Objective Quality Assessment of Interpolated Natural Images

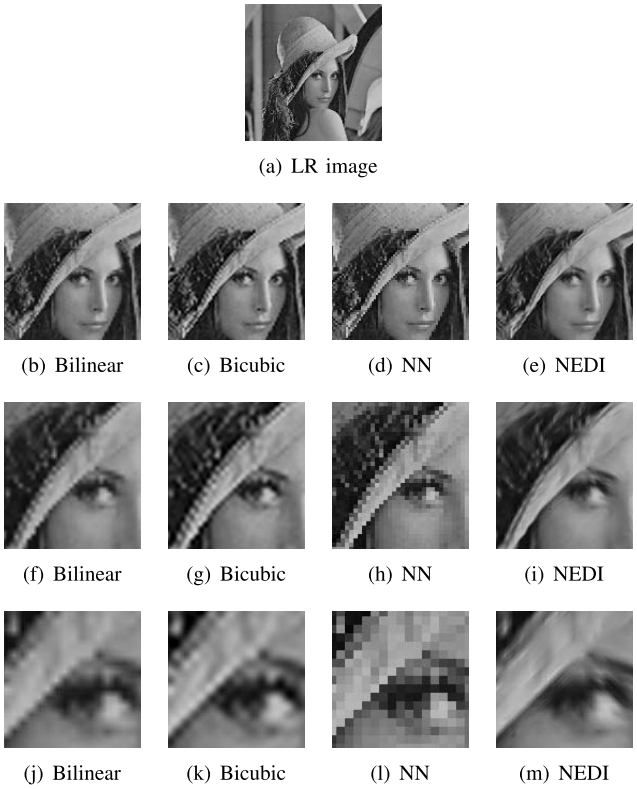
Hojatollah Yeganeh, *Member, IEEE*, Mohammad Rostami, *Student Member, IEEE*, and Zhou Wang, *Fellow, IEEE*

***Abstract*— Image interpolation techniques that create high-resolution images from low-resolution (LR) images are widely used in real world applications, but how to evaluate the quality of interpolated images is not a well-resolved issue. Subjective assessment methods are useful and reliable, but are also slow and expensive. Here, we propose an objective method to assess the quality of an interpolated natural image using the available LR image as a reference. Our method adopts a natural scene statistics (NSS) framework, where image quality degradation is gauged by the deviation of its statistical features from the NSS models trained upon high-quality natural images. Two distortion measures are proposed, namely, interpolated natural image distortion (IND) and weighted IND. Validations by subjective tests show that the proposed approach performs statistically equivalent or sometimes better than an average human subject. Moreover, we demonstrate the potential application of the proposed method in parameter tuning of image interpolation algorithms.**[1]

***Index Terms*— Image quality assessment, image interpolation, natural scene statistics.**

I. Introduction

IMAGE interpolation techniques that can improve the spatial resolution of given low-resolution (LR) images are extensively used in real-world devices and systems such as web browsers, media players, photo editors, and high-definition television (HDTV) [1]. Over the past decades, an increasing number of interpolation algorithms have been proposed. Early approaches such as nearest neighbor, bilinear, and bicubic interpolations [2] predict the values of missing pixels using spatially invariant fitting functions. More advanced recent methods are often adapted towards image edges or locally oriented structures [3]–[9]. Figure 1 depicts examples of reconstructed high resolution (HR) images created from the LR "Lena" image for scaling factors of 2, 4 and 8, by means of bilinear, bicubic, nearest neighbor (NN), and new edge-directed interpolation (NEDI) [3] methods, respectively. It can be observed that as the scaling factor increases, the perceptual differences between different interpolation methods become more pronounced. A natural question arises here is: with a variety of interpolation methods available, which one produces more natural-looking realistic HR images? To answer this question, image quality assessment (IQA) methods are highly desirable, without which, different interpolation methods cannot be compared and future improvement is pointless.



(a) LR image

(b) Bilinear (c) Bicubic (d) NN (e) NEDI

(f) Bilinear (g) Bicubic (h) NN (i) NEDI

(j) Bilinear (k) Bicubic (l) NN (m) NEDI

Fig. 1. (a): low-resolution (LR) image; (b-e): interpolated images by a scaling factor of 2; (f-i): interpolated images by a scaling factor of 4; (j-m): interpolated images by a scaling factor of 8. Column 1: bilinear interpolation; Column 2: bicubic interpolation; Column 3: nearest neighbor (NN) interpolation; Column 4: new edge-directed interpolation (NEDI) [3]. All interpolated images are cropped for better visualization.

Subjective evaluation provides a direct and reliable method in assessing the quality of interpolated images. In subjective IQA experiments, human subjects may be instructed to assign a "quality score" to each test image, and then the mean

Manuscript received September 30, 2013; revised July 29, 2014 and March 10, 2015; accepted June 24, 2015. Date of publication July 14, 2015; date of current version September 4, 2015. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada through the Discovery, Strategic, and in part by the Steacie Memorial Award Programs. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sam Hasinoff.

H. Yeganeh and Z. Wang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: hyeganeh@uwaterloo.ca; zhou.wang@uwaterloo.ca).

M. Rostami is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: m2rostam@uwaterloo.ca).

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.2015.2456638

[1]Partial early results of this work were presented at *IEEE International Conference on Image Processing*, Orlando, FL, USA, Oct. 2012.

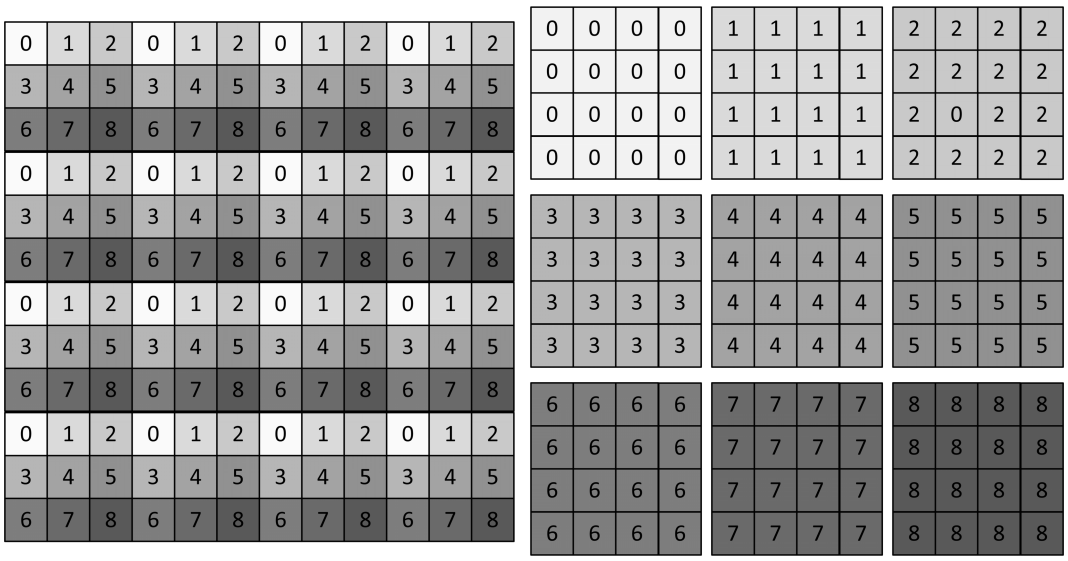


Fig. 2. Illustration of sub-image extraction from an HR image.

opinion scores (MOSs) as well as the variations between multiple subjects can be calculated [10], [11]. The subjects may also be asked to compare pairs of images and pick the one with better quality [12], [13]. Such a two-alternative-forced-choice (2AFC) approach has been shown to provide consistent results in the literature of visual psychophysics [14], though with low efficiency. A more efficient method is to ask the subjects to rank the quality of multiple images [15]. All these subjective evaluation methods are useful in comparing the performance of interpolation algorithms. However, they are often time-consuming and expensive, which largely constrains their applications when the volume of images becomes large or when one aims to incorporate them into the optimal design of interpolation algorithms.

Very limited progress has been made in automatic or objective quality assessment of interpolated images [15]. The difficulty lies in the fact that a perfect-quality HR image is unavailable to compare with. As a result, typical full-reference (FR) objective IQA approaches (where a perfect-quality reference HR image is available) such as peak signal-to-noise-ratio (PSNR) and the structural similarity (SSIM) index [16], [17] are not directly applicable. It is worth noting that the pixels in the LR image constitute a subset of the HR image pixels and are available to the IQA system. This well fits into the category of reduced-reference (RR) IQA, where only partial information about the perfect-quality original image is accessible [18], [19]. Recently, there has been a growing interest in using natural scene statistics (NSS) based approaches for RR and no-reference (NR, where no information about the reference image is available) IQA [20]. The basic assumption behind NSS approaches is that high-quality images captured from the natural visual environment have strong low-level statistical regularities [21], to which the biological visual apparatus has adapted and evolved over the millennia [20]. Consequently, any departure from such regularities creates "perceptual unnaturalness", which is assumed to be directly related to perceived image quality. In the past decade, NSS based approaches have been successfully used in a number of RR and NR IQA algorithms [19], [22]–[25], though have not been exploited in quality assessment of interpolated images.

In this work, we aim to develop an NSS-based objective quality assessment method for interpolated natural HR images using their corresponding LR images as references. Specifically, we build our NSS models using statistics of three features trained from high-quality natural images (details are given in Section II). The same set of features are calculated for the interpolated image being evaluated, and the deviations of these features from the NSS models are combined to produce an overall distortion measure of the interpolated HR image. A subjective experiment is carried out on interpolated natural images to validate the proposed method, which exhibits good agreement with subject opinions. Finally, we demonstrate the potential application of the proposed approach by applying it in the parameter tuning of existing image interpolation algorithms.

II. Proposed Method

Given an LR image, an image interpolation algorithm increases the spatial resolution to create an HR image by predicting and inserting new pixels between the existing LR image pixels. In practice, it is often desirable to rescale an image to arbitrary sizes. Strictly speaking, "interpolation" is not the correct term for such applications because the original positions and values of the LR image pixels cannot be fully maintained. In the current work, we only consider the case of interpolations by integer scaling factors and there is no fractional-pixel shift in the LR image. As such, the LR image can be viewed as a downsampled version of the HR image, where the pixel intensities remain unchanged at the sampling points. Another assumption we make here is that the LR image is noise free. This is reasonable because practical interpolation algorithms are noise sensitive, and denoising is typically applied before interpolation.

Fig. 3. An example of sub-images extracted from an interpolated HR image.

Let α be the interpolation scaling factor. Starting from the HR image, we could create α^2 downsampled sub-images, where one of them is the original LR image. An illustration of this sub-image extraction process is shown in Fig. 2 and a corresponding example of an interpolated image is given in Fig. 3, where $\alpha = 3$ and all the pixels marked with "0" in Fig. 2 constitute a sub-image that is exactly the same as the original LR image. All other sub-images are composed of pixels generated during the interpolation process. Our general assumption is that the statistical properties of all sub–images extracted from a high–quality HR image are similar, because all sub–images are nothing but downsampled versions of the same original image at different starting points. This may not be true for interpolated images, where only one of the sub-images is from the original HR image, while all others are artificially created by the interpolation algorithm. The key idea is to observe the relative difference in statistical properties across sub-images. For example, by close observation of the sub-images in Fig. 3, one can discern the differences between the original LR and the other sub-images, which often exhibits "unnaturally" smoothed edges and structural content. This observation inspires us to design statistical features based on sub-image statistics and to quantify how such features differ between high-quality natural HR images and interpolated HR images. Since the LR image is available as one of the sub-images, the design of the statistical features could make use of such information. The first two statistical features employed in the proposed IQA algorithm are developed by following these ideas and are elaborated in the following sub-sections.

A. Frequency Energy Falloff Statistics

It has long been discovered that the amplitude spectrum of natural images falls with the spatial frequency approximately proportional to $1/f^p$ [26], where f is the spatial frequency and p is an image dependent constant. This motivates us to develop a statistical feature based on frequency energy falloff. Specifically, we decompose the sub-images into dyadic scales using a steerable pyramid transform [27], which constitutes a tight frame and thus the energy in the spatial domain is preserved in the transform domain. For natural images, the energy in each scale, computed as the sum of squared transform coefficients, falls from the coarse to fine scales. Generally, the trends of scale or frequency energy falloff of natural images are fairly similar [26], while unnatural blurry images tend to have steeper slopes.

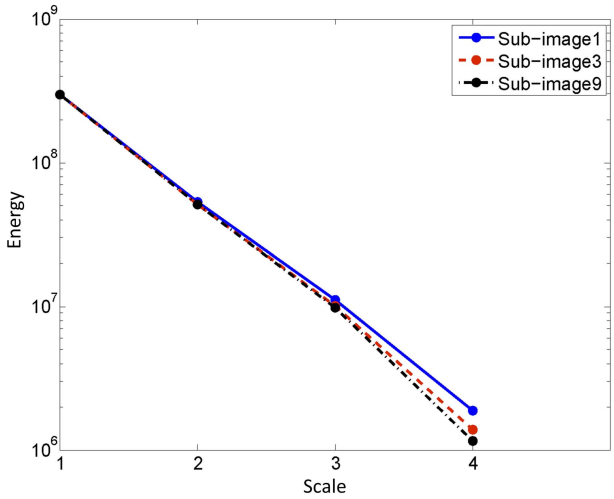


Fig. 4. Frequency energy falloffs of sub-images in Fig. 3.

Figure 4 shows the frequency energy falloff curves drawn for different sub-images in Figure 3. Two useful observations can be made here: First, the falloffs are approximately straight lines in log scale, which is consistent with the $1/f^p$ model; Second, the slope of energy falloff reflects the blurriness of the images. In particular, with the increase of blurriness from Sub-image1 to Sub-image3, and to Sub-image9, the slopes

of the falloffs become steeper. The substantial difference in the speed of falloff between different sub-images observed here is unlikely to happen in high-quality natural images, where statistics on sub-images are presumably similar. This observation suggests that measuring the statistical differences between sub-image frequency energy falloff could be a useful feature to distinguish interpolated images from high-quality natural images. As shown in Fig. 3, the deviation of the slope in the interpolated sub–image from that in the LR image maximizes at two finest scales. Therefore, we use s_i to denote the slope of frequency energy falloff between the two finest scales in the i-th sub-image, and define a frequency energy falloff feature as

$$e_f = \frac{\left[\frac{1}{\alpha^2-1}\sum_{k=2}^{\alpha^2}(s_k - s_1)^2\right]^{1/2}}{s_1}, \tag{1}$$

where s_1 computed from the first sub-image (which is the same as the reference LR image) is used as a reference, and the deviation of all other sub-images (for $k = 2, \ldots, \alpha^2$) are averaged and normalized by s_1. For a high-quality HR image, the deviation is expected to be small, leading to a small e_f feature (though unlikely to be zero, as will be shown in Section II-D). By contrast, for interpolated images that create blurry sub-images (as exemplified by Fig. 3), the e_f statistic is expected to be much larger.

B. Dominant Orientation Statistics

The frequency energy falloff feature described above is based on global energy measures, while interpolation processes often result in distortions in local image structures. In the literature, image gradient is widely used to study local structural details, particularly on edges and orientations [28]–[30]. Let $\mathbf{I}$ be an $N \times N$ image patch, and the gradient at pixel (x, y) in $\mathbf{I}$ be $\nabla\mathbf{I}(x, y) = \left[\nabla_x\mathbf{I}\ \nabla_y\mathbf{I}\right]_{(x,y)}$, where ∇_x and ∇_y denote the derivatives in horizontal and vertical directions, respectively. In the literature, careful study had been dedicated to the optimal design of multi–dimensional derivative filters [31]. Here we adopt the derivative kernels proposed in [31] for image gradient estimation. By pooling the gradients of all pixels in the patch, we obtain a gradient matrix given by

$$\nabla\mathbf{I} = \begin{bmatrix} \vdots & \vdots \\ \nabla\mathbf{I}_x(x, y) & \nabla\mathbf{I}_y(x, y) \\ \vdots & \vdots \end{bmatrix}_{N^2\times 2}. \tag{2}$$

We follow the well-known compact singular value decomposition (SVD) approach [28] to estimate the dominant orientation in the patch, together with the energy along the dominant and its orthogonal directions. The compact SVD of $\nabla\mathbf{I}$ can be written as

$$\nabla\mathbf{I} = \mathbf{U}\Lambda\mathbf{V}^T = \mathbf{U}\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}[\mathbf{v}_1\,\mathbf{v}_2]^T, \tag{3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, and the column vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ indicate the dominant gradient orientation and its orthogonal direction (dominant edge orientation), respectively. The matrix Λ is a 2×2 diagonal matrix, where the singular values $\lambda_1 \geq \lambda_2 \geq 0$ provide energy measures along $\mathbf{v}_1$ and $\mathbf{v}_2$ directions, respectively. A simple and elegant energy-independent orientedness measure is given by [28]

$$C = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}. \tag{4}$$

In case that the image patch is fully oriented with one dominant direction, $\lambda_1 \gg \lambda_2 \approx 0$, and thus $C \approx 1$. With the decrease of the gap between λ_1 and λ_2, the C measure declines. At the other extreme, when the energy along the two orthogonal orientations are equally strong or when the image patch is very smooth with little energy, the value of C is close to 0. This orientedness measure has found successful applications in identifying local dominant orientations of textures [28] and has been extended to multi-scale settings [29].

We compute the orientedness measure C over an 11×11 sliding window that runs across each sub-image. Let l_k be the mean of the C measure of the k-th sub-image, and we define our dominant orientation statistic features as

$$e_l = \frac{\left[\frac{1}{\alpha^2-1}\sum_{k=2}^{\alpha^2}(l_k - l_1)^2\right]^{1/2}}{l_1}. \tag{5}$$

Like in Eq. (1), here l_1 is computed from Sub-image 1 which is the same as the reference LR image. The value of l_1 is used as a normalization factor to quantify the deviation of orientation strength of all other sub-images (for $k = 2, \ldots, \alpha^2$). For a high-quality HR images, the deviation are expected to be small, resulting in small e_l measures, while interpolated images may generate much larger e_l, which will be shown in Section II-D.

C. Spatial Continuity Statistics

Many interpolation algorithms create unnatural looking discontinuities in the spatial domain. This inspires us to study spatial continuity statistics and relate them to the naturalness of images. Here we follow the general approach successfully used in previous work to detect spatial discontinuity [32], [33], based on which we develop a new statistical feature for spatial discontinuity. Let $f(i)$ for $i = 0, \ldots, N-1$ be one row (or column) of pixels extracted from the image, where N is the number of pixels in the row (or column). A straightforward method to examine signal continuity is to compute an absolute differencing signal given by

$$g(i) = |f(i+1) - f(i)| \quad \text{for} \quad 0 \leq i \leq N-2. \tag{6}$$

In the case of a high-quality HR natural image, the spatial continuity behavior is presumably close to uniform in statistical sense, while such behavior in an interpolated image may vary in a pattern related to the interpolation grid. To capture such variations, for a given interpolation factor α, we first measure the average spatial continuity at every α pixels by

$$k_j = \frac{1}{M}\sum_{i=0}^{M-1} g(\alpha i + j) \quad \text{for} \quad 0 \leq j \leq \alpha - 1. \tag{7}$$

where $M = \lfloor (N-1)/\alpha \rfloor$. This results in an length-α vector $\mathbf{k} = \{k_j | j = 0, \cdots, \alpha - 1\}$. We then use the ratio between the

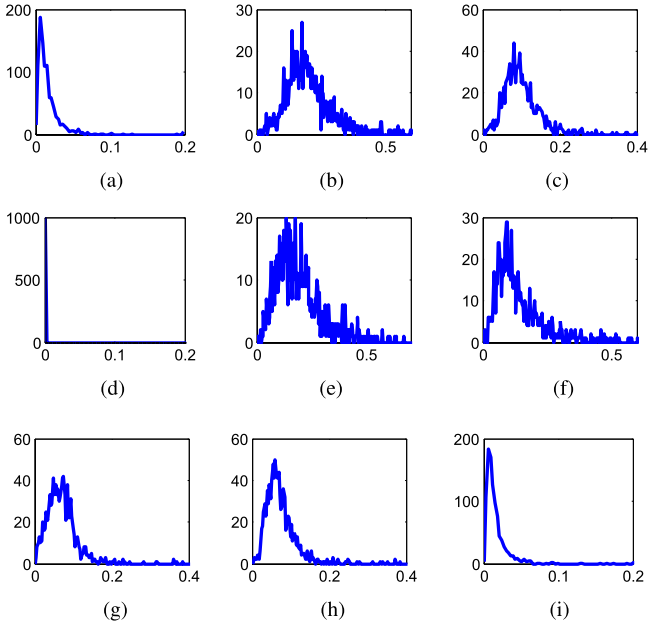


Fig. 5. Histograms of e_f feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [3], (f) (DFDF) [6], (g) ARSD [5], (h) NARM [4] and (i) ICBI [7] algorithms.

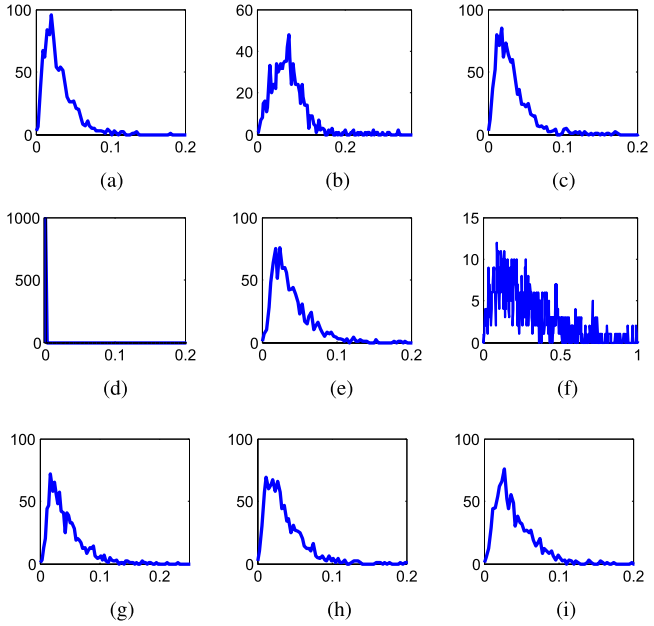


Fig. 6. Histograms of e_l feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [3], (f) (DFDF) [6], (g) ARSD [5], (h) NARM [4] and (i) ICBI [7] algorithms.

mean and std of the **k** vector as a measure of spatial continuity variation:

$$e_s = \frac{\text{std}(\mathbf{k})}{\mu(\mathbf{k})}. \tag{8}$$

This measure is averaged across all rows and columns in the image, resulting in an overall spatial continuity variation feature.

D. Statistical Modeling

At the heart of our approach is the NSS framework, where NSS models obtained from collections of natural images are essential in establishing the base reference to assess the naturalness and quality of the images being tested. Figures 5(a), 6(a), and 7(a) show the histograms of e_f, e_l and e_s features, respectively, obtained from 1000 high-quality original HR natural images. In addition, in Figures 5(b)-(i), 6(b)-(i), and 7(b)-(i), we show the corresponding histograms of interpolated HR images of scaling factor 2 generated using 8 interpolation approaches, which include bilinear, bicubic, nearest neighbor, NEDI [11], directional filtering and data fusion (DFDF) [6], adaptive autoregression and soft-decision estimation (ARSD) [5], nonlocal autoregressive modeling (NARM) [4], and iterative curvature-based interpolation (ICBI) [7] algorithms. It can be observed that for high quality natural HR images, the histograms of all three features are concentrated near zero but do not peak exactly at zero. By contrast, different interpolation methods introduce different types and levels of changes in e_f, e_l and e_s features. As a result, the statistics of these features deviate from those of natural images. For examples, interpolation algorithms that tend to create overly smooth images (such as bilinear interpolation, bicubic interpolation, NEDI and DFDF) significantly expand the dynamic ranges

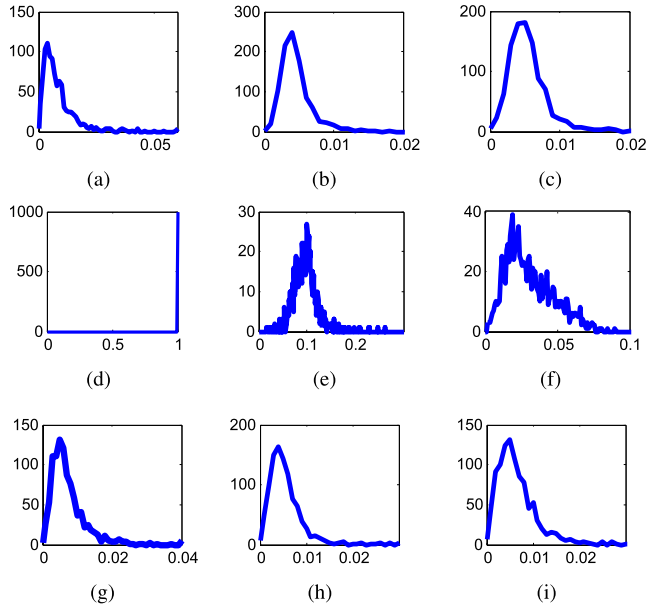


Fig. 7. Histograms of e_s feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [3], (f) (DFDF) [6], (g) ARSD [5], (h) NARM [4] and (i) ICBI [7] algorithms.

of the e_f feature. On the other hand, the nearest neighbor method repeats the originally LR pixel values to create HR images, resulting in large peaks at zero in e_f and e_s features. The edge directed NEDI method significantly affects spatial continuity statistics, whereas the bicubic or bilinear interpolation methods may strengthen spatial continuity. It is also interesting to observe that the feature histograms created by the most advanced algorithms such as ICBI exhibit the closest statistics to those of the natural images. All the above

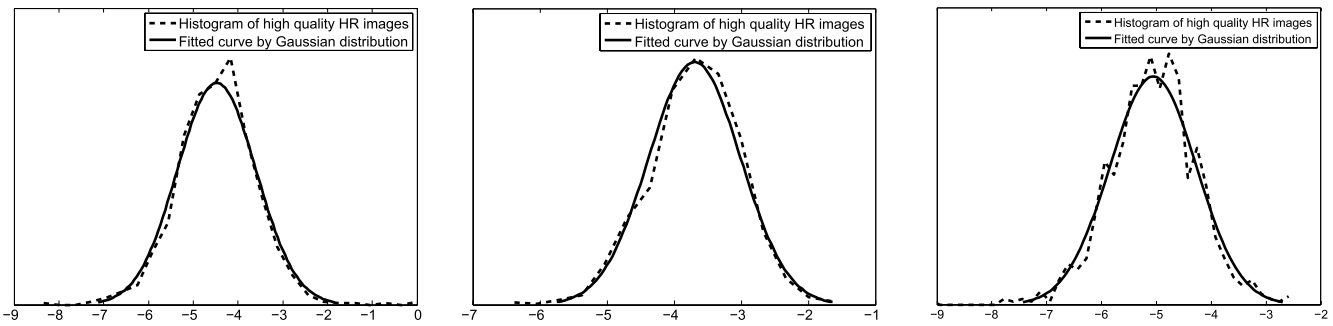


Fig. 8. Histograms of (a) $\ln e_f$, (b) $\ln e_l$ and (c) $\ln e_s$ features for $\alpha = 2$ drawn from original high-quality HR natural images, along with their corresponding Gaussian fitting functions.

TABLE I
FITTING PARAMETERS OF STATISTICAL MODELS

$\mu_f = -6.017\alpha^{-0.40}$	$\sigma_f = 0.72$
$\mu_l = -5.5\alpha^{-0.58}$	$\sigma_l = 0.62$
$\mu_s = -6.28\alpha^{-0.31}$	$\sigma_s = 1.1\alpha^{-2.2} + 0.53$

observations are intuitively sensible and demonstrate the potential usefulness of the proposed features, but in order to convert them to a quantitative image quality/distortion measure, we first need to build probability density models of these features.

All three features are non-negative by definition. In our study, we find it useful to observe them in the logarithm domain. Figures 8(a), 8(b) and 8(c), respectively, show the histograms of $\ln e_f$, $\ln e_l$ and $\ln e_s$ features for $\alpha = 2$ drawn from high-quality HR natural images. It is interesting to observe that all three histograms can be well fitted using Gaussian functions, which are also shown in the corresponding figures. This allows us to model the probability density of these features using simple models. In particular, for e_f feature, we have

$$p_{e_f}(e_f) = \frac{1}{Z_f} \exp\left[-\left(\frac{\ln e_f - \mu_f}{\sqrt{2}\sigma_f}\right)^2\right], \tag{9}$$

where Z_f is a constant normalization factor that ensures the density model integrate to 1. μ_f and σ_f are the logarithm domain mean and standard deviation parameters, respectively, for which the optimal values are obtained by maximal likelihood estimation. We found that the same probability density model provides good fittings to the feature histograms drawn from high-quality HR natural images, regardless of the values of α, though the optimal model parameters may change with α. By computing the optimal parameters for different α values between 2 and 8, we find empirical models that well summarize the optimal parameters, μ_f and σ_f, as functions of α. The empirical models are given in Table I.

Similarly, for e_l feature, we obtain the following model for high-quality HR natural images

$$p_{e_l}(e_l) = \frac{1}{Z_l} \exp\left[-\left(\frac{\ln e_l - \mu_l}{\sqrt{2}\sigma_l}\right)^2\right], \tag{10}$$

where Z_l is a normalization factor and the maximal likelihood estimations of the model parameters μ_l and σ_l are provided in Table I.

For e_s feature, the model for high-quality HR natural images is given by

$$p_{e_s}(e_s) = \frac{1}{Z_s} \exp\left[-\left(\frac{\ln e_s - \mu_s}{\sqrt{2}\sigma_s}\right)^2\right], \tag{11}$$

where Z_s is a normalization factor, and the maximum likelihood estimations of the model parameters μ_s and σ_s are summarized in Table I. Note that all the parameters introduced so far are purely determined by statistics of high-quality natural images only, without involving distorted images or human interference.

E. Quality Assessment Model

We use the p_{e_f}, p_{e_l} and p_{e_s} models built upon statistics of natural images in Section II-D as the basis to assess the naturalness of test images. A high-quality natural image is expected to have larger values of p_{e_f}, p_{e_l} and p_{e_s} than distorted unnatural images. Assuming statistical independence between e_f, e_l and e_s features, one can define a naturalness measure based on a joint probability model:

$$p_n = \frac{1}{K} p_{e_f}(e_f) p_{e_l}(e_l) p_{e_s}(e_s), \tag{12}$$

where a normalization factor $K = 1/(Z_f Z_l Z_s)$ is added such that the value of p_n is upper-bounded by 1. In information theory, self-information or "surprisal" is often employed as a measure of the information content associated with the outcome of a random variable. We adopt this approach and convert the probability-based measure in Eq. (12) into an interpolated natural image distortion (IND) measure given by

$$\text{IND} = -\ln p_n. \tag{13}$$

Plugging Eq. (9), Eq. (10), Eq. (11) and Eq. (12) into Eq. (13), we have

$$\begin{aligned} \text{IND} &= -\ln\left(Z_f p_{e_f}(e_f)\right) - \ln\left(Z_l p_{e_l}(e_l)\right) - \ln\left(Z_s p_{e_s}(e_s)\right) \\ &= \left(\frac{\ln e_f - \mu_f}{\sqrt{2}\sigma_f}\right)^2 + \left(\frac{\ln e_l - \mu_l}{\sqrt{2}\sigma_l}\right)^2 + \left(\frac{\ln e_s - \mu_s}{\sqrt{2}\sigma_s}\right)^2 \\ &\equiv D_f + D_l + D_s, \end{aligned} \tag{14}$$

where we define the first term, denoted by D_f, as the distortion of frequency energy falloff feature, the second term, denoted by D_l, as the distortion of dominant orientation statistical feature, and the third term, denoted by D_s, as the distortion of spatial continuity feature.

Although IND provides a simple and parameter-free (no training using distorted images or subjective testing data is involved) measure about the statistical unnaturalness of the test images, it lacks the flexibility to account for the variations in perceptual annoyance to different types of distortions. A natural extension of this approach is to assign different importance to different distortion features by linearly weighting the three distortion components. This results in a weighted IND measure (WIND) given by

$$\text{WIND} = w_f D_f + w_l D_l + w_s D_s, \tag{15}$$

Without loss of generality, we fix $w_l = 1$, and the remaining weighting parameters w_f and w_s are determined based on subjective evaluation, and the details are given in Section III.

III. Validation

To validate the proposed quality model, we built a database of interpolated images and carried out subjective quality assessment experiments. The database contains thirteen high-quality natural HR source images, representing different types of structural content, including indoor and outdoor scenes, humans, animals, natural scenes, and man-made architectures. None of the images used to build the NSS models in Section II is selected in this subjective test database. All source images have a size of 512 × 512. By directly downsampling the images by factors of 2, 4 and 8, we created 39 LR images with sizes of 256 × 256, 128 × 128 and 64 × 64, respectively. For each downsampled image, eight interpolation algorithms were employed to create interpolated HR images by scaling factors of 2, 4 and 8, respectively. The interpolation algorithms include classical and widely used bilinear, bicubic and nearest neighboor interpolation methods, as well as state-of-the-art algorithms such as NEDI [3], DFDF [6], ARSD [5], NARM [4], and ICBI [7]. Most of them interpolate an image by a scaling factor of 2, and were iteratively applied 2 and 3 times to achieve scaling factors of 4 and 8, respectively. Eventually, a total of 312 interpolated HR images were created, which are divided into 3 scaling factor levels and totally 39 image sets, each with 8 interpolated HR images.

Thirty subjects, including 17 males and 13 females, aged between 20 and 30, participated in the experiments. The subjects were either naïve or only have general knowledge about image processing, but no prior knowledge about the specific research work being carried out for this study. An HP ZR30w 30-inch monitor was used for the subjective test and the display spatial resolution is 2560 × 1600. This allows us to display a full set (out of the 39 sets) of 8 interpolated images, together with the original source HR image and the LR image, on the same screen. The viewing distance is adjusted to be approximately 32 pixels per degree of visual angle. A brief introduction and training session is conducted before the test. For each of the 39 image sets, the subjects were asked to use the source HR image and the LR image as references, and score each of the 8 interpolated images shown on the screen with a quality scale between 1 and 10. After the subjective test, a statistical analysis was performed and one subject was identified to be an outlier and the corresponding scores were removed. The remaining 29 subjective scores for each image were averaged to a mean opinion score (MOS). It is worth noting that we are using an absolute scale rating approach in the experiment, as opposed to paired comparison or direct ranking approach. We find that for this particular experiment, this approach gives more reliable results as compared to paired comparison method (which is slow and may cause transition problems) or direct ranking approach (which often leads to large variations between subjective rankings when the quality of two or more images are close).

In the subsequent analysis, we treat the MOS value obtained for each image as the "ground truth", which is used to compare against any other quality prediction method. Spearman's rank-order correlation (SRCC) and Pearson's linear correlation coefficient (PLCC) after a monotonic non-linear mapping are calculated to quantify the level of agreement between MOS and the quality prediction method being tested. Both evaluation criteria are upper-bounded by 1, which corresponds to perfect agreement, and higher values represent higher levels of agreement.

Before applying the SRCC and PLCC tests to assess objective quality models, we first evaluate how an average subject would perform in such tests. This is done by computing the SRCC and PLCC values between MOSs and the scores given by any particular individual subject. When this is done for all 29 subjects, we calculate the mean and standard deviation of the SRCC and PLCC values across all subjects. These average subject performance measures give useful baseline reference points on how an objective model behaves relative to a typical human subject. The average subject performance in terms of SRCC and PLCC is provided in Tables II, III for different scaling factors. In general, an average subject's SRCC and PLCC values are only moderately correlated with MOSs and are typically between the range of 0.5 and 0.8. This suggests that although subjects generally agree with each other on the quality of interpolated images, there exist significant variations between subject opinions. In the case of scaling factor 2, most interpolation algorithms perform quite well, making it difficult for the subjects to differentiate the relative quality of the interpolated images. This is reflected in the relatively low mean and high std of SRCC and PLCC values. With the increase of the scaling factor, the differences between the interpolation algorithms can be more easily discerned, making it a relatively easier task for the subjects. This leads to improved agreement between subjects, reflected by higher mean SRCC and PLCC values (Tables II and III).

One interesting observation is that the performance of an average human subject improves with the increase of the interpolation scale factor while an opposite pattern is observed for the purposed objective method. One possible theory that may be useful to provide an explanation is as follows. In the field of computational neuroscience, it has been widely hypothesized that the visual system is highly adapted to

TABLE II
SRCC EVALUATION ON INTERPOLATED IMAGES WITH DIFFERENT SCALING FACTORS

Image	$\alpha = 2$									
	mean/std subject performance	IND	WIND	NIQE (NR)	BIQI (NR)	DIVIINE (NR)	BLINDS-II (NR)	BRISQUE (NR)	SSIM (FR)	PSNR (FR)
Lena	0.66/0.32	0.50	0.52	0.45	0.90	-0.02	0.07	0.50	0.52	0.57
Butterfly	0.65/0.37	0.43	0.52	0.64	0.71	-0.24	0.03	0.26	0.38	0.69
Barbara	0.45/0.34	0.43	0.47	0.23	0.14	0	0.50	0.35	0.38	0.38
Boat	0.61/0.28	0.74	0.85	0.30	0.45	-0.16	0.30	-0.04	0.33	0.40
Einstein1	0.35/0.31	0.80	0.70	0.35	0.27	-0.11	0.67	0.05	-0.01	0.04
Statue	0.61/0.28	0.26	0.74	0.23	0.45	0	0.16	0.38	0.40	0.52
Lighthouse	0.34/0.34	0.55	0.59	0.38	0.42	0.06	0.15	0.13	0.20	0.24
Museum	0.54/0.32	0.47	0.43	0.59	0.54	-0.19	-0.11	0.14	0.01	0.14
Peppers	0.56/0.38	0.26	0.43	0.47	0.90	-0.14	0.20	0.26	0.3	0.23
Goldhill	0.32/0.30	0.54	0.71	-0.38	0	0.52	0.77	-0.66	0.21	0.21
Fruit	0.37/0.36	0.49	0.57	0.19	0.29	-0.01	0.06	0.14	0.28	0.10
Baboon	0.31/0.32	0.54	0.70	0.63	0.50	-0.11	0.08	0.31	0.21	0.16
Einstein2	0.57/0.26	0.28	0.69	-0.21	0.16	0.54	-0.40	0.23	0.69	0.64
Average	**0.50/0.32**	**0.48**	**0.61**	**0.30**	**0.37**	**0.01**	**0.15**	**0.15**	**0.30**	**0.32**
Image	$\alpha = 4$									
Lena	0.73/0.19	0.93	0.81	0.42	0.80	0	0.17	0.42	0.47	0.42
Butterfly	0.76/0.20	0.95	0.90	0.59	0.09	0.16	0.05	0.59	0.76	0.73
Barbara	0.71/0.20	0.47	0.59	0.23	0.54	-0.16	0.66	0.23	0.16	0.10
Boat	0.58/0.35	0.19	0.57	0.16	0.38	-0.64	0.71	0.16	0.38	0.23
Einstein1	0.66/0.25	0.64	0.43	0.19	0.28	-0.09	0.58	0.19	0.61	0.66
Statue	0.64/0.18	0.76	0.71	0.92	0.69	-0.09	0.02	0.92	0.38	0.38
Lighthouse	0.51/0.31	0.62	0.78	0.19	0.02	-0.26	0.81	0.19	-0.02	-0.04
Museum	0.51/0.30	0.62	0.69	0.30	0.28	-0.42	0.69	0.30	0.40	0.45
Peppers	0.75/0.17	0.57	0.83	0.23	0.78	0.47	0.23	0.23	0.16	0.02
Goldhill	0.79/0.16	0.54	0.64	-0.33	0.19	-0.38	0.40	-0.33	0.88	0.90
Fruit	0.70/0.22	0.64	0.36	0.42	-0.04	-0.09	0.47	0.42	0.35	0.33
Baboon	0.64/0.30	0.83	0.78	0.52	-0.45	-0.01	0.47	0.52	-0.16	-0.11
Einstein2	0.77/0.17	0.74	0.88	0.66	0.57	0	0.50	0.66	0.11	0.30
Average	**0.67/0.23**	**0.65**	**0.69**	**0.35**	**0.31**	**-0.20**	**0.44**	**0.35**	**0.34**	**0.33**
Image	$\alpha = 8$									
Lena	0.77/0.19	0.30	0.71	0.07	0.14	-0.38	0.09	-0.14	0.61	0.85
Butterfly	0.74/0.22	0.77	0.73	0.74	0.34	-0.03	0.13	0.34	0.62	0.34
Barbara	0.63/0.25	0.24	0.33	-0.23	0	-0.45	-0.04	-0.07	0.42	0.92
Boat	0.71/0.24	0.35	0.38	0.40	0.47	-0.64	0	0.07	0.59	0.33
Einstein1	0.73/0.26	0.0	0.12	0.40	0.54	0.04	0.40	0	0.42	0.47
Statue	0.61/0.28	0.40	0.50	0.19	0.42	-0.54	0.19	-0.33	0.33	0.35
Lighthouse	0.66/0.29	0.38	0.74	0.64	0.16	-0.09	-0.11	0.52	-0.26	-0.16
Museum	0.69/0.20	0.19	0.38	-0.09	0.02	-0.09	-0.28	-0.11	0.83	0.78
Peppers	0.58/0.24	0.65	0.62	0.68	0.71	-0.76	0.15	0.55	0.37	0.26
Goldhill	0.73/0.24	0.38	0.55	-0.01	0.22	0.17	-0.37	-0.17	0.88	0.70
Fruit	0.72/0.19	0.35	0.31	0.04	0.14	-0.90	0.19	0.14	0.38	0.54
Baboon	0.62/0.30	0.50	0.70	0.49	-0.03	-0.28	-0.16	0.49	0.04	-0.29
Einstein2	0.73/0.23	0.35	0.50	-0.19	0.07	-0.53	-0.33	-0.23	0.54	0.45
Average	**0.69/0.24**	**0.37**	**0.50**	**0.24**	**0.25**	**-0.36**	**-0.01**	**0.08**	**0.44**	**0.43**

optimally encode the information in the natural scenes [21]. In other words, prior models about what a natural scene should look like have already been built in the visual system. When evaluating the quality of an image, one makes use of two general cues. The first is signal fidelity, for which one can compare the distorted image with what is known about the reference image (in the case of image interpolation, the LR image provides a reduced reference). The second is about naturalness, for which one has to refer to the built-in prior models of natural scenes to judge the quality of the test image. When the interpolation scale factor is small, there is relatively more information about the reference image, and with an increasing scale factor, such information becomes less. This explains why the performance of the objective quality models like ours degrade. By contrast, the visual system has much stronger prior knowledge about statistical naturalness, and is thus able to make better judgement at higher scale factors.

Determining the optimal weighting parameters w_f and w_s in Eq. (15) based on subjective data is a straightforward linear regression problem. In addition to finding the

TABLE III
PLCC EVALUATION ON INTERPOLATED IMAGES WITH DIFFERENT SCALING FACTORS

Image	$\alpha = 2$									
	mean/std subject performance	IND	WIND	NIQE (NR)	BIQI (NR)	DIVIINE (NR)	BLINDS-II (NR)	BRISQUE (NR)	SSIM (FR)	PSNR (FR)
Lena	0.77/0.20	0.67	0.66	0.37	0.73	0.25	0.47	0.46	0.75	0.69
Butterfly	0.73/0.24	0.70	0.74	0.73	0.62	0.10	0.33	0.39	0.60	0.70
Barbara	0.63/0.27	0.50	0.50	0.35	0.25	0.81	0.34	0.36	0.48	0.51
Boat	0.67/0.27	0.72	0.86	0.54	0.62	0.63	0.53	0.03	0.63	0.63
Einstein1	0.56/0.40	0.93	0.92	0.59	0.78	0.04	0.54	0.04	0.81	0.86
Statue	0.74/0.27	0.87	0.86	0.29	0.83	0.79	0.68	0.19	0.76	0.87
Lighthouse	0.44/0.37	0.77	0.89	0.61	0.67	0.69	0.51	-0.01	0.69	0.69
Museum	0.66/0.25	0.82	0.82	0.76	0.77	0.18	0.39	0.04	0.73	0.75
Peppers	0.57/0.39	0.78	0.82	0.63	0.76	0.39	0.42	0.31	0.69	0.69
Goldhill	0.52/0.42	0.95	0.95	0.17	0.91	0.94	0.92	-0.18	0.92	0.94
Fruit	0.55/0.45	0.93	0.93	0.10	0.87	0.90	0.73	-0.30	0.87	0.90
Baboon	0.47/0.32	0.77	0.77	0.62	0.78	0.45	0.53	0.31	0.77	0.76
Einstein2	0.65/0.26	0.66	0.71	0.31	0.53	0.31	0.54	0.29	0.75	0.79
Average	**0.61/0.32**	**0.77**	**0.80**	**0.47**	**0.70**	**0.49**	**0.53**	**0.15**	**0.73**	**0.75**
Image	$\alpha = 4$									
Lena	0.86/0.07	0.90	0.93	0.67	0.72	0.05	0.77	0.72	0.76	0.76
Butterfly	0.84/0.12	0.93	0.92	0.69	0.69	0.34	0.30	0.78	0.81	0.88
Barbara	0.82/0.13	0.70	0.83	0.58	0.70	0.19	0.10	0.58	0.70	0.70
Boat	0.72/0.20	0.67	0.71	0.65	0.51	0.83	0.78	0.37	0.67	0.67
Einstein1	0.79/0.14	0.63	0.63	-0.02	0.60	0.46	0.76	0.40	0.71	0.73
Statue	0.84/0.10	0.92	0.93	0.80	0.88	0.87	0.58	0.89	0.88	0.87
Lighthouse	0.68/0.25	0.66	0.73	0.08	0.66	0.68	0.93	0.66	0.66	0.66
Museum	0.64/0.25	0.45	0.63	0.57	0.42	0.43	0.91	0.59	0.59	0.70
Peppers	0.79/0.15	0.74	0.90	0.46	0.88	0.78	0.40	0.42	0.65	0.65
Goldhill	0.83/0.13	0.67	0.60	0.35	0.67	0.57	0.10	0.36	0.80	0.87
Fruit	0.76/0.20	0.67	0.73	-0.15	0.60	0.60	0.73	0.56	0.73	0.73
Baboon	0.71/0.28	0.85	0.84	0.82	0.66	0.32	0.67	0.76	0.60	0.60
Einstein2	0.83/0.15	0.77	0.91	0.80	0.68	0.48	0.64	0.76	0.51	0.51
Average	**0.78/0.17**	**0.73**	**0.79**	**0.48**	**0.67**	**0.51**	**0.59**	**0.60**	**0.70**	**0.72**
Image	$\alpha = 8$									
Lena	0.82/0.16	0.52	0.75	0.59	0.09	-0.41	0.06	0.26	0.78	0.90
Butterfly	0.79/0.18	0.56	0.86	0.77	0.49	-0.27	0.30	0.37	0.78	0.67
Barbara	0.75/0.28	0.80	0.79	0.75	0.33	-0.55	- 0.04	0.44	0.85	0.87
Boat	0.74/0.24	0.75	0.70	0.53	0.38	-0.70	0.01	0.25	0.52	0.51
Einstein1	0.80/0.15	0.56	0.60	0.66	0.42	-0.14	0.43	0.17	0.81	0.81
Statue	0.74/0.18	0.72	0.70	0.40	0.68	-0.62	0.33	-0.04	0.72	0.72
Lighthouse	0.76/0.29	0.43	0.53	0.46	0.02	-0.53	-0.27	0.34	0.46	0.46
Museum	0.78/0.20	0.70	0.73	0.54	0.12	-0.74	-0.53	0.30	0.72	0.75
Peppers	0.69/0.20	0.68	0.78	0.64	0.65	-0.24	0.16	0.55	0.64	0.64
Goldhill	0.76/0.24	0.66	0.88	0.46	0.05	-0.72	-0.38	0.28	0.87	0.82
Fruit	0.79/0.18	0.40	0.63	0.37	0.25	-0.47	0.07	0.22	0.81	0.60
Baboon	0.69/0.28	0.65	0.82	0.38	-0.23	-0.59	-0.36	0.24	0.38	0.38
Einstein2	0.82/0.22	0.66	0.78	0.50	0.01	-0.62	-0.34	0.25	0.55	0.50
Average	**0.76/0.22**	**0.62**	**0.73**	**0.54**	**0.25**	**-0.51**	**-0.04**	**0.28**	**0.68**	**0.66**

optimal parameters, we also perform cross-validations to test the robustness of the process [34]. In particular, for each scaling factor, we randomly divide all images to a training group and a testing group. The weight parameters w_f and w_s are obtained by linear regression using the training group and are then tested using the testing group. Since the images from the same image set (out of 13 sets) are generated from the same source image, assigning these images to both training and testing groups violates the independence between them. Thus we divide the images based on image sets. As such, a leave-1-out method will use 12 image sets for training and 1 for testing, and a 7-6 split method uses 7 image sets for training and 6 for testing. In each case, the training and testing process is repeated multiple times, each with a different random split between training and testing groups. A comparison between leave-1-out and 7-6 split approaches are shown in Table IV, where we observe that the 7-6 split method performs slightly worse, which may be due to its limited coverage of image content in the training group. Therefore, we use leave-1-out method

TABLE IV
COMPARISON OF CROSS VALIDATIONS USING LEAVE-1-OUT AND 7-6 SPLIT IN WIND

Scaling factor	SRCC		PLCC	
	WIND leave-1-out	WIND 7-6 split	WIND leave-1-out	WIND 7-6 split
$\alpha = 2$	0.61	0.57	0.80	0.79
$\alpha = 4$	0.69	0.69	0.79	0.78
$\alpha = 8$	0.50	0.54	0.73	0.68

TABLE V
WEIGHTING FACTORS w_f AND w_s, ALONG WITH SRCC AND PLCC PERFORMANCE, OBTAINED FROM LEAVE-ONE-OUT TEST FOR DIFFERENT SCALING FACTOR α

α	mean/std of w_f	mean/std of w_s	SRCC	PLCC
2	1.17/0.07	0.09/0.02	0.612	0.802
4	1.26/0.04	0.16/0.01	0.700	0.800
8	3.20/0.10	0.40/0.05	0.511	0.732

to decide on the parameters in our WIND algorithm. Specifically, for each scaling factor, the leave-1-out process is repeated 13 times, each with a different set for testing. The mean and std of the w_f and w_s values obtained for scaling factors 2, 4 and 8 are given in Table V, together with the corresponding mean SRCC and PLCC performance test results over 13 trails. It appears that the weights obtained in all 13 trails are fairly close to each other, with low std values. In addition, the mean SRCC and PLCC results of the 13 leave-1-out trials are also close to those obtained using a fixed set of the average weighting parameters applied to all 13 image sets (which are shown in the WIND performance in the last rows in Tables II-III). All these suggest that the weights obtained through this procedure is robust. The mean w_f and w_s values for each scaling factor α given in Table V are the final values in all the subsequent tests.

In Table V, we also observe that with the increase of the scaling factor, the impact of frequency energy falloff and dominant orientation statistics increases (as compared to spatial continuity statistics). This is not surprising because with larger scaling factor, it becomes more difficult for an interpolation algorithm to maintain the original energy distributions at high frequencies as well as the orientations in local image structures. These are also the major factors that affect the perceptual quality of the interpolated images.

Table V provides the weighting parameters for scaling factors of 2, 4 and 8 only. A natural way to extend the weight selection approach at other integer scaling factors is to interpolate along the scaling factor axis, and we found the following functions are useful for this purpose:

$$w_f = 0.0002\alpha^{4.43} + 1.16, \tag{16}$$

$$w_s = 0.008\alpha^{1.7} + 0.06, \tag{17}$$

To the best of our knowledge, no other objective IQA method is directly applicable to the scenario we are interested in (where an LR image is used as a reference to assess the quality of an HR image). Therefore, in addition to comparing the proposed IND and WIND measures with an average human subject (as described earlier), we compare them with well-known FR-IQA measures, including PSNR and SSIM [16]. We have also included state-of-the-art NR-IQA methods NIQE [35], BIQI [36], DIVIINE [37], BLIINDS-II [38] and BRISQUE [24] in the comparison, which have shown promising performance when tested using a number of public image databases, but has never been tested on interpolated images. Note that the FR methods are included here as reference models only. They are not applicable in real applications because the perfect-quality original reference HR image is not available.

To evaluate the performance of different IQA measures, one could compute the SRCC and PLCC correlations of objective measures against MOS for all interpolated images and all scaling factors. However, such results could be misleading, because the most crucial job in real applications is to differentiate different interpolation methods for the same source image at the same scale factor. Therefore, a more relevant analysis approach is to validate the performance of different IQA methods on a set of images interpolated from the same source at the same scale factor. This approach is also fully consistent with how our subjective study was designed.

The performance evaluation results are summarized in Tables II and III for the cases of scaling factors 2, 4 and 8. It can be observed that for almost all image sets, the SRCC and PLCC values of the proposed IND and WIND measures are well within the range of ± 1 standard deviation from the SRCC and PLCC values of average human subjects. This indicates that the proposed methods behave quite similarly to an average subject. Unsurprisingly, between IND and WIND, WIND performs consistently better and often outperforms an average human subject. The FR SSIM and PSNR methods are general-purpose approaches without any prior knowledge or specific considerations of image interpolation application, thus they only achieve moderate correlations with subjective evaluations and are inferior to the proposed methods, which use novel features particularly designed to capture distortions in interpolated images. Moreover, none of the existing NR models exhibits competitive performance.

IV. APPLICATION: PARAMETER TUNING IN INTERPOLATION ALGORITHMS

The application domain of objective IQA measures is beyond comparing interpolated images and selecting the best interpolation algorithms. Many interpolation algorithms contain one or more parameters. Different selections of these parameters may result in interpolated HR images with very different perceptual quality, and the optimal parameters are often image-dependant. Without human interference, it is a challenging task to choose these parameters. An objective quality measure provides a useful tool to pick these parameters automatically. To demonstrate this, here we use the ICBI algorithm [7] as an example.

The ICBI method is a state-of-the-art "edge-directed" interpolation algorithms that upscale the image by keeping the original pixels in an enlarged grid and then estimate the

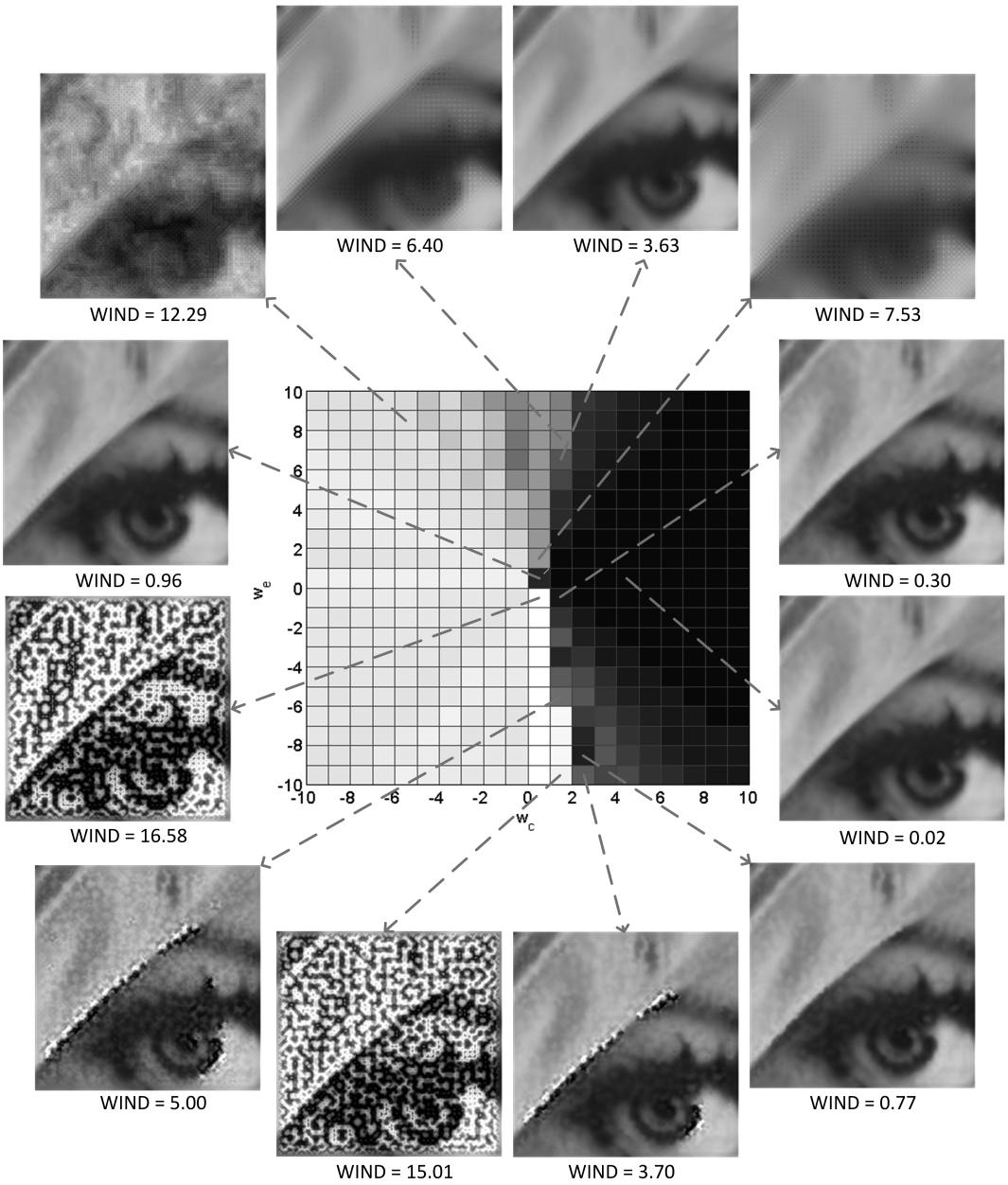


Fig. 9. ICBI [7] interpolated images over a wide range of w_c and w_e selections. Darker shade indicates lower WIND value or higher image quality.

missing pixels. The estimation is done using weighted averaging of neighboring pixels where the weights are determined based on local edge analysis. The ICBI algorithm uses some initial values for missing pixels, and then tries to define an energy term for each interpolated pixel. The ultimate goal is to minimize the energy term by small changes in the second order derivatives. In [7], the energy term is a weighted sum of three components: curvature continuity U_c, curvature enhancement U_e and isophote smoothing U_i. The energy of each interpolated pixel at (i, j) is given by

$$U(i, j) = w_c U_c(i, j) + w_e U_e(i, j) + w_i U_i(i, j), \quad (18)$$

According to [7], the U_c term is effective in removing artifacts but creates blurry images while the U_e term helps produce sharper edges. As a result, the relative values between w_c and w_e determines a tradeoff between edge sharpness and artifacts removal. It was found that the U_i component adds only a slight improvement to perceived image quality and thus w_i has relatively little influence on the performance of ICBI [7].

Since the perceptual quality of the interpolated image varies significantly with w_c and w_e, these parameters are typically chosen with try and error [7]. To visualize this, in Fig. 9, we plot the WIND measure as a 2D function of w_c and w_e for a test image interpolated by ICBI at scaling factor 4, where darker shade in the plot indicates a lower value of WIND or higher quality of the interpolated image. Sample interpolated images corresponding to different options of (w_c, w_e) values are also given. It is worth noting that the visual quality as a function of (w_c, w_e) is not smooth everywhere.

Indeed, it could be quite sensitive, such that in certain areas in the parameter space, small changes could lead to dramatic variations in perceived quality of the interpolated images, making it difficult to manually decide on the right parameters to use. Careful inspections and comparisons of the interpolated images as well as their corresponding WIND values suggest that WIND is a good perceptual quality indicator and provides a useful tool to automatically choose the best values of (w_c, w_e).

V. Conclusion and Future Work

We propose an NSS-based objective model to automatically assess the quality of interpolated natural HR images using LR images as references. Three statistical features are employed in the proposed approach, including sub-image frequency energy falloff statistics, sub-image local dominant orientation statistics and spatial continuity statistics. Statistical models are established based on high-quality natural HR images. The departures from such statistics are measured as the key indicators of perceptual unnaturalness, which is assumed to be closely related to perceived image quality. We build an image database of interpolated natural images and carried out subjective tests. Our experiments show that the proposed quality measure agrees well with the mean subjective opinions of interpolated image quality and often outperforms an average human subject. Furthermore, we demonstrate the extended potential applications of the proposed measure by applying it to automatic parameter tuning of state-of-the-art ICBI interpolation algorithm.

As one of the first attempts on a new research topic, our model has several limitations that should be improved in the future. First, the current method is applicable to the case of interpolations by integer factors only. In practice, users may resize an image by a fractional factor. Not all the statistical features used here can be directly applied and more feature extraction and statistical modeling work is necessary. Second, currently it is assumed that the three NSS features are independent to each other, which simplifies the model and produces promising results. In the future, how these features are related and how to optimally combine these features may be further exploited. Third, since the current models are built upon natural scene statistics, it may not properly generalize to the case of graphical images. Discovering new meaningful features for these images is a topic worth further investigating. Finally, many recent image super-resolution algorithms take one or multiple LR images as the input to create HR images, where the positions of the LR image pixels may be shifted by fractional factors from the integer pixel grid. This poses new challenges to IQA research and opens up new space for future exploration.

References

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Signal Process.*, vol. 26, no. 6, pp. 508–517, Dec. 1978.

[3] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[4] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.

[5] X. Zhang and X. Wu, "Image interpolation by adaptive 2D autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[6] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[7] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2760–2768, Oct. 2011.

[8] G. J. Grevera and J. K. Udupa, "Shape-based interpolation of multidimensional grey-level images," *IEEE Trans. Med. Imag.*, vol. 15, no. 6, pp. 881–892, Dec. 1996.

[9] G. T. Herman, J. Zheng, and C. A. Bucholtz, "Shape-based interpolation," *IEEE Comput. Graph. Appl.*, vol. 12, no. 3, pp. 69–79, May 1992.

[10] A. R. Reibman and T. Schaper, "Subjective performance evaluation of super-resolution image enhancement," in *Proc. 2nd Int. Workshop Video Process. Quality Metrics (VPQM)*, 2006.

[11] A. R. Reibman, R. M. Bell, and S. Gray, "Quality assessment for super-resolution image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 2017–2020.

[12] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 882–893, Oct. 2011.

[13] J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessment of stereoscopic images," *Multimedia Tools Appl.*, vol. 67, no. 1, pp. 31–48, 2013.

[14] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks," *Psychol. Rev.*, vol. 113, no. 4, pp. 700–765, 2006.

[15] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment for image super-resolution: A natural scene statistics approach," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1481–1484.

[16] 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.

[17] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.

[18] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Mateo, CA, USA: Morgan Kaufmann, Mar. 2006.

[19] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.

[20] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[21] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[22] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[23] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

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

[25] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[26] D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vis. Res.*, vol. 37, no. 23, pp. 3367–3383, 1997.

[27] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Oct. 1995, pp. 444–447.

[28] J. Bigun, G. H. Granlund, and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 775–790, Aug. 1991.

[29] X. Feng and P. Milanfar, "Multiscale principal components analysis for image local orientation estimation," in *Proc. Conf. Rec. 36th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Nov. 2002, pp. 478–482.
[30] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.
[31] H. Farid and E. P. Simoncelli, "Differentiation of discrete multidimensional signals," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 496–508, Apr. 2004.
[32] Z. Wang, A. C. Bovik, and B. L. Evan, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2000, pp. 981–984.
[33] 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, USA, Sep. 2002, pp. I-477–I-480.
[34] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1995, pp. 231–238.
[35] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 209–212, Mar. 2013.
[36] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.
[37] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[38] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

Hojatollah Yeganeh (S'10–M'14) received the B.S. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2006, the M.Sc. degree in electrical engineering with a specialization in speech processing and speech recognition from the Amirkabir University of Technology, Tehran, Iran, in 2009, and the Ph.D. degree in electrical engineering from the University of Waterloo, ON, Canada, in 2014. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His main research interests include image and video processing, quality assessment of high dynamic range images and videos, quality of experience of Internet videos, multimedia communications, and biomedical signal processing.

Mohammad Rostami received the B.S. degree in electrical engineering and the B.S. degree in pure mathematics from the Sharif University of Technology, Tehran, Iran, in 2010, and the M.Sc. degree in electrical and computer engineering from the University of Waterloo, Canada. He is currently pursuing the Ph.D. degree with the University of Pennsylvania, USA. His research interests include image and video processing/quality assessment, compressed sensing, computational neuroscience, and machine learning. He received the University of Waterloo Outstanding Achievement in Graduate Studies in Fall 2012 Convocation.

Zhow Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He has over 100 publications in his research fields with over 25 000 citations (Google Scholar). His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He was a member of the IEEE Multimedia Signal Processing Technical Committee (2013–2015). He was a recipient of the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Best Magazine Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as senior author). He served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), *Pattern Recognition* (2006–present), and the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009), the *EURASIP Journal of Image and Video Processing* (2009–2010), and *Signal, Image and Video Processing* (2011–2013).