

RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment

Rajiv Soundararajan* and Alan C. Bovik

Department of ECE, The University of Texas at Austin, Austin, TX 78712, USA

Email: rajivs@utexas.edu, bovik@ece.utexas.edu

Abstract

We study the problem of automatic “reduced reference” image quality assessment algorithms from the point of view of image information change. Such changes are measured between the reference image and natural image approximations of the distorted image. Algorithms that measure differences between the entropies of wavelet coefficients of reference and distorted images as perceived by humans are designed. The algorithms differ in the data on which the entropy difference is calculated and on the amount of information from the reference that is required for quality computation, ranging from almost full information to almost no information from the reference. A special case of this are algorithms that require just a single number from the reference for quality assessment. The algorithms are shown to correlate very well with subjective quality scores as demonstrated on the LIVE Image Quality Assessment Database and Tampere Image Database. The performance degradation as the amount of information is reduced is also studied.¹

I. INTRODUCTION

The role of visual media in everyday life has increased tremendously in recent years with ubiquitous applications. Given that images and videos in most applications are meant for human consumption, it is desirable to design systems in such a fashion so as to enrich the visual experience of these users. The field of image and video quality assessment (QA) seeks to partially address the question of how to quantitatively model user experience, and how to use these models to predict visual quality in accordance with visual perception.

¹EDICS category: SMR-HPM. This paper appeared in part at the IEEE International Conference on Acoustics, Speech and Signal Processing 2011.

Image and video QA algorithms can be broadly classified into full reference (reference available or FR) and no reference (reference not available or NR) algorithms. The mean squared error (MSE) has been used as a quality metric for a very long time, owing to its simplicity, despite having a very poor correlation with human perception [1]. The last decade has seen significant progress in the field of objective full reference image/video QA algorithms. The structural similarity index (SSIM) [2], visual information fidelity (VIF) [3], visual signal to noise ratio (VSNR) [4] and just noticeable difference (JND) metric [5], are examples of successful full reference algorithms which have been shown to perform well in predicting the quality scores of human subjects.

Progress on no reference QA, however, has been very slow. Indeed, the progress that has been possible, has been on account of relaxing the no reference assumption in various ways. One approach is to devise NR algorithms for a specific type of distortion only [6], [7], [8]. This approach can be refined by assuming that the distorted image is subjected to a set of possible distortions known *a priori*. Training based NR QA techniques have resulted in algorithms that perform at least as well as mean squared error, which has the benefit of a reference image [9], [10]. In [11], blind image quality assessment indices that measure the anisotropy in the distorted image through the Renyi entropy are introduced. However, performance of this index has not been thoroughly evaluated on a large and comprehensive dataset. Alternatively, partial information about the reference can be made available, which can be used along with the distorted image to predict quality. This paradigm is known as reduced reference (RR) QA, which may or may not require knowledge of the distortion type.

RR QA algorithms involve sending or supplying some amount of information about the reference along with the distorted image that is useful in quality computation. For example, the concept of *quality aware* images was proposed in [12], where partial reference image information is embedded within the image and can be extracted reliably despite distortions. The information embedded could for example, be the statistical parameters of the distribution of wavelet coefficients obtained by a multi scale-space-orientation decomposition of the reference image. Then, two parameters of a generalized Gaussian distribution and the error in approximating the empirical coefficients by this distribution are transmitted for every subband. The quality is based on computing the Kullback-Leibler (KL) divergence between the parametrized distribution of the reference and the empirical distribution of the distorted image. The performance of this algorithm is good only for certain individual distortion categories. This idea is further extended in [13], where an additional divisive normalization transform step is introduced before computing the KL divergence to improve performance. However, this algorithm depends on a number of parameters that need to be trained on databases. There has also been prior work on non-statistical reduced reference

methods that compute a generalized norm between selected wavelet coefficients of the reference and distorted images [14]. The coefficients are selected in a way that reduces the information required while still maintaining good performance.

Algorithms based on multiscale geometric analysis including curvelets, bandlets, wavelets and contourlets are developed in [15]. These algorithms depend on parameters that need to be tuned delicately on different databases. Moreover, their performances degrade severely with the reduction in the data rate required from the reference. Algorithms designed in [16] for color images, perform well when tested on images belonging to certain distortion classes such as JPEG or JPEG2000. While distributed source coding ideas are applied to approximate mean squared error in [17], RR QA algorithms based on color distribution of images are developed in [18], and training based approaches are used in [19]. These algorithms are either limited in their ability to achieve good performance across different distortion types or involve training on databases. The algorithm by [20] based on Weibull statistics of wavelet coefficients achieves good performance at a given data rate. However, what we desire in this work, is a family of algorithms that achieve graceful degradation in performance with the reduction in data rate and the ability to achieve better performance with increase in data rate.

In this paper, we develop a new framework of reduced reference QA algorithms that are information theoretic. We consider natural image approximations of the distorted image in the sense that wavelet coefficients of distorted images will be fitted with Gaussian scale mixture distributions. In effect, this paper approaches the problem of quality assessment from the perspective of measuring distances between the reference image and the projection of the distorted image on to the space of natural images. The algorithms compute the average difference between scaled entropies of wavelet coefficients of reference and projected distorted images that are obtained at the output of a neural noise channel. A family of algorithms are proposed depending on the subband in which the quality computation is carried out and the amount of information required from the reference image. This framework allows us to study how the performance of these information theoretic RR QA algorithms decays with reduction in the amount of information used from the reference. Since, the reference information scales with the size of the images, the algorithms are also applicable in scalable image and video QA [21]. Further, the algorithms allow for bidirectional computation of quality, by which we mean that the quality of the distorted image can also be computed at the reference if relevant information from the distorted image is made available. In other words, by supplying reduced information from the distorted image through a feedback channel, its quality can be computed at the reference without sending any information in the forward channel. This feature has potential applications in image/video quality monitoring in networks, which requires feeding

back the quality at different nodes in the network to the sender. Of course, since the quality index is based on the absolute difference, it does not indicate which of the two images is the reference image. In nearly any imaginable scenario, we know which of the two images is the reference. Another interesting feature of these algorithms is that they are not dependent on any parameters that need to be trained on databases. Depending on the bandwidth available for information supplied either from the reference or distorted image, one algorithm from this class may be picked for desired applications. The framework also allows users to choose an algorithm from this class for general purpose or distortion specific quality assessment.

The rest of the paper is structured as follows. In the following section, we describe prior information theoretic approaches that motivate the algorithms developed in this paper. In Section III, we present the main theory and resulting algorithms of the paper. We discuss perceptual interpretations of the algorithms in Section IV, and present numerical results in Section V. It is found that the performance attained is highly competitive on a large and comprehensive database of distorted images and subjective scores. Finally, we conclude the paper in Section VI.

II. INFORMATION THEORETIC APPROACHES TO QA

Information theoretic methods for quality assessment have produced some of the best performing *full reference* (FR) QA algorithms, including the information fidelity criterion (IFC) [22] and the visual information fidelity (VIF) index [3]. The motivation for such an approach is that image distortions tend to disturb the natural statistics of images and quantifying this disturbance can determine quality. It is also based on the assumption that such modifications of natural image statistics are perceptually noticeable. Mathematically, IFC and VIF compute the amount of mutual information shared between the reference and distorted images in the wavelet domain under a natural scene statistic (NSS) model. Moreover, both these algorithms possess psychovisual properties that are desirable for quality assessment. In particular, a number of similarities are drawn between these information theoretic methods and perceptual properties such as masking, suprathreshold effects, error pooling, scale-space-orientation decomposition, and so on, that make these algorithms perceptually and statistically appealing.

The mutual information terms in IFC and VIF can be shown to be functions of the correlation coefficients between patches of wavelet coefficients under the assumed NSS model, which is a Gaussian scale mixture (GSM) model [23]. This means that the entire reference image is required in order to compute the IFC/VIF indices (or the correlation coefficients). Hence these are both full reference QA algorithms and not useful in no reference and reduced reference scenarios. This leads to the question:

what information theoretic quantities can we compute using only reduced or no information from the reference?

The most successful general purpose reduced reference quality metrics are based on computing the KL divergence between the reference and test images [12]. A divisive normalization transform (DNT) step before computation of the KL divergence was shown to further improve performance [13]. The DNT resembles a divisive normalization stage used in visual neural models and also allows for local processing before computing the KL divergence, which is a global measure of distances. It is interesting to note that while the quality indices based on shared mutual information have perceptual interpretations such as masking, suprathreshold effects, error pooling, and so on [22], those based on KL divergence by themselves do not possess these features. Such psychovisual properties have to be additionally introduced, for example, by the divisive normalization transform as in [13] to improve performance. This leads to the question of whether we can design information theoretic QA indices in reduced reference scenarios that possess other desirable psychovisual properties in order to further improve their performance.

As a possible answer to questions posed in this section, the approach that we take in this paper, is to compute the average difference of scaled local entropies in the wavelet domain between the reference and the distorted image. In essence, the quality indices proposed compute the amount of local information difference between the reference and distorted images in a subband. Further, this quantity can be computed in a distributed fashion between the reference and distorted images, making it a reduced reference algorithm. The reason we compute the difference of entropies is primarily due to the constraint of the problem (reduced reference) than by choice. However, we show later on that this procedure still possesses desirable psychovisual properties leading to excellent performance of the algorithms. We present the details of the algorithms in the following section, reserving the perceptual interpretation of the algorithms for a later section.

III. RR QA ALGORITHMS

We now describe the main theory on which the algorithms proposed in the paper are based. The source model considered here closely follow the assumptions in [22] while we approximate the wavelet coefficients of the distorted image to also follow Gaussian scale mixture distributions. The resulting RR algorithms utilize the wavelet coefficients obtained by a steerable pyramid decomposition of the reference and distorted image into subbands at different orientations and scales [24]. Let K be the total number of subbands obtained as a result of this decomposition. The wavelet coefficients in subband k , $k \in \{1, 2, \dots, K\}$, are partitioned into M_k non-overlapping blocks, each block containing N coefficients

of size $\sqrt{N} \times \sqrt{N}$. Non-overlapping blocks are assumed independent and identically distributed (i.i.d.). Although, wavelet coefficients in adjacent blocks in scale, space or orientation may not be independent, we make this assumption in order to simplify the quality index.

A. Source Model

Wavelet coefficients of natural images are modeled well by Gaussian scale mixture distributions. Such models have also proved very useful in various image processing applications including quality assessment [3], denoising [25] and so on. Let $\bar{C}_{mk} = (C_{1mk}, C_{2mk}, \dots, C_{Nmk})$ denote the vector of coefficients in block m , $m \in \{1, 2, \dots, M_k\}$ of subband k of the reference image. Therefore,

$$\bar{C}_{mk} = S_{mk} \bar{U}_{mk},$$

where $\bar{U}_{mk} \sim \mathcal{N}(0, \mathbf{K}_{U_k})$ and S_{mk} is a scalar random variable that modulates the covariance matrix of the block \bar{C}_{mk} . Also, S_{mk} and \bar{U}_{mk} are independent. Subband k is associated with the covariance matrix \mathbf{K}_{U_k} and \bar{U}_{mk} for each block m , is distributed identically. Thus, the wavelet coefficient block \bar{C}_{mk} , when conditioned on the realization $S_{mk} = s_{mk}$, is distributed according to a Gaussian model with a covariance matrix $s_{mk}^2 \mathbf{K}_{U_k}$. Further, S_{mk} and \bar{U}_{mk} are each independent over m and k .

B. Distortion Model

Distortions introduced in a natural image may take it outside the space of natural images. As a result, it is possible that the wavelet coefficients of these images do not follow a Gaussian scale mixture distribution. We approach the problem of quality assessment by projecting the distorted image on to the space of natural images. This means that we model the wavelet coefficients of the distorted image as well as a Gaussian scale mixture distribution. If the distortion process were to retain the distorted image within the space of natural images, we would not make any error by employing such a model. We measure quality as a distance between the reference and a natural image approximation of the distorted image. We show later that the approach of approximating the wavelet coefficients of a distorted image by a GSM model results in RR QA algorithms that perform very well in predicting the quality scores of images. Denote $\bar{D}_{mk} = (D_{1mk}, D_{2mk}, \dots, D_{Nmk})$ as the vector of coefficients in block $m \in \{1, 2, \dots, M_k\}$ of subband k of the distorted image. We have

$$\bar{D}_{mk} = T_{mk} \bar{V}_{mk},$$

where $\bar{V}_{mk} \sim \mathcal{N}(0, \mathbf{K}_{V_k})$ and T_{mk} is the scalar pre-multiplier random variable as in the reference image. The independence assumptions are similar to the reference image. We now describe the RR index.

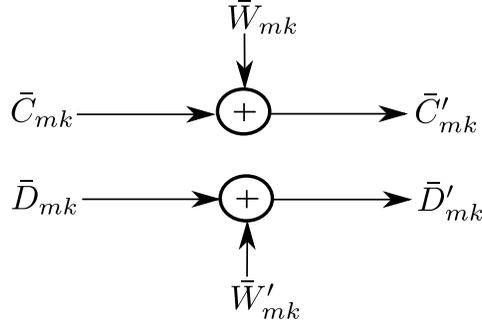


Fig. 1. System Model

C. RR Quality Index

We additionally model the perceived reference and distorted images as passing through an additive neural noise channel, where the noise is assumed to be a zero mean Gaussian random vector for each block of coefficients. The neural noise model accounts for uncertainty introduced by neural processing of the visual signal [3]. The resulting system model is shown in Figure 1. We have

$$\bar{C}'_{mk} = \bar{C}_{mk} + W_{mk}$$

$$\bar{D}'_{mk} = \bar{D}_{mk} + W'_{mk},$$

where $W_{mk} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$ and $W'_{mk} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$, W_{mk} and W'_{mk} are independent of each other, independent of \bar{C}_{mk} and \bar{D}_{mk} and independent across the indices m and k . The reduced reference quality indices that we introduce and which we term Reduced Reference Entropic Differences (RRED) indices are defined as the average of the absolute value of the difference between the scaled entropies of the neural noisy reference and distorted images, conditioned on the realizations of the respective premultiplier random variables in a subband.

Let the eigen values of \mathbf{K}_{U_k} be $\alpha_{1k}, \alpha_{2k}, \dots, \alpha_{Nk}$ and the eigen values of \mathbf{K}_{V_k} be $\beta_{1k}, \beta_{2k}, \dots, \beta_{Nk}$. In the following, assume that \mathbf{K}_{U_k} and \mathbf{K}_{V_k} are full rank matrices. If this is not true, then the index is calculated by using the positive eigen values alone. The entropy of a reference image block m in subband k conditioned on $S_{mk} = s_{mk}$ is given by ¹

$$h(\bar{C}'_{mk} | S_{mk} = s_{mk}) = \frac{1}{2} \log [(2\pi e)^N (s_{mk}^2 |\mathbf{K}_{U_k}| + \sigma_W^2 \mathbf{I}_N)] = \sum_{n=1}^N \frac{1}{2} \log [(2\pi e)(s_{mk}^2 \alpha_{nk} + \sigma_W^2)].$$

¹All logarithms in the paper are with respect to base 2.

Similarly, the entropy of the distorted image block conditioned on $T_{mk} = t_{mk}$ is given by

$$h(\bar{D}'_{mk} | \bar{T}_{mk} = t_{m,k}) = \sum_{n=1}^N \frac{1}{2} \log [(2\pi e)(t_{mk}^2 \beta_{nk} + \sigma_W^2)]. \quad (1)$$

Define scaling factors,

$$\gamma_{mk}^r = \log(1 + s_{mk}^2) \text{ and } \gamma_{mk}^d = \log(1 + t_{mk}^2).$$

The entropies conditioned on the realizations of the premultiplier random variables are multiplied by the above scalars before computing the difference. These scalars are increasing functions of the premultiplier random variables, that tend to zero as the premultiplier tends to zero and saturate at high values. The imposition of these scalars before computing the difference has many advantages. They lend a local character to the algorithm imposing additional local effects on the entropy terms. Moreover, the scalars may be interpreted as facilitating contrast masking while preventing over masking, as explained later in Section IV. The other benefit of these weights is in the context of those algorithms that operate with extremely small neural noise variance. In such a setting, these help saturate the entropy terms at locations having extremely small premultiplier random variable realizations, i.e. those that are close to zero. This helps avoid numerical instabilities in the computation of the entropy differences, especially when computing the logarithm of very small variance values.

We present a family of algorithms, by varying the subband in which quality is evaluated and the amount of information that is required from each subband for quality computation. First, we discuss algorithms obtained by varying the subband in which quality computation is carried out, alone. In these algorithms, the scaled entropies at each block in one particular subband k , $\{\gamma_{mk}^r h(\bar{C}'_{mk} | S_{mk} = s_{mk})\}_{m=1}^{M_k}$ of the reference image are used to evaluate quality. Since different subbands have different sizes, the number of blocks is reduced from the subbands at the finest to the subbands at the coarsest scales of the wavelet decomposition. Thus, the number of weighted entropy terms required is equal to the number of blocks in the corresponding subband, M_k .

The reduced reference QA index corresponding to subband k , when M_k scalars are available from the reference is given by

$$RRED_k^{M_k} = \frac{1}{L_k} \sum_{m=1}^{M_k} |\gamma_{mk}^r h(\bar{C}'_{mk} | S_{mk} = s_{mk}) - \gamma_{mk}^d h(\bar{D}'_{mk} | T_{mk} = t_{mk})|,$$

where L_k is the size (number of coefficients) of the subband k . The above index is a reduced reference index since M_k is less than the size of the image. We require all the M_k entropy terms since the absolute values of the differences are summed up. Note that the maximum size of M_k over all subbands is equal to the size of the image divided by N . Also, the information required reduces when the quality

is evaluated in the coarser bands as M_k reduces from subbands at finer scales to coarser scales. Either image can compute the index using the entropy information from the other image. The absolute value of the difference is calculated, since the nature of the distortion process could lead to either an increase or a decrease in entropy. We only wish to measure the magnitude of the difference to evaluate quality. This also implies that the RRED indices are always positive. Also, any enhanced image would show a difference in entropies and the difference should be interpreted as an improvement in quality.

The amount of information required from a subband can also be reduced by summing scaled entropy terms over patches and sending the sum of these scaled entropies instead of all the entropy terms. This is equivalent to filtering the image of weighted entropies in a subband using rectangular windows of sizes $b \times b$ and subsampling by b in each dimension, where b is a natural number that represents the size of the patches. This procedure results in loss of performance with subsampling as illustrated in Section V. Let Λ_k denote the number of subsampled blocks and let $\lambda \in \{1, 2, \dots, \Lambda_k\}$ index the block. Every $m \in \{1, 2, \dots, M_k\}$ belongs to one of the subsampled blocks $B_{\lambda k}$ in subband k . Define

$$g_{\lambda k}^r = \sum_{m \in B_{\lambda k}} \gamma_{mk}^r h(\bar{C}'_{mk} | S_{mk} = s_{mk}) \text{ and } g_{\lambda k}^d = \sum_{m \in B_{\lambda k}} \gamma_{mk}^d h(\bar{D}'_{mk} | T_{mk} = t_{mk}).$$

Then the RR quality index in subband k when Λ_k scalars are available from the reference is given by

$$RRED_k^{\Lambda_k} = \frac{1}{L_k} \sum_{\lambda=1}^{\Lambda_k} |g_{\lambda k}^r - g_{\lambda k}^d|.$$

Thus, by filtering and subsampling, we can reduce the information required from every subband for quality computation. For example, for $b = 2$, if the subband is filtered by windows of size 2×2 and subsampled by a factor of 2 in each dimension, then the number of entropy terms required reduces from M_k to $\Lambda_k = M_k/4$. This is another method of reducing the amount of information required as against evaluating quality in subbands at coarser scales. $RRED_k^1$ denotes the algorithm in which all the scaled entropy terms in the subband are added and only the sum, which is a single scalar, is required for quality computation. Since only a single number is needed, this may be considered as an almost reference free algorithm. Here, we do not imply that the almost reference free algorithm approximates a no reference algorithm, but rather that just a single number is required. The algorithm still requires the single number without which quality cannot be computed.

The two methods described above illustrate how the amount of information can be reduced gradually from an almost full reference scenario to an almost reference free scenario. Moreover, the filtering and subsampling procedure can be performed in coarser bands to further reduce the information. This results

in a family of algorithms with varying performance levels. There are other variations that could be performed on the class of algorithms discussed so far. Let \mathcal{K} denote the set of subbands from which quality indices are combined. For example, the quality index could be given as $\sum_{k \in \mathcal{K}} \mu_k RRED_k^b$ where μ_k are scalars used to weight the respective bands differently. We evaluate one such weighting strategy in Section V.

D. Estimation of Parameters

In order to compute the QA index, it is necessary to estimate \bar{s} , \bar{t} , \mathbf{K}_{U_k} and \mathbf{K}_{V_k} , $k \in \{1, 2, \dots, K\}$. The procedure outlined here is similar to the estimation of reference image parameters in [7]. We obtain maximum likelihood estimates for the above parameters. Without loss of generality, assume $\sum_{m=1}^{M_k} s_{mk}^2 = 1$ and $\sum_{m=1}^{M_k} t_{mk}^2 = 1$. Therefore, the ML estimates of \mathbf{K}_{U_k} and \mathbf{K}_{V_k} are given by

$$\hat{\mathbf{K}}_{U_k} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mk} \bar{C}_{mk}^T}{M_k} \quad \hat{\mathbf{K}}_{V_k} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mk} \bar{D}_{mk}^T}{M_k}.$$

Since the wavelet coefficients are conditionally Gaussian distributed, the ML estimates of s_{mk}^2 and t_{mk}^2 are given by

$$\hat{s}_{mk}^2 = \frac{\bar{C}_{mk}^T \mathbf{K}_{U_k}^{-1} \bar{C}_{mk}}{N} \quad \hat{t}_{mk}^2 = \frac{\bar{D}_{mk}^T \mathbf{K}_{V_k}^{-1} \bar{D}_{mk}}{N},$$

for $m \in \{1, 2, \dots, M_k\}$.

IV. PERCEPTUAL INTERPRETATION OF THE ALGORITHMS

We now cast the RR QA indices proposed in the previous section against perceptual principles often used in developing QA algorithms [26].

A. Scale-Space-Orientation Decomposition

The first step in the RR algorithm, similar to IFC/VIF, is a wavelet decomposition of the image at multiple scales and orientations. This step imitates the signal processing that happens in the primary visual cortex (area V1) of the human visual system. There are different wavelet transforms that could be used for the multi-scale multi-orientation decomposition. The Gabor family of wavelets are widely used in image processing owing to the fact that neural responses in the primary visual cortex are well modeled by Gabor filters [27]. Further, the Gabor functions achieve the lower bound on the uncertainty in space and spatial frequency and are thus simultaneously localizable. Successful video quality assessment algorithms such as the MOVIE index [28] employ Gabor filters to decompose video data prior to performing quality assessment.

On the other hand, orthogonal wavelet transforms used in multiscale wavelet analysis suffer from aliasing or lack of shift invariance. A small shift in space in the image signal could lead to a significant change in the subband responses. Steerable pyramid is an alternate multiscale multiorientation decomposition that has the advantage of being shift, scale and rotation invariant. Further, steerable pyramids have been successfully deployed in a variety of image processing applications including compression, denoising, deblurring, and so on. In the context of image quality assessment, full reference image quality indices such as IFC and VIF use the steerable pyramid wavelet decomposition. The quality indices are evaluated in the wavelet domain, using a Gaussian scale mixture distribution (GSM) model on the wavelet coefficients to compute information theoretic quantities. Likewise, we use the steerable pyramid decomposition to accomplish reduced reference image quality assessment.

B. Contrast Masking

The contrast masking principle refers to the phenomenon by which the visibility of distortions in a signal component is inhibited by the presence of a masker having a similar orientation or scale at a given location. Contrast masking has been modeled in various ways in the literature, the two main methods being threshold elevation [29], [30] and gain control through divisive normalization [31], [32], [33].

Divisive normalization refers to models of perceptual processing, whereby neural responses are divided by local (neural responses adjacent to the given location in space, orientation and scale) energy in the responses. Our approach to entropic RR QA algorithm design uses conditioning on the realizations of the premultiplier random variables, which is analogous to divisive normalization processes in the primary visual cortex (area V1) of the human visual system. Both divisive normalization and conditioning seek to reduce the amount of dependence in local blocks of wavelet coefficients or neural responses at the output of a multichannel decomposition. We use the notation from the previous section where a block of wavelet coefficients $\bar{C} = S\bar{U}$. While dividing by S results in the vector \bar{U} and the dependence structure within it, conditioning on S statistically achieves the same objective, leaving us with just the dependence in \bar{U} . Therefore, divisive normalization approximates conditioning and vice versa. Reduction in the dependence of local responses is also an important perceptual phenomenon in the human visual system which can be accomplished by both divisive normalization and conditioning.

The difference of weighted entropies also ensures numerical stability of the distortions at locations where the energy of the coefficients is extremely small, especially when the neural noise variance is also small. The weights at these locations tend towards zero, thereby saturating the difference between the reference and the distorted images at these locations. This effect is similar to the role that saturation

constants perform in the divisive normalization of errors [26]. We hypothesize that the neural noise variance is also related to the saturation constants in contrast sensitivity models. While saturation constants achieve numerical stability in areas of low signal energy, they also saturate the responses to zero in such areas. The neural noise variance precisely achieves this while computing the local entropy terms. These ensure that the local entropy terms are stable yet small in areas of low signal energy.

C. Suprathreshold Effects

Suprathreshold effect refers to the phenomenon by which distortions are perceivable only if they are at or above a threshold distortion level. Further, variations in the level of suprathreshold distortions decreases as the degree of distortion increases. The logarithm operation in the algorithm essentially accomplishes this desirable property. Note that the entropy of Gaussian random variables is expressed by the logarithm of the variance. Although we compute the logarithm of local variances of the reference and distorted image coefficients, scale them appropriately and then compute the difference, this is equivalent to computing the logarithm of the ratio of variances raised to powers (corresponding to the weights). Thus the ratio of variances raised to powers is the underlying distortion measure on which a nonlinear logarithmic function is applied.

D. Error Pooling

The error pooling strategy in these algorithms is a two step process depending on the amount of reference information on which the algorithms operate. The algorithms which send all the entropy terms from the reference image use a single pooling mechanism in which the absolute values of the differences of scaled entropies are averaged in the last step. The single number algorithms also follow a single pooling strategy where all the scaled entropy terms are first averaged and then the absolute values of these differences are calculated. In between the two extremes, the algorithms use a two step pooling strategy where some local entropy terms are pooled first and another pooling is performed on the absolute differences of the pooled entropy terms of the first stage. Pooling strategies in both stages can be thought of as a Minkowski error pooling strategy with exponent 1. The Minkowski pooling strategy is popularly used in human visual system based approaches to image quality assessment such as [26], although there is no evidence to support that such a pooling method is a good model for the aggregation of information in visual area MT (middle temporal) of the human visual system.

V. RESULTS AND DISCUSSION

We conducted experiments on the two largest and best-known image quality databases: LIVE Image Quality Assessment Database [34] and the Tampere Image Database [35] of distorted images and perceptual scores. The LIVE database contains five different distortion categories including JPEG2000, JPEG, additive white Gaussian noise, Gaussian blur and bit errors due to transmission of JPEG2000 images over a fast fading channel. There are a total of 779 distorted images across all distortion categories. Note that the reduced reference algorithm operates without knowledge of the distortion type. We present the results of various algorithms belonging to the framework described in this paper. One class of algorithms uses all of the entropy terms in every subband to compute quality. In another class of algorithms, the weighted entropies of every subband are filtered and subsampled at different rates, producing different algorithms for the same subband.

Both the reference and distorted images are decomposed into different subbands using a steerable pyramid wavelet decomposition using 6 orientations at 4 scales [24]. Thus there are a total of 25 subbands in the wavelet decomposition. The algorithm was implemented using blocks of size 3×3 in each subband, implying a value of $N = 9$. However, the algorithms are robust to the exact choice of the size of neighborhoods and we get approximately the same performance for different sizes.

In Table I, we show the performance results of the algorithm obtained by computing $RRED_k^{M_k}$ for all the vertical oriented subbands at different scales. The analysis in [22] suggests that human subjects are more sensitive to horizontal and vertical orientations than others. Further, we observe that the performance obtained by choosing the vertical subbands is marginally better than the horizontal subbands. The vertically oriented subbands are indexed by $k = 4, 10, 16, 22$ from the coarsest to the finest scale, i.e. at levels 1 through 4. We choose $\sigma_W^2 = 0.1$. Note that this is the same value of neural noise variance that has been used in prior algorithms [3] and we do not imply any training on the database. The performance of the algorithms in the RRED framework has proved to be robust to the choice of the neural noise variance, especially in the regime of high information from the reference.

In Table I, the Spearman rank order correlation coefficient (SROCC) between the scores of one class of the RRED indices and subjective (DMOS) scores from the LIVE image database are shown. The SROCC helps analyze how well the prediction monotonicity is preserved between the subjective ratings of quality and the scores output by the RR QA algorithm. The performance is shown for each distortion as well as over all on the database. A comparison is also drawn between the amount of information required from the reference for computation of the quality indices for the respective algorithms. Throughout this

section, L refers to the size of the image in pixels. The row labeled ‘No. of scalars’ indicates the amount of information required from the reference for quality computation. We also included the performance of FR QA algorithms such as PSNR and VIF for comparison. One important observation is that some of the RRED indices *perform nearly as well as the best performing FR QA algorithms* such as VIF. Further, most of the RRED indices considerably outperform PSNR for all distortion types (except images distorted by Gaussian noise) as well as on the entire database. Even for images distorted by Gaussian noise, the performance of the RRED indices is comparable with that of PSNR, which is an FR algorithm. The variation in performance of the RRED indices at different scales reveals that even though evaluating the quality in a coarser subband needs less information, it could potentially outperform the algorithm evaluated in a finer scale using more information. This suggests that transmitting the right information is crucial to obtain high performance QA algorithms.

The correlation analysis reveals that subbands at certain scales are more sensitive to particular distortions or artifacts. While blur and fast fading errors are better captured by subbands at finer scales, the coarser scales perform better for compressed and noisy images. This is because edge like artifacts introduced by these distortions are reflected better in the coarser subbands. The performance of fast fading errors goes down with the fineness of the scale even though the overall performance on the database is still very good. Given an application, a user could evaluate the algorithm in a subband or subbands that yield the best performance.

It is worth mentioning that an understanding of how subjective scores change with changes in the viewing distance or the size of the images is very limited. The work in [36] represents a subjective study on viewing distance conditions for JPEG and JPEG2000 images, while the work in [37] analyzes

TABLE I
SROCC BETWEEN RRED INDICES AT DIFFERENT SCALES AND LIVE IMAGE QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	$RRED_4^{M_4}$	$RRED_{10}^{M_{10}}$	$RRED_{16}^{M_{16}}$	$RRED_{22}^{M_{22}}$	PSNR	VIF (FR)[3]
JPEG2000	0.9536	0.9631	0.9580	0.9363	0.8951	0.9696
JPEG	0.9772	0.9777	0.9759	0.9405	0.8812	0.9846
AWGN	0.9763	0.9769	0.9780	0.9779	0.9853	0.9858
Gaussian Blur	0.9221	0.9595	0.9678	0.9236	0.7812	0.9728
Fast fading errors	0.7549	0.8523	0.9427	0.9377	0.8904	0.9650
Overall	0.8964	0.9343	0.9429	0.9149	0.8754	0.9636
No.of scalars	L/576	L/144	L/36	L/9	L	L

the effect of downsampling the image, prior to objective quality assessment. We remark that it is indeed possible that depending on the viewing conditions (in particular, the distance at which images are viewed), RRED evaluated at different scales could be more perceptually relevant.

We show the linear correlation coefficient (LCC) scores obtained between DMOS (difference of mean opinion scores) or subjective scores of quality and the RRED indices in Table II. A logistic nonlinearity is applied to the RRED indices before computing the linear correlation between the quality scores of the algorithm and the subjective scores available with the database. The nonlinearity relation is described by

$$\text{Quality}(x) = \beta_1 \text{logistic}(\beta_2, x - \beta_3) + \beta_4 x + \beta_5$$

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\tau x)}.$$

We observe the same trends in Table II as in Table I for SROCC. We also show the logistic fit of scores of the objective algorithms listed in the table in Fig 2. The flattening of the curves from (a) to (d) indicates that the RRED indices at coarser scales tend to cluster the quality indices of most images close to zero. As we move from (a) to (d), the RRED indices are now evaluated in finer scales, which provide a better separation of quality scores. The outliers in (a), (b) and (c) correspond to images containing a large area of high frequency textured regions. Since (a), (b) and (c) are evaluations of the RRED indices at coarser scales, they do not accurately capture the degradation in quality in these high frequency regions leading to deviations in the logistic fit.

TABLE II

LCC BETWEEN RR ALGORITHMS AT DIFFERENT SCALES AND LIVE IMAGE QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	$RRED_4^{M_4}$	$RRED_{10}^{M_{10}}$	$RRED_{16}^{M_{16}}$	$RRED_{22}^{M_{22}}$	PSNR	VIF (FR)[3]
JPEG2000	0.9600	0.9688	0.9629	0.9401	0.8995	0.9476
JPEG	0.9819	0.9820	0.9793	0.9457	0.8899	0.9873
AWGN	0.9813	0.9838	0.9845	0.9682	0.9861	0.9883
Gaussian Blur	0.9318	0.9661	0.9698	0.8980	0.7837	0.9745
Fast fading errors	0.7838	0.8688	0.9413	0.9220	0.8897	0.9696
Overall	0.9066	0.9349	0.9385	0.9099	0.8723	0.9604
No.of scalars	L/576	L/144	L/36	L/9	L	L

The algorithm evaluated in subband 10 yields excellent overall performance (in terms of both SROCC and LCC) as well as for compression, noise and blur distortions. The only drawback of the algorithm is the reduced performance in the category of images corrupted by fast fading errors. The overall performance

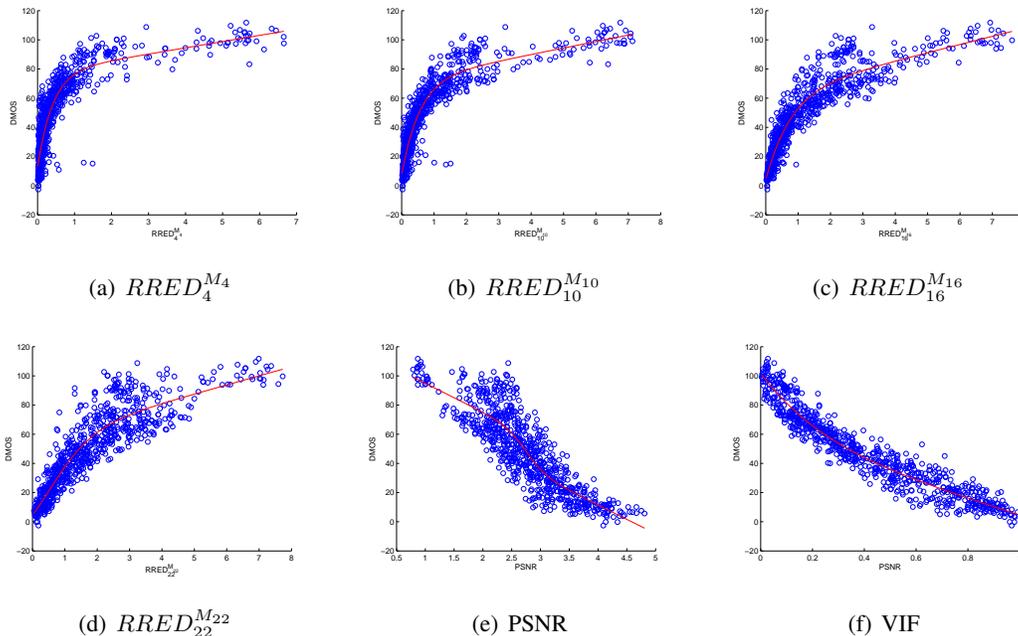


Fig. 2. Logistic fits used to compute linear correlation coefficient on the overall LIVE Image Quality Assessment Database for different objective algorithms

decays in the subband at the finest scale while the performance in the individual categories is still very good. Also, we observe that the performance of the subband at the second finest scale (subband 16) is uniformly very good for all distortion types and on the overall database. We now study the effect of filtering and subsampling the scaled entropies on subband 16. Similar trends are observed for the other subbands as well.

TABLE III
EFFECT OF FILTERING AND SUBSAMPLING ON RR ALGORITHMS - SROCC

Distortion Type	$RRED_{16}^{M_{16}}$	$RRED_{16}^{M_{16}/4}$	$RRED_{16}^{M_{16}/16}$	$RRED_{16}^{M_{16}/64}$	$RRED_{16}^{M_{16}/256}$
JPEG2000	0.9580	0.9600	0.9611	0.9580	0.9490
JPEG	0.9759	0.9760	0.9766	0.9726	0.9623
AWGN	0.9780	0.9748	0.9688	0.9626	0.9502
Gaussian Blur	0.9675	0.9429	0.9674	0.9679	0.9655
Fast fading errors	0.9427	0.9380	0.9338	0.9275	0.9202
Overall	0.9429	0.9359	0.9169	0.8865	0.8543
No.of scalars	L/36	L/144	L/576	L/2304	L/9296

Table III clearly demonstrates the degradation in the performance of the RR algorithms with sub-

sampling. Observe that even though the overall correlations are reduced by increases in the degree of subsampling, the correlation score for each individual distortion category remains very competitive with state of the art FR algorithms. The reduced performance on the overall database can be attributed to the different ranges of quality scores that occur for different distortion types. Certain distortions such as JPEG and JPEG2000 lead to different locations in the same subband having an increase and decrease in entropy simultaneously. For example, in JPEG, increases in entropy occur due to the introduction of discontinuous blocking artifacts, while decreases in entropy occur in smoother regions that are heavily quantized. As a result, when weighted entropies are summed up, the gain and loss of entropies tend to cancel each other leading to lower quality ranges. When the entropies are summed up after the absolute value of the difference is calculated for every block, the changes in the entropies are still preserved leading to an algorithm that possesses better overall performance.

As a limiting case of the above analysis, we analyze the performance of RRED indices that require only a single number from the reference. It turns out that these algorithms are more sensitive to the choice of the neural noise variance. Thus, we study the variation in performance of the single number algorithms as a function of the neural noise variance in Table IV.

TABLE IV
EFFECT OF NEURAL NOISE ON RRED INDICES - SROCC

Distortion Type	$\sigma_W^2 = 0$	$\sigma_W^2 = 0.1$	$\sigma_W^2 = 1$
JPEG2000	0.9468	0.9514	0.9455
JPEG	0.9113	0.9152	0.9179
AWGN	0.9452	0.9447	0.9314
Gaussian Blur	0.9698	0.9038	0.6108
Fast fading errors	0.9181	0.9183	0.8773
Overall	0.7682	0.7978	0.8877

Observe that when $\sigma_W^2 = 0$, the RRED indices simply compute the difference between the scaled entropies of the reference and distorted wavelet coefficients. While the overall performance increases as σ_W^2 increases, the performance within the category of images distorted by Gaussian blur decreases. Thus it is possible to trade off these performances and achieve a desirable operating point by choosing the corresponding value of σ_W^2 . This suggests that the neural noise variance of our model does impact the perception of blur apart from being able to better align scores belonging to different distortion categories. We could improve the overall performance for each value of σ_W^2 by sending one number

per subband of the wavelet decomposition. For example, by using 4 numbers from the four vertical subbands at different scales of the reference, we can improve the overall SROCC for $\sigma_W^2 = 0.1$ to 0.8606 by weighting the bands from coarser scales to the finer scales in the ratio 8:4:2:1. This algorithm still achieves excellent performance within each distortion category. Let $RRED^*$ denote this algorithm. We compare this algorithm against other popular RR QA algorithms in Table V.

TABLE V
COMPARISON OF $RRED^*$ WITH OTHER RR QA ALGORITHMS ON LIVE IMAGE DATABASE - SROCC

Distortion Type	$RRED^*$	Curvelet [15]	HWD2[15]	WNISM[12]
JPEG2000	0.9495	0.9170	0.9362	0.9135
JPEG	0.9512	0.9288	0.9543	0.9069
AWGN	0.9664	0.9585	0.9321	0.8703
Gaussian Blur	0.9453	0.9131	0.8282	0.9147
Fast fading errors	0.9310	0.9378	0.9386	0.9229
Overall	0.8606	0.9104	0.9418	0.7651
No. of scalars	4	24	16	18

We now present performance results for a couple of algorithms belonging to the framework of RRED indices on the Tampere Image Database 2008 (TID2008) [35]. TID2008 contains 17 types of distortions across 1700 distorted images. We use the same value of the neural noise $\sigma_W^2 = 0.1$ as used in all of the results presented in this section (other than the “single number” algorithms) on the LIVE Image Database. The results presented below demonstrate that the parameters of the algorithm do not require any training on databases. Table VI contains results on the SROCC obtained over all on the database while Table VII provides the SROCC for each type of distortion on the TID. $RRED_{16}^{M_{16}}$ performs better than some of the best full reference algorithms while $RRED_{22}^1$ is competitive with PSNR and outperforms [12].

We conclude this section with a brief note on the computational complexity of the RRED algorithms. The steerable pyramid decomposition of the image has a computational complexity of $O(N \log N)$ arithmetic operations per scale, where N is the total number of pixels in the image. The following calculations represent the computational cost per scale. The estimation of \mathbf{K}_U requires $O(N)$ operations, while the singular value decomposition of \mathbf{K}_U requires $O(M^3)$ operations, where \mathbf{K}_U is of size $M \times M$. Note that M represents the number of elements in a block of size $\sqrt{M} \times \sqrt{M}$. Further, the estimation of local variance parameters (premultipliers) requires $O(NM^2)$ operations, while sending the desired information from the reference or distorted image can be upper bounded by a cost of $O(N)$. Thus the

TABLE VI
OVERALL PERFORMANCE RESULTS ON TID2008

Algorithm	SROCC
MSSIM	0.853
$RRED_{16}^{M_{16}}$	0.824
VIF	0.750
VSNR	0.705
PSNR	0.553
$RRED_{22}^1$	0.521
WNISM [12]	0.512

TABLE VII
DISTORTION WISE PERFORMANCE RESULTS ON TID2008 - SROCC

Distortion type	$RRED_{16}^{M_{16}}$	MSSIM	VIF	$RRED_{22}^1$	PSNR	WNISM
Additive Gaussian noise	0.820	0.809	0.880	0.702	0.908	0.603
Additive noise in color components is more intensive than additive noise in the luminance component	0.850	0.806	0.876	0.684	0.897	0.604
Spatially correlated noise	0.842	0.820	0.870	0.712	0.917	0.599
Masked noise	0.833	0.816	0.868	0.744	0.851	0.633
High frequency noise	0.901	0.869	0.908	0.794	0.927	0.708
Impulse noise	0.741	0.687	0.833	0.549	0.872	0.593
Quantization noise	0.831	0.854	0.780	0.591	0.870	0.619
Gaussian blur	0.957	0.961	0.954	0.935	0.870	0.871
Image denoising	0.949	0.957	0.916	0.925	0.941	0.864
JPEG compression	0.933	0.935	0.917	0.819	0.873	0.834
JPEG2000 compression	0.968	0.974	0.971	0.948	0.813	0.935
JPEG transmission errors	0.870	0.874	0.859	0.782	0.751	0.875
JPEG2000 transmission errors	0.742	0.853	0.850	0.628	0.831	0.691
Non eccentricity pattern noise	0.713	0.734	0.762	0.277	0.581	0.452
Local block-wise distortions of different intensity	0.824	0.762	0.832	0.691	0.617	0.590
Mean shift (intensity shift)	0.538	0.737	0.510	0.418	0.694	0.292
Contrast change	0.542	0.640	0.819	0.723	0.587	0.701

overall computational complexity of the RRED algorithms is $O(N(\log N + M^2))$.

The time required for computing the index is calculated by running the algorithm on an Intel Pentium 4 processor with 2 GB RAM and 3.4 GHz speed. The algorithm was simulated using MATLAB version R2008a without any optimization of the code. The total time required to compute the index can be divided into four different phases at both the reference and distorted image. The first phase involves reading the image which takes around 0.2 seconds. One of the bottle necks in the computation of the RRED indices is the time required to perform a multiscale multiorientation decomposition using steerable pyramids, which requires 1.5-1.6 seconds. Computing the scaled entropies for a subband, which is also the information required to be transmitted takes 0.9-1 seconds for the largest subband. This could be more depending on the number of subbands in which the index is evaluated. Finally, computation of the index from the numbers transmitted from the reference and distorted requires time of the order of milliseconds. Overall the algorithm requires around 2.8 seconds to transmit the desired information from each (reference or distorted) image.

VI. CONCLUSION

We studied the problem of reduced reference image quality assessment by measuring the changes in suitably weighted entropies between the reference and distorted images in the wavelet domain. A distinguishing feature of the RRED indices is that these algorithms do not depend on any parameters that need to be trained on databases. The algorithms differ in the nature of the distortion measurement (by computing quality in different orientated subbands at different scales) and the quantity of the information required from the reference to compute quality (by filtering and subsampling in every subband). When the number of scalars required is around 1/40 of the image size, the algorithm achieves a performance which is nearly as good as the best performing full reference QA algorithms. Even when only a single scalar is obtained from the reference image, the algorithm achieves close to state of the art performance within each distortion category without knowing anything about the type of distortions that the image might have been subjected to. Moreover, the algorithms perform much better than mean squared error, which is an FR algorithm.

The overall performance of the single number algorithms may be improved further by better aligning the scores obtained for different distortion categories. This is a subject of future research. The use of a multiscale multiorientation decomposition before computing the index increases the complexity of the algorithm. Efficient implementations of this step can help reduce the time taken for implementation of the algorithm. The dependency on the viewing distance of the performance of such multiscale algorithms

also requires a better understanding. Future work also includes extending this idea to video where similar algorithms might be very useful in regimes that require a minimal amount of reference information.

REFERENCES

- [1] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? - A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [3] 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.
- [4] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
- [5] Sarnoff Corporation, "JNDmetrix technology," available at http://www.sarnoff.com/products_services/video_vision/jnd-metrix/downloads.asp.
- [6] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int Conf on Image Processing*, Rochester, NY, 2002.
- [7] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
- [8] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2.
- [9] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 587–599, May 2010.
- [10] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, Jun. 2010.
- [11] S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Amer. A*, vol. 24, pp. B42–B51, 2007.
- [12] 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. 5, pp. 1680–1689, 2006.
- [13] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal of Selected Topics in Signal Processing: Special issue on Visual Media Quality Assessment*, vol. 3, pp. 202–211, 2009.
- [14] M. Masry, S. S. Hemami, and Y. Sermadevi, "A scalable wavelet-based video distortion metric and applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 260–273, Feb. 2006.
- [15] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409–1423, 2009.
- [16] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 239–256, 2008.
- [17] K. Chono, Y. C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality assessment using distributed source coding," in *Proc. IEEE International Conference on Multimedia and Expo*, Hannover, Germany, 2008.

- [18] J. A. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, "Color distribution information for the reduced-reference assessment of perceived image quality," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1757–1769, 2010.
- [19] U. Engelke, M. Kusuma, H. J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525–547, 2009.
- [20] W. Xue and X. Mou, "Reduced reference image quality assessment based on Weibull statistics," in *Second International Workshop on Quality of Multimedia Experience*, Trondheim, 2010.
- [21] K. Seshadrinathan and A. C. Bovik, "Multi-scale and scalable video quality assessment," in *Proc. IEEE Int Conf on Consumer Electronics*, Las Vegas, NV, 2008.
- [22] 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.
- [23] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," in *Proc. IEEE Int Conf on Image Processing*, San Diego, CA, 2008.
- [24] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int Conf on Image Processing*, 1995.
- [25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, pp. 1338–1351, Nov. 2003.
- [26] P. C. Teo and D. J. Heeger, "Perceptual image distortion," *Proc. SPIE*, vol. 2179, pp. 127–141, 1994.
- [27] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [28] K. Seshadrinathan and A. C. Bovik, "Motion-tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, pp. 335–350, Feb. 2010.
- [29] G. Legge and J. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, Dec. 1980.
- [30] I. Ohzawa, G. Sclar, and R. D. Freeman, "Contrast gain control in the cat visual cortex," *Nature*, vol. 298, no. 5871, pp. 266–268, Jul. 1982.
- [31] J. Foley, "Human luminance pattern-vision mechanisms: masking experiments require a new model," *J. Opt. Soc. Amer. A (Optics and Image Science)*, vol. 11, no. 6, pp. 1710–1719, Jun. 1994.
- [32] A. Watson and J. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.
- [33] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nat. Neurosci.*, vol. 4, no. 8, pp. 819–825, Aug. 2001.
- [34] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," available at <http://live.ece.utexas.edu/research/quality>.
- [35] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.
- [36] S. H. Bae, T. N. Pappas, and B. Juang, "Subjective evaluation of spatial resolution and quantization noise tradeoffs," *IEEE Trans. Image Process.*, vol. 18, pp. 495–508, Mar. 2009.

- [37] M. D. Gaubatz and S. S. Hemami, "On the nearly scale-independent rank behavior of image quality metrics," in *Proc. IEEE Int Conf on Image Processing*, San Diego, CA 2008.