# Color Distribution Information for the Reduced-Reference Assessment of Perceived Image Quality

Judith A. Redi, Paolo Gastaldo, Ingrid Heynderickx, and Rodolfo Zunino, *Member, IEEE*

***Abstract*—Reduced-reference systems can predict in real-time the perceived quality of images for digital broadcasting, only requiring that a limited set of features, extracted from the original undistorted signals, is transmitted together with the image data. This paper uses descriptors based on the color correlogram, analyzing the alterations in the color distribution of an image as a consequence of the occurrence of distortions, for the reduced-reference data. The processing architecture relies on a double layer at the receiver end. The first layer identifies the kind of distortion that may affect the received signal. The second layer deploys a dedicated prediction module for each type of distortion; every predictor yields an objective quality score, thus completing the estimation process. Computational-intelligence models are used extensively to support both layers with empirical training. The double-layer architecture implements a general-purpose image quality assessment system, not being tied up to specific distortions and, at the same time, it allows us to benefit from the accuracy of specific, distortion-targeted metrics. Experimental results based on subjective quality data confirm the general validity of the approach.**

***Index Terms*—Computational intelligence, correlogram, image quality assessment.**

## I. Introduction

MODERN CONSUMER electronics devices often embed post-processing systems, which can automatically estimate the perceived quality of the incoming image and enhance it whenever required. To this end, they use objective quality assessment, based on algorithms that can automatically assess the quality of images or videos [1]–[7] in agreement with human quality judgments [8]. The latter is usually measured through subjective studies [9], [10].

Objective quality predictors are often designed to quantify the image degradation due to visible artifacts. The term "artifact" follows here Keelan's definition, and denotes an image attribute causing, if visible, degradation in quality [11]. Objective quality assessment algorithms are either targeted: 1) to cover a wide range of distortions (general-purpose systems), or 2) to deal with specific artifacts or distortions (distortion-oriented systems); the term *distortion* here is used to indicate some kind of "image distortion" (e.g., JPEG, JPEG2000, noise, blur, and so on), following the terminology adopted in seminal works such as [12], [13]. Additionally, objective quality assessment algorithms can be classified according to the availability of an original (distortion-free) image, with which the distorted image is to be compared. Full-reference (FR) methods [8], [12], [14]–[16] directly compare the received and the undistorted (reference) images. In spite of their accuracy [12], [13], FR methods prove impractical when the reference image is unavailable. No-reference (NR) methods [8], [16], [17]–[20] can assess perceived quality without any information about the original image, but are usually targeted to a predefined set of distortions, and therefore their applicability is limited.

In this paper, a reduced reference (RR) approach is preferred. RR methods [8], [16], [21]–[25] are designed to allow quality assessment being supported by only a limited number of numerical features extracted from the original image, paying the feasible additional cost of transmitting side information through the video-chain endpoints [3], [7]. In practice, the resulting small-sized information can be included in the video signal as metadata without major changes in existing broadcasting protocols. The challenging task in the design of a RR method is the selection of the smallest set of features that can support quality assessment effectively. Such an approach currently seems the most promising when addressing on-the-fly, real-time estimation of visual quality, as, compared to NR quality assessment, it achieves higher prediction accuracy.

This paper presents a general-purpose image quality assessment system implementing a RR strategy. The novelty aspects of this paper consist both in using color information in the assessment process, and in developing a fully adaptive methodology based on computational intelligence paradigms, which endow the overall framework with a notable flexibility, a remarkable accuracy, and the capacity of empirical adjustment.

The baseline for using color information in the RR method is that distortions due to digital processing can alter the original color distribution of an image significantly. Also,

Manuscript received July 9, 2009; revised April 15, 2010 and July 8, 2010; accepted July 11, 2010. Date of publication October 14, 2010; date of current version January 22, 2011. This paper was recommended by Associate Editor R. L. de Queiroz.

J. A. Redi is with the Mediamatics Department, Delft University of Technology, Delft 2628CD, The Netherlands (e-mail: j.a.redi@tudelft.nl).

P. Gastaldo and R. Zunino are with the Department of Biophysical and Electronic Engineering, University of Genoa, Genoa 16145, Italy (e-mail: paolo.gastaldo@unige.it; rodolfo.zunino@unige.it).

I. Heynderickx is with the Philips Research Laboratories, AA Eindhoven 5656, The Netherlands, and also with the Delft University of Technology, Delft 2628CD, The Netherlands (e-mail: ingrid.heynderickx@philips.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2010.2087456

it is known [26] that color plays a relevant role in quality perception. The approach presented in this paper uses luminance and hue information to numerically represent the images. Second-order histograms, and in particular color correlograms (CC) [27], combine such information in low-dimensional numerical descriptors.

The use of computational intelligence paradigms supports the critical task of recreating the highly non linear behavior of the human visual system (HVS). Traditional techniques [12], [23] usually decouple the quality assessment task into two steps, first defining an objective metric to obtain a feature-based representation of the image, and then mapping this lower-dimensional description into quality scores by means of regression [28]. These kinds of systems focus on the elaboration of reliable but complex metrics, often computationally expensive. In this paper the opposite approach is attempted: the modeling power of computational intelligence tools allows (partially) treating the HVS as a black-box, whose inputs define a numerical description of the image, and the actual mapping into quality scores, i.e., the perceptual mechanisms, is handled by the learning machines. Several examples are also available in the image quality literature on how to exploit learning machines to increase the accuracy of objective quality estimators [29]–[34]. In the proposed model, Computational Intelligence methods support the design of a two-layer, hybrid architecture that can handle differently distorted images, still taking advantage of the benefits in precision given by the use of distortion-oriented metrics. In the first layer, a neural classification algorithm determines the type of distortion affecting the image. The second layer integrates a set of dedicated quality predictors; each predictor hosts a neural regression machine and is trained to assess quality when images are affected by a specific distortion. Thus, the eventual quality assessment system can automatically use the best predictor without having any *a priori* information on the image to be evaluated.

The second release of the LIVE database [35] is used for the performance evaluation of the proposed approach. The results confirm the general validity of the framework based on computational intelligence paradigms, as well as of the use of color statistics to efficiently predict image quality by adopting a RR paradigm.

This paper is organized as follows. Section II gives an overview of the strategy adopted to implement the system. Section III analyzes and illustrates the use of color information in objective quality assessment, whereas Section IV briefly recalls the Computational Intelligence techniques adopted in this paper. The actual system implementation is then proposed in Section V and tested in Section VI. Concluding remarks are made in Section VII.

## II. Two-Layer Architecture for Distortion Classification and Quality Estimation

The RR strategy presented here adopts a two-layer, integrated approach to treat image distortions (Fig. 1). In the first layer, the distortion affecting the image is identified by means of a classification process. The image is then routed to a specialized predictor, trained to evaluate the effect of the identified distortion on quality. The architecture offers the major advantage of intrinsic flexibility, as each system module may include a dedicated, independent model.

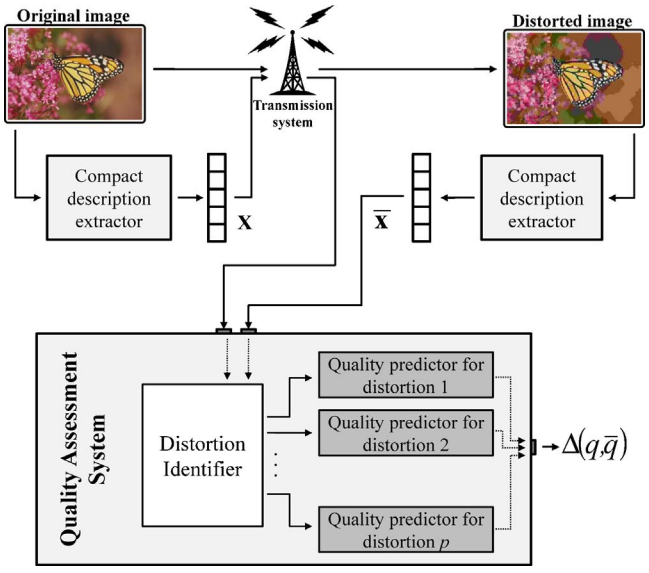


Fig. 1. Overview of the two-layer reduced reference quality assessment system.

The overall framework can be explained using the following notation. Let $I$ be an undistorted (reference) image, and let $\bar{I}$ be the image obtained by applying some distortion to $I$. Let $\mathbf{x}$ and $\bar{\mathbf{x}}$ define some feature-based representation of $I$ and $\bar{I}$, respectively. Finally, let $q$ and $\bar{q}$ be the subjective quality ratings associated to $I$ and $\bar{I}$, respectively. During the training phase, the system is taught to reproduce the distance, $\Delta(q, \bar{q})$, between the pair of subjective scores by processing the pair of objective descriptors, $\{\mathbf{x}, \bar{\mathbf{x}}\}$, associated with the two images. To achieve this goal, first the set $D = \{d_1, \ldots, d_p\}$ of the distortions to be addressed is established. Then: 1) a dedicated quality predictor $\Omega(d_i)$ is trained and implemented for each distortion, and 2) the corresponding distortion identifier is developed.

At *run time*, one cannot make any *a priori* assumption about the distortion, $d_i \in D$, affecting the received image, $\bar{I}$. The RR distortion identifier receives $\mathbf{x}$ as a part of the metadata, and computes $\bar{\mathbf{x}}$; the comparison between these descriptive vectors makes it possible to classify the distortion, $d_i$, actually present in $\bar{I}$. The pair of descriptive vectors enters the distortion-specific quality predictor, $\Omega(d_i)$, that is entrusted to quantify the effect of $d_i$ on the perceived quality. The eventual result is an objective estimate of the difference in subjective quality, $\Delta(q, \bar{q})$, between the reference and the distorted image.

In such a framework, the two layers can actually be designed either to support a distortion-oriented system or to support an artifact-oriented system. In the latter case, the quality predictors are specifically trained to deal with each distinct artifact and the first layer should be developed accordingly. Indeed, this paper addressed a distortion-oriented system following the problem setup adopted by several researches on image quality assessment proposed in literature (e.g., [12], [21], [23], [34]).

## III. Using Color Distribution Information for Objective Quality Assessment

The metrics that feed predictive models in objective approaches typically use the luminance component of the color information, as the relevance of luminance information in image quality assessment has already been proved extensively [12]. Recent subjective studies [26], however, showed that chrominance also plays a relevant role in quality perception, suggesting that quality measurements exclusively based on luminance information could provide overestimated (i.e., optimistic) scores. On the other hand, limiting the metric computation to color information only would prevent the metric from being used in a variety of applications involving grey-scale images (e.g., medical imaging). Therefore, our approach involves both luminance and hue information in the quality evaluation process.

RR approaches require the images to be encoded into a feature-based representation, which should be both informative and low-dimensional, as it must be attached to the original image as metadata. Hence, this paper uses second-order histograms to represent color information, under a twofold assumption: first, the distortions introduced by digital processing algorithms can seriously alter the original color distribution, and second, the alterations in color brought about by distortions are usually distinctive to the distortion itself, and can be modeled by comparing the statistics of the original and distorted images. Second-order histograms have been employed successfully in different recent applications, such as image indexing [27] and texture recognition [36]. Co-occurrence matrices [37] have been proved effective [24], [39], in the B/W domain, to convey significant information about perceived image quality. The research presented here adopts the *color correlogram* [27], which expresses how the spatial correlation of pairs of colors changes with distance, for extracting numerical features.

### A. Color Correlogram for Local Information Extraction

The color correlogram for an image region $Z$ (including, in general, $W \times H$ pixels) describes the spatial correlation of pairs of colors with respect to the distance $k$. Formally, given a set of $L$ bins $C = \{C_1, \ldots, C_L\}$ (typically $L = 256$ bins are used when color channels are represented by 8 bits), the element of the correlogram matrix is defined as follows:

$$G^k(i,j) = \left| \left\{ \begin{array}{l} (m,n), m < W, n < H, \quad \text{s.t.} \\ Z[m,n] = C_i;\ Z[p,q] = C_j; \\ dist(Z[m,n], Z[p,q]) = k \end{array} \right\} \right| \quad i, j = 1, ..., L \tag{1}$$

where the operator $|.|$ defines the cardinality of a set and $dist()$ is a measure of distance between a pair of pixels. Each element of the matrix gives the probability of finding a pixel having color $C_j$ at a distance $k$ from a pixel having color $C_i$. In the present work, $dist()$ embeds the L1-norm, hence only the pairs of pixels lying at a distance $k$ in the horizontal/vertical direction are considered in the histogram computation.

To preserve local information, every image is split into non-overlapping square regions (blocks), each holding $H \times H$ pixels. The block size, $H$, is crucial: small settings of $H$ yield a considerable number of values and ensure a significant statistical sampling; on the other hand, using larger blocks decreases the relative percentage of pixels that do not enter the computation of $G^k(i, j)$, and therefore limit possible border effects. Typically, this parameter is set to $H > 8$. For each block, the color correlogram is computed and a set of features, $\Phi$, is worked out. Table I gives the objective descriptors used in the feature-extraction process.

TABLE I
Objective Features Derived from the Color Correlogram

| Feature Name | Definition |
|---|---|
| Diagonal energy | $f_1 = \sum_i \left[G_Z^k(i,i)\right]^2$ |
| Entropy | $f_2 = -\sum_{i,j} G_Z^k(i,j) \log_2 G_Z^k(i,j)$ |
| Contrast | $f_3 = \sum_q q^2 \left( \sum_{i-j=q} G_Z^k(i,j) \right)$ |
| Homogeneity | $f_4 = \sum_{i,j} G_Z^k \Big/ \left[1 + (i-j)^2\right]$ |
| Energy ratio | $f_5 = \sum_{i,j} \left[G_Z^k(i,j)\right]^2 \Big/ \sum_i \left[G_Z^k(i,i)\right]^2$ |

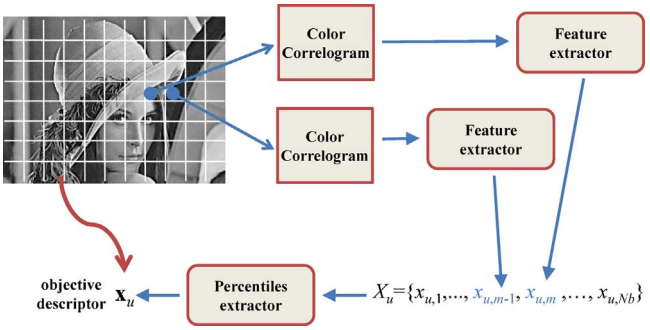


Fig. 2. Feature extraction. The procedure eventually yields the objective descriptor $\mathbf{x}_u$ for an image $I$ and a feature $f_u$.

### B. Global-Level Numerical Representation

The local-level feature extraction gives, for each feature $f_u \in \Phi$, as many values as the number of blocks composing the input image. Subjective scores, however, express the overall quality of a whole image. Therefore, it seems reasonable that the numerical image description consists of one vector of feature values, to be associated with the single quality score. Toward this end, the present approach aggregates block-based information by means of statistical percentiles, to render the distribution of a feature, $f_u \in \Phi$, over the image.

Algorithm 1 outlines the procedure to assemble the global descriptors for both the reference and the distorted image. The procedure is also schematized in Fig 2.

The eventual result of the process is, for each feature $f_u \in \Phi$, the description of an image by means of a global-level vector, $\mathbf{x}_u$, whose dimensionality depends on how many percentiles are worked out to characterize $X_u$. The global descriptor $\mathbf{x}$ for an image $\bar{I}$ results from gathering the objective vectors $\mathbf{x}_u$, being $n_f$ the number of features $f_u \in \Phi$: $\mathbf{x} = \{\mathbf{x}_u, u = 1, ..., n_f\}$.

Fig. 3 illustrates how the proposed global-level image descriptors can effectively capture the distortions in an image. The

**Algorithm 1** *Feature Extraction*

Inputs: a picture $\bar{I}$, a descriptive feature $f_u \in \Phi$, and a value for distance $k$

1. Block level feature extraction
   a. Split $\bar{I}$ into $N_b$ non-overlapping square blocks, and obtain the set: $\mathrm{B} = \{b_m;\ m = 1, ..., N_b\}$
   b. For each block $\mathrm{b_m} \in \mathrm{B}$: compute the associate color correlogram: $G_m^k(i, j)$
   c. For each $G_m^k$: compute the value $x_{u,m}$ of feature $f_u$
   d. Obtain the set $X_u$ that gathers the feature values for each block$\mathrm{b_m} \in \mathrm{B}$: $\mathrm{X}_u = \{x_{u,m};\ m = 1, .., N_b\}$
2. Global level numerical representation
   Assemble the objective descriptor vector, $\mathbf{x}_u$, for the feature $f_u$ on $\bar{I}$
   $\mathbf{x}_u = \{\varphi_{\alpha,u}\}$
   where $\varphi_{\alpha,u}$ is the $\alpha$th percentile of $\mathrm{X}_u$

Output: A global descriptor $\mathbf{x}_u$ for Image $\bar{I}$

figure shows one image content of the LIVE database [35] and addresses four image distortions: JPEG, JPEG2000, Gaussian Blur, and White Noise. The five charts report the behavior of $\mathbf{x}_u$ for the five features $f_u$, $u = 1, \ldots 5$, reported in Table I. The $y$-axis plots the $i$th percentile ($x$-axis) of the set $X_u$. It is worth stressing that the chart of the feature *energy ratio* uses a secondary $y$-axis to plot the set $X_u$ of the distortion "noise" in the same graph with the other distortions.

Fig. 4 exemplifies the sensitivity of the computed features to the distortion level. Distortion maps (second and forth rows) are reported for two distortions and features, comparing the original map for the image content *Bikes* (from LIVE) to those corresponding to the same image with medium and high level of distortion. The feature value/distortion level of each block of the image (in this case, $H = 32$) is represented by its color.

### C. Global-Level Numerical Representation

Using the whole feature set for quality assessment might prove too expensive in practice, since the global descriptors $\mathbf{x}$ lie in a huge space having dimension $d = n_f \times n_\alpha$, with $n_\alpha$ being the number of computed percentiles. Therefore, one applies a selection criterion to identify the most informative descriptors for each prediction module. The proposed unsupervised approach adopts Kolmogorov-Smirnov's test [38] (KS). The analysis selects from the complete set, $\Phi$, only the "active" features, i.e., those whose statistical properties depart significantly from their original values after applying the distortion $d_i$ at varying levels, $r$.

The data set is obtained by dividing training images, $\Psi = \{I^{(s)}, s = 1, \ldots, n_p\}$, into two disjoint subsets, $\Psi_1$ and $\Psi_2$, and applying distortion $d_i$ to $\Psi_2$, resulting in a set of processed images, $\overline{\Psi}_2^{(i,r)} = \{\bar{I}^{(s,i,r)},\ I^{(s)} \in \Psi_2\}$. For each objective feature $f_u \in \mathrm{F}$, and for each distortion level $r$, the analysis statistically

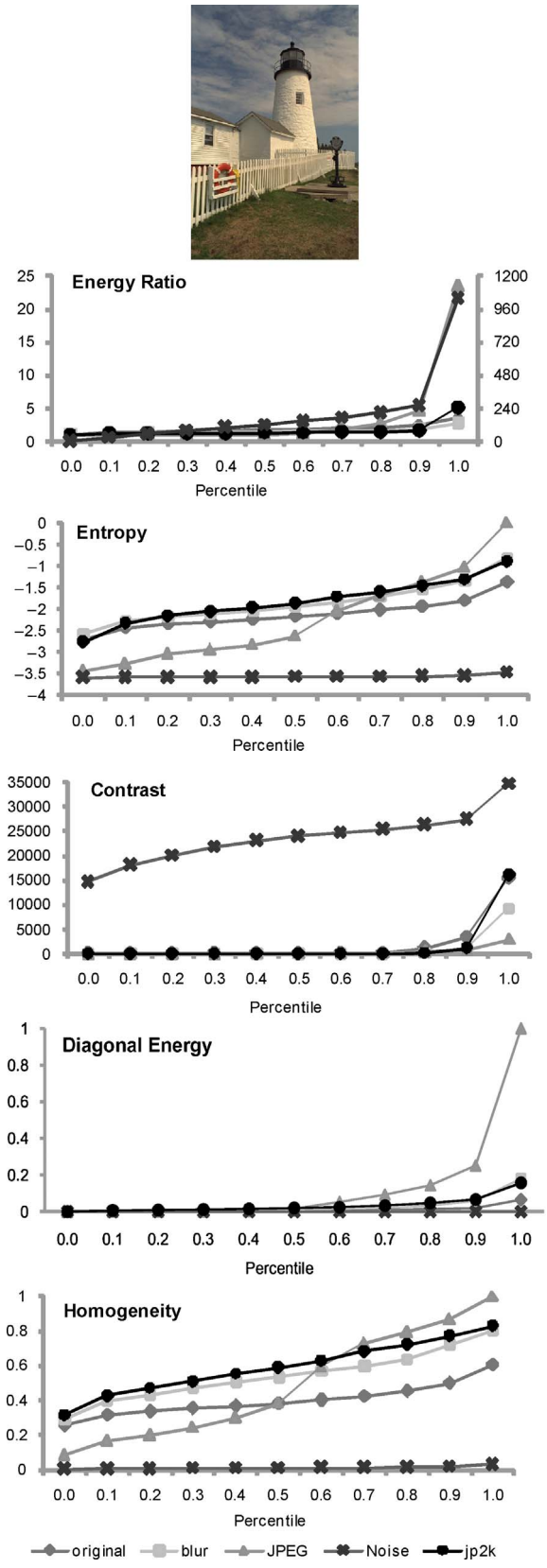


Fig. 3. Illustration of the sensitivity of the color correlogram features to different kinds of distortions. The graphs show the distribution of a feature (extracted according to the procedure as per Section III-B) for the image lighthouse and four distorted versions. The distorted images are taken from LIVE and are: image #44 of the JPEG set 1, image #96 of the White Noise set, image #106 of the JPEG2000 set 1, and image # 97 of the Gaussian Blur Dataset.

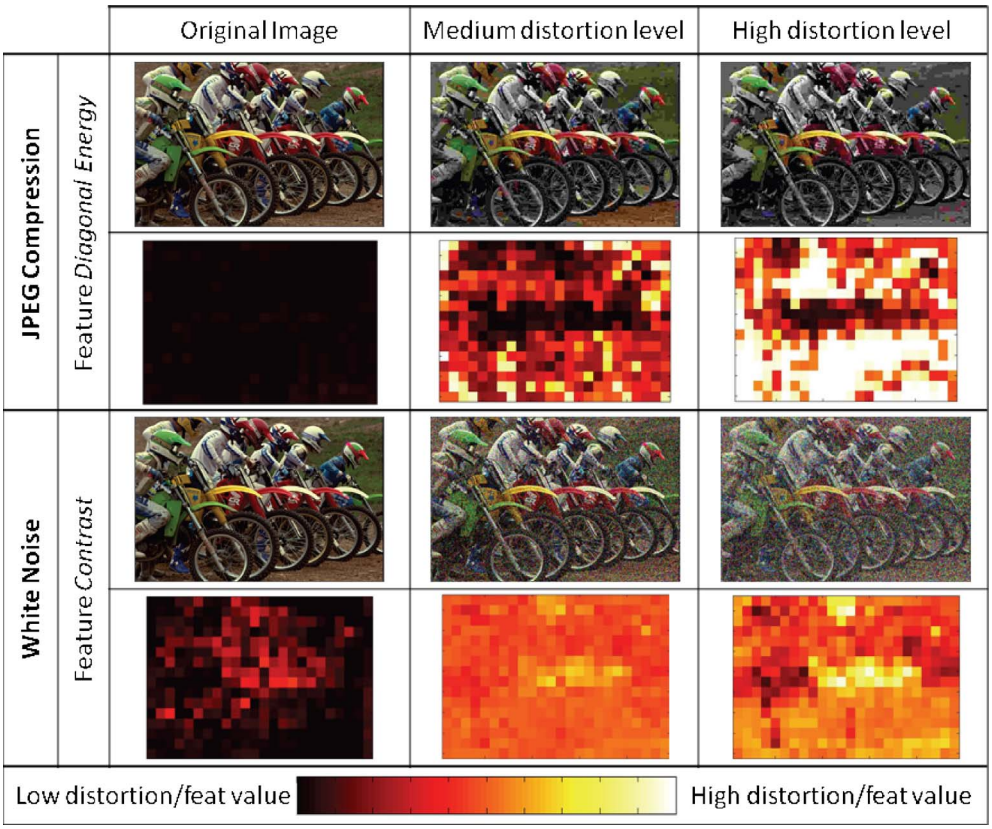


Fig. 4. Sensitivity of the color correlogram features to different distortion levels. JPEG compression and noise are taken as an example. For JPEG, the distortion maps correspond to images #29 and #77 of LIVE JPEG set 1, while for White Noise, the images #29 and #97 are analyzed.

compares two samples: one containing the values of $f_u$ for $Y_1$, the other containing the values of $f_u$ for $\overline{\Psi}_2^{(i,r)}$. Feature values are worked out on non-overlapping blocks of pixels randomly extracted from each image. The mutual independence of the data sets allows one to use the KS test to disprove the null hypothesis, i.e., to determine whether the two data sets for $f_u$ have not been drawn from the same distribution. In that case, $f_u$ is selected as an "active" feature. KS is preferred over parametric tests because one cannot assume a known distribution of data. The feature-selection algorithm is outlined as per Algorithm 2.

The set $\Theta_i$ includes the $N_{sf}$ features whose statistical properties are significantly altered by distortion $d_i$. The number of selected features depends on the expected dimension supported by $\Theta_i$ and on the particular architecture that the global descriptor will feed.

## IV. Computational-Intelligence Methods for Objective Quality Assessment

The methodology presented in this paper applies computational intelligence (CI) methods to both distortion detection and quality prediction. The basic advantage of this approach is the ability of CI-based models to deal with high-dimensional data characterized by complex perceptual relationships, especially when an explicit mathematical model of the underlying phenomenon is not available. The following sections overview the paradigms used in the system for classification and regression, namely, support vector machines (SVMs) [39] and circular back-propagation (CBP) [40] feed-forward neural networks.

### A. Support Vector Machines for Classification

Support vector machines are powerful tools to solve binary classification tasks. Given a set, $Z = \{(\mathbf{x}_l, y_l); l = 1, ..., n_p; \mathbf{x}_l \in X \subseteq \mathbb{R}^q; y_l \in Y = \{-1, +1\}\}$, of $n_p$ patterns, an SVM maps the input domain, $X \subseteq \mathbb{R}^q$, into a Hilbert space $F \subseteq \mathbb{R}^m$ ($q << m << \infty$), where a linear class separation is feasible. SVMs rely on the solution of the following quadratic programming problem, to find the optimal hyper-plane $\mathbf{w}$ separating the two classes:

$$\min_{\alpha} \left\{ \frac{1}{2} \sum_{l,m=1}^{n_p} \alpha_l \alpha_m y_l y_m K(\mathbf{x}_l, \mathbf{x}_m) - \sum_{l=1}^{n_p} \alpha_l \right\}$$
$$\text{subject to} \begin{cases} 0 \le \alpha_l \le C \quad \forall l \\ \sum_{l=1}^{n_p} y_l \alpha_l = 0 \end{cases} \tag{2}$$

where $\alpha_l$ are the SVM parameters setting the class-separating surface and the scalar quantity $C$ is a fixed regularization term that rules the tradeoff between accuracy and complexity. Problem setting (2) has the crucial advantage of involving a quadratic-optimization problem with linear constraints, ensuring that the solution is unique.

The kernel function $K()$ supports inner products among patterns in a higher dimensional, transformed space, without

**Algorithm 2** *KS based feature selection*

Inputs: a set $\Psi$ of undistorted images, a distortion $d_i$, a set of features $\Phi$

1. Data set construction
   For each distortion level ($r = 1, ..., n_r$):
   a. Create two sets, $\Psi_1 \cap \Psi_2 = \varnothing$, by randomly extracting $n_d \leq n_p/2$ images from $\Psi$.
   b. Apply $d_i$ to $\Psi_2$ to obtain: $\overline{\Psi}_2^{(r)} = \left\{ \bar{I}^{(s,r)}; \forall I^{(s)} \in \Psi_2 \right\}$.
   c. Compute each feature $f_u \in \phi$ for all images, and generate the sets $\Lambda_{1u}$, $\bar{\Lambda}_{2u}^{(r)}$: $\Lambda_{1u} = \left\{ x_{ub}; \forall I \in \Psi_1; b \in \mathrm{B} \right\}$ $\bar{\Lambda}_{2u}^{(r)} = \left\{ x_{ub}^{(r)}; \forall \bar{I} \in \overline{\Psi}_2^{(r)}; b \in \mathrm{B} \right\}$; $u$=1, ..., $n_f$; where B is the set of blocks extracted from an image $I$.
   d. Normalize the elements of $\Lambda_{1u}$ and $\bar{\Lambda}_{2u}^{(r)}$ into $[-1, 1]$.
2. Kolmogorov-Smirnov Test
   Assemble a probability vector, $p$, defined as: $p[u, r] = p_{\mathrm{KS}}(\Lambda_{1u}, \bar{\Lambda}_{2u}^{(r)})$; $u = 1, \ldots, n_f$; $r = 1, \ldots, n_r$; where $p_{\mathrm{KS}}(\cdot,\cdot)$ is the significance result of the KS test for the null hypothesis that the data sets $\Lambda_{1u}$ and $\bar{\Lambda}_{2u}^{(r)}$ have been drawn from the same distribution.
3. Feature ranking
   a. Set a reference confidence threshold, e.g., $p^* = 0.1$
   b. Compute the indicator vector, $t$, as: $t[u, r] = \begin{cases} 1 & p[u, r] \leq p^* \\ 0 & p[u, r] > p^* \end{cases}$; $u = 1, \ldots, n_f$; $r = 1, \ldots, n_r$;
   c. Assemble the occurrence vector, ($o$), whose $u$th element counts, over all possible distortion levels $n_r$, the event "the data sets $\Lambda_{1u}$ and $\bar{\Lambda}_{2u}^{(r)}$ are not drawn from the same distribution": $o[u] = \sum_{r=1}^{n_r} t[u, i]$; $u$=1, ..., $n_f$ ($0 \leq o[u] \leq n_r$)
4. Feature selection
   Set a number of features $N_{sf}$ to be included in the final feature set $\Theta_i \subset \Phi$, to be used to predict the effects of distortion $d_i$. Include in $\Theta_i$ the $N_{sf}$ features $f_u$, for which the highest values of $o[u]$ are obtained: $f_u \in \Theta \Leftrightarrow u = \arg\max o[u]$;

Output: a set $\Theta_i \subset \Phi$ of active features

requiring the specific mapping of each pattern. The radial basis function kernel, formulated as $K(\mathbf{x}_1, \mathbf{x}_2) = e^{-\|\mathbf{x}_1 - \mathbf{x}_2\|^2/\sigma^2}$, is a popular, effective choice and is used in this paper.

### B. Circular Back Propagation Networks for Quality Prediction

The multilayer perceptron (MLP) model [41] aims at implementing a stimulus-response behavior by arranging several elementary neurons into a layered network, which supports a unidirectional flow of information. Each neuron involves a nonlinear transformation of weighted inputs; theory proves that feed-forward networks embedding a sigmoidal nonlinearity can support arbitrary mappings [42], [43].

CBP networks [40] extend conventional MLPs with an additional input, computed as the sum of the squared values of all the network inputs. In the CBP architecture, a set of $n_i$ input values (as many as input features) connect to each of the $n_h$ "hidden" neurons. The estimation process supported by the network is expressed as

$$y_{\mathrm{CBP}}(\mathbf{x}) = sigm\left(w_0' + \sum_{u=1}^{n_h}\left[w_u' \cdot sigm\left(w_{u,0} + \sum_{k=1}^{d} w_{u,k}x_k + w_{u,d+1}\|\mathbf{x}\|^2\right)\right]\right). \quad (3)$$

The degrees of freedom of the nonlinear estimator (3) that must be fitted are the depth, nh, of the series expansion and the weighting coefficients within each neuron. The literature provides both theoretical [44] and practical criteria [45] to ensure prediction accuracy, while minimizing the risk of overfitting training data. This paper followed a practical, empirical approach [45] mainly because of its simplicity and proved effectiveness.

## V. Practical Implementation

This section describes the implementation of the RR system in practice. It is worth stressing that, thanks to the intrinsic flexibility of the system, each of its sub-modules may be implemented in a different way, not necessarily using computational intelligence methods. For the sake of coherence and without loss of generality, however, a uniform-paradigm for treating the two problems is proposed here.

### A. Numerical Description

Feature extraction is the first step in the proposed system (see Fig. 1). Both luminance and hue information of the original image $I$ and the received image $\bar{I}$ are used. The procedure outlined in Section III applies separately to: 1) the Y component, yielding the luminance descriptor $\mathbf{y}_{RR} = [\mathbf{y}, \bar{\mathbf{y}}]$ in the YCbCr color space, and 2) the hue component in the HSV color space, resulting in the hue descriptor, $\mathbf{h}_{RR} = [\mathbf{h}, \bar{\mathbf{h}}]$.

For calculating the features, the image is split in blocks of $32 \times 32$ pixels. For each block the features described in Table I are calculated from the color correlogram. For a given feature $f_u \in \Phi$, the objective vectors $\boldsymbol{y}_{u,RR}$ and $\boldsymbol{h}_{u,RR}$ are constructed by assembling six percentiles $\alpha = \{0, 20, 40, 60, 80, 100\}$ of the distributions $Y_u$ and $H_u$, respectively, over all blocks. As a result, the global descriptive vector for each feature spans a

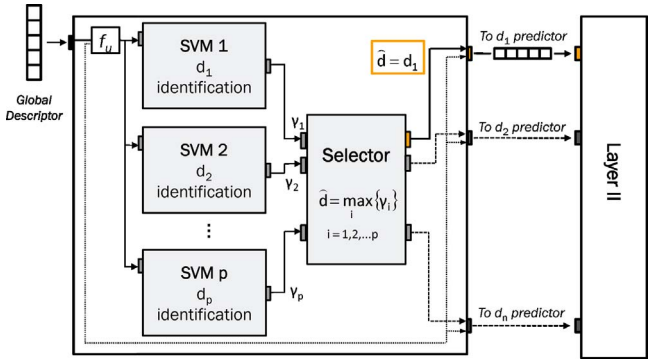


Fig. 5. Layer I implementation for distortion kind identification.

6-dimensional space. Eventually, both the original image and the received image are characterized by:

1) a luminance descriptor $\mathbf{y}$ ($\bar{\mathbf{y}}$ for $\bar{I}$) that is sized six times the number of features used to describe the Y component;
2) a hue descriptor $\mathbf{h}$ ($\bar{\mathbf{h}}$ for $\bar{I}$) that is sized six times the number of features used to describe the hue component.

Hence, the original image is described by a luminance descriptor $\mathbf{y}$ and a hue descriptor $\mathbf{h}$ that include at most 30 numerical quantities (six percentiles for each of the five features $f_u \in \Phi$). Under the hypothesis of adopting a conventional 32 bit representation for those quantities, the two vectors would result in a 240 byte packet of data to be sent as metadata.

### *B. First Layer; Distortion Identification*

The distortion identification task implies a multiclass problem spanning the set, $D = \{d_1, .., d_p\}$, of the distortions of interest. The goal of the eventual classification engine is to identify the distortion affecting the input image, thus redirecting the input to the dedicated quality predictor. The SVM model inherently solves a two-class problem; hence the multiclass problem is tackled by exploiting a conventional strategy [39]. The overall system exploits $p$ one-versus-rest classifiers, i.e., classifiers that are designed to solve the binary problem "one distortion versus the others." Then, given an input image, the system assigns such image to a class (i.e., a distortion) according to the following rules:

1) if one classifier ascribes the input image to a distortion, the image is assigned accordingly;
2) otherwise, the image is assigned according to the classifier whose decision function, $\gamma_i$, is highest (see [39] for details).

Fig. 5 presents a sketch of the overall layer architecture for distortion classification. The input of the distortion identifier layer (and therefore, of every SVM submodule in it) are the data included in the luminance descriptor $\mathbf{y}_{u,RR}$ and the hue descriptor $\mathbf{h}_{u,RR}$.

### *C. Second Layer: CBP Ensembles for Predicting the Loss/Gain in Quality*

The quality prediction layer includes as many elements as the number of distortions. Each element $\Omega(d_i)$ is implemented by an ensemble of several CBP networks, which are trained to quantify the effect of the specific distortion $d_i$ on quality. An "ensemble" [46] gathers parallel networks [46], [47] that have been trained on the same problem. The use of ensembles is justified by the statistical fluctuations in the empirical training set, which give rise to the problem of getting robust estimators. Indeed, by averaging the predictions of $\sigma$ independent estimators one reduces the variance $\sigma$, brought about by statistical noise in training data. Ideally, the decrease in variance reduces to $\bar{\sigma}^2 = \sigma^2/N$.

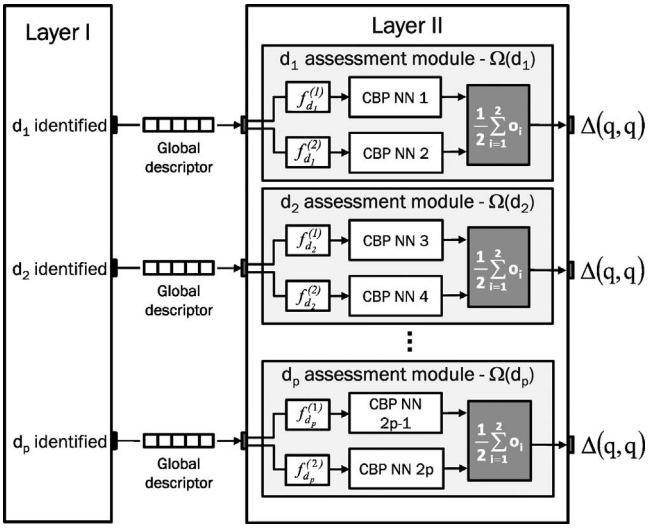


Fig. 6. Ensemble-based implementation of layer II.

Building independent estimators is crucial: when the number of patterns is small as compared to the data dimensionality (as is the case in quality estimation), an approach based on the theory of receptive fields [41] can be applied.

In the current implementation (Fig. 6), the $i$th ensemble, predicting the quality degradation for distortion $d_i$, involves a pair of features; thus, it includes a pair of CBP networks. The features should be as much informative as possible about the effect of the specific distortion on visual quality. As a consequence, the feature selection is performed independently for each distortion, since the ensembles of CBP networks are independent of one another. The eight neural networks employed share the same architecture and always use a sigmoid as a neuron-activation function. The input layer has the dimensionality of vector $\mathbf{y}_{u,RR}$. A single hidden layer is used, mostly to avoid the risk of over-fitting. Finally, one output neuron provides the objective estimate of gain/loss in quality of the distorted sample with respect to the original one.

## VI. Experimental Results

The second release of the LIVE database [35], including images distorted with JPEG 2000 Compression, White Noise, Gaussian Blur and JPEG compression is used as a test bed for the performance evaluation of the proposed system. As a consequence, the distortion classification layer contains four classifiers, and the quality assessment system eventually includes four dedicated assessment modules. The performance evaluation discusses: 1) the benefits of exploiting the two layer

TABLE II
OUTCOMES OF THE FEATURE SELECTION PROCEDURE FOR LAYER II PREDICTORS

| | JP2K | WN | GB | JPEG |
|---|---|---|---|---|
| Y | Entropy | Entropy | Entropy | Entropy |
| | Homog. | Contrast | Homog. | Homog. |
| H | Homog. | Contrast | Entropy | Diag. energy |
| | Contrast | En. ratio | Homog. | Entropy |

TABLE III
COMPOSITION OF THE FIVE TESTING FOLDS USED FOR THE CONTENT-BASED K-FOLD VALIDATION OF THE PROPOSED SYSTEM

| | Image Contents |
|---|---|
| Fold 1 | Bikes, Cemetry, House, Ocean, Sailing1, Stream |
| Fold 2 | Building2, ChurchAndCapitol, Lighthouse, PaintedHouse, Sailing2, StudentSculpture |
| Fold 3 | Buildings, CoinsInFountain, Lighthouse2, Parrots, Sailing3, Woman |
| Fold 4 | Caps, Dancers, ManFishing, Plane, Sailing4, WomanHat, |
| Fold 5 | CarnivalDolls, FlowersOnIh35, Monarch, Rapids, Statue |

structure, and 2) the single and joint contribution of luminance and hue information in the overall quality assessment. Experimental results are presented as follows. First, the overall experimental setup is outlined. Second, details concerning the setup of the layer I are presented. Third, the development of the distortion-oriented assessment modules is discussed. Finally, the performance of the eventual quality assessment system combining the two layers is analyzed.

### A. Experimental Setup

To evaluate the single and the joint contribution of luminance and hue information, the development of both the layer I and the layer II involved a separate analysis of the performance obtained by exploiting the luminance component $\mathbf{y}_{RR}$ and the Hue component $\mathbf{h}_{RR}$ of the global descriptor. The color correlogram based on the luminance component $\mathbf{y}_{RR}$ and on the hue component $\mathbf{h}_{RR}$ was computed for distance $k = 1$ using the L1 norm. The feature selection procedure presented in Section III-C was exploited to define the set of features $\Theta_{\mathrm{I}}$ and $\Theta_{\mathrm{II},d}$ needed by layer I and by each predictor $\Omega(d_i)$ of layer II, respectively. The reference confidence threshold was set to $p^* = 0.1$. The number of features to be included in the input vectors was set to $n_{sf} = 1$ for layer I and to $n_{sf} = 2$ for layer II. Since hue and luminance information are consistently different, features were selected separately for the two color components. The features selected for layer I were entropy for $\mathbf{y}_{RR}$ and energy ratio for $\mathbf{h}_{RR}$. The sets of features to be used for layer II are reported in Table II.

A k-fold test strategy was used for the overall performance evaluation of the double-layer quality estimator, mainly because this method proves effective to obtain reliable results when few data are available [48]. Toward that end, the database of 29 original images was split into five "folds." In each of the five experimental runs, four of the five folds (i.e., 23 original images and their distorted versions) were assigned to the training set, while the remaining fold (i.e., the remaining six original images and their distorted versions) were assigned to the test set. In this way, each sub-module of the system was proven to be able to generalize independently from the specific set of image contents presented in the training phase. Table III reports the specific composition of each fold.

### B. Accuracy in Distortion Classification

Since four types of distortions were involved in the present experimental session, a distortion identification layer based on four one-versus-rest SVM-based classifiers has been used. Table IV gives, for each of the five runs of the $k$-fold procedure, the percentage of misclassified patterns for each of the four classifiers for both the system trained with luminance statistics and the one trained with hue information. Furthermore, for each run, the eventual hyper-parameters of the SVMs are reported.

A first significant outcome of Table IV is that hue information appears more effective at identifying the distortion in an image, as the classification errors are significantly lower than when using luminance information. In fact, the luminance-based classifier proves effective only in separating images with White Noise from the rest of the dataset. This in turn strengthens the hypothesis that hue information can effectively capture the distortion affecting an image. The classifiers trained on hue information give satisfactory results for all four distortion categories in each of the five runs. The performance of the test set deteriorates only in two cases: the JPEG2000-versus-rest classifier in Run #1 and the JPEG-versus-rest classifier in Run #5. In general, however, the one-versus-rest classifiers attain a suitable performance independently of the specific composition of the training set and the test set. The classifier for the distortion Blur and for the distortion JPEG are the most critical ones, while separating images affected by White Noise from the others appears to be the easiest task.

In summary, the hue distribution statistics are, for distortion identification purposes, more informative than the corresponding luminance statistics. Thus, in the eventual complete framework, the second layer, based on distortion-specific predictors, should be supported by a distortion identifier exploiting only the hue component $\mathbf{h}_{RR}$ of the global descriptor.

### C. Accuracy in Predicting Gain/Loss in Quality

In the proposed framework, layer II includes as many predictors as the number of distortions. Hence, each module in the second layer was trained to target one of the four distortions addressed. It should be noted that the setup of the quality predictors assumes a perfect classification of the distortions in the first layer. The effect of a non-perfect classification will be discussed later. The JPEG2000 module was trained for the JP2K(1) and JP2K(2) sets, including 82 and 87 patterns, respectively. Note that these two sets of images resulted from separate subjective experiments (see [12]), and therefore, their quality ratings had to be treated separately. Analogously, the JPEG module was trained for the sets JPEG(1) and JPEG(2), including 87 and 88 patterns, respectively. Finally, the White Noise and Gaussian Blur datasets each contained 145 patterns. For each image, a difference mean opinion score (DMOS),

TABLE IV
VALIDATION ERROR (PERCENTAGE OF MISCLASSIFIED PATTERNS) AND HYPERPARAMETERS SETTINGS FOR THE DISTORTION IDENTIFICATION SUB-SYSTEM

| Run | Luminance | | | | | | Hue | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JP2K | Noise | Blur | JPEG | $C$ | $\sigma$ | JP2K | Noise | Blur | JPEG | $C$ | $\sigma$ |
| #1 | 21.5 | 0.7 | 6.1 | 13.0 | $1.10^4$ | 0.3 | 9.2 | 1.5 | 0.0 | 3.8 | $1.10^6$ | 50 |
| #2 | 21.5 | 0.7 | 10.0 | 11.5 | $1.10^5$ | 0.3 | 0.0 | 0.7 | 2.3 | 2.3 | $1.10^5$ | 0.3 |
| #3 | 17.9 | 0.7 | 17.1 | 15.6 | $1.10^5$ | 0.3 | 1.5 | 0.0 | 3.9 | 4.6 | $1.10^5$ | 0.3 |
| #4 | 22.2 | 3.1 | 15.0 | 14.2 | $1.10^6$ | 0.3 | 0.0 | 0.0 | 4.7 | 4.7 | $1.10^5$ | 0.3 |
| #5 | 18.2 | 0.9 | 16.3 | 13.4 | $1.10^4$ | 0.3 | 0.0 | 1.9 | 1.9 | 9.6 | $1.10^6$ | 50 |

indicating the perceived difference in quality between the original and processed image, and ranging between [1, 100] was provided in the LIVE database. For computational reasons, these scores were further remapped for this paper into the range $[-1, +1]$.

To evaluate the single and joint contribution of luminance and hue information, the distortion-oriented predictors were separately trained with the luminance component $\mathbf{y}_{RR}$ and with the hue component $\mathbf{h}_{RR}$. The approach proposed by Widrow and Lehr [45] provided an effective and practical method for tackling the sensitive problem of sizing the neural networks, i.e., to set the number of neurons to be included in the hidden layer of CBP. That method aims to ensure that the available training data effectively drives the adjustment of the network coefficients. Thanks to the flexibility of the system, it was possible to design the hidden layer of the CBP networks specifically for each predictor. The networks employed for predicting the effect of Noise and JPEG 2000 compression on quality were equipped with a 3-neurons hidden layer, while those used for blurred images were equipped with five hidden neurons. For the quality prediction of JPEG compressed images two CBP NN with seven hidden neurons were used.

Fig. 7 compares the performance obtained with the luminance component only, the hue component only, and the complete descriptor including both $\mathbf{y}_{RR}$ and $\mathbf{h}_{RR}$. Note that the performance is based on the average result of the 5-fold procedure (i.e., over the five runs, each on six test images, which were fully different in content from the images used for training); Appendix A reports the detailed results obtained in each run. The evaluation involves three descriptors, which measure the discrepancy between the change in quality, $\Delta(q, \bar{q})$, predicted in an objective manner, and the actual change in quality provided by the LIVE database. The three evaluation quantities are as follows:

1) Pearson's correlation coefficient, $\rho$;
2) Spearman's rank order correlation coefficient (SROCC);
3) mean percentage prediction error, $\%\mu_{|\mathrm{err}|}$, where $\%\mu_{|\mathrm{err}|}$ is the value of the absolute prediction error between $\bar{q}$ and $q$.

The first two indicators follow the VQEG recommendation [28]; the last one instead is particularly suited to evaluate the systems' generalization ability. Fig. 7 shows how the second layer is able to produce satisfactory quality predictions. The approach based on luminance information in general performs better than the one using hue information, although their performance is comparable for the prediction of the quality of JPEG compressed images. The quality prediction is worst for the JPEG 2000 compressed images, especially when using hue information. The corresponding mean percentage prediction error, however, is relatively low, especially for the luminance-based system, and can be considered acceptable for many applications.

When combining hue and luminance information, the performance of the system in general improves. For five out of the six datasets, the Pearson correlation increases and the mean percentage prediction error decreases with respect to systems based on hue or luminance information only. This outcome confirms the importance of including hue information in the quality prediction paradigm. Actually, only in the case of JPEG 2000 dataset 2 the Pearson correlation is slightly worse with respect to the luminance-only system. In the case of layer II therefore, the increase in complexity corresponding to the ensemble of the two systems is worth the cost, considering the general improvement.

### *D. Accuracy of the Combination of Distortion Classification and Quality Prediction*

The final step in the evaluation of the proposed framework is the analysis of the performance when the two layers are combined: layer I supports the distortion identification, thus redirecting the input image to the predictor that is entitled to evaluate the quality score. To this end, only one of the runs of the 5-fold procedure was used. Based on the partial results for layer I and layer II, run #3 was selected. Run #3 attained a satisfactory, but not the best performance. As such, it seems fair to estimate the overall performance of the proposed system based on this specific run.

For this final validation, the classification module based on hue only is chosen due to its effectiveness, as already discussed in Section VI-B. Conversely, the layer II predictors exploit both luminance and hue information, as a result of the performance analysis reported in section VI-C. For the JP2K and JPEG predictors, the final system includes those models guaranteeing higher performance: the model derived from set#1 for JP2K images, and the model derived from set#2 for JPEG images. All the test images entered layer I, and then followed the path indicated by the distortion identifier. The confusion matrix in Table V reports a suitable performance of layer I, since the number of misclassified patterns is limited to three. In particular, the confusion matrix shows that two images affected by JPEG distortion are classified as "Blur," while one image affected by JPEG2000 distortion is classified

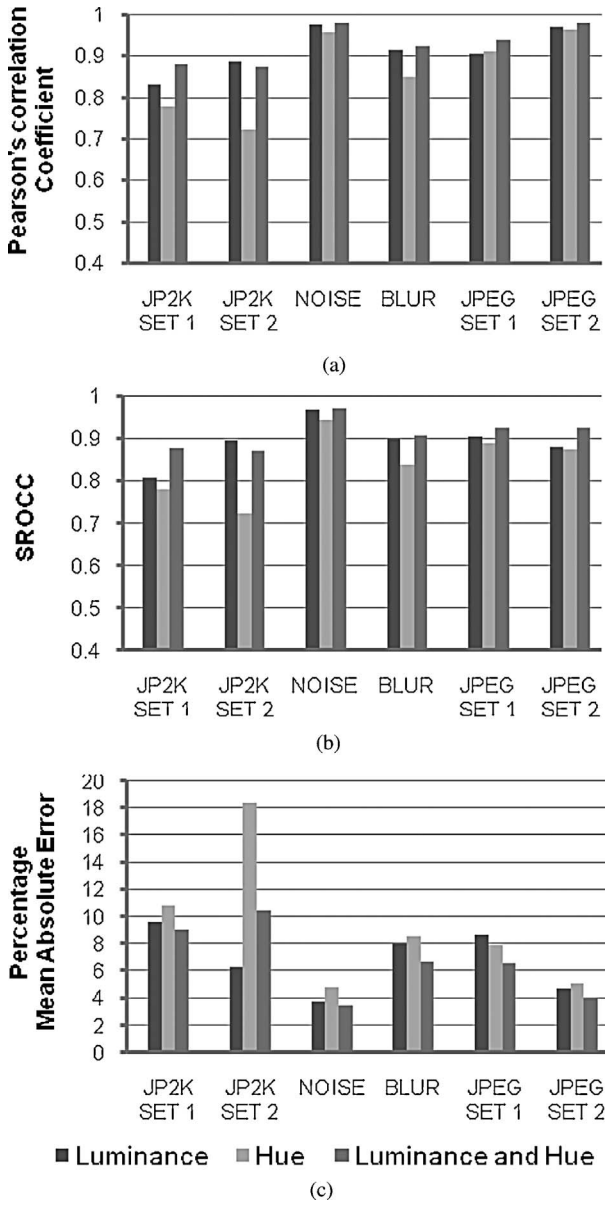


Fig. 7. Prediction accuracy for the LIVE DMOS. Results are reported for luminance only, hue only, and luminance and hue combined. (a) Pearson's correlation coefficient. (b) Spearman's correlation coefficient (SROCC). (c) Percentage mean absolute error.

as "Noise." As a major consequence, those images are assigned to the wrong quality predictor.

The consequent performance of the overall system is presented in Fig. 8. It compares the performance of the final quality predictor combining the two layers with: 1) the performance of the ideal system (i.e., the quality predictor that exploits a layer I not affected by misclassified patterns), and 2) the average performance of the quality predictors over the five folds. The three charts show that the performance attained by the overall system is quite satisfactory, and the misclassifications by layer I do not affect the performance significantly. As expected, the performance for the Blur distorted images slightly deteriorates, because two JPEG images are erroneously processed with the quality predictor designed to deal with the Blur distortion. Analogously, the misclassification of a JPEG2000 image causes an increase in the mean percentage prediction error for the Noise distorted images.

TABLE V
CONFUSION MATRIX FOR THE DISTORTION IDENTIFIER (LAYER I) USING RUN #3

| | | *Classified as* | | | |
|---|---|---|---|---|---|
| | | JPEG | JP2K | Noise | Blur |
| *Original* | JPEG | 30 | 0 | 0 | 2 |
| | JP2K | 0 | 35 | 1 | 0 |
| | Noise | 0 | 0 | 30 | 0 |
| | Blur | 0 | 0 | 0 | 30 |

TABLE VI
COMPARISON OF PREDICTION ACCURACY (PEARSON'S CORRELATION COEFFICIENT) BETWEEN EXISTING RR AND FR METHODS ON THE LIVE DATABASE

| | JP2K1 | JP2K2 | Noise | Blur | JPEG1 | JPEG2 |
|---|---|---|---|---|---|---|
| Proposed | 0.777 | 0.887 | 0.986 | 0.859 | 0.934 | 0.991 |
| Proposed (layer II) | 0.880 | 0.874 | 0.980 | 0.924 | 0.938 | 0.979 |
| Li and Wang | 0.948 | 0.965 | 0.965 | 0.956 | 0.820 | 0.957 |
| Wang *et al.* | 0.935 | 0.949 | 0.889 | 0.887 | 0.845 | 0.969 |
| Narwaria and Lin | 0.956 | | 0.989 | 0.951 | 0.947 | |
| MSSIM | 0.970 | 0.971 | 0.974 | 0.949 | 0.970 | 0.988 |
| PSNR | 0.933 | 0.874 | 0.985 | 0.783 | 0.886 | 0.916 |

Table VI finally compares the proposed approach with other image quality assessment methodologies. Three general purpose reduced-reference systems are considered: those proposed by Wang *et al.* [21], by Li and Wang [23] and by Narwaria and Lin [34]; the latter one recently presented a quality assessment framework based on machine learning technologies. Additionally, for reference purposes, also the well-known full-reference metrics MSSIM [49] and PSNR are given; although, one should take into account that the comparison of FR with RR metrics is not completely fair. The comparison is based on the Pearson's correlation coefficient. The data included in the table are: on the first row the prediction accuracy of the complete system (see Fig. 8), on the second row the average prediction accuracy of the ideal system (see Figs. 7 and 8), and in the following rows the metrics mentioned above. It should be noted that the work by Narwaria and Lin [34] does not give separate values of prediction accuracy for the two different datasets for the JPEG2000 and the JPEG distortions. As such, the table reports only one value of the Pearson's correlation coefficient for each distortion.

The table shows that the proposed method compares favorably with the alternative RR metrics, with the exception of JPEG2000 distorted images. One should notice, however, that the estimated performances of the proposed methodology appear quite robust, as they were estimated on five independent runs involving different compositions of the training and the test set. By contrast, the comparison RR methods measure prediction accuracy after applying a non-linear regression fit involving the entire image set LIVE DMOSs, hence they lack a test on the effectiveness of the regression function on images not included in the training process.

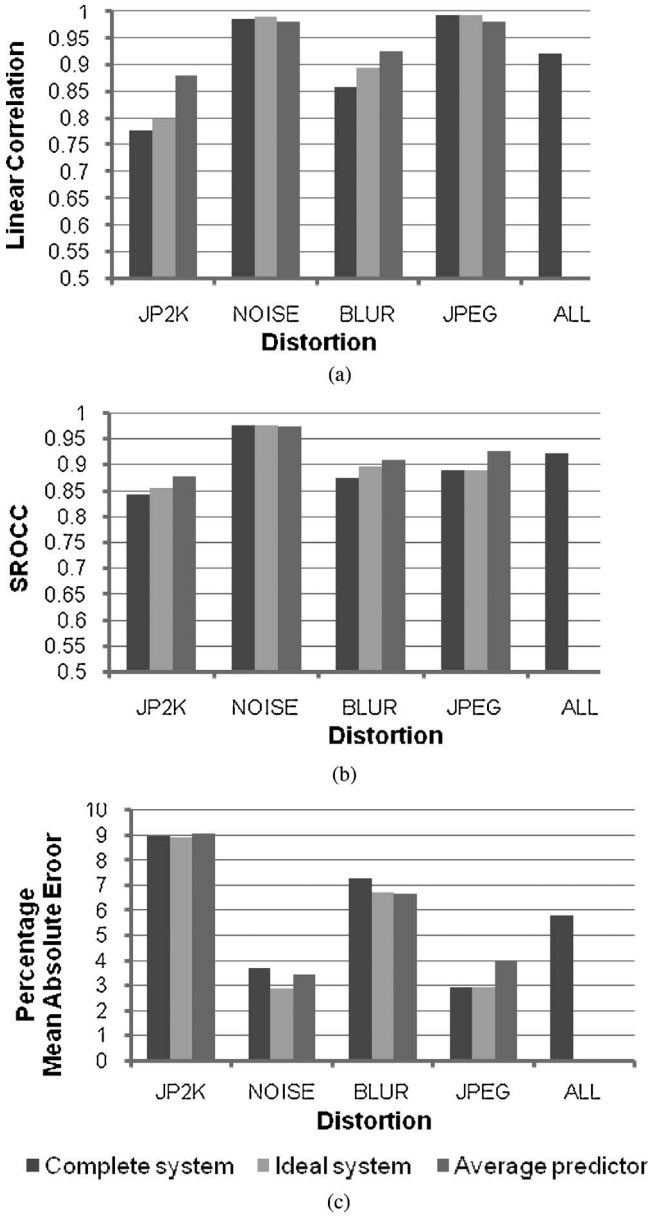


Fig. 8. Performance of the overall system evaluated with Run #3. The charts compare the results attained by the overall system with those attained by layer II on the same run only (i.e., an ideal system with no misclassification in layer I) and the average results attained by layer II over the five folds. (a) Pearson's correlation coefficient. (b) SROCC. (c) Percentage mean absolute error.

## VII. Conclusion

A double-layer reduced-reference system for image quality assessment was proposed. The system exploited the modeling power of computational intelligence methods to map color distribution information into a numerical expression for the perceived quality of images. The color distribution analysis took into account both luminance and hue information, granting an effective and small-sized description of the impact of distortions on the original color distribution.

The double layer framework consisted first in the identification of the distortion affecting the sample, and then the quantification of the degradation in quality by means of a specialized predictor. Such architecture was also easily expandable. The required actions needed for the system to deal with a new kind of distortion, were simply: 1) training a single SVM to distinguish the new distortion from the others, to be added to layer I, and 2) the design of a metric able to quantify the effect of the new distortion on quality, to be included in layer II. The proposed framework can also be efficiently implemented in hardware and embedded in real-life applications (i.e., consumer electronic devices).

Although its performance was satisfactory, the system can be consistently improved, both in terms of effectiveness and applicability. Future work might indeed be focused on two fronts: 1) the accuracy, with particular interest in including new distortions and consolidating the existing metrics, and 2) the usability, making the system more responding to real-life application problems.

To address the first issue, it would be of great interest to modify both layers in order to enable handling the co-presence of distortions or artifacts. Focusing on predicting the annoyance of single artifacts, in particular, would be of great beneficial, allowing the abstraction of their perceptual impact from the actual distortion producing them. Further effort should be then put in understanding how to combine the impact of different artifacts in estimating the overall quality of the picture. Needless to say, intensive subjective studies were required to make this development concrete.

The second issue mostly concerned the extension of the proposed framework to video quality measurement. In its current configuration, the system was not able to deal with video signals. An update of the system in that sense would require the introduction of the analysis of time-related distortions (e.g., packet loss, video freezing), together with the management of typical video perceptual issues, such as temporal masking of distortions. Another aspect to be addressed would be ability of generating continuous-time quality ratings. In principle, one might feed the system with the feature values continuously extracted from each sequence frame. However, known mechanisms specific for human perception should also be considered, such as the assessor's response time, masking phenomena and recency effects. As already shown in the literature [31], those mechanisms can be as well modeled by computational intelligence tools, if adequately taken into account when designing the learning procedure.

## Appendix A
## Detailed Results for Layer II

TABLE VII
Layer II Accuracy in Assessing the Quality of Image Compressed with JPEG2000 Algorithm, Set 1

| JP2K | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| SET 1 | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.970 | 5.67 | 0.729 | 12.206 | 0.904 | 8.767 |
| Run #2 | 0.621 | 14.32 | 0.915 | 8.468 | 0.891 | 10.268 |
| Run #3 | 0.753 | 11.62 | 0.749 | 8.958 | 0.797 | 8.881 |
| Run #4 | 0.934 | 6.70 | 0.698 | 13.854 | 0.888 | 8.337 |
| Run #5 | 0.877 | 9.65 | 0.802 | 10.621 | 0.920 | 8.908 |
| **Avg** | **0.831** | **9.59** | **0.779** | **10.822** | **0.880** | **9.032** |

TABLE VIII

LAYER II ACCURACY IN ASSESSING THE QUALITY OF IMAGE COMPRESSED WITH JPEG2000 ALGORITHM, SET 2

| JP2K | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| SET 2 | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.942 | 5.098 | 0.722 | 17.83 | 0.870 | 9.252 |
| Run #2 | 0.730 | 10.380 | 0.743 | 20.563 | 0.826 | 14.044 |
| Run #3 | 0.942 | 5.732 | 0.750 | 12.499 | 0.887 | 8.142 |
| Run #4 | 0.966 | 3.307 | 0.652 | 26.750 | 0.906 | 12.939 |
| Run #5 | 0.847 | 7.131 | 0.748 | 14.012 | 0.882 | 7.689 |
| **Avg** | **0.885** | **6.330** | **0.723** | **18.327** | **0.874** | **10.413** |

TABLE IX

LAYER II ACCURACY IN ASSESSING THE QUALITY OF IMAGE DISTORTED WITH WHITE NOISE

| NOISE | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.981 | 3.119 | 0.981 | 3.924 | 0.987 | 2.948 |
| Run #2 | 0.937 | 5.216 | 0.961 | 4.723 | 0.958 | 4.487 |
| Run #3 | 0.989 | 2.677 | 0.962 | 4.447 | 0.989 | 2.867 |
| Run #4 | 0.985 | 3.122 | 0.921 | 5.476 | 0.980 | 2.908 |
| Run #5 | 0.976 | 4.656 | 0.958 | 5.379 | 0.984 | 3.988 |
| **Avg** | **0.974** | **3.758** | **0.957** | **4.790** | **0.980** | **3.439** |

TABLE X

LAYER II ACCURACY IN ASSESSING THE QUALITY OF IMAGE DISTORTED WITH GAUSSIAN BLUR

| BLUR | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.923 | 5.410 | 0.913 | 6.670 | 0.935 | 5.092 |
| Run #2 | 0.952 | 6.469 | 0.938 | 6.580 | 0.965 | 4.927 |
| Run #3 | 0.957 | 6.023 | 0.744 | 10.843 | 0.893 | 6.678 |
| Run #4 | 0.832 | 10.490 | 0.823 | 9.532 | 0.903 | 8.424 |
| Run #5 | 0.901 | 11.593 | 0.824 | 9.135 | 0.921 | 8.255 |
| **Avg** | **0.913** | **7.997** | **0.848** | **8.552** | **0.924** | **6.675** |

TABLE XI

LAYER II ACCURACY IN ASSESSING THE QUALITY OF IMAGE COMPRESSED WITH JPEG ALGORITHM, SET 1

| JPEG | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| SET 1 | $\rho$ | $\%\mu_{\|err\|}$ | $P$ | $\%\mu_{\|err\|}$ | $P$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.938 | 7.674 | 0.925 | 9.731 | 0.964 | 6.529 |
| Run #2 | 0.748 | 13.248 | 0.904 | 7.659 | 0.872 | 9.914 |
| Run #3 | 0.932 | 5.395 | 0.934 | 5.441 | 0.939 | 4.920 |
| Run #4 | 0.951 | 5.985 | 0.820 | 9.764 | 0.940 | 4.729 |
| Run #5 | 0.955 | 11.211 | 0.967 | 6.906 | 0.976 | 7.054 |
| **Avg** | **0.905** | **8.703** | **0.910** | **7.900** | **0.938** | **6.629** |

TABLE XII

LAYER II ACCURACY IN ASSESSING THE QUALITY OF IMAGE COMPRESSED WITH JPEG ALGORITHM, SET 2

| JPEG | Luminance | | Hue | | L + H | |
|---|---|---|---|---|---|---|
| SET 2 | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ | $\rho$ | $\%\mu_{\|err\|}$ |
| Run #1 | 0.981 | 3.981 | 0.921 | 8.002 | 0.965 | 5.525 |
| Run #2 | 0.958 | 5.503 | 0.983 | 3.409 | 0.990 | 2.750 |
| Run #3 | 0.984 | 3.391 | 0.984 | 3.200 | 0.991 | 2.917 |
| Run #4 | 0.961 | 6.279 | 0.968 | 4.293 | 0.982 | 3.754 |
| Run #5 | 0.962 | 4.198 | 0.967 | 6.319 | 0.969 | 5.094 |
| **Avg** | **0.969** | **4.670** | **0.964** | **5.045** | **0.979** | **4.008** |

## REFERENCES


[1] *User Requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television*, ITU-T Rec. J.143, International Telecommunication Union, 2000.

[2] *Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference*, ITU-T Rec. J.144, International Telecommunication Union, 2004.

[3] *Full Reference and Reduced Reference Calibration Methods for Video Transmission Systems With Constant Misalignment of Spatial and Temporal Domains With Constant Gain and Offset*, ITU-T Rec. J.244, International Communication Union, 2008.

[4] *Perceptual Audiovisual Quality Measurement Techniques for Multimedia Services Over Digital Cable Television Networks in the Presence of a Reduced Bandwidth Reference*, ITU-T Rec. J.246, International Communication Union, 2008.

[5] *Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference*, ITU-T Rec. J.247, International Communication Union, 2008.

[6] *Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Reduced Reference*, ITU-T Rec. J.249, International Communication Union, 2010.

[7] *Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference*, ITU-R BT.1683 Rec., International Communication Union, 2004.

[8] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furth and O. Marques, Eds. Boca Raton, FL: CRC Press, 2003.

[9] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R BT.500 Rec., International Telecommunication Union, 1995.

[10] P. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Winchester, U.K.: Imcotek Press, 2000.

[11] B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction*. New York: Marcel Dekker, 2002.

[12] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[13] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008: A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, no. 1, pp. 30–45, 2009.

[14] G. Ginesu, F. Massidda, and D. D. Giusto, "A multi-factors approach for image quality assessment based on a human visual system model," *Signal Process. Image Commun.*, vol. 21, no. 4, pp. 316–333, 2006.

[15] Z. Gao and Y. F. Zheng, "Quality constrained compression using DWT-based image quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 910–922, Jul. 2008.

[16] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA: Morgan and Claypool, 2006.

[17] I. Marais and W. H. Steyn, "Robust defocus blur identification in the context of blind image quality assessment," *Signal Process. Image Commun.*, vol. 22, no. 2, pp. 833–844, 2007.

[18] Z. M. Parvezsazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Process. Image Commun.*, vol. 23, no. 4, pp. 257–268, 2008.

[19] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xua, "No-reference noticeable blockiness estimation in images," *Signal Process. Image Commun.*, vol. 23, no. 7, pp. 417–432, Jul. 2008.

[20] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[21] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[22] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Process. Image Commun.*, vol. 23, pp. 239–256, Apr. 2008.

[23] Q. Li and Z. Wang, "General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation," in *Proc. IEEE ICIP*, 2008, pp. 1192–1195.

[24] J. Redi, P. Gastaldo, R. Zunino, and I. Heynderickx, "Co-occurrence matrixes for the quality assessment of coded images," in *Proc. ICANN*, LNCS 5163. 2008, pp. 897–906.

[25] J. Redi, P. Gastaldo, R. Zunino, and I. Heynderickx, "Reduced reference assessment of perceived quality by exploiting color information," in *Proc. 4th Int. Workshop VPQM Consumer Electron.*, 2009 [Online]. Available: http://enpub.fulton.asu.edu/resp/vpqm/vpqm09/Proceedings_VPQM09/Papers/vpqmEPFL_FUB.pdf

[26] F. De Simone, F. Dufaux, T. Ebrahimi, C. Delogu, and V. Baroncini, "A subjective study of the influence of color information on visual quality assessment of high resolution pictures," in *Proc. 4th Int. Workshop VPQM Consumer Electron.*, 2009 [Online]. Available: http://enpub.fulton.asu.edu/resp/vpqm/vpqm09/Proceedings_VPQM09/Papers/VPQM2009_Redi.pdf

[27] J. Huang, S. Ravi Kumar, M. Mitra, W. J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Conf. CVPR*, Jun. 1997, pp. 762–768.

[28] VQEG. (2003). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment* [Online]. Available: http://www.vqeg.org

[29] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *J. Electron. Imaging*, vol. 14, no. 3, pp. 033004-1–033004-11, 2005.

[30] R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Process.*, vol. 87, no. 6, pp. 1493–1503, 2007.

[31] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316–1327, Sep. 2006.

[32] F.-H. Lin and R. M. Mersereau, "Rate-quality tradeoff MPEG video encoder," *Signal Process. Image Commun.*, vol. 14, no. 4, pp. 297–309, 1999.

[33] W. Lu, K. Zeng, D. Tao, Y. Yuan, and X. Gao, "No-reference image quality assessment in contourlet domain," *Neurocomputing*, vol. 73, nos. 4–6, pp. 784–794, 2010.

[34] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.

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

[36] M. Partio, B. Cramariuc, and M. Gabbouj, "An ordinal co-occurrence matrix framework for texture retrieval," *EURASIP J. Image Video Process.*, vol. 2007, no. 17358, p. 15, 2007.

[37] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[38] I. M. Chakravarti, R. G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*, vol. 1. New York: Wiley, 1967.

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

[40] S. Ridella, S. Rovetta, and R. Zunino, "Circular back-propagation networks for classification," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 84–97, Jan. 1997.

[41] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986.

[42] K. Hornik, M. Stinchcombe, and H. White, "Multilayered feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–66, 1989.

[43] B. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–14, 1989.

[44] E. B. Baum and H. David, "What size net gives valid generalization?" *Proc. Neural Comput.*, vol. 1, no. 1, pp. 151–60, 1989.

[45] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," *Proc. IEEE*, vol. 78, no. 9, pp. 1415–1442, Sep. 1990.

[46] M. Perrone, "Improving regression estimates: Averaging methods for variance reduction with extension to general convex measure optimization," Ph.D. dissertation, Dept. Phys., Brown Univ., Providence, RI, 1993.

[47] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, no. 1, pp. 1–48, 1992.

[48] P. Bartlett, S. Boucheron, and G. Lugosi, "Model selection and error estimation," *Mach. Learning*, vol. 48, nos. 1–3, pp. 85–113, 2001.

[49] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Nov. 2003, pp. 1398–1402.

**Judith A. Redi** received the M.S. degree in electronic engineering from Genoa University, Genoa, Italy, in 2006. In 2007, she joined the Smart Embedded Application Laboratory, University of Genoa, as a Ph.D. Student in Science and Technologies for Information and Knowledge. Her Ph.D. project focused on the study and modeling of visual quality perception mechanisms, and was developed in collaboration with Philips Research, AE Eindhoven, The Netherlands. She is currently an Assistant Professor with the Man-Machine Interaction Group of the Mediamatics Department, Delft University of Technology, Delft, The Netherlands.

Her main research interests include visual perception, visual quality assessment, machine learning, image forensics, and pattern recognition.

Dr. Redi received the Best Ph.D. Thesis Award from the University of Genoa for her Ph.D. project in 2010.

**Paolo Gastaldo** received the Laurea degree in electronic engineering and the Ph.D. degree in space sciences and engineering, both from the University of Genoa, Genoa, Italy, in 1998 and 2004, respectively.

Since 2004, he has been a Research Fellow with the Department of Biophysical and Electronic Engineering, University of Genoa. His main research interests include innovative systems for visual signal understanding, neural network-based methods for nonlinear information processing, text mining, and embedded electronics systems for advanced signal interpretation.

Dr. Gastaldo has been the Co-Chairman of two editions of the International Workshop on Computational Intelligence for Security in Information Systems (CISIS'08 and CISIS'09).

**Ingrid Heynderickx** received the Ph.D. degree in physics from the University of Antwerp, Antwerp, Belgium, in December 1986.

In 1987, she joined the Philips Research Laboratories, Eindhoven, The Netherlands, and meanwhile worked on different areas of research: optical design of displays, processing of liquid crystalline polymers, and functionality of personal care devices. Since 1999, she has been the Head of the Research Activities on Visual Perception of Display and Lighting Systems and in 2005 she was appointed a Research Fellow with the Group Visual Experiences. In 2005, she was appointed a Guest Research Professor with the Southeast University of Nanjing, Nanjing, China, and a Part-Time Full Professor with the University of Technology, Delft, The Netherlands.

Dr. Heynderickx is a member of the Society for Information Displays (SID), and for the SID, she was the Chairman of the Applied Vision Subcommittee from 2002 to 2007. In 2008, she became a Fellow of the SID and the Chairman of its European Program Committee.

**Rodolfo Zunino** (M'90) was born in Genoa, Italy, in 1961. He received the Laurea (*cum laude*) degree in electronic engineering from Genoa University, Genoa, Italy, in 1985.

From 1986 to 1995, he was a Research Consultant with the Department of Biophysical and Electronic Engineering (DIBE), Genoa University. He is currently an Associate Professor with DIBE, where he teaches electronics for embedded systems and electronics for security. He co-authored more than 170 scientific papers in international journals and conferences. His main scientific interests include intelligent systems for computer security, network security and critical infrastructure protection, embedded electronic systems for neural networks, efficient models for data representation and learning, massive-scale text-mining and text-clustering methods, and advanced techniques for multimedia data processing.

Mr. Zunino has been the Co-Chairman of the two editions of the International Workshop on Computational Intelligence for Security in Information Systems (CISIS'08 and CISIS'09). Since 2001, he has been an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS, and has participated in the scientific committees of several international events (ICANN'02, ICANN'09, IWPAAMS2004, IWPAAMS2005, Applied Computing 2006).