

Hybrid No-Reference Natural Image Quality Assessment of Noisy, Blurry, JPEG2000, and JPEG Images

Ji Shen, Qin Li, and Gordon Erlebacher

Abstract—In this paper, we propose a new image quality assessment method based on a hybrid of curvelet, wavelet, and cosine transforms called hybrid no-reference (HNR) model. From the properties of natural scene statistics, the peak coordinates of the transformed coefficient histogram of filtered natural images occupy well-defined clusters in peak coordinate space, which makes NR possible. Compared to other methods, HNR has three benefits: 1) It is an NR method applicable to arbitrary images without compromising the prediction accuracy of full-reference methods; 2) as far as we know, it is the only general NR method well suited for four types of filters: noise, blur, JPEG2000, and JPEG compression; and 3) it can classify the filter types of the image and predict filter levels even when the image is results from the application of two different filters. We tested HNR on very intensive video image database (our image library) and Laboratory for Image & Video Engineering (a public library). Results are compared to the state-of-the-art methods including peak SNR, structural similarity, visual information fidelity, and so on.

Index Terms—Blur, curvelet, discrete cosine transform (DCT), image quality assessment (IQA), JPEG, JPEG2000, log probability density function (pdf), natural scene statistics (NSS), no reference (NR), noise, wavelet.

I. INTRODUCTION

NOW that digital cameras and the Internet have both become popular media, millions of photos are taken daily and a substantial portion are posted online. They are stored on computers, web sites, cameras, even in cell phones, often through the intermediary of backend databases. Digital photos have large variance of quality as a result of the various distortions they undergo. The different distortions an image applied include picture shooting, image compression, transmission, and postprocessing. For example, when taking a photo using a digital camera, incorrect focus, low-quality lens, or camera shake create blur image, even in a high-quality camera. Long shutter exposure or high ISO speed (with higher electric current) expo-

sure increases the noise contamination of an image [1]. Lossy compression is another cause of quality degradation. In order to save storage, image data are often subjected to a lossy compression algorithm, with a quality level determined by a tradeoff between image quality and data size. Finally, many images are enhanced by image processing software, including denoising, deblurring, superresolution, etc. In the three scenarios above, the image is modified homogeneously, i.e., the quality degradation is statistically uniform across the image (compression may have different compression ratio if the image has different complexity level at different regions but the difference is small). There are other cases that are more complex. For example, photographs with shallow depth of field have blur that is spatially dependent. Alternatively, certain types of image enhancement might be localized, for example, edge enhancement. Any effect that modifies an original image affects its quality.

How to quantify the image quality? This research topic is referred to as image quality assessment (IQA). Such algorithms can help compare the quality of different cameras, evaluate enhancement algorithms by giving quality scores to images before and after processing, and benchmark image compression algorithms. Most IQA methods compare the distorted image to the original version. However, in every case, image quality is affected, and the original unmodified image is often no longer available (if it ever was). As a result, image quality estimation through deterministic and statistical algorithms has become increasingly important.

Since image quality relates to human perception, it is natural to assign subjective scores evaluated by human subjects to each image. However, compiling subjective scores is time consuming and large variances among human subjects are inevitable. On the other hand, we can evaluate image quality by objective metrics, i.e., image quality metric (IQM). The goal of IQA is to develop an objective IQM, whose scores have a consistent correlation with subjective scores. IQM quantifies the image quality by analyzing various image features, including noise level, blur level, compression level, information loss, and contrast difference.

Image quality is a characteristic of an image that can be perceived as image degradation with respect to some ideal. The image quality distortion is usually compared to the “ideal” or “perfect” image that refers to an image taken with an idealized high-quality digital camera without noise or blur, and prior to any eventual postprocessing. However, such a perfect quality image does not actually exist. Thus, in practice, we use an “original” image as a proxy of the ideal image, which refers to a high-quality image without compression or postprocessing.

Manuscript received April 13, 2010; revised August 21, 2010, and January 06, 2011; accepted January 10, 2011. Date of publication January 28, 2011; date of current version July 15, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler.

J. Shen and Q. Li are with the Department of Mathematics, Florida State University, Tallahassee, FL 32310 USA (e-mail: js04j@fsu.edu; qli@math.fsu.edu).

G. Erlebacher is with Visualization Laboratory, Department of Scientific Computing, Florida State University, Tallahassee, FL 32310 USA (e-mail: gerlebacher@fsu.edu).

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

Digital Object Identifier 10.1109/TIP.2011.2108661

With access to the original image, the distortion levels can be assessed by direct comparison between distorted and original images. Such metrics are called full- or reduced-reference (RR) IQM. They can achieve a high correlation with subjective quality scores. However, the original image required by the algorithm is often not available. This would be the case for Google images, or JPEG images taken by digital cameras. On the other hand, no-reference (NR) methods use alternative approaches without access to the original images. Such an approach, if successful, is more widely applicable, although more difficult to develop.

A real world image usually undergoes several kinds of distortions, such as a noisy and blurred JPEG image. However, as images that have undergone several types of distortions are particularly hard to analyze, most IQMs can only assess the quality when the image is assumed to have a single kind of distortion, i.e., either a noisy, a blurred, a JPEG, or a JPEG2000 image. A full-reference (FR) method usually works for all these four filter types, while an NR method usually works for only one out of the four filter types. Our proposed method is an NR general IQM that has been tested for four types of filters.

In this paper, when used alone, the word “filter” denotes Gaussian noise, Gaussian blur, JPEG2000, or JPEG compression. The filter level (FL) denotes the filter setting, i.e., the variance for Gaussian noise, the variance for Gaussian blur, bits per pixel (BPP) for JPEG2000, or the compression quality threshold for JPEG.

The remainder of this paper is organized as follows. In Section II, we introduce the different IQMs, natural scene statistics (NSS), and the curvelet transform. In Sections III and IV, we propose the hybrid NR (HNR) framework and our very intensive video image database (VIVID) image library. A special case curvelet NR (CNR) model [2] is first introduced, which is then generalized to HNR. Section V tests the performance of HNR on both VIVID and Laboratory for Image & Video Engineering (LIVE). Results are compared against several state-of-the-art FR, RR, and NR methods.

II. RELATED WORK

A. Subjective and Objective IQA

There are two types of image quality scores, classified as subjective and objective scores. Subjective scores are evaluated via human visual perception. For example, consider the five images displayed in Fig. 1; very likely, the reader can produce his own estimate of image quality, i.e., the subjective score, because the human visual system is a complex mechanism capable of distinguishing many subtle features within an image. On the other hand, objective IQMs do not involve human subjects, but analyze directly the distorted image content via well-defined algorithms, either by comparison with the original image or through examination of the specific distortion characteristics.

There are also two types of validations of objective IQM. One is to get the correlation between IQM scores and the exact FLs. The other is to get the correlation between IQM scores and the subjective scores evaluated by humans. The FLs are precise and known when creating the images, but might not correlate well with the real image quality. On the other hand, subjective scores



Fig. 1. Images 1–5 are the original, noisy (variance of 0.1), blurred (variance of 2.8), JPEG2000 (BPP of 0.1), and JPEG (quality threshold of 5) compressed images, respectively.

are ultimately the image quality of most interest but hard to compute accurately. For example, the subjective scores in the public image library LIVE [3] are called difference mean opinion score (DMOS) in the range of $[0, 100]$, but on an average, DMOS of each image has a standard deviation of more than 10 [4] across different humans. This is due to the difficulty for human subjects to produce a reliable image quality score.

Since objective and subjective scores both have their advantages and disadvantages, we test our algorithm with both of them. In our paper, we build our own objective image library VIVID to do the objective tests. In addition, we use LIVE to do the subjective tests and to compare our performance with other methods.

VIVID has 65 original images, each filtered by noise, blur, JPEG2000, and JPEG filters for 101 different FLs, resulting in 26 260 images in total. The subjective image library LIVE was created to calibrate objective methods to the subjective scores. It contains 779 images. There are 29 original images, each filtered by 5 types of filters with an average of 6 different FLs. The filters used include noise, blur, JPEG2000, JPEG, and JPEG2000 transmission error. LIVE provides both objective scores (FL) and subjective scores (DMOS) for each image.

B. Three Categories of IQA

IQMs are categorized as FR, RR, and NR, depending on the availability and use of the original image.

1) *Full- and Reduced-Reference*: The majority of existing IQMs are FR methods, which compare the original and distorted images at the pixel level, and quantify the difference into a quality score. Traditional FR algorithms, such as peak SNR (PSNR) and SNR, are based on the mean square error (MSE) between original and distorted images. The recent information fidelity criterion (IFC) [5] and visual information fidelity (VIF) [6] quantify the information loss intrinsic to the distortion. The structural similarity method measures the difference of luminance, contrast, and structure [7].

RR algorithms, on the other hand, only quantify the difference between some features extracted from the original and distorted images. If the feature has a small data size, it is sometimes embedded into the original image. Wang [8] discusses the quality aware image format, which embeds the wavelet histogram information of the original image into the image prior to distortion. For moderate distortion, the embedded data are unaffected. Once the embedded data can no longer be retrieved reliably, the method fails.

Both FR or RR algorithms require access to the original image during quality estimation. Of course, this is often not possible. For example, most of the images found on the Internet are in compressed format. Alternatively, images exported from

an average camera have inevitable built-in noise, blur, and compression. In both scenarios, the originals were either lost or never existed.

2) *No Reference*: NR methods dispense with the original image, alleviating the most serious drawback of the FR and RR approaches. However, good NR methods are quite difficult to develop, and most methods only work for a single type of filter [9].

The first category of NR IQM is based on the analysis of image content at the pixel level. Li [10] analyzes the difference between pixel intensities or local smoothness to predict the noise level. Marziliano *et al.* [11] analyzes the edge sharpness to predict the blur level. Wang *et al.* [12] measures the blocky artifacts at the edges of 8-by-8 blocks to predict JPEG quality. Sazzad *et al.* [13] measures the pixel distortion and edge data to predict JPEG2000 quality.

The second category of NR draws on the statistical properties of the natural images. Natural images [14] are images of the natural world taken from high-quality capture devices in the visual spectrum. This definition excludes computer-generated graphics, paintings, X-rays, random noise, etc. Previous study [15] reveals that the natural images are a very small part of the huge space of all possible images, and that they share similar statistical characteristics, i.e., the so-called NSS.

Research [16], [17] confirms that if one transforms natural images into some other space, such as discrete cosine transform (DCT) or wavelet space, the histograms of the basis/dictionary coefficients have qualitative similarities, which can usually be fit by a generalized Gaussian distribution. If the statistical properties of the histograms of the natural images are known, it is possible to infer a quantitative quality score for a distorted image.

For example, Brando and Queluz [18] used the Laplace probability density function to fit the histogram of DCT coefficients of JPEG images. The parameter λ of this Laplace probability density function (pdf) correlates well with the degree of blockiness, characteristic of the JPEG images. Sheikh *et al.* [19] computes the last two scales of horizontal, vertical, and diagonal wavelet coefficients. The joint histogram of these wavelet coefficients have good correlation with JPEG2000 quality.

Very recently, two NR approaches BLIINDS [20] and BIQI [21] based on NSS were developed as general frameworks for various filters. Saad *et al.* [20] converts the images to DCT space and extracts the feature vector from DCT coefficients. After a training process, the model can predict the quality of images processed by a variety of filters. Our proposed method HNR also belongs to this category, and we apply it successfully to the following four types of filters: noise, blur, JPEG2000, and JPEG.

C. Brief Introduction to the Curvelet Transform

The curvelet transform is at the basis of HNR. In this section, we briefly introduce curvelets, emphasizing their unique properties in representing curved singularities, which justifies their use for IQA.

Curvelets form an overcomplete dictionary (redundancy ≈ 3.6) with the property that they are nearly optimal for representing objects with C^2 singularities [22], [23]. Thus, edges within images have a sparse representation in curvelet space.

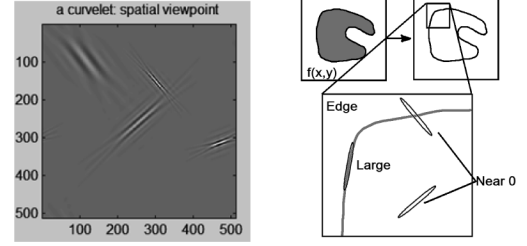


Fig. 2. Left: four curvelets in the spatial domain with different scales, angles, and locations. The figure is generated using CurveLab 2.0 [23]. Right: curvelets aligned with the curved singularity have large coefficients, otherwise their coefficients are close to zero.

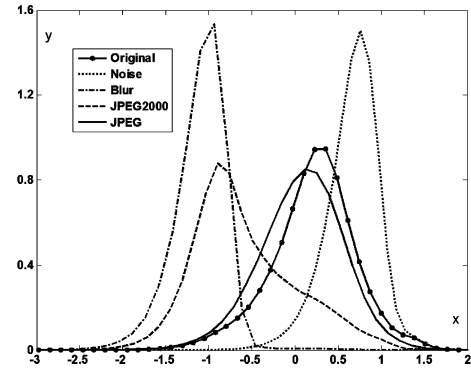


Fig. 3. F1-LPMCC of the images in Fig. 1.

Each curvelet [see Fig. 2(a)] has an approximately elliptical support, smooth along the major axis, which lies along the discontinuity, and oscillatory along the minor axis. There are three main attributes of 2-D curvelets, not shared by wavelets: 1) they are intrinsically 2-D (they have no tensor product representation); 2) they follow a parabolic scaling: curvelets of smaller scale contract as s normal to the discontinuity and as s^2 along it; and 3) they are parameterized not only by position and scale, but also by orientation.

Each coefficient in the curvelet expansion of an image is the result of the convolution of the associated curvelet and the image. If a curvelet of given scale, angle, and location is approximately aligned along some curve [see Fig. 2(b)], its curvelet coefficient is large, otherwise it is close to zero. Due to its needle-like support and its range of orientations, a very few curvelets are sufficient to approximate the curved singularities within images, i.e., the representation is sparse. Curvelet coefficients can also be large if the curvelet is small scale and the location is centered on point singularities. Because changes in noise or blur levels affect the properties of curved or point singularities within the image, the corresponding large curvelet coefficients will be strongly affected.

The discrete curvelet transform of a 2-D function $f[t_1, t_2]$ is defined as follows:

$$\theta(j, l, k) = \sum_{0 \leq t_1, t_2 < n} f[t_1, t_2] \overline{\varphi_{j,l,k}[t_1, t_2]} \quad (1)$$

where $\varphi_{j,l,k}$ is a curvelet of scale j at position index k with angle index l . The indices t_1 and t_2 denote coordinates in the physical domain [23].

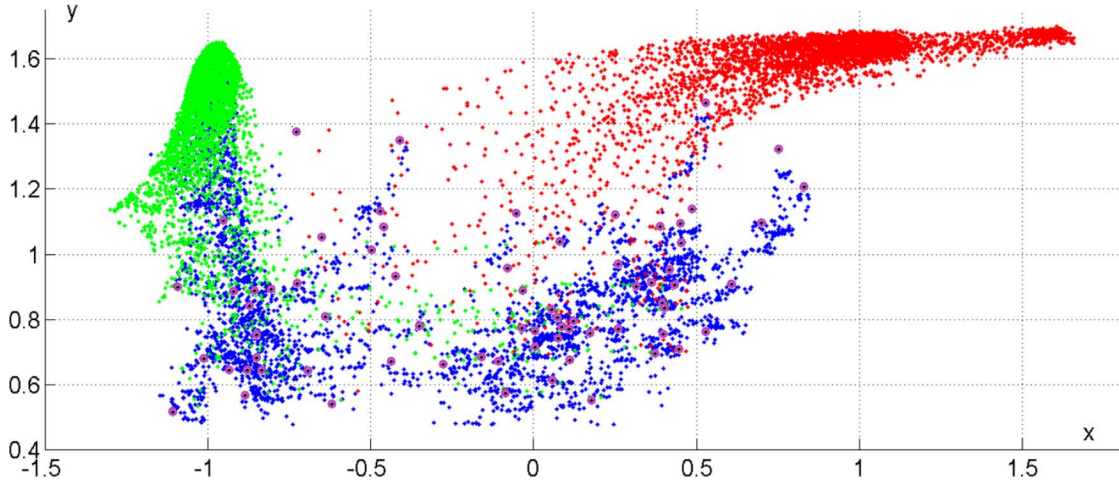


Fig. 4. Distribution of $\mathbf{c}_1 = (x_1, y_1)$ (the ICs of F1-LPMCC) of about 14 140 images. Red: noisy; Green: blurry; Blue: JPEG2000; and Magenta: original images.

III. CNR FRAMEWORK

A. Image Characteristic

The primary factor that explains why CNR works as an NR method is the particular choice of the characteristics extracted from the coefficients of the transformed images. An ideal characteristic (referred to as image characteristic or IC) should only depend on the FL. Independence of the IC from the image content is a key requirement. Moreover, the FL should be a continuous function of the IC. We choose to use the peak coordinate of LPMCC as our IC.

We define the LPMCC as the probability distribution of the logarithm (base 10) of the magnitude of the curvelet coefficients, i.e., the log-pdf of the magnitude of curvelet coefficients. Since curvelets have multiple scales, we also consider the LPMCC on a scale-by-scale basis. Thus, for scale j , the operator T_j calculates the pdf $h_j(x)$ for an image I

$$\begin{cases} h_j(x) \triangleq T_j(I) = \text{pdf}(\log_{10}(|\Theta_j|)) \\ \Theta_j = \{\theta(j, l, k)\} \end{cases} \quad (2)$$

where $h_j(x)$ is the probability density of $\log_{10}(|\Theta_j|)$, and $\theta(j, l, k)$ is the curvelet coefficient [see (1)]. The LPMCC that corresponds to the J th to finest scale curvelets is called FJ -LPMCC. Suppose curvelets have j_{\max} number of scales, then $j = j_{\max} + 1 - J$. Therefore, F1-LPMCC, F2-LPMCC, and F3-LPMCC are the LPMCC constructed from the finest, second to finest, and third to finest scale curvelet coefficients, respectively. Fig. 3 displays the five F1-LPMCCs of the images in Fig. 1. The four filters clearly affect the distributions in different ways. The x and y coordinates of the peak point in the LPMCC serve to characterize the entire image where the peak point corresponds to the global maximum of the pdf. Our experiment demonstrates that for each of 26 260 images, the LPMCC has a global maximum and no other local maximums. The pair of coordinates (x, y) define the IC. We denote IC for each scale as \mathbf{c}_J , and the global maximum operation as operator P . Therefore,

$$\mathbf{c}_J = (x_J, y_J) = P(h_j(x)) = \begin{cases} \arg \max_x h_j(x) \\ \max_x h_j(x) \end{cases} \quad (3)$$

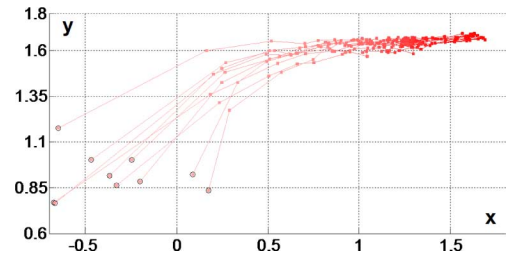


Fig. 5. IC distribution of 220 noisy images from F1-LPMCC. The color intensity of the dots increases with the noise level.

i.e., \mathbf{c}_J is the peak coordinates of the pdf $h_j(x)$, where $j = j_{\max} + 1 - J$.

Henceforth, to inspect the distribution of the IC, we use 14 140 images from VIVID, and represent each IC as a dot. From Fig. 4, it is apparent that the ICs belonging to each filter occupy a distinctive region. The noisy (red dots) and blurred (green dots) images are almost separated into two nonoverlapping regions. JPEG2000 (blue) and blurred (green) images are strongly correlated. The original images (larger magenta dots) occupy the central region of the figure.

Next, we consider whether there is a continuous relation between IC and FL. To this end, consider Fig. 5, where each dot represents the IC of one of the 220 noisy images. Out of these, ten are ICs of original images. Starting from an original image, a noise filter is applied with increasing noise levels. Successive ICs are joined by a dotted line. Thus, each dotted line in the figure corresponds to transformed images from a common original image.

The IC distribution of these 220 noisy images displays the following characteristics of NSS. First, both coordinates of IC can be modeled by quasi-monotonic continuous functions of the noise level, i.e., the mapping from IC to FL is a surjection. Second, all the ICs that correspond to a fixed value of the FL, across images, define local clusters. The cluster changes continuously with FL, which implies that the IC cluster is a function of FL but independent of the image content.

The continuous relation between IC and FL makes training-based NR method possible since similar ICs relate to similar

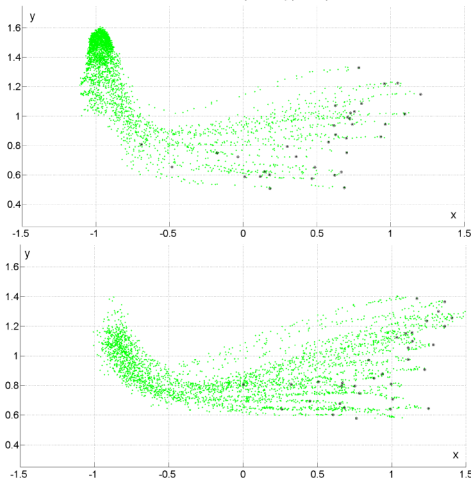


Fig. 6. IC distribution of F2-LPMCC and F3-LPMCC of blurry images.

FLs. Thus, if one considers a set of images with ICs in the neighborhood of an image under analysis, an approximation scheme can be devised to estimate the FL.

A small modification must be made for the blur filter, however. Indeed, past a certain FL threshold, the continuous relation between IC and FL breaks down when only considering the coefficients from F1-LPMCC. Beyond a critical blur level, the IC becomes almost stationary (see Fig. 4). In other words, F1-LPMCC is no longer affected by the blur filter once the image content at the finest scale has been filtered away. However, as demonstrated in Fig. 6, compared to F1-LPMCC, the ICs of the F2-LPMCC and F3-LPMCC are much more sensitive to the FL at high levels of blur. It is therefore useful to reinterpret the IC as a 6-D vector \mathbf{c} , which simply concatenates the ICs from each of three finest scales

$$\mathbf{c} = (x_1, y_1, x_2, y_2, x_3, y_3) \quad (4)$$

where (x_J, y_J) is the IC of the FJ -LPMCC.

B. Training, Prediction, and Error Estimation

During the training phase, the algorithm catalogs thousands of images with known quality scores. During the prediction phase, the algorithm compares an arbitrary image to similar images from training set and derives an estimate of the image quality.

The CNR model has two sets of images: the set S_T for training and the set S_P for prediction. S_T contains a combination of original images and their corresponding filtered versions, while S_P contains the remaining images. In our general framework, S_T consists of N_O “original” images filtered by one of N_F filters F_i ($i = 1, \dots, N_F$), each filter set to N_{F_i} FLs. We represent an individual image in this collection by I_{pik} (p th original image ($p = 1, \dots, N_O$), i th filter, and k th FL [$k = 1, \dots, N_{F_i}$]). The total number of images is $|S_T| = N_O \sum_{i=1}^{N_F} N_{F_i}$. Altogether, the images define N_F sets S_i that in turn contain $N_O N_{F_i}$ images processed by filter F_i .

During training, we extract the ICs for all images in S_i and form the pairs $\{(\mathbf{c}_i, l_i)\}$ between IC and FL. Altogether, the N_F mappings

$$l_i = f_i(\mathbf{c}_i) \quad i = 1, \dots, N_F \quad (5)$$

are established during the training phase. We will apply this framework to the following four filters: noise, blur, JPEG2000, and JPEG.

In the prediction phase, the images in set S_P are labeled identically to the images in S_T . For each image I^* in S_P with exact filter F_{i^*} and FL l^* , we predict its filter type \tilde{i} and FL \tilde{l} . The goal is that the estimates equal the exact values

$$\begin{cases} \tilde{i} = i^* \\ \tilde{l} = l^* \end{cases} \quad (6)$$

where “exact FL” of an altered image refers to the FL used to produce the altered image.

We wish to derive a formula that estimates the FL of the new image I^* as a function of the sample images in S_{i^*} , with stronger weighting for sample images “closer” to I^* in IC space. A simple Euclidean distance serves as a convenient means to measure the distance between two images I and I' with IC of \mathbf{c} and \mathbf{c}'

$$d(\mathbf{c}, \mathbf{c}') = \|\mathbf{c} - \mathbf{c}'\| = \sqrt{\sum_{i=1}^n (c_i - c'_i)^2}. \quad (7)$$

Next, we define an exponentially decreasing weight (with respect to distance)

$$w_i(d) = e^{-a_i \cdot \|d\|} \quad (8)$$

where the parameter a_i controls its rate of decay. We choose a_i to maximize the linear correlation coefficient (CC) (see Section V-A) between the set of computed and exact FLs using images different from the prediction set.

If we assume that I^* is filtered by filter F_i , the approximate FL \tilde{l}_i is a linear combination of the FLs of all the images I in S_i

$$\tilde{l}_i = \tilde{f}_i(\mathbf{c}^*) = \frac{\sum_{I \in S_i} f_i(\mathbf{c}_i) \cdot w_i(\|\mathbf{c}_i - \mathbf{c}^*\|)}{\sum_{I \in S_i} w_i(\|\mathbf{c}_i - \mathbf{c}^*\|)} \quad (9)$$

where \mathbf{c}^* is the IC of the new testing image I^* , and $f_i(\mathbf{c}_i)$ is the FL of the image I associated with \mathbf{c}_i from the training set in (5).

We use the weighted approximation for each of the N_F filters to generate FL predictions $\tilde{l}_i = \tilde{f}_i(\mathbf{c}^*)$, where $i \in [1, N_F]$, i.e., we predict the FL for each filter. Since I^* is actually only filtered by F_{i^*} , we expect $\tilde{l}_i = \delta_{i,i^*} l^*$, where $\delta_{i,j}$ is the Kronecker delta function. In reality, this ideal situation rarely occurs as the images processed with different filters always have some degrees of correlation, which leads to $\tilde{l}_i \neq 0$ but near zero when $i \neq i^*$. Thus, we take the FL prediction as the largest \tilde{l}_i over i , and the filter-type prediction as the associated i

$$\begin{cases} \tilde{i} = \arg \max_i \{\tilde{l}_i\} \\ \tilde{l} = \max_i \{\tilde{l}_i\} \end{cases} \quad (10)$$

The FL prediction has an absolute error

$$e = \left| \tilde{l} - l^* \right|. \quad (11)$$

Note that since most related IQM does not classify the filter type, in all the following experiments that measure the accuracy of FL predictions, we assume that the filter type F_{i^*} of the image is already known. Therefore, in (10), we assume $\tilde{i} = i^*$ is known, i.e.,

$$\tilde{l} = \tilde{l}_{i^*}. \quad (12)$$

C. Alternative Training, Predictions, and Error Estimation

Previous research [24] has found that the FL is not always the most reliable means of estimating image quality. An alternative approach lies in the prediction of subjective scores of image quality. The framework established in the previous paragraph remains valid. Although the range of values might change (for example, DMOS $\in [0, 100]$ while FL $\in [0, 1]$), since (9) is a linear function with respect to l , we can replace all the FLs (l) by subjective scores (d) and compute estimates of the subjective scores, independent of any specified range.

D. Reasons Behind Our Decisions

CNR can predict noise and blur levels rather well when tested on our image library (see Section III), with accuracy on par with the best FR methods. This is a consequence of some of our modeling decisions, including the use of curvelets, log-pdf, IC, and multiscale coefficients.

In our model, we plot the log-pdf of the curvelet coefficients of the natural images instead of the pdf used in many NSS-based IQM. The peak of the log-pdf corresponds to the larger values of the coefficients, rather than the smaller ones in the case of the pdf. These larger curvelet coefficients identify the curved singularities in the image, which are sensitive to the image distortions.

The peak coordinates of LPMCC, used as IC of each image, are sensitive to the FL of the image. In Fig. 3, the x coordinate of each peak coordinate is approximately the mean of the logarithm of the magnitude of the curvelet coefficients, while the y coordinate is proportional to the probability of this mean value. High-frequency components, such as noise, are well captured by curvelets, and hence, the related coefficients have larger magnitude. Thus, the peak coordinate shifts right, as shown in Fig. 3. Smoothing (blurring and compression) the edges of the image reduces the strength of curved singularities, which decreases the magnitude of related curvelet coefficients and shifts the peak coordinate to the left.

Since curvelets have higher angular discrimination than wavelets, they are more sensitive to distortions oblique edges. In addition, curvelets have a redundancy factor of about 3.6. The additional coefficients increase the sparsity of the coefficients and also improve the smoothness and stability of the log-pdf. To our knowledge, CNR is the first IQM using the curvelet transform.

IV. HNR FRAMEWORK

The CNR model performs well for noisy and blurred images but not for compressed images. Therefore, we explore new types of transforms to better predict the quality of compressed images.

A. Wavelet NR Model

Wavelet NR (WNR) is designed to improve the prediction of JPEG2000 images. We use the Cohen–Daubechies–Feauveau 9/7 wavelet transform, adopted for JPEG2000, to replace the curvelet transform. The IC is the first peak of log-pdf of the magnitude of wavelet coefficients. Except for some differences on computing IC, WNR and CNR are identical.

B. DCTNR Model

Similar to what was done in WNR, the DCTNR replaces the curvelet transform with DCT to better predict JPEG quality. Following the JPEG compression algorithm, we apply DCT to each 8-by-8 block of the image followed by quantization of the DCT coefficients. We compute the log-pdf of the magnitude of DCT coefficients (LPMDC) and extract the IC as the local maximum of LPMDC, which now has several peak values. We choose the second peak from the left of the pdf due to its stronger sensitivity to the blocky artifacts of JPEG. Note that DCTNR results in a 2-D IC.

C. Hybrid Transform No-Reference Model

We found that when testing LIVE images, neither CNR nor WNR is the best choice. Choosing other kinds of wavelets or adding more dimensions to the IC sometimes increases the prediction accuracy. We therefore consider yet another revised model, in which we use an n -dimensional ($n = 6k, k \in \mathbb{N}^+$) IC instead of the 6-D IC used in CNR, WNR, or DCTNR. The $6k$ -dimensional IC is a simple concatenation of k vectors of 6-D IC computed from k types of transforms. We choose the combination of the transforms, which has the best prediction accuracy for a certain filter to form hybrid transform NR (HTNR).

D. Hybrid No Reference

To achieve the best correlation with image quality, we determine the combination of transforms used in the training and prediction phases of HNR for each filter. In HNR, for each filter F_i , after computing the coefficients of transform t_i , we draw the log-pdf of the magnitude of the transformed coefficients (LPMTC). We define the k_i^{th} maximum value (if there is only a single maximum, $k_i = 1$) as the IC for each single transform: $k_i = 1$ for curvelets or wavelets, and $k_i = 2$ for DCT, respectively.

If n transforms are used for filter F_i (such as in HTNR), the IC vector at a higher dimension is a concatenation of ICs corresponding to each transform t_i . The transforms are denoted as set $T_i = \{t_1, \dots, t_n\}$. Since (7)–(9) and (11) are all expressed in terms of the IC, the formulations remain identical to CNR.

Let the set $\{(F_i, T_i)\}$ denote the hybrid transform profile (HTP). We form the profile by choosing, for each filter F_i , the best performing model T_i among CNR, WNR, DCTNR, and

TABLE I
HYBRID TRANSFORM PROFILE (HTP) 1 FOR TESTING ON VIVID

F_i	T_i	k_i	IC Dim
Noise	Curvelet	global max	6
Blur	Curvelet	global max	6
JP2K	CDF9/7	1	6
JPEG	DCT	2	2

TABLE II
HYBRID TRANSFORM PROFILES

Profile	HTP 1	HTP 2	HTP 3
Target	VIVID FL	LIVE DMOS	LIVE FL
Noise	Curvelet	Curvelet	Curvelet
Blur	Curvelet	Curvelet+Coiflet+Battle	Curvelet
JP2K	CDF9/7	Daubechies	Coiflet+Daubechies
JPEG	DCT	DCT	DCT

HTNR. Table I gives results when testing on VIVID. The chosen profile for different target data sets are listed in Table II.

We do not decide on the profile based on the characteristic of the image, but rather based on the question posed by the user. If the user is interested in a quality prediction with the best correlation with the FL, Profile 1 is the best. On the other hand, if the user desires a prediction close to the LIVE subjective score, Profile 2 is more appropriate.

E. VIVID Image Library

The VIVID includes both videos and images. The videos are not discussed in this paper. The images are built to calibrate HNR with the exact image FLs. The FL denotes the filter setting, i.e., the variance for Gaussian noise, the variance for Gaussian blur, BPP for JPEG2000 (0–24), or the compression quality threshold for JPEG (0–100). In this paper, the maximum variances for noise and blur are 2 and 7, respectively. and all FLs are normalized to the range $[0, 1]$, where 0 corresponds to the original image. VIVID has 65 original images, each filtered by 4 filters (noise, blur, JPEG2000, and JPEG) with 101 different FLs, resulting in a total of $65 \times 4 \times 101 = 26\,260$ images.

VIVID has three objectives. First, it is specifically designed for training-based IQMs, such as HNR. A large number of original images and FLs helps to reduce the detrimental effect of a few outliers to the model fit, making it more reliable. Of course, more samples allows more extensive covering of parameter space. Second, VIVID aims to help develop an algorithm that surpasses the ability of human vision, such as discriminating images with tiny FL difference, undetectable by eyesight alone. As a result, VIVID has a substantially higher number of original images and more FLs for each original image than found in the LIVE database.

Each 512×512 grayscale original image is scaled down from a 12 megapixel raw format color image taken by a Nikon D90. The color version of all the 65 original images is previewed in Fig. 7. In order to increase the consistent reliability of HNR applied to a wide range of images, this testing library includes many categories of natural images, including animals, architectures, landscapes, plants, portraits, vehicles, and more.



Fig. 7. Original images of VIVID.

TABLE III
BENCHMARKS OF HNR TESTING ON VIVID

Metric	Noise	Blur	JPEG2000	JPEG
CC	0.993	0.981	0.973	0.975
SROCC	0.993	0.981	0.976	0.975
RMS	0.0360	0.0573	0.0695	0.0651
ACI	0.011	0.018	0.020	0.017

TABLE IV
FILTER CLASSIFICATION PERCENTAGE

Filter	Noise	Blur	JPEG2000	JPEG
FCP (%)	94.4	76.7	93.4	91.7

V. TESTING AND RESULTS

A. Benchmark Measurements

We define the training set S_T to contain approximately half of the original and their corresponding filtered images from each image category of VIVID. In the prediction phase, the remaining images from VIVID form the prediction image set S_P . For each image in S_P , prefiltered by filter F_{i^*} at level l^* , we compute a predicted FL \tilde{l} from (9) and (12), and a predicted filter type \tilde{i} from (10). For each filter F_i , the predicted (\tilde{l}) and exact (l^*) FLs are analyzed with three statistical error measures, and the two sets of filter types (predicted \tilde{i} and i^*) are analyzed with a successful rate measure.

Following the Video Quality Experts Group (VQEG) Phase II test [25], the two sets of FL are analyzed by 1) the Pearson's linear CC, 2) the Spearman rank-order CC (SROCC), and 3) the rms error. These three measurements calculate the prediction accuracy, monotonicity, and average absolute error, respectively. We also calculate the average 95% confidence interval (ACI) for each filter, which is the average of the 95% confidence intervals of all the FLs.

In order to quantify how many testing images can be classified to the correct filter type, we define the filter classification percentage (FCP) as follows:

$$\text{FCP} = \frac{|\{\tilde{i} \mid \tilde{i} = i^*\}|}{|\{\tilde{i}\}|} \times 100\% \quad (13)$$

TABLE V

LINEAR CCs BETWEEN PREDICTED SCORES AND SUBJECTIVE SCORES (DMOS) FROM THE LIVE IMAGE DATABASE. THE BENCHMARKS OF FR METHODS ARE REFERENCED FROM [4], AND THE OTHERS ARE REFERENCED FROM THEIR OWN PAPERS. NOTE THAT FOR NOISE AND BLUR PREDICTION, THE FIRST NUMBER OF HNR SHOWS RESULTS TESTED ON FIRST HALF OF LIVE, AND THE SECOND NUMBER SHOWS THOSE OF THE SECOND HALF. IQMs WITHOUT A STAR USE VARIOUS NONLINEAR FITTINGS OPTIMIZED FOR EACH FILTER, WHILE THE COLUMN "ALL" IS THE CC USING A SAME NONLINEAR FITTING FOR ALL THE PREDICTED SCORES CROSSING DIFFERENT FILTERS. "N/A" UNDER THE FILTER-TYPE COLUMNS INDICATES THAT THE IQM DOES NOT WORK FOR THE FILTER

	CC	Noise	Blur	JP2#1	JP2#2	JP#1	JP#2	All
NR	HNR* proposed	0.948/0.976	0.922/0.950	0.925	0.883	0.879	0.780	0.921
	HNR (fitted)	0.948/ 0.980 (3rd)	0.926/ 0.950 (5th)	0.952 (3rd)	0.893	0.879	0.780	0.930
FR	PSNR	0.986	0.783	0.933	0.874	0.886	0.917	0.871
	SSIM [7]	0.974	0.949	0.970	0.971	0.970	0.988	0.940
	IFC [5]	0.977	0.969	0.942	0.963	0.921	0.974	0.944
	VIF [6]	0.988	0.976	0.979	0.979	0.971	0.989	0.953
RR	QAI [8]	0.889	0.887	0.935	0.949	0.845	0.970	N/A
NR	Choi [26]	0.914	N/A	0.914		N/A	N/A	N/A
	JNB [27]	N/A	0.932	N/A	N/A	N/A	N/A	N/A
	Marziliano* [11]	N/A	0.96	0.85		N/A	N/A	N/A
	Sazzad [13]	N/A	N/A	0.93	0.93	N/A	N/A	N/A
	Sheikh [28]	N/A	N/A	0.920	0.892	N/A	N/A	N/A
	Wang [24]	N/A	N/A	N/A	N/A	0.951	0.888	N/A
	BLINDS [20]	0.974	0.957	0.922		0.839		N/A
	BIQI [21]	0.954	0.829	0.809		0.901		N/A

where \hat{i} and i^* are the predicted and exact filters of image set S_i in S_P [defined in (10)], respectively. FCP refers to the percentage of images in S_i with successful filter-type prediction.

B. HNR Tests on VIVID

Table III summarizes the error estimates of the HNR model using VIVID. Approximately half of the original and their corresponding filtered images from each image category of VIVID are picked to form the training set. For each filter, predicted and exact FLs are compared using three metrics.

Both CC and SROCC are above 0.97 for all four filters. Since the best FR methods have a CC around 0.98 (although tested on a different image library), 0.97 indicates a very high correlation between predicted and exact FLs. RMS indicates that the square root of the average square error is below 7%. The ACI is less than 0.02 for all the filters, which means that for a given predicted FL, we have 95% confidence that the error is less than 2%.

The FCP across all images for each filter is above 95%, except in the case of blur filter (see Table IV). Blurry images are often mistaken as JPEG2000 images since in IC space (see Fig. 4), low-level blur and high-level JPEG2000 ICs share the same region. In this shared region, FLs of JPEG2000 images are typically higher than those of blurry images, leading to the misclassification of blurred images.

Note that our model requires a large number of image samples for training. The prediction accuracy not only depends on the approximation algorithm and model parameters, but also strongly depends on the training set through the choice of the included image categories (animals, architectures, landscapes, plants, portraits, vehicles, etc.), the number of original images, and the number of FLs. Our experiments (not shown due to limited space) confirm that reducing any of the three numbers results in less accurate predictions. For example, using only one category of images can reduce the CC between the exact and predicted scores up to 0.06. On the other hand, changing the number of FLs from 101 to 6 reduces the CC by about 0.03.

C. HNR Tests on LIVE

In order to compare the proposed HNR with other IQMs, we train our HNR with DMOS subjective scores from the public LIVE image database and predict the DMOS values following the process described in Section III-C.

We compare HNR against some of the best approaches from the FR, RR, and NR categories (see Table V). HNR provides good CC for noise, blur, and JPEG2000, except JPEG. Although HNR does not improve on any of the FR methods, one notes that HNR, as an NR method, does not require an original image. In addition, HNR and the latest BLINDS and BIQI are the only general NR methods that succeed for all four filters (albeit not well for JPEG). A strength of HNR is its ability to provide competitive estimations of DMOS in a consistent manner for different classes of filter, changing only the image transforms.

Note that such a test is quite unfair for HNR. The limited number of LIVE images available for training decreases the reliability of HNR. In addition, most IQMs strongly depend on using different nonlinear fittings of quality scores for each filter. However, these methods are not capable of classifying the filters, therefore, a nonlinear fitting for all the prediction data makes the improvement of CC much smaller. For example, VIF, which is the best IQM, does not actually have an advantage over HNR for noise, blur, and JPEG2000 prediction if using fitting over all data.

The LIVE database provides two sets of quality scores. One is the DMOS subjective score and the other is the FL objective score (except JPEG). Therefore, we perform three additional tests (see Table I). The images to be predicted are always LIVE images, but the quality scores to be trained and predicted are either DMOS scores or FLs, which decides the specific HTP to be used. The CCs are the average values of the two CCs testing on the first half and then the second half of LIVE images. Logistic fittings are used for each filter.

A comparison between Tests II and Test demonstrates that all the CCs in Test III are higher than those in Test II. We can draw two conclusions. First, the large number of FLs available



Fig. 8. From left to right: original image, original image with noise, compressed image, and compressed image with noise.

TABLE VI
CCs OF THREE TESTS ON LIVE IMAGES

Test	HTP	Train	Type	Noise	Blur	JP2K
I	2	LIVE	DMOS	0.962	0.936	0.923
II	3	LIVE	FL	0.969	0.968	0.915
III	3	VIVID	FL	0.979	0.973	0.934

TABLE VII
CCs OF NOISE AND BLUR LEVEL PREDICTIONS OF COMPRESSED
(VARIOUS LEVELS) IMAGES

Images (PSNR)	Noise	Blur
JPEG (27.7)	0.990	0.985
JPEG (25.7)	0.992	0.987
JP2K (27.1)	0.993	0.981

in the training set from VIVID results in more accurate predictions. Second, although there must be differences between the two image databases when generating the images, training on VIVID and predicting FLs of LIVE images produces good results.

A comparison between Test I (predicts DMOS) and Test II (predicts FLs) shows that our model has a better correlation with FL than that with DMOS. This is also confirmed by the very high CC when predicting the VIVID FLs shown in Section V-B.

D. Noise and Blur Level Prediction of Compressed Images

In VIVID, we apply individual filters directly to the original images (see Fig. 8, first row). However, most real world images result from multiple filter applications. For example, noise or blur is often found together in compressed images (see Fig. 8, second row). We, therefore, consider whether HNR can predict noise level from compressed noisy images or blur level from compressed blurry images.

To this end, we compressed the 65 original images in VIVID with a specific compression ratio, and then applied the noise or blur filter onto the compressed images. The training and prediction processes are recomputed on these noisy compressed images and blurred compressed images. We performed three tests, each with a specific compression level. To give an idea of the quality of these compressed images, the average PSNR values of the images are given. Table VII indicates that HNR can predict noise and blur levels at the same accuracy, with or without precompression of the image. This extends the applicability of HNR to more real world images. However, the tests only confirm that HNR can work for compressed images at selected compression ratios. Additional work is required to obtain a complete relationship between the whole range of compression levels, noise or blur levels, and the ICs of HNR model.

VI. CONCLUSION

In this paper, we presented HNR, a new NR IQM, and VIVID, an objective image library. The major contributions of the paper are as follows.

- 1) We studied the NSS of the log-pdf of the transformation coefficients in the transformed space. We used tens of thousands of natural images and various transform methods, including the new curvelet transform, several types of wavelet transforms, DCT, and some combinations thereof. A reliable statistical relation between the image quality levels and ICs (carrying some invariance property of NSS) was obtained, which makes NR assessment possible.
- 2) Compared to FR and RR methods, HNR does not require a reference image, yet it has CCs above 0.97 for all the four filters when tested on VIVID, and comparable levels of performance (except for JPEG) when tested on LIVE. In contrast to most NR methods designed to work for one or two filters, HNR is the only general approach that handles the four filters, while simultaneously achieving top or close to best performance of each filter (except LIVE JPEG).
- 3) HNR is capable of classifying whether an image is a noisy, blurred, JPEG2000, or JPEG compressed image with average success rate of 89%. In addition, HNR has been tested to assess the quality of images that have two filters applied to them. It predicts the noise or blur level of compressed images (with certain compression ratios) with CCs of over 0.98. To the best of our knowledge, no other methods address these issues.
- 4) The new VIVID image library has 65 high-quality natural images and 26 260 filtered images. It is very helpful for developing and testing IQMs, especially for training-based methods.

In the future, we would like to calibrate HNR to subjective scores using VIVID to improve the correlation with JPEG DMOS scores. To do so, we would use VIF (having the best correlation with subjective DMOS scores) to give approximate subjective quality scores for all the images in VIVID. Thereafter, we can use the large number of images in VIVID to examine, which IC and which transform would be best suited to estimate the real image quality. In addition, we are considering to assess the noise, blur, and compression levels when they are simultaneously applied to a single image.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for valuable comments.

REFERENCES

- [1] S. Kelby, *The Digital Photography Book, Volume 1*. Berkeley, CA: Peachpit, 2008.
- [2] J. Shen, Q. Li, and G. Erlebacher, "Curvelet-based no-reference image quality assessment," in *Proc. Pict. Coding Symp.*, May 2009, pp. 1–4.
- [3] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2" 2003 [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [4] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

- [5] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
- [6] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [7] 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.
- [8] Z. Wang, G. Wu, H. Sheikh, E. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.
- [9] K. Seshadrinathan and A. C. Bovik, "New vistas in image and video quality assessment," *SPIE Proc. Human Vis. Electron. Imag.*, vol. 5150, no. 1, pp. 649202–649202, Jan. 2007.
- [10] X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, vol. 1, pp. 449–452.
- [11] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. Int. Conf. Image Process.*, 2002, vol. 3, no. III, pp. 57–60.
- [12] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, 2000, vol. 3, pp. 981–984.
- [13] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "Spatial features based no reference image quality assessment for JPEG2000," in *IEEE Image Process.*, Sep. 2007, vol. 3, pp. III-517–III-520.
- [14] E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Ann. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.
- [15] D. L. Ruderman, "The statistics of natural images," *Netw.: Comput. Neural Syst.*, vol. 5, pp. 517–548, 1994.
- [16] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 18, pp. 1661–1666, Oct. 2000.
- [17] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2091–2110, Dec. 1989.
- [18] T. Brando and M. P. Queluz, "Image quality assessment based on DCT domain statistics," *Signal Process.*, vol. 88, no. 4, pp. 822–833, Apr. 2008.
- [19] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Blind quality assessment for JPEG2000 compressed images," in *Proc. IEEE Asilomar Conf. Signals, Systems, Comput.*, Nov. 2002, vol. 2, pp. 1735–1739.
- [20] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 583–586, Jun. 2010.
- [21] 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.
- [22] E. J. Candes and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with C2 singularities," *Commun. Pure Appl. Math.*, vol. 57, no. 2, pp. 219–266, Nov. 2003.
- [23] E. J. Candes, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model Simul.*, vol. 5, pp. 861–889, 2006.
- [24] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, 2002, pp. 477–480.
- [25] Final report from the video quality experts group on the validation of objective models of video quality assessment phase II VQEG, 2003 [Online]. Available: <http://www.vqeg.org>
- [26] M. G. Choi, J. H. Jung, and J. W. Jeon, "No-reference image quality assessment using blur and noise," in *Proc. World Acad. Sci., Eng. Technol.*, Feb. 2009, vol. 38, pp. 163–167.
- [27] 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.
- [28] 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.



Ji Shen received the B.S. degree in applied math from Nanjing University, Jiangsu, China, in 2003. He is currently working toward the Ph.D. degree at the Department of Mathematics, Florida State University, Tallahassee.

In spring 2004, he was a visiting scholar in visual computing group at Microsoft Research Asia, Beijing, China. His research interests include image processing, image/video quality assessment, and image compression.



Qin Li received the B.S. degree in computational math from Sichuan University, Chengdu, China, in 2005. She is currently working toward the Ph.D. degree at the Department of Mathematics, Florida State University, Tallahassee.

Her research interests include sparse approximation and compressive sensing.



Gordon Erlebacher received the B.S. and M.S. degrees from Free University of Brussels, Brussels, Belgium, in 1979, and the Ph.D. degree in applied mathematics from Columbia University, New York, NY, in 1983.

He is currently the Director of the Visualization Laboratory of Department of Scientific Computing, Florida State University, Tallahassee, where he is also a Full Professor in the same department. From 1983 to 1989, he was a Research Scientist in NASA Langley Research Center, Hampton, VA. From 1989 to 1996, he was a Senior Staff Scientist and Research Fellow at the Institute for Computer Applications in Science and Engineering, Hampton, VA. He is the author or coauthor of more than 80 published papers and technical reports. His research interests include the applications of modern graphics processing units to simulation, analysis, and visualization, and the applications of gaming to education.

Dr. Erlebacher has received five achievement awards from NASA. He is the Editor of *Wavelets: Numerical Methods and Applications*.