JAISCR, 2011, Vol.1, No.1, pp. 27–41

A TWO-LAYER NEURAL SYSTEM FOR REDUCED-REFERENCE VISUAL QUALITY ASSESSMENT

Judith Redi, Paolo Gastaldo, Rodolfo Zunino
Department Biophysical and Electronic Engineering
University of Genoa
Via Opera Pia 11/, 16145 Genoa, Italy

Abstract

Real-time assessment of visual quality can be efficiently supported by reduced-refe-rence paradigms, which require a very limited amount of information on the original signal, easily embeddable in the signal itself. In this paper, a reduced-reference system for image quality assessment is proposed, based on a small sized numerical description of images encoding the luminance distribution and its variations due to visual distortions. The assessment paradigm is implemented exploiting machine learning tools and articulates in two phases: first, a Support Vector Machines-based classifier identifies the kind of distortion affecting the image; then, the actual quality level of the distorted image is computed by a specifically trained SVM regressor. The general validity of the approach is supported by experimental validations based on subjective quality data.

1 Introduction

As the fruition of video and multimedia contents becomes wider, exploiting new media technologies such as the internet, electronic imaging systems are more than ever required to guarantee the high quality of the displayed signal, regardless to the distortions originated during the transmission and/or the displaying processes. Such a scenario calls for the need of accurate, on-board signal post-processing systems, able to detect the artifacts brought about by distortions, to estimate the perceived quality of the received images and to apply enhancement algorithms, in order to appropriately correct and enhance the finally displayed signal. The study and modeling of visual quality perception covers then a crucial role in the development of cutting-edge video technology applications.

Up to now, the most reliable tool for estimating perceived quality is subjective testing, directly involving humans and their judgment [5, 31, 20, 1]. Real-time post-processing chains cannot rely on such expensive and time-consuming techniques, hence requiring automated (objective) quality assessment systems (OQAs) [6, 21, 36]. OQAs rely on the computation of objective metrics, which can either exploit low to high level Human Visual System salient features, or relevant statistical descriptions of the signal. Full-Reference (FR) paradigms [33, 11, 37, 3, 7, 8] perform quality assessments by computing features as a comparison of the original with the distorted image, hence requiring full access to both signals. Such techniques find several applications (i.e. coding algorithms performance tests) but cannot be exploited in real-time contexts. Reduced Reference (RR) methods provide instead a good trade-off between blind (No Reference, [39, 35, 19, 9, 18, 40, 16]) and FR quality assessment, only involving a limited amount of numerical features characterizing the original signal [30, 34, 38, 17, 15, 23, 24]. This overcomes a common drawback of No Reference algorithms, namely their limited applicability to a single type of distortion. Thus, the RR paradigm can provide a successful approach for supporting real-time modeling of perceived quality.

Several existing FR and RR algorithms [37, 3,

34, 38] can be defined as "general-purpose" (as opposed to distortion-specific), and offer the great advantage of using a single feature-based description for assessing the quality losses due to different kinds of distortions (e.g. noise, compression artifacts, reduced sharpness). Usually, this non-linear mapping is computed distortion-dependent, adapting the general-purpose metric to the specific effects of one distortion on quality. Such a strategy is proved to be successful; however, it requires some a priori knowledge on the kind of distortion affecting the signal, which is often considered as given. The research proposed in this paper addresses such central issue by introducing a two-layer OQA system that can automatically identify the kind of distortion affecting the signal, and then apply the most effective objective metric in the quality assessment phase.

The present work further investigates on assessment models based on connectionist paradigms [24]. A two-layer RR system based on Support Vector Machines (SVMs) [32] is designed to map a numerical description of the image into quality scores. To handle different distortions, first a classifier determines the type of distortion affecting the image. According to the output of the classifier, a regression machine is chosen among a bank of predictors trained to evaluate the effects on the quality of different distortions. This machine eventually performs the non linear mapping required to quantify the loss in quality of the image with respect to its original version.

The proposed system is fed with a feature-based representation of the distorted image and its original version. Luminance distribution information supports image representation, as artifacts brought about by digital processing affect the original luminance content of the image, each in a peculiar way. Hence, the rationale of the present approach is that by comparing the statistics of the original and distorted image one can identify both the kind and the extent of the distortion. Previous works [23, 24] showed that second order statistics can apply successfully toward that end; therefore, this research adopts a set of features derived from the co-occurrence matrix [25].

In this paper, bandwidth and computational constraints are also considered as parameters to evaluate the effectiveness of the approach. Hence, with respect to [24], a different strategy, involving changes in the metric computation and in the system setup, is used. Experimental validation is provided on the LIVE database [10], including three types of distortion: White noise, Gaussian Blur and JPEG compression. Empirical results confirm the validity of the connectionist paradigm and the effectiveness of luminance statistics for predicting the image quality. Furthermore, the changes applied to [24] further prove the flexibility of the system and the robustness of the overall approach.

2 System Overview

Reduced reference OQAs represent a promising solution for on-board, real time image quality assessment in consumer multimedia systems. The major advantage that such methods offer is the possibility of assessing quality by exploiting some information about the original image, provided that such information is sufficiently small-sized with respect to the video signal to be transmitted. As a major consequence, that information can be included in the signal as metadata without affecting the bandwidth occupation. In this regard, RR approaches improves over full-reference approaches, which are actually unsuitable for most applications, as they require full access to the original signal.

This study proposes a Reduced Reference OQA, based on a double-layer approach that allows handling the effects of different distortions. The first layer tackles the task of distortion identification. The second layer is made of a set of dedicated predictors, specifically trained to understand the quality losses in the image due to the presence of the detected distortion.

In this research, emphasis is indeed put on limiting the computational cost of the objective metric and the amount of information required to be computed from the original signal and set along the transmission channel. Three main factors allow to tackle this purpose effectively, namely : (1) the suitability of luminance distribution derived features to describe quality losses; (2) a feature selection procedure designed to discard non informative features, which in turn would inflate the size of the metadata vector to be transmitted; (3) the non-linear modeling power of Machine Learning tools (i.e. Support Vector Machines), which precisely

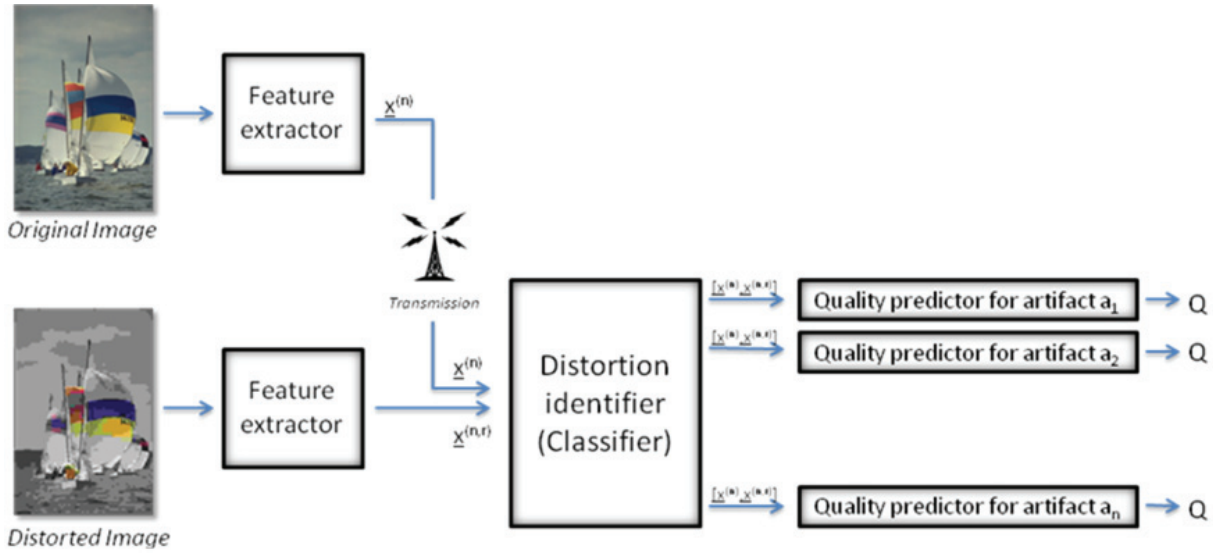

Figure 1. Overview of the Reduced Reference General Purpose Objective Qulaity Assessment System.

mimic human visual perception without requiring a detailed model of the HVS.

2.1 Outline of the Reduced-Reference Quality Prediction Framework

Let $I^{(n)}$ be the reference image, and $\bar{I}^{(n,r)}$ the image resulting from the insertion of some distortion to$I^{(n)}$, being r the distortion level. Let $x^{(n)}$ and $x^{(n,r)}$ be the numerical representations of $I^{(n)}$ and $\bar{I}^{(n,r)}$, respectively. Finally, let $q^{(n)}$ and $q^{(n,r)}$ be the quality levels for $I^{(n)}$ and $\bar{I}^{(n,r)}$, respectively, determined via subjective testing. The proposed system (see figure 1) compares the numerical descriptors, $\{x^{(n)},x^{(n,r)}\}$, and estimates the discrepancy,$d_S\left(q^{(n)},q^{(n,\,r)}\right)$, between the subjective scores associated with the images. At runtime, vector $x^{(n)}$ is worked out at the signal source, while $x^{(n,r)}$ is computed on the receiver side, and the two vectors are eventually processed to obtain quality estimates.

Having defined the set A=$\{a_1, a_2, \ldots, a_l\}$ of distortions of interest, the system relies on a dedicated quality predictor for each of them. As no a-priori knowledge can be assumed on the nature of the distortion affecting the input signal, a distortion detector is first needed to identify the distortion ($a_i \in$A) applied to the image$I^{(n)}$. This first module forwards the image descriptors to the appropriate quality predictor $\Omega(a_i)$, which finally provides an estimation for the difference in quality $d_S\left(q^{(n)},q^{(n,\,r)}\right)$ between the reference and incoming signal, distorted by a_i.

3 Objective Quality Metric

The color distribution across one image can be consistently altered due to the distortion impact. To describe such changes, second order histograms [25, 13] are powerful tools, representing the joint occurrence of a pair of colors throughout the image. Hence, the present research exploits features derived from the co-occurrence matrix to construct an objective description of the image. To represent the color, the luminance (Y) layer of the YC_bC_r colorspace is chosen. The relevance of luminance in quality assessment has been already extensively proved [11]. Furthermore, video streams are usually encoded in the YC_bC_r colorspace, hence, in a real time perspective, the luminance channel would be immediately available for computation.

The luminance distribution-based objective metric is computed in three steps: first locally, on a block-by-block basis, then gathering the obtained values into a single statistical global descriptor, to reduce the number of relevant values to be processed by the Machine Learning based assessment system. Finally, a feature selection procedure is applied, to limit information redundancy in the global descriptor.

3.1 Co-occurrence Matrices

The co-occurrence matrix measures local correlation among gray tones within an image subregion. Defining a region a, including $H_a \times W_a$

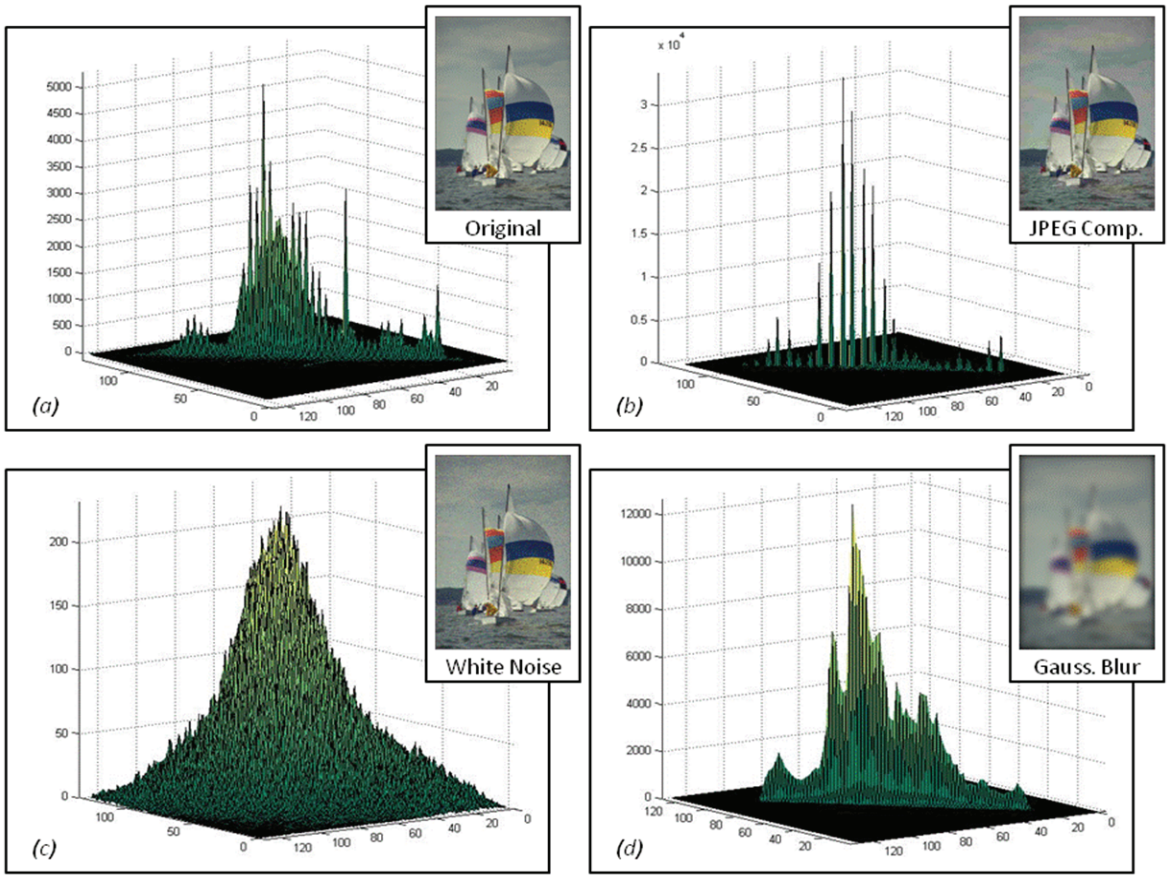

Figure 2. Effects of distortions on the color distribution. The co-occurrence matrix is shown for the original image (a), and the same image with Jpeg Compression (b), White Noise (c) and Gaussian Blur (d) applied, respectively.

pixels, each matrix element $C_{i,j}^{(a)}(\lambda,\theta)$ of the co-occurrence matrix $\mathbf{C}_a$ describes a joint occurrence of luminance levels, and counts the pairs of pixels in a that: (1) have gray levels i and j, and (2) are separated by λ radial units at an angle q to the horizontal axis. Formally, C_a is defined as:

$$C_{i,j}^{(a)}(\lambda,\theta)=\left|\left\{(m,n),\ \begin{matrix}0\le m\le W_a-1\\ 0\le n\le H_a-1\end{matrix},\ \text{s.t.}\ \begin{matrix}a[m,n]=i\\ a[m+\Delta h,n+\Delta v]=j\end{matrix}\right\}\right|$$

$$with\ i,j\ =0,\ldots,N_g \qquad (1)$$

where Dh and Dv are the horizontal and vertical displacements in the q direction, respectively:

$$\Delta h=\lambda\lceil\cos\theta\rceil,\Delta v=\lambda\lceil\sin\theta\rceil,0\circ<q\le 90\circ \quad (2)$$

$$\Delta h=\lambda\lfloor\sin\theta\rfloor,\Delta v=\lambda\lceil\cos\theta\rceil,90\circ<q\le 180\circ \quad (3)$$

$\mathbf{C}_a$ is eventually a $N_g\times N_g$ matrix, where N_g is the number of luminance quantization levels.

Figure 2 illustrates the Co-occurrence matrix extracted from the original image *Sailing_2* of LIVE database and from three differently distorted versions of it. Axis X and Y are indexed by color bins, while the Z axis indicates the joint occurrence of the given pair of colors at a predefined distance 1 pixel in the horizontal direction (i.e., $l=1$ and $q=0$). It is easily noticeable how different distortions cause peculiar changes to the second-order histogram when compared with the original signal (figure 2(a)). For example, JPEG compression (figure 2(b)) causes a severe quantization of the histogram, while white noise (figure 2(c)) spreads the informative content across the whole matrix. Such a marked behavior suggests the possibility that a metric based on the analysis of the co-occurrence matrix variations could efficiently reveal both the distortion affecting the image and its amount.

3.2 Local Metric Computation

Luminance distribution information is collected locally, on adjacent non-overlapping subregions of the image sized N_axN_a pixels. For each region, a Co-occurrence matrix is computed, and, from it, local features values.Indeed, for optimization purposes, co-occurrence matrices can also be characterized by using a set of scalar descriptors (features), statistically-based and image-description oriented [33]. All features are implicitly indexed by the image region, a, from which the matrix is calculated. To minimize computational cost the present research adopts a subset, $\Phi = \{f_u; u = 1,..,N_f\}$, of N_f=10 features that have already been tested to be successful [23, 22].

The block size, N_a, plays a crucial role in the reliability of the image description, and the image size should be taken into account to ensure proper sampling by a sufficient number of measures. Moreover, the features derived from the co-occurrence matrix $C^{(a)}(\lambda,\theta)$ may suffer from border effects. The percentage of pixels that do not enter the computation of $C^{(a)}_{i,j}(\lambda,\theta)$ decreases as N_a increases, hence one should avoid to use small block sizes; a typical setting is N_a>8.

3.3 Distortion Distribution Representation at a Global Level

The second step compresses the information obtained, and aggregates block-level data into one objective vector that characterizes the whole image. This procedure is performed consistently with the actual measuring procedure, since human assessors usually generate one overall quality score per image. Also, for computational and limited bandwidth reasons, block-based information must be reduced into a single, small-sized vector per image. To accomplish this, a percentile-based description of the distribution of each co-occurrence matrix feature is taken. As each feature reports on the effects of distortions on color distribution, the percentile based global vector can be considered an expression of the distortion action across the image.

For a parameter setting (l*,q*), the image $\bar{I}^{(n,r)}$ is represented by a set of N_f objective vectors, $x^{(n,r)}_{u,(\lambda*,\theta*)}$, u=1,..,N_f, which contain detail-related information. In the remainder of this paper, the index pairs (n,r) and (l*,q*) will be omitted wherever possible, in order to simplify the notation. The procedure to construct the objective vectors can be summarized as follows:

Inputs: a picture $\bar{I}^{(n,r)}$, a descriptive feature f_u and a the set of values $\mathrm{X}_u = \{x_{u,m};\ m = 1,..,N_b\}$, computed for each block m of the N_b blocks in the image.

- Compute a percentile-based description of the sample set X_u; let p_a be the a-th percentile:

$$\varphi_{\alpha,u} = p_\alpha(\mathrm{X}_u)$$

- Assemble the objective descriptor vector, x_u, for the feature f_u on the image $\bar{I}^{(n,r)}$ as

$$x_u = \{\varphi_{\alpha,u}; \alpha = 0,20,40,60,80,100\} \quad (4)$$

3.4 Feature Selection

The paper considers the statistical approaches to feature selection proposed in [22], which exploits Kolmogorov-Smirnov's test.

The proposed method tackles feature selection empirically; thus, the data set is obtained by applying the image-processing algorithm, $\zeta_q(\cdot)$, at different settings to a library of training images, Ω = $\{I^{(s)}$, s=1,...,$n_p\}$ and collecting the sample of processed images, $\bar{\Omega} = \{\bar{I}^{(s,q)}; s = 1,..,n_p; q = q_1,..,q_n\}$. Applying the feature-extraction process (as per Sect. 3.2) to each element in $\bar{\Omega}$, gives the eventual sample set V, which holds n_s=$q_n n_p$ elements and is given by:

$$V = \left\{x^{(s,q)}; s = 1,..,n_p; q = q_1,..,q_n\right\}. \quad (5)$$

The analysis selects from the complete set of candidate features, Φ, only the 'active' ones, i.e., those whose statistical properties depart significantly from their original values after applying a processing algorithm, $\zeta_q(\cdot)$. Thus, for each objective feature $f_k \in \Phi$, the analysis compares statistically two samples: one contains the values of f_k for a set of original images, the other holds the values of f_k for a set of processed images. To guarantee the independence of the two samples, the two sets of pictures are disjoint. The feature values are worked out on non-overlapping blocks of pixels randomly extracted from each image.

Table 1. Co-occurrence Matrix Descriptive Features

Feature name	Definition	Feature name	Definition
Absolute value	$f_1 = \sum_z z P_z(\lambda,\theta)$	Inverse Difference	$f_2 = \sum_{i,j} \frac{C_a(i,j,\lambda,\theta)}{1+(i-j)^2}$
Correlation	$f_3 = \left[\sum_{i,j} ij C_a(i,j,\lambda,\theta) - \mu_i^2\right] / \sigma_i^2$	Autocorrelation	$f_4 = \sum_{i,j} ij C_a(i,j,\lambda,\theta)$
Energy	$f_5 = \sum_{i,i} [C_a(i,j,\lambda,\theta)]^2$	Diagonal Energy	$f_6 = \sum_{i,j}^{i=j} [C_a(i,j,\lambda,\theta)]^2$
Entropy	$f_7 = -\sum_{i,j} C_a(i,j,\lambda,\theta) \log_2 C_a(i,j,\lambda,\theta)$	Differential Variance	$f_8 = \left[\sum_z (z-f_1)^2 P_z(\lambda,\theta)\right]^{1/2}$
Contrast	$f_9 = \sum_z z^2 P_z(\lambda,\theta)$	Differential Entropy	$f_{10} = -\sum_z P_z(\lambda,\theta) \log_2 P_z(\lambda,\theta)$
IMC	$f_{11} = \left(\sum_{i,j} C_a(i,j,\lambda,\theta) \log C_a(i,j,\lambda,\theta) - \sum_{i,j} C_a(i,j,\lambda,\theta) \log\left[\sum_j C_i^{(a)}\right]^2\right) \Big/ C_i^{(a)}$		

With $C_i^{(a)} = \sum_j C_a(i,j,\lambda,\theta)$, μ_I *and* σ_I *mean and standard deviation of* $C_i^{(a)}$, *respectively, and* $P_z(\lambda,\theta) = \sum_{i,j,|i-j|=z} C_a(i,j,\lambda,\theta)$

The mutual independence of the data sets allows one to use the Kolmogorov-Smirnov test [12] to disprove the null hypothesis, that is, to determine whether the two data sets for f_k have not been drawn from the same distribution. In that case, f_k is selected as an 'active' feature. KS has been preferred over parametric tests because one usually cannot assume a known distribution of the data involved.

The full pseudo-code of the feature selection algorithm is outlined in [22].

4 Connectionist Paradigms for Objective Quality Assessment

The system described in section 2 consists of two steps. Firstly, the distortion affecting the image has to be identified; secondly, the numerical representation of the image has to be mapped into a quality score by a dedicated predictor, which is specifically trained to assess image quality for a given distortion. The first layer is required to solve a classification problem. When aiming to detect the distortion a_i affecting the sample $I^{(n)}$, the set A=$\{a_1, a_2, \ldots, a_l\}$, the system is required to relate the input vector $\mathbf{x}_u$to a discrete value, representing a_i. On the other hand, the second layer maps the numerical descriptor $\mathbf{x}_u$ into a quality score, which cannot be expressed by discrete values to achieve acceptable accuracy. Therefore, this module can be designed to solve a regression problem.

In both cases, the use of connectionist paradigms is appealing. The machine learning world provides excellent tools able to handle both classification and regression supervised problems. Moreover, from a modeling point of view, such methods appear particularly suitable to model a highly non-linear context such as perception. Among others, Support Vector Machines (SVM) proved to be effective both for classification and regression tasks.

In a most general setting, one has a data sample, X, holding n patterns: each pattern includes a data vector, $\mathbf{x}\in R^m$, and its associate 'target' label, y. Classification problems involve a binary setting $y\in\{-1,+1\}$, whereas a regression problem is tackled when target values are expressed by continuous values, e.g., $y\in[-1,1]$. The learning phase uses

both **x** and y to build up a decision rule; at run-time, the trained machine processes unseen data and associates every input with a prediction of its target,$\hat{y}$.

The regression strategy implements the decision function, $\hat{y}= f(\mathbf{x})$ as a weighted series, whose basic terms, $f(\mathbf{x})$, typically embed nonlinear functions:

$$\hat{y} = f(x) = \sum_i \beta_i \phi_i(x) + \beta_0 \qquad (6)$$

Classification machines just yield a binary output by applying the operator $sign(.)$ to $f(\mathbf{x})$.

In the practical design of any estimator, the training set TG = $\{(\mathbf{x}_i,\mathrm{y}_i);\ i$=1,..,$n_p\}$ gives a sample-based formulation of the desired input-output mapping. For any empirical paradigm, the training procedure implements that mapping by fitting the degrees of freedom of the supported nonlinear estimator as per (6).

SVMs, in particular, tackle the pattern recognition problem within the Statistical Learning Theory [32] framework. A crucial element of it is the so-called kernel trick [29]: the kernel function $K(\cdot,\cdot)$ allows inner products of patterns in a higher dimensional, transformed space, though not involving the specific mapping of each single pattern. Given the points $\phi(x_1)$ and $\phi(x_2)$ in the feature space that are associated with $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, then their dot product can be written as $\langle \phi(x_1), \phi(x_2) \rangle = K(x_1, x_2)$.

4.1 Support Vector Machines for Classification

In the case of binary classification problems, SVM relies on the solution of the following Quadratic Programming problem to set the free parameters in (6):

$$\min_{\alpha} \left\{ \frac{1}{2} \sum_{l,m=1}^{np} \alpha_l \alpha_m y_l y_m K(x_l, x_m) - \sum_{l=1}^{np} \alpha_l \right\}$$
$$subject\ to: \begin{cases} 0 \le \alpha_l \le C, \forall l \\ \sum_{l=1}^{np} y_l \alpha_l = 0 \end{cases} \qquad (7)$$

In (7), α_l are the SVM parameters setting the class-separating surface and C is a fixed regularization term that rules the trade-off between accuracy and complexity.

Problem setting (7) has the crucial advantage of involving a quadratic-optimization problem with linear constraints, ensuring that the solution is unique. Actually, the specific choice for the kernel parameters $\{C, \sigma\}$ has an impact on the eventual generalization performance of the SVM. Both theoretical [32] and empirical [2] approaches can be adopted to determine the generalization limits. The present research follows an empirical approach involving k-fold cross validation [2].

4.2 Support Vector Machines for Regression

A SVM is used in the proposed system also to map feature-based image descriptions into scalar values that represent the perceived image quality. As such, the objective quality assessment model can be regarded as a regression problem, in which learning evolves according to an empirical sample.

In the present application, the vector, **x**, contains a feature-based description of an image, while the value y_i represents the (normalized) quality score associate with that image. SVMs regression models approximate the target function for an input vector, **x**, as

$$\hat{\gamma}_{SVM}(x) = \sum_{i=1}^{n_{SV}} (\alpha_i - \alpha_i^*) y_i K(\tilde{x}_i, x) + b \qquad (8)$$

where a_i, $a^*{}_i$ are positive parameters and b is a bias. The patterns $\{(\tilde{x}_i, y_i), i = 1, ..., n_{SV}\}$ used are a subset of the training set and are called 'support vectors'. Expression (8) shows that $\hat{\gamma}_{SVM}(x)$ is a series expansion having the kernel function $K(\cdot,\cdot)$ as a basis and involving part or all of the training patterns. Inner products can still be handled in the transformed space independently of the mapping of the original patterns; therefore, the use of the kernel trick also remains valid in regression problems.

The coefficients α_i, α_i^* and b in expression (8) must be adjusted in compliance with the input sample distribution so as to minimize some cost function measuring the deviation resulting from the approximation. To this end, Vapnik [32] suggested the use of ε-insensitive loss functions, which penalize the error whenever the absolute approximation error remains smaller than ε.

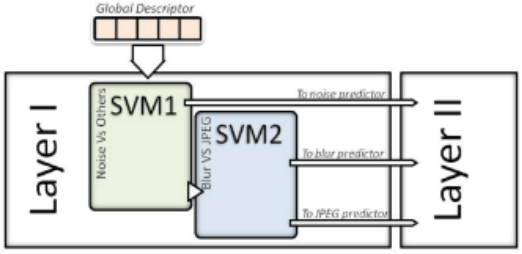

Figure 3. First Layer scheme. A first SVM detects Noisy images, while the second discerns between compressed and blurred ones.

5 Practical Implementation

In this section, a possible implementation of the double-layer system is proposed. A SVM-based classifier provides the distortion identification problem, while a SVM-based architecture maps feature-based description of images into quality scores.

As compared with [24], the present paper proposes a RR system that saves bandwidth and improves the computational cost of the assessment tool. Such goals are achieved 1) by exploiting the co-occurrence matrix for the feature-based representation of the image and 2) by developing an effective objective metric based only on two features.

5.1 Objective Metric Settings and Feature Selection

The following settings are applied to the metric described in section 3 to compute the objective descriptors $x_u^{(n)}$and$x_u^{(n,\,r)}$, corresponding to the reference image $I^{(n)}$ and the target image $\bar{I}^{(n,r)}$ respectively. The input image is divided into square sub-regions of 32x32 pixels and the co-occurrence matrix is computed on the luminance component (Y-layer) of the blocks, with settings λ=1 (neighboring pixels) and θ = 0 (horizontal direction). The set of features Φ defined in table (1) is then extracted for each block. Finally, the global descriptor is assembled by computing 6 percentiles of the distribution of each feature f_u, and combining them in the vector $x_u^{(n,r)} = \left\{\varphi_{\alpha,u}^{(n,r)}; \alpha = 0,20,40,60,80,100; u = 0,1,...,11\right\}$.

The feature selection procedure presented in section 3.3 is applied to Φ to select the two most significant features to be elaborated by the SVMs in layers I and II, independently on the specific problem to be treated. This choice is made in order not to inflate the bandwidth required by the RR model. In practice, the procedure described in 3.3 is applied to each of the learning tasks the system is supposed to tackle: distortion identification, quality mapping for Noise, Blur and JPEG. The features f_k resulting as active in the majority of the tasks are finally selected as the most effective for the whole system performance.

Eventually, the features Entropy and IMC (see table I) are selected. For each feature, the input of the SVM-based quality assessment system is obtained simply by combining the descriptors of the original and the distorted image:

$$z_u^{(n,\,r)} = \left[x_u^{(n)}, x_u^{(n,\,r)}\right] \tag{9}$$

As a result, the system processes 24 values in total, of which only twelve are required to be sent through the transmission channel together with the signal.

5.2 SVM-Based Quality Loss/Gain Quantification

The Support Vector machine technology is exploited for the prediction system implementation. In the first stage, the system is required to recognize which distortion is affecting the image under test. Hence, the role of the distortion identifier module is to solve a multiclass problem, associating each $\mathbf{z}_u$ (as per eq. 8) to distortion $a_i \in$A. We propose to implement a multiclass machine by connecting binary predictors in series, adopting a one-vs.-all strategy. The first SVM is trained to identify images distorted by a_1, and forward them to the a_1 distortion quantifier in layer II. The second SVM module recognizes images affected by a_2, and so on. Eventually, layer I will be made of l-1 SVMs, given l distortions of interest. In the present implementation, distortion caused by White Noise, Gaussian Blur and Jpeg Compression were considered (l

= 3). Layer I is implemented as in figure 3, including a first SVM handling noisy images and a second one dividing blurred from compressed samples. To privilege the system simplicity, a single feature is elaborated by layer I. The outcome of the feature selection is restricted to the better performing feature among the two classification tasks, resulting in feature Entropy.

The layer delegated for objective quality prediction (layer II) instead replicates as many independent predictors as the number of considered distortion (figure 4). Following an approach already experimented in [23], each module (also trained independently) is made of several SVM regressors, each fed by a different feature. These estimators are further integrated within an *ensemble* structure [14, 27]. Based on previous research [23, 24, 27] a simple output averaging strategy is chosen, being the most effective method for integrating the predictions of all estimators:

$$\hat{d}_S\left(q^{(n)}, q^{(n,r)}\right) = \frac{1}{U}\sum_u \hat{d}_S\left(q^{(n)}, q^{(n,r)}\right)\Big|_{f=f_u} \quad (10)$$

For the ensemble strategy to be successful, a basic requirement is to build independent estimators, based on the receptive fields theory [26]: the input space is divided into several, lower-dimensional subspaces, and a predictor is dedicated to each subspace.

Applying the coordinate-partitioning principle to the quality assessment domain leads to the specialization of each predictor on a single feature of the considered set. This setting not only validates the hypothesis of disjoint subspaces required for ensembles effectiveness, but also decreases the dimensionality of the input space, enhancing the SVM generalization ability.

Layer II involves therefore in its final configuration l quality prediction modules made each of U single SVMs gathered in an ensemble, being U the number of features selected for the task. In this research, layer II is made of three quality prediction modules (one for quantifying the effects of White noise, one for Gaussian Blur and one for JPEG compression), each including two SVMs fed by $\mathbf{z}_{Entropy}$ and $\mathbf{z}_{IMC}$(thus, $U = 2$), respectively, consistently with the feature selection procedure output.

5.3 Comparison with a Previous Implementation

The study presented in this paper represents an extension and improvement of a previously proposed work [24]. The previous implementation of the system (from now on 2LQA$_{old}$, as opposed to the one here presented, indicated with 2LQA$_{new}$) will be taken as term of comparison for the experimental validation. Nonetheless, the proposed setup already presents some peculiar advantages with respect to 2LQA$_{old}$, of which a short description will be given in the following.

2LQA$_{old}$ was characterized by an objective metric based on the color-correlogram [13] features. The color-correlogram is a second-order histogram which still measures the joint occurrence of colors at a given offset, but instead of considering only the co-occurrences along direction θ, involves in the computation color pairs in every possible direction, according to a predefined norm. By adopting the co-occurrence matrix, a very reliable distortion description is still available (see figure 2); nonetheless, one can lower the overall computational cost of the feature-extraction procedure, since at least 50% of the pixel pairs are excluded from the computation (when comparing with a correlogram computed with offset=1 and norm L1, being the less computationally heavy configuration).

The second relevant difference can be found in the feature selection procedure, which was previously performed empirically. For 2LQA$_{old}$ color-correlogram features were selected in order to maximize the generalization ability of each of the distortion-dedicated quality assessors in layer II. That choice led to the use of 4 features, corresponding to a 48-dimensional input vector for the system and to a double transmission overhead for the original image description.

Finally, while the implementation of the classification layer holds, the second layer of 2LQA$_{old}$ is based on CBP neural networks [28] ensembles. The effectiveness of this choice will have to be confirmed evaluating experimental results.

6 Experimental Validation

The second release of the LIVE database [11] was used as a testbed for the performance evalu-

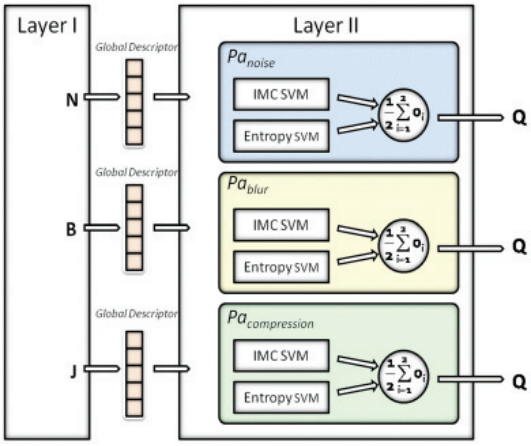

Figure 4. Layer II implementation. Three different modules are built to be specialized on the effect of one of the considered distortions. Each module includes an ensemble of 2 SVMs, each trained on a different descriptive feature.

ation of the proposed model, being a recognized benchmark in the image quality assessment field. LIVE database is based on 29 original images (from now on "image contents"). Each of this content is altered with different levels of five tipologies of distortions, originating then five subsets including both the original images and their impaired versions. For each of the 729 images in the dataset, a subjective score is provided, originated from panel sessions directly involving humans. Subjective scores, namely DMOSs, Differential Mean Opininon Scores, express the difference in quality between each impaired picture and its undistorted equivalent. Such values are the targets for layer II. For layer I, manual annotation of the distortion affecting the samples was sufficient, being independent from subjective evaluation.

The three datasets including images impaired with White noise, Gaussian Blur and JPEG compression were considered for validation.

To ensure robustness and avoid image content-related learning effects, a k-fold-like testing strategy was adopted. Five groups of images were created, each containing all the distorted versions of disjoint subsets of image contents. Both layers were then tested performing 5 runs, in each of which alternatively 4 of the 5 folds were used as training data and the remaining one was used as test data. In this way, the machines were tested on image contents never processed during the training.

Experimental results are presented as follows. First, details concerning the set up of the layer I are presented. Then, the development of the distortion-oriented assessment modules is discussed. Finally, the performance of the eventual quality assessment system combining the two layers is compared with other approaches proposed in the literature.

6.1 Precision in Distortion Identification

The two SVMs of the first layer were trained independently. The *k*-fold cross-validation technique was applied to tune the kernel parameters. For the first task, noisy images recognition, the SVM was trained on a dataset resulting from the merge of the three LIVE sets. A linear kernel handled the problem successfully, as shown in table II. The parameter C was finally set to $2.8*10^2$. The second SVM was trained on a subset of the previous dataset, including only blurred and compressed images. A normalized second order polynomial kernel as formulated in [4] was preferred for this task. Based on the cross-validation output, the parameter C was set to 1.310^5.

Table II reports the classification errors for each run and for both SVM classifiers. While the performance of the first classifier is almost perfect, the second SVM lacks in precision, due to the intrinsically more complex problem. The perceptual overlap between compression and blurring artifacts (JPEG compression causes also blur) is reflected also in the model: in this case a non-linear kernel was necessary, and the setting of the parameter C indicates increased complexity. Nonetheless, on average the percentage of misclassified images

is less than 6.5%, gaining more than 1% in accuracy with respect to the 2LQA$_{old}$ system. This result validates, at least for layer I, the choice of a co-occurrence matrix-based metric and of the feature selection output. Also, the use of a polynomial kernel seems to be more appropriate for the second task tackling.

6.2 Accuracy in Quality Loss Prediction

The three ensembles of SVMs implementing the second layer were each trained on a different dataset. The datasets for White Noise and Blurred images contained 145 patterns; the remaining testbed included 159 JPEG compressed images. DMOSs, originally ranging between [0,100] were remapped for computational reasons into the range [-1, +1]. The six SVMs were all equipped with a RBF kernel; thanks to the intrinsic flexibility of the system it was possible to select optimum models independently for every predictor. The final settings for the Noise predictor were $\{ C^{Noise}_{Entropy} = 100, \quad \sigma^{Noise}_{Entropy} = 2\}$ for the Entropy based SVM and $\{ C^{Noise}_{IMC} = 100, \quad \sigma^{Noise}_{IMC} = 1\}$ for the IMC-based SVM. Following the previous notation, the setting used for the Blur predictor were $\{ C^{Blur}_{Entropy} = 10, \quad \sigma^{Noise}_{Entropy} = 1\}$, $\{ C^{BLur}_{IMC} = 10, \quad \sigma^{Blur}_{IMC} = 0.5\}$; and for the JPEG Compression predictor $\{ C^{JPEG}_{Entropy} = 10, \quad \sigma^{JPEG}_{Entropy} = 1\}$, $\{ C^{JPEG}_{IMC} = 100, \quad \sigma^{JPEG}_{IMC} = 0.5\}$.

To evaluate the second layer performance, we report several parameters which measure the discrepancy between the estimated change in quality, $\hat{d}_S\left(q^{(n)}, q^{(n,\,r)}\right)$, and the actual variation provided by the LIVE database, $d_S\left(q^{(n)}, q^{(n,\,r)}\right)$. Four quantities are considered:

- The Pearson's Correlation Coefficient, ρ;
- The Spearman's Rank Order Correlation Coefficient, SROCC;
- The mean percentage prediction error, $\%\mu_{|err|}$, where $\mu_{|err|}$ is the value of the absolute prediction error between d_s and $\hat{d}_S$.
- The Root Mean Square prediction Error, RMSE, between d_s and $\hat{d}_S$.

The first two indicators are recommended by the VQEG committee for objective metric performance evaluation, being a measure of prediction monotonicity, i.e. of the consistency between the rank ordering of the samples given by the OQA and that provided by humans in subjective tests. Complimentary, $\mu_{|err|}$ and RMSE are given as measures of the prediction accuracy. Tables 3 to 5 show the output of the second layer of the proposed system compared to the performance of 2LQA$_{old}$.

In general, the system achieves considerably high accuracy, particularly when dealing with artifacts caused by Gaussian Noise, for which the percentage prediction absolute error is lower than 6% and the Correlation of the assessments with the subjective scores is 0.96. Dealing with compressed images, the RMSE is lower than 0.18 on a two points scale, which is acceptable for real-world applications. Finally, as for 2LQA$_{old}$, some lack in precision is presented by the Blur Predictor.

With respect to 2LQA$_{old}$, the proposed system gains in prediction monotonicity, for both Blur and Compression effects prediction, while a slight decrease in accuracy occurs for the three quality assessors. The differences in the systems setup should be taken in account: although the proposed system consistently simplifies the metric computation and the system requirements, it still allows obtaining an increase in correlation of the predicted scores, just slightly loosing in accuracy.

6.3 Comparison with Other Approaches

As a further validation of the system, table 6 and 7 compare the proposed approach with several well known OQAs. General Purpose RR OQAs are actually very few [34, 15], hence further comparison is provided with the two well-known FR metrics MSSIM and PSNR. The proposed system compares satisfactorily with the metric proposed by Li and Wang (yet not including distortion identification), and outperforms the method proposed in [34] for all distortion types except for Gaussian Blur Prediction, already recognized as a weak point of the model. In the comparison with MSSIM and PSNR, two details should be taken into account, namely (1) the difference in the original image availability for the computation of FR and RR metrics and (2) the fact the quantities reported were computed after a non-linear regression of all patterns for each dataset and the generalization ability of the obtained models was tested only using images that had been used

Table 2. Performance of each SVM machine for distortion identification in terms of % of misclassified patterns

	Proposed system		**2LQA$_{old}$ [27]**	
	Noise vs. All	**Blur vs. JPEG**	**Noise vs. All**	**Blur vs. JPEG**
Run #1	0.00%	1.49%	0.00%	1.49%
Run #2	1.04%	3.18%	1.07%	1.58%
Run #3	0.00%	3.22%	1.09%	4.84%
Run #4	0.00%	14.51%	0.00%	19.35%
Run #5	0.00%	10.00%	0.00%	12.76%
Average	**0.26%**	**6.48%**	**0.26%**	**7.56%**
Kernel	Linear	Polynom.	RBF	RBF
C	$2.8 * 10^2$	$1.3 * 10^5$	10^4	10^5

Table 3. Performance of the quality estimator for noisy images in terms of correlation (Pearson and Spearman Coefficients and errors (absolute and RMSE) between predicted and subjective quality scores.

	Proposed system				**2LQA$_{old}$ [27]**			
	ρ	SROCC	$\mu_{\|err\|}\%$	rmse	ρ	SROCC	$\mu_{\|err\|}\%$	rmse
Run #1	0.954	0.921	6.376	0.143	0.981	0.971	3.119	0.082
Run #2	0.942	0.951	5.864	0.149	0.937	0.944	5.216	0.163
Run #3	0.979	0.969	4.591	0.115	0.989	0.984	2.677	0.070
Run #4	0.954	0.956	5.777	0.142	0.985	0.982	3.122	0.079
Run #5	0.965	0.963	4.439	0.116	0.976	0.964	4.656	0.110
Average	**0.959**	**0.952**	**5.409**	**0.133**	**0.974**	**0.969**	**3.758**	**0.101**

Table 4. Performance of the quality estimator for blurred images in terms of correlation (Pearson and Spearman Coefficients and errors (absolute and RMSE) between predicted and subjective quality scores.

	Proposed system				**2LQA$_{old}$ [27]**			
	ρ	SROCC	$\mu_{\|err\|}\%$	rmse	ρ	SROCC	$\mu_{\|err\|}\%$	rmse
Run #1	0.933	0.897	7.867	0.189	0.946	0.946	5.365	0.136
Run #2	0.915	0.894	7.724	0.183	0.668	0.643	12.328	0.334
Run #3	0.839	0.842	8.653	0.213	0.966	0.972	3.758	0.100
Run #4	0.936	0.933	7.139	0.169	0.914	0.912	6.469	0.154
Run #5	0.812	0.834	14.270	0.392	0.892	0.868	10.244	0.240
Average	**0.887**	**0.880**	**9.130**	**0.229**	**0.877**	**0.868**	**7.633**	**0.193**

Table 5. Performance of the quality estimator for JPEG Compressed images in terms of correlation (Pearson and Spearman Coefficients) and errors (absolute and RMSE) between predicted and subjective quality scores.

	Proposed system				**2LQA$_{old}$ [27]**			
	ρ	SROCC	$\mu_{\|err\|}\%$	rmse	ρ	SROCC	$\mu_{\|err\|}\%$	rmse
Run #1	0.914	0.902	7.078	0.181	0.944	0.922	5.367	0.138
Run #2	0.886	0.872	7.522	0.190	0.857	0.864	7.186	0.207
Run #3	0.944	0.894	6.217	0.165	0.920	0.860	7.094	0.179
Run #4	0.934	0.923	6.287	0.151	0.939	0.912	4.579	0.121
Run #5	0.910	0.908	7.992	0.207	0.882	0.868	8.712	0.225
Average	**0.917**	**0.900**	**7.019**	**0.179**	**0.908**	**0.885**	**6.588**	**0.174**

in the training process.

7 Conclusions

A Reduced-reference, double layer system for objective image quality assessment is proposed. This general purpose system is designed to first recognize which distortion is affecting the image, and then to quantify the quality loss caused by the presence of such distortion. Both layers are supported by Support Vector Machines, trained in the first case to correctly classify images according to the distortion affecting them, and in the second case to understand the mapping between a numerical representation of the image and the quality impairment brought about by the applied distortion. The numerical description of the image (Objective Metric) is designed to minimize both the computational cost and the transmission bandwidth requirements.

The proposed implementation of the system allows consistent savings in computational time and bandwidth. With respect to the system presented in [24], a co-occurrence matrix is used in place of the most expensive Color Correlogram, reducing the computational time up to 50%. The stricter feature selection, applied offline, selects only two values to be computed from the original and distorted signals, compared to the four required before. This brings a twofold benefit. Firstly, it decreases the computational effort necessary for the numerical description of the images. Secondly, it allows characterizing one image with 24 values, i.e., with 96 bytes, on common 32-bits architectures. As a consequence, on the sender side, only two quantities have to be computed, and an overhead $rr_{info} <$ 0.05 KB is sent throughout the channel as metadata. This is of the major importance for emerging multimedia technologies, such as video streaming on mobile phones, for which not always large bandwidth availability can be assumed.

The performance of the system is not compromised with respect to its less efficient version [24], also well comparing with the state of the art of available Reduced-Reference quality assessment metrics.

A possible limitation of the proposed system is the current inability of handling the combined effect of different distortions. A heavily blurred image presenting traces of noise would be at present processed as if no noise was applied. However, different artifacts contribute in a different way to the final quality evaluation; therefore, all of them should be taken into account. Focusing on predicting the annoyance of different types of visual artifacts, would be greatly beneficial, allowing the abstraction of their perceptual impact from the actual distortion producing them. Further effort should then be put in understanding how to combine the impact of different distortions in estimating the overall quality of the picture. Needless to say, intensive subjective studies are required to make this development concrete.

References

[1] V. Baroncini. New tendencies in subjective video quality evaluation. *IECIE Transactions on Fundamentals*, 11(89):2933–2937, 2006.

[2] Lugosi G. Bartlett P., Boucheron S. Model selection and error estimation. *Machine Learning*, 48(1–3):85–113, 2002.

[3] H.R. Sheikh.and A.C. Bovik. Image information

and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[4] A. Boni D. Anguita, S. Ridella, F. Rivieccio, and D. Sterpi. *Theoretical and Practical Model Selection Methods for Support Vector Classifiers*, pages 159–179. Springer, 2005.

[5] P. Engeldrum. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek Press, Winchester, 2000.

[6] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *IEEE Trans. on Communications*, 43(12):2959–2965, 1995.

[7] D.D. Giusto G. Ginesu, F. Massidda. A multifactors approach for image quality assessment based on a human visual system model. *Signal Processing: Image Communication*, 21:316–333, 2006.

[8] Z. Gao and Y.F. Zheng. Quality constrained compression using dwt-based image quality metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(7):910–922, 2008.

[9] A.C. Bovik H.R. Sheikh and L. Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE Trans. Image Process*, 14(11):1918–1927, 2005.

[10] L. Cormack H.R Sheikh, Z. Wang and A.C. Bovik. Live image quality assessment database at http://live.ece.utexas.edu/research/quality. Technical report.

[11] M.F. Sabir H.R. Sheikh and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithm. *IEEE Trans. Image Processing*, 15(11):3441–3452, 2006.

[12] R.G. Laha I.M. Chakravarti and J. Roy. *Handbook of Methods of Applied Statistics*, volume I. John Wiley, 1967.

[13] S. Ravi Kumar J. Huang, M. Mitra, W.J. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proc. IEEE CVPR '97*, pages 762–768, 1997.

[14] R.P.W. Duin J. Kittler, M. Hatef and J. Matas. On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[15] Q. Li and Z. Wang. General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation. In *IEEE International Conference on Image Processing*, San Diego, CA, 2008.

[16] H. Liu and I. Heynderickx. A perceptually relevant no-reference blockiness metric based on local image characteristics. *EURASIP Journal on Advances in Signal Processing*, 2009.

[17] D. Barba M. Carnec, P. Le Callet. Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Processing: Image Communication*, 23:239–256, 2008.

[18] I. Van Zyl Marais and W.H. Steyn. Robust defocus blur identification in the context of blind image quality assessment. *Signal Processing: Image Communication*, 22:833–844, 2007.

[19] S. Winkler P. Marziliano, F. Dufaux and T. Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. *Signal Processing: Image Communication*, 19:163–172, 2004.

[20] ITU-T Recommendation P.911. *Subjective audiovisual quality assessment methods for multimedia applications*. Geneva, 1998.

[21] T.N. Pappas and R.J. Safranek. *Perceptual criteria for image quality evaluation*. Stateplace New York: Academic, 2000.

[22] R.Muijs P.Gastaldo, R. Zunino and I. Heynderickx. Building neural systems for no-reference quality assessment. In *Proc. of the First International Workshop on Video Processing and Quality Metrics(VPQM'05)*, 2005.

[23] Zunino R. Redi J., Gastaldo P. and Heynderickx I. Co-occurrence matrixes for the quality assessment of coded images. In *International Conference on Artificial Neural Networks (ICANN'08)*, 2008.

[24] Zunino R. Redi J., Gastaldo P. and SnplaceHeynderickx SnI. Reduced reference assessment of perceived quality by exploiting color information. In *Proc. International Conference on Artificial Neural Networks (ICANN'09)*, 2009.

[25] K. Shanmugam R.M. Haralick and I. Dinstein. Textural features for image classification. *IEEE Trans. On Systems, Man and Cybernetics SMC-3*, pages 610–621, 1973.

[26] D.E. Rumelhart and J.L. McClelland. *Parallel distributed processing*. MIT Press, Cambridge, MA, 1986.

[27] E. Bienenstock S. Geman and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–48, 1992.

[28] S. Rovetta S. Ridella and R. Zunino. Circular back-propagation networks for classification. *IEEE Trans. on Neural Networks*, 8(1):84–97, 1997.

[29] B. Schlkopf and A. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[30] H.-J. Zepernick T.M. Kusuma. On perceptual objective quality metrics for in-service picture quality monitoring. In *Third ATcrc Telecommunications and Networking Conference and Workshop*, Melbourne, Australia, 2003.

[31] International Telecommunication Union. *Methodology for the subjective assessment of the quality of television pictures ITU-R BT.500*. 1995.

[32] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[33] S. Winkler. Issues in vision modeling for perceptual video quality assessment. *Signal Processing*, 78:231–252, 1999.

[34] E.P. Simoncelli Z. Wang. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Proceedings of SPIE Human Vision and Electronic Imaging X*, volume 5666, pages 149–159, San Jose, CA, 2005.

[35] H.R. Sheikh Z. Wang and A.C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *IEEE Int. Conf. Image Processing*, pages 477–480, Rochester, NY, 2002.

[36] H.R. Sheikh Z. Wang. and A.C. Bovik. *Objective video quality assessment*. CRC Press, Boca Raton, FL, 2003.

[37] H.R. Sheikh Z. Wang, A.C. Bovik and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 2004.

[38] H.R. Sheikh Z. Wang, G. Wu, E.P. Simoncelli, E.H. Yang, and A.C. Bovik. Quality-aware images. *IEEE Transactions on Image Processing*, 15(6):1680–1689, 2006.

[39] S.Winkler Z. Yu, H.R. Wu and T. Chen. Vision-model-based impairment metric to evaluate blocking artifact in digital video. In *Proc. IEEE*, volume 90, pages 154–169, 2002.

[40] Y. Horita Z.M. Parvez Sazzad, Y. Kawayoke. No reference image quality assessment for jpeg2000 based on spatial features. *Signal Processing: Image Communication*, 23:257–268, 2008.