No-Reference Image Quality Assessment using Visual Codebooks

Peng Ye, *Student Member, IEEE* David Doermann, *Senior Member, IEEE*

***Abstract*—The goal of no-reference objective image quality assessment (NR-IQA) is to develop a computational model that can predict the human perceived quality of distorted images accurately and automatically without any prior knowledge of reference images. Most existing NR-IQA approaches are distortion-specific (DS) and are typically limited to one or two specific types of distortions. In most practical applications, however, information about the distortion type is not really available. In this paper, we propose a general-purpose NR-IQA approach based on visual codebooks. A visual codebook consisting of Gabor filter based local features extracted from local image patches is used to capture complex statistics of a natural image. The codebook encodes statistics by quantizing the feature space and accumulating histograms of patch appearances. This method does not assume any specific types of distortions, however, when evaluating images with a particular type of distortion, it does require examples with the same or similar distortion for training. Experimental results demonstrate that the predicted quality score using our method is consistent with human perceived image quality. The proposed method is comparable to state-of-the-art general purpose NR-IQA methods and outperforms the full-reference image quality metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) on the LIVE image quality assessment database.**

***Index Terms*—no-reference image quality assessment, visual codebook, texture analysis, Gabor filter**

I. Introduction

Assessing the quality of visual information plays an important role in numerous image/video processing and computer vision applications. In recent years, a large number of approaches have been developed to measure the perceptual quality of images. Previous work on objective image quality assessment (IQA) can be broadly classified into content-based and degradation-based. Content-based IQA algorithms aim to evaluate the inherent property of an image. For example, Ke *et al.* [1] designs high level semantic features to classify between high quality professional photos and low quality snapshots. Li and Chen [2] study the aesthetic visual quality of paintings. In this scenario, images are usually assumed to be noise-free, but a degradation-based approach focuses on evaluating various degradations that arise from the production process. Based on the availability of a reference image, objective degradation-based IQA algorithms can be further classified into: full-reference (FR), no-reference (NR) and reduced-reference (RR) approaches [3]. While there have been some well established FR [4]–[8] and RR [3], [9], [10] methods which correlate well with human perception of quality, there is still considerable room for the improvement of NR-IQA methods. This paper proposes a general purpose NR-IQA metric.

NR-IQA methods aim to predict the quality of distorted images with respect to human perception automatically without prior knowledge of reference images. Most of the existing NR-IQA algorithms [11]–[15] are distortion-specific (DS) and assume the type of distortion is known. They are typically limited to one or two specific types of distortions. For example, Wang *et al.* [11] introduce blockiness measures for JPEG compressed images; Gastaldo and Zunino [12] use neural networks to learn a mapping from feature space to quality scores for JPEG compressed images; Brandão *et al.* [13] evaluate noise in block-based DCT domain arising from JPEG or MPEG encoding based on natural scene statistics of the DCT coefficients; Marziliano *et al.* [14] introduce blur and ringing measures for JPEG2K compressed images and Sheikh *et al.* [15] develop natural scene statistics based approach for JPEG2K compressed images. However, in most practical applications, information about distortion type is not available. This underlying assumption limits the application domain of these approaches. It is desirable to design general purpose non-distortion-specific (NDS) NR-IQA methods that do not examine the exact prior knowledge of distortion.

Existing general purpose NR-IQA algorithms [16]–[25] assume that examples which possess the same or similar types of distortions as in testing images are available and apply machine learning techniques for quality estimation. The NR-IQA problem is usually transformed to a regression or classification problem, where a regressor or classifier is trained using features related to image quality. Based on the types of features used, previous approaches usually follow one of the following two trends (1) Natural scene statistics (NSS) based approaches [16]–[21] and (2) Training-based approaches [23]–[25]. NSS based approaches assume that natural scenes possess certain statistical properties and the presence of distortion will affect these properties. Current state-of-the-art NR-IQA algorithms explore NSS based features [16]–[19], [23]. Statistical models such as generalized Gaussians are used to characterize statistical properties of wavelet coefficients [16], [18], [23], cosine coefficients [17], [19] or contourlet coefficients [20]. Estimated model parameters are used as features to perform regression. The second approach relies on a large number

P. Ye is with the Institute for Advanced Computer Studies, 3348 AV Williams Building, University of Maryland, College Park, MD 20742. E-mail: pengye@umiacs.umd.edu

D. Doermann is with the Institute for Advanced Computer Studies, 3451 AV Williams Building, University of Maryland, College Park, MD 20742. E-mail: doermann@umiacs.umd.edu

of features which are designed to capture relevant factors affecting image quality. Then different regression techniques such as Support Vector Regression (SVR) [23] and Neural Networks [24] can be applied to learn the mapping from feature space to image quality. These features, however, may not be easily interpreted. All of these previous approaches use human knowledge of the distorted and non-distorted images in feature design, but no information about training images was explored. A summary of previous IQA methods cited in this paper can be found in Table I[1].

This paper presents a simple yet effective learning-based approach for general purpose NR-IQA. The proposed algorithm differs from previous approaches in the following aspects: First, the previous NSS based approaches construct statistical models at pixel level in some transform domain and then derive patch level and image level features related to image quality based on domain knowledge. Our method extracts patch level features and explores information of training images for local descriptor encoding. This process requires less domain knowledge compared to previous approaches. Differences between our method and previous methods are shown in Fig. 1. Second, instead of explicitly building a statistical model for image patches in high dimensional feature space, our method uses a visual codebook based method for feature space quantization and then learns the mapping from the quantized feature space to image quality scores. Specifically, we use the visual codebook method and Gabor filter based features to effectively capture image statistics. Visual codebooks constructed from descriptors extracted from local image patches have been widely used in texture analysis and visual recognition [26], [27]. They can effectively capture the complex statistics of real images in a convenient local form. There are various methods for modeling textures and extracting texture features, and we choose to use Gabor filter based features which have been successfully applied to texture discrimination [28]–[31]. The proposed method is block-based and solely based on local image features, thus has the potential to be used in real-time applications with a parallel implementation.

Extensive experiments have been conducted to demonstrate the effectiveness of this new framework. Our experimental results on the LIVE IQA database [6], [32] show that the proposed method is comparable to state-of-the-art general purpose NR-IQA methods and outperforms the full-reference image quality metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

The remainder of this paper is organized as follows. In Section II, we introduce visual codebook based NR-IQA methods. Experimental results and a thorough analysis of our results are presented in Section III. Finally, Section IV concludes with a summary of our work.

II. Assessment using a Visual Codebook

The proposed method is based on the assumptions that cues to distortion type and level can be captured from local image patches. Ideally, we want to extract features that are

[1]This is by no means a complete list of previous works.

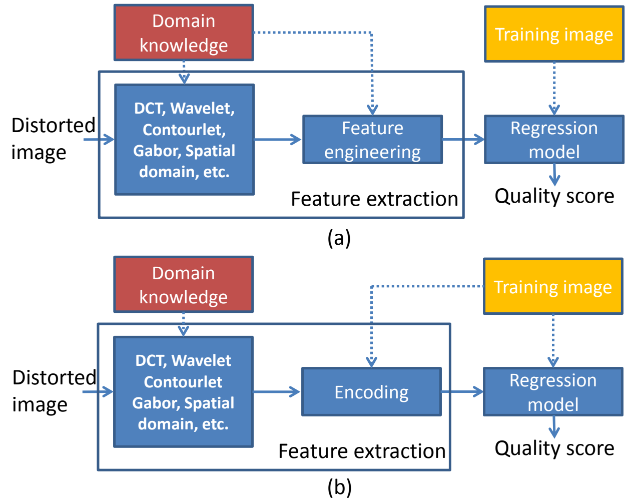


Fig. 1: Overview of general purpose NR-IQA systems. (a) Previous systems. (b) Our method.

Fig. 2: "House": an example image from LIVE database.

highly correlated with distortion but have no correlation with image content. However, the content of the image will alter the observation of distortion and there are distortions which are dependent on image content. Thus we need to look at the features extracted from local patches together to gain a better understanding of the quality of an image and for this purpose we build a codebook of image patches and examine the codeword distribution to capture "quality information". The proposed framework consists of local feature extraction, codebook construction, image representation and regression.

A. *Local feature extraction*

We use Gabor filtering for local feature extraction. The use of Gabor filter based feature is motivated by the observation that images with the same type of distortion and with similar quality share similar "texture" and Gabor filters are particularly appropriate for texture representation and discrimination. Examples of images blocks (cropped from Fig. 2) with different types and levels of distortions are shown in Fig. 3. As can be seen, the appearance of local image patches varies with the level of degradations and patches within the same image block share similar distortion patterns. Gabor filter based features

FR	**Reference**	**Measure**		
	Wang *et al.* [4]	Structural Similarity Index Metric (SSIM)		
	Sheikh *et al.* [5], [7]	Visual Information Fidelity (VIF)		
	Li and Bovik [8]	Modified SSIM		
RR	**Reference**	**Analysis Domain**		
	Wang and Bovik [3]	Wavelet		
	Gao *et al.* [9]	Multiscale geometric analysis in wavelet, contourlet, curvelet, etc		
	Tao *et al.* [10]	Contourlet		
DS-NR	**Reference**	**Distortion**	**Features**	**Regression**
	Wang *et al.* [11]	JPEG	Blockiness and activity metrics	Nonlinear function fitting
	Gastaldo and Zunino [12]	JPEG	Various	Neural networks
	Brandão *et al.* [13]	JPEG	DCT coefficient statistics	Nonlinear function fitting
	Marziliano *et al.* [14]	JPEG2K	Blur and ringing metrics	Nonlinear function fitting
	Sheikh *et al.* [15]	JPEG2K	Wavelet coefficient statistics	Nonlinear function fitting
NDS-NR	**Reference**	**Features**	**Regression**	
	Moorthy and Bovik [16], [18]	Wavelet coefficient statistics	SVM for Classification + SVR for regression	
	Saad *et al.* [17], [19]	DCT coefficient statistics	Probabilistic model / SVR	
	Lu *et al.* [20]	Contourlet domain statistics	Nonlinear function fitting	
	Tong *et al.* [22]	Raw image patch	Binary classifier + Adaboost	
	Tang *et al.* [23]	Various	SVR	
	Li *et al.* [24]	Various	Neural network	
	Ye and Doermann [25]	Gabor	Codebook based method	

TABLE I: Summary of previous IQA methods cited in this paper. FF: full reference IQA, RR: reduced reference IQA, DS-NR: distortion-specific no reference IQA, NDS-NR: non-distortion-specific/general-purpose no reference IQA. "Various" indicates that various features with different properties are used, which cannot be summarized in just a few words.

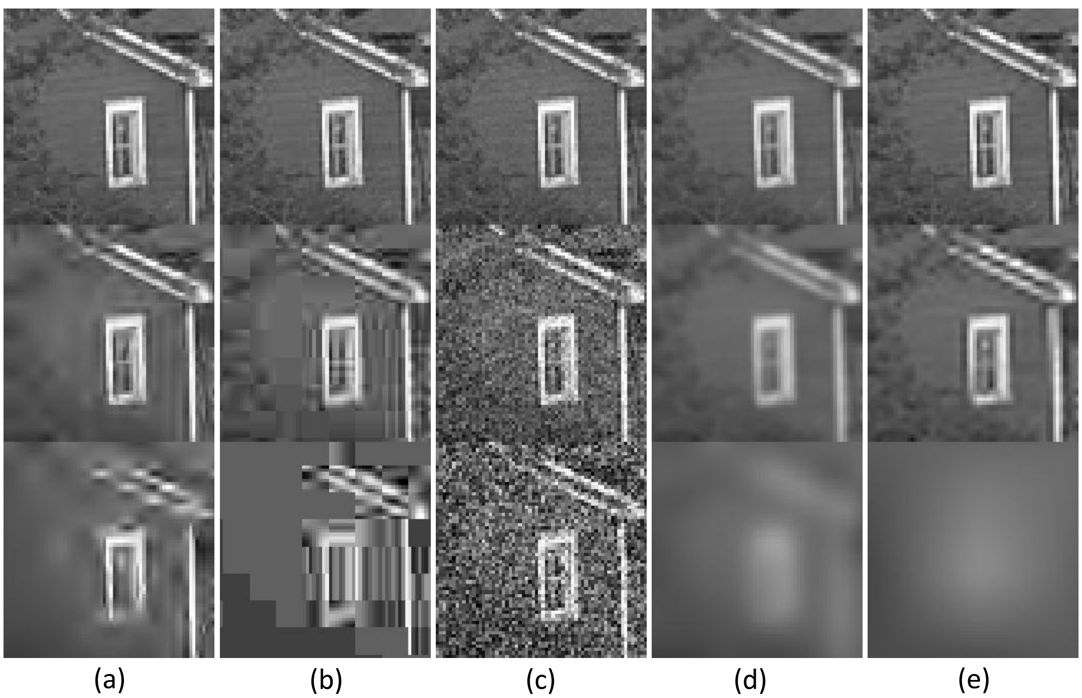


Fig. 3: Examples of image blocks with different types and levels of distortions. (a) JPEG2K compression (b) JPEG compression (c) white Gaussian noise (d) Gaussian blur (e) fast fading distortion. (block size=64×64)

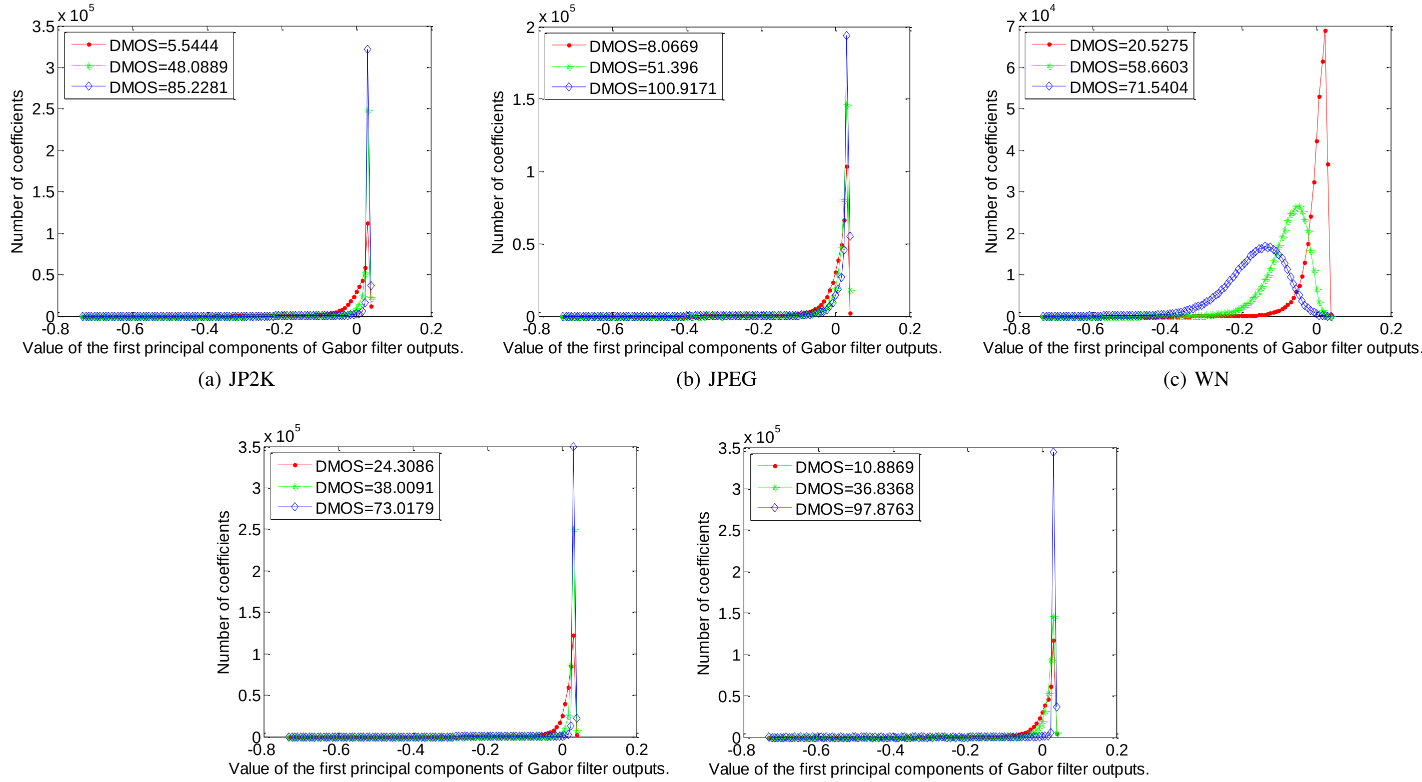
have the following positive properties: (1) they are optimal in time and frequency or space and spatial-frequency in two dimensions, (2) the frequency and orientation representations of Gabor filters are similar to those of human visual system, and (3) simple operations on Gabor filters can be established to achieve illumination, rotation, and translation invariance [30], [31]. A number of different Gabor filter based methods have been developed for extracting local image features [30]. In this paper, we use the simple Gabor features introduced in [31].

A normalized 2-D Gabor filter function in the continuous spatial domain is defined as follows:

$$\begin{cases} \psi(x,y,f,\theta) = \frac{f^2}{\pi\gamma\eta} exp[-(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2) + j2\pi f x'] \\ x' = xcos\theta + ysin\theta \\ y' = -xsin\theta + ycos\theta \end{cases} \tag{1}$$

where f is the frequency of sinusoidal plane wave, θ is the rotation of the Gaussian envelop and the sinusoidal, γ and η are the spatial widths of the filter along the major and the minor axis respectively.

Suppose the image function is $\xi(x,y)$, the response of



(a) JP2K (b) JPEG (c) WN

(d) BLUR (e) FF

Fig. 4: Perform Gabor filtering at five frequencies (1, $1/\sqrt{2}$, $1/2$, $1/(2\sqrt{2})$, $1/4$) in four orientations ($0°$,$45°$,$90°$,$135°$) and extract the first principal component associated with each pixel.

Gabor filter to ξ is given by the convolution between ξ and ψ:

$$g(x, y; f, \theta) = \psi(x, y, f, \theta) * \xi(x, y) \tag{2}$$

The feature matrix $G[g_{i,j}]$ for point (x_0, y_0) in ξ is given by:

$$G = \begin{bmatrix} g(x_0, y_0; f_0, \theta_0) & \cdots & g(x_0, y_0; f_0, \theta_{n-1}) \\ g(x_0, y_0; f_1, \theta_0) & \cdots & g(x_0, y_0; f_1, \theta_{n-1}) \\ \vdots & \ddots & \vdots \\ g(x_0, y_0; f_{m-1}, \theta_0) & \cdots & g(x_0, y_0; f_{m-1}, \theta_{n-1}) \end{bmatrix} \tag{3}$$

where m is the number of frequencies and n is the number of orientations. Illumination invariance is achieved by using a normalized feature matrix G'.

$$G' = \frac{G}{\sqrt{\sum_{i,j} |g_{i,j}^2|}} \tag{4}$$

To show that the distribution of Gabor filter responses varies with the levels of distortions, we perform Gabor filtering on the degraded versions of the "House" image at five frequencies (1, $1/\sqrt{2}$, $1/2$, $1/(2\sqrt{2})$, $1/4$) in four orientations ($0°$,$45°$,$90°$,$135°$) and extract the magnitude of responses at each pixel. Image blocks extracted from this group of degraded images are shown in Fig. 3. One pixel is associated with a 20-by-1 feature vector. We extract the first principal components associated with the collections of feature vectors. Plots of the distribution of the first principal components are shown in Fig. 4, where DMOS stands for *differential mean opinion score* which is a well established subjective image quality measure [6]. On the LIVE database, DMOS is generally in the range from 0 to 100 where a smaller DMOS indicates higher image quality. As can be seen, for JP2K, JPEG, BLUR and FF distortions, images with higher quality tend to have a lower peaked distribution but for WN, higher quality images tend to have a sharp peaked distribution. Fig. 4 shows the distribution of pixel level features.

Instead of examining pixel level feature statistics, we extract patch level features. Given an image patch, simple Gabor features are extracted from each point in the patch, then the mean and variance of the magnitude of the elements in the feature vector over all points in the patch are computed to form a $2mn$-by-1 vector, which is referred to as *Gabor feature vector*.

B. Codebook Construction

The next stage consists of building a "codebook" of image patches, that is used for feature space quantization. Given one training image, we randomly sample M different $B \times B$ small patches, where patches may overlap. Currently, most image or video compression algorithms (e.g. JPEG and JPEG2k) are block-based and often a block size of 8×8 is used, our patch size is chosen to be larger than 8 so that we can capture variation at block edges. In our experiments, we use $B = 11$.

Constant patches are removed since they do not contain any information for quality estimation. *Gabor feature vectors* are computed for the rest of image patches. By repeating this operation on all training images, we obtain a large set of *Gabor feature vectors* Ω. A codebook C is then created from this set using a clustering algorithm such as k-means. Codewords $C(i), i = 1, ..., D$ in C are the learned cluster centers.

C. Image Representation

Images are represented in terms of codewords from the codebook obtained in the previous stage. As is shown in Fig. 5, an effective image representation is obtained with the following steps (1) extracting local descriptors, (2) hard-assignment encoding of local descriptors and (3) average pooling of encoded local descriptors. Specifically, given an image I, we randomly sample M different $B \times B$ non-constant patches and extract *Gabor feature vectors* from them. The collection of *Gabor feature vectors* is denoted as G_1 in Fig. 5. Then, an encoded version of G_1 is obtained using the trained codebook and is denoted as G_2. In G_2, each column contains only one non-zero element which equals one and represents the nearest neighbor (in the codebook) of the corresponding column in G_1. Then the image level feature H_I for I is obtained by averaging over different local descriptors and can be considered the normalized (scaled to total sum 1) histogram of occurrence counts for the different codewords. The probability of the occurrence of a codeword $C(i)$ can be approximated by $H_I(i)$, and it is referred to as the *codeword histogram*. H_I is used as the input to regression program for estimating quality score of I.

D. Regression

In this section, we introduce two different approaches for regression. The first approach constructs a linear mapping from H_I to its associated quality score and the second approach learns the mapping using an off-the-shell regressor.

Approach 1: Example based method [25]

The basic idea of example based method is to approximate the quality score of a testing image I by the weighted average of the quality scores of training images. Given a collection of training images $\{I_1, I_2, ..., I_N\}$, this approximation can be written as follows:

$$Qm(I) = \sum_{k=1}^{N} w_k(I) DMOS(I_k) \tag{5}$$

where $w_k(I)$ is a similarity measure between I and the k-*th* training image I_k and $\sum_{k=1}^{N} w_k = 1$. $DMOS(I_k)$ is human evaluated quality score for I_k. To find the similarity measure, we apply clustering on individual training images. The set of cluster centers obtained from training image $I_k, k = 1, ..., N$ is denoted S_k. S_k can be considered as a "summary" of I_k. We combine cluster centers from all training images to form a codebook C. Then the similarity measure $w_k(I)$ can be written as:

$$w_k(I) = \frac{1}{M} \sum_{i=1}^{M} H_I(i) \delta(C(i) \in S_k) \tag{6}$$

where M is the number of local descriptors extracted from I and $\delta(C(i) \in S_k)$ is an indicator function which equals one if $C(i) \in S_k$ is true and zero otherwise. To demonstrate that the codeword histogram is a good indicator of image quality and to show how codewords are distributed differently for images with different levels of degradations, let's assume the quality score of a codeword $C(i)$ equals the DMOS of the training image which generates $C(i)$, i.e., $DMOS(C(i)) = DMOS(I_k)$ if $C(i) \in S_k$, then we can compute a histogram of codeword quality score as shown in Fig. 6, where the x-axis represents quality score of the codeword and y-axis represents the percentage of local image descriptors that are mapped to codewords with a particular quality score. These plots are obtained from images with the same content but with different levels and types of degradations. It can be seen that images of higher visual quality weigh more on "good" codewords and highly distorted images weigh more on "bad" codewords. This implies that a high quality image is more similar to high quality training images. Therefore the similarity measure defined in Eq. 6 is reasonable. The averaging scheme tends to smooth the predication, thus compared to DMOS, Qm is usually in a smaller range, but this does not affect the performance if we use correlation as evaluation measure.

Approach 2: Support vector regression (SVR)

Instead of explicitly finding the linear relationship between codeword histogram and quality score, we can use off-the-shell regressors to learn this mapping. Specifically, we use SVR with linear kernel in our experiment. Unlike the example based method, here we do not associate a quality score with a codeword, and thus are not required to use labeled training images for codebook construction. Codeword histogram is directly used as the input to SVR. Compared to approach 1 where the similarity measure between testing image and training image is defined ad-hoc, using SVR, a linear regression function which maps feature vector to quality score is discriminatively learned on training data.

The first approach is referred to as CBIQ-I and the second approach is referred as CBIQ-II in the remainder of this paper.

III. Experiments

A. Protocol

Database for evaluation: For evaluation, we use two databases.

(1) *LIVE IQA database [6], [32]:* We tested the proposed algorithm on the *Laboratory for Image & Video Engineering* (LIVE) image quality assessment database. The LIVE IQA database consists of 29 reference images and their degraded version with five different types of distortions, i.e, JPEG2k compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blurring (BLUR) and fast fading (FF). DMOS associated with distorted images are provided.

(2) *CSIQ IQA database [33]:* The CSIQ database consists of 30 reference images and their degraded versions with 6 different types of distortions at 4 to 5 different levels. DMOS associated with distorted images is provided and is in the range $[0, 1]$, where lower DMOS indicates higher quality. We used

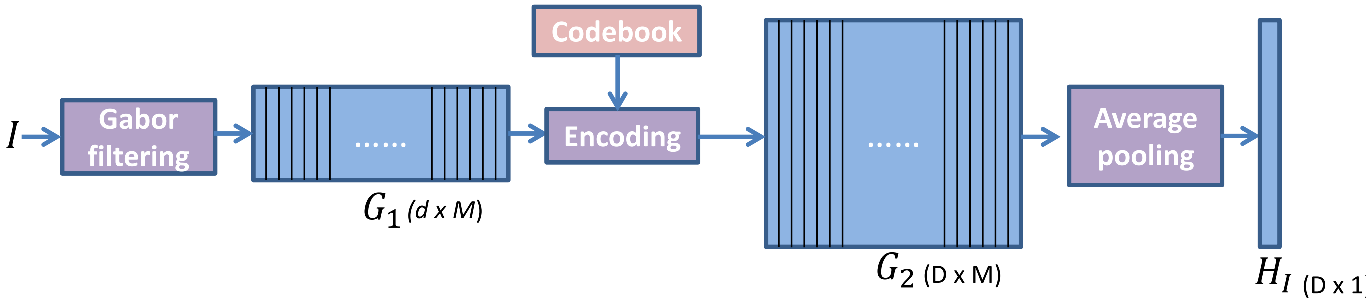
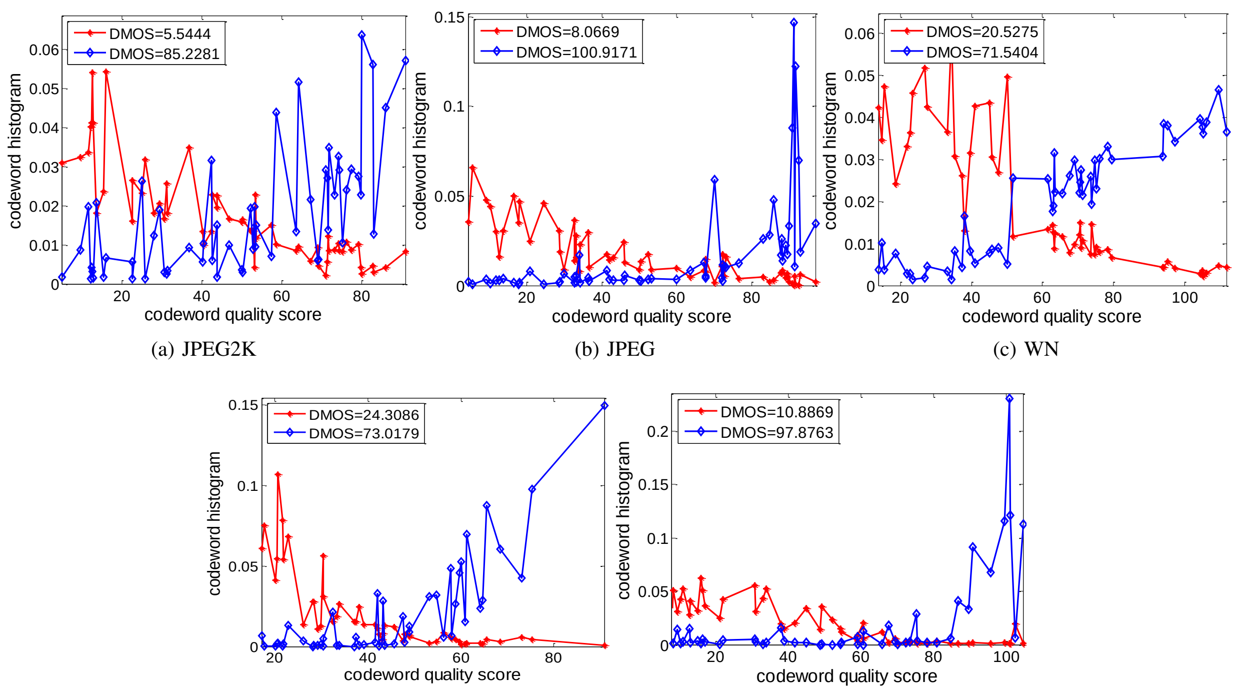


Fig. 5: Feature extraction. (Parameters: M:number of image patches, d:dimension of local Gabor feature, D: codebook size. Columns in G_1 and G_2 represent features extracted from different local image patches.)



(a) JPEG2K (b) JPEG (c) WN

(d) BLUR (e) FF

Fig. 6: Codeword histogram of images with different types and levels of distortions.

the CSIQ database to demonstrate database independence test, where we train on the LIVE database and test on the CSIQ database. We tested images with four types of distortions that are present in both the CSIQ database and the LIVE database - JPEG, JP2K, WN and BLUR.

Local Gabor feature extraction: Five filter frequencies (1, $1/\sqrt{2}$, $1/2$, $1/(2\sqrt{2})$, $1/4$) at four orientations ($0°$,$45°$,$90°$,$135°$) are used for extracting multi-scale features. The block size B used in our experiment is 11. The number of patches sampled from each image is 5000.

Codebook construction: For CBIQ-I, we extract 200 cluster centers from each training image and combine them to form a codebook. For CBIQ-II, in order to demonstrate that the codebook construction can be independent of the database, we construct a codebook of length 10,000 using distorted images with five types of distortions: WN, JPEG, JP2k, BLUR and additive pink Gaussian noise from CSIQ database. Due to memory problem, we could not directly perform k-means clustering on all the feature vectors from training set. Instead, we use hierarchy clustering by first running k-means on individual training images which results in 200 cluster centers for each training image and then performing k-means on the set of cluster centers from all training images. Thus the constructed codebook is optimal. For computing codeword histograms, the L_2 Euclidean distance is used in nearest neighbor search.

Cross validation: For cross-validation, we randomly split all images in the dataset according to their reference images into two parts - 23 (reference images and associated distorted images) for training and 6 for testing and then repeated the experiments. All results reported in this section are obtained after 1000 train-test iterations. We use the *nu*-SVR with linear kernel in libsvm package [34] for regression.

Evaluation: Two metrics [35] are used to evaluate the performance of the objective quality assessment model. The first metric is the Spearman rank-order correlation coefficient (SROCC) between predicted quality score and DMOS. It is

	JP2K	JPEG	WN	BLUR	FF	ALL
PSNR	0.872	0.885	0.941	0.764	0.875	0.867
SSIM	0.939	0.946	0.965	0.909	0.941	0.914
VIF	0.967	0.982	0.984	0.973	0.963	0.964
DIIVINE	0.913	0.910	0.984	0.921	0.863	0.916
BLIINDS-II	0.929	0.942	0.969	0.923	0.889	0.931
CBIQ-I	0.912	0.963	0.959	0.918	0.885	0.896
CBIQ-II	0.919	0.965	0.933	0.944	0.912	0.930

TABLE II: MEDIAN SPEARMAN CORRELATIONS (DMOS VERSUS PREDICTED DMOS, *Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

	JP2K	JPEG	WN	BLUR	FF	ALL
PSNR	0.873	0.874	0.928	0.774	0.869	0.855
SSIM	0.920	0.955	0.982	0.891	0.939	0.906
VIF	0.979	0.988	0.992	0.976	0.972	0.961
DIIVINE	0.922	0.921	0.988	0.923	0.888	0.917
BLIINDS-II	0.935	0.968	0.980	0.938	0.896	0.930
CBIQ-I	0.913	0.943	0.940	0.939	0.906	0.897
CBIQ-II	0.920	0.967	0.954	0.949	0.939	0.928

TABLE III: MEDIAN LINEAR CORRELATIONS (DMOS VERSUS PREDICTED DMOS, *Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

related to prediction monotonicity of a model. The second metric is the Pearson linear correlation coefficient (LCC) between predicted quality score and DMOS. For the LIVE database, we use the realigned DMOS scores [6] and report results only on the distorted images.

B. Performance evaluation

1) Experiments on the LIVE database: In this subsection, we compare the performance of CBIQ-I and CBIQ-II with two recent general purpose NR-IQA algorithms DIIVINE [18], BLIINDS-II [19] and three FR metrics[2] PSNR, SSIM [4] and VIF [7]. Results on the LIVE database are shown in Table II and Table III. The first five columns show results from distortion-specific (DS) experiments on different distortion subsets in the LIVE database. The objective of the DS experiment is to see how the algorithm will perform if we only have images with one particular type of distortion. The last columns in Tables II and III are obtained by performing train-test runs on images with all five types of distortions in LIVE. As can be seen, CBIQ-II outperforms PSNR and SSIM and is comparative to DIIVINE and BLIINDS-II on LIVE. The proposed method is slightly worse than the VIF, but VIF is a state-of-the-art full reference method while ours is no reference. We performed two sample T-test with 95% confidence level between SROCC generated by PSNR, SSIM and our algorithms CBIQ-I and CBIQ-II in 1000 iterations of experiments on the LIVE database. Test results are shown in Table IV. From this table, we can see that CBIQ-II is statistically superior to PSNR and SSIM, CBIQ-I is statistically superior to PSNR on LIVE.

To show the consistency of the performance of the proposed method, we computed the standard deviation of the SROCC and LCC obtained from the 1000 runs of experiments on the LIVE database and report results in Table V. The box plot of SROCC and LCC distributions of different quality measures from 1000 runs of experiments on the LIVE database are shown in Fig. 7. According to this result, performances of both CBIQ-I and CBIQ-II are more consistent than PSNR but less consistent than SSIM.

	PSNR	SSIM	CBIQ-I	CBIQ-II
PSNR	0	-1	-1	-1
SSIM	1	0	1	-1
CBIQ-I	1	-1	0	-1
CBIQ-II	1	1	1	0

TABLE IV: Results of the two sample T-test performed between SROCC values obtained by different measures. **1** (-**1**) indicates the algorithm in the row is statistically superior (inferior) than the algorithm in the column. 0 indicates the algorithm in the row is statistically equivalent to the algorithm in the column.

	PSNR	SSIM	BLIINDS-II	CBIQ-I	CBIQ-II
SROCC	0.0348	0.0159	0.0277	0.0317	0.0199
LCC	0.0326	0.0167	0.0252	0.0302	0.0215

TABLE V: Standard deviation of SROCC and LCC for 1000 iterations of experiments on the LIVE database.

2) Database independence experiment: Additionally, we tested CBIQ-I and CBIQ-II by performing training on the LIVE database and testing on the CSIQ database[3]. We use the CSIQ database for database independence test, because there is no overlap between images in the CSIQ database and images in the LIVE database. We only report results on the four distortions - JPEG2k, JPEG, WN and BLUR, which are present in both the LIVE database and the CSIQ database. In Table VI and VII, the first four columns show results from DS experiments on different distortion subsets in the LIVE database and CSIQ database. The last columns in Table VI and VII are obtained by performing train-test runs on images with all four types of distortions. To compare with DIIVINE and BLIINDS-II (with single scale) whose trained prediction models on the entire LIVE (including the FF distortion) are available online, we also train CBIQ-I and CBIQ-II on the entire LIVE and then test on the four distortions in CSIQ database. Results are shown in Table VIII. As can be seen, the proposed methods performs fairly well in the database independence test.

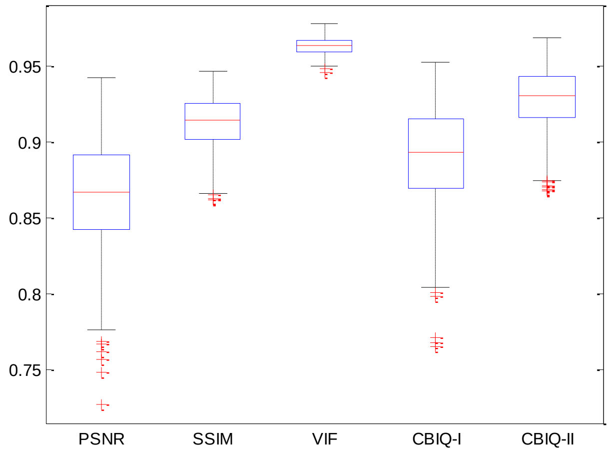
	JP2k	JPEG	WN	BLUR	ALL
SROCC	0.832	0.929	0.809	0.903	0.842
LCC	0.824	0.934	0.803	0.914	0.855

TABLE VI: Database independence test I: CBIQ-I was trained on different distortion subsets of LIVE database and tested on the CSIQ database.

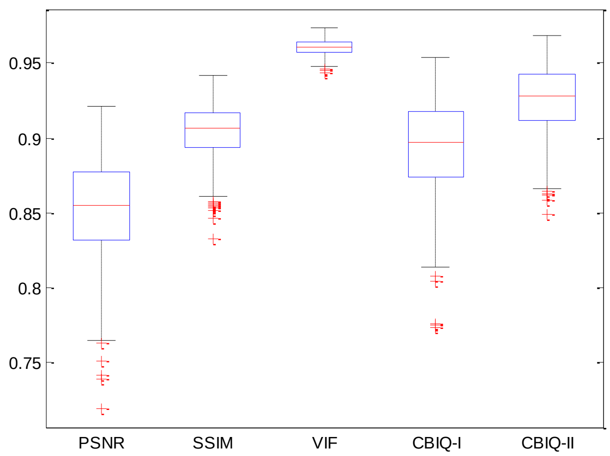
[2]A logistic nonlinear mapping as suggested in [6] was applied to FR measures before comparing them with DMOS.

[3]In previous section, we use codebook trained on CSIQ database for CBIQ-II. Since we are evaluating on CSIQ database now, we use codebook constructed from images in LIVE.

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.



(a) Spearman Correlation



(b) Pearson Correlation

Fig. 7: Box plot of SROCC and LCC distributions of NR-IQA algorithms from 1000 runs of experiments on the LIVE database. (On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles.)

	JP2k	JPEG	WN	BLUR	ALL
SROCC	0.850	0.942	0.828	0.896	0.884
LCC	0.873	0.953	0.807	0.925	0.908

TABLE VII: Database independence test I: CBIQ-II was trained on different distortion subsets of LIVE database and tested on the CSIQ database.

	DIIVINE	BLIINDS-II	CBIQ-I	CBIQ-II
SROCC	0.828	0.873	0.827	0.879
LCC	0.845	0.900	0.833	0.905

TABLE VIII: Database independence test II: Prediction models were trained on the entire LIVE database and tested on the CSIQ database.

3) Test on Lena image: To further verify the effectiveness of our method, we also test our method on images with distortions that have no corresponding examples in training set. Specifically, we use the 512×512 Lena image for testing and train on the first ten reference images with their degraded versions in LIVE database using CBIQ-I and CBIQ-II.

We consider distortions which have no examples in training set which include salt-pepper noise (noise density = D), speckle noise (multiplicative noise with mean 0 and variance V) and motion blur (linear motion of a camera by L pixels and with an angle of T degrees). As shown in Fig. 8 and 9, the CBIQ-II provides consistent prediction for all three types of distortions and CBIQ-I makes one mistake in ranking images with salt-and-pepper noise. This results show that for images with unknown distortions if unknown distortions share similar texture with some samples in our training set, for example, the motion blurring distortion looks similar to Gaussian blurring distortion; salt-pepper noise and speckle noise are somehow similar to White Gaussian noise, the proposed method can be potentially applied for IQA task.

C. Effect of Codebook Size

One attention is how the codebook size affects the performance of the proposed method, so we tested CBIQ-II using codebook with different sizes. Results are shown in Fig. 10. In general, performance improves as D increases. The fluctuation of the curve may be due to the fact that the codebook construction process is not well optimized. To achieve greater than 90% SROCC on each distortion subset and entire LIVE, a codebook with 2500 or more codewords is required when using current codebook construction scheme.

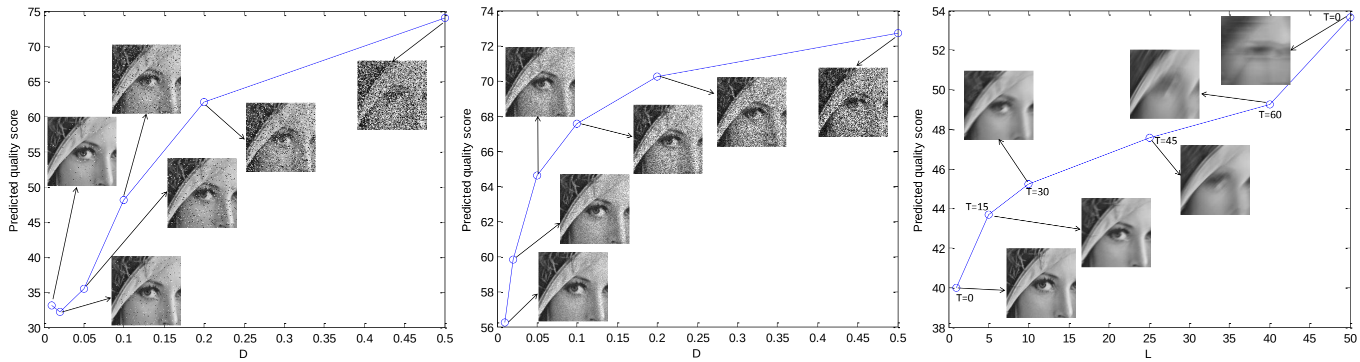
D. Discussion

Implementation Complexity: There are a number of issues related to the speed of our method that are worth noting. First, the most time-consuming part of this algorithm is the process of extracting Gabor filter based feature. We use the Multi-resolution Gabor Feature Toolbox [36] for feature extraction. It takes about 58 second to extract Gabor features on 5000 11-by-11 patches. The entire process for computing $DMOS_p$ for one 512×768 image using CBIQ-II with a codebook of size 10000 takes around 60 seconds[4] by running an un-optimized MATLAB program on a Intel Xeon 2.4GHz machine. Second, all these operations are block-based, i.e. they are performed independently on non-overlapping images blocks. Thus the proposed method has the potential to be applied in real-time application through parallel computing. Less computational expensive local feature extraction methods may be used to improve speed of the proposed method.

Modularity: The codebook based method is modular in that it can be extended to any number of distortions. To deal with a new type of distortion, we only need to add samples with this new type of distortion to our training set.

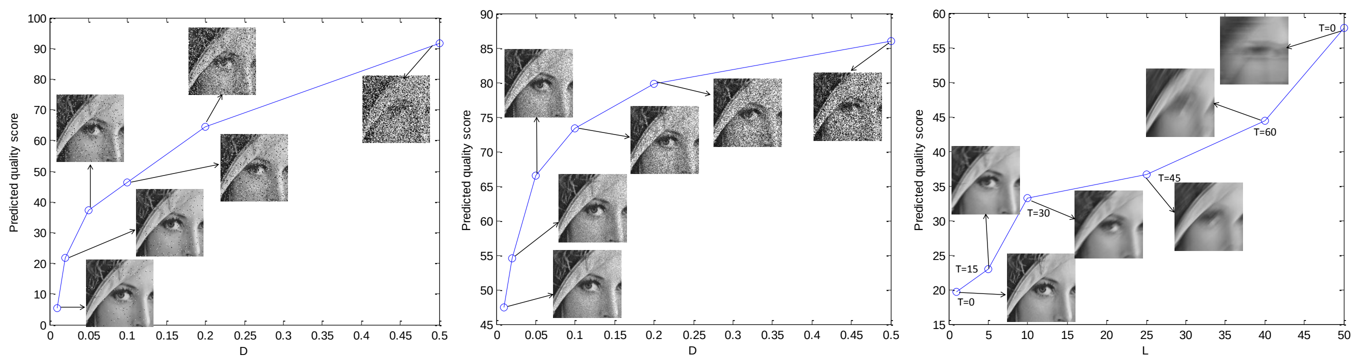
Future Improvement: First, the codebook construction process we have implemented is not fully optimized. Efficient quantization methods such as the Hierarchical K-Means [27] and the Extremely Randomized Forest [37] should be explored

[4]This does not include time for loading svm model and codebook.



(a) Salt-Pepper noise (b) Speckle noise (c) Motion Blurring

Fig. 8: Trend plots of Lena (cropped from 512×512 to 128×128 for visibility) with different types of distortions using CBIQ-I.



(a) Salt-Pepper noise (b) Speckle noise (c) Motion Blurring

Fig. 9: Trend plots of Lena (cropped from 512×512 to 128×128 for visibility) with different types of distortions using CBIQ-II.

in our future work. With an optimized codebook, the performance of the proposed method may be improved. Second, we use hard-assignment encoding and average pooling to obtain an image representation, more advanced encoding schemes, such as soft-assignment encoding or sparse coding and pooling schemes may be applied to optimize the proposed framework.

IV. Conclusions

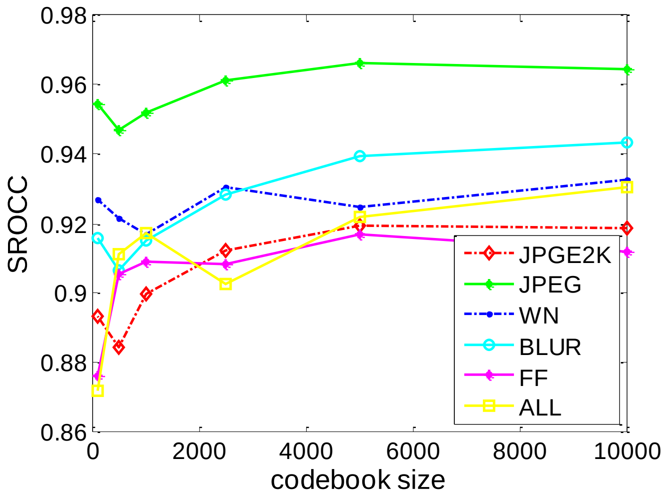
We have presented a simple and effective general purpose NR-IQA algorithm which estimates image quality without having a reference image and without any assumption on the types of distortion. When evaluating images with a particular type of distortion, the proposed method does require examples with the same or similar distortion for training. The proposed method is based on visual codebooks, which are a collection of Gabor features extracted from local image patches. Information from the training images are explored in the feature extraction stage. Our experimental results on the LIVE database show that the proposed method is comparable to state-of-the-art general purpose NR-IQA methods and outperforms the full-reference image quality metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

Acknowledgment

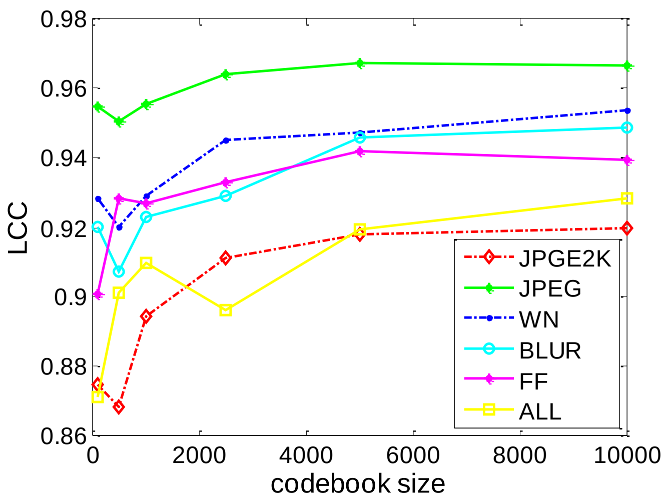
The partial support of this research by the US Government through NSF Award IIS-0812111 is gratefully acknowledged. The authors would like to thank the anonymous reviewers for their helpful and constructive comments on this paper.

References

[1] Y. Ke, X. Tang, and F. Jing, “The design of high-level features for photo quality assessment,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 419–426.
[2] C. Li and T. Chen, “Aesthetic visual quality assessment of paintings,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 236–253, Apr. 2009.
[3] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool, 2006.
[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[5] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
[6] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
[7] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.



(a) SROCC



(b) LCC

Fig. 10: Effect of different codebook sizes for CBIQ-II.

[8] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing-image Communication*, vol. 25, pp. 517–526, 2010.

[9] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[10] D. Tao, X. Li, W. Lu, and X. Gao, "Reduced-reference IQA in contourlet domain," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 6, pp. 1623–1627, Dec. 2009.

[11] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *IEEE International Conference on Image Processing*, vol. 1, Rochester, New York, Sep. 2002, pp. I–477 – I–480.

[12] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *J. Electron. Imaging*, vol. 14, 033004, Aug. 2005.

[13] T. Brandão and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Processing*, vol. 88, pp. 822–833, Apr. 2008.

[14] 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.

[15] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

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

[17] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

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

[19] M. A. Saad, A. C. Bovik, and C. Charrier, "Model-based blind image quality assessment using natural DCT statistics," *IEEE Transactions on Image Processing*, 2011 (to appear).

[20] W. Lu, K. Zeng, D. Tao, Y. Yuan, and X. Gao, "No-reference image quality assessment in contourlet domain," *Neurocomput.*, vol. 73, pp. 784–794, Jan. 2010.

[21] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment: The natural scene statistic model approach," *IEEE Signal Processing Magazine*, vol. 28, 2011.

[22] H. Tong, M. Li, H. Zhang, C. Zhang, J. He, and W. Ma, "Learning no-reference quality metric by examples," in *Proceedings of the 11th International Multimedia Modelling Conference*, Jan. 2005, pp. 247–254.

[23] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 305–312.

[24] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, May 2011.

[25] P. Ye and D. Doermann, "No-reference image quality assessment based on visual codebook," in *IEEE International Conference on Image Processing*, 2011, pp. 3150–3153.

[26] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *International Conference on Computer Vision*, 2005, pp. 604–610.

[27] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 2161–2168.

[28] A. Bovik, M. Clark, and W. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 55 –73, Jan. 1990.

[29] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[30] P. Kruizinga, N. Petkov, and S. Grigorescu, "Comparison of texture features based on Gabor filters," in *International Conference on Image Analysis and Processing*, 1999, pp. 142 –147.

[31] V. Kyrki and J. Kamarainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25, pp. 311–318, 2004.

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

[33] 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, 011006, 2010.

[34] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, May 2011.

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

[36] J. Ilonen, J.-K. Kamarainen, and H. Kä lviä inen, "Efficient computation of Gabor features," Research Report 100, Lappeenranta University of Technology, Department of Information Technology, 2005.

[37] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1632–1646, Sep. 2008.

Peng Ye received the B.Eng. degree in the Department of Electronic Engineering at the Tsinghua University, China, in 2007 and a M.Eng. degree in Electrical Engineering from the University of Delaware, in 2009. She is pursuing the Ph.D. degree in Electrical Engineering at the University of Maryland, College Park, where she is a Research Assistant in the Laboratory for Language and Media Processing in the University of Maryland's Institute for Advanced Computer Studies. Her area of research is image processing and computer vision, with a focus on quality assessment of scene and document image.

David Doermann is a senior research scientist in UMIACS. He received a B.Sc. degree in Computer Science and Mathematics from Bloomsburg University in 1987, and a M.Sc. degree in 1989 in the Department of Computer Science at the University of Maryland, College Park. He continued his studies in the Computer Vision Laboratory, where he earned a Ph.D. 1993. Since 1993, he has served as co-director of the Laboratory for Language and Media Processing in the University of Maryland's Institute for Advanced Computer Studies and as an adjunct member of the graduate faculty. His team of researchers focuses on topics related to document image analysis and multimedia information processing. Recent intelligent document image analysis projects include page decomposition, structural analysis and classification, page segmentation, logo recognition, document image compression, duplicate document image detection, image based retrieval, character recognition, generation of synthetic OCR data, and signature verification. In video processing, projects have centered on the segmentation of compressed domain video sequences, structural representation and classification of video, detection of reformatted video sequences and the performance evaluation of automated video analysis algorithms. In 2002 he received an Honorary Doctorate of Technology Sciences from the University of Oulu for his contributions to digital media processing and document analysis research. He is a founding co-editor of the International Journal on Document Analysis and Recognition, has the General Chair or Co-Chair of over a half dozen international conferences and workshops and is the General Chair of the International Conference on Document Analysis and Recognition (ICDAR) to be held in Washington DC in 2013. He has over 30 journal publications and over 125 refereed conference papers.