No-reference image and video quality assessment: a classification and review of recent approaches

Muhammad Shahid*, Andreas Rossholm, Benny Lövström and Hans-Jürgen Zepernick

Abstract

The field of perceptual quality assessment has gone through a wide range of developments and it is still growing. In particular, the area of no-reference (NR) image and video quality assessment has progressed rapidly during the last decade. In this article, we present a classification and review of latest published research work in the area of NR image and video quality assessment. The NR methods of visual quality assessment considered for review are structured into categories and subcategories based on the types of methodologies used for the underlying processing employed for quality estimation. Overall, the classification has been done into three categories, namely, pixel-based methods, bitstream-based methods, and hybrid methods of the aforementioned two categories. We believe that the review presented in this article will be helpful for practitioners as well as for researchers to keep abreast of the recent developments in the area of NR image and video quality assessment. This article can be used for various purposes such as gaining a structured overview of the field and to carry out performance comparisons for the state-of-the-art methods.

Keywords: No-reference; Image quality assessment; Video quality assessment; Perceptual quality

1 Review

1.1 Introduction

There has been a tremendous progress recently in the usage of digital images and videos for an increasing number of applications. Multimedia services that have gained wide interest include digital television broadcasts, video streaming applications, and real-time audio and video services over the Internet. The global mobile data traffic grew by 81% in 2013, and during 2014, the number of mobile-connected devices will exceed the number of people on earth, according to predictions made by Cisco. The video portion of the mobile data traffic was 53% in 2013 and is expected to exceed 67% by 2018 [1]. With this huge increase in the exposure of image and video to the human eye, the interest in delivering quality of experience (QoE) may increase naturally. The quality of visual media can get degraded during capturing, compression, transmission, reproduction, and displaying due to the distortions that might occur at any of these stages.

The legitimate judges of visual quality are humans as end users, the opinions of whom can be obtained by subjective experiments. Subjective experiments involve a panel of participants which are usually non-experts, also referred to as test subjects, to assess the perceptual quality of given test material such as a sequence of images or videos. Subjective experiments are typically conducted in a controlled laboratory environment. Careful planning and several factors including assessment method, selection of test material, viewing conditions, grading scale, and timing of presentation have to be considered prior to a subjective experiment. For example, Recommendation (ITU-R) BT.500 [2] provides detailed guidelines for conducting subjective experiments for the assessment of quality of television pictures. The outcomes of a subjective experiment are the individual scores given by the test subjects, which are used to compute mean opinion score (MOS) and other statistics. The obtained MOS, in particular, represents a ground truth for the development of objective quality metrics. In ITU-R BT.500 and related recommendations, various types of subjective methods have been described. These types include either single

*Correspondence: muhammad.shahid@ieee.org
Blekinge Institute of Technology, Karlskrona SE-37179, Sweden

stimulus or double stimulus-based methods. In single stimulus methods, the subjects are shown variants of the test videos and no reference for comparison is provided. In some situations, a hidden reference can be included but the assessment is based only on a no-reference scoring of the subjects.

Due to the time-consuming nature of executing subjective experiments, large efforts have been made to develop objective quality metrics, alternatively called as objective quality methods. The purpose of such objective quality methods is to automatically predict MOS with high accuracy. Objective quality methods may be classified into psychophysical and engineering approaches [3]. Psychophysical metrics aim at modeling the human visual system (HVS) using aspects such as contrast and orientation sensitivity, frequency selectivity, spatial and temporal pattern, masking, and color perception. These metrics can be used for a wide variety of video degradations but the computation is generally demanding. The engineering approach usually uses simplified metrics based on the extraction and analysis of certain features or artifacts in a video but do not necessarily disregard the attributes of the HVS as they often consider psychophysical effects as well. However, the conceptual basis for their design is to do analysis of video content and distortion rather than fundamental vision modeling.

A set of features or quality-related parameters of an image or video are pooled together to establish an objective quality method which can be mapped to predict MOS. Depending on the degree of information that is available from the original video as a reference in the quality assessment, the objective methods are further divided into full reference (FR), reduced reference (RR), and no-reference (NR) as follows:

- FR methods: With this approach, the entire original image/video is available as a reference. Accordingly, FR methods are based on comparing distorted image/video with the original image/video.
- RR methods: In this case, it is not required to give access to the original image/video but only to provide representative features about texture or other suitable characteristics of the original image/video. The comparison of the reduced information from the original image/video with the corresponding information from the distorted image/video provides the input for RR methods.
- NR methods: This class of objective quality methods does not require access to the original image/video but searches for artifacts with respect to the pixel domain of an image/video, utilizes information embedded in the bitstream of the related image/video format, or performs quality assessment as a hybrid of pixel-based and bitstream-based approaches.

1.2 Applications of no-reference image and video quality assessment

In recent years, there has been increasing interest in the development of NR methods due to the widespread use of multimedia services in the context of wireless communications and telecommunication systems. Applications of NR methods include the following areas:

- Network operators and content providers have a strong interest to objectively quantify the level of service quality delivered to the end user and inside the network nodes. NR methods will provide the data needed to adopt network settings such that customer satisfaction is secured and hence churn can be avoided.
- The involvement of multiple parties between content providers and the end users gives rise to establish service-level agreements (SLA) under which an agreed level of quality has to be guaranteed. In this respect, NR methods are a suitable choice for in-service quality monitoring in live systems.
- In general, NR methods are well suited to perform real-time objective quality assessment where resources are limited such as frequency spectrum in wireless communications. In such cases, RR methods have limited application as an ancillary channel is required to transmit the required features of the original video.
- Real-time communication and streaming services require quality adaptations using NR methods for collecting statistics of the delivered quality.

1.2.1 Related work: published reviews of objective visual quality methods

According to the framework introduced in [4] for NR visual quality estimation, three stages are present in an NR quality estimation approach. These stages are *measurement* of a physical quantity relevant for visual quality, also called as feature, *pooling* the measured data over space and/or time, and *mapping* the pooled data to an estimate of perceived quality. A survey of the measurement stage, which is essentially the main focus in much of the work done in NR quality estimation, has been provided in the same contribution. The survey in [4] divides the literature review into two main categories. In the first category, the methods estimating mean square error (MSE) caused by block-based compression, MSE caused by packet loss errors, and noise estimation methods to compute MSE have been discussed. The second category encompasses the approaches that are termed as feature-based. The feature-based methods are based on either a model developed for particular artifacts related to a visible degradation, or a model developed to quantify the impact of degradations on a specific set of attributes of

the original uncorrupted image or video. A brief survey of NR methods of image quality assessment (IQA) based on the notion of quantifying the impact of distortions on natural scene statistics (NSS) is provided in [5]. Some NR methods of visual quality are discussed in [6] also under the categorization of features and artifacts detection. Similarly, a review of the objective methods of video quality assessment (VQA) is provided in [7] including a classification of objective methods in general without specifying it for no-reference methods. In [7], the objective methods are classified as data metrics, pictures metrics, and packet or bitstream-based metrics. The review and performance comparison of video quality assessment methods in [8] present a classification of FR and RR methods only. A survey on visual quality assessment methods that are based on information theory is given in [9]. It was observed that information theory-based research for the development of NR methods is rather limited. The type of NR methods surveyed in [9] relies on an approach that employs Rényi entropy for determining the amount of randomness in the orientation of local structures in an image. NR methods have been reviewed in [10] by classifying them following three approaches. Firstly, a review of NR methods has been performed by classifying them based on the type of distortion that is estimated to formulate a quality value. The second approach used for the classification is based on methods that are designed for quantifying the artifacts produced by a specific compression standard. Lastly, a review of methods that are not designed specifically for a particular distortion has been performed. A broad survey of image and video quality methods, as well as a classification of the methods, was published during 2007 in [11]. This includes both NR and RR methods, and our article focuses on a classification and review of NR methods of IQA and VQA published after [11].

1.2.2   Our proposed classification

The current literature in the area of methods of NR image/video quality assessment is quite diverse. Hence, it is a challenging task to classify these methods into a well-structured and meaningful categorization. A good categorization of such methods should be concise enough to be properly understandable and also comprehensive enough to present most of the relevant methodologies. The aforementioned types of classifications cover a range of NR methods, but there is a need to broaden the categorization approaches in order to review currently existing methods in this area. Reibman et al. [12] classify NR methods as either stemming from statistics derived from pixel-based features and call them NR pixel (NR-P) type or computed directly from the coded bitstream and call them NR bitstream (NR-B) type. We believe that this is a useful classification which can serve as an effective basis for constructing a broader classification.

In the case of NR-P-based methods, one relevant method to classify available approaches is to investigate these in terms of the employment of certain artifacts that are related to a specific kind of degradation of the visual quality. Quantification of such artifacts has been used as a measure for the quality assessment. The quality values may depend only on a single artifact or it may depend upon a combination of many artifacts. It is common that single artifact measure-based methods are developed by considering a given model of degradation, often simulated artifacts, and sometimes their performance remains unknown for realistic or more general scenarios. For example, most of the available blur methods are based on Gaussian or linear blur models, which may not adequately measure the blur produced by a complex relative motion between image capturing device and the object. Moreover, single-artifact-based quality methods may not have satisfactory performance in the assessment of the overall quality, in the presence of other artifacts. Therefore, methods have been introduced where estimation of a combination of artifacts is fused to generate a single quality score. Also, in the domain of NR-P-based methods, there are many methods which work beyond simple artifacts computation and the quality assessment is derived from the impact of distortions upon NSS (referring to statistical characteristics commonly found in natural images). Moreover, some quality-relevant features can be computed from the image/video pixels to formulate an estimation of the perceptual quality.

The NR-B-based methods are relatively simpler to compute than NR-P-based methods, and the quality values can often be computed in the absence of a full decoder. However, such methods can have limited scope of application as they are usually designed for a particular coding technique and bitstream format, e.g., H.264/AVC standard. Such methods are based on either the encoding information derived from the bitstream or the packet header information or a combination of both. These methods are quite suitable for network video applications such IPTV and video conferencing.

Quality assessment performance can be compromised in NR-B-based methods to gain reduction in the computational complexity as compared to the NR-P-based methods. The performance of NR-B-based methods of quality assessment can be improved by adopting an approach of adding some input from NR-P-based quality assessment. Such composites of NR-P- and NR-B-based methods are called hybrid methods. These methods inherit the computational simplicity of NR-B-based methods and depend on NR-P-related data to gain further robustness.

In light of the aforementioned discussion, our approach of a meaningful classification of NR objective visual quality methods is outlined in Figure 1. This classification is formulated by considering the type and granularity of usage of the image or video data for the design of an objective method of quality. Thus, it offers the opportunity to present a discussion of most of recently published techniques of the NR visual quality assessment. It is to be noted that the main focus of this article is to review, in a systematic and structured manner, recent advancements in this area. Hence, a performance comparison of the reviewed methods on a comprehensive test database is out of the scope of this paper.

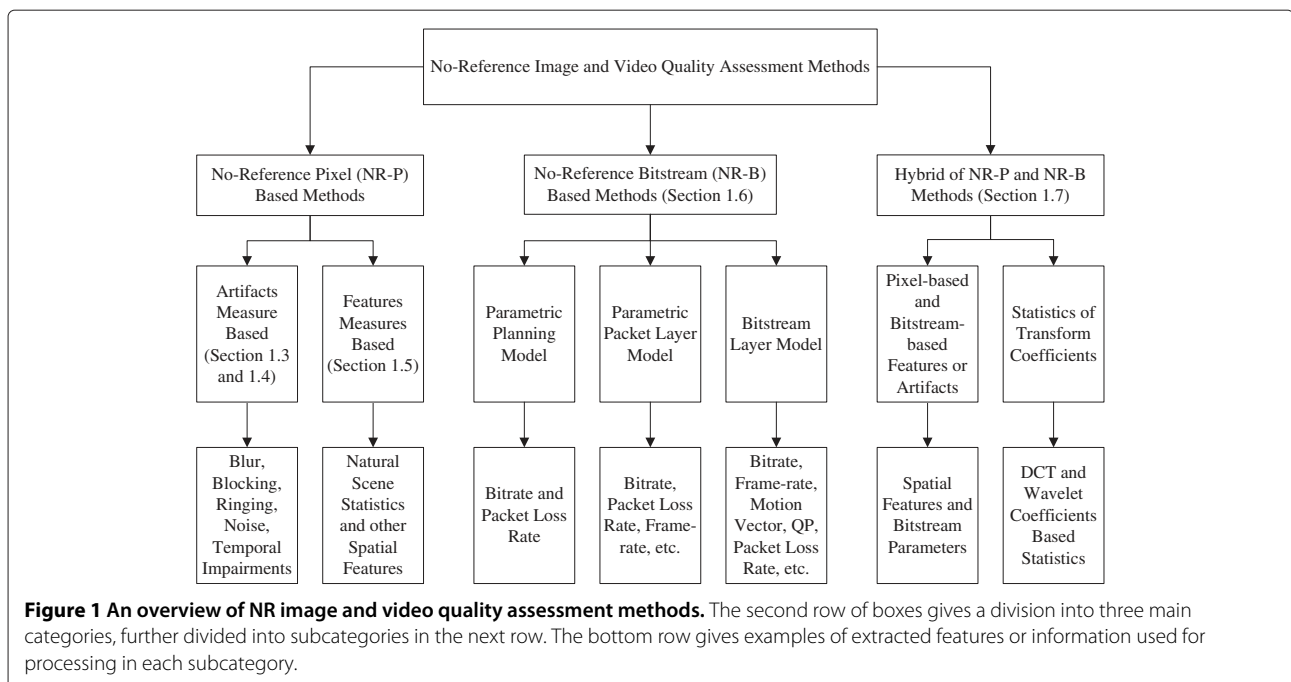### 1.2.3 The preliminaries and organization of this paper
Most of the existing NR quality methods fall into NR-P or NR-B type methods or a hybrid of these two approaches. As shown in Figure 1, the following sections present an overview of the different classes of NR methods of IQA and VQA. In each section, we have presented a general idea used in computation of various types of methods of quality estimation using block diagrams. Summaries of most of the discussed methods are shown in tables throughout the paper and in dedicated discussion sections. Mostly, the performance of an objective quality prediction model is reported by using measure of prediction accuracy, i.e., Pearson's linear correlation coefficient, and measure of monotonicity, i.e., Spearman's rank order correlation coefficient, as recommended by Video Quality Expert Group (VQEG) [13]. These measures have been used to report the performance of the reviewed methods

in the tables. In these tables, some cells have been marked with a hyphen (-) in cases where the corresponding value has not been reported in the reference or some uncommon measure of performance has been used. Other than the explicit numerical values of the number of pixels used for stating the resolution of the test data, the following short forms are used:

- QCIF, Quarter Common Intermediate Format (176×144)
- CIF,   Common Intermediate Format (352×288)
- SIF,   Standard Interchange Format (320×240)
- SD,   Standard Definition (720×480 or 720×576)
- HD,   High Definition (1920×1080 or 1280×720)

For validation of the proposed method, some publicly available databases of images and videos have been used in most of the reference papers. In this article, the reference to a public database of test media indicates that either a subset or the complete set of the available media has been used. These sources of the test media include the following:

- Laboratory for Image and Video Engineering (LIVE): LIVE offers databases of compressed images and videos with the corresponding data of the subjective assessment. The images have been encoded using Joint Photographic Experts Group (JPEG) and JPEG2000 standards. Moreover, some images have been generated using simulated conditions of certain artifacts such as Gaussian blur and white noise. The video database contains sets of videos encoded

Figure 1 An overview of NR image and video quality assessment methods. The second row of boxes gives a division into three main categories, further divided into subcategories in the next row. The bottom row gives examples of extracted features or information used for processing in each subcategory.

using Moving Picture Experts Group (MPEG)-2 and H.264/AVC. While we refer to the usage of test data from LIVE in the tables, the standard used for encoding shown in the column *Processing* indicates whether the used data is an image or a video. References to the publications based on the use of these databases have been provided at the source website [14].

- Video Quality Experts Group (VQEG): VQEG has released its test data for public use which is available on their website [15]. The data contains standard definition television videos and the corresponding values of the subjective assessment.
- Tampere Image Database 2008 (TID2008): This database contains test data produced from 17 different types of distortion introduced in the given 25 reference images. The test images have been provided with the corresponding subjective assessment scores and values of many objective methods of quality estimation. More information on it is found in [16].
- Images and Video Communications (IVC): The IVC database contains a set of ten original images distorted by four types of processing and is supported by the corresponding quality scores as available in [17].
- Toyoma: This database consists of subjective assessment data and test stimuli generated through processing of 14 reference images using JPEG and JPEG2000 [18].

This article is organized as follows. For the pixel-based approaches, the methods that apply direct estimation of single and multiple artifacts are reviewed in Sections 1.3 and 1.4, respectively. The methods based on computation of various features and an evaluation of impacts of pertinent artifacts upon NSS are discussed in Section 1.5. Bitstream-based NR methods are reviewed in Section 1.6. The methods constructed as hybrids of pixel and bitstream-based approaches are discussed in Section 1.7. Finally, some conclusive remarks and a brief outlook of possible future works in this area are presented in Section 2.

1.3 Single artifact NR-P-based methods

Blurring, blocking, and ringing are considered to be the most commonly found spatial domain artifacts in images/videos compressed by lossy encoders [19]. Moreover, noise is also a common source of annoyance in images and videos. Transmission of videos over lossy networks gives rise to temporal artifacts such as frame freeze. In the following, we examine the recent methods which adopt the approach of quantifying a single artifact for perceptual quality estimation. The section is divided into subsections for each of these artifacts, and an overall discussion is provided at the end.

1.3.1 Blurring

Winkler defines blur as an artifact which appears as a loss of spatial detail and a reduction of edge sharpness [20]. The reasons for the occurrence of blur can be many, originating in the acquisition, processing, or compression [21]. The primary source of blur in compression techniques is the truncation of high-frequency components in the transform domain of an image. Other possible reasons of the blurring of an image or video can be out-of-focus capturing, relative motion between the camera and the object being captured, or limitations in the optical system. Traditional no-reference blur methods usually focus on a particular coding artifact for quality prediction and hence their performance is compromised in circumstances of more general blur. Moreover, there has been little work carried out to build methods which have the capability of assessing blur in natural scenarios, rather, most of the work is focused on the simulated blur. A basic schematic of NR blur assessment is shown by the flowchart given in Figure 2. In many NR methods of estimating the impact of blur on visual quality, the computations begin with measuring the spread of pixels present on the edges in an image. Usually, it involves the application of commonly used edge detectors such as Sobel and/or Canny for finding the edges in the image. The next step is typically the computation of the edge distortion value that can be used towards finding an estimate of the blur. Some methods, however, make use of HVS adaptation to the value of edge distortion to classify it as perceivable or not perceivable by a human subject.
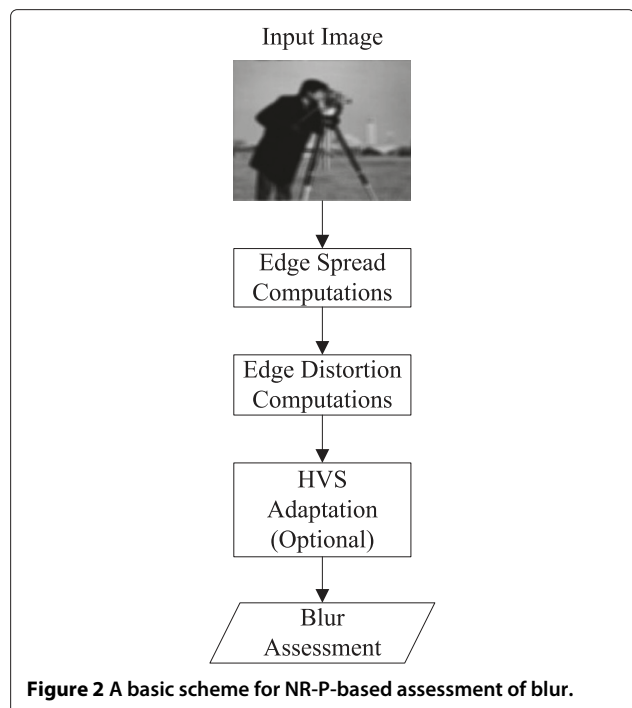
Figure 2 A basic scheme for NR-P-based assessment of blur.

A paradigm for blur evaluation has been presented in [22] that is mainly composed of four methods of blur quantification, given in [23-25] and [26], which have been integrated by an artificial neural network (ANN) powered multifeature classifier. In the method given in [23], an image quality measurement method in terms of global blur has been proposed. The method relies on histograms of discrete cosine transform (DCT) coefficients present in MPEG and JPEG encoded data to qualitatively encompass the distribution of null coefficients, given the fact that blurred images usually end up having a lot of high-frequency coefficients set to zero. This algorithm provides results which align with subjective assessment but it focuses only on out-of-focus blur and it does not perform well when there is a uniform background present or when an image is over-illuminated. The blur assessment algorithm proposed in [24] exploits the ability of the Haar wavelet transform (HWT) to distinguish edge types, and the method works both for out-of-focus and linear-motion blur. This method is however not tested for realistic blur. The method proposed in [25] presents a framework where global blur is measured in terms of averaged edge lengths. The authors considered only a small set of Gaussian blurred images for its evaluation. Nonetheless, the method has good correlation with subjective scores. An improved version of [25] is found in [26] where HVS properties have been added to get weighted edge lengths. It is to be noted that none of these four reference methods quantify realistic blur situations, but Ciancio et al. [22] have shown their method to be useable for measuring naturally occurring blur. Overall, [22] uses local phase coherence, mean brightness level, and variance of the HVS frequency response and contrast as additional inputs, together with the earlier mentioned four methods, to various ANN models designed for quality estimation. For input calibration, a five-parameter nonlinear mapping function was used for the types of blur including simulated Gaussian, simulated linear motion, a combination of both, and real blur. The proposed method outperforms the given four reference methods when tested on a fairly large database of 6,000 images corrupted by blur. Although the proposed method does not correlate so well with subjective scores in realistic blur scenarios, with a Pearson's correlation coefficient of approximately 0.56, it performs better than the reference methods with respect to subjective rating. In an earlier paper, the same authors have used the idea of estimating image blur using local phase coherence [27] and a similar method proposed by Hassen et al. is found in [28].

It has been argued in [29] that blur below a certain threshold value remains unperceived by the HVS and such a threshold value is termed as just noticeable blur (JNB). By incorporating the response of the HVS to sharpness at various contrast levels, the authors have proposed a measure of image sharpness. It is suggested that most of the existing no-reference blur assessment methods do not perform well for a variety of images and are rather limited to assess varying blur in a certain image. They have validated this argument by testing a set of 13 contemporary reference methods, which are based on different techniques of blur assessment used for quality assessment such as pixel-based techniques, statistical properties, edge-detection-based, and derivative-based techniques. The proposed method has higher correlation with subjective MOS than the given 13 objective methods of quality assessment when it has been tested on a public database of test images. In [29], the block size used for finding edge pixels is 64×64, and a similar contribution based on JNB from the same authors is reported in [30] where a block size of 8×8 has been used for finding the edge pixels. The method proposed in [30] has been improved in [31] by adding the impact of saliency-weighting in foveated regions of an image. Specifically, more weighting is given to the local blur estimates that belong to salient regions of an image, while spatial blur values are pooled together to compute an overall value of blur for the whole image.

A similar method found in [21,32] improves [29] by addition of the concept of commutative probability of blur detection (CPBD) so that the method should estimate the quality by including the impact of HVS sensitivity towards blur perception at different contrast levels. Testing the proposed method upon three public image databases having different blur types reveals that the method performance is considerably better than some of the contemporary sharpness/blur methods. However, this method gives a quality index in a continuous range of 0 to 1 and the authors have modified it in [33] where it gives a quality value on a discrete scale of 1 to 5, the usual five quality classes which are described from *Bad* to *Excellent*. Given that blur estimation methods most often work on the idea of measurement of edge-spread and blur manifests itself in smooth or diminished edges, some edges may remain undetected. Varadarajan et al. [34] improved the method proposed in [29] by incorporating an edge refinement method to enhance the edge detection and hence outperformed the blur assessment. The authors achieved as much as 9% increase in Pearson's correlation coefficient.

In contrast to usual schemes of blur detection at the edges, the method proposed in [35] does an estimation of blur at the macroblock (MB) boundaries. The overall blur of an image can be calculated by averaging the block level measure for the whole image. The authors have also used a content-sensitive masking approach to compensate the impact of image texture. As the method was designed for videos encoded following the H.264/AVC standard, it mainly quantifies the blurring effects from quantization

and de-blocking filter. This method is essentially based on a method proposed for images [36] where an estimation of the blur in a video is made by taking an average measure of blur values for each frame.

A wavelet-based noise-resilient color image sharpness method is presented in [37]. The procedure is to compute a multiscale wavelet-based structure tensor which represents the multiscale gradient information of local areas in a color image (image gradient is defined as the directional change in the intensity or color in an image). The proposed tensor structure preserves edges even in the presence of noise. Thus, the sharpness method is defined by calculating the eigenvalues of the multiscale tensor once edges have been identified. A competitive correlation with subjective MOS is achieved when the proposed method is tested on LIVE image database [14], in comparison to a similar sharpness method.

Out-of-focus blur estimation without using any reference information has been given in [38] using the point spread function (PSF) which is derived from edge information. As the proposed algorithm works in the spatial domain, avoiding any iterations or involvement of complex frequencies, it is expected to operate fast and possible to be deployed in real-time perceptual applications. Based on the similar approach in [39], the method has been made workable to assess blurriness of conditions like added blur, realistic blur, and noise contamination.

Chen et al. [40] have claimed that their method works for any kind of blurriness, without being sensitive to the source of the blur. A gradient image is calculated from the given image pixel array. A Markov model is used and a transition probability matrix is computed. Finally, a pooling strategy is applied to the probabilistic values to obtain the blurriness measure.

Some of the other recently introduced no-reference blur assessment methods include the following: In [41] a method based on multiscale gradients and wavelet decomposition of images is given, an image sharpness based on Riemannian tensor mapping into a non-Euclidean space has been found in [42], radial analysis of blurred images in frequency domain is done in [43] to set an image quality index for blur estimation, and reference [44] presents a perceptual blur method to assess quality of Gaussian blurred images. A method based on blur measure in salient regions has been presented in [45]. The perceptually relevant areas in an image are identified through elements of visual attention, namely, color contrast, object size, orientation, and eccentricity. Quality values in correlation with subjective scores are produced by localizing the degradation measure in these elements.

1.3.2 Blocking

Blocking is an artifact which manifests itself as a discontinuity between adjacent blocks in images and video

frames [3]. It is a predominant degradation that occurs after employment of block-based processing and compression techniques at high compression ratio conditions. In such techniques, transform is usually followed by quantization of each block individually leading to incoherent block boundaries in the reconstructed images or frames. Blockiness can be estimated in a region of an image, in general, by computing the difference between neighboring blocks and the amount of brightness around those blocks as shown in Figure 3. After the value of blockiness is determined in a certain region, it is important to estimate whether it would be significant for human perception or not by taking into account the impact from masking effects. This way, certain features that represent the input from HVS can be calculated. In general, blocking perception is affected by various factors including the blockiness strength (i.e., the difference between adjacent blocks), the local brightness around the blocks, and the local texture present in an image.

A frequency domain pixel-based bi-directional (horizontal and vertical) measure used to gauge blocking in images is presented in [46]. The authors claim that the proposed method can be used for any image or video format. Unlike the traditional no-reference blocking measures, this method does not require any *a priori* information about block origin, block offset or block-edge detection. The method has been evaluated on a large set of LIVE image and video database available as JPEG encoded images and MPEG-2 encoded videos. It outperforms a set of 13 contemporary blockiness methods in terms of prediction accuracy and monotonicity.

Liu et al. [47] presented an HVS-based blocking method to assess image quality using a grid detector to locate
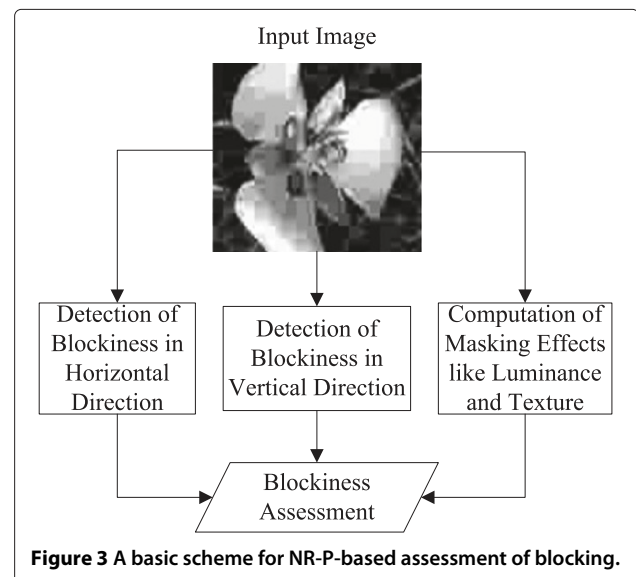

Figure 3 A basic scheme for NR-P-based assessment of blocking.

blocking. A local pixel-based blockiness measure which is calculated on the detected degraded regions is averaged to provide a blockiness value for the whole image. The main strength of this method in terms of computational efficiency and relevance to HVS response lies in the application of visual masking which makes the calculations perform only in the areas of blockiness visible to human perception. The authors took up the same method for further extensive evaluation in [48] under various conditions of comparison of performance where, for example, HVS models and grid detector are omitted or included. The results show that the proposed method performs better than some contemporary methods and can be a good candidate for real-time applications due to its simplified HVS model.

In [49], a blockiness assessment method is presented for block-based discrete cosine transform (BDCT) coded images. It is based on the estimation of noticeable blockiness. The so-called noticeable blockiness map is derived from luminance adaptation and texture masking in line with HVS response combined with a discontinuity map to quantify the visual quality. Along with its validated usability for deblocking of JPEG images, it has the potential of optimizing the codec parameters and similar other post-processing techniques.

Babu et al. presented their HVS related features-based blocking method in [50]. Blockiness as perceived by humans in JPEG encoded images is affected by a number of features such as edge amplitude around the borders of DCT blocks and edge length; the value of these increase in amount as compression rate is increased. It is also affected by the amount of background activity and background luminance as these have masking impact on possible blocking artifacts. The authors have used a sequential learning algorithm in a growing and pruning radial basis function (GAP-RBF) network to estimate the relationship between the mentioned features and the corresponding quality measure. Babu et al. also proposed a method of determining block-edge impairment [51] using the idea that edge gradients of blocks in the regions of low spatial details would contribute towards the overall blocking in an image. The level of spatial details is estimated through edge activity that is computed through standard deviation measurement of each edge.

Other methods in this area include the blind measurement of blocking in low bit rate H.264/AVC encoded videos based on temporal blocking artifact measure between successive frames of a video presented in [52]. A weighted Sobel operator-based blocking method is presented in [53], in which the computation involves luminance gradient matrices of DCT-coded images. A method where a rather simple approach of taking abrupt change in pixel values as a signal of blocking has been proposed in [54] and it can be implemented both in pixel and DCT domain, and a method of blockiness estimation in natural scene JPEG compressed images has been presented in [55] which was influenced by the impact of multineural channels pattern of HVS for vision sensing.

1.3.3 Ringing

The ringing artifact is associated with Gibbs phenomenon and is observed along edges in otherwise smooth texture areas [20]. It has yet been relatively less investigated for NR perceptual quality measurements. This kind of degradation is caused by rough quantization of the high-frequency transform coefficients and is observed in the form of ripples around high contrast edges. A schematic block diagram of commonly used approaches for the estimation of perceptual ringing is shown in Figure 4. Certain features can be extracted from the edge maps to classify the image areas in terms of relevance towards ringing artifact. Masking effects of textured regions can be examined to check if the ringing would be visible to HVS perception. From the obtained data, a ringing map is generated for various regions and an overall value of perceptual ringing is obtained for the whole image. We have not found any publication on the NR estimation of ringing in videos.

Liu et al. have put forward HVS-based quality assessment methods which quantify ringing in compressed images in [56,57]. The work in [56] does not incorporate the masking effects of HVS properties. However, in [57], Liu et al. have improved the already existing method
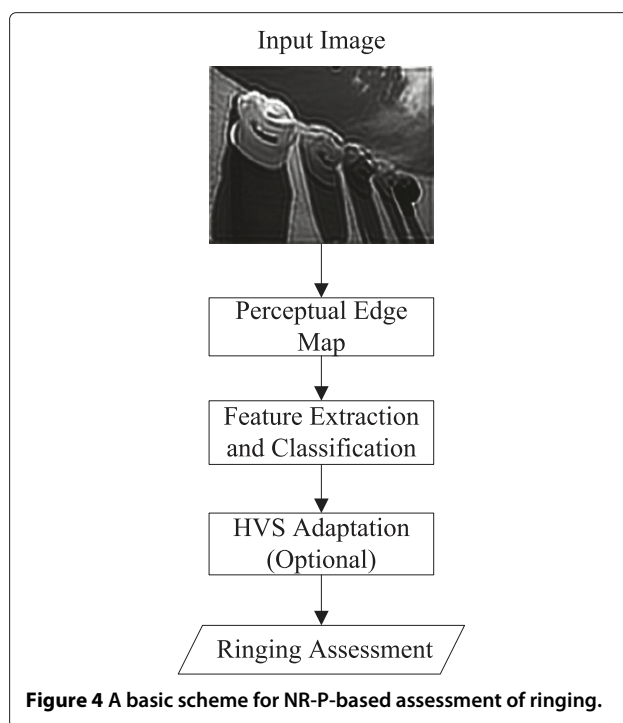

Figure 4 A basic scheme for NR-P-based assessment of ringing.

in multiple aspects. Edge detection is crucial for locating the possible ringing artifact and is used along with consideration of HVS masking in designing of such a method. The HVS masking is integrated by adding human visibility index of ringing nuisance estimate inside the already detected distorted regions. This method has a performance level comparable to a full reference method and it outperforms the two given no-reference methods of ringing assessment while tested on JPEG compressed images. As the method does not use coding parameters like DCT coefficients, the authors argue that a slightly tuned version of the same method should perform similarly well when employed on other types of compressed images, e.g., JPEG2000.

Ringing may occur also as a result of an image restoration process, unlike the other artifacts which usually occur during compression. The ringing that occurs due to image restoration has different characteristics as compared to the one that occurs due to compression. Iterations of blind deconvolution in the image restoration process are likely to result in the generation of ringing [58]. A quality method to assess perceived ringing as a result of application of blind deconvolution methods for image restoration is proposed in [58] and in [59]. The authors claim that these methods evaluate ringing with no sensitivity to the image content type and any specific ringing process. In the method proposed in [58], a 2D Gabor wavelet filter and a line detector were used to quantify ringing in restored images. A similar approach with enhancement is found in [59] where the authors have proposed to assess the degradation on image boundaries and image edges separately and then fuse the weighted results of the two values to have the overall ringing value. A 2D Gabor filter response image is used to calculate the perceived ringing at boundaries, and a Canny edge detector is used for locating ringing around edges in the image. The proposed method was tested on gray scale images restored from simulated blur. It has been found that the reported results are in line with subjective scores of quality assessment.

1.3.4 Noise

Besides the aforementioned unwanted components of an image or video that affect the perceptual quality, there can be other types of spatial noise as well. The mostly occurring types of spatial noise include salt and pepper noise, quantization noise, Gaussian noise, and speckle in coherent light situations. Mostly, the noise is considered to be an additive component, e.g., Gaussian noise, but in some situations the noise component is multiplicative, e.g., speckle noise [60]. Noise can be introduced during the image/video acquisition, recording, processing, and transmission [61]. Estimation of noise is required due to numerous reasons and applications in image processing

such as denoising, image filtering, image segmentation, and feature extraction. For the estimation of noise signal, in most cases, it is assumed to be *independent, identically distributed additive and stationary zero-mean signal*, i.e., white noise [62]. Image noise estimation methods can be categorized into either smoothing-based approaches, where noise is computed using the difference between the input image and a smoothed version of it, or block-based approaches, where block variances of the most homogenous block in a set of image blocks is taken as noise variance [63]. Similar to the approaches used for estimation of other artifacts, computation of noise characteristics depends on the extraction of some features that are affected by noise. Figure 5 shows the basic scheme of a block-based approach of noise estimation where an image is divided into smooth areas. A variance higher than a certain threshold in those areas gives an estimate of the noise.

A block-based approach proposed in [64] uses statistical analysis of a histogram of local signal variances to compute an estimation of image noise variance. However, this method is challenged by high computational requirements due to its iterative processing, and [65] simplifies this technique by taking image structure into consideration. It uses high-pass directional operators to determine the homogeneity of blocks besides using average noise variances. The performance of the improved method has
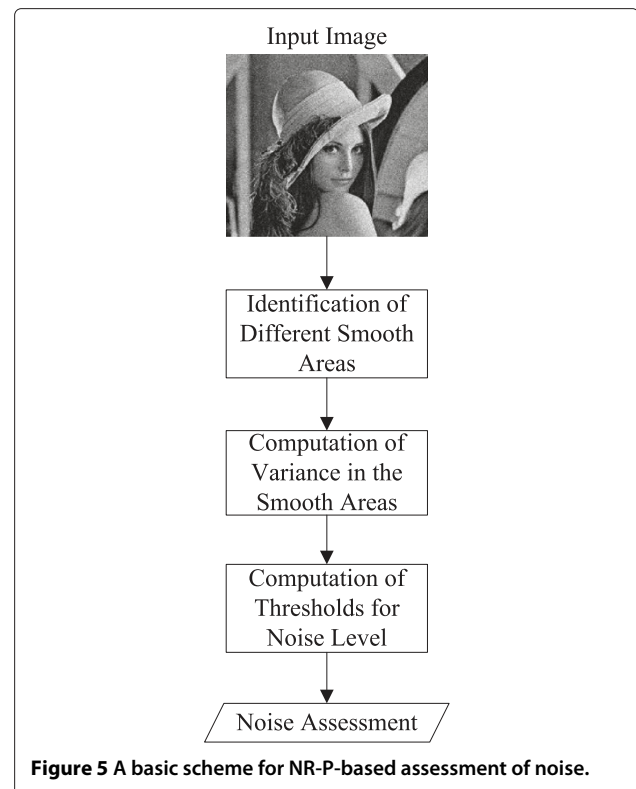

Figure 5 A basic scheme for NR-P-based assessment of noise.

been verified using highly noisy as well as good quality images. This method requires a full search of an image to determine the homogeneous areas in it. At the expense of decreased accuracy, spatial separations between blocks can be used to reduce the computational complexity. This approach has been adopted in [66] where particle filtering techniques have been used in the process of localization of the homogeneous regions. It has been shown that the proposed method reduces the number of required computations for homogeneity measurements while it outperforms [65] in accuracy. More examples of block-based approaches are found in [67-69] where noise level is computed by performing principal component analysis (PCA) of the image blocks.

1.3.5   Temporal impairments

Temporal impairments can be divided into two main categories: impairments caused by the encoding process and impairments caused by network perturbations. The typical temporal impairments caused by the encoding process come from temporal downsampling which can be performed uniformly or non-uniformly, depending on different underlying reasons. The impairments generated by network perturbations come from delay or packet loss [70]. These different impairments can be categorized as the following [3,4,71,72]:

- Jerkiness: non-fluent and non-smooth presentation of frames as a result of temporal downsampling
- Frame freeze: frame halts as a result of unavailability of new frames to present due to network congestion or packet loss etc.
- Jitter: perceived as unnatural motion due to variations in transmission delay as a result of, e.g., fluctuations in the available bandwidth or network congestion
- Flickering: noticeable discontinuity between consecutive frames as a result of a too-low frame rate together with high texture, coding artifacts, or motion content
- Mosquito noise: appears as temporal shimmering seen mostly in smooth textured areas produced by ringing and prediction error due to motion compensation mismatch

Jerkiness is the impairment perceived by the user, while jitter and frame freezes are the technical artifacts which produce jerkiness. Figure 6 presents an overview of how temporal impairments are computed in most of the contemporary methods. Generally, the first step is to compute the inter-frame difference of pixel intensities (usually the luminance channel only) and the obtained value can be used as it is or a mean square value can be calculated. Afterwards, various techniques can be applied to determine the location and possibility of frame freeze or frame drops. Some kind of thresholding is then useful to obtain
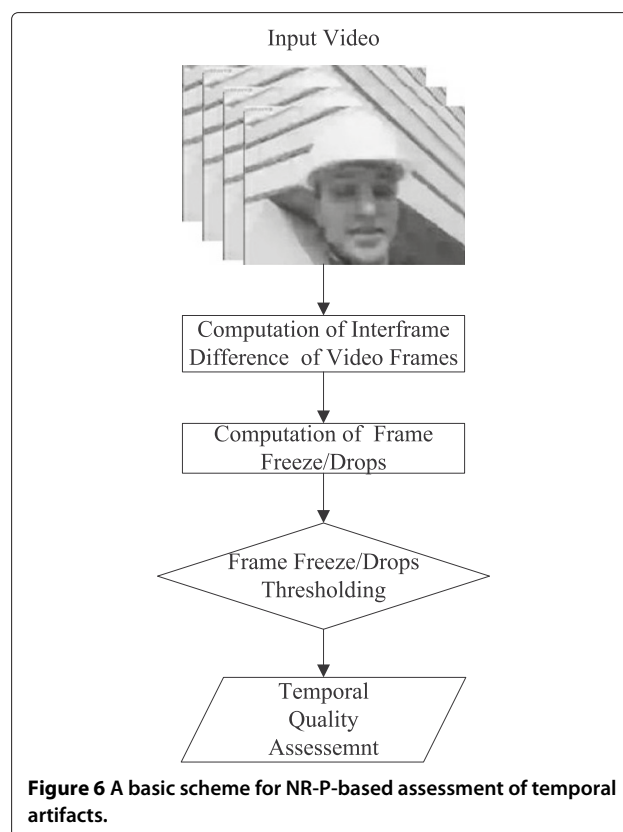

Figure 6 A basic scheme for NR-P-based assessment of temporal artifacts.

more information about the occurrence of a potential temporal artifact. Finally, a suitable pooling mechanism is used to compute an overall value of the artifact under consideration.

Borer [71] presented a model based on the mean square difference (MSD) of frames for measuring jerkiness (both frame jitter and frame freeze) which proved its potential for quality assessment of videos with resolution ranging from QCIF up to HD. This model calculates jerkiness as an accumulative result of multiplication of three functions called relative display time, a monotonic function of display time, and motion intensity of all frames. The display time and motion intensity values are parameterized through a mapping S-shaped function, which is equivalent to a sigmoid function. Besides the fact that the proposed model has reasonable correlation with MOS, it does not take into account the value of the motion intensity at the start of a freezing interval.

An earlier proposed temporal quality method which is centered around measuring the annoyance of frame freeze duration is given in [72]. This method uses MSD value to mark freeze events and builds a mapping function based on such durations of freeze to estimate the subjective MOS. The method is a part of ITU-T Recommendation J.247 Annex C [73] for the objective perceptual quality measurement of video. Although the quality method

has not been compared for performance against other methods, it has promising values of correlation with the subjective scores. However, the blind frame freeze detection system proposed in [74] claims to outperform the model [72] in terms of precision of correctly signaling a zero MSD event as a frame freeze event or not. They have presented an algorithm for thresholding such as zero MSD events to be classified as frame freeze events or the absence of it. The proposed method is reported to be equally good in performance for videos encoded using low or high quantization parameter (QP) values.

Wolf proposed an approach to accurately detect video frame dropping in [75]. One of the salient features of this approach is its use in a RR method where an adaptive threshold value is determined to avoid detection of very low amount of motion (e.g., lips movement) as a potential frame drop event. Similar to the temporal-artifact-based methods discussed before, this method also derives its computations from the basic difference in pixel values between frames to check for possible frame drops.

A method for visual quality distortion due to arbitrary frame freeze is discussed in [76]. It recursively aggregates arbitrary freeze distortions in the video under test using a method which they proposed earlier in [77]. Essentially, the approach presented in [77] replaces the sum of various freeze distortions with an equivalent single freeze length for predicting the video quality.

Yang et al. targeted their research to assess both consistent and inconsistent frame drops as a measure of perceptual quality in their contribution found in [78]. The constituents of the quality method are the amount of frame drops, motion of various objects in a video, and localized contrast of temporal quality. Instead of relying on frame rate to be used as a basis for temporal quality method, the event length of frame losses has been used. The proposed method correlates well with subjective MOS for test sequences with a range of frame rates and a variety of motion contents.

A rather general model was proposed in [79] for several fluidity break conditions: isolated, regular, irregular, sporadic, and varying discontinuity durations with different distributions and densities. Similarly, the temporal quality method proposed in [80] accounts for the impact of various frame dropping situations and spatio-temporal luminance variations due to motion.

In [81], the authors have shared their preliminary findings on estimation of the effects of lost frames on visual quality by analyzing the inter-frame correlation present at the output of the rendering application. As the lost frames are replaced by a repetition of the previous frame, this results in high temporal correlation at those locations. Analysis of this correlation results in temporal and distortion maps.

1.3.6 *Discussion*

Except for temporal impairments, most of the methods reviewed in this section have been proposed and tested for images and not for videos. For example, blockiness is a common artifact at high compression rates and some coding standards such as H.264/AVC include the use of a deblocking filter while the videos are being processed by the codec. The blockiness methods proposed for images can be used in the case of videos as well where a suitable temporal pooling scheme needs to be used. We believe that development and testing of more NR methods of blockiness estimation for videos would be beneficial. For the case of spatial-artifacts-based methods, it is evident that most of the research focus has been aimed at the development of techniques that are based on a specific coding technique or image compression standard. This fact necessitates the focus towards unraveling cross-encoder methodologies. Considering the available methods related to the quantification of perceptual impacts of various temporal artifacts, it is noted that more diverse methods are required in this area that can be applied for a variety of video resolutions and frame rates. It has also been observed that many methods employ some commonly used test database of images and videos which in turn gives an opportunity to compare the performance of competitive methods on the common benchmarks of quality. One important strength of the methods that are tested for the performance using test databases such as LIVE (image or video) is their higher applicability because the media present in such databases have been assessed for overall perceptual quality and not for a particular artifact. However, the test databases should be enriched with new demanding areas such as higher-resolution images and videos (HD and above). Besides declaring the performance of the proposed methods, finding some common approaches for reporting the computational complexity would be interesting. Table 1 presents a summary of the methods discussed in the subsections regarding blurring, blocking, ringing, and temporal artifacts. It is noted that a very low number of methods have been tested for HD resolution images. Competitive methods can be seen at a glance by observing the significantly high values of the performance indicators.

1.4 Multiple artifacts NR-P-based methods

Various artifacts found in images and videos, incurred due to compression or other reasons, can be combined to predict the overall perceived quality. As shown in Figure 7, an image or video can be processed for the extraction of features relevant to different artifacts. A suitable pooling mechanism can be employed to combine the results of different artifact measurements, to make an estimate of overall perceptual quality.

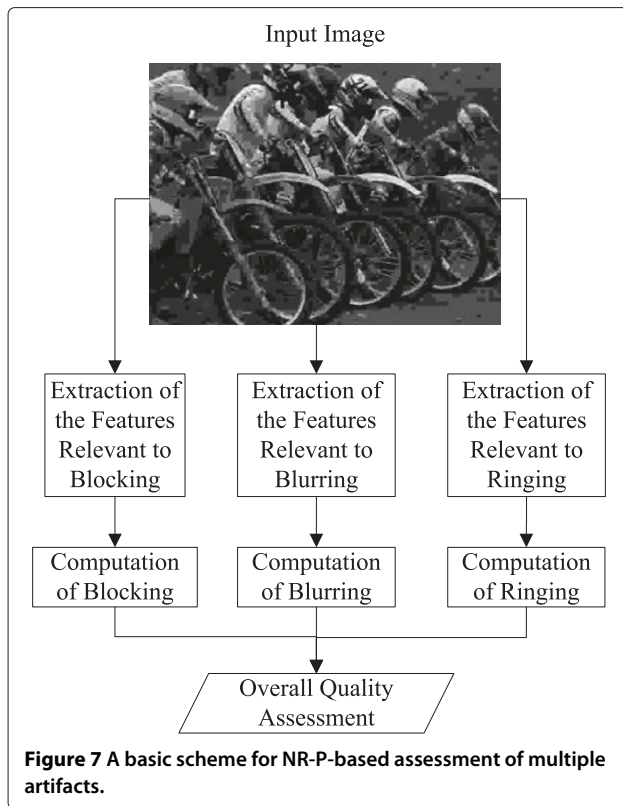**Table 1 Characteristic summary of single-artifact-based metrics**

Method	Reference	Processing	Resolution	Test data[a]	Performance[b]
Blurring	Narvekar et al. [21]	JPEG2000	768 × 512	LIVE image, TID2008,	(LIVE) PC = 0.88, SC = 0.88
		Gaussian blur	768 × 512	IVC, Toyoma	(LIVE) PC = 0.91, SC = 0.94
	Ciancio et al. [22]	Multiple	Multiple	6,000 test images	Multiple
	Hassen et al. [28]	Multiple	768 × 512	LIVE image	PC = 0.92, SC = 0.93
	Ferzli et al. [29]	JPEG2000	768 × 512	LIVE image	PC = 0.88, SC = 0.87
		Gaussian blur	768 × 512	LIVE image	PC = 0.93, SC = 0.94
	Varadarajan et al. [34]	Gaussian blur	-	24 test images	PC = 0.75
	Debing et al. [35]	H.264/AVC	HD	1,176 test images	PC = 0.85
	Maaloof et al. [37]	Multiple	768 × 512	LIVE image	PC = 0.96, SC = 0.95
	Chen et al. [41]	Gaussian blur	768 × 512	LIVE image	SC = 0.61
	Chetouani et al. [43]	Gaussian blur	Multiple	LIVE image, IVC	PC = 0.86 (LIVE), 0.94 (IVC)
	Hua et al. [44]	Gaussian blur	768 × 512	LIVE image	PC = 0.96, SC = 0.95
	Oprea et al. [45]	JPEG2000	768 × 512	29 test images	PC = 0.85
Blocking	Chen et al. [46]	JPEG	768 × 512	LIVE image	PC = 0.96, SC = 0.94
		MPEG-2	768 × 512	LIVE video	PC = 0.95, SC = 0.85
	Liu et al. [47]	JPEG	768 × 512	LIVE image	PC = 0.73, SC = 0.91
	Liu et al. [48]	JPEG	768 × 512	LIVE image	PC = 0.79, SC = 0.91
	Zhai et al. [49]	JPEG	768 × 512	LIVE image	PC = 0.96, SC = 0.91
	Babu et al. [50]	JPEG	768 × 512	LIVE image	R-square = 0.95
	Zhang et al. [52]	H.264/AVC	CIF	50 test videos	-
	Hua et al. [53]	JPEG	768 × 512	LIVE image	PC = 0.92, SC = 0.83
	Suthahran et al. [55]	JPEG	768 × 512	LIVE image	PC = 0.94
Ringing	Liu et al. [56]	JPEG	768 × 512	16 test images	-
	Liu et al. [57]	JPEG	768 × 512	55 test images	PC = 0.80, SC = 0.73
	Zuo et al. [59]	-	171 × 256	LIVE image	PC = 0.94
Temporal impairments	Borer [71]	H.264/AVC	CIF, HD	-	-
	Huynh-Thu et al. [72]	-	QCIF	-	PC = 0.95
	Wolf [75]	H.264/AVC	SD	VQEG phase 1	-
	Yang et al. [78]	-	QCIF	6 SRCs	PC = 0.94

[a]The used test database; [b]performance denotes the correlation with subjective assessment, unless stated otherwise; SRC, source sequence; PC, Pearson's correlation coefficient; SC, Spearman's rank order correlation coefficient.

Blurring and ringing are the main associated degradations when JPEG2000 coding is operated at low bitrate conditions. The quality method proposed in [19] predicts quality of JPEG2000 coded images by combining blur and ringing assessment methods. Based on the local image structures, a gradient profile sharpness histogram is calculated for evaluation of a blur estimation method, and a ringing method is generated from regions associated with gradients profiles. Here, a gradient profile is essentially the distribution of the gradient magnitude along the gradient direction. It has been argued that the underlying proposed

blur method is insensitive to the inherent blur found in natural images, e.g., out-of-focus blur. The performance of the method is similar to or better than a number of competitive methods while tested on LIVE JPEG2000 and TID2008 datasets.

A rule-based VQA method given in [82] relies on a group of pixel domain features of a video. It includes blockiness and blurriness as well as spatial activity, temporal predictability, edge continuity, motion continuity, and color continuity. The authors have used already available methods to measure the first five features and have

Input Image

Extraction of the Features Relevant to Blocking

Extraction of the Features Relevant to Blurring

Extraction of the Features Relevant to Ringing

Computation of Blocking

Computation of Blurring

Computation of Ringing

Overall Quality Assessment

Figure 7 A basic scheme for NR-P-based assessment of multiple artifacts.

proposed their own methods for the estimation of the motion continuity and color continuity. A multivariate data analysis technique has been used to combine all the features for computing a single quality score. The earliest mentioned three features (blockiness, blurriness, and spatial activity) are measured on a single frame and the rest is calculated on an inter-frame basis. The used approach is to segregate the given set of videos into one of the given feature models and then estimate an initial prediction of the quality measure. After that, using a low-quality version of the video, a correction value is added to the initial quality estimate of the original videos. The authors claim that at the time of publication, this is the first reference-free quality method for H.264/AVC encoded videos which have been tested on a relatively large test database.

Noise is an artifact found in images in the form of a random variation of brightness and color information (see Section 1.3.4 for more details on noise). An empirical formulation of the objective measure of image quality based on blur and noise has been proposed in [83]. The method is based on the level of image intensity variations around its edges. The authors argue that in modern digital cameras, the image signal processor (ISP) enhances the image by removing noise but in doing so, it may deteriorate the image texture. Hence, there is a need of finding a

trade-off between noise and blur and it provides a rationale for combining the estimation of noise and blur in the same method. Specifically, this method considers simulated conditions of white noise as source of the noise artifact in the test stimuli.

Another joint method for noise and blur estimation is found in [84]. It estimates the amount of degradations introduced by additive white noise, Gaussian blur, and defocus blur on the quality of an image. Given the fact that noise disturbs virtually all the spatial frequencies of an image and causes an adverse rise in the higher frequencies while blur has attenuation effect on them, it is justified to study the joint impact of noise and blur on the perceptual quality. The authors have evaluated the impact of noise in both spatial and frequency domain while only the frequency domain is used for blur estimation. The central idea is to influence the power spectra of the image in order to highlight the impact of the distortions on the spectrum properties. The source of noise in the test stimuli used in this work is also white noise. The proposed method has not been tested for its correlation with subjective assessment but it has a competitive performance in comparison with a contemporary method [85] of blind image quality evaluation.

In [86], a sharpness method is presented which is sensitive to the prevailing blur and noise in an image. The mathematical formulation of the method is based on image gradients computed through singular value decomposition (SVD) rather than edge detection as commonly found in contemporary pixel-based structure measures. However, it requires a prior estimate of noise variance. This issue has been resolved in the authors' later contribution [87]. Simulations on realistic noise data have substantiated the potential usage of this method in parameter optimization issues of image restoration such as applications for denoising. The support vector regression (SVR)-based method reported in [88] uses singular vectors from the SVD data instead of using singular values as in [87]. Various artifacts would modify the singular vectors, and hence the geometry of the distorted image will be changed leading to visual annoyance as perceived by humans. The usefulness of the method was tested on multiple image databases with a variety of artifacts. The results were found to be in accordance with subjective ratings.

Another quality method based on gradient statistics of JPEG and JPEG2000 images, degraded by blockiness and blur, is presented in [89]. This method differs from the methods given above in one way that it does not combine the estimated amount of artifacts to yield a single quality score. Instead, it uses the same approach of calculation of local features in gradient domain for both of JPEG and JPEG2000 images and then estimates the quality of the two sets separately. The obtained results lie in

accordance with some contemporary methods of blocking estimation in JPEG images and blur estimation in JPEG2000 images. Further, an artificial neural network has been used in [90] to combine a blocking method, a blurring method, and a ringing method to estimate the overall quality of an image. Quality estimators targeted for images encoded by JPEG2000 usually quantify ringing only, but such images may contain blur as well. The method proposed in [91] first determines the type of distortion by using an ANN classifier and then, depending on these results, either uses a ringing [92] or blur [43] method for the quality assessment.

Different from the aforementioned IQA methods, another example of a composite method has been proposed for videos [93]. This method is based on blocking and flickering measure of H.264/AVC encoded videos. It correlates well with subjective quality assessment and also with the structural similarity (SSIM) index [94].

Most of the VQA methods are processed in the luminance plane only to simplify the computational complexity. However, the method proposed in [95] computes three artifacts both in the luminance and chrominance planes of a video. In this method, they compute the significance of the direction in which an artifact is calculated for determining its contribution to perceptual quality assessment. Hence, for example, the value of blur in vertical direction has been given more weighting than the same in horizontal direction. In this method, blocking is measured by computing boundary smoothness between 8×8 blocks and block visibility detection. The third impairment which is considered is jitter/jerkiness. Finally, a multiple regression scheme is employed for weighted integration of the six feature values towards the corresponding quality value. The suggested quality predictor bears competitive correlation with subjective MOS when compared with some contemporary methods as tested on standard-definition television (SDTV) sequences found in VQEG Phase 1 database.

A modular method of combining artifacts both from spatial and temporal domain for quality estimation has been proposed in [80]. The method accounts for frame freeze/jerkiness and clearness/sharpness in MPEG-4 encoded videos. It has been claimed that the combined model is an estimator of global visual quality.

1.4.1 Discussion

Given the fact that a certain type of processing, e.g., JPEG2000 coding, can introduce more than one kind of artifacts, it is imperative to have quality estimators that can assess the impact of more than one artifact. The application of the estimation of multiple artifacts becomes even more interesting when a certain processing that involves removal of an artifact, such as denoising, can produce another artifact due to its underlying methodology. The

popularity of digital cameras in the recent years increases the demand of a quality estimation mechanism to compute multiple artifacts that can be used as an aid to improve the photography experience. Global visual quality estimators such as in [80] are a useful contribution towards making an overall assessment of a video signal as it can be impaired by spatial artifacts like blurring and temporal artifacts like jerkiness at the same time. Table 2 presents a summary of some of the existing methods of quality assessment that are based on the estimation of multiple artifacts. Overall, it is noted that these methods should be tested on higher-resolution images/videos to account for the requirements of the new display devices with capability of presenting resolutions of HD and above.

1.5 Features measures-based methods

An image or video signal can be decomposed to obtain various features that may be used in the process of estimating the perceptual quality of an image or a video. Generally, such features can represent a particular aspect of the visual signal and its relation to the corresponding perceptual quality. Depending upon the nature of the feature with regards to its relation to perceptual quality, a certain feature can be a desired or an unwanted component of an image or video. For instance, the presence of sharpness in an image can be perceptually preferred in many cases and hence it may be considered as a wanted feature. On the other hand, an image with pixel distortions could be considered as of low quality. In addition, certain features represent different characteristics of an image or video and can be used as complementary information besides other features for making an estimate of quality. For example, the amount of spatio-temporal information content of a video can be used to characterize the masking effect on various artifacts that may be present in the signal. More examples of visual quality relevant features include local contrast, brightness, colorfulness, and structural activity [96,97].

Moreover, it has been described in [98] that natural images possess a common statistical behavior. This behavior has been termed as NSS, and it has been found to be a useful feature for the description of image quality. There have been numerous applications of NSS including image segmentation, denoising, and texture analysis and synthesis. Although it was concluded in [98] that the major usage of scene statistics would be in the investigation of visual sensory processing, these have recently been proved to be quite useful in the design of no-reference quality methods. It has been found that such common statistical characteristics get distorted by image processing applications like image compression, and a quantitative measure of this distortion can yield the relevant variations in the image quality. Thus, an NSS-driven NR quality assessment method would provide the measure of the

Table 2 Characteristic summary of multiple-artifacts-based and features measures-based metrics

Method	Reference	Processing	Resolution	Test data[a]	Performance[b]
Multiple	Liang et al. [19]	JPEG2000	768 × 512	LIVE image, TID2008	(LIVE) PC = 0.92, SC = 0.94
artifacts	Pastrana et al. [80]	MPEG-4	QCIF, CIF	6 SRCs	PC = 0.9, SC = 0.9
NR-P-based	Oelbaum et al. [82]	H.264/AVC	CIF	300 test videos	PC = 0.82, SC = 0.75
	Choi et al. [83]	JPEG2000, noise	768 × 512	LIVE image	PC = 0.91
	Cohen et al. [84]	Noise and blur	256 × 256	75 test images	-
	Zhu et al. [86]	JPEG2000	512 × 768	LIVE image	-
	Narwaria et al. [88]	Multiple	Multiple	Multiple	(LIVE) PC = 0.8894
	Liu et al. [89]	JPEG, JPEG2000	768 × 512	LIVE image	PC = 0.92
Natural scene	Zhou et al. [99]	JPEG2000	768 × 512	LIVE image	-
statistics	Lu et al. [100]	Multiple	768 × 512	LIVE image	Multiple
	Shen et al. [102]	Multiple	512 × 512	LIVE +26, 260 test images	Multiple
	Moorthy et al. [103]	Multiple	768 × 512, 512 × 384	LIVE image, TID2008	Multiple
Pixel-based	Gastaldo et al. [111]	JPEG	480 × 720, 768 × 512	LIVE image	PC = 0.94
features	Li et al. [112]	Multiple	768 × 512	LIVE image	PC = 0.87, SC = 0.87
	Zhang et al. [113]	JPEG2000	768 × 512	LIVE image	PC = 0.93, SC = 0.92
	Zhang et al. [97]	JPEG, JPEG2000	768 × 512	LIVE image	PC = 0.92, SC = 0.92
	Yao et al. [119]	-	-	VQEG phase 1	PC = 0.86, SC = 0.85
	Ries et al. [121]	H.264/AVC	SIF	10 SRCs	PC = 0.93
	Ries et al. [122]	H.264/AVC	SIF	10 SRCs	PC = 0.93
Pixel-based	Sazzad et al. [114]	JPEG2000	768 × 512	LIVE image	PC = 0.93, SC = 0.96
features and	Jiang et al. [127]	MPEG-2	HD	72 test images	PC = 0.91
artifacts	Keimel et al. [129]	H.264/AVC	HD	7 SRC videos	PC = 0.86, SC = 0.85
	Sazzad et al. [130]	JPEG	640 × 480	490 test image pairs	PC = 0.93

[a]The used test database; [b]performance denotes the correlation with subjective assessment, unless stated otherwise; SRC, source sequence; PC, Pearson's correlation coefficient; SC, Spearman's rank order correlation coefficient.

unnaturalness introduced into the natural scene statistics under the effect of image distortions. Figure 8 shows a basic schematic block diagram of feature-based methods. We have divided the review of such methods into three subsections: (i) Natural scene statistics, (ii) Pixel-based features, and (iii) Pixel-based features and artifacts.
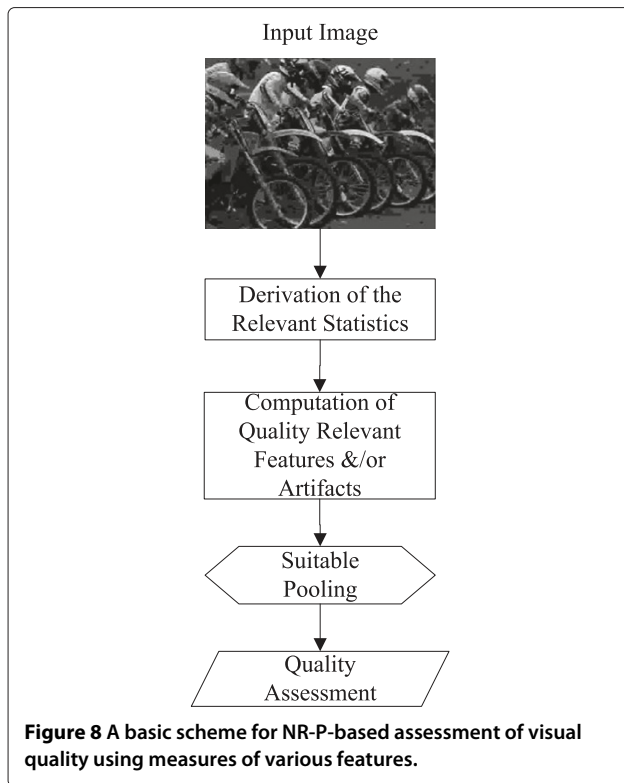
1.5.1 Natural scene statistics

It has been claimed in [92] that the distortion introduced in the nonlinear dependencies found in natural images can be quantified for making an estimate of perceptual quality. Based on that notion, the authors presented an NSS-driven approach for quality assessment of images processed by wavelet-based compression standards like JPEG2000.

Similarly, the NSS-based image quality prediction approach presented in [99] is also limited to be applicable only to JPEG2000. The authors have used a neural network to regress between inputs from NSS-based spectral amplitude fall-off curves in combination with positional similarity measure of wavelet coefficients and the corresponding quality value.

Harnessed by the measures to keep the model attributes unaffected by image content variations, the method proposed in [100] uses a contourlet transform [101] to quantify the degradations incurred on NSS. The authors show that wavelet transform does not completely exhibit the artifacts present in the image and the effect of degradations is visible in all the subbands of the contourlet domain. Hence, the contourlet domain can be more effective in image quality assessment. The proposed method has a clear advantage in precisely predicting the image quality while tested for images degraded by JPEG2000 and JPEG compression and distortions like Gaussian blur, fast fading channel, and white noise. Similarly, a statistical relationship between the characteristics of NSS in

Input Image

Derivation of the
Relevant Statistics

Computation of
Quality Relevant
Features &/or
Artifacts

Suitable
Pooling

Quality
Assessment

Figure 8 A basic scheme for NR-P-based assessment of visual quality using measures of various features.

images and the corresponding quality values was studied in [102] to engineer a reference-free quality method. In order to provide the quality ranking of the filtered natural images, a histogram of a combination of image transforms, namely, curvelet, wavelet, and cosine transform is computed. The considered distortions include noise, blur, and artifacts introduced by compression using JPEG2000 and JPEG. As the authors pointed out, this is one of the few quality methods which can quantify the perceptual impact of such a broad range of degradation types. The additional advantage of this method is its ability to classify images on the basis of the presence of one or more of these artifacts. The proposed method was tested on a large set of images from the LIVE image database as well as authors' own test set of images. As a result, a promising level of correlation with subjective quality assessment was obtained.

The distortion identification-based image quality estimation method proposed in [103] offers an NSS-based approach of image quality prediction framework and algorithm. Firstly, the pertinent distortion is identified. Then, NSS features are used to quantify the relevant quality value which is largely independent of the distortion type present in the image. The used feature set describes (i) scale and orientation selective statistics, (ii) orientation selective statistics, (iii) correlations across scales, (iv) spatial correlation, and (v) across orientation statistics.

Support vector regression is used to train the model, and the proposed method is proved to be comparable in precision of assessment to full reference methods such as peak signal-to-noise ratio (PSNR) and SSIM. The method was evaluated on images found in TID2008 and LIVE databases. It was found quite closely correlated to subjective assessment of image quality and hence proved itself to be test-set independent.

The idea of the impact of distortions on NSS has been used in [104] for prediction of video quality where each frame of the video is decomposed into a Laplacian pyramid of a number of subbands. Intra-subband statistics including mean, variance, skewness, kurtosis, energy, entropy, and inter-subband statistics, namely, Jensen Shannon divergence, SSIM, and smoothness are computed. A Minkowski pooling scheme is adopted to yield a single value out of the aforementioned statistics. The proposed method is reported to perform better than some FR metrics while tested on the LIVE video quality database.

Similar to NSS, a basic model is presented in [105] to develop an NR quality method based on temporal statistics of videos called as natural motion statistics (NMS). The theory of independent component analysis (ICA) has been applied in order to compute NMS. The authors have shown that independent components calculated from the optical flow vectors of a video signal follow the Laplacian distribution. Consequently, it has been observed that the root mean square (RMS) error of the fit between the extracted independent components and Laplacian distribution can be used as an indicator of video quality.

Saad et al. have presented their DCT statistics-based image integrity index in [106]. The central idea is to track the change in particular statistics of an image while it traverses from being original to a distorted one. The proposed framework is mainly DCT based. Owing to the perceptual relevance, some features representing structural information and contrast of an image have been extracted from the DCT values at two levels of spatial sampling. An improved version of this approach is found in [107] where the impact of NSS features for various perceptual degrees of degradation has been added.

In contrast to most of the approaches mentioned before that involve transformation of an image into another domain such as DCT, the NSS-based quality estimator presented in [108] performs in the spatial domain. Locally normalized luminance and its products-based empirical distribution is used to compute quality relevant features for building a spatial NSS model. The performance of the proposed method has been found to be better than FR methods such as PSNR and SSIM. The authors have validated the NR application of this method by employing it in an image denoising system. A similar approach has been

adopted in [109] to define latent quality factors that were used to estimate the image quality.

The idea of NSS features-based quality estimator has been used in the case of stereoscopic images as well. In reference [110], 2D- and 3D-based statistical features are extracted from stereopsis to estimate the image quality. A support vector machine model has been trained using these features, and the model has been tested using the LIVE 3D database.

1.5.2 Pixel-based features

There are some methods of no-reference quality estimation which rely on certain statistics, mainly spatial features, derived from pixels of an image or video to perform the corresponding perceptual quality evaluation. In [111], the authors present an example where they have used objective features related to energy, entropy, homogeneousness, and contrast from the color correlogram of an image. These features have been used to train an ANN which serves as a prediction model. Li et al. [112] have also deployed an ANN-based model to devise a quality estimation method using perceptually relevant image features including phase congruency, entropy of the degraded image, and gradient of the degraded image. The importance of phase of an image for its fidelity representation is well known, and the gradient of an image is an implication of changes in the luminance of an image. An ANN model is also used in the image semantic quality method presented in [96] where a variety of quality descriptive features have been used. The authors argue that the overall visual quality can be seen in terms of the usefulness and naturalness of an image. Sharpness and clarity are considered as the representatives of usefulness of an image, whereas brightness and colorfulness represent naturalness. These four representations of usefulness and brightness are further branched into a large set of pixel-based features; edge pixel distribution, contrast, mean brightness, and color dispersion are a few of the used 14 features. The advantage of using higher number of features has been shown by better performance of the predictor.

Compared to the aforementioned methods that rely on the process of training a particular model by using an extracted set of features, the pixel-activity-based method proposed in [113] does not use such methodology. The focus here is on the activity map of an image, essentially controlled by features, namely, monotone-changing, zero-crossing (ZC), and the existence of inactive pixels, which are calculated for non-overlapping image blocks. The concept of ZC has been used to refer to the places in the Laplacian of an image where the value of the Laplacian passes through zero, i.e., the points where the Laplacian changes sign. Such points often occur at edges in an image. The use of ZC as a constituent of an activity map

is justified as the method was proposed for JPEG2000 encoded images; and ringing, which can be caused by JPEG2000-based compression, has the potential of generating ZC around contours. Moreover, spatial features consisting of edge information and pixel distortion have been used to predict quality of JPEG2000 encoded images in [114]. Pixel distortion is computed using standard deviation of a central pixel and a measure of difference between central pixel and its closest neighbor pixels. Edge information relies on zero-crossing rate and a histogram measure. Particle swarm optimization has been employed to integrate these features into a single quality index. The authors have presented a similar method in their contribution [115].

The notion of quality estimation with regards to structural changes in images as a result of distortions has gained widespread attention. The FR method SSIM [94] is a commonly used representative method of this area. Zhang et al. [97] have put forward a similar approach of quality estimation based on structural changes. However, the nature of the particular distortion should be known beforehand. This method can be used to evaluate degradation caused by the following artifacts but one set at a time: (i) Gaussian blur and noise, (ii) blur and ringing, and (iii) blockiness. In a nutshell, local structural activity is taken in the form of direction spread whereas structural activity weight is computed through a measure of structural strength and zero-crossing activity.

Some feature-based methods make use of the properties of HVS to govern the performance of the method for better correlation with subjective assessment. A 3D multispectral wavelet transform-based method of NR quality estimation for color video has been given in [116]. Various channels of the HVS have been represented by wavelet decomposition of the video. To invoke the impact of the HVS, a perceptual mask of sensitivity with integrated impacts of spatio-temporal contrast and luminance has been applied to all wavelet bands. The final step is to draw a perceptual mask weighted flow tensor between successive frames to define the method. An ANN has been used in [117] with an extreme learning machine (ELM) algorithm for determining the relationship between spatial statistics of an image and its corresponding quality. These statistics are mainly HVS-inspired features, namely, edge amplitude, edge length, background activity, and background luminance of an image. As the proposed method is basically targeted at JPEG encoded images, some of the underlying methodologies which help calculate these features are focused on computation of blockiness. Since DCT coding is used in video coding also, the proposed algorithm can be generalized to be workable for video quality assessment.

In the experiments on determining the visual interest for different objects and locations inside an image, it has

been found that HVS perception is not spatially uniform. Instead, there are specific areas called region of interest (ROI), which draw more attention and hence contribute more towards overall quality assessment of an image. Treisman et al. [118] observed that the visual system notices and codes different features in parallel channels before the observer actually recognizes the objects in an image. Features such as color, brightness, and orientation can be pooled together to form a unique entity to be observed. Based on this observation, there exist IQA methods which assess perceptual quality of an image by focusing mainly on those ROIs. One such method is proposed in [119] where the impact of importance of various ROIs in a video frame has been integrated into a wavelet-based just noticeable difference (JND) profile of visual perception. The proposed method works better than some contemporary methods when it was tested on the VQEG Phase I database.

In order to estimate the impact of packet loss impairments on video quality, a method based on edge strength around macroblock boundaries is proposed in [51]. Edge strength values are processed through a low-pass filter, and a threshold value is applied to compute the edge maps of adjacent rows. Finally, the impact of packet loss is computed through a difference between these edge maps.

In order to quantify the quality of enhanced images, the method given in [120] divides an image into smooth and textured areas. A JND formulation of perception is derived based on the local average brightness and local spatial frequency. The effect of enhancement is monitored through a comparison of local brightness and a JND threshold. The performance of the proposed method is reported to be better than that of conventional average local variance-based methods.

Features-based assessment of the content of an image or video can be used in the estimation of perceptual quality. Ries et al. have shown the relevance of the content class of videos in the process of determination of the visual quality in [121]. The authors classify a given set of videos into five groups based on the content. One of such group, called class here, contains videos which have a small moving ROI with a still scene in the background. Another content class has videos with huge spread of angle of movie capturing device and is called panorama. These content classes are created based on the statistics that are mainly related to motion dynamics of a video. Values of zero motion vector ratio, mean size of motion vector, uniformity of the movement, horizontalness of movement, and greenness are the classification parameters which are used to segregate the set of videos into different content classes. The central idea of the method is to first check the content class of a video and then estimate the visual quality based on bitrate and frame rate. The authors continued working on the same idea in their

contribution found in [122] where they have presented a method aimed at the most common content classes of videos for handheld devices. Khan et al. have proposed a content-based method to combine encoding and transmission level parameters to predict video quality in [123]. Based on spatio-temporal features, the videos are first divided into content-based groups using cluster analysis. Adaptive network-based fuzzy inference system (ANFIS) and a regression model have been used separately to estimate the quality score. As per their results, transmission parameters like packet error rate have more impact on the quality than the compression parameters such as frame rate etc. The underlying techniques of ANFIS model and content clustering have been used in the authors' other contributions as given in [124,125].

1.5.3 Pixel-based features and artifacts

Some of the existing no-reference perceptual quality assessment methods are composed of a set of spatial features combined with some measurement of artifacts. A set of spatial artifacts has been combined with some spatial image features to estimate perceptual image quality in [126]. An ANN model was trained with these features for the quality prediction. Working on a similar approach, the method presented in [127] integrates spatial features such as picture entropy (represents the amount of information of a picture) and frequency energy (distribution of frequency energy in images) with artifacts, namely, blur and blockiness. The proposed method seems prominent because of its use of the chrominance information also while most of the contemporary quality measures are based on statistics from the luminance channel only. In this contribution, it has been shown that extraction of these features from ROI further improves the value of correlation with subjective scores. Five features of quality significance have been used to model an ANN-based quality predictor in [128] where the features set constitutes a measure of artifacts such as blocking and ringing and spatial statistics such as zero-crossing rate, edge activity measure, and *z*-score. Another method built on similar principle is found in [129] where the amount of blurring and blocking has been combined with spatial activity in an image and predictability of an image. A partial least square regression (PLSR) approach has been used to determine the function between these features and the quality value.

The approach given in [130] uses local segmented features related to degradation and dissimilarity for quality estimation of 3D images. In fact, the essential methodology used in [114] for 2D images have been extended to be employed for 3D images in [130]. One of the key means used to check disparity in left and right images of a stereoscopic image is the block-based edge information measure.

The authors in [131] propose a method for the assessment of facial image quality. Eye-detection, sharpness, noise, contrast, and luminance values of a test image are calculated. A weighted sum of these quantities constitutes the value of the quality method. In view of the discussion presented in [132], relatively more weighting has been given to sharpness and eye-detection as they are more important for determining facial image quality.

In [133], a set of artifacts, namely, blocking, ringing, truncation of the number of bits for image values, and noise is combined with a set of features including contrast and sharpness for designing a video quality prediction method. Each of these parameters is fitted separately in a functional relationship with subjective assessment of quality such that the correlation between the parameter values and subjective scores is maximized. Subsequently, these individual fitting functions are merged together to form a joint relationship with the perceptual quality. The data used for training includes original videos as well as different combinations of sharpness-enhanced and noise-contaminated videos. The trained model is tested on another data set which reveals a promising correlation with subjective scores.

Unlike the aforementioned NR-P-based artifacts or features measures-based methods, the mean square error distortion due to network impairments for a H.264/AVC encoded video is computed in [134]. An estimate of MSE is computed using the pattern of lost macroblocks due to an erroneous transmission of the video. Information about the lost macroblocks is estimated through the traces of the error concealment process. The same methodology has been enhanced in [135] for more general application scenarios such as no assumption is done about a certain error concealment algorithm and it does not require the knowledge of exact slicing structure.

1.5.4 Discussion
From the review of the features measures-based methods, we can make some general observations. The approach of estimating visual quality by quantifying the impact of distortions on natural scene statistics has gained a wide interest to gauge degradations due to different image processing techniques including compression. However, more of such approaches should be tested in the case of videos as well. Moreover, assessment of quality degradation due to network impairments using NSS-based approaches could be useful. The pixel-based and features-based approaches can be seen as composed of techniques that rely on a variety of spatial features including those related to edges, contrast, and some measures of structural information. The performance of these approaches can be enhanced by adapting the computational procedure with regards to the input of HVS preferences. Additionally, including the impact of mostly occurring artifacts such as blurring,

blocking, or noise could be an advantage. We observe that most of the pixel domain features-based approaches have been designed for images and it is desirable to generalize the relevant methods for applications in the case of videos. Temporal pooling quality methods such as Minkowski summation or other methods such as adaptive to perceptual distortion [136] can be used for this purpose. Table 2 presents a summary of some of the methods discussed in this section. It is evident that most of the methods in this category exhibit very promising performance, with correlation coefficient values equal to or higher than 0.85.

1.6 Bitstream-based methods
An estimate of the quality of an encoded video can be made by parsing the coded bitstream to deliver readily available features such as encoding parameters and network quality of service (QoS)-related parameters. The methods that adopt the usage of the bitstream data for quality estimation avoid the computational complexity of processing the full video data, as full decoding of the input video is not usually required in the case of bitstream-based methods. Another advantage of this type of methods is the use of readily available information from the bitstream that is significant for the quality estimation, for example, the motion vectors, coding modes, and quantization parameter values. However, these methods are inherently coding standard specific as different encoders have different formats of bitstream. There is a range of quality relevant features that can be extracted by partial decoding or primary analysis of the bitstream data. The performance of such methods significantly depends upon the level of access to the bitstream [137]. A block diagram of general framework in bitstream-based methods is given in Figure 9. We have divided the discussion of these methods into three categories based on the level of information used for processing, in accordance with the standardized models recommended by telecommunication standardization sector of International Telecommunication Union (ITU-T), as discussed in [138,139]. This includes parametric models (parametric planning model and parametric packet-layer model) and bitstream layer model. In the former type, extrinsic features of a video that are of parametric nature such as bitrate, frame rate, and packet loss rate are used. Bitstream layer models have detailed access to the payload and intrinsic features related to a video such as coding modes, quantization parameter, and DCT coefficients. The standardization of these models includes the methods designed for estimation of audio quality as well, but our discussion is limited to video quality only.

1.6.1 Parametric planning model
The parametric planning models have rather low complexity as they do not access the bitstream and utilize
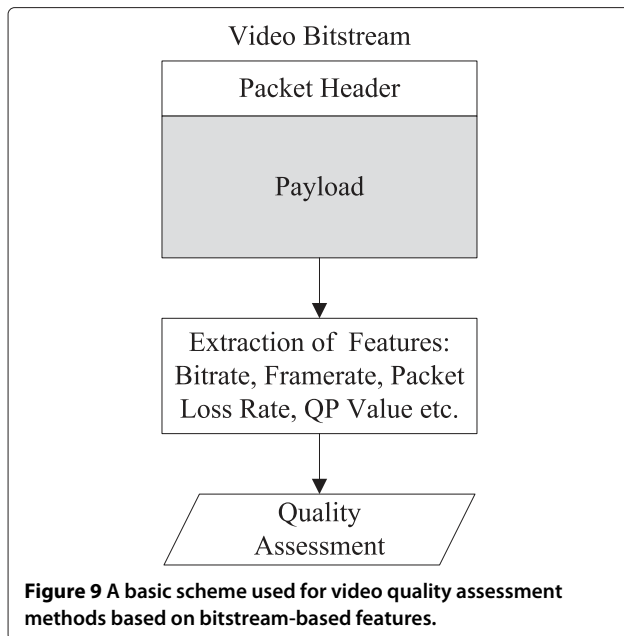
Figure 9 A basic scheme used for video quality assessment methods based on bitstream-based features.

bitrate, codec type, and packet loss rate for making a crude estimation of video quality. The work item related to this category in ITU-T is known as Opinion model for video-telephony applications, G.1070 [140]. ITU-T Recommendation G.1070 proposes a method for the assessment of videophone quality, based on speech and video parameters, that can be used by the network performance planners for ensuring the given level of end-to-end quality of the service. A quality prediction model for MPEG-2 and H.264/AVC encoded videos for IPTV is presented in [141]. The model takes some parameters related to encoding information, packet information and client information to assess the overall quality. In reference [142], a parametric model is proposed that is based on a simple method of estimating MSE that occurs due to a given pattern of packet loss. The authors derived a relationship between average motion vector length and MSE and this relation gives a fair estimate of the actual MSE.

1.6.2 Parametric packet-layer model

The packet layer models have access to the packet header of the bitstream and can extract a limited set of parameters including bitrate on sequence or frame level, frame rate and type, and packet loss rate. Parametric packet-layer models are also known as QoS-based methods. The work item related to this category in ITU-T is known as non-intrusive parametric model for the assessment of performance of multimedia streaming (P.NAMS) [143]. The visual quality estimation method proposed in [144] presents an approach where it is not required to decode the video at any level, suitable for situations where

the encoded video is encrypted. Given the observation that error concealment is more effective when there is less motion in the video, an estimation of the motion dynamics of a particular video is required to assess the effectiveness of an error concealment strategy. In this method, the ratio of the average of the B (bi-predictive coded) frame data size to the average of the size of all frames is compared with a predetermined threshold to adjust the value of the video quality score. The results obtained from the effectiveness of error concealment are refined by adjusting the values in accordance with the importance of the region in which the error has occurred.

The models in [141,145] are designed for H.264/AVC coded SD and HD videos where a support vector machine (SVM) classifier has been used to assess the video quality based on the visibility of packet loss. By the same research group, the packet layer model presented in [146] uses video resolution, bitrate, packet loss rate, and some information of the codec settings to design a quality estimator for H.264/AVC- and MPEG-2-based encoded videos. An improvement on such statistical parameters-based models is found in [147] where temporal and spatial characteristics of a video are estimated from the packet header to build a content-adaptive model for quality assessment. The no-reference method presented in [148] is based on a nonlinear relationship between an objective quality metric and the quality-related parameters. To make it computationally simple, the authors have used only two parameters, namely, packet loss rate and the value of the interval between intra-frames of a video.

In [149], the authors have presented preliminary results of their investigation into streamlining the impacts of different conditions of packet loss over visible degradation to classify packet loss as visible or invisible. The parameters used in the decision making are extracted from the encoded bitstream. This model was tested for SD resolution H.264/AVC coded videos. If 25% or less subjects perceived an artifact, such a packet loss event was classified as invisible. If 75% or more subjects perceived an artifact, the corresponding packet loss event was classified as visible. In this case, the artifacts perceived by subjects between 25% and 75% were not accounted for at all. This issue was addressed in the authors' later contribution [150] where all artifacts perceived by less than 75% subjects were classified as invisible. Moreover, they extended the model by including more quality-relevant parameters and generalized it by testing it on HD videos. The authors applied the same model for High Efficiency Video Coding (HEVC) encoded videos to examine its cross-standard performance, as reported in [151]. It was observed that the artifact visibility slightly increases while changing from H.264/AVC to HEVC-based video coding.

1.6.3 Bitstream layer model

In the bitstream-based methods, bitstream layer models have access to most of the data that can be used for the video quality estimation. The work item parametric non-intrusive bitstream assessment of video media streaming quality (P.NBAMS) [152] in its mode 1 (parsing mode) is related to the bitstream layer models. In this mode, it is allowed to do any kind of analysis of the bitstream except the usage of the pixel data. The input information includes parameters extracted from the packet header and payload. Besides the parameters included in the parametric models, this model uses QP, DCT coefficients of the coded video, and pixel information. This makes the model comparatively more complex but it generally offers better performance. A low-complexity solution of video quality prediction based on bitstream extracted parameters is found in [153]. The features used are mainly related to the encoding parameters and are taken on sequence level. Low complexity has been achieved by using a simple multilinear regression system for building the relationship between the parameters and quality values. An improvement of this approach is presented in [154] where the required number of parameters has been reduced for computational efficiency and the prediction accuracy has been improved by the virtue of the usage of an ANN. A further improvement is found in [155] where a larger features set is used and the prediction of subjective MOS is also performed. A set of 48 bitstream parameters related to slice coding type, coding modes, various statistics of motion vectors, and QP value was used in [156] to predict the quality of high-definition television (HDTV) video encoded by H.264/AVC. PLSR was used as tool for regression between the feature set and subjective assessment. This method outperformed the authors' earlier contribution [129] and some contemporary objective methods of visual quality assessment.

H.264/AVC employs an in-loop filter to suppress blocking, and this filter has a specific parameter called boundary strength (BS) assigned to transform blocks. Statistics of BS combined with QP and average bitrate has been used in [157] to predict quality of H.264/AVC encoded videos. The proposed method formulates a linear combination of these parameters and a linear regression was conducted to determine its relationship with the subjective assessment scores. A motion-based visual quality estimation method was proposed in [158] for H.264/AVC encoded videos. In this method, some statistical features related to motion vectors along with bitrate and frame rate are calculated. PCA is used to identify the parameters most influential in the estimation of video quality value. Finally, the selected set of features is fed to an equation of quality computation. The inclusion of motion features into the reference-free quality assessment

is justified by the fact that the reduction in visual quality is less for a certain level of compression when the motion is low, for example, the case of videos with static scenes.

A PSNR estimator for H.264/AVC encoded video is presented in [159] where bitrate, QP value, and coding mode are used as the features for quality prediction. The method given in [160] uses QP and block coding mode parameters for quality prediction of H.264/AVC encoded videos.

Based on an opinion model from ITU-T [140], an automatic QoE monitoring method is proposed in [161]. It depends on the network level information derived from packet loss pattern and loss rank of a frame in a group of pictures (GOP) and a measure of motion vectors to represent motion activity to train an ANN model against subjective scores of expert viewers.

In [162], the authors proposed a framework for quality estimation where a QoS parameter, packet loss rate, is combined with spatial and temporal complexities of a video. Usually, a complete decoding of the video is required to estimate its spatial and temporal complexity as these complexity values are generally obtained by an average measure of the pixel variance of codeblocks in a frame. However, the authors have proposed a method of estimating spatial and temporal complexity from the bitstream information only. Specifically, they have developed a rate-distortion model for QP value and bitrate which helps in estimating the complexity measure. Combining this complexity estimate with effects of packet loss delivers a measure of frame quality. Temporal domain quality degradation is computed through occurrences of frame freeze or frame loss. An overall estimate of the video quality is made by a pooling scheme which integrates the spatial and temporal quality indicators. The authors have argued that the suggested method can be used for real-time video services due to its fair accuracy and efficiency in computational cost.

In [163], the impact of compression on quality estimated through MSE prediction using DCT coefficients data [164] is combined with (i) a packet loss model similar to the one presented in ITU-T Recommendation G. 1070 [140], (ii) a frame type-dependent packet loss model, and (iii) a frame type- and error pattern-dependent model separately. It was concluded from the obtained results that a combination of [164] and (iii) offers the best prediction of visual quality of these three combinations.

Bitstream layer methods can also utilize the DCT coefficients data of the encoded image or video, as it can be obtained by partial decoding [138]. There are several such methods which make a quality estimate based on the statistics of the DCT coefficient values. Eden [165] has proposed an algorithm for estimation of PSNR using the assumption that the probability distribution function

(pdf) of DCT coefficients follows Laplacian distribution for H.264/AVC encoded videos. A modified Laplacian model for estimation of DCT coefficients distribution has been presented in [166] for JPEG images. The authors proposed to use maximum likelihood with linear prediction estimates to compute the parameter λ (lambda) of the Laplacian pdf, where λ is a parameter of the distribution. Investigation of the correlation between distribution parameters at adjoining frequencies and integration of the prediction results using maximum-likelihood parameters are the key components of this method. They have also used Watson's model [167] for perceptual weighting of local error estimates in an image. The method given in [166] has been upgraded to be workable for videos in [168]. Here, the video quality predictor has a local error assessment unit, besides having statistics from motion vectors. These values are passed to a perceptual spatio-temporal model that incorporates the HVS sensitivity to produce the visual quality score. Two more methods based on the similar approach from these authors are PSNR estimation for H.264/AVC encoded videos [169] and PSNR estimation for MPEG-2 encoded videos as given in [170].

Contrary to the assumption of Laplacian distribution to model DCT coefficients, it has been argued in [171] that a Cauchy distribution better suits the H.264/AVC encoded data in the process of quality estimation. The proposed approach has been found to be better than the Laplacian distribution [165] in terms of bias between the actual and estimated values of PSNR.

The authors in [172] have used DCT basis functions to evaluate kurtosis measure of images for quality assessment. Three different kinds of kurtosis measures have been made, namely, 1D kurtosis based on frequency band, basis function-based 1D kurtosis, and 2D kurtosis. However, the proposed scheme is meant only for images degraded by blur and it has been tested on LIVE [14] data set JPEG2000 encoded images.

Nishikawa et al. presented a PSNR estimation method of JPEG2000 coded videos in [173] which is actually a no-reference version of their earlier article that needed reference information [174]. The method estimates the PSNR by using wavelet coefficients from the neighboring frames of the frame which has lost some compressed codeblocks. It is assumed that the effect of packet loss upon codeblocks is possible to compute at the receiver end, given that only packet loss occurs and no bit errors exist.

1.6.4 Discussion
Bitstream-based methods of VQA have recently received a significant attention for their computational simplicity and applications in the online quality monitoring. Potentially, the main advantage of these methods is the variety in choice of the features which can be used for quality estimation that in turn means the privilege of adapting to the desired level of complexity. As compared to pixel-based processing, the bitstream-based methods have special advantage of having access to readily available information such as bitrate, frame rate, QP, motion vectors, and various types of information regarding the impacts of network impairments. However, these methods are coding scheme specific that makes them less generally applicable. In the case of parametric planning models, the performance of quality estimation remains limited due to the constraints of the information that can be obtained from the allowed level of access to the bitstream. Packet layer models have better performance with popular application in intermediate nodes of a network as they do not need complex processing and decryption of the data. Bitstream layer models are superior in the performance and the complexity can be flexible depending upon the desired level of accuracy. For possible future works in this area, some comparative performance reports of various models, such as the ones presented in [139,175] would be useful to further accelerate the research in designing better bitstream-based VQA approaches. As we notice in the summary of bitstream-based methods in Table 3, the research community has mostly embraced H.264/AVC-based coding for the design of such methods. It would be advantageous to develop such methods for other popular coding standards as well. Moreover, analysis of the features relevant for quality estimation for the recently approved ITU-T standard of video coding, namely, H.265/HEVC [176] would be useful. For example, in [177], it has been shown that the existing methods of MSE estimation are not feasible for HEVC as it has significantly different coding structure as compared to the previous standards.
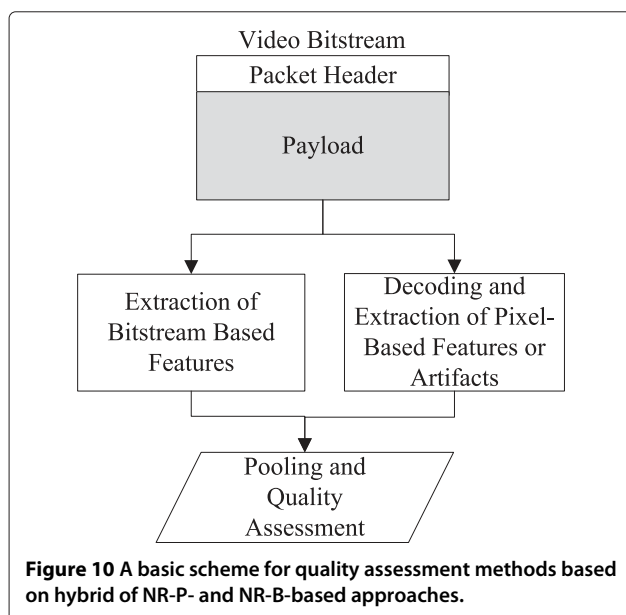
1.7 Hybrid of NR-P and NR-B methods
There are no-reference visual quality estimation methods which combine features from the coded bitstream and some statistics from the decoded media. This type of methods inherits the simplicity of computation from the bitstream-based approaches, and further accuracy in quality estimation is achieved by adding input from the pixel-based approaches. Therefore, such methods can avoid some of the difficulties involved in the pixel and bitstream-based methods [178]. One such example is the fusion of artifacts like blocking or blurring with parameters derived from motion vectors to build up a quality estimation method. The work item P.NBAMS [152] in its mode 2 (full decoding mode) is related to the hybrid models where the information from the coded bitstream as well as reconstructed video can be used. Figure 10 gives an overview of the methodology used in this type of methods. Essentially, the choice of the features for extraction from

Table 3 Characteristic summary of NR-B and hybrid metrics

Method	Reference	Processing	Resolution	Test data[a]	Performance[b]
Bitstream-based	Saad et al. [106]	Multiple	768×512	LIVE image	SC = 0.8
	Saad et al. [107]	Multiple	768×512	LIVE image	SC = 0.93, PC = 0.93
	Yamada et al. [144]	H.264/AVC	$1,440 \times 1,080$	-	PC = 0.85
	Rossholm et al. [153]	H.264/AVC	QCIF	288 test videos	PC (PEVQ) = 0.95
	Shahid et al. [154]	H.264/AVC	QCIF	288 test videos	PC (PEVQ) = 0.98
	Shahid et al. [155]	H.264/AVC	QCIF, CIF	120 test videos	PC = 0.98
	Keimel et al. [156]	H.264/AVC	HD	-	PC = 0.93
	Lee [157]	H.264/AVC	QCIF	13 SRCs	PC = 0.9
	Ries et al. [158]	H.264/AVC	QCIF, SIF, CIF	-	PC = 0.80
	Yang et al. [162]	MPEG-4	QCIF	-	PC = 0.93, SC = 0.93
	Eden [165]	H.264/AVC	HD	5 SRCs	PC (PSNR) = 0.99
	Brandão et al. [166]	JPEG	768×512	LIVE image	PC = 0.97, SC = 0.97
	Brandão et al. [168]	H.264	-	50 test videos	PC = 0.93, SC = 0.95
	Zhang et al. [172]	JPEG2000	768×512	LIVE image	PC = 0.9
	Ichigaya et al. [191]	MPEG-2	SD, HD	26 test videos	PC (PSNR) = 0.95
Hybrid of NR-P and NR-B	Farias et al. [181]	H.264/AVC	CIF	5 SRCs	Regression (PSNR) = 0.79
	Shanableh [182]	MPEG-2	CIF	-	PC = 0.93
	Shanableh [183]	MPEG-2	CIF	-	NRMSE = 0.046
	Davis et al. [184]	H.264/AVC	-	18 test videos	PC = 0.91
	Yamada et al. [186]	H.264/AVC	SD	-	PC = 0.95
	Sugimoto et al. [189]	H.264/AVC	HD	14 SRCs	PC = 0.91

[a]The used test database; [b]performance denotes the correlation with subjective assessment, unless stated otherwise. SRC, source sequence; NRMSE, normalized root mean square error; PC, Pearson's correlation coefficient; SC, Spearman's rank order correlation coefficient.

Figure 10 A basic scheme for quality assessment methods based on hybrid of NR-P- and NR-B-based approaches.

bitstream or pixel domain depends on the design requirements of a method, the availability of a certain type of data for quality estimation, and the encoding scheme. The discussion on this class of methods is divided into two categories, namely, pixel-based and bitstream-based features or artifacts, and statistics of transform coefficients.

1.7.1 Pixel-based and bitstream-based features or artifacts

Video quality-related features and measures of artifacts can be computed both from the pixel and bitstream data and can be pooled for an overall quality estimate. One such method which focuses on quantifying the perceptual quality of H.264/AVC encoded videos degraded by loss of packets in the IP networks is presented in [179]. The error incurred due to packet loss becomes propagative due to the two types of coding predictions involved in H.264/AVC encoders, namely, intra-prediction (spatial) and inter-prediction (temporal) at the encoder end. Even more errors can be introduced while the decoder tries to conceal for the prediction residuals and/or motion vectors lost due to missing packets in the IP bitstream. For simulating the packet loss conditions, a packet loss rate in the range [0.1, 20]% with error patterns generated

using a two-state Gilbert model set for average burst length of three packets was used. Quantitatively, the measures involved in the modeling of the proposed method encompass the impact of errors due to concealment, errors propagated due to loss of reference MBs, and the channel-induced degradation due to H.264/AVC-specific coding techniques. The calculations of these distortions are done on the macroblock level, and the resulting values are summed up to frame and sequence levels. It has been observed that the proposed method yields results which bear good correlation with SSIM [94]. Another method was presented by the same authors in their earlier published contribution [180] where the effects of loss of motion vector information and prediction residuals were incorporated for quality estimation. A method in which transmission and compression artifacts are integrated for VQA is presented in [181]. The constituents of the method are estimations of blockiness, blurring, and packet loss ratio.

Two MPEG-4 encoded video quality prediction methods based on several MB level statistics, derived from bitstream and reconstructed videos, are reported in [182] for PSNR and in [183] for SSIM. A plethora of bitstream-based and pixel-based features at macroblock level have been used in these two methods. One of the distinctive aspects of these two contributions is the usage of different models for system identification between the parameters and the corresponding quality index. In the method targeted for PSNR estimation, spectral regression and reduced model polynomial network have been employed. A multipass prediction system based on stepwise regression has been used in the estimation of SSIM method. The statistical features in both of the methods constitute mainly the coding information of an MB, some relative measures of motion vector of neighboring MBs, and some numerical values related to the texture of an MB.

Average QP values were combined with pixel difference contrast measure values to offer a visual quality method in [184]. The authors have shown that the method outperforms PSNR for a wide range of bitrates of H.264/AVC encoded videos. Similarly, two parametric models have been combined in [185] to design a hybrid model of perceptual quality for H.264/AVC encoded videos. This method uses average value of QP and an average measure of contrast from the decoded video, besides having input from noise masking property of the video content.

A hybrid of bitstream-based and pixel domain quality estimator is proposed in [186]. It has been argued that a video quality estimation merely based on the amount of impaired macroblocks could be erroneous as, in modern video decoders, some error concealment methods are applied to cure the impaired macroblocks and this concealment is not accounted for in such estimations. As the error concealment may not always be effective,

the proposed method uses motion intensity and luminance discontinuity measures to estimate the number of impaired macroblocks for which error concealment remains ineffective. In essence, the visual quality, in terms of MSE, is estimated directly based on the macroblocks for which the error concealment could not perform well. The same authors have generalized this approach for three methods of error concealment and a different value of packet length in [187].

In order to estimate the impact of transmission errors on the quality of H.264/AVC encoded videos, a saliency map-based method is proposed in [188]. Color, contrast, and luminance information has been used to compute spatial saliency map, while motion vector information, readily available in coded bitstream of a video, has been exploited for the computation of temporal saliency maps. A squared sum of spatial and temporal saliency maps has been used to pool them together for computing the overall spatio-temporal map. Accordingly, this map is used for weighting of an error map for each video frame to calculate the value of the proposed model.

Another hybrid method of perceptual quality measurement, which is based on information from the bitstream and spatio-temporal image features, is presented in [189]. The weighted Minkowski method is employed to integrate the average quantization scale with their proposed measures of flickering and blocking for H.264/AVC encoded videos.

A framework for a hybrid method for videos transmitted over long term evolution (LTE) networks is proposed in [190]. It suggests to include parameters from packet layer (packet loss rate, packet size), bitstream layer (frame error, frame duration), and media layer (blurring, blocking) for estimation of the quality. However, a suitable pooling scheme to integrate these parameters into a quality indication value remains as a future work.

1.7.2 Statistics of transform coefficients

In some cases, the transform coefficients can be obtained through partial decoding of the coded bitstream data and features from bitstream as well as pixel domain can be combined for the quality estimation. One such example is found in [191] where an estimate of PSNR has been computed for MPEG-2 coded videos using DCT coefficients. This is actually an improved version of the authors' earlier contribution [192] in which they modeled the distribution of DCT coefficients as a Laplacian pdf to calculate PSNR of the video frames one-by-one for all types, i.e., I, P, and B frames. However, it lacks in accuracy of assessment for B frames. Therefore, the authors conjectured that this happens as a result of fall in the amount of DCT coefficients information which is available for B frames due to processes of rate control and motion compensation. Henceforth, a hybrid approach to resolve this issue

has been found in [191] where picture energy has also been used in addition to DCT coefficients. There is a significant improvement of correlation with estimated and actual PSNR, in the case when the proposed method was tested on SDTV and HDTV sequences.

1.7.3 Discussion

The hybrid methods use not only pixel-based information but also bitstream-based information, which in turn makes the hybrid framework having a potential of being the most accurate quality estimator as compared to the other approaches [193]. Thus, the importance of careful combination of the features from pixel and bitstream domains is evident. Further studies are needed to investigate the interaction among various types of artifacts due to compression and transmission and the joint impact towards the video quality.

Various approaches exist on how to combine the impact of various sources of degradation into one representative value for all the artifacts under consideration. In the recommendation ITU-T P.1202.1 that presents a complementary algorithm of NR quality assessment for the recommendation P.NBAMS [152], four types of degradations are sorted with respect to their impact on quality. The values of the two most significant artifact types are pooled together through a linear combination. A higher weighting is applied to the artifact value that is found to be the most significant out of the four types. As different artifact types can exist in different range of values, it is important that all of them are aligned to the same scale before the sorting is applied. Besides linear combination, some contributions [189] adopt the Minkowski metric [3] for pooling the values of different artifacts into a single quantity.

With regards to the preference on which a pooling strategy should be chosen it may depend on many factors including relative severity of different artifacts, spatio-temporal characteristics of the contents, and the presence of masking effects. Linear combination is more valid if the constituents can be related to the quality value through a linear regression. While combining different artifacts through a linear relation, different artifacts can be given different significance. For example, more weight is given to the impact of bitstream layer features than to media-layer features in the hybrid model given in [194]. On the other hand, the Minkowski metric of summation has its roots in additive properties of low-level vision. Therefore, it is required to find the suitable value of its exponent through measurements.

Most of the aforementioned hybrid methods make an assessment of quality in terms of MSE but this measure of quality is known to be rather inaccurate in representing the perceptual quality [195]. This fact necessitates the desire of enhancing such methods for better correlation

with subjective quality assessment. As can be seen from the summary of hybrid methods in Table 3, the main focus of the development of hybrid methods has been on videos.

2 Conclusions

Motivated by the growing interest in NR methods of quality assessment for images and videos, we have presented a classification and review of recent approaches proposed in this research area. The available contributions have been classified into different categories based on the methodologies used in the design. Recognized classifications and standardizations in the area have been extrapolated to introduce our approach of classification. The new classification enabled us to present a review of a large amount of recently published work. On the highest tier, three categories have been identified to group the existing NR methods. The NR methods that employ pixel domain approaches for quality assessment are called NR-P-based methods, and the methods that employ encoded bitstream and parametric information of the media signal are called NR-B-based methods. The third category is called hybrid methods which are designed by a composite of NR-P- and NR-B-based methods. A further subcategorization has been presented to organize the discussion of the review.

It is observed that the majority of the publications introduce methods that are processed in the pixel domain. This trend can be attributed to the rich heritage of work in the image processing area. In most cases, pixel-based methods require more processing power than bitstream-based methods. NR quality estimation is a widely adopted application in the area of online quality monitoring. It is thus required to employ computationally less complex methods. This fact necessitates to focus towards designing bitstream-based or hybrid methods. The distortions present in a network can introduce a variety of temporal perturbations in a video transmitted through it. Such perturbations have to be monitored by service providers to ensure a given threshold of visual quality at the end users' premises. This can be performed using NR metrics which estimate the impact of degradation in the temporal domain. Unfortunately, most of the existing methods are designed to account for a single or a limited set of degradations. Therefore, it is not easy to make an estimate of the overall loss of visual quality. Hence, methods which can make a complete assessment of the quality are desirable. Similarly, attention can be drawn towards designing less complex methods which are not application specific, such as the methods that are not limited to a particular coding standard.

In the context of the reviewed methods, it is interesting to compare the approaches adopted in case of IQA and VQA methods. In case of IQA, the main focus has been on addressing the most occurring spatial artifacts

such as blocking, blurring, and ringing as a result of popular compression techniques such as JPEG and JPEG2000. Besides many methods that are specifically designed for particular artifacts and hence have limited application, it is exciting to see many methods that are not restricted to a specific artifact type and have a wider application area. In such methods of global application, the mostly adopted approach is based on computing the impact of distortions on natural scene statistics of natural images. This also suggests that such approaches may not be applied to artificial images (such as animations and cartoons). This issue can be considered as a challenge for future work in IQA. More focus has been seen on the development of bitstream-based approaches in the case of VQA methods. This is of advantage in the sense that bitstream-based approaches have relatively low computational complexity. However, they face the drawbacks of being coding scheme specific and sometimes less accurate. We believe that the development of more robust approaches based on hybrids of NR-P and NR-B methods may be beneficial to meet these challenges associated with the NR VQA area.

We observe that many of the existing contributions in NR IQA and VQA have reported the results of the proposed methods therein by doing the relevant performance tests on the publicly available test databases. This is useful for independent benchmarking and performance comparison tests of these methods by other researchers. Therefore, more variety in the content and resolution of the media available through public test databases would be of great value. On the other hand, one general drawback of many existing methods of NR quality assessment lies in the limited use of the test data, as the data used for the designing or training of a metric is often also used for its performance verification. This drawback actually does not allow to draw meaningful results from such studies. Also, it has been observed that most of the existing methods for video quality assessment are limited to one encoder implementation or rather one particular setting of an encoder. Hence, cross-encoder design of VQA metrics would be a useful contribution. Moreover, we enlist the following trends and directions towards future research in the area of NR quality assessment:

- The trend of contribution in the NR quality estimation has been settling towards finding approaches of lesser complexity as shown by the growing interest in the bitstream-based methods. However, bitstream-based methods face the challenge of being limited for a specific codec. Given the fact that such methods have been shown to have promising performance by having reasonable values of correlation with subjective quality assessment, it would be advantageous to generalize the methodologies of these methods for diverse coding schemes.

- The performance of the bitstream-based methods has been found to be largely content dependent, as the spatio-temporal perceptual complexities vary with varying content, and, in turn, the nature of the features used for quality estimation also changes. However, in the case of pixel-based methods, it is relatively easier to differentiate the content characteristics. Thus, it is required for bitstream-based models to be trained on a sufficiently high variety in content, enabling them to be used in practice. Future inventions can focus on the development of methods that can be applied in more general scenarios with the desired amount of variety in content.

- The existing NR methods are usually designed and tested for cases where the quality difference is beyond the threshold of perception of an artifact, i.e., rather clearly visible. However, attention needs to be paid to scenarios where the test stimuli may already be of high quality. Future developments should therefore envisage the degradations that are considered in the category of subthreshold artifacts. The need of such methods becomes even more important with regards to the newly approved HEVC standard that supports ultra-high definition video resolutions of 4K and beyond.

- It has been observed that emphasis is being put towards making the quality estimation more in line with the perceived quality by HVS. In the future, the NR quality assessment methods should continue to adapt for HVS parameters and further advancements in the understanding of HVS, such as attention-driven foveated quality assessment models [196] should be taken into consideration.

- A robust combination of audio-visual quality estimators can be devised for designing scenario-dependent models. For example, in quality monitoring of sports video, more emphasis can be put on the visual component than audio as the viewers might be interested more in video. For example, a video of a football match would draw more focus towards visual scene than audio, as compared to news or head and shoulder scenario. Moreover, audio-visual quality estimation is challenging due to the complex interplay of HVS preferences. In terms of the mutual impact of audio-visual subjective quality, some studies report an average cross-modal interaction level of 0.5 MOS [197] to above 2 MOS points [198] on a scale of 1 to 5 quality rating.

- Given the presented comprehensive literature review, it has been observed that developments of NR methods that consider visual attention are rather limited, especially, in the case of videos. As noted in [199], visual attention models can be integrated into existing NR methods to make them more robust. Generally,

the advantage of including visual attention-based modeling appears to be larger for methods designed for video quality assessment than for image quality assessment methods. Visual attention becomes more significant in scenarios of audio-visual stimuli, as it is required to account for the cues from visual channels as well as auditory channels.

- To make the quality estimation closer to the subjective assessment, intelligent methods are needed that consider the content preference and expectations of humans in a given scenario. For example, the subjective quality assessment results mentioned in [200] indicate that desirable content is rated significantly higher than the undesirable and neutral contents.
- The task of finding the optimal trade-off between temporal and spatial resolution, and the level of quantization for its impacts on the perceptual quality in different scenarios of application, is challenging [201]. This issue should be taken into consideration for the future development of NR methods.
- In order to combine independent and isolated approaches for the development of hybrid NR VQA methods, a five-point agenda has been identified by joint effort group (JEG) at VQEG [202]. We believe that such collaborative works will be instrumental in paving the ways of NR VQA towards a measurable evolution.

We believe that our contribution in this article can be utilized and extended in various ways. One can use this review as a systematic literature review to perform comparisons on a class of NR methods using the same image or video test database to highlight the state-of-the-art. Furthermore, this review can be very useful for the beginner researchers in this area to achieve a concise yet comprehensive overview of the field. This way, we expect this contribution to be instrumental for future research and development in the area of NR visual quality assessment. Moreover, a possible future work is to survey the contributions for audio-visual quality assessment based on NR paradigm, similar to [203] that deals with FR methods of audio-visual quality assessment.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MS prepared the database of the publications in the area, streamlined the publications to be reviewed in the paper, planned the structure of the paper in consultation with co-authors, and wrote the main draft of the paper. AR participated in writing some parts of the paper and involved in the discussions on improving the manuscript. BL took part in the initial discussions on the paper topic and gave feedback during the writing process on the structure, the content and its revisions, and wrote the first draft of the conclusions. HJZ helped with refining and structuring the content of the review; participated in drafting, correcting, and revising the article; and helped with assessing the reviewed approaches. All authors read and approved the final manuscript.

References
1. Cisco Visual Networking Index, Global mobile data traffic forecast update, 2013-2018. Cisco white paper (2014)
2. ITU, ITU-R Recommendation BT.500-13 Methodology for the subjective assessment of the quality of the television pictures. http://www.itu.int/rec/R-REC-BT.500-13-201201-I/en. Accessed 4 November 2013
3. HR Wu, KR Rao, *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. (CRC, Boca Raton, 2005)
4. SS Hemami, AR Reibman, No-reference image and video quality estimation: applications and human-motivated design. Signal Process. Image Commun. **25**(7), 469–481 (2010)
5. Z Wang, AC Bovik, Reduced- and no-reference image quality assessment. IEEE Signal Process. Mag. **28**(6), 29–40 (2011)
6. W Lin, C-c Jay Kuo, Perceptual visual quality metrics: a survey. J. Vis. Commun. Image Representation **22**(4), 297–312 (2011)
7. M Vranješ, S Rimac-Drlje, K Grgić, Review of objective video quality metrics and performance comparison using different databases. Signal Process. Image Commun. **28**(1), 1–19 (2013)
8. S Chikkerur, V Sundaram, M Reisslein, LJ Karam, Objective video quality assessment methods: a classification, review, and performance comparison. IEEE Trans. Broadcasting **57**(2), 165–182 (2011)
9. R Soundararajan, A Bovik, Survey of information theory in visual quality assessment. Signal, Image Video Process. **7**(3), 391–401 (2013)
10. DM Chandler, Seven challenges in image quality assessment: past, present, and future research. ISRN Signal Processing 53 (2013). (doi:10.1155/2013/905685)
11. U Engelke, H-J Zepernick, Perceptual-based quality metrics for image and video services: a survey, in *EuroNGI Conference on Next Generation Internet Networks* (Trondheim, 21–23 May 2007), pp. 190–197
12. AR Reibman, S Sen, JV der Merwe, Analyzing the spatial quality of internet streaming video, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, January 2005)
13. VQEG, Final report from the Video Quality Experts Group on the validation of objective models of multimedia quality assessment, phase I (2008). http://www.its.bldrdoc.gov/vqeg/projects/multimedia-phase-i/multimedia-phase-i.aspx. Accessed 11 April 2014
14. HR Sheikh, Z Wang, L Cormack, Bovik AC, LIVE Image Quality Assessment Database. http://live.ece.utexas.edu/research/quality. Accessed 11 April 2014
15. Video Quality Experts Group (VQEG). http://www.its.bldrdoc.gov/vqeg/downloads.aspx. Accessed 11 April 2014
16. N Ponomarenko, V Lukin, A Zelensky, K Egiazarian, M Carli, F Battisti, TID2008 - a database for evaluation of full reference visual quality assessment metrics. Adv. Modern Radioelectron. **10**, 30–45 (2009)
17. PL Callet, F Autrusseau, Subjective quality assessment IRCCyN/IVC database (2005). http://www.irccyn.ec-nantes.fr/ivcdb/. Accessed 11 April 2014
18. ZMP Sazzad, Y Kawayoke, Horita Y, Image Quality Evaluation Database. http://mict.eng.u-toyama.ac.jp/mictdb.html. Accessed 11 April 2014
19. L Liang, S Wang, J Chen, S Ma, D Zhao, W Gao, No-reference perceptual image quality metric using gradient profiles for JPEG2000. Signal Process. Image Commun. **25**(7), 502–516 (2010)
20. S Winkler, *Digital Video Quality: Vision Models and Metrics*. (Wiley, Chichester, 2005)
21. N Narvekar, L Karam, A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). IEEE Trans. Image Process. **20**(9), 2678–2683 (2011)
22. A Ciancio, ALNT da Costa, EAB da Silva, A Said, R Samadani, P Obrador, No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans. Image Process. **20**(1), 64–75 (2011)
23. X Marichal, W-Y Ma, H Zhang, Blur determination in the compressed domain using DCT information, in *IEEE International Conference on Image Processing*, vol. 2 (Kobe, 24–28 October 1999), pp. 386–390
24. H Tong, M Li, H Zhang, C Zhang, Blur detection for digital images using wavelet transform, in *IEEE International Conference on Multimedia and Expo*, vol. 1 (Taipei, 27–30 June 2004), pp. 17–20

25. P Marziliano, F Dufaux, S Winkler, T Ebrahimi, Perceptual blur and ringing metrics: application to JPEG2000. Signal Process. Image Commun. **19**(2), 163–172 (2004)

26. R Ferzli, LJ Karam, Human visual system based no-reference objective image sharpness metric, in *IEEE International Conference on Image Processing*, (2006), pp. 2949–2952

27. A Ciancio, ALN da Costa, EAB da Silva, A Said, R Samadani, P Obrador, Objective no-reference image blur metric based on local phase coherence. Electron. Lett. **45**(23), 1162–1163 (2009)

28. R Hassen, Z Wang, M Salama, No-reference image sharpness assessment based on local phase coherence measurement, in *IEEE International Conference on Acoustics Speech and Signal Processing* (Dallas, 14–19 March 2010), pp. 2434–2437

29. R Ferzli, LJ Karam, No-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans. Image Process. **18**(4), 717–728 (2009)

30. R Ferzli, LJ Karam, A no-reference objective image sharpness metric based on just-noticeable blur and probability summation, in *IEEE International Conference on Image Processing*, vol. 3 (San Antonio, TX, USA, 16–19 September 2007), pp. 445–448

31. NG Sadaka, LJ Karam, R Ferzli, GP Abousleman, A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling, in *15th IEEE International Conference on Image Processing* (San Diego, 12–15 October 2008), pp. 369–372

32. ND Narvekar, LJ Karam, A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection, in *International Workshop on Quality of Multimedia Experience* (San Diego, 29–31 July 2009), pp. 87–91

33. N Narvekar, LJ Karam, An improved no-reference sharpness metric based on the probability of blur detection, in *Workshop on Video Processing and Quality Metrics* (Scottsdale, 13–15 January 2010)

34. S Varadarajan, LJ Karam, An improved perception-based no-reference objective image sharpness metric using iterative edge refinement, in *IEEE International Conference on Image Processing* (San Diego, 12–15 October 2008), pp. 401–404

35. L Debing, C Zhibo, M Huadong, X Feng, G Xiaodong, No reference block based blur detection, in *International Workshop on Quality of Multimedia Experience* (San Diego, 29–31 July 2009), pp. 75–80

36. P Marziliano, F Dufaux, S Winkler, T Ebrahimi, A no-reference perceptual blur metric, in *Proceedings of the International Conference on Image Processing 2002*, vol. 3 (Rochester, New York, 22–25 September 2002), pp. 57–60

37. A Maalouf, M-C Larabi, A no reference objective color image sharpness metric, in *The European Signal Processing Conference* (Aalborg, 23–27 August 2010), pp. 1019–1022

38. S Wu, W Lin, Z Lu, EP Ong, S Yao, Blind blur assessment for vision-based applications, in *IEEE International Conference on Multimedia and Expo* (Beijing, 2–5 July 2007), pp. 1639–1642

39. S Wu, W Lin, S Xie, Z Lu, EP Ong, S Yao, Blind blur assessment for vision-based applications. J. Vis. Commun. Image Representation **20**(4), 231–241 (2009)

40. C Chen, W Chen, JA Bloom, A universal reference-free blurriness measure. SPIE J. Image Q. Syst. Performance VIII **7867**, 78670–7867014 (2011)

41. M-J Chen, AC Bovik, No-reference image blur assessment using multiscale gradient, in *International Workshop on Quality of Multimedia Experience* (San Diego, 29–31 July 2009), pp. 70–74

42. R Ferzli, L Karam, A no reference objective sharpness metric using riemannian tensor, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 25–26 January 2007), pp. 25–26

43. A Chetouani, A Beghdadi, M Deriche, A new reference-free image quality index for blur estimation in the frequency domain, in *IEEE International Symposium on Signal Processing and Information Technology* (Ajman, 14–17 December 2009), pp. 155–159

44. Z Hua, Z Wei, C Yaowu, A no-reference perceptual blur metric by using OLS-RBF network, in *Pacific-Asia Workshop on Computational Intelligence and Industrial Application* (Wuhan, 19–20 December 2008), pp. 1007–1011

45. C Oprea, I Pirnog, C Paleologu, M Udrea, Perceptual video quality assessment based on salient region detection, in *Advanced International Conference on Telecommunications* (Venice, 24–28 May 2009), pp. 232–236

46. C Chen, JA Bloom, A blind reference-free blockiness measure, in *Proceedings of the Pacific Rim Conference on Advances in Multimedia Information Processing: part I* (Shanghai, 21–24 September 2010), pp. 112–123

47. H Liu, I Heynderickx, A no-reference perceptual blockiness metric, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Las Vegas, 31 March to 4 April 2008), pp. 865–868

48. H Liu, I Heynderickx, A perceptually relevant no-reference blockiness metric based on local image characteristics. EURASIP J. Adv. Signal Process. **2009**, 1–15 (2009)

49. G Zhai, W Zhang, X Yang, W Lin, Y Xu, No-reference noticeable blockiness estimation in images. Signal Process. Image Commun. **23**(6), 417–432 (2008)

50. RV Babu, S Suresh, A Perkis, No-reference JPEG-image quality assessment using GAP-RBF. Signal Process. **87**, 1493–1503 (2007)

51. RV Babu, AS Bopardikar, A Perkis, OI Hillestad, No-reference metrics for video streaming applications, in *International Workshop on Packet Video* (Irvine, 13–14 December 2004)

52. Z Zhang, H Shi, S Wan, A novel blind measurement of blocking artifacts for H.264/AVC video, in *International Conference on Image and Graphics* (Xian, 20–23 September 2009), pp. 262–265

53. Z Hua, Z Yiran, T Xiang, A weighted Sobel operator-based no-reference blockiness metric, in *Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, vol. 1 (Wuhan, 19–20 December 2008), pp. 1002–1006

54. C-S Park, J-H Kim, S-J Ko, Fast blind measurement of blocking artifacts in both pixel and DCT domains. J. Math. Imaging Vis. **28**, 279–284 (2007)

55. S Suthaharan, No-reference visually significant blocking artifact metric for natural scene images. Signal Process. **89**(8), 1647–1652 (2009)

56. H Liu, N Klomp, I Heynderickx, A perceptually relevant approach to ringing region detection. IEEE Trans. Image Process. **19**(6), 1414–1426 (2010)

57. H Liu, N Klomp, I Heynderickx, A no-reference metric for perceived ringing artifacts in images. IEEE Trans. Circuits Syst. Video Technol. **20**(4), 529–539 (2010)

58. B-X Zuo, J-W Tian, D-L Ming, A no-reference ringing metrics for images deconvolution, in *International Conference on Wavelet Analysis and Pattern Recognition*, vol. 1 (Hong Kong, 30–31 August 2008), pp. 96–101

59. B-X Zuo, D-L Ming, J-W Tian, Perceptual ringing metric to evaluate the quality of images restored using blind deconvolution algorithms. SPIE J. Opt. Eng. **48**(3), 037004 (2009)

60. AC Bovik, *Handbook of Image and Video Processing*, 2nd edn. (Elsevier Academic, Burlington, 2005)

61. A Amer, E Dubois, Fast and reliable structure-oriented video noise estimation. IEEE Trans. Circuits Syst. Video Technol. **15**(1), 113–118 (2005)

62. J Tian, L Chen, Image noise estimation using a variation-adaptive evolutionary approach. IEEE Signal Process. Lett. **19**(7), 395–398 (2012)

63. SI Olsen, Estimation of noise in images: an evaluation. CVGIP: Graph. Models Image Process. **55**(4), 319–323 (1993)

64. K Rank, M Lendl, R Unbehauen, Estimation of image noise variance. IEE Proc. Vis. Image Signal Process. **146**(2), 80–84 (1999)

65. A Amer, E Dubois, Fast and reliable structure-oriented video noise estimation. IEEE Trans. Circuits Syst. Video Technol. **15**(1), 113–118 (2005)

66. M Ghazal, A Amer, Homogeneity localization using particle filters with application to noise estimation. IEEE Trans. Image Process. **20**(7), 1788–1796 (2011)

67. X Liu, M Tanaka, M Okutomi, Noise level estimation using weak textured patches of a single noisy image, in *19th IEEE International Conference on Image Processing (ICIP)* (Orlando, 30 September to 3 October 2012), pp. 665–668

68. X Liu, M Tanaka, M Okutomi, Single-image noise level estimation for blind denoising. IEEE Trans. Image Process. **22**(12), 5226–5237 (2013)

69. S Pyatykh, J Hesser, L Zheng, Image noise level estimation by principal component analysis. IEEE Trans. Image Process. **22**(2), 687–699 (2013)

70. Q Huynh-Thu, M Ghanbari, Temporal aspect of perceived quality in mobile video broadcasting. IEEE Trans. Broadcasting **54**(3), 641–651 (2008)

71. S Borer, A model of jerkiness for temporal impairments in video transmission, in *International Workshop on Quality of Multimedia Experience* (Trondheim, 21–23 June 2010), pp. 218–223

72. Q Huynh-Thu, M Ghanbari, No-reference temporal quality metric for video impaired by frame freezing artefacts, in *IEEE International Conference on Image Processing* (Cairo, 7–10 November 2009), pp. 2221–2224

73. ITU-T Recommendation J.247: objective perceptual multimedia video quality measurement in the presence of a full reference. http://www.itu.int/rec/T-REC-J.247/en. Accessed 4 November 2013

74. G Yammine, E Wige, F Simmet, D Niederkorn, A Kaup, Blind frame freeze detection in coded videos, in *Picture Coding Symposium* (Krakow, 7–9 May 2012), pp. 341–344

75. S Wolf, A no reference (NR) and reduced reference (RR) metric for detecting dropped video frames, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 15–16 January 2009)

76. K Watanabe, J Okamoto, T Kurita, Objective video quality assessment method for evaluating effects of freeze distortion in arbitrary video scenes. Proc. SPIE-Image Qual. Syst. Perform. IV **6494**, 64940–1649408 (2007)

77. K Watanabe, J Okamoto, T Kurita, Objective video quality assessment method for freeze distortion based on freeze aggregation. Proc. SPIE- Image Qual. Syst. Perform. III **6059**, 1–8 (2007)

78. K-C Yang, CC Guest, K El-Maleh, PK Das, Perceptual temporal quality metric for compressed video. IEEE Trans. Multimedia **9**(7), 1528–1535 (2007)

79. RR Pastrana-Vidal, J-C Gicquel, Automatic quality assessment of video fluidity impairments using a no-reference metric, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 22–24 January 2006)

80. RR Pastrana-Vidal, J-C Gicquel, A no-reference video quality metric based on a human assessment model, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 25–26 January 2007)

81. F Battisti, ANM Carli, No-reference quality metric for color video communication, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 19–20 January 2012)

82. T Oelbaum, C Keimel, K Diepold, Rule-based no-reference video quality evaluation using additionally coded videos. IEEE J. Select. Topics Signal Process. **3**(2), 294–303 (2009)

83. MG Choi, JH Jung, JW Jeon, No-reference image quality assessment using blur and noise. Int. J. Comput. Sci. Eng. **3**(2), 76–80 (2009)

84. E Cohen, Y Yitzhaky, No-reference assessment of blur and noise impacts on image quality. Signal, Image Video Process. **4**(3), 289–302 (2010)

85. S Gabarda, G Cristóbal, Blind image quality assessment through anisotropy. J. Opt. Soc. Am. A **24**(12), 42–51 (2007)

86. X Zhu, P Milanfar, A no-reference sharpness metric sensitive to blur and noise, in *International Workshop on Quality of Multimedia Experience* (San Diego, 29–31 July 2009), pp. 64–69

87. X Zhu, P Milanfar, A no-reference image content metric and its application to denoising, in *IEEE International Conference on Image Processing* (Hong Kong, 26–29 September 2010), pp. 1145–1148

88. M Narwaria, W Lin, Objective image quality assessment based on support vector regression. IEEE Trans. Neural Netw. **21**(3), 515–519 (2010)

89. H Liu, J Redi, H Alers, R Zunino, I Heynderickx, ed. by BE Rogowitz, TN Pappas, No-reference image quality assessment based on localized gradient statistics: application to JPEG and JPEG2000, in *SPIE Proceedings-Human Vision and Electronic Imaging*, vol. 7527, (2010), p. 75271

90. R Ferzli, L Karam, A novel free reference image quality metric using neural network approach, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (Scottsdale, 13–15 January 2010), pp. 1–4

91. A Chetouani, A Beghdadi, A new image quality estimation approach for JPEG2000 compressed images, in *IEEE International Symposium on Signal Processing and Information Technology* (Bilbao, 14–17 December 2011), pp. 581–584

92. HR Sheikh, AC Bovik, L Cormack, No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Trans. Image Process. **14**(11), 1918–1927 (2005)

93. P Romaniak, L Janowski, MI Leszczuk, Z Papir, Perceptual quality assessment for H.264/AVC compression, in *IEEE International Workshop on Future Multimedia Networking* (Las Vegas, 14–17 January 2012)

94. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

95. X Liu, M Chen, T Wan, C Yu, Hybrid no-reference video quality assessment focusing on codec effects. Korea Internet Inform. Soc. Trans. Internet Inform. Syst. **5**(3), 592–606 (2011)

96. S Ouni, E Zagrouba, M Chambah, M Herbin, No-reference image semantic quality approach using neural network, in *IEEE International Symposium on Signal Processing and Information Technology* (Bilbao, 14–17 December 2011), pp. 106–113

97. J Zhang, TM Le, SH Ong, TQ Nguyen, No-reference image quality assessment using structural activity. Signal Process. **91**(11), 2575–2588 (2011)

98. DL Ruderman, The statistics of natural images. Network: Comput. Neural Syst. **5**(4), 517–548 (1994)

99. J Zhou, B Xiao, Q Li, A no reference image quality assessment method for JPEG2000, in *IEEE International Joint Conference on Neural Networks* (Hong Kong, 1–8 June 2008), pp. 863–868

100. W Lu, K Zeng, D Tao, Y Yuan, X Gao, No-reference image quality assessment in contourlet domain. Neurocomputing **73**, 784–794 (2010)

101. MN Do, M Vetterli, The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Process. **14**(12), 2091–2106 (2005)

102. J Shen, Q Li, G Erlebacher, Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images. IEEE Trans. Image Process. **20**(8), 2089–2098 (2011)

103. AK Moorthy, AC Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)

104. K Zhu, K Hirakawa, V Asari, D Saupe, A no-reference video quality assessment based on Laplacian pyramids, in *20th IEEE International Conference on Image Processing (ICIP)* (Melbourne, 15–18 September 2013), pp. 49–53

105. MA Saad, AC Bovik, Natural motion statistics for no-reference video quality assessment, in *International Workshop on Quality of Multimedia Experience* (San Diego, 29–31 July 2009), pp. 163–167

106. MA Saad, AC Bovik, C Charrier, A DCT statistics-based blind image quality index. IEEE Signal Process. Lett. **17**(6), 583–586 (2010)

107. MA Saad, AC Bovik, C Charrier, DCT statistics model-based blind image quality assessment, in *IEEE International Conference on Image Processing* (Brussels, 11–14 September 2011), pp. 3093–3096

108. A Mittal, A Moorthy, A Bovik, No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)

109. A Mittal, GS Muralidhar, J Ghosh, AC Bovik, Blind image quality assessment without human training using latent quality factors. IEEE Signal Process. Lett. **19**(2), 75–78 (2012)

110. M-J Chen, LK Cormack, AC Bovik, No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. **22**(9), 3379–3391 (2013)

111. P Gastaldo, G Parodi, J Redi, R Zunino, No-reference quality assessment of JPEG images by using CBP neural networks, in *17th International Conference Artificial Neural Networks* (Porto, 9–13 September 2007). Lecture Notes in Computer Science, vol. 4669, (Springer, Heidelberg), pp. 564–572

112. C Li, AC Bovik, X Wu, Blind image quality assessment using a general regression neural network. IEEE Trans. Neural Netw. **22**(5), 793–799 (2011)

113. J Zhang, TM Le, A new no-reference quality metric for JPEG2000 images. IEEE Trans. Consumer Electron. **56**(2), 743–750 (2010)

114. ZMP Sazzad, Y Kawayoke, Y Horita, No reference image quality assessment for JPEG2000 based on spatial features. Image Commun. **23**, 257–268 (2008)

115. ZMP Sazzad, Y Kawayoke, Y Horita, Spatial features based no reference image quality assessment for JPEG2000, in *IEEE International Conference*

on Image Processing, vol. 3 (San Antonio, 16–19 September 2007), pp. 517–520

116. A Maalouf, M-C Larabi, A no-reference color video quality metric based on a 3D multispectral wavelet transform, in *International Workshop on Quality of Multimedia Experience* (Trondheim, 21–23 June 2010), pp. 11–16

117. S Suresh, R Venkatesh Babu, HJ Kim, No-reference image quality assessment using modified extreme learning machine classifier. Appl. Soft Comput. **9**, 541–552 (2009)

118. AM Treisman, G Gelade, A feature-integration theory of attention. Cognit. Psychol. **12**(1), 97–136 (1980)

119. S Yao, E Ong, MH Loke, Perceptual distortion metric based on wavelet frequency sensitivity and multiple visual fixations, in *IEEE International Symposium on Circuits and Systems* (Seattle, 18–21 May 2008), pp. 408–411

120. X-H Wang, Z Ming, A new metric for objectively assessing the quality of enhanced images based on human visual perception. J. Optoelectron. Laser **19**(2), 254–262 (2008)

121. M Ries, C Crespi, O Nemethova, M Rupp, Content based video quality estimation for H.264/AVC video streaming, in *IEEE Conference on Wireless Communications and Networking* (Kowloon, 11–15 March 2007), pp. 2668–2673

122. M Ries, O Nemethova, M Rupp, Performance evaluation of mobile video quality estimators, in *European Signal Processing Conference* (Poznan, 3–7 September 2007)

123. A Khan, L Sun, E Ifeachor, Content-based video quality prediction for MPEG4 video streaming over wireless networks. J. Multimedia **4**(4), 228–239 (2009)

124. A Khan, L Sun, E Ifeachor, An ANFIS-based hybrid video quality prediction model for video streaming over wireless networks, in *International Conference on Next Generation Mobile Applications, Services and Technologies* (Cardiff, 16–19 September 2008), pp. 357–362

125. A Khan, L Sun, E Ifeachor, Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks, in *IEEE International Conference on Communications* (Dresden, 14–18 June 2009), pp. 1–5

126. U Engelke, H-J Zepernick, An artificial neural network for quality assessment in wireless imaging based on extraction of structural information, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1 (Honolulu, 15–20 April 2007), pp. 1249–1252

127. X Jiang, F Meng, J Xu, W Zhou, No-reference perceptual video quality measurement for high definition videos based on an artificial neural network, in *International Conference on Computer and Electrical Engineering* (Phuket, 20–22 December 2008), pp. 424–427

128. D Čulibrk, D Kukolj, P Vasiljević, M Pokrić, V Zlokolica, Feature selection for neural-network based no-reference video quality assessment. Lecture Notes Comput. Sci. **5769**, 633–642 (2009)

129. C Keimel, T Oelbaum, K Diepold, No-reference video quality evaluation for high-definition video, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Taipei, 19–24 April 2009), pp. 1145–1148

130. ZMP Sazzad, S Yamanaka, Y Kawayokeita, Y Horita, Stereoscopic image quality prediction, in *International Workshop on Quality of Multimedia Experience*, (2009), pp. 180–185

131. D Bhattacharjee, S Prakash, P Gupta, No-reference image quality assessment for facial images, in *International Conference on Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence,* Zhengzhou, vol. 6839 (Springer Heidelberg, Germany, 11–14 August 2011), pp. 594–601

132. OYG Castillo, Survey about facial image quality. Technical Report. Fraunhofer Institute for Computer Graphics Research (2007), Darmstadt, Germany, (2006)

133. JE Caviedes, F Oberti, No-reference quality metric for degraded and enhanced video, in *Proceedings of SPIE 5150, Visual Communications and Image Processing*, vol. 5150 (Lugano, 8 July 2003), pp. 621–632

134. G Valenzise, S Magni, M Tagliasacchi, S Tubaro, Estimating channel-induced distortion in H.264/AVC video without bitstream information, in *Second International Workshop on Quality of Multimedia Experience (QoMEX)* (Trondheim, 21–23 June 2010), pp. 100–105

135. G Valenzise, S Magni, M Tagliasacchi, S Tubaro, No-reference pixel video quality monitoring of channel-induced distortion. IEEE Trans. Circuits Syst. Video Technol. **22**(4), 605–618 (2012)

136. J Park, K Seshadrinathan, S Lee, AC Bovik, Video quality pooling adaptive to perceptual distortion severity. IEEE Trans. Image Process. **22**(2), 610–620 (2013)

137. AR Reibman, VA Vaishampayan, Y Sermadevi, Quality monitoring of video over a packet network. IEEE Trans. Multimedia **6**(2), 327–334 (2004)

138. A Takahashi, D Hands, V Barriac, Standardization activities in the ITU for a QoE assessment of IPTV. IEEE Commun. Mag. **46**(2), 78–84 (2008)

139. F Yang, S Wan, Bitstream-based quality assessment for networked video: a review. IEEE Commun. Mag. **50**(11), 203–209 (2012)

140. ITU-T, ITU-T Recommendation G.1070: opinion model for video-telephony applications (2012). http://www.itu.int/rec/T-REC-G.1070. Accessed 11 April 2014

141. A Raake, MN Garcia, S Moller, J Berger, F Kling, P List, J Johann, C Heidemann, T-V-model: parameter-based prediction of IPTV quality, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Las Vegas, 31 March to 4 April 2008), pp. 1149–1152

142. J Han, Y-H Kim, J Jeong, J Shin, Video quality estimation for packet loss based on no-reference method, in *IEEE International Conference on Advanced Communication Technology*, vol. 1 (Phoenix Park, Dublin, 7–10 February 2010), pp. 418–421

143. ITU, ITU-T Recommendation P.1201: Parametric non-intrusive assessment of audiovisual media streaming quality (2012). http://handle.itu.int/11.1002/1000/11727. Accessed 11 April 2014

144. T Yamada, S Yachida, Y Senda, M Serizawa, Accurate video-quality estimation without video decoding, in *IEEE International Conference on Acoustics Speech and Signal Processing* (Dallas, 14–19 March 2010), pp. 2426–2429

145. S Argyropoulos, A Raake, M-N Garcia, P List, No-reference bit stream model for video quality assessment of H.264/AVC video based on packet loss visibility, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Prague, 22–27 May 2011), pp. 1169–1172

146. M-N Garcia, A Raake, Parametric packet-layer video quality model for IPTV, in *10th International Conference on Information Sciences Signal Processing and Their Applications (ISSPA)* (Kuala Lumpur, 10–13 May 2010), pp. 349–352

147. F Yang, J Song, S Wan, HR Wu, Content-adaptive packet-layer model for quality assessment of networked video services. IEEE J. Select. Topics Signal Process. **6**(6), 672–683 (2012)

148. I Sedano, K Brunnström, M Kihl, A Aurelius, Full-reference video quality metric assisted the development of no-reference bitstream video quality metrics for real-time network monitoring. EURASIP J. Image Video Process (2014). doi:10.1186/1687-5281-2014-4

149. N Staelens, N Vercammen, Y Dhondt, B Vermeulen, P Lambert, R Van de Walle, P Demeester, VIQID: a no-reference bit stream-based visual quality impairment detector, in *International Workshop on Quality of Multimedia Experience* (Trondheim, 21–23 June 2010), pp. 206–211

150. N Staelens, G Van Wallendael, K Crombecq, N Vercammen, J De Cock, B Vermeulen, R Van de Walle, T Dhaene, P Demeester, No-reference bitstream-based visual quality impairment detection for high definition H.264/AVC encoded video sequences. IEEE Trans. Broadcasting **58**(2), 187–199 (2012)

151. G Van Wallendael, N Staelens, L Janowski, J De Cock, P Demeester, R Van de Walle, No-reference bitstream-based impairment detection for high efficiency video coding, in *Fourth International Workshop on Quality of Multimedia Experience* (Yarra Valley, 5–7 July 2012), pp. 7–12

152. ITU, ITU-T Recommendation P.1202: parametric non-intrusive bitstream assessment of video media streaming quality (2012). http://handle.itu.int/11.1002/1000/11730. Accessed 11 April 2014

153. A Rossholm, B Lövström, A new low complex reference free video quality predictor, in *IEEE Workshop on Multimedia Signal Processing* (Cairns, 8–10 October 2008), pp. 765–768

154. M Shahid, A Rossholm, B Lövström, A reduced complexity no-reference artificial neural network based video quality predictor, in *International Congress on Image and Signal Processing*, vol. 1 (Shanghai, 15–17 October 2011), pp. 517–521

155. M Shahid, A Rossholm, B Lövström, A no-reference machine learning based video quality predictor, in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)* (Klagenfurt, 3–5 July 2013), pp. 176–181

156. C Keimel, M Klimpke, J Habigt, K Diepold, No-reference video quality metric for HDTV based on H.264/AVC bitstream features, in *IEEE*

International Conference on Image Processing (Brussels, 11–14 September 2011), pp. 3325–3328

157. S-O Lee, K-S Jung, D-G Sim, Real-time objective quality assessment based on coding parameters extracted from H.264/AVC bitstream. IEEE Trans. Consumer Electron. **56**(2), 1071–1078 (2010)

158. M Ries, O Nemethova, M Rupp, Motion based reference-free quality estimation for H.264/AVC video streaming, in *International Symposium on Wireless Pervasive Computing* (San Juan, 5–7 February 2007)

159. M Slanina, V Ricny, R Forchheimer, A novel metric for H.264/AVC no-reference quality assessment, in *EURASIP Conference on Speech and Image Processing Multimedia Communications and Services* (Maribor, 27–30 June 2007), pp. 114–117

160. I Park, T Na, M Kim, A noble method on no-reference video quality assessment using block modes and quantization parameters of H.264/AVC, in *Proceedings of SPIE-Image Quality and System Performance VIII*, vol. 7867 (San Francisco Airport, 23 January 2011), pp. 78670–7867011

161. KD Singh, G Rubino, No-reference quality of experience monitoring in DVB-H networks, in *Wireless Telecommunications Symposium* (Tampa, 21–23 April 2010), pp. 1–6

162. F Yang, S Wan, Q Xie, HR Wu, No-reference quality assessment for networked video via primary analysis of bit stream. IEEE Trans. Circuits Syst. Video Technol. **20**(11), 1544–1554 (2010)

163. M Chin, T Brandão, MP Queluz, Bitstream-based quality metric for packetized transmission of H.264 encoded video, in *International Conference on Systems, Signals and Image Processing* (Vienna, 11–13 April 2012), pp. 312–315

164. T Brandão, M Chin, MP Queluz, From PSNR to perceived quality in H.264 encoded video sequences, in *Proceedings of EuroITV* (Lisbon, 29 June to 1 July 2011)

165. A Eden, No-reference estimation of the coding PSNR for H.264-coded sequences. IEEE Trans. Consumer Electron. **53**(2), 667–674 (2007)

166. T Brandão, PQ Maria, No-reference image quality assessment based on DCT domain statistics. Signal Process. **88**(4), 822–833 (2008)

167. AB Watson, DCT quantization matrices visually optimized for individual images, in *Proceedings of SPIE, Human Vision, Visual Processing, and Digital Display IV*, vol. 1913 (San Jose, 31 January 1993), pp. 202–216

168. T Brandão, PQ Maria, No-reference quality assessment of H.264/AVC encoded video. IEEE Trans. Circuits Syst. Video Technol. **20**(11), 1437–1447 (2010)

169. T Brandão, PQ Maria, No-reference PSNR estimation algorithm for H.264 encoded video sequences, in *European Signal Processing Conference* (Laussane, 25–29 August 2008)

170. T Brandão, PQ Maria, Blind PSNR estimation of video sequences using quantized DCT coefficient data, in *Picture Coding Symposium* (Lausanne, 12–13 November 2007)

171. S-Y Shim, J-H Moon, J-K Han, PSNR estimation scheme using coefficient distribution of frequency domain in H.264 decoder. Electron. Lett. **44**(2), 108–109 (2008)

172. J Zhang, SH Ong, TM Le, Kurtosis-based no-reference quality assessment of JPEG2000 images. Image Commun. **26**, 13–23 (2011)

173. K Nishikawa, K Munadi, H Kiya, No-reference PSNR estimation for quality monitoring of motion JPEG2000 video over lossy packet networks. IEEE Trans. Multimedia **10**(4), 637–645 (2008)

174. K Nishikawa, S Nagawara, H Kiya, QoS estimation method for JPEG2000 coded image at RTP layer. IEICE Trans. **E89-A**(8), 2119–2128 (2006). doi:10.1093/ietfec/e89-a.8.2119

175. J Joskowicz, R Sotelo, JCL Ardao, Towards a general parametric model for perceptual video quality estimation. IEEE Trans. Broadcasting **59**(4), 569–579 (2013)

176. ITU, ITU-T Recommendation H.265/high efficiency video coding. http://www.itu.int/rec/T-REC-H.265-201304-I. Accessed 11 April 2014

177. B Lee, M Kim, No-reference PSNR estimation for HEVC encoded video. IEEE Trans. Broadcasting **59**(1), 20–27 (2013)

178. S Winkler, P Mohandas, The evolution of video quality measurement: from PSNR to hybrid metrics. IEEE Trans. Broadcasting **54**(3), 660–668 (2008)

179. M Naccari, M Tagliasacchi, S Tubaro, No-reference video quality monitoring for H.264/AVC coded video. IEEE Trans. Multimedia **11**(5), 932–946 (2009)

180. M Naccari, M Tagliasacchi, F Pereira, S Tubaro, No-reference modeling of the channel induced distortion at the decoder for H.264/AVC video coding, in *IEEE International Conference on Image Processing* (San Diego, 12–15 October 2008), pp. 2324–2327

181. MCQ Farias, MM Carvalho, HTM Kussaba, BHA Noronha, A hybrid metric for digital video quality assessment, in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (Nuremberg, 8–10 June 2011), pp. 1–6

182. T Shanableh, No-reference PSNR identification of MPEG-4 video using spectral regression and reduced model polynomial networks. IEEE Signal Process. Lett. **17**(8), 735–738 (2010)

183. T Shanableh, Prediction of structural similarity index of compressed video at a macroblock level. IEEE Signal Process. Lett. **18**(5), 335–338 (2011)

184. AG Davis, D Bayart, DS Hands, Hybrid no-reference video quality prediction, in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (Bilbao, 13–15 May 2009), pp. 1–6

185. K Yamagishi, T Kawano, T Hayashi, Hybrid video-quality-estimation model for IPTV services, in *IEEE Global Telecommunications Conference, (GLOBECOM)* (Honolulu, 30 November to 4 December 2009), pp. 1–5

186. T Yamada, Y Miyamoto, M Serizawa, No-reference video quality estimation based on error-concealment effectiveness, in *International Workshop on Packet Video* (Lausanne, 12–13 November 2007), pp. 288–293

187. T Yamada, Y Miyamoto, T Nishitani, No-reference quality estimation for video-streaming services based on error-concealment effectiveness. IEICE Trans. Fundamentals Electron. Commun. Comput. Sci. **E95-A**(11), 2007–2014 (2012)

188. H Boujut, J Benois-Pineau, T Ahmed, O Hadar, P Bonnet, A metric for no-reference video quality assessment for HD TV delivery based on saliency maps, in *IEEE International Conference on Multimedia and Expo (ICME)* (Barcelona, 11–15 July 2011), pp. 1–5

189. O Sugimoto, S Naito, S Sakazawa, A Koike, Objective perceptual video quality measurement method based on hybrid no reference framework, in *IEEE International Conference on Image Processing* (Cairo, 7–10 November 2009), pp. 2237–2240

190. S Zhao, H Jiang, Q Cai, S Sherif, A Tarraf, Hybrid framework for no-reference video quality indication over LTE networks, in *23rd Wireless and Optical Communication Conference (WOCC)* (Newark, 9–10 May 2014), pp. 1–5

191. A Ichigaya, Y Nishida, E Nakasu, Nonreference method for estimating PSNR of MPEG-2 coded video by using DCT coefficients and picture energy. IEEE Trans. Circuits Syst. Video Technol. **18**(6), 817–826 (2008)

192. A Ichigaya, M Kurozumi, N Hara, Y Nishida, E Nakasu, A method of estimating coding PSNR using quantized DCT coefficients. IEEE Trans. Circuits Syst. Video Technol. **16**(2), 251–259 (2006)

193. VQEG, Draft VQEG Testplan: Hybrid Perceptual/Bitstream Group (2012). http://www.its.bldrdoc.gov/vqeg/projects/hybrid-perceptual-bitstream/hybrid-perceptual-bitstream.aspx. Accessed 11 April 2014

194. C Keimel, J Habigt, K Diepold, Hybrid no-reference video quality metric based on multiway PLSR, in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* (Bucharest, 27–31 August 2012), pp. 1244–1248

195. Z Wang, AC Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process. Mag. **26**(1), 98–117 (2009)

196. J You, T Ebrahimi, A Perkis, Attention driven foveated video quality assessment. IEEE Trans. Image Process. **23**(1), 200–213 (2014)

197. B Belmudez, S Möller, Audiovisual quality integration for interactive communications. EURASIP J. Audio Speech Music Process. **2013**(1) (2013). 10.1186/1687-4722-2013-24

198. AN Rimell, MP Hollier, RM Voelcker, The influence of cross-modal interaction on audio-visual speech quality perception, in *Audio Engineering Society Convention 105* (San Francisco, 26–29 September 1998)

199. U Engelke, H Kaprykowsky, H-J Zepernick, P Ndjiki-Nya, Visual attention in quality assessment. IEEE Signal Process. Mag. **28**(6), 50–59 (2011)

200. P Kortum, M Sullivan, The effect of content desirability on subjective video quality ratings. Hum. Factors: J. Hum. Factors Ergonom. Soc. **52**, 105–118 (2010)

201. A Rossholm, M Shahid, B Lövström, Analysis of the impact of temporal, spatial, and quantization variations on perceptual video

quality, in *IEEE Network Operations and Management Symposium (NOMS)* (Krakow, 5–9 May 2014), pp. 1–5

202. M Barkowsky, I Sedano, K Brunnström, M Leszczuk, N Staelens, Hybrid video quality prediction: re-viewing video quality measurement for widening application scope. Multimedia Tools Appl (2014). doi:10.1007/s11042-014-1978-2

203. J You, U Reiter, MM Hannuksela, M Gabbouj, A Perkis, Perceptual-based quality assessment for audio-visual services: a survey. Signal Process. Image Commun. **25**(7), 482–501 (2010)