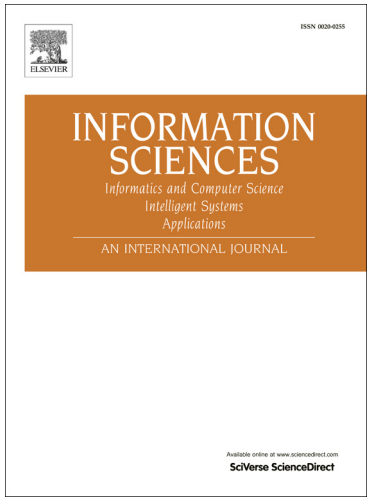

Non-Distortion-Specific no-reference image quality assessment: A survey

Redzuan Abdul Manap, Ling Shao

PII: S0020-0255(15)00007-9
DOI: http://dx.doi.org/10.1016/j.ins.2014.12.055
Reference: INS 11346

To appear in: *Information Sciences*

Received Date: 11 August 2014
Revised Date: 11 December 2014
Accepted Date: 31 December 2014

Please cite this article as: R.A. Manap, L. Shao, Non-Distortion-Specific no-reference image quality assessment: A survey, *Information Sciences* (2015), doi: http://dx.doi.org/10.1016/j.ins.2014.12.055

Information Sciences

Non-Distortion-Specific No-Reference Image Quality Assessment: A Survey

Redzuan Abdul Manap[1, 2, 4] and Ling Shao[1, 3]

[1] College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] Department of Electronic and Electrical Engineering, the University of Sheffield, Sheffield, S1 3JD, UK

[3] Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

[4] Department of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal, Melaka, 76100, Malaysia

Corresponding Author: Ling Shao. Email: ling.shao@ieee.org. Phone: +44 (0)191 349 5220

ABSTRACT

Over the last two decades, there has been a surge of interest in the research of image quality assessment due to its wide applicability to many domains. In general, the aim of image quality assessment algorithms is to evaluate the perceptual quality of an image using an objective index which should be highly consistent with the human subjective index. The objective image quality assessment algorithms can be classified into three main classes: full-reference, reduced-reference, and no-reference. While full-reference and reduced-reference algorithms require full information or partial information of the reference image respectively, no reference information is required for no-reference algorithms. Consequently, a no-reference (or blind) image quality assessment algorithm is highly preferred in cases where the availability of any reference information is implausible. In this paper, a survey of the recent no-reference image quality algorithms, specifically for non-distortion-specific cases, is provided in the first half of this paper. Two major approaches in designing the non-distortion-specific no-reference algorithms, namely natural scene statistics-based and learning-based, are studied. In the second half of this paper, their performance and limitations are discussed before current research trends addressing the limitations are presented. Finally, possible future research directions are proposed towards the end of this paper.

KEYWORDS

Image quality assessment, learning-based, natural scene statistics, no-reference image quality assessment, blind image quality assessment, non-distortion-specific.

1. INTRODUCTION

Due to tremendous growth and near-ubiquitous presence of digital images, it is becoming essential to have efficient and reliable methods to assess the quality of these images. Today, image quality assessment (IQA) has become a crucial aspect in various computer vision and image processing applications such as image acquisition, transmission, restoration and enhancement [66],[67], image search and retrieval

[25],[26],[46],[78],[79],[86], image recognition [77],[115],[116] as well as image tagging [87]. For examples, parameters of image transmission systems may have been fine-tuned to reflect the quality of the transmitted image, image quality can be utilized to rank images in image retrieval systems and image quality measures can be used in image processing algorithms for evaluation purposes [109]. In measuring the quality of an image, a human can easily give the subjective score to the observed image. However, embedding such a mechanism into computer vision or image processing systems is a difficult task. Therefore, an IQA method that can automatically provide an objective measurement which is consistent with human subjective results is highly desired.

Depending on the volume of information available about the reference image, objective IQA algorithms can be classified into three categories: full-reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA [54], which is also known as blind IQA. In FR-IQA algorithms, full information of the reference image is needed to predict the quality of degraded or distorted images. The simplest approach in implementing an FR-IQA algorithm is by measuring local pixel-wise disparity between reference and distorted images, then collapsing these measurements into a scalar representing the total or global quality metric of the distorted image. Mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are the most extensively used quality metrics in this case. However, it is generally known that they do not correlate well with subjective quality measures [21], [89]. Consequently, a large variation of more advanced FR-IQA algorithms have been proposed ranging from quality estimation based on the human visual system (HVS) to quality estimation based on image structure and image statistics.

Primary visual cortex (V1) computational neural models have been employed by most HVS-based FR-IQA algorithms. Three key stages exist in these models [6]: (1) frequency-based decomposition to model initial linear responses of visual neurons, (2) gain control adjustment taking into account non-linearity of the decomposition coefficients and (3) summation of responses to estimate the quality. Visual signal-to-noise ratio (VSNR) [7] and most apparent distortion (MAD) [38] algorithms are well known examples of HVS-based methods.

Meanwhile, an assumption that a good quality image has structure which is similar to the original image is exploited in image structure-based FR-IQA. By measuring changes in a local image's structure, such as luminance, contrast, phase or gradient, the quality of the image can be estimated. Examples of established image structure-based FR-IQA quality metrics include universal image quality index (UQI) [88], structural similarity

index (SSIM) [91], multi-scale SSIM (MS-SSIM) [95], complex wavelet SSIM (CW-SSIM) [62], feature similarity index (FSIM) [112], edge strength SSIM (ESSIM) [113], hybrid phase congruency IQA (IQA-HPC) [11], visual gradient similarity (VGS) [117] and matching pursuit based metric (MP_Q) [28].

The third approach, image statistics-based, is based on statistical measures and often supplemented by machine learning techniques. Information fidelity criterion (IFC) [71], visual information fidelity (VIF) [65], [69], singular value decomposition (SVD) [73] and machine learning image quality measure (MLIQM) [9] are the examples of current FR-IQA algorithms for this approach.

In contrast to FR-IQA algorithms, only parts of the reference image information are necessary for RR-IQA algorithms. A minimal set of reference image parameters are extracted and then used with the distorted image to predict image quality. Generally, the RR-IQA algorithms can be categorized into three groups. The first RR-IQA group is based on the image source models. These models often capture low-level statistical properties of natural images in transform domains such as discrete wavelet transform (DWT) [42],[58],[94],[96] or discrete cosine transform (DCT) [47]. The second RR-IQA algorithms class is aligned to capture image distortions. Sufficient information of the distortion process undergone by the images such as standard image or video compression is required by the algorithms in [12],[24],[36],[97] to estimate the images' quality. The final category of RR-IQA algorithms is based on the image receiver models where the physiological and/or psychophysical vision studies models are utilized [4], [5].

Higher correlation with subjective assessment of image quality is achieved by the above-mentioned algorithms when full or partial information of the reference image is available. However, in many situations, the availability of any reference information may be implausible. For example, in quality of service (QoS) monitoring of the image content transmitted over different types of network, the original signals are often not available at the middle or the end parts in the network. In photo and film restoration application, it is possible that a degraded print is the only available record of a photo or a film. In such cases, an NR-IQA algorithm is highly desired.

The objective of the NR-IQA algorithms is to estimate the quality of distorted images with respect to subjective perceptual measures without having to use any reference images. In general, these algorithms can be further classified into two categories: distortion-specific (DS) and non-distortion-specific (NDS), depending on the prior knowledge of the distortion type. Distortion that affects the image is assumed to be known in the DS NR-IQA, where it is quantified in isolation of other factors. For example, the quality of JPEG2000 compressed

images is estimated by algorithms in [43],[63],[70],[110],[111] while JPEG compressed images' quality estimation is proposed in [2],[3],[16],[22],[61],[93]. In addition, the quality of an image distorted by blocking artifacts is predicted in [56],[83],[90],[98],[102] whereas the effect of blur and noise is studied in [10],[14],[15],[17],[34],[54], [103]. Unfortunately, application domains of the algorithms might be limited by this assumption. Multiple types of distortion may present in the distorted image, thus universal or generic NR-IQA algorithms which are responsive to multiple distortions are preferred in real-world applications.

In contrast to DS NR-IQA, the prior knowledge of distortion type is not considered by NDS NR-IQA algorithms. Instead, the quality score is given through assumption that the image to be assessed has similar distortion type to those in the training database. Most of the NDS NR-IQA algorithms are designed to follow one of these two approaches: (1) natural scene statistics (NSS) based approach and (2) learning or training based approach. Towards this end, the general classification of IQA algorithms can be illustrated as in Figure 1.

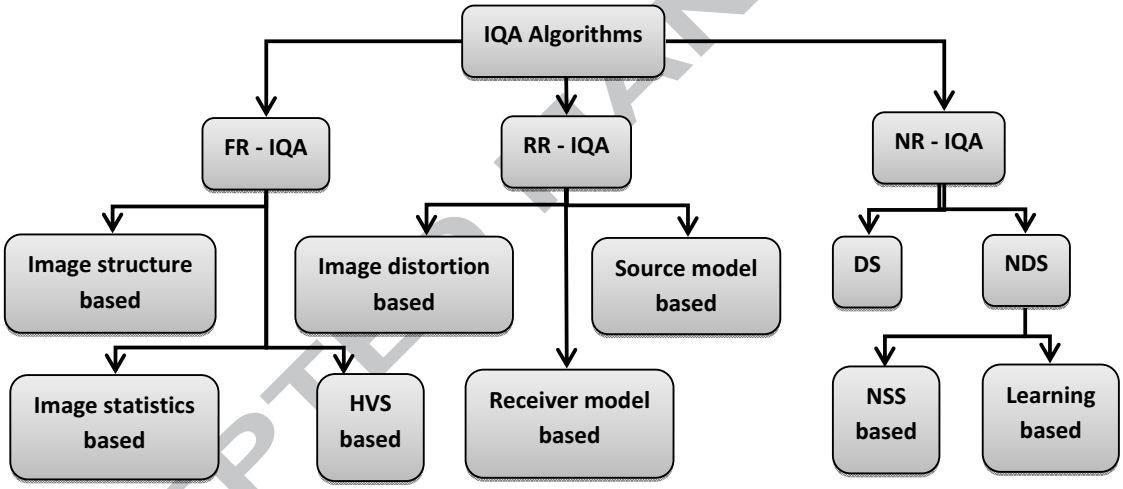

Figure 1: General classification of IQA algorithms.

This paper is meant to provide a survey of the recent advances in IQA, specifically for NDS NR-IQA algorithms. The organization of this paper is as follows. Several established NDS NR-IQA algorithms based on NSS techniques are first described in Section 2. Previous approaches based on learning techniques in designing NDS NR-IQA algorithms are then studied in Section 3. In Section 4, the performance and the limitations of the algorithms are first discussed before current research trends addressing the limitations are presented. Possible future directions are then proposed in Section 5 before the paper is concluded in Section 6.

2. NATURAL SCENE STATISTICS-BASED NON-DISTORTION-SPECIFIC NO-REFERENCE IMAGE QUALITY ASSESSMENT ALGORITHM OVERVIEW

NSS-based methods depend on an assumption that natural scenes contain certain statistical properties that will be altered by the existence of distortion. Therefore, the image quality can be predicted by obtaining features which illustrate the extent to which these statistics deviate in the distorted image. Different approaches in deriving suitable features from the statistical properties of natural images have been proposed and utilized by these NSS-based NR-IQA algorithms. The approaches can generally be classified into two major groups which are transform-based approaches and transform-free approaches.

2.1 Transform-Based Approach

2.1.1 Blind Image Quality Index Algorithm

In [51], the blind image quality index (BIQI) algorithm is proposed where the quality is estimated based on statistical features extracted using the DWT. A wavelet transform is first applied to an image over three scales and three orientations using the Daubechies 9/7 wavelet basis. Then, the resulting sub-band coefficients are modelled by a generalized Gaussian distribution (GGD) as:

$$f_X(x;\mu,\sigma^2,\gamma) = ae^{-[b|x-\mu|]^\gamma} \quad x \in \Re \tag{1}$$

$$a = \beta\gamma/2\Gamma(1/\gamma),$$

$$b = (1/\sigma)\sqrt{\Gamma(3/\gamma)/\Gamma(1/\gamma)}$$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \quad x > 0,$$

where μ, σ^2 and γ are the mean, variance and shape parameter of the distribution respectively and $\Gamma(\cdot)$ is the gamma function. The distribution parameters are estimated using a moment-based method proposed in [68]. Since wavelet bases serve as band-pass filters, zero-mean response are achieved, leaving two estimated parameters (σ^2 and γ) to be used as features at each sub-band.

The resulting 18-dimensional feature vector f_i (3 scales x 3 orientations x 2 parameters) is next used to represent the distortion in the image, i.e., the distortion identification stage. Specifically, a multi-class support vector machine (SVM) with radial-basis function (RBF) kernel is utilized with f_i as an input to estimate the presence of five different distortion categories – JPEG, JPEG2000, white noise (WN), Gaussian blur and fast fading (FF) in the image. In other words, the amount of each distortion in the image is obtained from the probability estimates $p_i, \{i = 1, \dots, 5\}$ produced by the SVM.

At the distortion-specific quality assessment stage, the computed feature vector is reused to evaluate the quality of the image at each category of these distortions where the quality is denoted as $q_i, \{i = 1, \dots, 5\}$. Prior to that, in order to obtain a quality index, a ν-SVM is first trained by quality scores from the training set to learn the relationship between the feature space and the subjective quality for each distortion. Thus, given similar image from the category that the SVM is trained for, this regression produces a representative quality score of the image. Finally, the quality of the image is represented as a probability-weighted summation:

$$BIQI = \sum_{i=1}^{5} p_i \cdot q_i \tag{2}$$

2.1.2 Distortion Identification-based Image Verity and Integrity Evaluation Algorithm

The distortion identification-based image verity and integrity evaluation (DIIVINE) algorithm, which improves upon BIQI, is later proposed in [52]. Wavelet decomposition is first performed on an image using a steerable pyramid over two scales and six orientations resulting in 12 sub-bands. In contrast to the BIQI algorithm, a divisive normalization procedure [84] is then utilized at each coefficient in each sub-band in order to account for non-linearity of the decomposition coefficients. Next, the sub-band coefficients are parameterized using GGD as in Equation (1) producing a total of 24 features (2 scales x 6 orientations x 2 parameters).

At this stage, rather than using these features straightaway, a larger set of features that indicate the dependencies between sub-band coefficients across scales and orientations are added to characterize the distortion in the image. 7 additional features are computed by utilizing the GGD fit to reflect relationship between sub-bands on the same orientation but across different scales. Another 12 features indicating the correlations across scales are also extracted. This is done by comparing each band-pass sub-band with the high-pass residual band (obtained from the steerable pyramid transform) through windowed structural correlation.

In addition, spatial correlation statistics across the sub-bands is also considered. The spatial correlation coefficient $\rho(\tau)$ is first computed and plotted for a varied distance τ. The obtained curve is then parameterized by a third order polynomial fitting. 30 more features are then formed from the polynomial coefficients and the error between the fit and the real $\rho(\tau)$. Finally, 15 more features are captured by computing windowed structural correlation between all possible pairs of sub-bands at the coarsest level. These features are meant to represent the statistical correlation that natural images exhibit across orientations. In all, a total of 88 features are extracted as summarized in Table 1. These features are then used to estimate image quality through the same two-stage classification/regression framework deployed in the BIQI algorithm.

Table 1. List of Features and Computational Methods for the DIIVINE Algorithm [52]

Feature ID	Description	Computational Method
$f_1 - f_{12}$	Variance of sub-band coefficients	Fitting GGD to sub-band coefficients
$f_{13} - f_{24}$	Shape parameter of sub-band coefficients	Fitting GGD to sub-band coefficients
$f_{25} - f_{31}$	Shape parameter across sub-band coefficients	Fitting GGD to orientation sub-band coefficients
$f_{32} - f_{43}$	Correlations across scales	Computing windowed structural correlation between filter responses
$f_{44} - f_{73}$	Spatial correlation across sub-bands	Fitting a polynomial to the correlation function
$f_{74} - f_{88}$	Across orientation statistics	Computing windowed structural correlation between adjacent orientations at the same scale

2.1.3 Blind Image Integrity Notator using DCT Statistics Algorithm

Meanwhile, an image quality estimation algorithm based on a single-stage DCT framework, known as blind image integrity notator using DCT statistics (BLIINDS), is introduced in [59]. In this algorithm, DCT is first employed to 17×17 image patches centered at each pixel in the given image. Four features are then extracted at two scales representing the information of image contrast and image structure. The first feature, i.e., the mean value of the local DCT contrast, is computed by averaging the local DCT contrast values from all image patches. The local DCT contrast is computed as the average of the ratio of the non-DC DCT coefficient magnitudes in the local patch normalized by the DC coefficient of that patch [59]. The second feature is determined by first calculating the kurtosis of each DCT image patch and then averaging the lowest tenth percentile of the computed value. The remaining two features, the DCT coefficient entropy variance across four orientations and the maximum DCT coefficient entropy across four orientations, are defined based on anisotropy measurement on each of the DCT patches.

Specifically, DCT image patches are initially generated at four different orientations: $0^{\circ}, 45^{\circ}, 90^{\circ}$ and 135°. Each DCT patch is then normalized as follows:

$$\tilde{P}_{\theta}[n,k] = \frac{P_{\theta}[n,k]^2}{\sum_k P_{\theta}[n,k]^2} \tag{3}$$

where θ is one of the four orientations, $P[n,k]$ denotes DCT coefficients of a certain patch; k indicates the DCT coefficient frequency index $(1 < k \leq 17)$, and n represents the spatial index where the patch is computed. Next, the Renyi entropy for the particular image patch is computed as:

$$R_\theta[n] = -\frac{1}{2}\log\left(\sum_k \tilde{P}_\theta[n,k]^3\right) \quad (4)$$

By defining $E[R_\theta]$ as the average per orientation for all patches of orientation θ, the variance across all four orientations and the maximum entropy across the four orientations are then determined as $var\ (E[R_\theta])$ and $\max(E[R_\theta])$ respectively [59].

Once the features have been extracted, multivariate Gaussian distribution is used as a probabilistic predictive model to train the features. By denoting the vector of extracted features and the subjective DMOS associated with image i as X_i and $DMOS_i$ respectively, the distribution of the pair $(X_i, DMOS_i)$ is modeled. In this work, the model is trained on a subset of the LIVE image database [72] to find the model parameters by distribution fitting. The model is then used to perform prediction by maximizing $P(DMOS_i/X_i)$. The operation is equivalent to maximizing the joint distribution of X and $DMOS$.

2.1.4 Improved Blind Image Integrity Notator using DCT Statistics Algorithm

Later on, the use of DCT statistics for NR-IQA is further advanced in [60]. The proposed algorithm, known as BLIINDS-II, provides improvement to the BLIINDS algorithm in the use of a general statistical model and in the performance. The framework of BLIINDS-II algorithm is illustrated in Figure 2.

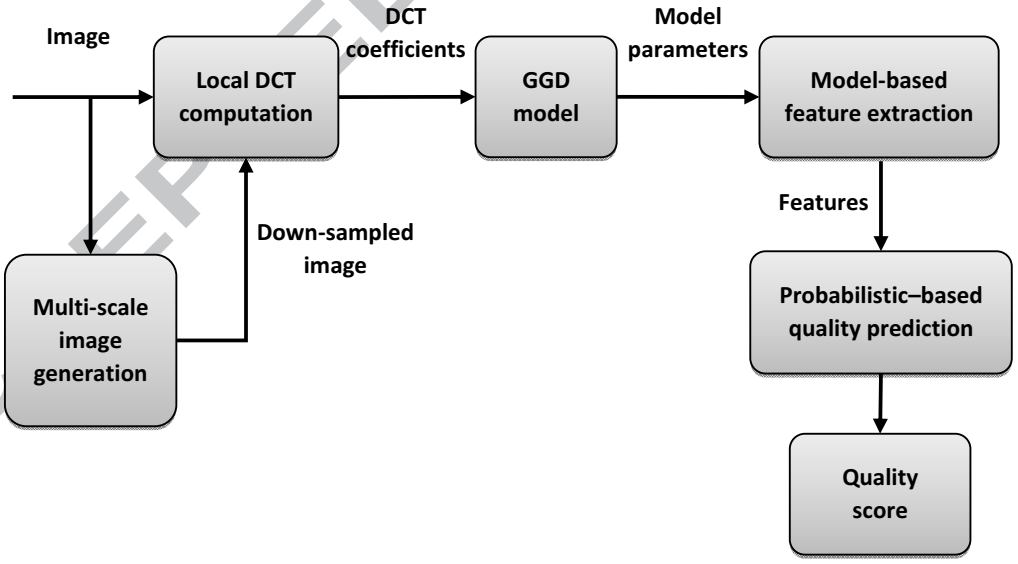

Figure 2: BLIINDS-II framework [60].

An input image is first divided into 5×5 blocks with a 2-pixel overlap between the blocks. Then, the blocks are subjected to local DCT computation where the coefficient extraction is carried out locally in the

spatial domain. The DCT block is then partitioned into two configurations: first, the block is partitioned into three oriented sub-regions to indicate directional information and second, the block is partitioned in such a way to reflect the low-frequency, mid-frequency and high-frequency sub-bands respectively.

While sample statistics such as kurtosis and entropy are used as features in the previous BLIINDS algorithm, the BLIINDS-II algorithm refines the technique by modelling image features using a generalized Gaussian distribution as in Equation (1). In other words, a generalized Gaussian density model is applied to each block as well as to particular partitions within the block at the second stage of the framework.

At the third stage of the framework, the derived model parameters are utilized to extract eight distinct features that are relevant for perceptual image quality score estimation process. The first two model-based features are derived by computing the generalized Gaussian model shape parameter γ across all blocks in the given image. The computed shape parameter is then pooled in two ways: 10% and 100% pooling, yielding the lowest 10th percentile of the shape parameter and the mean of the shape parameter as the extracted features. The third and the fourth features are based on the coefficient of frequency variation which is defined as:

$$\zeta = \frac{\sigma|X|}{\mu|X|} \tag{5}$$

where $\sigma|X|$ and $\mu|X|$ are the standard deviation and mean of the DCT coefficient magnitudes $|X|$. Again, the feature ζ is first calculated for all blocks in the image before being pooled by averaging over the highest 10th percentile and over all (100th percentile) of the local block scores across the image. The computation results in the highest 10th percentile of the coefficient of frequency variation and the mean of the coefficient of frequency variation being selected as the extracted features.

Using the second configuration of DCT block partitioning, the next two features are then designed based on local DCT energy sub-band ratio measure R_n which is defined as:

$$R_n = \frac{\left|E_n - \frac{1}{n-1}\sum_{j<n} E_j\right|}{E_n + \frac{1}{n-1}\sum_{j<n} E_j} \tag{6}$$

where $E_n = \sigma_n^2$ is the average energy in frequency band $n, n = 1,2,3$ (lower, middle, high) and σ_n^2 is the model variance corresponding to the band. The mean of R_2 and R_3 is calculated for all blocks in the image and then pooled by computing the highest 10th percentile average and the 100th percentile average of the local block scores across the image. These two energy sub-band ratio measures are then chosen as the fifth and the sixth model-based extracted features.

The remaining two features, the highest 10th percentile of the orientation feature and the mean of the orientation feature, are generated by modelling the block DCT coefficients along three orientations (the first partition configuration). The variance of ζ is first calculated at each of the three orientations. Then the variance of ζ across the three orientations from all the blocks in the image is pooled in the same way as before to produce the two extracted features.

A total of 24 model-based features, extracted over three scales, are then used at the final stage of the framework to predict the quality score. A simple Bayesian inference method, similar to the one used in the BLIINDS algorithm, is utilized to maximize the probability that the image has a certain quality score given the model-based features extracted from the image [60].

2.2 Transform-Free Approach

2.2.1 Blind / Reference Less Image Spatial Quality Evaluator Algorithm

An NSS-based NR-IQA algorithm that works in the spatial domain is presented in [48]. In contrast to previous NR-IQA algorithms, no transformation (DCT, DWT, etc.) is required for the algorithm referred as blind/reference less image spatial quality evaluator (BRISQUE). Using a spatial NSS framework, the empirical distribution of locally normalized luminance coefficients and pairwise products of these coefficients is utilized to design 18 features for image quality prediction. These features are selected based on observations that the coefficients possess statistical properties which are varied with the existence of distortion and the signs of adjacent coefficients also show a regular structure which can change in the presence of distortion. By quantifying these changes, it is assumed that the type of distortion affecting the image and its perceptual quality can be predicted.

Given an image, locally normalized luminance is first obtained by computing local mean subtraction and divisive normalization to an intensity image $I(i,j)$, producing the transformed luminance $\hat{I}(i,j)$ defined as:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{7}$$

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} \big(I_{k,l}(i,j) - \mu(i,j)\big)^2},$$

where $i \in 1,2..M, j \in 1,2..N$ are spatial indices, M, N are image height and width respectively, $C = 1$ is a constant to prevent the denominator of Equation (7) from falling to zero, $\omega = \{\omega_{k,l}\, k = -K, \dots, K, l = -L, \dots, L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled with 3 standard deviations and rescaled to unit volume and $K = L = 3$.

The normalized luminance coefficients empirical distribution is then modelled by GGD as before. Similarly, the shape parameter γ and the variance parameter σ^2 from the GGD fit of the coefficients are then estimated using the moment-based method presented in [68]. These form the first two features to be used in capturing image distortion.

The remaining features are next extracted by computing the empirical distribution of pairwise products of neighbouring normalized luminance coefficients on four orientations: horizontal, vertical, main-diagonal and secondary-diagonal as shown in Figure 3.

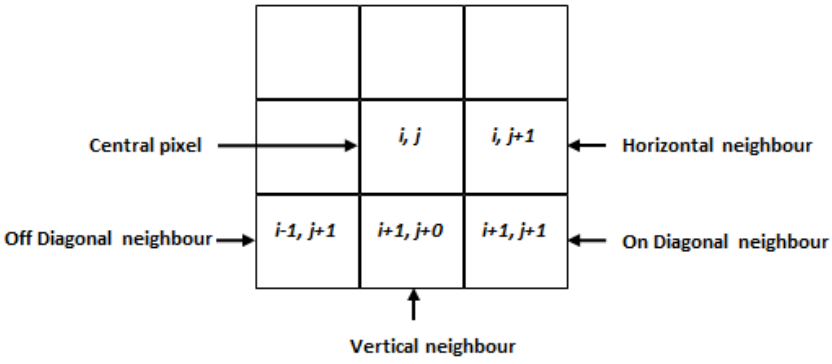

Figure 3: Four orientations' pairwise products computation in BRISQUE [48].

Instead of GGD, asymmetric generalized Gaussian distribution (AGGD) is adopted to model the distribution of the pairwise coefficients. The AGGD with the zero mode is given by [39]:

$$f(x; \nu, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l+\beta_r)\Gamma(1/\nu)} \exp(-(-x/\beta_l)^\nu) & x < 0 \\ \frac{\nu}{(\beta_l+\beta_r)\Gamma(1/\nu)} \exp(-(x/\beta_r)^\nu) & x \geq 0 \end{cases} \tag{8}$$

$$\beta_l = \sigma_l \sqrt{\Gamma\big((1/\nu)/\Gamma(3/\nu)\big)} \quad \beta_r = \sigma_r \sqrt{\Gamma\big((1/\nu)/\Gamma(3/\nu)\big)}$$

where ν is the shape parameter of the distribution while σ_l^2 and σ_r^2 are the left and the right variance parameters of the distribution respectively. As before, the parameters are estimated using the moment-based method. The three parameters plus the mean of the best AGGD fit are then selected as the extracted features at each

Table 2. List of Extracted Features and Computational Methods for the BRISQUE Algorithm [48]

Feature ID	Description	Computational Method
$f_1 - f_2$	Shape and variance	Fitting GGD to normalized luminance coefficients
$f_3 - f_6$	Shape, mean, left variance and right variance	Fitting AGGD to horizontal pairwise products
$f_7 - f_{10}$	Shape, mean, left variance and right variance	Fitting AGGD to vertical pairwise products
$f_{11} - f_{14}$	Shape, mean, left variance and right variance	Fitting AGGD to main-diagonal pairwise products
$f_{15} - f_{18}$	Shape, mean, left variance and right variance	Fitting AGGD to secondary-diagonal pairwise products

orientation. A total of 16 parameters (4 parameters x 4 orientations) are processed to yield the remaining set of features. The utilized features in the BRISQUE algorithm are summarized in Table 2. Overall, 36 features, which are extracted over two scales, are then utilized as input to a support vector regression (SVR) module to learn the mapping from the feature space to the image quality, yielding the final quality score. The algorithm is also tested using the aforementioned two-stage classification/regression framework deployed in BIQI and DIIVINE algorithms.

2.2.2 Derivative Statistics Based Image Quality Evaluator Algorithm

In [114], a new algorithm known as derivative statistics-based image quality evaluator (DESIQUE) expands the feature extraction stage in BRISQUE to include operation in the frequency domain. The algorithm is based on the observation that some derivative-based statistical properties in both the spatial and frequency domains will vary in the presence of distortion. By characterizing these variations, image quality can be evaluated. The feature extraction framework of DESIQUE is shown in Figure 4.

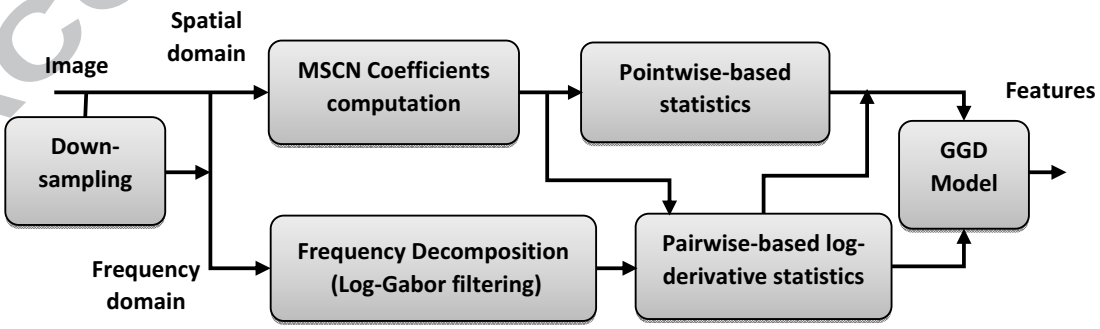

Figure 4: Feature extraction for DESIQUE [114].

The first set of features in the spatial domain, point-wise based statistics, are designed in the same way as in BRISQUE. The locally normalized luminance coefficients, also known as the mean-subtracted contrast-normalized (MSCN) coefficients, are modelled by a zero-mean GGD where its estimated parameters (shape parameter γ and variance parameter σ^2) are chosen as the extracted features.

The second set of features is developed by modelling the relationship of neighbouring MSCN coefficient pairs based on log-derivative statistics. Five types of log-derivatives representing five possible orientations of derivatives between pairs of pixels (horizontal, vertical, main-diagonal, secondary-diagonal and combined-diagonal) are first computed. Under the Gaussian coefficient model, these log-derivative values also adhere to the GGD. Therefore, the same two estimated parameters are chosen to form 10 additional features in the spatial domain (5 derivative types x 2 parameters).

In the frequency domain, an image is first decomposed using a log-Gabor filter in two orientations (horizontal and vertical) where the high spatial frequency layer of the image sub-band is used for analysis. The same five log-derivatives types are then computed. Similarly, the coefficients' log-derivative statistics for the corresponding sub-bands obey the GGD and consequently their estimated parameters yield another 20 features (2 sub-bands x 5 derivative types x 2 parameters).

Due to the facts that images are typically multi-scale and distortions change image structure across scale, the features are extracted at two scales. Thus, a total of 64 features are utilized to determine distortion and to perform distortion-specific quality assessment. The same classification / regression framework used in previous algorithms is employed by DESIQUE to predict the final quality metric.

3. LEARNING-BASED NON-DISTORTION-SPECIFIC NO-REFERENCE IMAGE QUALITY ASSESSMENT ALGORITHM OVERVIEW

In contrast to the NSS-based approaches where the relevant features are specified using natural scene statistics, the features are discovered through machine learning process in the learning-based approaches. Most of learning-based NR-IQA algorithms usually depend on a huge number of features that capture appropriate factors affecting the quality of an image. Once the features have been designed, a regression technique such as SVR or neural networks is then used to learn the mapping from the feature space to the image quality.

3.1 General Learning-based Approach

3.1.1 Leaning Quality Metric by Examples Algorithm

In [80], the quality score modelled directly through learning method is proposed. Their work is based on the assumption that similar quality images have some common traits in their low-level features and these can be learned from the given examples. General flowchart of this method is shown in Figure 5.

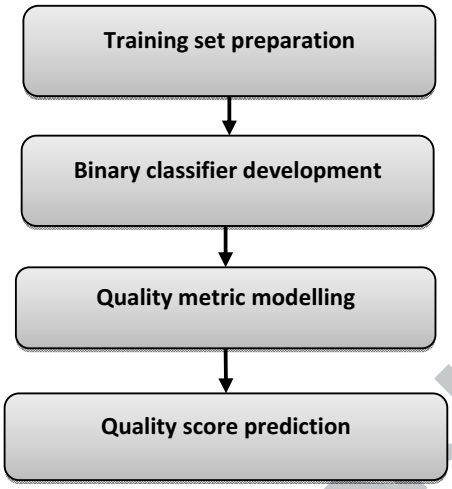

Figure 5: The flowchart of the method in [80].

First, the training set is composed of both high quality and low quality examples $E(i)$. The corresponding initial feature vector $F(i)$ is then formed by these examples as:

$$E(i) \rightarrow F(i) \quad i = 1,2,\dots,N \tag{9}$$

where N is the total number of examples.

Next, a binary classifier is developed which set apart the positive (good quality) and the negative (poor quality) examples as far as possible. Two different approaches are used to build the classifier. The first approach, which is based on a Real-AdaBoost method [18], computes both feature extraction and classifier training concurrently. In the second approach, more dominant features are first re-extracted from the initial feature vector using principal component analysis (PCA) before maximum marginal diversity (MMD) [81] being adopted in the feature selection process. The selected features are then fed to a Bayesian classifier for classification.

Once the binary classifier is developed, the quality score $Q_m(j)$ of a new example j can be modeled by its probabilities of being positive or being negative:

$$Q_m(j) = f\big(P(S^+|F(i)), P(S^-|F(i))\big) \tag{10}$$

where S^+ is the subset for positive examples; S^- is the subset for negative examples; $P(S^+|F(i))$ and $P(S^-|F(i))$ represent the posteriors of the example being positive and negative respectively obtained from the trained classifier; and $f: R^2 \rightarrow R$ is a function that maps the posteriors to a quality score. The overall distortion metric $Q_m(I)$ for an image I is then determined by averaging $Q_m(j)$ over the whole image.

Finally, the quality score $P_s(I)$ of the image can then be predicted as the following:

$$P_s(I) = \alpha + \beta . Q_m(I)^\gamma \tag{11}$$

where α, β and γ are determined by minimizing the MSE between the predicted score obtained in Equation (11) and the mean human scores:

$$MSE = \frac{1}{N_{aho}} \sum_{I=1}^{N_{aho}} (P_s(I) - Mhs(I))^2 \tag{12}$$

where $Mhs(I)$ is the mean human score for the I^{th} image and N_{aho} is the number of images used to determine the parameters.

3.1.2 Learning Based Image Quality Algorithm

Meanwhile, a learning based image quality (LBIQ) algorithm proposed in [76] utilizes a set of low-level features in a machine learning architecture to learn the mapping from these features to subjective image quality scores. These features are designed based on several main observations regarding image quality which are: a good image prior for image enhancement is a good image quality indicator; texture statistics can be a good distortion artifacts indicator; and noise and blur are the two significant degradation processes that appear in a variety of distortion types. The designed features, in the forms of natural image statistics, distortion texture statistics, or blur and noise statistics, are then combined and divided into three distinct groups of features as indicated in Table 3.

In the learning stage, PCA is first performed for each group of features to reduce the dimension of the features. The resulting low-dimensional projections are then utilized to train a ϵ-SVM regression model [64] for each group. By denoting the low dimension projection of j^{th} feature for image i as x_i^j, the following optimization problem is solved:

$$w^j = \arg\min_w \frac{1}{2}\|w\|^2 + C\sum_i \xi_i + C\sum_i \xi_i^* \tag{13}$$

subject to

Table 3. List of extracted features in LBIQ [76]

Group 1: Marginal distribution of wavelet coefficients	
Feature	Dimension No.
Negative log histogram of magnitude	720
Negative log histogram of real	720
Negative log histogram of phase	720
Maximal likelihood estimation (MLE) estimates of Generalized Gaussian distribution (GGD) parameter / likelihood of real	36
MLE estimates of Weibull distribution (WBD) parameter / likelihood of magnitude	36
Group 2:Cross-scale joint distribution of wavelet coefficients	
Feature	Dimension No.
Negative log histogram of phase	7200
Negative log histogram of magnitude	7200
Group 3: Blur / noise statistics	
Feature	Dimension No.
Patch PCA singular values	25
Negative log histogram of alpha value	10
Negative log histogram of residual	20
Step edge based blur / noise estimation	2

$$\sum_n \omega_n k\left(x_i^j, x_n^j\right) + b^j - y_i \leq \epsilon + \xi_i, \xi_i \geq 0 \tag{14}$$

$$y_i - \sum_n \omega_n k\left(x_i^j, x_n^j\right) - b^j \leq \epsilon + \xi_i, \xi_i, \xi_i^* \geq 0 \tag{15}$$

y_i is the subjective image quality of the I^{th} image (similar to $Mhs(I)$ in Equation (12)) and $k(.,.)$ is the kernel function. Here, RBF is used as a kernel. The test image quality is computed once the optimization is performed.

$$\bar{y}^j = \sum_i k\left(x^j, x_i^j\right)\omega_i^j + b^j \tag{16}$$

The results of those three respective SVM regression outputs are then combined using a weighted linear combination of the kernel SVM outputs to yield the final quality metric:

$$LBIQ = \sum_j \mu_j . \bar{y}^j \tag{17}$$

As in Equation (12), the weights of the linear combination μ_j are learned by minimizing the prediction error on the validation set:

$$\mu^* = \arg\min \sum_i (LBIQ_i - y_i)^2 \tag{18}$$

3.1.3 General Regression Neural Network Based Algorithm

Due to its excellent prediction performance, the use of general regression neural network (GRNN) in developing an NR IQA algorithm is investigated in [41]. Several image features which contain relevant information-bearing, perceptual content are first designed based on three major attributes of the image: phase congruency, entropy and gradient. The facts that most of perceptual information in an image signal is contained

in the Fourier phase [30], [55] and that perceptually meaningful image features occur at spatial locations where the important Fourier components are maximally in-phase with one another [53] are utilized in developing phase congruency features.

Meanwhile, image anisotropy, identified by the use of image entropy, has been shown to be sensitive to noise and blur [19]. Based on this finding, image entropy is also proposed as a feature to be extracted from the image. Image gradient, representing the perceptually relevant rate of change in image luminance, is another feature used in this work.

These perceptually motivated features are then used as inputs to the GRNN. Given an input vector X, the output Y of the GRNN can be calculated as [74]:

$$Y(X) = \frac{\sum_{i=1}^{n} Y_i \exp(-D_i^2/2\sigma^2)}{\sum_{i=1}^{n} \exp(-D_i^2/2\sigma^2)} \tag{19}$$

where n indicates the number of sample observations; $D_i^2 = (X - X_i)^T(X - X_i)$; X_i and Y_i are sample values; and σ is the spread parameter controlling the smoothness of functional approximation.

The proposed GRNN architecture is illustrated in Figure 6, where it consists of four layers: input layer, pattern layer, summation layer and output layer. The number of inputs in the input layer is determined by the number of extracted features at the previous stage while each node in the pattern layer represents a training pattern. Two nodes are included in the summation layer: the first node determines the numerator of Equation (19) and the second node determines the denominator of Equation (19). As indicated by Equation (19), the weight on the link between node i in the pattern layer and the first node of the summation layer is represented by Y_i while the weight on the connection between node i in the pattern layer and the second node of the summation layer is set to unity. The quotient of the two outputs of the summation layer is then computed at the output layer, producing the predicted value of the dependent feature.

3.2 Codebook-based Approach

3.2.1 Codebook Based Image Quality Algorithm

A block based NDS NR-IQA that utilizes a visual codebook technique for feature space quantization is proposed in [106]. The proposed algorithm, named codebook based image quality (CBIQ), is designed based on the assumption that distortion type and level can be attained from local image patches. Thus, a codebook of

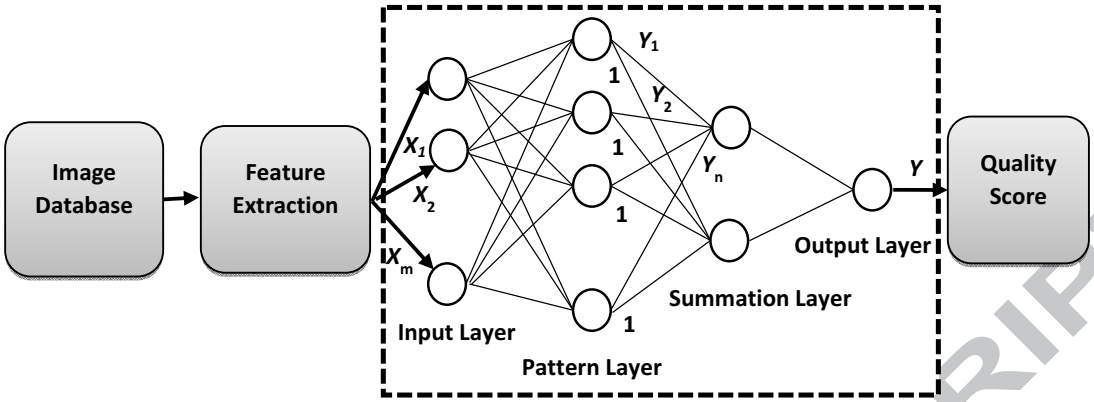

Figure 6: GRNN architecture for image quality assessment [41].

image patches is built and the codeword distribution is examined to obtain knowledge regarding the image quality.

The architecture of CBIQ consists of local feature extraction, codebook construction, image representation and regression [106]. At the local feature extraction stage, Gabor-filter-based features are adopted. It is inspired by the observation that similar quality images or images with similar type of distortion have similar texture and Gabor filters are suitable for texture representation. Specifically, simple Gabor features [37] are first obtained from each point in a given image patch using the Gabor filter. Then, the mean and variance of the magnitudes of the features across all points in the patch are calculated to produce the Gabor feature vector.

A codebook of image patches to be utilized for feature space quantization is built in the second stage. M different $B \times B$ small patches in a given training image are first randomly sampled where B is the patch size. Uniform patches are eliminated as they do not have any information useful for quality evaluation. Gabor feature vectors are then calculated for the remaining of the image patches. A Gabor feature vectors set is obtained by rerunning this step on all training images. Then the codebook C is produced from this set via a k-means clustering algorithm. Codewords $C(i), i = 1,2, \ldots\ldots, D$ are the learned cluster centres where D represents the codebook size.

At the next stage, images can be characterized by the codewords from the codebook acquired in the earlier step. This is done by deriving local descriptors, hard-assignment encoding of the descriptors and then average pooling of the encoded local descriptors, as shown in Figure 7. Given an image I, the collection of Gabor feature vectors, denoted by G_1, is extracted from M randomly sampled image patches. Next, an encoded

version of G_1 is acquired via the trained codebook and is expressed as G_2. The global feature H_I is then computed by averaging over different local descriptors. It can be treated as the normalized histogram of occurrence counts for the different codewords. In short, the probability of the codeword $C(i)$ occurrence can be approximated by the codeword histogram $H_I(i)$. H_I is later used as the input to the regression stage to estimate the quality score of the image.

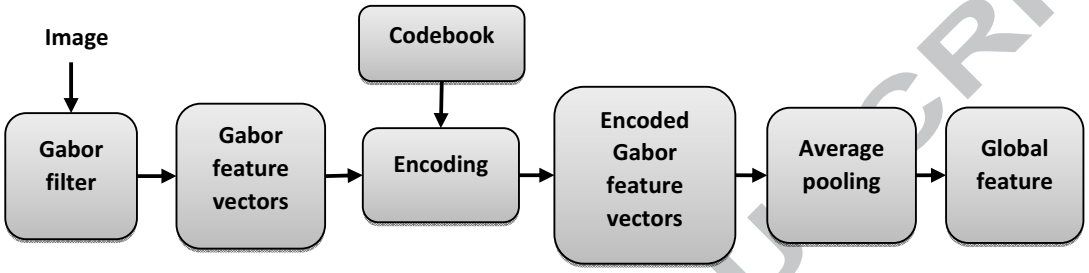

Figure 7: Image representation of CBIQ [106].

Two different regression approaches to evaluate the relationship between H_I and its associated quality score are proposed in this work. In the first approach, example-based method [105], the quality score of an image is approximated by the weighted average of the quality scores of training images. The approximation can be represented as:

$$Q_m(I) = \sum_{k=1}^{N} w_k(I) DMOS(I_k) \tag{20}$$

$$w_k(I) = \frac{1}{M} \sum_{i=1}^{M} H_I(i) \delta(C(i) \in S_k) \tag{21}$$

where $w_k(I)$ is a similarity score between testing image I and the kth training image I_k, $DMOS(I_k)$ is the subjective DMOS for I_k and S_k is the set of cluster centres obtained from the training image. $\delta(C(i) \in S_k)$ is an indicator function that is equal to 1, if $C(i) \in S_k$ is true and 0, otherwise.

Meanwhile, SVR with a linear kernel is utilized to learn the linear relationship between a codeword histogram and the quality score in the second approach. Unlike the example-based method, the quality score is not associated with a codeword and labelled training images are not required to construct the codebook. Codeword histograms are employed as an input to SVR. Using SVR, a regression function to linearly maps feature vectors to quality scores is concretely learned on training data.

3.2.2 Codebook Representation for No-Reference Image Assessment Algorithm

This visual codebook method for the NR-IQA problem is later improved by an algorithm called codebook representation for no-reference image assessment (CORNIA) [109]. In general, there are two major differences between CBIQ and CORNIA. First, instead of using hand-crafted Gabor-filter based features, raw image patches are used as features in the learning framework in CORNIA leading to more efficient and easier computation. Hence, unlabelled data available on the Internet can also be used in this unsupervised feature learning approach. Second, soft-assignment coding with max pooling is used in CORNIA rather than hard-assignment coding with average pooling at the image representation stage. This is to enable raw image patches to be used as features. Max pooling is chosen since it has demonstrated greater performance in image classification compared to average pooling [45].

3.2.3 Saliency Based Feature Learning for No-Reference Image Assessment Algorithm

Following the promising results achieved by the unsupervised feature learning approach in CORNIA, an extended algorithm is later proposed in [29]. The algorithm, named saliency-based feature learning for no-reference image quality assessment (SFLNIA), is designed to consider and acknowledge the importance of salient parts of the image. A saliency detection method is adopted prior to the feature extraction stage where the regions are found through a saliency map calculated based on low-level local image features. Raw image patches are then extracted mainly from these salient parts of the image rather than uniformly or randomly extracted from the image itself as in CORNIA. The learning framework of SFLNIA is shown in Figure 8.

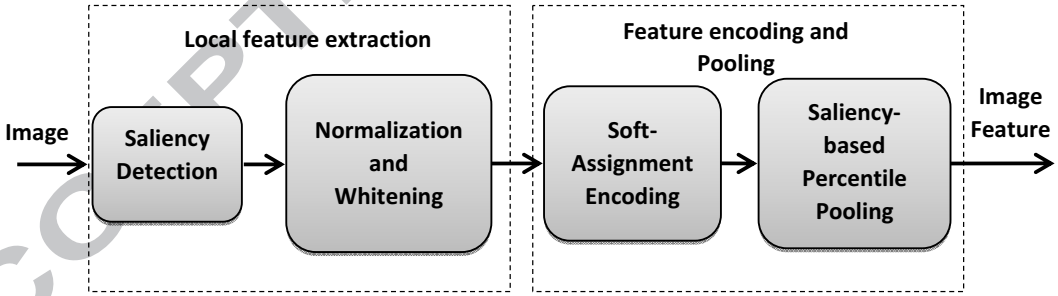

Figure 8: Learning framework of SFLNIA [29].

Specifically, the saliency map $s(x)$ is first constructed on the given image by a spectral residual approach [99]. A bilinear interpolation is then performed to produce a new saliency map $s'(x)$ that has the same size as the input image. With this map, the saliency s_{pi} is then calculated for each image patch p_i:

$$s_{pi} = \sum_{x \in p_i} s'(x), \quad i = 1,2, \dots, N \tag{22}$$

where N represents the number of image patches. Next, local descriptors $H = [h_1, h_2, \dots h_M]$ of the input image are chosen from these patches:

$$\{h_j\} = \left\{p_i | s_{pi} > \frac{1}{N}\sum_k^N s_k\right\},\ i = 1,2,..,N; j = 1,2,\dots M \quad (23)$$

The chosen patches are then normalized and whitened using zero component analysis (ZCA) at the second stage of the framework. At the third stage, similar soft-assignment coding as used in CORNIA is performed before saliency-based percentile pooling is used to produce the image level feature vector. A similar regression approach, SVR, is adopted in SFLNIA to learn the mapping between a feature vector and its associated quality score.

4. ALGORITHMS' PERFORMANCE, LIMITATIONS AND CURRENT RESEARCH TRENDS

Having described various frameworks and approaches of the state-of-the-art NDS NR-IQA algorithms in the previous two sections, in this section, we will first analyse the quality assessment performance of the algorithms. Certain limitations in those algorithms and current research trends addressing those limitations are then discussed towards the end of this section.

4.1 Performance Analysis

4.1.1 Evaluation on LIVE IQA database

In this work, most of the previously-mentioned algorithms are tested on the popular LIVE IQA database [72], which consists of 29 reference images and 779 distorted images that span various distortion categories: JPEG compression, JPEG2000 compression, WN, Gaussian blur and FF. DMOS associated with distorted images is also provided by the database. In order to evaluate algorithms' performance, the well-established Spearman's rank ordered correlation coefficient (SROCC) and the Pearson's linear correlation coefficient (LCC) between the predicted score from these algorithms and the DMOS are normally used as performance metrics. A SROCC and LCC scores close to 1 indicates better performance in terms of correlation with human perception.

Several of these NR-IQA algorithms are implemented based on their publicly available codes (BIQI, DIIVINE, BLIINDS-II, BRISQUE, CBIQ and CORNIA) while predicted quality scores from other algorithms (LBIQ, GRNN and SFLNIA) are obtained from their corresponding publications. For fair comparison between the algorithms, the LIVE database is partitioned by standard setting: 80% images are used as training set and the remaining 20% as testing set. This setting is then randomly conducted and the performance of several

Table 4. Median SROCC across 1000 Train-Test Combinations on the LIVE IQA Database

	JPEG2K	JPEG	WN	Blur	FF	All
PSNR	0.870	0.885	0.942	0.763	0.874	0.866
SSIM	0.939	0.946	0.964	0.907	0.941	0.913
MS_SSIM	0.963	0.979	**0.977**	0.954	0.939	0.954
FSIM	**0.970**	**0.981**	0.967	**0.972**	**0.949**	**0.964**
BIQI	0.799	0.891	0.951	0.846	0.707	0.820
BLIINDS-II	**0.929**	0.942	0.969	0.923	**0.889**	0.931
DIIVINE	0.913	0.910	**0.984**	0.921	0.863	0.916
BRISQUE	0.914	**0.965**	0.979	**0.951**	0.877	**0.940**
LBIQ	0.904	0.929	0.970	0.898	0.827	0.906
GRNN	0.816	0.872	**0.979**	0.833	0.735	0.827
CBIQ	0.919	**0.965**	0.933	0.944	**0.912**	0.930
CORNIA	0.943	0.955	0.976	**0.969**	0.906	**0.942**
SFLNIA	**0.951**	0.947	0.972	0.952	0.911	0.923

Table 5. Median LCC across 1000 Train-Test Combinations on the LIVE IQA Database

	JPEG2K	JPEG	WN	Blur	FF	All
PSNR	0.873	0.876	0.926	0.779	0.870	0.856
SSIM	0.921	0.955	0.982	0.893	**0.939**	0.906
MS_SSIM	**0.975**	0.979	**0.977**	0.954	**0.939**	0.954
FSIM	0.910	**0.985**	0.976	**0.978**	0.912	**0.960**
BIQI	0.809	0.901	0.954	0.829	0.733	0.820
BLIINDS-II	**0.935**	0.968	0.980	0.938	0.896	0.930
DIIVINE	0.922	0.921	**0.988**	0.923	0.888	0.917
BRISQUE	0.923	**0.973**	0.985	**0.951**	**0.903**	**0.942**
LBIQ	0.910	0.935	0.976	0.910	0.838	0.909
GRNN	0.828	0.880	**0.989**	0.825	0.819	0.837
CBIQ	0.920	**0.967**	0.954	0.949	**0.939**	0.928
CORNIA	0.951	0.965	0.987	**0.968**	0.917	**0.935**
SFLNIA	**0.957**	0.958	0.978	0.955	0.920	0.916

algorithms from Sections 2 and 3 as well as several FR-IQA algorithms is shown in Tables 4 and 5. Note that the FR-IQA algorithms are italicized while the best performing algorithms (FR-IQA and both NSS-based and learning-based NR-IQA) are in bold.

The results on the DS experiments are presented by the first five columns in these tables. This is to evaluate how well the algorithm performs in any particular distortion. The last column on these tables, representing the NDS experiment, is obtained by performing the train-test (or labelled-test) run on all distorted images regardless their distortions. As shown on these two tables, good SROCC and LCC scores are achieved by these NDS-NR-IQA algorithms indicating their acceptable performance in terms of consistency with the

human perception. In addition, most of the algorithms outperform PSNR and SSIM algorithms on NDS experiments while obtaining comparable performance to the MS-SSIM and FSIM algorithms.

Among the NSS-based NR-IQA algorithms, BRISQUE has the best DS performance when tested on JPEG, Blur and FF distorted images. In the WN and JPEG2K cases, DIIVINE and BLIINDS-II produce the best correlation scores respectively. For the NDS experiment, BRISQUE has outperformed other NSS-based NR-IQA algorithms. Meanwhile, it can be seen that CBIQ performs the best among learning-based NR-IQA algorithms in the JPEG and FF cases. When tested on WN and JPEG2K distorted images, GRNN and SFLNIA have the best performance respectively. Finally, CORNIA has the highest correlation score on blurred images as well as the best performance in the NDS experiment.

4.1.2 Statistical Significance

In terms of statistical significance, the hypothesis testing is conducted based on the one-sided t-test. It is performed using the 1000 SROCC values of all pairs of the selected algorithms at the 5% significance level. The null hypothesis is that the SROCC values for two algorithms are taken from equal-mean populations whereas the alternative hypothesis is that the mean of one algorithm is greater than that of the other algorithm. To guarantee the independency assumption of the t-test, the LIVE database is split into training and testing sets randomly.

The results of one-sided t-test are tabulated in Table 6. A score of '1' in the table implies that the algorithm in row is statistically superior to the algorithm in column, whereas a score of '-1' implies that the row algorithm is statistically worse than the column algorithm. A score of '0' indicates the row and the column algorithms are statistically equivalent. Thus, based on the table, it can be concluded that most of the recent NR-IQA algorithms are statistically superior compared to the FR-IQA algorithms PSNR and SSIM while at the same time achieving close performance to the latest FR-IQA algorithms MS-SSIM and FSIM.

4.1.3 Database Independence

Although these algorithms are normally trained and tested on a single database, i.e. the LIVE IQA dataset, most of the algorithms can also be considered database independent. Once trained, the algorithms are capable of evaluating the quality of images over the distortions that it is trained for. In this work, several NR-IQA algorithms are first trained entirely on LIVE database and then being tested on two other major databases,

Table 6. Results of One-Sided T-Test Performed between SROCC Values of Various Algorithms

	A	B	C	D	E	F	G	H	I	J	K	L	M
PSNR (A)	0	-1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	-1
SSIM (B)	1	0	-1	-1	1	-1	-1	-1	1	1	-1	-1	-1
MS-SSIM (C)	1	1	0	-1	1	1	1	1	1	1	1	1	1
FSIM (D)	1	1	1	0	1	1	1	1	1	1	1	1	1
BIQI (E)	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1
BLIINDS-II (F)	1	1	-1	-1	1	0	1	-1	1	1	1	-1	1
DIIVINE (G)	1	1	-1	-1	1	-1	0	-1	1	1	-1	-1	-1
BRISQUE (H)	1	1	-1	-1	1	1	1	0	1	1	1	-1	1
LBIQ (I)	1	-1	-1	-1	1	-1	-1	-1	0	1	-1	-1	-1
GRNN (J)	-1	-1	-1	-1	1	-1	-1	-1	-1	0	-1	-1	-1
CBIQ (K)	1	1	-1	-1	1	-1	1	-1	1	1	0	-1	1
CORNIA (L)	1	1	-1	-1	1	1	1	1	1	1	1	0	1
SFLNIA (M)	1	1	-1	-1	1	-1	1	-1	1	1	-1	-1	0

Table 7. SROCC on the TID2008 and CSIQ Databases

ALGORITHM	TID2008	CSIQ
PSNR	0.525	0.806
SSIM	0.767	0.876
MS-SSIM	0.896	0.914
FSIM	0.881	0.924
BLIINDS-II	0.840	0.873
DIIVINE	0.891	0.828
BRISQUE	0.898	0.900
DESIQUE	0.913	0.918
CORNIA	0.890	0.885

TID2008 [57] and CSIQ [38]. The results are tabulated in Table 7 where a competitive performance is produced in comparison to FR-IQA algorithms.

4.1.4 Computational Complexity

Computational requirement is another important aspect to be considered when evaluating the performance of an NR-IQA algorithm. In this section, a brief computational complexity analysis of the tested algorithms is presented. The average processing time required by the algorithms during the test stage as well as their corresponding computational complexity is summarized in Table 8. These algorithms are tested under the MATLAB2011b programming environment on an Intel i5 3.20 GHz processor, 8 GB RAM computer.

BIQI is the fastest among the tested NR-IQA algorithms, requiring 0.076s to predict a quality score of a distorted image. Unfortunately, it has the worst accuracy performance. On the other hand, BLIINDS-II,

Table 8. Computational complexity analysis

ALGORITHM	Runtime (s)	Complexity [100]	Remarks
BIQI	0.08	$O(N)$	N: number of pixels in test image
BLIINDS-II	95.24	$O\left((1/d^2)N\log(N/d^2)\right)$	N: number of pixels in test image, d:block size
DIIVINE	28.20	$O\left(N(\log N + m^2 + N + 392b)\right)$	N: number of pixels in test image, m: neighbour size in DNT, b: number of bins in the 2-D histogram
BRISQUE	0.18	$O(Nd^2)$	N: Number of pixels in test image, d:filter window size
CBIQ	59.80	$O(Nd^2K)$	N: Number of patches in test image, d:patch size, K: codebook size
CORNIA	2.43	$O(Nd^2K)$	N: Number of patches in test image, d:patch size, K: codebook size

DIIVINE and CBIQ give more accurate predictions compared to BIQI in expense of higher computational loads. As it can be seen from the table, BRISQUE and CORNIA are the two best algorithms that have high correlation scores while at the same time requiring an acceptable runtime to process an image.

In fact, it is difficult to identify the best NDS NR-IQA algorithm as the algorithms are designed through different approaches and frameworks as well as having different types of features. Depending on application domains, having a two-stage classification / regression framework such as in BIQI and DIIVINE proves to be useful as they enable the identification of image distortions. For example, in an image restoration system, it is easier to repair a distorted image if the distortion afflicting the image is known beforehand. This is achieved at the expense of higher computational requirements involving a larger set of features and sophisticated learning mechanisms.

Another example can be seen in an application where the number of distortion types examined can be increased. In this case, an algorithm with a modular framework, such as CBIQ and LBIQ, is preferred as it can cater a higher number of distortion types. Again, this is accomplished at the expense of a higher computational load. In contrast, fast computation is essential for an IQA algorithm that is required to judge an image quality instantly such as on mobile devices. In such a scenario, BRISQUE and CORNIA are the better options.

4.2 Limitations and Current Research Trends

Based on those brief analyses, it can be concluded that highly competitive performance has been shown by these algorithms with respect to all present day FR-IQA algorithms. These results certainly are great accomplishment providing that the FR-IQA algorithms require both the reference and the distorted image to estimate image quality. They also serve as the state-of-the-arts of NDS NR-IQA research.

Having said that, there are few limitations of the algorithms need to be further addressed. In this sub-section, limitations of the NSS-based algorithms as well as current efforts in addressing them are first explained

followed by a discussion on the drawbacks of the learning-based algorithms and their corresponding current solutions. At the end of this sub-section, common limitations shared by both approaches are then described as well as latest research work to overcome them.

4.2.1 NSS-based approaches

Most of the current NSS-based NDS NR-IQA algorithms use GGD to model the non-Gaussianity (NG) of the wavelet, Gabor or DCT coefficients of natural images. In other words, the non-Gaussian marginal distribution of these coefficients is predicted using GGD. However, the prediction error introduced by GGD makes the extracted features unable to reflect the change in the NG accurately. Moreover, none of these NSS-based algorithms appear to use the exponential decay characteristic (EDC) property of natural images, which is highly correlated with image degradations during the feature extraction stage.

Potential solutions to these two limitations are recently discussed in [20]. First, instead of using GGD parameters to model the NG of the coefficients, the original marginal distribution of the coefficients is used. Such a strategy avoids the prediction error produced by GGD modelling of the NG in these algorithms. Second, based on the finding that the energy of a natural image decays exponentially as the scale becomes finer, the use of EDC statistics in designing extracted features is also exploited. Better performance is achieved through these approaches in comparison with the other NDS NR-IQA algorithms though still inferior to the best FR-IQA algorithm.

Through the discussion in Section 2, it can be observed that the statistical distributions of wavelet or DCT decomposition of images are typically being used to design the image quality metric. Having said that, other image features closely correlated to the responses of visual cortical neurons can also be explored. Subsequently, potential use of a divisive normalization transform (DNT) in designing an NR-IQA metric is investigated in [13]. This approach is motivated by the capability of DNT to mimic the nonlinear mechanism in the biological visual system. This is done by simulating the behaviour of visual neurons to derive the independent components of natural images. Qualitatively, by analysing the statistical independence between neighbouring DNT coefficients of the images, it is found that the degree of statistical independence changes in the DNT domain is in agreement with the level of distortion. Thus, useful features for quality assessment can be designed through this approach. Encouraging results are achieved by this statistical independence based learning framework suggesting its great potential to be further expanded.

4.2.2 Learning-based approaches

The success of codebook-based and learning-based approaches such as CBIQ, CORNIA and SFLNIA largely relies on using a large set of codewords at the feature extraction stage. However, the size of the codebook has a trade-off. On the one hand, a large codebook makes these algorithms perform well but computationally expensive. On the other hand, the performance of these algorithms drops dramatically when a limited set of codewords are utilized. In addition, the feature extraction and the regression model training are treated independently by these algorithms making them somewhat questionable.

To address these limitations, the use of a supervised filter learning method is recently proposed in [108]. Rather than using a large codebook, a compact set of linear filters is used for feature encoding. Specifically, each local descriptor/raw image patch is encoded by its feedback to the set of linear filters. The maximal and the minimal values of filter responses are then used as an input feature vector to the regression model. At the same time, the filters are learned in a supervised way based on back-projection where the regression model and the feature extractor are jointly optimized. This way, more discriminative features are yielded by the learned filters. Competitive performance with fast computation is achieved by this technique providing good potential of filter learning methods in designing NDS NR-IQA algorithms.

4.2.3 Common limitations of both approaches

Certain applications can cause images to be concurrently subjected to multiple types of distortions. In such cases, an IQA algorithm should consider the collective effects of these distortions on the image as well as the effects of these distortions on each other. As demonstrated in previous studies on the joint effects of these distortions on image quality [8], [33], and [44], multiple types of distortions can intuitively interact with each other when they are added to an image. Interaction with the image itself is also possible in ways that might be difficult to predict based on their physical combinations. Unfortunately, no performance evaluation against images with multiple distortions is available from the previously mentioned algorithms in Sections 2 and 3.

Initial research to address this situation is presented in [31] where a new LIVE multiply distorted image quality database is produced. Two groups of multiply distorted images, namely blur plus JPEG and blur plus noise, as well as their associated subjective quality scores are provided by the database. The performance of several IQA algorithms on this database is then analysed where unsurprising average performance is achieved. Later on, a new NR-IQA algorithm attempting to determine the quality of a multiply distorted image is presented in [23]. The algorithm is developed based on several image processing blocks to mimic the HVS image perceiving process. Specifically, the noise of the input image is first computed as a prior, followed by a

possible denoising operation according to its estimated noise level. The image is then assessed independently by both the blur and JPEG metrics before the final image quality score is attained by combining the estimated results of noise, blur and JPEG stages. Promising results are achieved where the algorithm performs better than mainstream NDS NR-IQA algorithms when tested on the database [31] though higher performance can be produced by replacing one or several parts of the algorithm with more powerful models. Thus, NR-IQA algorithms could definitely profit from additional research on these multiple distortions' effects.

Another limitation shared by those NR-IQA algorithms lies in the way that they use the same similarity measure for different features to learn the universal quality metrics. Since different features have different properties, this approach is questionable. Consequently, the idea to measure the similarity of different features using different kernels is put forward [20]. Unlike previous algorithms where a single kernel method (SVM or logistic regression) is used to learn the mapping between the feature space and the image quality, multiple kernel learning (MKL) is utilized. A better representation of the numerical connection between the features and the image quality is achieved by MKL in comparison to single kernel learning methods, indicating the great potential of MKL in developing effective NDS NR-IQA algorithms.

In evaluating the image quality, these algorithms usually learn through regression methods from human subjective scores of training samples. Thus, a huge amount of human scored images are needed for training making them dataset dependent and their results heavily dependent on the size of the training dataset. Furthermore, learning a mapping function such as SVR to map the extracted features to a single perceptual quality score makes the relationship between the features and the quality scores implicit. Consequently, those algorithms cannot produce a local quality map of the distorted image that is preferable in identifying different quality scores in different regions of an image. Towards this end, an interesting question can then be raised: can an NR-IQA algorithm be developed without using human scored images for training?

The ideas of developing such algorithms in addressing this limitation are recently explored in [49], [50], [100], [107]. In [50], probabilistic latent semantic analysis (pLSA) on the statistical features of pristine and distorted image patches is first performed. The latent quality factors are then applied to the image patches of the test image to produce a quality score. Meanwhile, a multivariate Gaussian (MVG) model developed for NSS features of sharp image regions in the undistorted images is proposed in [49]. At the testing stage, the distance between the model of the test image and the model constructed from the undistorted images is calculated and used as the quality score.

In [100], another algorithm based on a quality-aware clustering (QAC) method is investigated. Through this method, a set of quality-aware centroids is learned before being used as the codebook to deduce the quality of an image patch. At the testing stage, each of the distorted image patches is compared to the learned centroids before being assigned a quality score. Subsequently, the quality score of the overall image can be determined by totalling over all patches.

Meanwhile, an algorithm called blind learning of image quality using synthetic scores (BLISS) is proposed in [107]. As indicated by the name, the algorithm trains the quality assessment model using synthetic scores rather than using human opinion scores. It is based on the idea that the current FR-IQA metrics yield high correlation with human opinion scores, thus can be used as approximations to human opinion scores. Several state-of-the-art metrics such as gradient magnitude similarity deviation (GMSD) [101], VIF, FSIM and information content weighted SSIM (IW-SSIM) [92] are combined in an unsupervised way to produce the synthetic scores. The scores are then used to replace the human opinion scores in training standard NR-IQA algorithms.

By implementing these approaches, it is shown that an NR-IQA algorithm can be developed without having to use human scored images for training. Though slightly worse performance is produced, these approaches provide alternative future directions for NDS NR-IQA research.

5. OTHER POSSIBLE FUTURE RESEARCH DIRECTIONS

Despite tremendous success achieved by the current research, it is worth to note that further research can still be done to improve NDS NR-IQA algorithms. Apart from several possible future research directions indicated in Section 4, other alternative theories and techniques can also be explored. For example, in most of the previous NDS NR-IQA algorithms, the appropriate feature space where the input image data are mapped into is either described explicitly using hand-crafted features or determined implicitly with a kernel function. In both cases, a massive cost of computational time and expert knowledge can incur. An alternative way to do this is by automatically learning the features through deep architectures.

One advantage that a deep architecture has over a shallow architecture such as SVM is that some highly non-linear functions can be expressed more compactly in terms of the number of parameters with deep architectures. Moreover, the curse of dimensionality affecting shallow architectures is addressed by deep architectures through the use of distributed representations [1]. As a result, deep learning architectures have been successfully applied to other application domains such as audio classification [40], visual classification

tasks [82] and image retrieval [35]. Thus, the use of deep learning architectures such as deep belief network (DBN), deep kernel machine and deep convolutional network to automatically learn the features for NDS NR-IQA algorithms can be considered as one possible future research direction. The initial work utilizing deep architecture in designing NR-IQA algorithm can be seen in the two latest papers [32], [75].

Besides that, these algorithms are designed for distorted images where they are often based on the assumption that a good quality image is the one which is most identical to the original image. Yet, the concept of similarity is less applicable in the enhanced images such as artwork images [85], fused images [27] or user-generated images [104]. Thus, a different quality assessment approach is needed when dealing with these types of images.

Another possible direction for future NDS NR-IQA research is dealing with geometric changes in images. Geometric distortions such as rotation, translation, scaling or viewpoint adjustments can lead to huge variations in pixel intensities. Unfortunately, most of the current algorithms are unable to handle such geometric distortions where the quality of these geometrically altered images are estimated to be lower than the quality marked with the actual human ratings. With the emergence of applications that can produce images containing extreme changes in geometrical properties, further research in this area is clearly needed.

Finally, as these algorithms move into more mainstream applications, issues regarding their effectiveness in terms of memory requirements and operational speed become more and more important. Though they triumph in terms of estimation accuracy, often they need relatively large memory bandwidth and runtime. As it can be seen from previous discussions, most of the computation and runtime occur in computing the local statistics of the image transform coefficients. Hence, future investigation on how to accelerate image transforms and local statistics computation can definitely provide important insights for enhancing these IQA algorithms.

6. CONCLUSION

In this paper, a review of recent advances on NDS NR-IQA research is provided. Generally, these types of algorithms try to redefine the IQA tasks as a classification or regression task in which certain extracted features are used to train the classifiers/regressors. These features can be extracted based on two main approaches: via machine learning (learning-based techniques) or specified by using statistical properties of natural images (NSS-based techniques).

Several recent algorithms utilizing the statistical properties of natural images and their relationship with perceptual image quality are described in Section 2. The methods are based on a hypothesis that natural scenes exhibit certain statistical properties which will be altered by the existence of distortion. Therefore, by obtaining appropriate features which can illustrate the amount of deviation for these statistics in the distorted image, the quality can be estimated. Different natural image statistical properties in identifying suitable features to be extracted as well as different schemes in constructing quality metrics are explored and utilized by these algorithms.

Meanwhile, state-of-the-art algorithms based on learning or training are discussed in detail in Section 3 where a huge number of features are first created to acquire appropriate factors affecting the image quality. Once the features have been designed, several different regression techniques such as SVR and neural networks are then utilized to learn the feature space to the image quality mapping.

The performance of these two types of algorithms is then analysed in Section 4. Based on these analyses, one main conclusion that can be made is that these algorithms can work remarkably well at estimating image quality in such a similar way to human. Having said that, it is also noticeable that these algorithms still have some limitations before they can be considered complete. These limitations are subsequently discussed in the later part of Section 4 along with several current research works to address them.

Recognizing the fact that IQA research is far from finished, several other future research directions are proposed in Section 5. Potential use of deep learning architectures, challenges related to enhanced images and geometric changes as well as computational efficiency issues are some of identified research areas to be further investigated. In return, these efforts will undoubtedly lead to further advancements in IQA research.

REFERENCES

[1] L. Arnold, S. Rebecchi, S. Chevallier and H. P. Moisy, An introduction to deep learning, Proc. Euro. Symposium Artificial Neural Network, ESANN2011, vol. 1, Bruges, Belgium, 2011, pp. 477-488.

[2] R. V. Babu and A. Perkis, An HVS-based no-reference perceptual quality assessment of JPEG coded images using neural networks, Proc. IEEE Conf. Image Processing, ICIP2005, vol. 1, Genoa, Italy, 2005, pp. 433-436. DOI: http://dx.doi.org/10.1109/ICIP.2005.1529780

[3] T. Brandão and M. P. Queluz, No-reference image quality assessment based on DCT domain statistics, Signal Process. 88(4) (2008) 822–833. DOI: http://dx.doi.org/10.1016/j.sigpro.2007.09.017

[4] M. Carnec, P. Le Callet, and D. Barba, An image quality assessment method based on perception of structural information, Proc. IEEE Conf. Image Processing, ICIP2003, vol. 3, Barcelona, Spain, 2003, pp. 185-188. DOI: http://dx.doi.org/10.1109/ICIP.2003.1247212

[5] M. Carnec, P. Le Callet, and D. Barba, Visual features for image quality assessment with reduced reference, Proc. IEEE Conf. Image Processing, ICIP2005, vol. 1, Genoa, Italy, 2005, pp. 421-424. DOI: http://dx.doi.org/10.1109/ICIP.2005.1529777

[6] D. M. Chandler, Seven challenges in image quality assessment: Past, present, and future research, ISRN Signal Process. 2013 (2013) 1-53. DOI: http://dx.doi.org/10.1155/2013/905685

[7] D. M. Chandler and S. S. Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16(9) (2007) 2284-2298. DOI: http://dx.doi.org/10.1109/TIP.2007.901820

[8] D. M. Chandler, K. H. Lim, and S. S. Hemami, Effects of spatial correlations and global precedence on the visual fidelity of distorted images, Proc. SPIE Human Vis. Electron. Imag. 6057 (2006) 1-15. DOI: http://dx.doi.org/10.1117/12.655442

[9] C. Charrier, O. l'ezoray, and G. Lebrun, Machine learning to design full-reference image quality assessment algorithm, Signal Process.: Image Comm. 27(3) (2012) 209-219. DOI: http://dx.doi.org/10.1016/j.image. 2012.01.002

[10] M. J. Chen and A. C. Bovik, No-reference image blur assessment using multiscale gradient, Proc. Int. Workshop Quality Multimedia Experience, QOMEX2009, vol. 1, San Diego, CA, 2009, pp. 70-74. DOI: http://dx.doi.org/10.1109/QOMEX.2009.5246973

[11] Q. Chen, Y. Xu, C. Li, N. Liu, and X. Yang, An image quality assessment metric based on quaternion wavelet transform, Proc. IEEE Conf. Multimedia and Expo Workshops, ICMEW2013, vol. 1, San Jose, CA, 2013, pp. 1-6. DOI: http://dx.doi.org/10.1109/ICMEW.2013.6618378

[12] K. Chono, Y. –C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, Reduced-reference image quality assessment using distributed source coding, Proc. IEEE Conf. Multimedia and Expo, ICME2008, vol. 1, Hannover, Germany, 2008, pp. 609-612. DOI: http://dx.doi.org/10.1109/ICME.2008.4607508

[13] Y. Chu, X. Mou, W. Hong, and Z. Ji, A novel no-reference image quality assessment metric based on statistical independence, Proc. IEEE Visual Communications Image Processing, VCIP2012, vol. 1, San Diego, CA, 2012, pp. 1-6. DOI: http://dx.doi.org/10.1109/VCIP.2012.6410790

[14] A. Ciancio, A. L. N. T. da Costa, E. A. B. da Silva, A. Said, R. Samadani, and P. Obrador, No-reference blur assessment of digital pictures based on multifeature classifiers, IEEE Trans. Image Process. 20(1) (2011) 64-75. DOI: http://dx.doi.org/10.1109/TIP.2010.2053549

[15] E. Cohen and Y. Yitzhaky, No-reference assessment of blur and noise impacts on image quality, Signal, Image and Video Process. 4(3) (2010) 289-302. DOI: http://dx.doi.org/10.1007/s11760-009-0117-4

[16] H. Dong and C. Wang, A no-reference method for JPEG image compression impairment evaluation, Proc. Int. Congress Image and Signal Processing, CISP2011, vol. 3, Shanghai, China, 2011, pp. 1703-1706. DOI: http://dx.doi.org/10.1109/CISP.2011.6100467

[17] R. Ferzli and L. J. Karam, A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB), IEEE Trans. Image Process. 18(4) (2009) 717-728. DOI: http://dx.doi.org/10.1109/TIP.2008.2011760

[18] J. Friedman, T. Hastie, and R. Tibshirani, Additive logistic regression: a statistical view of boosting, The Annual of Statistics, 28(2) (2000) 337-407. DOI: http://dx.doi.org/10.1214/aos/1016218223

[19] S. Gabarda and G. Cristobal, Blind image quality assessment through anisotropy, J. Opt. Soc. Amer. A 24(12) (2007) B42-B51. DOI: http://dx.doi.org/10.1364/JOSAA.24.000B42

[20] X. Gao, F. Gao, D. Tao, and X. Li, Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning, IEEE Trans. Neural Network and Learn. Syst. 24(12) (2013) 2013-2026. DOI: http://dx.doi.org/10.1109/TNNLS.2013.2271356

[21] B. Girod, What's wrong with mean-squared error?, Digital Images and Human Vision, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993.

[22] S. A. Golestaneh and D. M. Chandler, No-reference quality assessment of JPEG images via a quality relevance map, IEEE Signal Process. Lett. 21(2) (2014) 155-158. DOI: http://dx.doi.org/10.1109/LSP.2013.2296038

[23] K. Gu, G. Zhai, M. Liu, X. Yang, W. Zhang, X. Sun, W. Chen, and Y. Zuo, FISBLIM: A Five-step blind metric for quality assessment of multiply distorted images, Proc. IEEE Workshop Signal Processing Systems, SiPS2013, vol. 1, Taipei City, Taiwan, 2013, pp. 241-246. DOI: http://dx.doi.org/10.1109/SiPS.2013.6674512

[24] I. P. Gunawan and M. Ghanbari, Reduced-reference picture quality estimation by using local harmonic amplitude information, Proc. London Communication Symposium, London, UK, 2003, pp. 137-140.

[25] J. Han, X. Ji, X. Hu, J. Han and T. Liu, Clustering and retrieval of video shots based on natural stimulus FMRI, Neurocomputing, 144 (2014) 128-137. DOI: http://dx.doi.org/10.1016/j.neucom.2013.11.052

[26] J. Han, X. Ji, X. Hu, D. Zhu, K. Li, X. Jiang, G. Cui, L. Guo, and T. Liu, Representing and retrieving video shots in human-centric brain imaging space, IEEE Trans. Image Process. 22(7) (2013) 2723-2736. DOI: http://dx.doi.org/10.1109/TIP.2013.2256919

[27] R. Hong, W. Cao, J. Pang, and J. Jiang, Directional projection based image fusion quality metric, Information Sciences (281) (2014) 611-619. DOI: http://dx.doi.org/10.1016/j.ins.2014.03.046

[28] R. Hong, J. Pan, S. Hao, M. Wang, F. Xue, and X. Wu, Image Quality Assessment based on Matching Pursuit, Information Sciences (273) (2014) 196-211. DOI: http://dx.doi.org/10.1016/j.ins.2014.03.009

[29] Z. Hong, F. Ren, and Y. Ding, Saliency-based feature learning for no-reference image quality assessment, Proc. IEEE Conf. Green Computing, Communications, Internet of Things, Cyber, Physical and Social Computing, vol. 1, Beijing, China, 2013, pp. 1790-1794. DOI: http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.329

[30] T. Huang, J. Burnett, and A. Deczky, The importance of phase in image processing filters, IEEE Trans. Acoust., Speech, Signal Process. 23(6) (1975) 529-542. DOI: http://dx.doi.org/10.1109/TASSP.1975.1162738

[31] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, Objective quality assessment of multiply distorted images, Proc. 46th IEEE Asilomar. Conf. Signals, Systems and Computers, ACSSC2012, vol. 1, Pacific Grove, CA, 2012, pp. 1693-1697. DOI: http://dx.doi.org/10.1109/ACSSC.2012.6489321

[32] L. Kang, P. Ye, Y. Li, and D. Doermann, Convolutional neural networks for no-reference image quality assessment, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2014, vol. 1, Columbus, Ohio, 2014, pp. 1-8.

[33] V. Kayargadde and J. B. Martens, Perceptual characterization of images degraded by blur and noise: experiments, J. Opt. Soc. Amer. A 13(6) (1996) 1166-1177. DOI: http://dx.doi.org/10.1364/JOSAA.13.001166

[34] K. Kemalkar and V. K. Bairaqi, A no-reference image quality assessment, Proc. IEEE Conf. Emerging Trends Computing, Communications and Nanotechnology, ICE-CCN2013, vol. 1, Tirunelveli, India, 2013, pp. 462-465. DOI: http://dx.doi.org/10.1109/ICE-CCN.2013.6528543

[35] A. Krizhevsky and G. E. Hinton, Using very deep autoencoders for content-based image retrieval, Proc. Euro. Symposium Artificial Neural Network, ESANN2011, vol. 1, Bruges, Belgium, 2011, pp. 489-494.

[36] T. Kusuma and H. –J. Zepernick, A reduced-reference perceptual quality metric for in-service image quality assessment, Proc. Joint 1st Workshop Mobile Future and Symposium Trends in Communications, vol. 1, Bratislava, Slovakia, 2003, pp. 71-74. DOI: http://dx.doi.org/10.1109/TIC.2003.1249092

[37] V. Kyrki and J. Kamarainen, Simple Gabor feature space for invariant object recognition, Pattern Recog. Lett. 25(3) (2004) 311-318. DOI: http://dx.doi.org/10.1016/j.patrec.2003.10.008

[38] E. C. Larson and D. M. Chandler, Most apparent distortion: Full-reference image quality assessment and the role of strategy, Electron Imag. 19(1) (2010) 011006.1-011006-21. DOI: http://dx.doi.org/10.1117/1.3267105

[39] N. E. Lasmar, Y. Stitou, and Y. Berthoumieu, Multiscale skewed heavy tailed model for texture analysis, Proc. IEEE Conf. Image Processing, ICIP2009, vol. 1, Cairo, Egypt, 2009, pp. 2281-2284. DOI: http://dx.doi.org/10.1109/ICIP.2009.5414404

[40] H. Lee, Y. Largman, P. Pham, and A. Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, Proc. Neural Information Processing Systems, NIPS2009, vol. 1, BC, Canada, 2009, pp. 1-15.

[41] C. Li, A. C. Bovik, and X. Wu, Blind image quality assessment using a general regression neural network, IEEE Trans. Neural Network 22(5) (2011) 793-799. DOI: http://dx.doi.org/10.1109/TNN.2011.2120620

[42] Q. Li and Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Select. Topics Signal Process. 3(2) (2009) 202-211. DOI: http://dx.doi.org/10.1109/JSTSP.2009.2014497

[43] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, No-reference perceptual image quality metric using gradient profiles for JPEG2000, Signal Process.: Image Comm. 25(7) (2010) 502-516. DOI: http://dx.doi.org/10. 1016/j.image.2010.01.007

[44] L. Linde, Similarity of distorted pictures: on the interaction between edge blur and random noise, FOA Rep. C 53004-H9, Swedish National Defense Research Institute, 1981.

[45] L. Liu, L. Wang, and X. Liu, In defence of soft-assignment coding, Proc. IEEE Conf. Computer Vision, ICCV2011, vol.1, Barcelona, Spain, 2011, pp. 2486-2493. DOI: http://dx.doi.org/10.1109/ICCV.2011.6126534

[46] W. Liu and D. Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process. 22(7) (2013) 2676-2687. DOI: http://dx.doi.org/10.1109/TIP.2013.2255302

[47] L. Ma, S. Li, F. Zhang, and K. N. Ngan, Reduced reference image quality assessment using reorganized DCT-based image representation, IEEE Trans. Multimedia 13(4) (2011) 824-829. DOI: http://dx.doi.org/10.1109/TMM. 2011.2109701

[48] A. Mittal, A. K. Moorthy, and A. C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21(12) (2012) 4695-4708. DOI: http://dx.doi.org/10.1109/TIP.2012.2214050

[49] A. Mittal, G. Muralidhar, and A. C. Bovik, Making a complete blind image quality analyser, IEEE Signal Process. Lett. 20(3) (2013) 209-212. DOI: http://dx.doi.org/10.1109/LSP.2012.2227726

[50] A. Mittal, G. Muralidhar, J. Ghosh, and A. C. Bovik, Blind image quality assessment without human training using latent quality factors, IEEE Signal Process. Lett. 19(2) (2012) 75-78. DOI: http://dx.doi.org/10.1109/LSP.2011.2179293

[51] A. K. Moorthy and A. C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17(5) (2010) 513-516. DOI: http://dx.doi.org/10.1109/LSP.2010.2043888

[52] A. K. Moorthy and A. C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20(12) (2011) 3350-3364. DOI: http://dx.doi.org/10.1109/TIP.2011.2147325

[53] M. C. Morrone and R. A. Owens, Feature detection from local energy, Pattern Recog. Lett. 6(5) (1987) 303-313. DOI: http://dx.doi.org/10.1016/0167-8655(87)90013-4

[54] N. D. Narvekar and L. J. Karam, No-reference image blur metric based on the cumulative probability of blur detection (CPBD), IEEE Trans. Image Process. 20(9) (2011) 2678-2683. DOI: http://dx.doi.org/10.1109/TIP. 2011.2131660

[55] A. V. Oppenheim and J. S. Lim, The importance of phase in signals, Proc. IEEE 69(5) (1981) 529-541. DOI: http://dx.doi.org/10.1109/PROC.1981.12022

[56] F. Pan, X. Lin, S. Rahardja, E. P. Ong, and W. S. Lin, Measuring blocking artifacts using edge direction information [image and video coding], Proc. IEEE Conf. Multimedia and Expo, ICME2004, vol. 2, Taipei, Taiwan, 2004, pp. 1491-1494. DOI: http://dx.doi.org/10.1109/ICME.2004.1394519

[57] 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, Adv. Modern Radioelectronic 10(4) (2009) 30-45.

[58] A. Rehman and Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21(8) (2012) 3378-3389. DOI: http://dx.doi.org/10.1109/TIP.2012.2197011

[59] M. A. Saad, A. C. Bovik, and C. Charrier, A DCT statistics-based blind image quality index, IEEE Signal Process. Lett. 17(6) (2010) 583-586. DOI: http://dx.doi.org/10.1109/LSP.2010.2045550

[60] 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. 21(8) (2012) 3339-3352. DOI: http://dx.doi.org/10.1109/TIP.2012. 2191563

[61] Z. Y. –Sai and C. Z. –Jun, No-reference JPEG image quality assessment based on support vector regression neural network, Proc. Int. Workshop Database Technology and Applications, DBTA2010, vol. 1, Wuhan, China, 2010, pp. 1-4. DOI: http://dx.doi.org/10.1109/DBTA.2010.5658934

[62] M .P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, Complex wavelet structural similarity: A new image similarity index, IEEE Trans. Image Process. 18(11) (2009) 2385-2401. DOI: http://dx.doi.org/10. 1109/TIP.2009.2025923

[63] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, No reference image quality assessment for JPEG2000 based on spatial features, Signal Process.: Image Comm. 23(4) (2008) 257-268. DOI: http://dx.doi.org/10.1016/j.image. 2008.03.005

[64] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, New support vector algorithms, Neural Computation 12(5) (2000) 1207-1245. DOI: http://dx.doi.org/10.1162/089976600300015565

[65] K. Seshadrinathan and A. C. Bovik, Unifying analysis of full reference image quality assessment, Proc. IEEE Conf. Image Processing, ICIP2008, vol. 1, San Diego, CA, 2008, pp. 1200-1203. DOI: http://dx.doi.org/10.1109/ICIP.2008.4711976

[66] L. Shao, R. Yan, X. Li and Y. Liu, From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms, IEEE Trans. Cybernetics 44(7) (2014) 1001-1013. DOI: http://dx.doi.org/10.1109/TCYB.2013.2278548

[67] L. Shao, H. Zhang and G. de Haan, An overview and performance evaluation of classification based least squares trained filters, IEEE Trans. Image Process. 17(10) (2008) 1772-1782. DOI: http://dx.doi.org/10.1109/TIP.2008.2002162

[68] K. Sharifi and A. Leon-Garcia, Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video, IEEE Trans. Circuits Syst. Video Technol. 5(1) (1995) 52-56. DOI: http://dx.doi.org/10. 1109/76.350779

[69] H. R. Sheikh and A. C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15(2) (2006) 430-444. DOI: http://dx.doi.org/10.1109/TIP.2005.859378

[70] H. R. Sheikh, A. C. Bovik, and L. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Process. 14(11) (2005) 1918-1927. DOI: http://dx.doi.org/10.1109/TIP.2005. 854492

[71] 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. 14(12) (2005) 2117-2128. DOI: http://dx.doi.org/10.1109/TIP.2005.859389

[72] H. R. Sheikh, Z. Wang, L. Cormark, and A. C. Bovik, LIVE Image Quality Assessment Database Release 2 [Online]. Available: http://live.ece.utexas.edu/research/quality

[73] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, An SVD-based grayscale image quality measure for local and global assessment, IEEE Trans. Image Process. 15(2) (2006) 422-429. DOI: http://dx.doi.org/10.1109/TIP. 2005.860605

[74] D. F. Specht, A general regression neural network, IEEE Trans. Neural Network 2(6) (1991) 568-576. DOI: http://dx.doi.org/10.1109/72.97934

[75] H. Tang, N. Joshi, and A. Kapoor, Blind image quality assessment using semi-supervised rectifier networks, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2014, vol. 1, Columbus, Ohio, 2014, pp. 1-8.

[76] H. Tang, N. Joshi, and A. Kapoor, Learning a blind measure of perceptual image quality, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2011, vol. 1, Colorado Springs, Colorado, 2011, pp. 305-312. DOI: http://dx.doi.org/10.1109/CVPR.2011.5995446

[77] D. Tao, X. Li, X. Wu and S. J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29(10) (2007) 1700-1715. DOI: http://dx.doi.org/10.1109/TPAMI.2007.1096

[78] D. Tao, X. Li, X. Wu and S. J. Maybank, Geometric mean for subspace selection, IEEE Trans. Pattern Anal. Mach. Intell. 31(2) (2009) 260-274. DOI: http://dx.doi.org/10.1109/TPAMI.2008.70

[79] D. Tao, X. Tang, X. Li and X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 28(7) (2006) 1088-1099. DOI: http://dx.doi.org/10.1109/TPAMI.2006.134

[80] H. Tong, M. Li, H. J. Zhang, C. Zhang, J. He, and W. Y. Ma, Learning no-reference quality metric by examples, Proc. 11th Int. Multimedia Modelling Conf., MMMC2005, vol. 1, Melbourne, Australia, 2005, pp. 247-254. DOI: http://dx.doi.org/10.1109/MMMC.2005.52

[81] N. Vasconcelos, Feature selection by maximum marginal diversity: optimality and implications for visual recognition, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2003, vol. 1, Madison, Wisconsin 2003, pp. 762-769. DOI: http://dx.doi.org/10.1109/CVPR.2003.1211430

[82] P. Vincent, H. Larochelle, I, Lajoie, Y. Bengio, and P. A. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010) 3371-3408.

[83] T. Vlachos, Detection of blocking artifacts in compressed video, Electron. Lett. 36(13) (2000) 1106–1108. DOI: http://dx.doi.org/10.1049/el:20000847

[84] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, Natural image statistics and divisive normalization: Modeling nonlinearities and adaptation in cortical neurons, Statist. Theories of the Brain, MIT Press, 2002, 203-222.

[85] M. Wang, R. Hong, X. T. Yuan, S. Yan, and T. S. Chua, Movie2Comics: Towards a lively video content presentation, IEEE Trans. Multimedia. 14(3) (2012) 858-870. DOI: http://dx.doi.org/10.1109/TMM.2012.2187181

[86] M. Wang, X. S. Hua, R. Hong, J. Tang, G. J. Qi, and Y. Song, Unified video annotation via multi-graph learning. IEEE Trans. Circuits Syst. Video Technol. 19(5) (2009) 733-746. DOI: http://dx.doi.org/10.1109/TCSVT.2009.2017400

[87] M. Wang, B. Ni, X. S. Hua, and T. S. Chua, Assistive tagging: A survey of multimedia tagging with human-computer joint exploration, ACM Computing Surveys 4(4) (2012) Article 25. DOI: http://dx.doi.org/10.1145/2333112.2333120

[88] Z. Wang and A. C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9(3) (2002) 81-84. DOI: http://dx.doi.org/10.1109/97.995823

[89] Z. Wang and A. C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26(1) (2009) 98-117. DOI: http://dx.doi.org/10.1109/MSP.2008.930649

[90] Z. Wang, A. C. Bovik, and B. L. Evans, Blind measurement of blocking artifacts in images, Proc. IEEE Int. Conf. Image Processing, ICIP2000, vol. 3. Vancouver, BC, Canada, 2000, pp. 981–984. DOI: http://dx.doi.org/10.1109/ICIP.2000.899622

[91] 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. 13(4) (2004) 600-612. DOI: http://dx.doi.org/10.1109/TIP.2003. 819861

[92] Z. Wang and Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20(5) (2011) 1185-1198. DOI: http://dx.doi.org/10.1109/TIP.2010.2092435

[93] Z. Wang, H. R. Sheikh, and A. C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, Proc. IEEE Int. Conf. Image Processing, ICIP2002, vol. 1, Rochester, NY, 2002, pp. 477–480. DOI: http://dx.doi.org/10.1109/ICIP.2002.1038064

[94] Z. Wang and E. P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, Proc. SPIE Human Vis. Electron. Imag. 5666 (2005) 149-159. DOI: http://dx.doi.org/10.1117/12.597306

[95] Z. Wang, E. P. Simoncelli, and A. C. Bovik, Multiscale structural similarity for image quality assessment, Proc. 37th IEEE Asilomar. Conf. Signals, Systems and Computer, ACSSC2003, vol. 2. Pacific Grove, CA, 2003, pp. 1398-1402. DOI: http://dx.doi.org/10.1109/ACSSC.2003.1292216

[96] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. H. Yang, and A. C. Bovik, Quality-aware images, IEEE Trans. Image Process. 15(6) (2006) 1680-1689. DOI: http://dx.doi.org/10.1109/TIP.2005.864165

[97] S. Wolf and M. H. Pinson, Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system, Proc. SPIE Multimedia Syst. and Applications II 3845 (1999) 266-277. DOI: http://dx.doi.org/10.1117/12.371210

[98] H. R. Wu and M. Yuen, A generalized block-edge impairment metric for video coding, IEEE Signal Process. Lett. 4(11) (1997) 317–320. DOI: http://dx.doi.org/10.1109/97.641398

[99] H. Xiaodi and Z. Liqing, Saliency detection: A spectral residual approach, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2007, vol. 1, Minneapolis, MN, 2007, pp. 1-8. DOI: http://dx.doi.org/10.1109/CVPR.2007.383267

[100] W. Xue, L. Zhang, and X. Mou, Learning without human scores for blind image quality assessment, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2013, vol. 1, Portland, OR, 2013, pp. 995-1002. DOI: http://dx.doi.org/10.1109/CVPR.2013.133

[101] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Trans. Image Process. 23(2) (2014) 684-695. DOI: http://dx.doi.org/10.1109/TIP.2013.2293423

[102] G. Yammine, E. Wige, and A. Kaup, A no-reference blocking artifacts visibility estimator in images, Proc. IEEE Conf. Image Processing, ICIP2010, vol. 1, Hong Kong, 2010, pp. 2497-2500. DOI: http://dx.doi.org/10.1109/ICIP.2010.5652927

[103] Q. Yan, Y. Xu, and X. Yang, No-reference image blur assessment based on gradient profile sharpness, Proc. IEEE Symposium Broadband Multimedia System and Broadcasting, BMSB2013, London, UK, 2013, pp. 1-4. DOI: http://dx.doi.org/10.1109/BMSB.2013.6621727

[104] Y.Yang, X. Wang, T. Guan, J. Shen, and L. Yu, A multi-dimensional image quality prediction model for user-generated images in social networks, Information Sciences (281) (2014) 601-610. DOI: http://dx.doi.org/10.1016/j.ins.2014.03.016

[105] P. Ye and D. Doermann, No-reference image quality assessment based on visual codebook, Proc. IEEE Conf. Image Processing, ICIP2011, vol. 1, Brussels, Belgium, 2011, pp. 3089-3092. DOI: http://dx.doi.org/10.1109/ICIP.2011.6116318

[106] P. Ye and D. Doermann, No-reference image quality assessment using visual codebooks, IEEE Trans. Image Process. 21(7) (2012) 3129-3138. DOI: http://dx.doi.org/10.1109/TIP.2012.2190086

[107] P. Ye, J. Kumar, and D. Doermann, Beyond human opinion scores: blind image quality assessment based on synthetic scores, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2014, vol. 1, Columbus, Ohio, 2014, pp. 1-8.

[108] P. Ye, J. Kumar, L. Kang, and D. Doermann, Real-time no-reference image quality assessment based on filter learning, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2013, vol. 1, Portland, OR, 2013, pp. 987-994. DOI: http://dx.doi.org/10.1109/CVPR.2013.132

[109] P. Ye, J. Kumar, L. Kang, and D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR2012, vol. 1, Providence, Rhode Island, 2012, pp. 1098-1105. DOI: http://dx.doi.org/10.1109/CVPR.2012. 6247789

[110] J. Zhang and T. M. Le, A new no-reference quality metric for JPEG2000 images, IEEE Trans. Consumer Elec. 56(2) (2010) 743-750. DOI: http://dx.doi.org/10.1109/TCE.2010.5505996

[111] J. Zhang, S. H. Ong, and T. M. Le, Kurtosis-based no-reference quality assessment of JPEG2000 images, Signal Process.: Image Comm. 26(1) (2011) 13-23. DOI: http://dx.doi.org/10.1016/j.image.2010.11.003

[112] L. Zhang, X. Mou, and D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20(8) (2011) 2378-2386. DOI: http://dx.doi.org/10.1109/TIP.2011.2109730

[113] X. Zhang, X. Feng, W. Wang, and W. Xue, Edge strength similarity for image quality assessment, IEEE Signal Process. Lett. 20(4) (2013) 319-322. DOI: http://dx.doi.org/10.1109/LSP.2013.2244081

[114] Y. Zhang and D. M. Chandler, An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes, Proc. SPIE-IS&T Elec. Imaging 8653 (2013) 86530J-1-86530J-10. DOI: http://dx.doi.org/10.1117/12.2001342

[115] Z. Zhang and D. Tao, Slow feature analysis for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 34(3) (2012) 436-450. DOI: http://dx.doi.org/10.1109/TPAMI.2011.157

[116] F. Zhu and L. Shao, Weakly-supervised cross-domain dictionary learning for visual recognition, International Journal Computer Vision (IJCV), 109(1-2) (2014) 42-59. DOI: http://dx.doi.org/10.1007/s11263-014-0703-y

[117] J. Zhu and N. Wang, Image quality assessment by visual gradient similarity, IEEE Trans. Image Process. 21(3) (2012) 919-933. DOI: http://dx.doi.org/10.1109/TIP.2011.2169971