Efficient neural-network–based no-reference approach to an overall quality metric for JPEG and JPEG2000 compressed images

Hantao Liu
Judith Redi
Hani Alers
Rodolfo Zunino
Ingrid Heynderickx

Efficient neural-network–based no-reference approach to an overall quality metric for JPEG and JPEG2000 compressed images

Hantao Liu
Judith Redi
Hani Alers

Delft University of Technology
Department of Mediamatics
Delft, The Netherlands
E-mail: hantao.liu@tudelft.nl

Rodolfo Zunino

University of Genoa
Department of Biophysical and Electronic Engineering
Genoa, Italy

Ingrid Heynderickx

Delft University of Technology
Department of Mediamatics
Delft, The Netherlands
and
Philips Research Laboratories
Group Visual Experiences
Eindhoven, The Netherlands

Abstract. *Reliably assessing overall quality of JPEG/JPEG2000 coded images without having the original image as a reference is still challenging, mainly due to our limited understanding of how humans combine the various perceived artifacts to an overall quality judgment. A known approach to avoid the explicit simulation of human assessment of overall quality is the use of a neural network. Neural network approaches usually start by selecting active features from a set of generic image characteristics, a process that is, to some extent, rather ad hoc and computationally extensive. This paper shows that the complexity of the feature selection procedure can be considerably reduced by using dedicated features that describe a given artifact. The adaptive neural network is then used to learn the highly nonlinear relationship between the features describing an artifact and the overall quality rating. Experimental results show that the simplified feature selection procedure, in combination with the neural network, indeed are able to accurately predict perceived image quality of JPEG/JPEG2000 coded images. © 2011 SPIE and IS&T. [DOI: 10.1117/1.3664181]*

1 Introduction

Understanding and evaluating image quality has become increasingly important for a broad range of applications, such

as the optimization of digital imaging systems, the benchmarking of image and video coding algorithms, and the quality monitoring and control in displays.^{1,2} Traditionally, image quality has been evaluated by human subjects, and a mean opinion score (MOS) represents the image quality perceived by the average viewer. When conducted properly, subjective experiments are considered as the most reliable means of assessing image quality. However, performing subjective experiments is very time-consuming, very expensive, and often too slow to be useful in real-world applications. Therefore, during the last decades, a lot of research effort has been devoted to the development of objective metrics that can automatically and quantitatively predict perceived image quality.

Objective metrics reported in literature range from dedicated metrics that measure a specific image distortion to general metrics that assess the perceived overall quality. The various approaches can be classified into full-reference (FR), reduced-reference (RR), and no-reference (NR). FR metrics measure the similarity or fidelity between the distorted image and its original version, where the latter is considered as a distortion-free reference. The most widely used FR metrics are the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR), both aiming at an overall quality assessment.³ Improved alternatives for these two basic metrics include models as described in Refs. 4–6, and some

Paper 11013PR received Jan. 25, 2011; revised manuscript received Oct. 8, 2011; accepted for publication Nov. 7, 2011; published online Dec. 1, 2011.

1017-9909/2011/20(4)/043007/15/\$25.00 © 2011 SPIE and IS&T

recently developed metrics as the structural similarity (SSIM) index⁷ and the visual information fidelity (VIF) index.⁸ Since FR metrics require access to the original, which is mostly not available at the receiver end of an imaging chain, their applicability is limited to in-lab (off-line) testing of image and video processing algorithms. RR metrics are mainly used in scenarios where the reference is not fully available, e.g., in complex communication networks. They make use of certain features extracted from the reference, which are then employed as side information to evaluate the quality of a distorted image (see, for example, Refs. 9–11). Instead, in imaging systems with broadcasted content, NR metrics, in which the quality prediction is based on the distorted image only, i.e., without any reference, are more practical. Designing NR metrics, however, is still challenging, partly due to the limited understanding of how humans assess image quality.

Recently, considerable progress has been made in the development of NR metrics. Most NR metrics (see, e.g., Refs. 12–19) are dedicated metrics measuring a specific type of artifact created by a specific image distortion process. Examples are a metric measuring sensor noise, a metric measuring ringing or blockiness as a consequence of signal compression, or a metric measuring blur generated during acquisition. In such a scenario, the design of the NR metric can make use of the specific characteristics of the artifact, and therefore, generally obtains a higher reliability with respect to the related perceived quality degradation. Specific NR metrics are, for example, used to tune the setting of various parameters in the algorithms of a video chain in current TVs (see, e.g., Refs. 20–22). In addition, they can be combined to predict the perceived overall image quality. Various examples of this approach are given in the literature (see, e.g., Refs. 23–25); a ringing metric and a blur metric are often combined to assess the overall image quality of wavelet-based compression.²⁴ This approach, however, largely depends on the reliability of each of the artifact specific models, and on the efficiency of their combination in a perceptually meaningful way.

An alternative approach for combining individual, dedicated metrics to an estimate for overall image quality is given in Ref. 26, in which natural scene statistics are used to blindly determine the overall quality of images compressed by JPEG2000. The approach relies on the assumption that natural images usually exhibit strong statistical regularities, and therefore, reside in a tiny area of the space containing all possible images. Based on this assumption, it quantifies overall image quality by detecting variations in the statistics of image features in the wavelet domain. The approach is promising, but relies heavily on the sophisticated and computationally expensive modeling of natural scene statistics.

Instead of precisely modeling specific artifacts or natural scene statistics, NR image quality assessment has also been formulated as a machine learning problem. Based on pioneering work (such as in Refs. 27 and 28) and some more recent research (as, for example, in Refs. 29–32), this approach has proved to be effective for the overall quality prediction of a specific distortion type, e.g., JPEG and MPEG-2 compression. It treats the human visual system (HVS) as a black box, whose input–output relationship between image characteristics and a quality rating is to be learned by compu-

tational intelligent tools, such as a neural network (NN). The problem is generally formulated as a regression or function approximation, and the data needed for training are obtained from subjective experiments. During training the error between the desired output (i.e., the subjective quality rating) and the model prediction is minimized. At run time, the properly trained machine implements the resulting model without requiring further computational effort. The critical step in this approach, however, is the selection of the active features, effectively describing the perceived quality. In general, a considerable number of image features is extracted as input to the NN. These so-called common features may be pixel-based as in Refs. 27–31, or HVS-based as in Ref. 32. In both cases, however, the feature selection requires considerable effort toward optimization, and it is hard to guarantee minimization of the model's complexity at sufficiently high prediction accuracy.

In this paper, we propose to combine the advantages of the two approaches mentioned so far, i.e., the use of (aspects of) artifact-specific metrics as features, and the use of an NN to assess the overall perceived quality. We apply this approach to overall quality estimation of JPEG/JPEG2000 compressed images. JPEG/JPEG2000 compression is widely used in current digital imaging systems, involving the internet. Having a reliable NR metric embedded in these systems is of fundamental importance to optimize their performance from a human perception point of view. For example, in the receiving end of an imaging chain, an NR metric can be adopted to adjust the parameter settings of image enhancement strategies, thus optimizing the output of the chain. In addition, the perceived quality needs to be modeled at a complexity that allows implementation in a real-time imaging chain. In practice, our approach has two components: first, we calculate the feature(s) describing the most relevant artifact in JPEG/JPEG2000 compressed images, and second, we use an adaptive NN to learn the highly nonlinear relationship between the feature(s) and the overall quality rating. The use of features dedicated to a single artifact is motivated from results in literature reporting a high correlation between a specific artifact type and the overall quality of JPEG/JPEG2000 compressed images.^{33–36} Our novel approach is highly efficient for two reasons. First, it calculates features based on artifact characteristics, and so, this avoids a lengthy and tedious feature selection procedure. In addition, the usefulness of the selected features for predicting image quality is already known from literature. Second, it leaves the simulation of the HVS for the perceived overall image quality to the NN, and as such this part of the model is reduced after training to the implementation of a simple algorithm at run-time. It should be noted that the whole process only uses the luminance component of the images, thus further reducing the computational load.

Section 2 of this paper discusses the feature-extraction process, and derives the numerical descriptors for the NN that are based on simple yet efficient metrics for both blockiness and blur artifacts. Section 3 describes the actual quality prediction tool, which relies on empirical training of a neural network. Section 4 presents the overall performance of the proposed NR JPEG and JPEG2000 metrics and a comparison with metrics existing in literature. Section 5 is devoted to a discussion of the specific added value of the proposed approach.

Fig. 1 Illustration of the occurrence of compression artifacts: (a) blocking artifacts in an image JPEG compressed with a bit rate of 0.21 bits per pixel (bpp), and (b) blur artifacts in an image JPEG2000 compressed with a bit rate of 0.099 bpp.

2 Feature Extraction and Description

The literature shows that the overall quality of JPEG compressed images is highly correlated with the occurrence of blocking artifacts,³³ while the overall quality of JPEG2000 compressed images is highly correlated with the occurrence of blur.²⁴ A blocking artifact manifests itself as an artificial discontinuity in the image content, which is a direct consequence of the fact that the quantization in JPEG is block-based and that the blocks are quantized independently. A blur artifact occurs in JPEG2000 compressed images mainly due to the loss of high frequency transform coefficients in the wavelet-based coding, a result of which the image signal is smoothened. Figure 1 illustrates the occurrence of blocking artifacts in a JPEG compressed image, and of blur artifacts in a JPEG2000 compressed image, respectively.

Quality degradation as a consequence of compression should then easily be predictable from the extraction of blockiness/blur related image features, provided that an adaptive NN is used to empirically learn the highly nonlinear relationship between these artifact-oriented features and the overall quality rating. To efficiently characterize the local behavior of artifacts and thus feed the neural network with relevant features for image quality prediction, a gradient-based feature extraction scheme is proposed. It contains three basic

components: 1. the localization of the artifacts, 2. the local feature extraction using local gradients in relation to their neighborhood, and 3. the assembling of a global statistical descriptor as input to the neural network. The implementation of each of these steps is detailed below.

2.1 Local Feature Extraction: JPEG

Due to the underlying coding algorithm for JPEG compression, the spatial location of blocking artifacts is very regular. In principle, they occur on a grid of blocks of 8×8 pixels, starting at the top-left corner of an image. In common applications, however, grid sizes may differ and starting positions may shift, either due to perturbations in the incoming signal or as a consequence of spatial scaling. In such a scenario, a (naïve) NR metric might run the risk of calculating blockiness at wrong pixel positions, and therefore might incur a dramatic degradation in accuracy.¹⁷ To ensure that the metric is calculated exactly at block boundaries, a grid detector can be adopted. The research presented in this paper implements the blocking grid detection method proposed in Ref. 17. It is, however, worth stressing that the feature-extraction approach is independent of the particular choice of grid detector, hence any alternative approach (e.g., the one described in Ref. 22)

Fig. 2 Illustration of the template for calculating the local blockiness: (a) two adjacent 8×8 blocks (i.e., A and B) extracted from a real JPEG image, and (b) the gradient profile of the image patch of (a).

can be applied. The blocking grid detector first maps an image onto a one-dimensional (1D) signal profile, in which the periodic property of blocking artifacts is maintained. Then the exact block size, as well as the grid offset, is easily extracted from the discrete Fourier transform of this 1D signal profile.

When the blocking artifacts are (exactly) located, the related feature can be extracted. In this paper, the feature for the JPEG compressed images is based on the visual strength of a blocking artifact within a local area of the image content.¹⁷ Since a blocking artifact is a local edge that stands out from its spatial vicinity, it can be simply defined as relating to the energy present in the gradient at the artifact to the energy present in the gradient in its neighboring pixels. When the luminance channel of an image of $M \times N$ (height $\times$ width) pixels is denoted as $I(i, j)$ for $i \in [1, M]$, $j \in [1, N]$, the local blockiness $L_{\text{blockiness-}h}$ along the horizontal direction at location (i, j) is quantified as

$$L_{\text{blockiness-}h}(i, j) = \frac{G_h(i, j)}{\frac{1}{2n} \sum_{x=-n, \dots, n, x \neq 0} G_h(i, j+x)}, \quad (i, j) \in \{\text{blocking grid}\}, \quad (1)$$

where $G_h(i, j)$ indicates the gradient map along the horizontal direction, and is computed as

$$G_h(i, j) = |I(i, j+1) - I(i, j)|, \quad j \in [1, N-1], \quad (2)$$

and where n determines the size of the template used to describe the local content. The size is determined as a balance between sufficient information of the local content, while avoiding noise from content too far away. In our experiments we used $n = 3$, being equal to half the amount of pixels between two blocking edges.

An example of the template for calculating $L_{\text{blockiness-}h}$ is shown in Fig. 2, where two adjacent blocks of 8×8 pixels (i.e., A and B) are extracted from a real JPEG image. The local blockiness along the vertical direction $L_{\text{blockiness-}v}$ can be calculated similarly. The higher the values of $L_{\text{blockiness-}h}$ and $L_{\text{blockiness-}v}$ are, the larger the distortion of the blocking artifact is. It should, however, be noted that this does not necessarily mean that the blocking artifact is also more visible. The local visibility of a blocking artifact may be affected by texture and luminance masking, which typically occur in the HVS. It has been shown in literature that taking into account these masking effects can be greatly beneficial for the prediction performance of a dedicated NR blockiness metric.¹⁷ However, modeling the HVS introduces more computational power. So, in this paper we avoid the calculation of masking,

and rely on the NN to learn the unknown functional relationship between the extracted gradient-based features and the rating of overall image quality.

2.2 Local Feature Extraction: JPEG2000

In JPEG2000 compression, blur artifacts are perceptually prominent along edges or in textured areas. Hence, in this paper, the local feature extracted for the JPEG2000 compressed images calculates the degree of blur at an edge within a local area of image content. Literature offers a wide variety of techniques to detect strong edges, and consequently to identify the spatial location of blur artifacts (see, for example, Ref. 37, and references therein). The approach implemented here uses a straightforward Sobel edge detector, resulting in a gradient image. The location of strong edges is then extracted by applying a threshold to this gradient image (as such removing noise and insignificant edges). The threshold value is automatically set depending on the image content (e.g., using the mean of the gradient magnitude squared image).

We then use a novel, simple, yet efficient measure for the blur of all detected edges. Instead of calculating the distance between the start and end position of an edge (as proposed in Ref. 12), edge blur is locally defined in the gradient domain as the sharpness of the edge related to its surrounding content within a limited extent. When describing blur simply as the relative gradient energy of an edge compared to its direct vicinity, it can be quantified in the same manner as used in Eq. (1), i.e.,

$$L_{\text{blur-}h}(i, j) = \frac{G_h(i, j)}{\frac{1}{2n} \sum_{x=-n, \dots, n, x \neq 0} G_h(i, j+x)}, \quad (i, j) \in \{\text{strong edges}\}, \quad (3)$$

where $L_{\text{blur-}h}$ indicates the local blur along the horizontal direction and n , representing the size of the template, has the same value as in Eq. (1). $L_{\text{blur-}v}$, i.e., the local blur in the vertical direction, can be calculated similarly. The lower the value of $L_{\text{blur-}h}$ and $L_{\text{blur-}v}$, the larger the distortion of the blur artifact is. Figure 3 explains the reasoning behind the proposed approach of using gradient energy to detect blur. Figures 3(a) and 3(c) show a detected edge [i.e., at location (113, 259)] in the JPEG2000 compressed image of Fig. 1(b), and its intensity profile over the pixels in its direct vicinity, respectively. Figures 3(b) and 3(d) show the corresponding edge in the original uncompressed image of Fig. 1(b), and its intensity profile over the pixels in its direct vicinity. The difference in sharpness between the two edges is clearly revealed in the gradient domain [see Figs. 3(e) and 3(f)]. In correspondence, the values of $L_{\text{blur-}h}$ indicate that the edge of Fig. 3(a) is more blurred than the edge of Fig. 3(b).

It should be noted that we used a very simple edge detector and blur feature extraction in our metric. More complex implementations, such as a careful threshold criterion for the edge map or calculation of blur perpendicular to the detected edge, may be implemented as well. Such implementations, however, make the metric more complicated, while the additional complexity is not needed from a performance point of view (as we will illustrate later in the paper).

Fig. 3 Illustration of the calculation of local blur: (a) image patch extracted from the JPEG2000 compressed image of Fig. 1(b) [the red dot indicates the location of the detected edge at (113, 259) in Fig. 1(b), and the template indicates the area in which the local blur is calculated for this edge], (b) the image patch of the original uncompressed image corresponding to (a), (c) the intensity profile over the pixels within the template of (a), (d) the intensity profile over the pixels within the template of (b), (e) the gradient profile of (c), and (f) the gradient profile of (d).

Fig. 4 Visualization of the extracted features: (a) the JPEG compressed image of Fig. 1(a), (b) the feature map of (a) showing the location and the degree of distortion of the blocking artifacts in the horizontal direction, (c) the JPEG2000 compressed image of Fig. 1(b), and (d) the feature map of (c) showing the location and the degree of distortion of the blur artifacts in the horizontal direction. In (b) and (d), the higher the intensity, the larger the distortion is.

2.3 Global Descriptor of the Image Features

Once the local features related to blocking/blur artifacts are explicitly extracted and calculated for each JPEG/JPEG2000 compressed image, the results can be visualized in a spatially varying feature map. An example is given in Fig. 4 for a JPEG and JPEG2000 compressed image, respectively. Figures 4(b) and 4(d) illustrate the location of the artifacts in the horizontal direction, and the intensity at each pixel indicates the local degree of distortion; i.e., the higher the intensity, the larger the distortion is. The location and intensity of the artifacts in a vertical direction can be obtained in a similar way.

Direct application of all extracted feature values as input to an NN is problematic since the dimension of the space of these values is often too large, and as such inappropriate for the network in terms of training. Existing approaches to reduce the number of feature values (see, e.g., Refs. 29–32) usually calculate a statistical descriptor that characterizes the whole image. This descriptor is a single vector, which needs to be associated with the single quality score generated by human subjects. In this paper, the statistical description of an image feature as proposed in Refs. 29–31 is adopted. It unifies the local feature values of an image to a single vector using percentiles. Having computed the feature values f_i ($i = 1, \dots, N_F$) per image (i.e., $L_{\text{blockiness}}$ calculated in both the

horizontal and vertical direction on the blocking grid or L_{blur} calculated in both the horizontal and vertical direction on the detected edges), these values are sorted in ascending order of magnitude. The envelope of the obtained distribution is then expressed in a global descriptor $\mathbf{f}$ by taking 11 of its percentiles φ :

$$\mathbf{f} = \{\varphi_\alpha; \alpha \in \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}\};$$

$$\varphi_\alpha = \left\lceil \frac{N_F}{100} \alpha + \frac{1}{2} \right\rceil. \quad (4)$$

Figure 5 illustrates the formation of the global descriptor of an image feature. Compared to simply taking the average of the feature values, this spatial pooling strategy allows feeding the nonlinear regression with a more complete overview of the amount and behavior of the considered distortion in the image.

3 NR Image Quality Estimator Based on a Neural Network

As reported in literature already (see, e.g., Refs. 27–32 and 38), we implement an NN to approximate the functional relationship between the image features and the related quality score. The main difference with earlier contributions to literature is that in our case the input feature vector to

Fig. 5 Illustration of the formation of the global descriptor of an image feature: (a) the feature values (i.e., the blocking artifacts along the horizontal and vertical direction) extracted from the JPEG compressed image in Fig. 1(a) and sorted in ascending order of magnitude, and (b) the global descriptor of the image feature, taking 11 percentiles of the distribution of (a).

the NN contains descriptors of the actually occurring artifacts (see Sec. 2). The NN aims to mimic the mechanism of quality perception and avoids an explicit model of the HVS, thus reducing the number of assumptions typically required to model perceived quality analytically. In this paper, a feed-forward NN is employed to operate on the feature vector extracted from JPEG/JPEG2000 images. The implementation of this NN is already described in more detail in Refs. 29–31, and is only briefly repeated here, and illustrated in Fig. 6.

A feed-forward NN aims at implementing a stimulus-response behavior by arranging several elementary units (“neurons”) into a layered structure, which does not allow any feedback between layers. Each neuron involves a simple, nonlinear transformation of weighted inputs, and the nonlinearity is often performed by a sigmoidal function. The multilayer perceptron (MLP) paradigm³⁹ belongs to this type of network, and it has been proved to be able to perform effectively in scenarios where the target mapping function can be determined by a few computing units with global scope. It

Fig. 6 Schematic overview of the architecture of a CBP network. The CBP model includes one additional input to the standard MLP, consisting of the sum of the squares of all the network inputs.

Fig. 7 Schematic overview of the proposed NR metric for the perceived overall quality assessment of JPEG/JPEG2000 compressed images: (a) the NR JPEG metric, and (b) the NR JPEG2000 metric.

intrinsically implements a series expansion of n_h basis functions a_h (i.e., sigmoids), which can be generally expressed as:

$$y(\mathbf{x}) = w_0 + \sum_{h=1}^{n_h} w_h a_h(\mathbf{x}), \quad (5)$$

where $\mathbf{x}$ indicates the stimulus vector with its output value $y(\mathbf{x})$, and the coefficients of $\mathbf{w}$ are called the network “weights” and need to be adjusted during the training phase. The basic scheme of Eq. (5) is usually enhanced by applying a sigmoidal nonlinearity to the output value.

The circular back-propagation (CBP) network⁴⁰ improves the conventional MLP paradigm by adding one more input value, which is the sum of the squared values of all the network inputs. As illustrated in Fig. 6, for an input stimulus vector $\mathbf{x} = \{x_1, \dots, x_{n_i}\}$, the input layer connects the n_i values to each neuron of the “hidden” layer. The j ’th hidden neuron performs a nonlinear transformation of a weighted combination of the input values with coefficients (weights) $w_{j,i}$ ($j = 1, \dots, n_h$, and $i = 1, \dots, n_i$):

$$a_j = \text{sigm} \left(w_{j,0} + \sum_{i=1}^{n_i} w_{j,i} \cdot x_i + w_{j,n_i+1} \sum_{i=1}^{n_i} x_i^2 \right), \quad (6)$$

where $\text{sigm}(z) = (1 + e^{-z})^{-1}$, $w_{j,0}$ is a bias term, and a_j is the neuron activation (i.e., the output of the basis function). The output layer provides the final network response, y_k , ($k = 1$ in the case of image quality assessment):

$$y_k = \text{sigm} \left(w_{k,0} + \sum_{j=1}^{n_h} w_{k,j} \cdot a_j \right), \quad (7)$$

where $w_{k,j}$ and $w_{k,0}$ represent the output coefficients and the output bias, respectively.

The resulting CBP network can map both linear and circular separation boundaries.⁴⁰ The additional input value enhances the overall representation ability of the network, while not affecting the properties of the MLP structure (e.g., $w_{j,n_i+1} = 0$ reduces a CBP network to a classical MLP). Since the actual coefficients of w_{j,n_i+1} are determined by the empirical training process, the selection between a conventional MLP and a CBP model is entirely data-driven and does not require any *a priori* assumption. Such an adaptive behavior makes CBP networks appropriate for perception related problems, whose underlying structure is often obscure.

The degrees of freedom of the NN that need to be fitted are the depth n_h of the series expansion and the weighting coefficients within each neuron. To determine the former quantity, literature provides both theoretical⁴¹ and practical⁴² criteria to ensure prediction accuracy, while minimizing the risk of over-fitting training data. In this paper, we follow an empirical approach⁴² mainly due to its simplicity and proved effectiveness. Once the number of network neurons (i.e., n_h) is decided, a fitting process tunes the set of weights in such a way that the network optimizes the desired input–output mapping, minimizing a cost function which implements the mean square error between the predictions and the subjective quality scores.

4 Evaluation of the Overall Metric Performance

4.1 Test Environment

Figure 7 illustrates the schematic overview of the proposed NR metric for the perceived overall quality assessment of JPEG/JPEG2000 compressed images. It should be mentioned that the approach discussed in this paper still treats the JPEG and JPEG2000 images separately, since the NR metric is feasible only when the prior knowledge about the image distortion process is available.¹ But, we envision that by

Fig. 8 Experimental setup for the K -fold cross-validation: (a) source images of the LIVE database (Ref. 45) assigned over six groups, and (b) distribution of the source images assigned over the training and test set for each of the six runs.

including a classification algorithm of JPEG and JPEG2000 (see, e.g., Ref. 43) the system can automatically select the appropriate metric to use on any compressed image. This is, however, outside the scope of this paper. In our experiments, for each image, a vector containing 1 percentiles of the distribution of the local blockiness/blur features was calculated as the input to the NN. The CBP network was equipped with three hidden neurons and trained with the back-propagation⁴⁴ algorithm.

To evaluate the performance of the proposed approach, the LIVE image quality assessment database⁴⁵ was used. It consists of a set of 29 high-resolution and high-quality color source images that reflect adequate diversity in image content. These images were compressed using JPEG at a bit rate ranging from 0.15 bits per pixel (bpp) to 3.34 bpp, resulting in a database of 233 JPEG compressed stimuli (including the originals). The same source images were also compressed using JPEG2000 at a bit rate ranging from 0.028 to 3.15 bpp, yielding a database of 227 JPEG2000 compressed stimuli (including the originals). An extensive psychovisual experiment (performed at the University of Texas and described in more detail in Ref. 45) was conducted to assign a difference mean opinion score (DMOS) to each stimulus. The DMOS was measured on a continuous linear scale that was divided into five intervals marked with the adjectives “Bad,” “Poor,” “Fair,” “Good,” and “Excellent.”

So far, empirically measuring the error on the test data (e.g., by cross-validation^{46–48}) was proved to be the most reliable method to achieve an accurate approximation of the performance of an NN system. In our performance evaluation, a K -fold cross-validation method⁴⁸ was adopted (see also our previous experimental setup reported in Ref. 49). It randomized the statistical design problem by repeatedly splitting the available data in a training set and a test set. Figure 8 illustrates the experimental setup, in which the source images were divided into six groups. The entire procedure included six different trials, and for each trial (hereafter referred to as “run”) five groups of source images were used for training and the remaining one group of source images was used for testing. It is noteworthy that none of the stimuli used for testing ever entered any step of the training process. This way of working served to assess the generalization of the system performance empirically. It resulted in (an average of) 194 stimuli for training and (an average of) 39 stimuli for testing in the JPEG database for evaluating our JPEG metric. Additionally, we had (an average of) 189 stimuli for training and (an average of) 38 stimuli for testing in the JPEG2000 database for evaluating our JPEG2000 metric.

Compared to practical applications, in which the NN-based model can only be trained on a relatively small set of images as compared to the set of images it should be able to generalize, our approach to evaluate the generalization

Table 1 Performance of the proposed NR JPEG metric per run and averaged over all runs in terms of Pearson correlation coefficient and RSME.

NR JPEG metric	Pearson correlation coefficient	RMSE (normalized score scale [0, 1])
Run 1	0.963	0.119
Run 2	0.956	0.127
Run 3	0.962	0.115
Run 4	0.953	0.116
Run 5	0.975	0.084
Run 6	0.965	0.096
MEAN	0.962	0.109

Table 2 Performance of the proposed NR JPEG2000 metric per run and averaged over all runs in terms of Pearson correlation coefficient and RSME.

NR JPEG2000 metric	Pearson correlation coefficient	RMSE (normalized score scale [0, 1])
Run 1	0.942	0.115
Run 2	0.934	0.138
Run 3	0.926	0.144
Run 4	0.924	0.155
Run 5	0.923	0.145
Run 6	0.925	0.139
MEAN	0.930	0.139

performance has limitations. To mimic realistic implementations, we should have chosen to train the NN on about 1/5 of all stimuli and then test its generalization performance on all remaining stimuli. But, since an NN-based model requires sufficient data for training, we could not implement the more realistic evaluation of generalization performance with the small test set (i.e., limited number of stimuli) we had available. Of course, to better evaluate the generalization performance of the model, a larger-scale subjective database is preferred,⁵⁰ however, the content-oriented K -fold cross-validation method is known to provide the most reliable alternative when the available data set is relatively small. In compliance with cross-validation theory, one might estimate the generalization error by first directly checking whether the model attains satisfactory performance in all runs, and second, by indirectly checking whether a small variance among different runs is obtained.

4.2 Overall Metric Performance

As prescribed by the video quality experts group (VQEG, Refs. 51 and 52), the performance of our approach was evaluated with respect to its ability to predict subjective quality ratings (the DMOS). Two statistical tools usually employed in literature were adopted to characterize the prediction ability: i.e., the Pearson linear correlation coefficient, and the root mean square error (RMSE). The corresponding correlation coefficients and RMSE are listed in Tables 1 and 2 for the JPEG and JPEG2000 compressed images, respectively. In both cases, our proposed metric consistently resulted in a high prediction performance over all (six) runs. The NR JPEG metric yielded an averaged Pearson correlation coefficient of 0.9623 (with a highest value of 0.975 and a lowest value of 0.953), and an averaged RMSE of 0.109 on a normalized scale [0, 1] (with a highest value of 0.127 and a lowest value of 0.084). The NR JPEG2000 metric provided an averaged Pearson correlation coefficient of 0.930 (with a highest value of 0.942 and a lowest value of 0.923), and an averaged RMSE of 0.139 on a normalized scale [0, 1] (with a highest value of 0.155 and a lowest value of 0.115).

4.3 Comparison to Alternative Metrics

In the image quality community, researchers are accustomed to compare their metrics to alternatives available in the literature. It is, however, important to note that the performance of these metrics needs to be evaluated in a comparative setting, so that their strengths and weaknesses are fairly analyzed. In this respect, apart from only listing the numerical results (e.g., Pearson correlation coefficient) of the metrics in comparison, we also address some important issues behind these values.

An objective metric is conventionally validated through quantifying the correlation between its predicted values and the subjective quality scores. This correlation, however, can be calculated under three different testing conditions (TCs) as reported in literature. The first one is to directly calculate the linear correlation between the metric's predictions and the subjective data (see, e.g., Refs. 15 and 17, and 53, and hereafter referred to as TC1). This method is often used in metric comparison to better visualize differences in performance. The second method is suggested by the VQEG,⁵¹ and applies a nonlinear mapping function (e.g., a logistic function) to fit the metric's results to the DMOS before computing the correlation (referred to as TC2). A sophisticated nonlinear fitting (often including various parameters) accounts for a possible saturation effect, and usually yields higher correlation coefficients in absolute terms. However, in terms of performance comparison, it generally keeps the relative differences between the metrics as computed under TC1.¹⁵ A more demanding testing condition is cross-validation, in which the dataset is partitioned into complementary subsets: one for model calibration and the other for validation (see, e.g., Refs. 31 and 32, and 38, and hereafter referred to as TC3). Therefore, it should be noted that simply comparing the reported correlation coefficients of different metrics is not meaningful, unless they are evaluated with the same database under the same testing condition. Even under the same testing condition, the quantitative comparison between metrics may be biased due to a different selection of the disjoint sets for training and testing (e.g., under TC3).

Here, we compare the proposed JPEG and JPEG2000 metrics to state-of-the-art NR metrics in terms of performance.

Table 3 Performance of state-of-the-art NR JPEG metrics.

	JPEG metric	Pearson correlation coefficient	Testing environment
NR	Liu et al. (Ref. 17)	0.918	LIVE JPEG – TC1
	Wang et al. (Ref. 23)	0.931	LIVE JPEG – TC2
	Gastaldo et al. (Ref. 31)	0.943	LIVE JPEG – TC3
	Babu et al. (Ref. 32)	0.932	LIVE JPEG – TC3
	Proposed	0.962	LIVE JPEG – TC3
RR	Li et al. (Ref. 9)	0.889	LIVE JPEG – TC2
	Carnec et al. (Ref. 10)	0.972	LIVE JPEG – TC2
FR	PSNR (Ref. 3)	0.901	LIVE JPEG – TC2
	SSIM (Ref. 7)	0.979	LIVE JPEG – TC2
	VIF (Ref. 8)	0.980	LIVE JPEG – TC2

Issues related to computational complexity will be discussed in Section 5.1. For practical reasons the NR metrics used for comparison are limited to four JPEG metrics and three JPEG2000 metrics. To further compare the performance of our NR metric with respect to RR and FR metrics, we also include two RR metrics and three FR metrics well-known in literature. However, it should be noted that a direct comparison of NR metrics to RR/FR metrics is not completely fair, since the design of a reliable NR metric is more challenging, and often an NR metric is the only available option in real-time applications. Tables 3 and 4 list the Pearson correlation coefficient and the corresponding testing environment for these metrics. It can be seen that the prediction performance

Table 4 Performance of state-of-the-art NR JPEG2000 metrics.

	NR JPEG2000 metric	Pearson correlation coefficient	Testing environment
NR	Marziliano et al. (Ref. 24)	0.850	LIVE JP2K – TC3
	Sheikh et al. (Ref. 26)	0.910	LIVE JP2K – TC3
	Sazzad et al. (Ref. 25)	0.930	LIVE JP2K – TC3
	Proposed	0.930	LIVE JP2K – TC3
RR	Li et al. (Ref. 9)	0.957	LIVE JP2K – TC2
	Carnec et al. (Ref. 10)	0.957	LIVE JP2K – TC2
FR	PSNR (Ref. 3)	0.904	LIVE JP2K – TC2
	SSIM (Ref. 7)	0.971	LIVE JP2K – TC2
	VIF (Ref. 8)	0.979	LIVE JP2K – TC2

of our proposed JPEG and JPEG2000 metrics is slightly better than currently leading NR metrics. For the JPEG metric, our proposed approach even outperforms some of the existing RR and FR metrics, and comes close in performance to the best in class of these metrics. For the JPEG2000 metric, our approach is slightly underperforming with respect to RR and FR metrics currently known in literature.

5 Evaluation of Specific Metric Components

5.1 Reduction in Computational Complexity

From a practical point of view, it is highly desirable to develop an NR metric that is easy to implement, computationally efficient, and uses only a few parameters. In this section, we specifically elaborate on the reduction in computational complexity of the proposed approach compared to existing alternatives in the literature. Since the actual quantification of computational costs may largely depend on the specific implementation of a metric,^{15,17} we prefer to discuss the issue in qualitative rather than quantitative terms in order to avoid biased conclusions resulting from actual implementation choices.

Many NR metrics for JPEG compression are reported in literature, and the most successful ones are listed in Table 3. Our proposed metric, however, outperforms these alternatives in terms of simplicity, still obtaining a high reliability, as shown in Table 3. The metric of Ref. 17 explicitly models the HVS via texture and luminance masking. To keep the computational load of the metric within reasonable limits, both masking processes were heavily simplified, and that limited the metric's performance. Compared to the metrics of Refs. 23 and 31, and 32 which involved an extensive feature computation or selection stage (for example, in Ref. 23 both blockiness and signal activities were calculated, in Ref. 31 a large number of general pixel-based features were extracted, and in Ref. 32 a variety of HVS related features were computed), our metric clearly shows its advantage by only simply calculating the local blockiness in a computationally efficient way.

Progress in NR metrics for JPEG2000 compression is limited, mainly because the various artifacts are inherently content-dependent, and so, difficult to be detected and modeled. Researchers have taken different approaches to this problem. The metric of Ref. 24 is a well-known metric that attempts to predict the overall quality of JPEG2000 compressed images by modeling their most relevant artifact, which is blur. It is a very simple metric, however, its reported performance is limited, since the averaged blur is mapped to the quality scores with only a simple nonlinear transformation. Our metric clearly outperforms the metric of Ref. 24 (see Table 3), yet without introducing additional computational cost. The metric of Ref. 26 adopts natural scene statistics for measuring the image quality, which, however, often requires sophisticated modeling to achieve a reliable metric. A recently proposed NR metric for JPEG2000 compression²⁵ reports a high correlation with the LIVE database,⁴⁵ but contains an intensive feature extraction stage (i.e., eight different spatial features) and a complex parameter optimization procedure (i.e., nine model parameters) to combine these features. Thus, our metric outperforms the alternatives of Refs. 25 and 26 in implementation complexity and computational efficiency.

Table 5 Performance when using only the averaged blockiness feature value for the three testing conditions described in Sec. 4.3.

Testing environment		Pearson correlation coefficient	RMSE
LIVE JPEG - TC1		0.661	0.349
LIVE JPEG - TC2		0.906	0.247
LIVE JPEG - TC3	RUN1	0.875	0.250
	RUN2	0.929	0.188
	RUN3	0.912	0.245
	RUN4	0.926	0.237
	RUN5	0.872	0.238
	RUN6	0.907	0.211
MEAN		0.904	0.228

Within the framework of a NR metric, the performance of our NN-based metrics is very high. We could have considered improving the performance of the metrics even further by including combinations of artifacts as input features to the NN. For example, it could have been meaningful to include color artifacts, co-existing in JPEG/JPEG2000 compression. However, the features we selected as input to the NN (being blocking for JPEG and blur for JPEG2000) seem to be sufficiently representative for overall image quality. Adding additional artifact features would have made the metric more complicated without adding significant improvement.

5.2 Added Value of Using a Neural Network

The promising performance of the NR metrics, proposed in this paper, is primarily achieved by the combination of two essential components: 1. a simplified feature extraction that largely reduces the computational complexity and avoids multiple-feature modeling, and 2. a powerful NN to map the extracted feature to a quality rating. To validate the added value of including an NN, additional experiments were conducted, in which the neural network was omitted. Instead, the averaged feature value (i.e., the mean of the calculated local blockiness for the JPEG metric, and the mean of the calculated local blur for the JPEG2000 metric) was used as the metric's output. To make a fair comparison to our original NN-based metric, we evaluated the resulting metrics under three different testing conditions, as mentioned in Sec. 4.3. Tables 5 and 6 list the correlation coefficient and the RMSE for each of the testing conditions. It should be mentioned that the logistic function suggested by VQEG (Ref. 51) is conventionally used in both TC2 and TC3. The nonlinear regression function maps the metric's predictions to a set of predicted MOS values (i.e., $DMOS_p$), which are then compared to the actual DMOS values. The three-parameter logistic function is expressed as:

$$DMOS_p = \frac{b1}{1 + \exp[-b2 \times (Metric - b3)]}. \quad (8)$$

Table 6 Performance when using only the averaged blur feature value for the three testing conditions described in Sec. 4.3.

Testing environment		Pearson correlation coefficient	RMSE
LIVE JPEG2000 - TC1		0.715	0.257
LIVE JPEG2000 - TC2		0.741	0.221
LIVE JPEG2000 - TC3	RUN1	0.811	0.198
	RUN2	0.779	0.230
	RUN3	0.769	0.214
	RUN4	0.818	0.213
	RUN5	0.669	0.259
	RUN6	0.675	0.255
MEAN		0.754	0.228

The experimental results show that the simple, single feature without the use of an NN hardly achieves a reliable metric, not even after a complex fitting of the metric's output values to the corresponding subjective ratings. Of course, the performance of these simple metrics can be improved by adding properties of the HVS and by explicitly modeling the inherent artifacts (see, for example, Refs. 13 and 16, and 17). This generally yields metrics that can be used to assess the overall image quality, but at the expense of complex HVS modeling and an extensive parameter optimization process. For real-world applications, our proposed approach including an NN tends to be an efficient and inexpensive solution.

5.3 Performance on the TID Database

To further evaluate the robustness of the proposed approach against different databases, we repeat the experiment in Sec. 4.2 on another publicly available image database [i.e., TID (Ref. 54)] that is widely recognized in the image quality community. It includes distorted images generated from 25 reference images with 17 distortion types at four distortion levels. The subsets of 100 JPEG compressed images and of 100 JPEG2000 compressed images are used in our experiment. Details on the subjective test and data processing can be found in Ref. 55. To conduct a K -fold cross-validation, the reference images were divided into 5 groups, which resulted in 80 stimuli for training and 20 stimuli for testing for both metrics. The correlation coefficients and RMSE are listed in Tables 7 and 8 for JPEG and JPEG2000, respectively.

It shows that our proposed approach consistently results in a high prediction performance for all runs of the JPEG/JPEG2000 subsets of the TID database. Moreover, the overall performance we obtained for the TID database is largely comparable to the performance for the LIVE database. Of course, we can also make an even more critical evaluation of our metric by training it with one database, and testing it on the other database. We obtained an average correlation coefficient of **0.93** (i.e., training with LIVE and testing on TID, and the other way around) for the NR

Table 7 Performance of the proposed NR JPEG metric per run and averaged over all runs in terms of Pearson correlation coefficient and RSME, on the TID database (Ref. 54).

NR JPEG metric	Pearson correlation coefficient	RMSE (normalized score scale [0, 1])
Run 1	0.976	0.070
Run 2	0.976	0.070
Run 3	0.966	0.082
Run 4	0.968	0.077
Run 5	0.953	0.092
MEAN	0.968	0.078

JPEG metric, and an average correlation of **0.91** for the NR JPEG2000 metric. It should, however, be noted that this kind of cross database evaluation is not straightforward, and the soundness of the results may largely be affected by differences in the use of the quality scales between subjective experiments.⁵⁰ To compensate for this issue an additional scale realignment experiment is needed, but is considered outside the scope of this paper.

5.4 Limitations and Future Research

The NR metrics, proposed in this paper, aim to assess the perceived overall quality of JPEG and JPEG2000 compressed images. To achieve a simple yet efficient metric, especially for real-time processing, we kind of neglect the occurrence of and interaction between various artifacts that may occur simultaneously in an image, thus affecting the perceived overall quality. In our case, only the most relevant artifact is extracted to predict the overall image quality, and we fully rely on the NN to approximate the unknown relationship between this single feature and the quality rating. As a consequence, the proposed NR metrics intrinsically exhibit two major drawbacks: 1. the local distortion values (simply calculated) are not necessarily in agreement to what the human eye perceives, and thus cannot be used to precisely reflect

Table 8 Performance of the proposed NR JPEG2000 metric per run and averaged over all runs in terms of Pearson correlation coefficient and RSME, on the TID database (Ref. 54).

NR JPEG2000 metric	Pearson correlation coefficient	RMSE (normalized score scale [0, 1])
Run 1	0.962	0.096
Run 2	0.922	0.114
Run 3	0.950	0.095
Run 4	0.938	0.107
Run 5	0.940	0.117
MEAN	0.942	0.106

the local annoyance of a perceived artifact, and 2. the perceived annoyance of other artifacts, e.g., ringing in JPEG and JPEG2000 compression, cannot be assessed by the overall metric. Being able to quantify the annoyance of more types of artifacts is of fundamental importance to noise reduction in image/video enhancement. It requires research in the design of dedicated metrics (see, e.g., Refs. 12–17) to detect and estimate the local annoyance of a specific artifact type. However, in current visual communication systems, predicting the perceived overall quality in real-time without compromising the system's complexity is very valuable, but still challenging.

We feel that designing an NR metric based on an NN is promising in terms of predicting the overall image quality. We are continuing our efforts into designing NR metrics for more types of distorted images, such as blur, noise, and wireless channel errors. We also plan the design of an NR metric to assess the quality of images after post-processing, i.e., at the very end of an imaging chain. The latter application is of great value for feedback control, i.e., to optimize the output of an imaging chain. The most demanding challenge is to extend the approach to the more complex problem of video quality assessment, which also is very relevant, given the ubiquity of streaming videos for digital television, digital cinema, and wireless broadcasting. At the lowest level of complexity, video quality assessment may be approached on a frame-by-frame basis. The approach discussed in this paper can be used to do so, albeit that a redesign is needed to specifically address the artifacts in video specific codecs, such as MPEG and H.264. At the next level of complexity temporal integration of artifacts over frames needs to be added. To achieve this, not only the NN-based approach needs a significant extension, but also a large database on video quality needs to become available.

6 Conclusions

In this paper, we provide an efficient NR approach for the perceived overall quality assessment of JPEG/JPEG2000 compressed images. Its reliable prediction ability at a largely reduced computational cost is achieved by skillfully combining a simplified feature extraction strategy with an adaptive neural network. The first component efficiently selects and calculates the most relevant feature representative for the overall image quality, and thus avoids explicitly modeling the occurrence of and interaction between various artifacts inherent in a distorted image. The latter component, subsequently, is used to empirically learn the highly nonlinear relationship between the relevant feature and the overall image quality rating. The resulting NR JPEG and JPEG2000 metrics are validated with subjective data under a critical cross-validation condition, and are fairly compared to several alternative metrics existing in literature.

References

1. Z. Wang and A. C. Bovik, *Modern Image Quality Assessment. Synthesis Lectures on Image, Video, & Multimedia Processing*, Morgan & Claypool, San Rafael, California (2006).
2. P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Winchester, Massachusetts (2000).
3. Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? – A new look at signal fidelity measures," *IEEE Signal Process. Mag.* 26(1), 98–117 (2009).

4. S. Daly, *The Visible Differences Predictor: An algorithm for the Assessment of Image Fidelity, Digital Images and Human Vision*, A. B. Watson, Ed., MIT Press, Cambridge, Massachusetts (1993).
5. J. Lubin, *The Use of Psychophysical Data and Models in the Analysis of Display System Performance, Digital Images and Human Vision*, A. B. Watson, Ed., MIT Press, Cambridge, Massachusetts (1993).
6. A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A* **14**, 2378–2390 (1997).
7. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
8. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
9. Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Top. Signal Process.* **3**(2), 202–211 (2009).
10. M. Carnec, P. Le Calleta, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Process. Image Commun.* **23**(4), 239–256 (2008).
11. U. Engelke, M. Kusuma, H. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Process. Image Commun.* **24**(7), 525–547 (2009).
12. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings of IEEE International Conference on Image Processing*, Vol. 3, pp. 57–60 (2002).
13. R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.* **18**, 717–728 (2009).
14. X. Feng and J. P. Allebach, "Measurement of ringing artifacts in JPEG images," *Proc. SPIE* **6076**, 74–83 (2006).
15. H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.* **20**, 529–539 (2010).
16. H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.* **4**(11), 317–320 (1997).
17. H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," *EURASIP J. Appl. Signal Process.* **2009**, 263540 (2009).
18. L. M. J. Meesters and J. B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Process.* **82**(3), 369–387 (2002).
19. J. B. Martens, *Image Technology Design: A Perceptual Approach*, Kluwer Academic, Dordrecht (2003).
20. K. Zou and W. Ali, "Automated video chain optimization," *IEEE Trans. Consum. Electron.* **47**, 593–603 (2001).
21. C. K. Koh, S. K. Mitra, J. M. Foley, and I. Heynderickx, "Annoyance of individual artifacts in MPEG-2 compressed video and their relation to overall annoyance," in *Human Vision and Electronic Imaging X*, *Proc. SPIE* **5666**, 595–606 (2005).
22. I. O. Kirenko, R. Muijs, and L. Shao, "Coding artifact reduction using no-reference block grid visibility measure," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 469–472 (2006).
23. Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proceedings of the IEEE International Conference on Image Processing*, Vol. 1, pp. 477–480 (2002).
24. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process. Image Commun.* **19**, 163–172 (2004).
25. Z. Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Process. Image Commun.* **23**, 257–268 (2008).
26. H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.* **14**(11), 1918–1927 (2005).
27. I. R. L. Davis, D. Rose, and R. Smith, "Automated image quality assessment," *Human Vision, Visual Processing, and Digital Display IV*, *Proc. SPIE* **1913**, 27–36 (1993).
28. F. Lin and R. M. Mersereau, "A constant subjective quality MPEG encoder," in *IEEE ICASSP*, pp. 2177–2180 (1995).
29. P. Gastaldo, S. Rovetta, and R. Zunino, "Objective quality assessment of MPEG-2 video streams by using CBP neural networks," *IEEE Trans. Neural Netw.* **13**(4), 939–947 (2002).
30. P. Gastaldo, R. Zunino, I. Heynderickx, and E. Vicario, "Objective quality assessment of displayed images by using neural networks," *Signal Process. Image Commun.* **20**, 643–661 (2005).
31. P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *J. Electron. Imag.* **14**(3), 033004 (2005).
32. R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Process.* **87**(6), 1493–1503 (2007).
33. H. de Ridder and M. C. Willemsen, "Percentage scaling: a new method for evaluating multiply impaired images," in *Human Vision and Electronic Imaging V*, *Proc. SPIE* **3959**, 68–77 (2000).
34. M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Process.* **70**(3), 247–278 (1998).
35. M. C. Q. Farias, M. S. Moore, J. M. Foley, and S. K. Mitra, "Perceptual contributions of blocking, blurring, and fuzzy impairments to overall annoyance," in *Human Vision and Electronic Imaging IX*, *Proc. SPIE*, **5292**, 109–120 (2004).
36. J. Xia, Y. Shi, K. Teunissen, and I. Heynderickx, "Perceivable artifacts in compressed video and their relation to video quality," *Signal Process. Image Commun.* **24**, 548–556 (2009).
37. H. Liu, N. Klomp, and I. Heynderickx, "A perceptually relevant approach to ringing region detection," *IEEE Trans. Image Process.* **19**, 1414–1426 (2010).
38. J. Redi, P. Gastaldo, R. Zunino, and I. Heynderickx, "Reduced reference assessment of perceived quality by exploiting color information," in *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (2009).
39. D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge, Massachusetts (1986).
40. S. Ridella, S. Rovetta, and R. Zunino, "Circular back-propagation networks for classification," *IEEE Trans. Neural Netw.* **8**, 84–97 (1997).
41. E. B. Bau and H. David, "What size net gives valid generalization?" *Neural Comput.* **1**, 151–160 (1989).
42. B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, madaline and back propagation," *Proc. IEEE* **78**(9), 1415–1442 (1990).
43. J. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, "Color distribution information for the reduced-reference assessment of perceived image quality," *IEEE Trans. Circuits Syst. Video Technol.* **20**, 1757–1769 (2010).
44. T. P. Vogel, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back propagation method," *Biol. Cybern.* **59**, 257–263 (1988).
45. H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," <http://live.ece.utexas.edu/research/quality>.
46. Y. Liu, "Unbiased estimate of generalization error and model selection in neural network," *Neural Networks* **8**, 215–219 (1995).
47. L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural Networks* **4**, 761–767 (1998).
48. R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, Reading, Massachusetts (1989).
49. H. Liu, J. Redi, H. Alers, R. Zunino, and I. Heynderickx, "No-reference image quality assessment based on localized gradient statistics: application to JPEG and JPEG2000," *IS&T/SPIE Electronic Imaging 2010, Human Vision and Electronic Imaging XV* (2010).
50. J. Redi, H. Liu, H. Alers, R. Zunino, and I. Heynderickx, "Comparing subjective image quality measurement methods for the creation of public databases," *IS&T/SPIE Electronic Imaging 2010, Image Quality and System Performance VII* (2010).
51. VQEG (2003, Aug.): *Final report from the video quality experts group on the validation of objective models of video quality assessment*. Available: <http://www.vqeg.org>.
52. VQEG, Report on the Validation of Video Quality Models for High Definition Video Content, Video Quality Experts Group (VQEG), 2010. www.vqeg.org.
53. S. Winkler, "Vision models and quality metrics for image processing applications," Ph.D. dissertation, Dept. Elect., EPFL, Lausanne (2002).
54. N. Ponomarenko and K. Egiazarian, "Tampere image database 2008 TID2008," [Online]. Available: <http://www.ponomarenko.info/tid2008.htm>.
55. N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "Metrics performance comparison for color image database," in *Proceedings of the 4th International Workshop Video Process. Quality Metrics Consum. Electron.*, Scottsdale, Arizona (2009).

Hantao Liu received an MSc degree in signal processing and communications from The University of Edinburgh, United Kingdom, in 2005. In 2011, he received a PhD degree from the Delft University of Technology, The Netherlands. His PhD research was supported by the Philips Research Laboratories in Eindhoven, The Netherlands, focusing on modeling perceived quality for imaging applications. Currently, he is a postdoctoral researcher at the Department of Mediamatics, Delft University of Technology, The Netherlands, working on medical image perception. His research interests include image analysis, human visual perception, and signal processing.

Judith Redi obtained an MSc degree in electronic engineering in 2006 from Genoa University, Italy, and her PhD in 2010 from the same university. Her PhD thesis "Visual overall quality assessment: a perception-driven framework based on computational intelligence techniques" received the "best engineering thesis award" from the University of Genoa in the same year. After being appointed as a post-doc at Eurecom, where she focused on digital image forensics and computer vision, she moved to the Man-Machine Interaction group of Delft University of Technology, where she is currently working as an assistant professor. Her main research interests concern the understanding of visual perception and of the quality of the visual experience. Her research activity in this area ranges from empirical studies and psychophysics to the use of computational intelligence for modeling visual perception mechanisms.

Hani Alers received an MSc degree in media and knowledge engineering from the Delft University of Technology in September 2007, where he is currently pursuing a PhD. His research interests include modeling human image and video quality perception, as well as employing Internet technology in educational systems.

for data representation and learning, massive-scale text-mining

Rodolfo Zunino obtained the "Laurea" degree cum laude in electronic engineering from Genoa University in 1985. Since 1995 he is with DIBE, where he now serves as associate professor teaching "Embedded Systems" and "Cyber Security and Electronic Instrumentation." His main scientific interests include intelligent systems for computer security, network security, and critical infrastructure protection, embedded electronic systems for neural networks, efficient models

and text-clustering methods, and advanced techniques for multimedia data processing. He co-authored more than 200 scientific papers in International Journals and Conferences; he co-chaired two editions of the International Workshop on Computational Intelligence for Security in Information Systems (CISIS'08 and CISIS'09). From 2001 to 2009 he contributed as associate editor of the IEEE Transactions on Neural Networks, and has participated in the Scientific Committees of several International Events (IJCNN, ICANN, etc.). He is a senior member of IEEE (CIS- Computational Intelligence Society).

Ingrid Heynderickx received her PhD degree in physics at the University of Antwerp (Belgium) in December 1986. In 1987 she joined the Philips Research Laboratories in Eindhoven (The Netherlands), and meanwhile worked in different areas of research: optical design of displays, processing of liquid crystalline polymers, and functionality of personal care devices. Since 1999 she is head of the research activities on visual perception of display and lighting systems and in 2005 she is appointed research fellow in the group visual experiences. She is member of the Society for Information Displays (SID), and for the SID, she was chairman of the applied vision subcommittee from 2002 until 2007. In 2008, she became a fellow of the SID and chairman of the European Program Committee of the SID. In 2005, she was appointed guest research professor at the Southeast University of Nanjing (China) and part-time full professor at the University of Technology in Delft (the Netherlands).