0.8674	**0.9239**	**0.9043**	**0.8210**
CORNIA-10K	0.2407	0.4719	0.2841	**0.9128**	0.8536	0.8042	0.7076
CORNIA-100	0.6147	0.3488	0.3148	0.8759	0.8447	0.8118	0.6191
HOSA	0.7310	0.5056	0.7426	0.8763	0.8654	0.8333	0.7930

TABLE X
PERFORMANCE EVALUATION (SROCC) FOR LEAVE ONE DISTORTION TYPE OUT TEST.

Database	JP2K	JPEG	WN	GB	FF
LIVE	0.9041	0.8243	0.8578	0.9488	0.8438
TID2013	0.9301	0.8648	0.6374	0.8982	–
CSIQ	0.8852	0.8043	0.2312	0.8966	–

test on LIVE, TID2013 and CSIQ databases. According to the image content, the database is again randomly divided into two non-overlapping subsets 1000 times. Each subset contains distorted images from 50% of reference images. Then we choose images with one type of distortion in a subset for testing, and images with other types of distortions in another subset for training. Therefore the training and testing set include images with different types of distortions and have no content overlapping. The results are presented in Table X. We can see that except for WN, the algorithm performs well on other distortions. This can be attributed that WN has very different characteristics.

5) Database Independence Analysis: It is expected that the BIQA model learned from one database should be applicable to images from another database. Therefore, to demonstrate the generality and robustness of one BIQA model, it is necessary to see if satisfying results could still be obtained by applying the BIQA model learned from one database to another. In this subsection, we conducted the following experiments. Since LIVE, CSIQ, TID2013 and Toyama only contain natural images with similar distortion types, we trained HOSA on all images from LIVE database and tested it on the other three databases. In Table XI, we report the performance of BIQA models across all distortion categories and the whole database. The best two results are bolded. Again, it is notable that HOSA performs well on all three databases. Those performance indices above have confirmed the validness and robustness of the proposed model across a broad range of images.

6) Computation Complexity: The computational complexity and running cost of competing BIQA models are presented in Table XII. Experiments were performed on a SunFire X4170 with a 2.8GHz Intel Xeon X5560 CPU with non optimized MATLAB code. The software platform was MATLAB R2013a. The cost time includes feature extraction time and quality prediction time which is consumed by each BIQA model for evaluating the quality of 100 fixed color images with resolution of 512×512 (taken from CSIQ database) by the model trained on LIVE database. With a smaller codebook, HOSA has a much lower running time then CORNIA-10K but without any loss of algorithm performance.

B. Experiments on images with realistic distortions

1) Database: In the previous subsection, we conducted experiments on the images with simulated distortions. However, these distortions are less relevant for real world applications. And there are some newly released databases containing images with realistic non-simulated distortions. Therefore we tested the proposed method on these realistic distortions in this subsection. The datasets information is described as follows.

LIVE In the Wild Image Quality Challenge Database (Challenge DB) [56]: This database contains 1162 test images with diverse authentic distortions. The images were captured with different devices under various conditions. The MOS range is $[0, 100]$.

CID2013 [57]: It includes 6 sets of photos, and within each set, 6 image scenes are captured by different types of devices. Since the subjective experiment is different for Sets 1-3 and Sets 4-6. We choose Sets 4-6 to perform algorithm comparison. There are 234 images in total and the MOS range is $[0, 100]$.

We also tested on two document image datasets to further show its generalization ability. Unlike natural image quality evaluation, the task of document image quality assessment is usually to predict the OCR accuracy not MOS/DMOS. The datasets information is described as follows.

Newspaper database [27]: This database contains 521 grayscale text zone images with different resolutions. These

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2585880, IEEE Transactions on Image Processing

TABLE XI
DATABASE INDEPENDENCE TEST: TRAINED ON LIVE AND TESTED ON TID2013, CSIQ AND TOYAMA.

Database	Dist.	BIQI	DIIVINE	BLIINDS-2	BRISQUE	GM-LOG	ILNIQE	CORNIA-10K	CORNIA-100	HOSA
TID2013	WN	0.8127	0.8643	0.8232	0.8448	**0.8950**	**0.8859**	0.7790	0.6340	0.8385
	GB	0.8463	0.8817	0.8938	0.8814	0.8893	0.8349	**0.9176**	0.8572	**0.9192**
	JPEG	0.8864	0.8629	**0.8992**	0.8909	0.8972	0.8681	0.8854	0.8213	**0.9150**
	JP2K	0.8390	0.8773	0.8650	0.8979	**0.9349**	0.9121	0.9092	0.8424	**0.9337**
	All	0.8389	0.8824	0.8644	0.8832	**0.9071**	0.8770	0.8847	0.8070	**0.9037**
CSIQ	WN	0.7815	0.8517	**0.9309**	0.9185	**0.9196**	0.8497	0.7507	0.6968	0.8847
	JPEG	0.6794	0.8698	**0.9160**	0.9035	0.9092	0.8993	0.8983	0.8341	**0.9176**
	JP2K	0.6897	0.8671	0.8612	0.8656	0.8764	0.9060	**0.9079**	0.8620	**0.9097**
	GB	0.8330	0.8827	0.9033	0.8983	0.8621	0.8579	**0.9142**	0.8930	**0.9038**
	All	0.8025	0.8707	**0.9030**	0.8957	0.8966	0.8799	0.8956	0.8440	**0.9125**
Toyama	JPEG	0.5878	**0.8964**	0.8685	0.8601	0.8147	0.7073	0.7660	0.6053	**0.8775**
	JP2K	0.7376	0.8332	0.8253	0.8698	**0.8748**	0.7383	0.8475	0.6207	**0.9005**
	All	0.6642	**0.8606**	0.8234	0.8513	0.8363	0.7114	0.8049	0.6113	**0.8861**

TABLE XII
COMPUTATIONAL COMPLEXITY AND AVERAGE RUN TIME COMPARISON FOR EACH BIQA METHOD.

Method	Computational complexity	Time cost (Second)
BIQI	$O(N)$	0.1119
DIIVINE	$O(N(log(N)+m^2+N+392b))$ m: neighborhood size in DNT; b: bin number of 2D histogram	24.9523
BLIINDS-2	$O((N/(d^2))log(N/(d^2)))$ d: block size	68.4110
BRISQUE	$O(Nd^2)$ d: block size	0.0768
GM-LOG	$O(N(h+k))$ h: filter size; k: probability matrix size	0.0677
ILNIQE	$O(N(d^2+h+gh))$ d: block size; h: filter size; g: log-Gabor filter size	12.1478
CORNIA-10K	$O(Nd^2K)$ d: block size; K: codebook size	1.6216
CORNIA-100	same as CORNIA-10K	0.2271
HOSA	$O(Nd^2K)$ d: block size; K: codebook size	0.3529

images are a subset of historical document images with machine printed English and Greek. A commercial OCR software (ABBYY Fine Reader) is used to obtain OCR results and ISRI-OCR evaluation tool [58] is used to generate OCR accuracies in the range $[0,1]$. This database contains character level distortions.

Sharpness-OCR-Correlation (SOC) database [59]: It contains camera-captured document images with blurriness and noisiness. A 8 mega-pixel cell phone camera was used to generate different versions of 25 non-distorted document images. Finally a total of 175 color images with fixed resolution (3264×1840) were created. The OCR results were obtained by ABBYY Fine Reader and OCR accuracies were also generated by the ISRI-OCR evaluation tool.

2) Evaluation: We first compare competing methods on real photos. For Challenge DB, 80% images are randomly selected for training and the rest 20% images for testing. For CID2013, there are 6 image scenes (6 types of image content) in total. The detailed scene information can be found in [60]. We performed a leave one out cross validation and each image scene was tested once. Since there is only one type of content for testing one time, homogeneous content could result in a higher median SROCC. Therefore every combination of 2 different image scenes was taken and the MOS predicted from the cross validation was used to compute a median SROCC. This is to ensure that more than one type of image content is present in the testing phase. The median SROCC scores on Challenge DB and CID2013 are presented in Fig. 6. Obviously HOSA outperforms feature learning based method CORNIA and is competitive to handcrafted feature based BIQA methods on both databases.

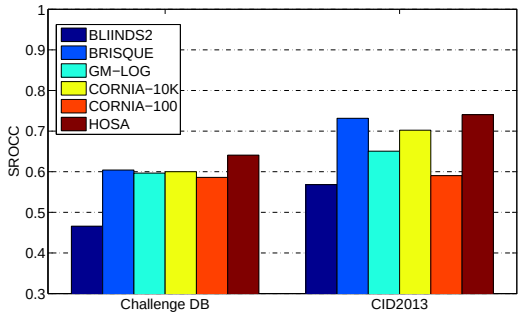


Fig. 6. Performance evaluation (SROCC) on two image databases with realistic distortions.

For the document image quality assessment, there is no reference image in Newspaper database, therefore we randomly choose 80% of images for training and the remaining for testing. Five methods were selected for comparison, including BRISQUE [18], GM-LOG [20], CORNIA-10K [26], SFL-100 [27] and a CNN based method [61]. Experimental results on the Newspaper and SOC databases are shown in Fig. 7. We can observe that HOSA significantly outperforms CORNIA-10K and SFL-100. And it is very comparable to CNN which utilize sophisticated training approach.

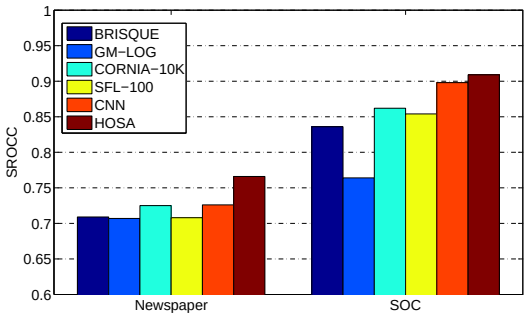


Fig. 7. Performance evaluation (SROCC) on two document image databases.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2585880, IEEE Transactions on Image Processing

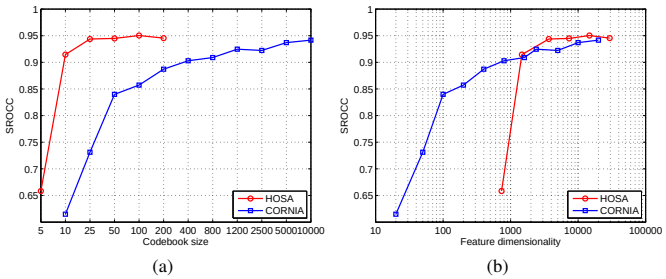


Fig. 8. Comparison of HOSA and CORNIA as a function of the codebook size (left) and feature dimensionality (right) on LIVE.

C. Discussion

The performance advance of HOSA is due to three components: the comprehensive codebook, the high order statistics aggregation scheme and the alleviation of image content variations. First, former feature learning based BIQA methods only provide the mean of cluster to build the codebook, while the valuable high order statistics (i.e., mean, dimension wise variance and skewness) of clusters are generally inobservant. However, we generated a comprehensive codebook from K-means clustering in this paper. Second, the differences of high order statistics between clusters and images are aggregated to form the quality aware representation. It produces more abundant information with a quite small codebook. And it describes how the parameters of the codebook from all kinds of local image features should be modified to better fit the given image features. In other words, it calculates the statistical differences between local features and quality aware codewords. Therefore, it is appropriate to discriminate images of different distortion severities for quality assessment. Third, several schemes are utilized to alleviate the influence of image content, e.g., the normalization and whitening on local patches, and the signed power normalization.

We also made a detailed comparison between HOSA and CORNIA to demonstrate the superiority of HOSA. We believe this comparison to be completely fair, since we use the same local features and the same codebook. We show the results on LIVE of SROCC in Fig. 8 both as a function of the codebook size and as a function of the feature dimensionality. For the same codebook size, HOSA significantly outperforms CORNIA all the time. This is not surprising because HOSA extracts more abundant information than CORNIA. And for a small codebook, the difference is particularly impressive. For example, with 10 codewords, HOSA still obtains a SROCC performance over 0.91 while CORNIA only gets 0.61. For the same feature dimensionality, CORNIA performs slightly better for small number of dimensions but HOSA performs better for larger number of dimensions. And with similar feature dimensionality, the computation speed for HOSA is much faster than CORNIA (see Table XII).

Another interesting issue is that almost all handcrafted feature based BIQA methods performed poorly on unnatural images. This is mainly because these features are generated based on the assumption that natural scene images share similar statistical regularities, while this is not suitable to screen content images which contain text and graphics and document images which only contain text information. The statistical distribution varies severely with diverse screen contents [51]. Conversely, HOSA extracts a comprehensive codebook which contains various types of patterns. Although the codebook constructed from a set of natural images, it represents image primitive microstructure, for example, spots, lines and corners, which are also the basic elements of characters and graphics. Consequently HOSA has a better generalization ability for these unnatural images.

IV. Conclusion

Previous feature learning based BIQA methods typically utilize zero order statistics for image quality evaluation. And they require large codebook and codebook updating to achieve satisfied performance. Former handcrafted feature based BIQA methods rely on the natural scene statistics which are not applicable to images with artificial components. To solve these problems, in this paper, we made first attempt to incorporate high order statistics for image perceptual quality evaluation. By aggregating the soft weighted first, second and third order statistics differences between normalized image patches and quality aware clusters, the effective image quality assessment model HOSA is generated. With a small codebook, it is superior to former feature learning based BIQA methods and comparable to state-of-the-art handcrafted feature based BIQA methods in terms of prediction accuracy, generalization ability and computation complexity.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful and constructive comments on this paper. The MATLAB source code of our algorithm is publicly available at http://lampsrv02.umiacs.umd.edu/pubs/LampMedia.php.

References

[1] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*, vol. 101, pp. 2008–2024, Sep. 2013.
[2] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Jun. 2011.
[3] 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.
[4] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[5] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
[6] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[7] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Conf. Image Processing (ICIP)*, 2002, pp. 474–477.
[8] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
[9] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal processing: Image communication*, vol. 19, no. 2, pp. 163–172, 2004.
[10] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[11] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.

[12] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Jan. 1995.

[13] D. L. Ruderman, T. W. Cronin, and C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[14] J. Geusebroek and A. Smeulders, "A six-stimulus theory for stochastic texture," *Int. J. Comp. Vis.*, vol. 62, no. 1, pp. 7–16, Apr. 2005.

[15] A. Vo, S. Oraintara, and N. Nguyen, "Vonn distribution of relative phase for statistical image modeling in complex wavelet domain," *Signal Processing*, vol. 91, no. 1, pp. 114–125, Jan. 2011.

[16] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[17] M. A. Saad and A. C. Bovik, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

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

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

[20] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[21] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, Feb. 2015.

[22] H. Tang, N. Joshi, and A. Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 2877–2884.

[23] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[24] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[25] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2011.

[26] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2012, pp. 1098–1105.

[27] ——, "Real-time no-reference image quality assessment based on filter learning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2013, pp. 987–994.

[28] L. Zhang, Z. Gu, X. Liu, H. Li, and J. Lu, "Training quality-aware filters for no-reference image quality assessment," *IEEE Multimedia*, vol. 22, no. 4, pp. 67–75, Apr. 2014.

[29] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 1733–1740.

[30] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *ECCV Workshop on Statistical Learning for Computer Vision*, 2004, pp. 1–22.

[31] A. Coates, H. Lee, and A. Y. Ng., "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2011, pp. 215–223.

[32] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 2486–2493.

[33] J. Xu, Q. Li, P. Ye, H. Du, and Y. Liu, "Local feature aggregation for blind image quality assessment," in *Proc. IEEE Conf. Visual Communication and Image Processing (VCIP)*, 2015 (in press).

[34] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2007, pp. 1–8.

[35] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2010, pp. 3304–3311.

[36] D. Picard and P. Gosselin, "Improving image similarity with vectors of locally aggregated tensors," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2011, pp. 669–672.

[37] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 1784–1791.

[38] S. Lyu and E. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2008, pp. 1–8.

[39] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[40] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods," in *Proc. British Machine Vision Conf. (BMVC)*, 2011, pp. 1–8.

[41] S. Lloyd, "Least square quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[42] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006:1–21, Jan.-Mar. 2010.

[43] D. Pelleg and A. Moore, "Accelerating exact k-means algorithms with geometric reasoning," in *Proc. 5th ACM Int. Conf. on Knowledge discovery and data mining (SIGKDD)*. ACM, 1999, pp. 277–281.

[44] Y. Jia and T. Darrell, "Heavy-tailed distances for gradient based image descriptors," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2011, pp. 397–405.

[45] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *Proc. Euro. Conf. Computer Vision (ECCV)*, 2010, pp. 143–156.

[46] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "LIBLINEAR: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, Sep. 2008.

[47] H. Sheikh, M. Sabir, and A. 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.

[48] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Europian Workshop on Visual Information Processing (EUVIP)*, Paris, France, Jun. 2013, pp. 106–111.

[49] Y. Horita, K. Shibata, Y. Kawayoke, and Z. Sazzad. MICT image quality assessment database. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html

[50] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. IEEE Conf. Record of the 46th Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 1693–1697.

[51] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.

[52] "Final report from the video quality experts group on the validation of objective models of video quality assessment," VQEG, Tech. Rep., Jun. 2000. [Online]. Available: http://www.vqeg.org/

[53] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–841, Jul. 2015.

[54] M. A. Saad, P. Corriveau, and R. Jaladi, "Objective consumer device photo quality evaluation," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1516–1520, Oct. 2015.

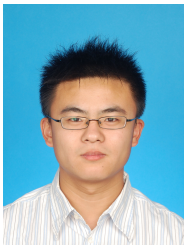
[55] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.

[56] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.

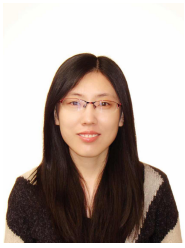
[57] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Hakkinen, "Cid2013: a database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.

[58] R. Smith, "ISRI-OCR evaluation tool," http://code.google.com/p/isri-ocr-evaluation-tools/.

[59] J. Kumar, P. Ye, and D. Doermann, "A dataset for quality assessment of camera captured document images," in *Camera-Based Document Analysis and Recognition (CBDAR)*, 2013, pp. 113–125.
[60] M. Nuutinen, T. Virtanen, and P. Oittinen, "Image feature subsets for predicting the quality of consumer camera images and identifying quality dimensions," *Journal of Electronic Imaging*, vol. 23, no. 6, pp. 061 111:1–17, Nov.-Dec. 2014.
[61] L. Kang, P. Ye, Y. Li, and D. Doermann, "A deep learning approach to document image quality assessment," in *Proc. IEEE Conf. Image Processing (ICIP)*, 2014, pp. 2570–2574.



Jingtao Xu (S'14) received his B.Sc. degree in Electronic Information Science and Technology from Beijing Normal University in 2008. He is currently pursuing his Ph.D. degree in Beijing University of Posts and Telecommunications. From 2013 to 2015, he was a Visiting Student with the University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland, College Park. His research interests include image & video quality assessment, computer vision and machine learning.



Peng Ye received the B.Eng. degree from Tsinghua University, Beijing, China, in 2007, the M.Eng. degree in electrical engineering from the University of Delaware, Newark, in 2009 and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2014. She is currently a Software Engineer with Airbnb in California, USA. Her area of research is image processing, computer vision and machine learning, with a focus on image quality assessment.

Qiaohong Li received the B.Eng. degree and M.Eng. degree in School of Information and Communication Engineering from Beijing University of Posts and Telecommunications, Beijing, China. She is now pursuing the Ph.D. degree with the School of Computer Engineering, Nanyang Technological University, Singapore. Her research interests include image quality assessment, speech quality assessment, computer vision, and visual perceptual modeling.

Haiqing Du received the B.Eng. degree and M.Eng. degree from Liaoning Technical University in 1999 and 2002 respectively, and the Ph.D. degree in communication and information system from Beijing University of Posts and Telecommunications in 2010. She is currently a Lecturer in School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. From 2013 to 2014, she was a Visiting Scholar with the Department of Electrical Engineering of the University of Washington in Seattle. Her research interests include image & video quality assessment and network optimization.

Yong Liu received his B.Eng. degree, M.Eng. degree and Ph.D. degree in telecommunications and information systems from Beijing University of Posts and Telecommunications in 1984, 1987 and 2001 respectively. He is a Full Professor in School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include mobile communications and multimedia communications.

David Doermann (SM'09-F'13) received the B.Sc. degree in computer science and mathematics from Bloomsburg University, Bloomsburg, PA, in 1987, the M.Sc. degree from the University of Maryland, College Park, in 1989, the Ph.D. degree from the University of Maryland, in 1993, and the honorary doctorate of technology sciences from the University of Oulu, Oulu, Finland, in 2002, for his contributions to digital media processing and document analysis research.

He is a senior research scientist with the University of Maryland Institute for Advanced Computer Studies (UMIACS) and has more than 30 journal publications and 125 refereed conference papers. Dr. Doermann is a founding co-editor of the International Journal on Document Analysis and Recognition, was the general chair or co-chair of more than a half dozen international conferences and workshops, and was the general chair of the International Conference on Document Analysis and Recognition in 2013. And he is a fellow of the IEEE and IAPR. He is currently on assignment serving as Program Manager at DARPA.