

# Toward a No-Reference Image Quality Assessment Using Statistics of Perceptual Color Descriptors

Dohyoung Lee, *Student Member, IEEE*, and Konstantinos N. Plataniotis, *Fellow, IEEE*

**Abstract**—Analysis of the statistical properties of natural images has played a vital role in the design of no-reference (NR) image quality assessment (IQA) techniques. In this paper, we propose parametric models describing the general characteristics of chromatic data in natural images. They provide informative cues for quantifying visual discomfort caused by the presence of chromatic image distortions. The established models capture the correlation of chromatic data between spatially adjacent pixels by means of color invariance descriptors. The use of color invariance descriptors is inspired by their relevance to visual perception, since they provide less sensitive descriptions of image scenes against viewing geometry and illumination variations than luminances. In order to approximate the visual quality perception of chromatic distortions, we devise four parametric models derived from invariance descriptors representing independent aspects of color perception: 1) hue; 2) saturation; 3) opponent angle; and 4) spherical angle. The practical utility of the proposed models is examined by deploying them in our new general-purpose NR IQA metric. The metric initially estimates the parameters of the proposed chromatic models from an input image to constitute a collection of quality-aware features (QAF). Thereafter, a machine learning technique is applied to predict visual quality given a set of extracted QAFs. Experimentation performed on large-scale image databases demonstrates that the proposed metric correlates well with the provided subjective ratings of image quality over commonly encountered achromatic and chromatic distortions, indicating that it can be deployed on a wide variety of color image processing problems as a generalized IQA solution.

**Index Terms**—No-reference (NR) image quality assessment (IQA), chromatic information, color invariance descriptors, circular distribution models.

## I. INTRODUCTION

QUALITY assessment of visual data plays a central role in various sectors of image processing technologies, e.g., the evaluation of image acquisition/reproduction systems, the quality monitoring of multimedia visual data and the optimization of visual processing algorithms. Although subjective assessment is the most reliable means to measure

Manuscript received January 11, 2016; revised April 29, 2016; accepted June 2, 2016. Date of publication June 9, 2016; date of current version June 28, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aljosa Smolic.

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: dohyoung.lee@utoronto.ca; kostas@ece.utoronto.ca).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The file presents the proof of invariant properties of color invariance descriptors discussed in the paper. In addition, the computation of circular kurtosis from the observed samples of circular data is presented, which is necessary for the extraction of quality-aware features. The total size of the file is 133 kB. Contact dohyoung.lee@utoronto.ca for further questions about this work.

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/TIP.2016.2579308

visual quality, it is not always feasible in practical applications, e.g., real-time quality control systems. Therefore, automated solutions that can precisely predict the subjective opinions of human observers, e.g., Mean Opinion Scores (MOS), have gained considerable research attentions in recent years. To this end, a plethora of *image quality assessment* (IQA) metrics has been proposed and successfully deployed in applications such as image/video coding [1], information hiding [2], image rendering [3], tone mapping [4] and color gamut mapping [5]. Depending on the availability of a reference image, existing IQA metrics can be classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) approaches [6]. Since the accessibility of reference data allows for the accurate prediction of perceived quality, there have been several well-established FR IQA metrics, such as the Structural SIMilarity (SSIM) index [7]. However, reference images are rarely available in real-world visual communication environments. Therefore, NR IQA metrics, which are more easily deployable in broader applications, have been considered as practical alternatives to FR IQA solutions.

Existing NR metrics can be categorized into *distortion-specific* and *general-purpose* approaches. Early NR metrics are developed to handle a pre-defined type of distortions, e.g., blurriness [8] or blockiness/ringing artifacts introduced by compression solutions [9]. Despite their simplicity, distortion-specific metrics have limited applicability as information about the distortion type is not always well defined for practical scenarios. Hence, general-purpose metrics which are designed to consistently work well across multiple classes of distortions have become pervasive recently. Almost all these metrics estimate perceived visual quality by quantifying the deviation of distorted image statistics from statistical models that majority of natural images<sup>1</sup> conform to, often referred to as *Natural Scene Statistics* (NSS). They generally consist of two processing stages: feature extraction followed by quality score prediction. Initially, general-purpose metrics extract *Quality-Aware Features* (QAF) from an input image, which are features conveying relevant information for inferring the perceived image quality. As shown in Table I, commonly employed QAFs in existing metrics are derived from the statistical distributions of image coefficients in frequency domain (e.g., wavelet or discrete cosine transforms) [10]–[12], normalized luminance coefficients [13], [14] or local contrast features (e.g., gradient magnitude) [15]. The majority of NR metrics,

<sup>1</sup>The “natural images” are the ones captured by an optical camera which can include scenes of natural indoor/outdoor environments, as well as man-made subjects taken under photopic or scotopic viewing conditions. They are pristine images that have not been subjected to artificial processing [10].

TABLE I

LIST OF THE REPRESENTATIVE NO-REFERENCE (NR) IMAGE QUALITY ASSESSMENT (IQA) METRICS IN LITERATURE.  
FOR DISTORTION-SPECIFIC METRICS, THE TARGETED DISTORTION TYPES ARE INDICATED IN PARENTHESIS

Class	Metric	Incorporated Features	Features to Score Mapping
Distortion specific NR metrics	Ferzli & Karam (sharpness) [8]	local contrast and edge width	probability summation
	Sheikh et al. (JPEG2000) [9]	statistics of wavelet coefficients	weighted linear combination
General purpose (distortion agnostic) NR metrics	DIIVINE [10]	statistics of subband coefficients obtained from steerable pyramid decomposition	Support vector regression (SVR)
	C-DIIVINE [12]	DIIVINE features + complex phase statistics	SVR
	BLIINDS-II [11]	local statistics of discrete cosine transform (DCT) coefficients	Bayesian prediction model
	BRISQUE [13]	statistics of mean subtracted contrast normalized (MSCN) coefficients	SVR
	GM-LOG [15]	statistics of gradient magnitude (GM) and Laplacian of Gaussian (LOG)	SVR
	CBIQ [16]	statistics of Gabor features	linear mapping / SVR
	NIQE [14]	statistics of mean subtracted contrast normalized (MSCN) coefficients	multivariate Gaussian model
	IL-NIQE [17]	NIQE features + statistics of log-Gabor and log opponent color features	multivariate Gaussian model
	FRIQUEE [18]	statistics of luminance + statistics of chrominance features from CIELAB, LMS and $\ell\alpha\beta$ color spaces	Deep belief net (DBN) / SVM

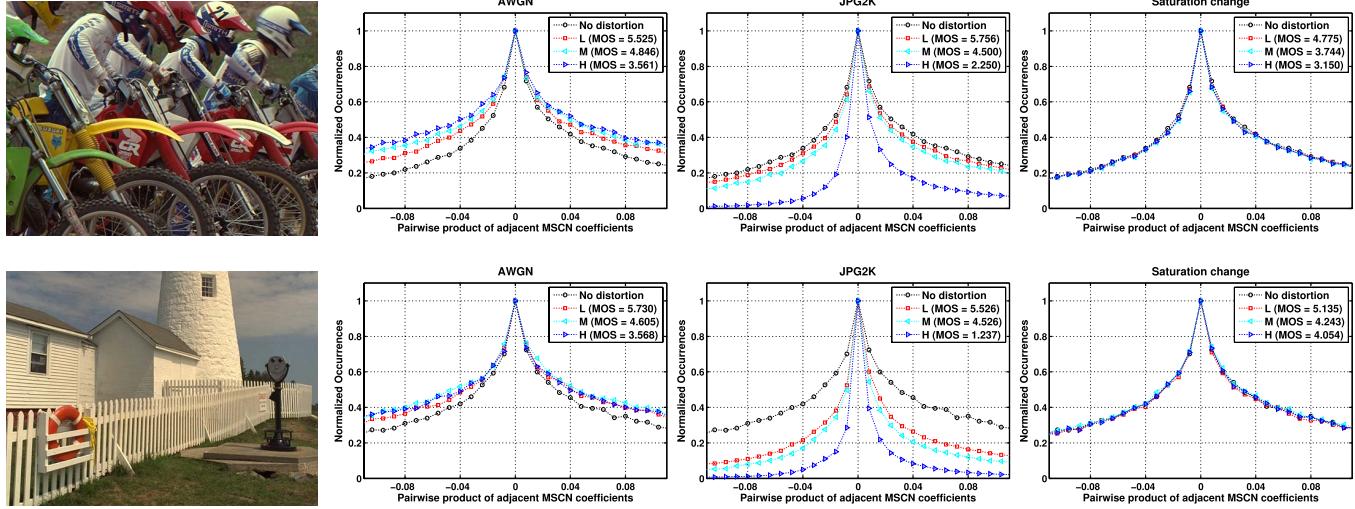


Fig. 1. The histograms of the pairwise products of adjacent MSCN coefficients for two images (“bike” and “lighthouse”) and their degraded versions from TID2013 DB [19], subjected to achromatic-structural distortions (Gaussian blur and JPEG2000 compression) and chromatic distortion (saturation change) with varying severities. “L”, “M” and “H” represent low, medium and high degrees of distortions, respectively. The MOS values indicate the subjective ratings of perceived image quality; low MOS being poor quality and high MOS being good quality.

then apply machine learning techniques to predict image quality from the extracted QAFs. The accuracy of a NR metric is largely determined by the relevance of utilized QAFs to visual quality perception; thus, identifying a representative set of QAFs that properly mimics the characteristics of the human visual system (HVS) is crucial for the design of NR metrics.

However, most NR metrics in literature only emphasize the statistical aspects of luminance in color images while disregarding ones of chromatic data. It is mainly because luminance has been considered as a dominant factor for understanding the human visual perception [6]. Nevertheless, NR IQA solely based on luminance can be suboptimal since it underestimates the visual disturbance caused by chromatic distortions unless they influence luminance statistics. Consider Fig. 1 which demonstrates the distributions of luminance-based features, known as the Mean Subtracted Contrast Normalized (MSCN) coefficients [13], for two pristine images and their degraded versions. The marginal histograms of the pairwise products of adjacent MSCN coefficients effectively capture distortion-relevant information for additive white Gaussian noise (AWGN) and JPEG2000 distortions. For instance, one can identify distortion types and quantify discomfort levels by analyzing the tail behavior of histograms (i.e., kurtosis) for such achromatic-structural distortions [13].

However, Fig. 1 clearly visualizes that their distribution patterns for a representative chromatic distortion, saturation change, are nearly identical to those of pristine images regardless of varying distortion levels. This suggests that existing luminance-based QAFs may not be reliable to handle distortions that primarily involve chromatic components.

Our previous work [20] highlighted the significance of chromatic data for FR IQA by exploiting local features of hue and chroma channels for quantifying the color similarities between reference and distorted images. In this work, we advocate the use of chromatic data for NR IQA since the intrinsic performance degradation of NR metrics (compared to FR metrics) due to the absence of reference data can be minimized by utilizing complementary chromatic information. To this end, we introduce a novel set of chromatic QAFs tailored for NR IQA tasks.

The proposed QAFs are parameters of four parametric models that characterize the statistical distributions of low-level color information of natural images. Two models are derived from the zeroth-order (RGB value-based) image representations that are highly relevant to color perception, i.e., *hue* and *saturation*, whereas the remaining two are derived from the first-order (derivative-based) representations, i.e., *opponent angle* and *spherical angle*. The aforementioned representations

encode independent descriptions of color images while sharing the common property that they are computed using *color invariance descriptors* [21], [22]. Contrary to luminances, these descriptors provide robust pixel-level descriptions of image scenes which are less influenced by *scene-accidental factors*, e.g., viewing geometry and illumination variations. Incorporation of them along with luminance allows for a closer approximation of human color perception as the HVS is naturally able to discount scene-accidental effects. Given that human eyes are highly adapted for extracting structural information from viewing field [7], we postulate that establishing statistical models relevant to the spatial correlation of robust color representation will lead us to develop an accurate NR metric. We make following contributions in this article:

- 1) We devise parametric models based on color invariance descriptors describing the general characteristics of chromatic data in natural images. The proposed models characterize the spatial dependence of chromatic data by analyzing the empirical distributions of differences between invariance descriptor values obtained from spatially adjacent pixels. We demonstrate that the parameters of established models provide reliable cues for quantifying visual discomfort caused by the presence of image distortions (mainly for chromatic ones). To our knowledge, there exists no previous studies that investigate the relationship between the image statistics of color invariance descriptors and the human judgment of image quality.
- 2) We propose a general-purpose NR IQA framework, namely the Invariance DDescriptor-based ALgorithm (IDEAL), that augments the parameters of the proposed models of color statistics with existing luminance-based QAFs. We experimentally validate that by exploiting chromatic data, the proposed metric accords well with the human perception of visual quality against a wide range of traditional and chromatic distortions commonly occurred in practical applications.

This manuscript is organized as follows. Section II presents preliminary color theories related to image formation and color invariance descriptors. Section III introduces new parametric models that represent the statistical characteristics of chromatic data in natural images using color invariance descriptors. Section IV demonstrates the proposed NR IQA metric using the derived statistical models. Section V is devoted to experimental results and conclusions are drawn in Section VI.

## II. COLOR INVARIANT REPRESENTATIONS

### A. Color Image Formation

Color conveys rich information that allows for the extraction of image descriptors providing more robust structural information (i.e., less influenced by viewing geometry and illumination variations) of image scenes than its luminance counterpart [23]. The *Dichromatic Reflection Model* (DRM) [24] provides a simple physical model to describe how recorded RGB values are affected by various viewing conditions. According to the DRM, the measured color value,  $C \in \{R, G, B\}$ , at the spatial coordinate  $\mathbf{x} \in \mathbb{R}^2$  is represented by a linear combination of the body reflection (intrinsic object

color) and the interface reflection (highlight) as follows:

$$C(\mathbf{x}) = m^b(\mathbf{x})b^C(\mathbf{x})e^C + m^i(\mathbf{x})e^C \quad (1)$$

where  $e$  is the color of the light source (assumed constant throughout the scene),  $b$  is the surface albedo,  $m^b$  and  $m^i$  are the scale factors of the body and the interface reflections, respectively, whose values are depending on the geometric structure at  $\mathbf{x}$ , e.g., camera viewpoint, light source direction, and object orientation. For matte surfaces, the model in (1) further simplifies by ignoring the second term as there is no interface reflection, i.e.,  $m^i = 0$ .

The spatial derivatives of the image provide information related to object boundaries. The spatial derivative of the RGB color values  $C_{\mathbf{x}}$  can be derived from (1) as follows [22]:

$$C_{\mathbf{x}}(\mathbf{x}) = m_{\mathbf{x}}^b(\mathbf{x})b^C(\mathbf{x})e^C + m^b(\mathbf{x})b_{\mathbf{x}}^C(\mathbf{x})e^C + m_{\mathbf{x}}^i(\mathbf{x})e^C \quad (2)$$

where the subscript  $\mathbf{x}$  is used to indicate spatial differentiation.

### B. Color Invariance Descriptors

The DRM in (1) has been applied in the derivation of various invariance descriptors that provide robust color representation of scene objects against scene-accidental effects [23]. For instance, invariance against highlights can be obtained by representations that cancel out the term  $m^i$ , while invariance against viewing geometry can be obtained by removing the term  $m^b$ . Invariance against illumination intensity, such as those caused by shadows and shading, can be achieved by canceling out the light intensity  $e$ .

We first present two color invariance descriptors based on the zeroth-order image representation (i.e., RGB value-based), the saturation  $S$  and the hue  $H$ . These descriptors are known to correlate well with how humans perceive the color information. Although there exist several different ways to define them from RGB values, we consider the ones that possess invariance properties in this work. The *saturation*  $S$ , an attribute of the purity of a color, is given by:

$$S = 1 - \left[ \frac{3 \times \min(R, G, B)}{R + G + B} \right] \quad (3)$$

The opponent color space is known to provide decorrelated effects for RGB color channels. Two opponent channels in the opponent color space [23] are computed by:  $opp_1 = (R - G)/\sqrt{2}$  and  $opp_2 = (R + G - 2B)/\sqrt{6}$ , where  $opp_1$  and  $opp_2$  corresponds to the red-green and the yellow-blue channels. The *hue*  $H$ , an attribute related to the dominant wavelength of color signal, is defined as follows [25]:

$$H = \tan^{-1}(opp_1/opp_2) = \tan^{-1} \left[ \frac{\sqrt{3}(R - G)}{R + G - 2B} \right] \quad (4)$$

Subsequently, we present two color invariance descriptors based on the first-order image representations (i.e., derivative-based) [22], the opponent angle  $O$  and the spherical angle  $S$ . The computation of these two angular descriptors requires the transformation of color derivatives  $C_{\mathbf{x}}$  to a color space which is uncorrelated with respect to photometric events [26]. The derivatives of the opponent colors are obtained by transforming color derivatives  $C_{\mathbf{x}}$  using an opponent color transformation as follows:  $opp_{1,\mathbf{x}} = (R_{\mathbf{x}} - G_{\mathbf{x}})/\sqrt{2}$  and  $opp_{2,\mathbf{x}} = (R_{\mathbf{x}} +$

TABLE II  
OVERVIEW OF INVARIANT PROPERTIES OF IMAGE DESCRIPTORS  
(INVARIANCE IS DENOTED BY +, WHEREAS SENSITIVITY  
TO THE IMAGING CONDITION IS DENOTED BY -)

Descriptor	Formula	Viewing geometry	Highlights	Illuminant intensity
$S$ (Saturation)	(3)	+	-	+
$H$ (Hue)	(4)	+	+	+
$O$ (Opponent angle)	(5)	-	+	+
$A$ (Spherical angle)	(6)	+	-	+
$Y$ (Luminance)	(8)	-	-	-

TABLE III  
DEFINITIONS OF TERMINOLOGIES

Term	Definition
<i>Descriptor</i>	an entity that provides pixel-wise color description of image data (e.g., luminance, hue, saturation, opponent/spherical angles)
<i>Feature</i>	an informative cue extracted from image data that allows for discriminating the type of image distortions present in image data or quantifying the severity of visual disturbances in image data.
<i>Model</i>	an entity embodying the general image characteristics shared by (undistorted) natural images that are violated in the presence of image distortions.

$G_x - 2B_x)/\sqrt{6}$ . The *opponent angle*  $O$  is defined as a ratio of  $opp_{1,x}$  and  $opp_{2,x}$  as follows:

$$O = \tan^{-1} (opp_{1,x}/opp_{2,x}) \quad (5)$$

The *spherical angle*  $A$  is defined as a ratio of image derivatives in the spherical color space as follows:

$$A = \tan^{-1} (sph_{1,x}/sph_{2,x}) \quad (6)$$

where the derivatives of spherical channels  $sph_{1,x}$  and  $sph_{2,x}$  are obtained by transforming color derivatives  $C_x$  using a spherical transformation [26]:

$$\begin{aligned} sph_{1,x} &= \frac{G_x R - R_x G}{\sqrt{R^2 + G^2}} \\ sph_{2,x} &= \frac{R_x RB + G_x GB - B_x R^2 - B_x G^2}{\sqrt{(R^2 + G^2)(R^2 + G^2 + B^2)}} \end{aligned} \quad (7)$$

Table II summarizes the invariant characteristics of the incorporated color invariance descriptors. Please refer to the *Supplementary Material* for the detailed proof of invariant properties. Note that the luminance  $Y$  is computed by:

$$Y = 0.299R + 0.587G + 0.114B \quad (8)$$

and is sensitive to aforementioned scene-accidental effects. The color invariance descriptors presented in this section were originally introduced for matching and retrieval applications, since their robustness against accidental imaging conditions allows for more accurate performance in vision tasks [22]. However, their effectiveness in visual quality prediction is largely uninvestigated, and it is the main focus of this paper.

For the convenience of discussion, we clarify three frequently used terminologies throughout the paper — *descriptor*, *feature*, and *model* — in Table III.

### III. STATISTICAL MODELING OF NATURAL IMAGES USING COLOR INVARIANT REPRESENTATIONS

The research on the statistical modeling of natural images hypothesizes that visual quality perception is highly adapted to NSS models conforming to general characteristics of natural images [27], [28]. Although such general characteristics have been analyzed in different image domains (e.g., frequency

or spatial), the spatial correlation between image pixels has shown to convey meaningful information for understanding the HVS properties [29]. In addition, spatial domain approaches have demonstrated their efficiency and accuracy in practical NR IQA metrics [13], [14]; hence, we mainly focus on spatial domain NSS models in this paper.

Natural images are highly structured in the sense that there are strong dependencies between luminance values of adjacent pixels [28]. The derivation of NSS models in spatial domain can be seen as an operation to eliminate spatial image redundancies and to reveal the contents-independent cues of perceived quality [13]. It has been shown that an operation of local mean subtraction followed by divisive variance normalization on image luminances, i.e., MSCN coefficients, has a spatial-decorrelating effect. The NSS models related to the spatial arrangement of MSCN coefficients have been exploited in existing NR metrics and shown their effectiveness in differentiating between traditional distortions and quantifying visual discomfort [13], [14]. However, as exemplified in Fig. 1, such luminance-based NSS models are suboptimal in dealing with chromatic distortions; hence, it necessitates complementary NSS models to characterize the statistics of chromatic components.

Only few works can be found in literature pertaining to the statistical modeling of chromatic data in natural images. Su et al. [30] proposed NSS models of the mean and standard deviation of the Gabor magnitude responses against spatial frequency and orientation using two chrominance components  $a^*$  and  $b^*$  of CIELAB color space. The authors exploited the proposed models into a stereoscopic vision application and demonstrated their effectiveness in predicting human binocular perception. However, these models are not spatial domain NSS models,<sup>2</sup> nor are they comprehensively validated on IQA of general color images. Ruderman et al. [31] showed that the distributions of logarithmic-scale opponent color response<sup>3</sup> of natural images can be well modelled using a Gaussian distribution. Recently, Ruderman's models have been deployed in the IL-NIQE [17], one of the few general-purpose NR metric explicitly making use of chromatic data. The authors confirmed that a marginal performance gain<sup>4</sup> is achieved by integrating Ruderman's models with luminance-based models.

In this section, we develop new parametric NSS models of chromatic data in natural images, providing reliable cues for IQA in the forms of model parameters. Four spatial domain NSS models are derived from color representations discounting viewing geometry and illumination variations; thus, are well correlated with how human perceive structural information from color images. The first two models capture the statistical properties of intuitive color attributes, *hue* and *saturation*; whereas the remaining two models capture the ones of color derivative information, *opponent angle* and *spherical*

<sup>2</sup>They rely on computationally expensive 2D Gabor transform.

<sup>3</sup>The use of logarithm was motivated by psychophysical principles (the Weber Fechner law). These responses are obtained in two stages. Initially, the mean subtracted logarithmic signals of individual RGB channels are obtained; followed by the opponent color transformation onto the resultant signals.

<sup>4</sup>The weighted-average SRCC over four image databases, e.g., TID2013 DB, LIVE DB, and CSIQ DB, are slightly increased from 0.6369 to 0.6448.

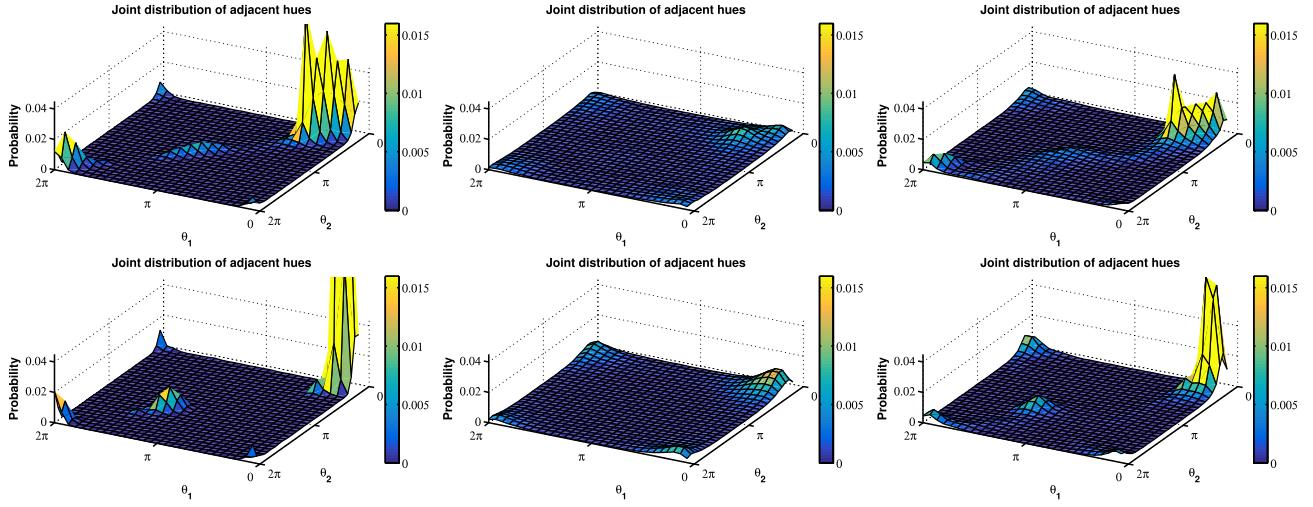


Fig. 2. The joint distributions of two hue values of adjacent pixels for pristine images in Fig. 1 and their degraded versions subjected to different distortions – from left to right: original, AWGN, and chromatic aberration. The upper and lower distributions correspond to “bike” and “lighthouse” images, respectively.

*angle.* The proposed spatial domain NSS models of chromatic data complement existing luminance models since fine details of image data are best detected in the luminance whereas the chromatic channels preserve coarse details [23]. Also, recall from Table II that each representations are systematically chosen to emphasize different physical aspects of color descriptions. We postulate that the structural information of image data captured by individual invariance descriptors has a distinctive influence on the perceived image quality. As will be discussed in Section V-C, combining QAFs from the proposed NSS models along with the ones from existing luminance-based NSS models allows for devising an effective NR metric that generalizes well against a wide variety of image contents and distortions.

#### A. Statistical Model for Hue Component

The hue plays a central role in IQA; visible hue distortions often result in unnatural/unrealistic reproduction of color data, having a strong visual impact [32]. Therefore, establishing a hue-based NSS model is important in applications involving adjustments of chromatic components, e.g., gamut mapping and tone mapping. In Fig. 2, the joint distributions of hue values of neighboring pixels for two natural images are demonstrated. The joint densities are highly concentrated in a main diagonal axis, indicating hue values of two adjacent pixels are highly correlated. With the presence of a specific image distortion, the joint hue distributions of two natural images are altered in a characteristic manner. For instance, additive white Gaussian noise (AWGN) produces more scattered distributions, while chromatic aberration introduces an asymmetry along the diagonal axis. These observations reveal that the spatial correlation of hues between neighbor pixels provides distortion-relevant information. In order to investigate general characteristics of hue data in terms of their spatial correlations, we consider the hue differences between adjacent pixels, i.e., *relative hue*.

A relative hue  $\Delta H$  at a spatial location  $(i, j)$  of a color image is defined as the *angular difference* of two adjacent hue

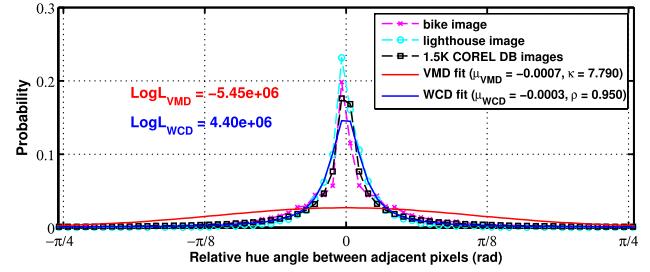


Fig. 3. The empirical probability histograms of relative hue for “bike” and “lighthouse” images in Fig. 1, as well as for 1.5K natural images from COREL DB. Two fitted circular distribution models for COREL DB images are also illustrated with the associated log likelihood values.

values. Given that the values of  $H$  lie in the range  $[0, 2\pi)$ , the relative hue in horizontal orientation<sup>5</sup> is defined as:

$$\Delta H^{ho}(i, j) = \Psi(H(i + 1, j), H(i, j)) \quad (9)$$

where the angular difference operator  $\Psi(\cdot)$  produces value in the range  $[-\pi, \pi]$ :

$$\Psi(\theta_1, \theta_2) = \begin{cases} \theta_1 - \theta_2, & -\pi \leq \theta_1 - \theta_2 \leq \pi \\ \theta_1 - \theta_2 + 2\pi, & -2\pi \leq \theta_1 - \theta_2 < -\pi \\ \theta_1 - \theta_2 - 2\pi, & \pi < \theta_1 - \theta_2 \leq 2\pi \end{cases} \quad (10)$$

Fig. 3 shows the empirical histograms of  $\Delta H$  samples for two pristine images in Fig. 1. To formulate more reliable model,  $\Delta H$  samples are obtained from 1.5K images from COREL DB,<sup>6</sup> a large collection of natural images containing a wide range of scene complexity and color variations.

As depicted in Fig. 3, the  $\Delta H$  histograms for natural images exhibit unimodal circular distributions; thus, can be modeled with *circular distribution models*. We examined the suitability of two representative circular distributions for fitting the  $\Delta H$

<sup>5</sup>For the sake of simplicity, only the horizontal difference is discussed here, but it can be generalized to all orientations.

<sup>6</sup>The COREL DB images used in our experiment are 24-bit RGB images with  $512 \times 768$  resolution. Incorporated categories include ancient architecture, beaches, castles, children, flora, forests, garden, lakes, Mediterranean, models, skies, statues, and sunset. For more descriptions on the COREL DB, see [33].

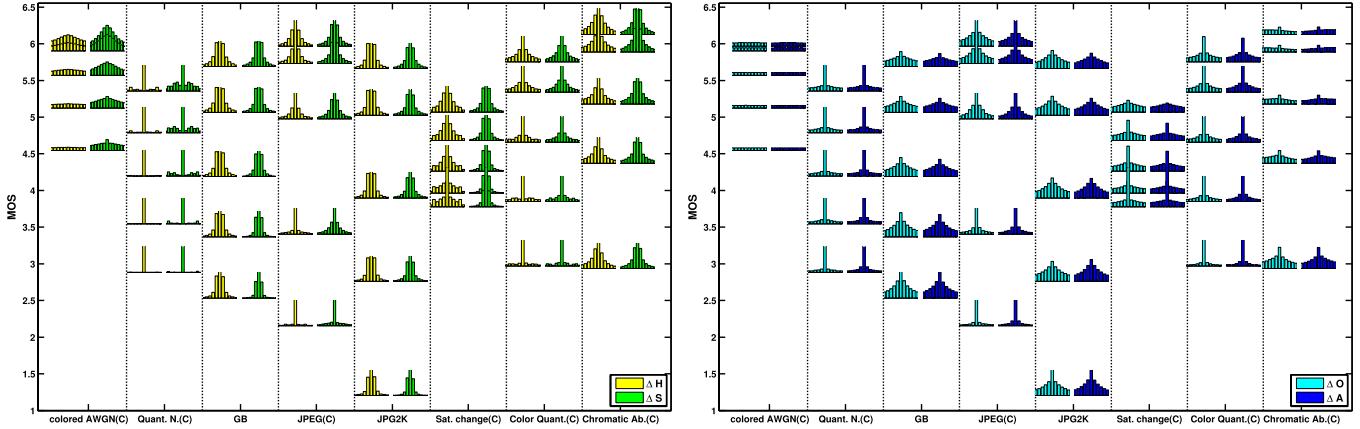


Fig. 4. The empirical marginal histograms of  $\Delta H$ ,  $\Delta S$ ,  $\Delta O$  and  $\Delta A$  obtained from distorted images in TID2013 DB. Chromatic distortion types are indicated with (C). For individual distortion types, five levels of degradation exist. The vertical axis represents the Mean Opinion Score (MOS) value, where the higher MOS corresponds to the better visual quality. Individual bar graphs are obtained by averaging results from the 24 images of different scene contents.

histogram of COREL DB images, the von Mises (VMD) and the wrapped Cauchy distributions (WCD). It is evident from the visual inspection of Fig. 3 that the WCD model produces more accurately fitted curve with the  $\Delta H$  histogram than the VMD model. The numerical comparisons between the empirical histogram and the fitted models using log-likelihood values also produced the same results (i.e., the WCD model results in higher log-likelihood value). The reason behind this is that the WCD model fits well with a peaky and heavy-tailed circular distribution compared to the VMD model.

Consequently, we model the probability density function (pdf) of the relative hue, denoted by a random variable  $R_h$ , using the following WCD model:

$$f_{R_h}(r_h; \mu_h, \rho_h) = \frac{1}{2\pi} \frac{1 - \rho_h^2}{1 + \rho_h^2 - 2\rho_h \cos(r_h - \mu_h)} \quad (11)$$

where  $\mu_h$  is the location parameter indicating the axis of symmetry and  $\rho_h$  is the scale parameter ( $0 \leq \rho_h < 1$ ). The larger the value of  $\rho_h$ , the denser the clustering around the  $\mu_h$ .

In Fig. 4, we plot the empirical histograms of  $\Delta H$  for TID2013 DB images impaired by eight types of distortions with varying severity levels. It is observed that the shape of  $\Delta H$  distribution provides an informative cue pertaining to the type of distortions as well as their severity. For instance, the  $\Delta H$  histograms for colored AWGN exhibit more wider distributions than the ones of pristine images; the ones for JPEG lossy encoding exhibit denser distributions; the ones for chromatic aberration produce asymmetric distributions. Therefore, the  $\rho_h$  and  $\mu_h$  parameters of (11) are effective QAFs for handling the aforementioned chromatic distortions. Conversely, for traditional distortions such as GB and JPEG2000, they are not reliable standalone indicators for identifying distortions; hence, complementary QAFs from other color attributes are required for accurate quality prediction.

#### B. Statistical Model for Saturation Component

The saturation of color stimuli is related to the sensation of colorfulness. Increased colorfulness typically leads to an enhanced perceived quality since images of low colorfulness appear to be washed-out, whereas images of extremely high

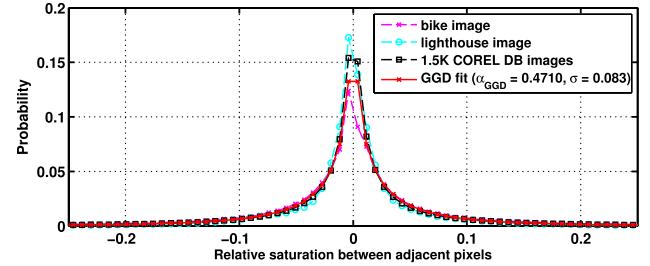


Fig. 5. The empirical histograms of relative saturation for “bike” and “lighthouse” images in Fig. 1 as well as for 1.5K images from COREL DB. The fitted GGD model for COREL DB images are also shown.

colorfulness may appear unrealistic [34]. This observation leads us to focus on spatial dependencies between saturation values of adjacent pixels which can be altered by the presence of distortions rather than directly assessing the level of colorfulness. It is noteworthy that visual impairments caused by changes in spatial structure of saturation can arise by a gamut mapping process, or a severe color quantization in JPEG compression. In order to capture the spatial correlation in saturation component, we remove image redundancies by computing the *relative saturation*  $\Delta S$  as follows:

$$\Delta S^{ho}(i, j) = S(i+1, j) - S(i, j) \quad (12)$$

The empirical histograms of  $\Delta S$  samples in natural images (Fig. 5) exhibit generally symmetrical distributions around zero. The probability histograms of  $\Delta S$  can be well fitted using a generalized Gaussian distribution (GGD), and thus we model the pdf of relative saturation, denoted by a random variable  $R_s$ , using the following GGD model:

$$f_{R_s}(r_s; \alpha_s, \sigma_s^2) = \frac{\alpha_s}{2\beta\Gamma(1/\alpha_s)} \exp\left[-\left(\frac{|r_s|}{\beta}\right)^{\alpha_s}\right] \quad (13)$$

where  $\beta = \sigma_s \sqrt{\Gamma(\frac{1}{\alpha_s}) / \Gamma(\frac{3}{\alpha_s})}$  and  $\Gamma(\cdot)$  denotes the gamma function:  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ . The  $\alpha_s$  of (13) controls the shape of distributions while the  $\sigma_s^2$  controls the variance.

The GGD model is found to be versatile enough to represent the  $\Delta S$  histograms of images affected by various distortion types (Fig. 4). For instance, as quantization noise gets more

visually distracting, the distribution gets denser around zero. Since the  $\Delta S$  histograms of different distortion types and severity levels exhibit varying levels of peakedness and spread, two GGD parameters ( $\alpha_s, \sigma_s^2$ ) are effective QAFs.

### C. Statistical Model for Derivative-Based Color Descriptors

Image derivatives convey crucial information for local image structures (e.g., edges, textures) and have correspondences with the center-surround mechanism of the HVS [23]. Therefore, prior studies have suggested that derivative-based image features should be combined with zeroth-order (RGB pixel-based) features to devise accurate perceptual models of color saliency [26] or color constancy [35]. To this end, we further investigate the general characteristics of natural images by means of derivative-based descriptors, opponent angle  $O$  and spherical angle  $A$ . Contrary to hue  $H$  and saturation  $S$  which are directly related to artists' intuition of color mixing, these derivative-based descriptors are less intuitive. However, they are physically meaningful in the sense that they provide invariant representation to scene-accidental effects (see Table II), extracted from perceptually significant image derivatives. It is noteworthy that the local structural information extracted by  $O$  and  $A$  descriptors is distinguished from the one extracted based on luminance  $Y$ . For instance, the image boundary information captured by MScN coefficients [13] may contain both the boundaries generated by scene-accidental factors (e.g., edges caused by the shape or position of an object with respect to light sources) and perceptually more significant boundaries (e.g., edges between two objects or the transitions between objects and background), whereas the  $O$  and  $A$  emphasizes the latter.

Both  $O$  and  $A$  are defined as angular quantities; again, we define the relative terms,  $\Delta O$  and  $\Delta A$ , by computing the angular difference of descriptor values between adjacent pixels as follows (to remove spatial dependencies):

$$\begin{aligned}\Delta O^{ho}(i, j) &= \Psi(O(i+1, j), O(i, j)) \\ \Delta A^{ho}(i, j) &= \Psi(A(i+1, j), A(i, j))\end{aligned}\quad (14)$$

In Fig. 6, the  $\Delta O$  and  $\Delta A$  histograms of COREL DB images are fitted with the VMD and WCD models, and the latter is observed to be suitable to represent them. Hence, we model the pdfs of  $\Delta O$  and  $\Delta A$ , denoted by two random variables  $R_o$  and  $R_s$ , using the WCD models as follows:

$$\begin{aligned}f_{R_o}(r_o; \mu_o, \rho_o) &= \frac{1}{2\pi} \frac{1 - \rho_o^2}{1 + \rho_o^2 - 2\rho_o \cos(r_o - \mu_o)} \\ f_{R_a}(r_a; \mu_a, \rho_a) &= \frac{1}{2\pi} \frac{1 - \rho_a^2}{1 + \rho_a^2 - 2\rho_a \cos(r_a - \mu_a)}\end{aligned}\quad (15)$$

where  $\mu_o$  and  $\mu_a$  are the location of symmetry parameters;  $\rho_o$  and  $\rho_a$  are the scale parameters.

The empirical histograms of  $\Delta O$  and  $\Delta A$  samples for various distortion types are illustrated in Fig. 4. As can be seen, they generally exhibit unimodal distributions with varying degrees of peakedness; thus the  $\rho$  parameter of WCD model is an effective QAF. The  $\mu$  parameter of WCD model is found to be useful for detecting chromatic aberration since the deviation of  $\mu$  from zero reflects the skewness of marginal distributions.

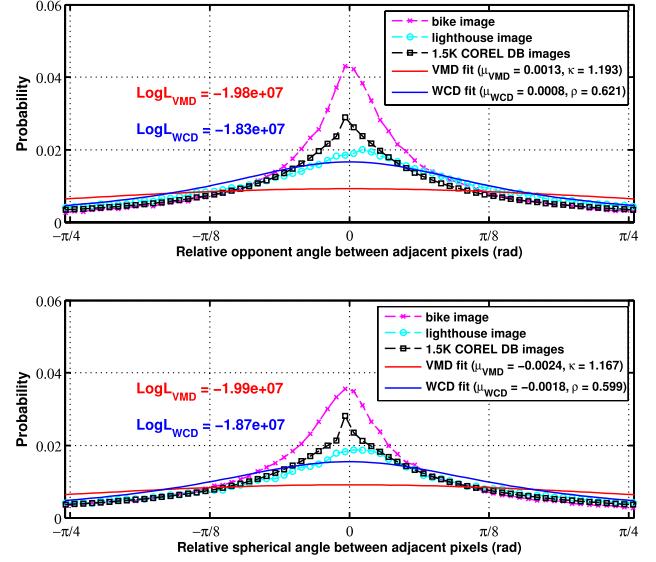


Fig. 6. The empirical histograms of  $\Delta O$  and  $\Delta A$  for “bike” and “lighthouse” images in Fig. 1 as well as for 1.5K images from COREL DB. The fitted circular distribution models for 1.5K COREL DB images are also shown.

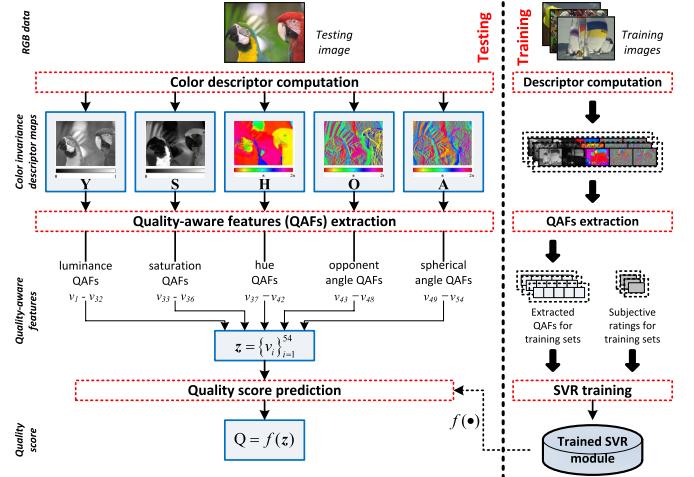


Fig. 7. The workflow of the proposed NR IQA metric.

## IV. APPLICATION OF THE STATISTICAL MODELS OF INVARIANT DESCRIPTORS IN IMAGE QUALITY METRIC

In Section III, we have developed parametric NSS models of chromatic data in natural images and have shown that the presence of image distortions leads to the deviation of image statistics from such models. In this section, the practical utility of proposed NSS models are validated by unifying them within a general-purpose NR IQA framework, namely, the Invariance DDescriptor-based ALgorithm (IDEAL). It is composed of two main stages: feature extraction followed by quality score prediction as illustrated in Fig. 7. Initially, five descriptor maps are retrieved from an input RGB image representing independent aspects of color perception (as in Table II). From individual maps, we extract QAFs which can be viewed as a process of removing redundancies to reveal the low dimensional statistical description of input image data for IQA task. The set of QAFs is systematically chosen to precisely approximate the quality perception mechanism (i.e., the predicted quality scores correlates well with human judgments

across various image scenes and distortion types). Apart from the proposed chromatic QAFs in Section III, the luminance-based QAFs [13] are adopted since they have proven effective in dealing with achromatic structural distortions.

The extracted QAFs are then gathered together to form a vector, denoted by  $\mathbf{z}$ , which becomes an input to a quality prediction module. In the proposed metric, the quality prediction function  $f(\cdot)$  is trained using a machine learning technique in order to reproduce a complex underlying relationship between QAFs and perceived quality.

### A. Luminance Quality-Aware Feature Extraction

The statistics of MSCN coefficients have been widely deployed in NR metrics [13], [14], [17] since the presence of achromatic structural distortions changes the statistics of MSCN coefficients, and the degree of these changes becomes a reliable basis for IQA. Hence, we adopt QAFs based on MSCN coefficients in the proposed metric. The MSCN coefficients  $\hat{Y}$  are extracted from the luminance  $Y$  of image data as follows:

$$\hat{Y}(i, j) = \frac{Y(i, j) - \mu_Y(i, j)}{\sigma_Y(i, j) + 1} \quad (16)$$

where  $(i, j)$  is the spatial location,  $\mu_Y$  is the local luminance mean and  $\sigma_Y$  is the local luminance variance.<sup>7</sup>

The histograms of pairwise products of neighboring MSCN coefficients along four orientations, i.e.,  $M^{ho}(i, j) = \hat{Y}(i, j)\hat{Y}(i+1, j)$ ,  $M^{d1}(i, j) = \hat{Y}(i, j)\hat{Y}(i+1, j-1)$ ,  $M^{ve}(i, j) = \hat{Y}(i, j)\hat{Y}(i, j-1)$ , and  $M^{d2}(i, j) = \hat{Y}(i, j)\hat{Y}(i-1, j-1)$ , are observed to follow a zero mode asymmetric generalized Gaussian distribution (AGGD) given by [13]:

$$f_M(m; \sigma_l, \sigma_r, \gamma) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp\left[-\left(\frac{-m}{\beta_l}\right)^\gamma\right], & m < 0 \\ \frac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp\left[-\left(\frac{m}{\beta_r}\right)^\gamma\right], & m \geq 0 \end{cases} \quad (17)$$

where  $\beta_l = \sigma_l \sqrt{\Gamma(\frac{1}{\gamma})/\Gamma(\frac{3}{\gamma})}$  and  $\beta_r = \sigma_r \sqrt{\Gamma(\frac{1}{\gamma})/\Gamma(\frac{3}{\gamma})}$ . The parameter  $\gamma$  adjusts the shape of the distribution and the scale parameters  $\sigma_l$  and  $\sigma_r$  control the spread of the AGGD to the left and right of the origin. The mean parameter  $\eta$  of the best AGGD fit is computed as follows:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)} \quad (18)$$

The AGGD parameters  $(\sigma_l, \sigma_r, \gamma, \eta)$  for paired products of four orientations are extracted as QAFs in the proposed metric due to their relevance to quality perception. They can be efficiently estimated by means of the moment-matching approach [36].

The luminance-based QAFs are extracted over two spatial resolutions since the multi-scale extraction of luminance QAFs improves the accuracy of quality prediction when dealing with changes in the image resolution or variations in viewing

<sup>7</sup>Inheriting the authors' recommendation [13],  $\mu_Y$  and  $\sigma_Y$  at each pixel are computed by taking a  $7 \times 7$  local window with a 2D circularly-symmetric Gaussian weighting functions sampled out to 3 standard deviations and re-scaled to unit volume.

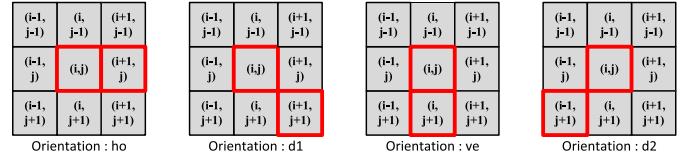


Fig. 8. Definition of the four orientations in pixel arrangement.

distance [13]. In other word, the feature extraction is repeated after low-pass filtering the image and subsampling it by a factor of 2. Consequently, 32 features —  $\{\sigma_{l,i}^d, \sigma_{r,i}^d, \gamma_i^d, \eta_i^d\}$ , where  $i \in \{1, 2\}$  is the index for the multi-scale extraction and  $d \in \{ho, ve, d1, d2\}$  is the orientation index (Fig. 8) — constitutes the luminance-based QAF vector  $\mathbf{v}_Y$ .

### B. Chromatic Quality-Aware Feature Extraction

In this section, we present chromatic QAFs to be used in the proposed IDEAL metric. This process begins with the extraction of relative quantities ( $\Delta H, \Delta S, \Delta O$  and  $\Delta A$ ) from a given distorted image. Regarding the number of orientations for extracting relative quantities, having more than two primary orientations does not substantially increase the correlation between the metric scores and subjective ratings; thus we only consider orientations  $d \in \{ho, ve\}$ . This is consistent with a physiological phenomenon called the *oblique effect* that the sensitivity for a majority of vision tasks is superior for stimuli aligned in horizontal or vertical orientations, as compared to stimuli in oblique orientations [37]. The chromatic QAFs are extracted only from a full-scale image as there is no prominent gain in performance by having more than one scale.

1) *Hue-Based QAFs*: For the extraction of hue-based QAFs, we compute  $\Delta H$  along primary orientations as follows:

$$\begin{aligned} \Delta H^{ho}(i, j) &= \Psi(H(i+1, j), H(i, j)) \\ \Delta H^{ve}(i, j) &= \Psi(H(i, j+1), H(i, j)) \end{aligned} \quad (19)$$

Then, two parameters of the WCD model ( $\mu_h, \rho_h$ ) in (11) fitted into the empirical histograms of  $\Delta H$  are extracted as QAFs using the recursive estimation method in [38].

We found that the circular kurtosis  $k_h$  of  $\Delta H$  samples provides additional gain for accurate quality prediction. It is because the  $k_h$  provides reliable information about the central tendency of circular distribution when the empirical distribution deviates largely from the fitted WCD curve. The  $k_h$  is also estimated for two orientations. Consequently, we obtain the hue-based QAF vector  $\mathbf{v}_H$  of six elements:

$$\mathbf{v}_H = [\mu_h^{ho}, \rho_h^{ho}, k_h^{ho}, \mu_h^{ve}, \rho_h^{ve}, k_h^{ve}] \quad (20)$$

2) *Saturation-Based QAFs*: Initially, we compute the  $\Delta S$  for two primary orientations as follows:  $\Delta S^{ho}(i, j) = S(i+1, j) - S(i, j)$ ,  $\Delta S^{ve}(i, j) = S(i, j+1) - S(i, j)$ . Subsequently, two parameters  $(\alpha_s, \sigma_s^2)$  in (13) are attained from the GGD fits on the histograms of  $\Delta S^{ho}$  and  $\Delta S^{ve}$  using the estimation approach in [39]. It results in the saturation-based QAF vector  $\mathbf{v}_S$  of four elements:

$$\mathbf{v}_S = [\alpha_s^{ho}, \sigma_s^{2,ho}, \alpha_s^{ve}, \sigma_s^{2,ve}] \quad (21)$$

TABLE IV

THE EXTRACTED QAFs AND FEATURE IDs OF THE PROPOSED METRIC.  
(NUMBERS IN PARENTHESES INDICATE THE INDEX OF FORMULA)

Descriptor	Model	Feature ID	Feature elements
Y	AGGD (17)	$v_1 - v_{32}$	$\sigma_{l,1}^{ho}, \sigma_{r,1}^{ho}, \gamma_1^{ho}, \eta_1^{ve}, \sigma_{l,1}^{ve}, \sigma_{r,1}^{ve}, \gamma_1^{ve}, \eta_1^{ve}, \sigma_{d,1}^{ho}, \sigma_{d,1}^{ve}, \eta_1^{d1}, \eta_1^{d2}, \sigma_{d,2}^{d1}, \sigma_{d,2}^{d2}, \eta_2^{d1}, \eta_2^{d2}, \sigma_{r,1}^{ho}, \sigma_{r,1}^{ve}, \eta_1^{r1}, \eta_1^{r2}, \sigma_{r,2}^{ho}, \sigma_{r,2}^{ve}, \eta_2^{r1}, \eta_2^{r2}, \sigma_{r,2}^{d1}, \sigma_{r,2}^{d2}, \gamma_2^{d1}, \eta_2^{d1}, \sigma_{r,2}^{d2}, \gamma_2^{d2}, \eta_2^{d2}$
S	GGD (13)	$v_{33} - v_{36}$	$\alpha_s^{ho}, \sigma_s^{2,ho}, \alpha_s^{ve}, \sigma_s^{2,ve}$
H	WCD (11)	$v_{37} - v_{42}$	$\mu_h^{ho}, \rho_h^{ho}, k_h^{ho}, \mu_h^{ve}, \rho_h^{ve}, k_h^{ve}$
O	WCD (15)	$v_{43} - v_{48}$	$\mu_o^{ho}, \rho_o^{ho}, k_o^{ho}, \mu_o^{ve}, \rho_o^{ve}, k_o^{ve}$
A	WCD (15)	$v_{49} - v_{54}$	$\mu_a^{ho}, \rho_a^{ho}, k_a^{ho}, \mu_a^{ve}, \rho_a^{ve}, k_a^{ve}$

3) *Opponent/Spherical Angle-Based QAFs:* The retrieval of the descriptor maps of opponent angle  $O$  and spherical angle  $A$  requires the computation of image derivatives, e.g.,  $R_x$ ,  $G_x$  and  $B_x$ . Please note they are computed by convolving individual RGB channels with Gaussian derivative filter with a standard deviation  $\sigma = 1$  [40]. Subsequently, the  $\Delta O$  and  $\Delta A$  are computed for two primary orientations analogous to (19). Then, the model parameters  $(\mu, \rho)$  of the fitted WCD models in (15) as well as the circular kurtosis are extracted as QAFs from the histograms of  $\Delta O^{ho}$ ,  $\Delta O^{ve}$ ,  $\Delta A^{ho}$ , and  $\Delta A^{ve}$ . Consequently, we obtain the opponent angle- and spherical angle-based QAF vectors, denoted by  $\mathbf{v}_O$  and  $\mathbf{v}_A$ , as follows:

$$\begin{aligned} \mathbf{v}_O &= \left[ \mu_o^{ho}, \rho_o^{ho}, k_o^{ho}, \mu_o^{ve}, \rho_o^{ve}, k_o^{ve} \right] \\ \mathbf{v}_A &= \left[ \mu_a^{ho}, \rho_a^{ho}, k_a^{ho}, \mu_a^{ve}, \rho_a^{ve}, k_a^{ve} \right]. \end{aligned} \quad (22)$$

### C. Prediction of Quality From Quality-Aware Features

Let  $\mathbf{z} = \{v_i\}_{i=1}^{54}$  be the concatenated vector of all extracted QAFs, where individual QAFs are summarized in Table IV.<sup>8</sup> The quality score  $Q$  can be represented as a function of  $\mathbf{z}$ :

$$Q = f(\mathbf{z}) \quad (23)$$

where  $f(\cdot)$  is an operator relating the elements of  $\mathbf{z}$  to the final quality score. Explicit modeling of the  $f(\cdot)$  is nontrivial because the interactions between a set of QAFs and perceptual quality is complex and highly nonlinear. In this work, we use the Support Vector Regression (SVR)<sup>9</sup> to estimate  $f(\cdot)$  since it is a well-established machine learning solution for handling high dimensional data [43]. In particular, the kernel-based SVR is adopted since it allows for mapping of input data into high dimensional feature space so that nonlinear problem can be solved as a simpler linear regression problem.

Suppose that the training data  $\{(\mathbf{z}_1, s_1), \dots, (\mathbf{z}_N, s_N)\}$  is presented to the SVR module, where  $N$  is the number of

<sup>8</sup>The QAFs considered in the proposed metric are chosen to perform reasonably well regardless of the image contents and distortion types (Section V-C). Better prediction accuracy may be achieved by introducing a feature dimensionality reduction process (reducing the redundancies present in the extracted QAFs) or a feature selection process (identifying more influential QAFs from the complete list of QAFs) as presented in [41], but it is beyond the scope of this paper.

<sup>9</sup>Among two popular versions of SVR, the  $\epsilon$ -SVR and  $\nu$ -SVR, we used the former since it is the original formulation of SVR [42] and is commonly used one in existing IQA metrics which allows for a fair comparison.

TABLE V

DESCRIPTION OF EVALUATION DATABASES ( $N_R/N_T$ : NUMBER OF UNDISTORTED/DISTORTED IMAGES; BPC: BITS PER COLOR CHANNEL)

Database	$N_R/N_T$	Format	Resolution	Types of distortions
LIVE	29 / 779	8-BPC BMP	480x720 to 768x512	5 simulated distortion types
CSIQ	30 / 866	8-BPC PNG	512x512	6 simulated distortion types
TID2013	24 / 2880	8-BPC BMP	512x384	6 chromatic (colored AWGN, quantization, JPEG, saturation change, quantization with dither, chromatic aberration) + 18 achromatic
CID2013	0 / 480	8-BPC JPG	1600x1200	realistic distortions from multiple concurrent sources (combination of camera sensors, optics and image processing pipelines)

training images,  $\mathbf{z}_i$  and  $s_i$  are the extracted feature vector and the subjective rating associated with the  $i$ -th training image. For the kernel-based SVR, the  $f(\cdot)$  to be learned is given by:

$$f(\mathbf{z}) = \sum_{i=1}^N (a_i - a_i^*) K(\mathbf{z}_i, \mathbf{z}) + b \quad (24)$$

where  $K(\mathbf{z}_i, \mathbf{z})$  is the kernel function;  $a_i$ ,  $a_i^*$  and  $b$  are the variables to be optimized during the training. The SVR finds a  $f(\cdot)$  that has the deviation of at most  $\epsilon$  from the target  $s_i$  for all the training data and at the same time is as flat as possible. This can be formulated as a following optimization problem:

$$\begin{aligned} \text{maximize: } & -\frac{1}{2} \sum_{i,j=1}^N (a_i - a_i^*)(a_j - a_j^*) K(\mathbf{z}_i, \mathbf{z}_j) \\ & - \epsilon \sum_{i=1}^N (a_i + a_i^*) + \sum_{i=1}^N s_i (a_i - a_i^*) \\ \text{subject to: } & \sum_{i=1}^N (a_i - a_i^*) = 0 \text{ and } a_i, a_i^* \in [0, C] \end{aligned} \quad (25)$$

where  $C > 0$  is the tradeoff parameter (a larger value of  $C$  leads to over-fitting).

For SVR solutions, selecting an appropriate kernel function  $K(\cdot)$  is of vital importance to ensure good performance. We use the Radial Basis Function (RBF) as  $K(\cdot)$  because it requires less number of parameters to estimate than other alternatives, e.g., polynomial or sigmoid, while producing competitive performance in many scenarios [44]. The RBF kernel is given by:  $K(\mathbf{z}_i, \mathbf{z}) = \exp(-\gamma \|\mathbf{z}_i - \mathbf{z}\|^2)$ , where  $\gamma > 0$  is a parameter controlling the radius of kernel.

During the training phase, the unknown terms in (24) are estimated. In the test phase, the trained system is presented with the test feature vector  $\mathbf{z}$  and predicts the final quality score  $Q$ . There are two parameters  $(C, \gamma)$  to configure for the RBF kernel-based SVR. Adopting the recommendation from [44], the optimal values of  $(C, \gamma)$  are identified for individual tests by performing a grid-search, i.e., examining various combinations of  $(C, \gamma)$  and choosing the one that yields the highest performance (i.e., high SRCC/PLCC values).

## V. PERFORMANCE EVALUATION

### A. Databases and Evaluation Criteria

We incorporate four publicly available databases in Table V to evaluate the performance of the proposed metric, e.g., TID2013 DB [19], LIVE DB [45], CSIQ DB [46] and

CID2013 DB [47]. They provide a comprehensive ground for evaluation as they contain diversifying visual scenes and distortion types. In particular, the TID2013 DB is useful for our analysis as it contains a balanced mixture of chromatic and achromatic distortion types. The recently introduced CID2013 DB [47] provides the images of eight types of scenes captured by different digital devices (ranging from low to high quality cameras, including mobile phones, compact cameras, and SLR cameras); thus, it offers photographs of more real-world scenarios with many concurrent distortion sources.

In order to evaluate whether a metric is statistically consistent with visual perception, predicted metric scores are compared with subjective ratings using three evaluation criteria suggested by the Video Quality Experts Group (VQEG) [48]: i) Pearson Linear Correlation Coefficient PLCC, ii) Spearman Rank-order Correlation Coefficient (SRCC) and iii) Root Mean Squared Error (RMSE). The PLCC measures the *prediction accuracy* of a metric, i.e., the ability of predicting the subjective ratings with low error, whereas the SRCC measures the *prediction monotonicity*, i.e., the degree to which a metric agrees with the rank of the subjective ratings. The RMSE is defined as the root mean square difference of predicted metric scores and subjective scores. In order to remove the non-linearity of subjective rating process, the PLCC and RMSE are measured after applying a non-linear regression using the logistic function  $f$  [45].

$$\tilde{p} = f(p) = \beta_1 \left[ \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(p - \beta_3))} \right] + \beta_4 p + \beta_5 \quad (26)$$

where  $p$  is the metric score,  $\tilde{p}$  is the mapped score, and  $\{\beta_i\}_{i=1}^5$  are the parameters to be determined by minimizing the sum of squared differences between  $\tilde{p}$  and subjective score. Please note that the SRCC is not affected by the non-linear regression. A metric correlates well with subjective human scores if SRCC and PLCC are close to 1 and RMSE is close to 0.

### B. Metric Configuration

Since the proposed approach requires a training process to calibrate a SVR module, we divide each database into two randomly chosen non-overlapping subsets: 80 % training and 20 % testing adopting the general practice in NR metric validation [10], [13], [15] (See Section V-D for the influence of varying training/testing partition sizes on performance). To ensure that the results generalize across different training/test sets, we repeat this test procedure 1000 times and the medians of the performance indicators, i.e., SRCC, PLCC or RMSE, across 1000 trials are reported.

### C. Importance of Individual Quality-Aware Feature Types

The single and joint contributions of QAF types on the quality prediction accuracy are investigated in this section. The subsets of QAFs are selectively chosen and trained on the individual databases via the protocol described in Section V-B. Note that the two subsets of TID2013 DB are considered in this analysis, one containing images of six chromatic distortions and the other containing images

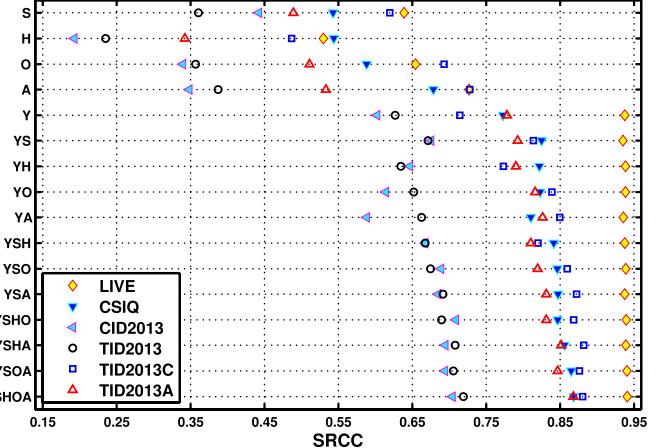


Fig. 9. SRCC performance of the proposed metric with various QAF combinations on evaluation databases.

of eleven commonly encountered achromatic-structural distortions, referred as TID2013(C) and TID2013(A), respectively (denoted by “color” and “actual” subsets by original authors [19]). We use SRCC as a performance indicator but PLCC and RMSE yield very similar results.

Fig. 9 shows that the most effective QAF type is  $v_Y$  in the most evaluation databases. This indicates that when dealing with the entire database, i.e., mixture of distortion types, the luminance statistics contributes the most for quality prediction, which is consistent with the fact that the HVS is generally more sensitive to achromatic than to chromatic signal [49]. However,  $v_A$  is slightly superior to  $v_Y$  for TID2013(C), indicating the enhanced significance of chromatic data against chromatic distortions. It is notable that the luminance QAF  $v_Y$  performs favorably on the LIVE DB by yielding nearly the maximum achievable SRCC. One possible explanation is that the distortions in the LIVE DB are traditional types where solely luminance-based statistics are sufficient to precisely predict the visual quality perception. This result is in accordance with the observations in [20], where luminance-based similarity measures yield high correlation with subjective ratings for the LIVE DB in FR IQA.<sup>10</sup> Conversely, complementary effects of chromatic QAFs are relatively significant for remaining databases, i.e., the SRCC gain of a complete set of QAFs compared to solely luminance QAF is approximately 0.1. The results imply that different distortions manifest by modifying different image statistics; hence, we need to combine all proposed QAFs to properly estimate quality when different types of distortions are examined together.

### D. Performance Dependency on Training Set Size

In order to demonstrate the performance dependency of the proposed metric on training set size, we report the performance of the metric for a varying ratio of train/test sets split  $R_{tr}$  from 0.1 to 0.9 for individual databases in Fig. 10. For this

<sup>10</sup>Note that seemingly insignificant difference of median SRCC values in Fig. 9 may have a statistically significant meaning. For instance, as will be discussed in Table VII, the result of one-sided t-test indicates that the improvement of SRCC performance achieved by YSHOA over Y (denoted as “V3” and “V1” in Table VII, respectively) is statistically significant even though the gain of median SRCC in Fig. 9 is insignificant.

TABLE VI

PERFORMANCE INDICES OF IQA METRICS (MEDIAN SRCC, PLCC, AND RMSE VALUES ACROSS 1000 TRAIN-TEST TRIALS) ON DATABASES. TWO BEST NR METRICS ARE BOLDED FOR INDIVIDUAL DATABASES. A BETTER IQA METRIC HAS HIGHER PLCC/SRCC, WHILE LOWER RMSE. THE INDICES OF FR IQA METRICS, i.e., SSIM AND DSCSI, FOR CID2013 DB ARE NOT REPORTED SINCE REFERENCE IMAGES ARE NOT PROVIDED IN CID2013 DB

Metrics	LIVE (779 images)			CSIQ (866 images)			CID2013 (480 images)		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
SSIM	0.9462	0.9482	8.8403	0.8643	0.8770	0.1305	N/A	N/A	N/A
DSCSI	0.9403	0.9448	9.4390	0.9317	0.9385	0.1028	N/A	N/A	N/A
BRISQUE	0.9311	0.9251	10.0182	0.8441	0.7897	0.1387	0.6485	0.6168	13.7595
GM-LOG	<b>0.9545</b>	<b>0.9497</b>	<b>8.3010</b>	0.8583	0.8155	0.1357	0.6392	0.5790	15.5363
BLIINDS-II	0.9272	0.9214	10.1450	0.8207	0.7588	0.1529	0.5756	0.4940	15.2963
DIIVINE	0.8659	0.8596	13.6274	0.7316	0.6572	0.1806	0.5658	0.5291	16.5004
C-DIIVINE	0.9000	0.8948	11.6783	0.7900	0.7437	0.1594	0.6856	0.6523	12.8484
NIQE	0.9122	0.9111	11.2062	0.7306	0.6290	0.1779	0.6878	0.6643	13.5463
IL-NIQE	0.9095	0.9055	11.5895	0.8634	0.8208	0.1300	0.5341	0.3356	16.1812
V1	0.9424	0.9374	9.0802	0.8400	0.7728	0.1430	0.6662	0.6022	13.8646
V2	0.9416	0.9376	9.2170	<b>0.8728</b>	<b>0.8415</b>	<b>0.1280</b>	<b>0.7393</b>	<b>0.6676</b>	<b>12.6574</b>
V3	<b>0.9462</b>	<b>0.9409</b>	<b>8.7796</b>	<b>0.8913</b>	<b>0.8683</b>	<b>0.1199</b>	<b>0.7496</b>	<b>0.7040</b>	<b>11.6749</b>
<hr/>									
Metrics	TID2013 (2880 images)			TID2013C (720 images)			TID2013A (1320 images)		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
SSIM	0.7992	0.7490	0.7459	0.8050	0.5532	0.6992	0.8933	0.8884	0.5445
DSCSI	0.8824	0.8814	0.6724	0.9235	0.9053	0.4611	0.9105	0.9207	0.5557
BRISQUE	0.6821	0.5970	0.9139	0.7991	0.7186	0.7162	0.8089	0.7669	0.7183
GM-LOG	<b>0.7673</b>	<b>0.7127</b>	<b>0.8023</b>	0.8218	0.7458	0.6810	<b>0.8628</b>	<b>0.8569</b>	<b>0.5984</b>
BLIINDS-II	0.6982	0.6072	0.8946	0.7363	0.6441	0.8114	0.7959	0.7552	0.7439
DIIVINE	0.6313	0.5296	0.9731	0.6412	0.5583	0.9299	0.7245	0.6783	0.8492
C-DIIVINE	0.6693	0.5563	0.9836	0.6635	0.6095	0.8969	0.8013	0.7423	0.7421
NIQE	0.4308	0.3249	1.1176	0.6774	0.5432	0.8709	0.4355	0.4063	1.0940
IL-NIQE	0.6494	0.5252	0.9437	0.7073	0.5513	0.8269	0.8070	0.7807	0.7144
V1	0.7010	0.6267	0.8981	0.7953	0.7143	0.7172	0.8065	0.7783	0.7328
V2	0.7269	0.6676	0.8498	<b>0.8480</b>	<b>0.8201</b>	<b>0.6173</b>	0.8367	0.8103	0.6731
V3	<b>0.7674</b>	<b>0.7190</b>	<b>0.8021</b>	<b>0.8910</b>	<b>0.8806</b>	<b>0.5327</b>	<b>0.8823</b>	<b>0.8679</b>	<b>0.5725</b>

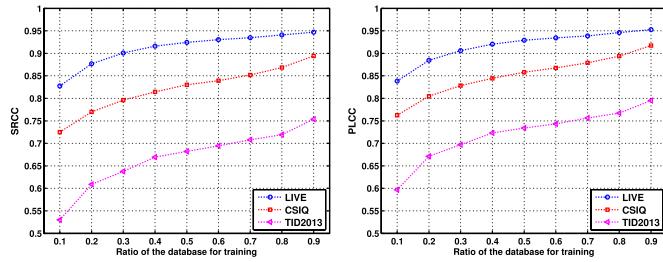


Fig. 10. Variation of metric performance for different ratio of train/test set split for three image databases.

experiment, all five types of QAFs are considered. It is notable that the metric performance does not deteriorate substantially along with the reduction of the training set size (i.e., the SRCC reduction caused by decreasing  $R_{tr}$  from 0.8 to 0.5 are less than 0.05 for all three databases), indicating the robustness of the metric against the small sample problem.

#### E. Metric Performance on Individual Databases

In this section, we compare the performance of the proposed metric with representative general-purpose IQA metrics in literature. The NR metrics in consideration are: i) BRISQUE [13], ii) GM-LOG [15], iii) BLIINDS-II [11], iv) DIIVINE [10], v) C-DIIVINE [12], vi) NIQE [14] and vii) IL-NIQE [17]. Implementations of these metrics are publicly available. All compared NR metrics are luminance-based except for IL-NIQE. For DIIVINE and C-DIIVINE, the one-stage version, i.e., employing SVR to train a regression module which directly maps the feature vector to a final quality score, is used. For a fair evaluation, we used the default parameters configured by the original authors, and optimized the SVR parameters ( $C, \gamma$ ) for the SVR-based metrics by means of a grid-search. Although direct comparisons between

FR and NR metrics are not fair, the performance of the two competitive FR solutions, the SSIM [7] and the DSCSI [20], is also reported for a reference purpose; where the formal is a luminance-based while the latter is a color-based approach. Please note that NIQE, IL-NIQE, SSIM and DSCSI metrics do not require a partition of training and testing sets as other metrics; for a consistent evaluation, the correlations of metric scores with subjective ratings are only reported on the test set.

We consider three variations of the proposed metric by varying the feature vector  $\mathbf{z}$  in (23): i)  $V_1$ : solely luminance-based QAFs, i.e.,  $\mathbf{z} = \{v_i\}_{i=1}^{32}$ ; ii)  $V_2$ : the combination of luminance-, hue- and saturation-based QAFs, i.e.,  $\mathbf{z} = \{v_i\}_{i=1}^{42}$ ; and iii)  $V_3$ : the combination of all QAFs, i.e.,  $\mathbf{z} = \{v_i\}_{i=1}^{54}$ .

Experimental results show that the  $V_3$  consistently remain competitive with existing NR metrics on most databases: yielding higher median SRCC/PLCC and lower median RMSE values (Table VI), as well as more compact distributions around the median SRCC values (Fig. 11). In particular, the observed high correlation of the  $V_3$  with human subjective ratings on CID2013 DB implies that the proposed chromatic QAFs are reliable in terms of accurately predicting perceived image quality for realistic distortion types.<sup>11</sup> One exceptional case is observed for LIVE DB where the GM-LOG slightly outperformed the proposed metric. It is because the local contrast features exploited in GM-LOG metric (e.g., gradient magnitude and Laplacian of Gaussian) perform more favorably than MSCN-based QAFs for LIVE DB images while the proposed chromatic QAFs provide only an insignificant gain as shown in Fig. 9.

An interesting observation is that the  $V_3$  outperforms the SSIM on the TID2013(C) in terms of prediction accuracy.

<sup>11</sup>Again, recall that CID2013 DB offers images acquired from more real-world scenarios with concurrent distortion sources while other databases mainly contain images contaminated by single simulated distortions.

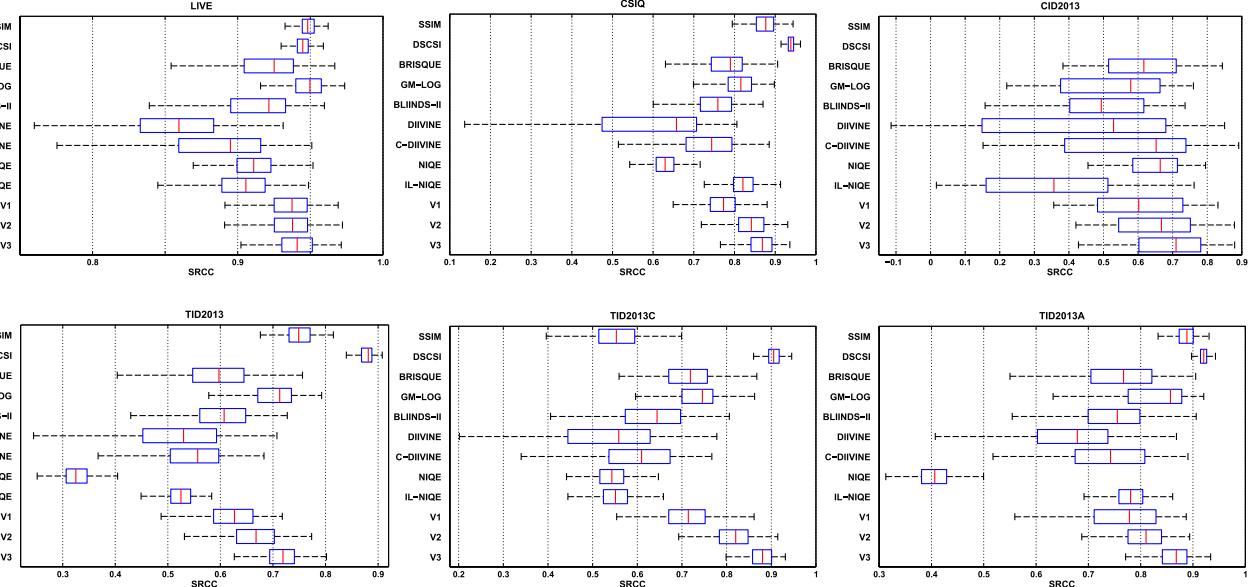


Fig. 11. The box plots of SRCC for IQA metrics across 1000 train-test trials on databases. The left and right edges of the blue box indicate the first quartile (Q1) and third quartile (Q3), respectively. The red vertical line inside the blue box is the median SRCC of 1000 trials. The left and right ends of the whiskers represent the lowest SRCC within 1.5 interquartile range (IQR: distance between Q1 and Q3) of the Q1 and the highest SRCC within 1.5 IQR of the Q3. The SRCC values of FR IQA metrics, i.e., SSIM and DSCSI, for CID2013 DB are not reported since reference images are not provided in CID2013 DB.

TABLE VII

STATISTICAL SIGNIFICANCE TEST (ONE SIDED t-TEST) ON DATABASES. “1” AT LOCATION  $(i, j)$  INDICATES METHOD  $i$  IS STATISTICALLY BETTER THAN  $j$  AT 95% CONFIDENCE LEVEL. “-1” INDICATES THE OPPOSITE AND “0” INDICATES THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN TWO METRICS. (a) LIVE DB. (b) CSIQ DB. (c) CID2013 DB. (d) TID2013 DB. (e) TID2013(C). (f) TID2013(A)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	1	1	1	1	-1	-1	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	-1	0	-1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	-1	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	-1	-1	-1	-1	-1
6. NIQE	-1	-1	1	1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	0	1	-1	0	-1	-1	-1	-1
8. V1	1	-1	1	1	1	1	0	0	0	-1
9. V2	1	-1	1	1	1	1	0	0	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(a)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	1	1	1	1	-1	-1	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	1	-1	-1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
6. NIQE	-1	-1	-1	-1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	-1	1	-1	1	0	-1	-1	-1
8. V1	1	-1	1	1	1	1	0	0	0	-1
9. V2	1	-1	1	1	1	1	0	0	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(b)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	1	1	1	1	-1	-1	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	1	-1	-1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
6. NIQE	-1	-1	-1	-1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	-1	-1	-1	1	0	-1	-1	-1
8. V1	1	-1	1	1	1	1	0	0	0	-1
9. V2	1	-1	1	1	1	1	0	0	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(c)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	0	1	1	1	-1	-1	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	1	-1	-1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
6. NIQE	-1	-1	-1	-1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	-1	-1	-1	1	0	-1	-1	-1
8. V1	1	-1	1	1	1	1	0	0	0	-1
9. V2	1	-1	1	1	1	1	0	0	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(d)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	1	1	1	1	0	-1	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	1	-1	-1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	1	-1	-1	-1	-1
6. NIQE	-1	-1	-1	-1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	-1	-1	-1	1	0	-1	-1	-1
8. V1	1	-1	1	1	1	1	0	0	0	-1
9. V2	1	-1	1	1	1	1	0	0	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(e)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. BRISQUE	0	-1	1	1	1	1	-1	0	-1	-1
2. GM-LOG	1	0	1	1	1	1	1	1	1	-1
3. BLIINDS-II	-1	-1	0	1	1	1	0	1	-1	-1
4. DIVINE	-1	-1	-1	0	-1	0	-1	-1	-1	-1
5. C-DIVINE	-1	-1	-1	0	-1	0	-1	-1	-1	-1
6. NIQE	-1	-1	-1	-1	0	1	-1	-1	-1	-1
7. IL-NIQE	-1	-1	-1	-1	-1	1	0	-1	-1	-1
8. V1	0	-1	1	1	1	1	1	0	-1	-1
9. V2	1	-1	1	1	1	1	1	1	0	-1
10. V3	1	-1	1	1	1	1	1	1	1	0

(f)

Given that our metric does not require a reference image to compare with, this is a meaningful improvement; indicating that a NR metric with a carefully chosen set of QAFs could even outperform a well-established FR metric. Considering the metric variation, it is found that the  $V_3$  performs better than the  $V_1$  and  $V_2$  in most databases. The solid performance of the  $V_3$  can be attributed to the combinatorial use of QAFs based on perceptually relevant color representations. For instance, the  $S$  or  $H$  descriptor-based QAFs capture statistical regularities of color attributes highly correlated with color perception, which led the  $V_2$  more effective than the  $V_1$ . The  $A$ -based and  $O$ -based QAFs complement other QAFs by considering the statistics of image derivatives, which serve as more reliable

quality indicators for highly textured image contents and artifacts involving structural distortions (e.g., damaged image edges).

In order to evaluate the statistical significance of metric performance, we perform one-sided t-test with a 95% confidence level between SRCC values generated by metrics across 1000 train-test trials. The null hypothesis of this analysis assumes that the SRCC values of the metrics in comparison are drawn from populations with equal means. Table VII validates the aforementioned observations: the  $V_3$  outperforms the  $V_1$  and  $V_2$  in terms of a statistical sense; and overall, the proposed metric is highly competitive with all considered NR metrics. These results indicate that the proposed IDEAL

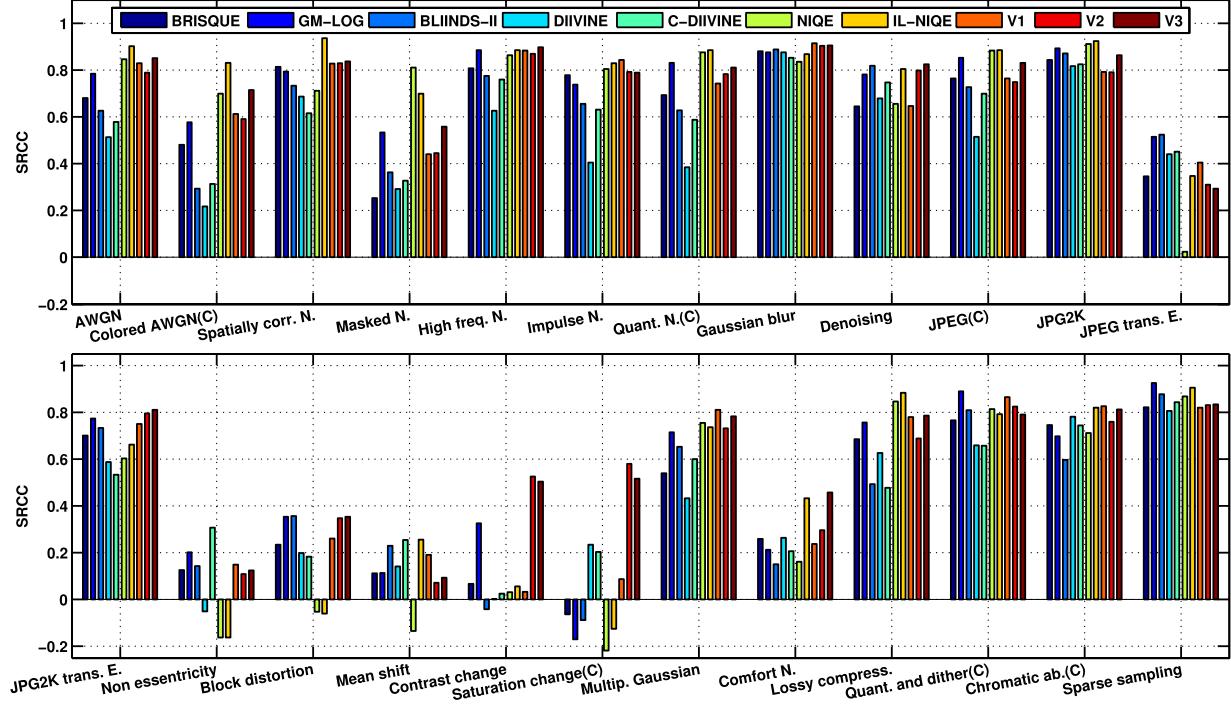


Fig. 12. Median SRCC performance of NR metrics on individual distortion types in TID2013 DB. Chromatic distortions are indicated with (C). Three criteria indicate: i) the number of times that the SRCC for individual distortions are greater or equal to 0.8; ii) the number of times that the SRCC for individual distortions are ranked on top three among metrics; iii) the number of times that the SRCC for individual distortions are ranked on bottom three among metrics.

metric has a good potential as a general-purpose NR metric because of its stable performance over databases of commonly encountered distortion types.

#### F. Robustness of the Metric Against Distortion Types

The proposed metric is further validated on individual distortion types since maintaining stable performance on any specific distortion is a necessary condition for a general-purpose metric. We made use of the TID2013 DB since it contains a wide variety of chromatic and achromatic distortions. It is evident from Fig. 12 that the  $V_3$  maintains consistently stable prediction accuracy across distortions; it shows the highest number of times that it remains on top three among compared metrics. It should be also emphasized that the  $V_3$  do not fail severely on any specific distortions as it is rarely ranked as bottom three metrics. The IL-NIQE slightly outperforms the proposed metric in terms of the number of times that it achieves larger SRCC than 0.8; however, such enhancement can be achieved at the cost of large computational overhead (see Section V-H).

The result shows that the proposed chromatic QAFs not only provide a performance gain for chromatic distortions (e.g., colored AWGN, quantization, JPEG and saturation variation); they also improve accuracy on achromatic-structural distortions (e.g., JPG2K and block distortions), indicating that they are effective for approximating quality perception over a wide variety of distortions. This observation is consistent with

Table VI, Fig. 11 and Table VII, where the  $V_3$  is ranked as the best performer on TID2013(C) and TID2013(A).

#### G. Cross-Database Metric Performance

In order to demonstrate that the metric performance is not bounded to a specific database, a cross-database validation is performed. The metrics are trained on images of training database, then are tested on images on separated testing database. We only considered test images of distortion categories that are covered in the training database for this experiment. Note that the performance of the NIQE and IL-NIQE are reported for a reference purpose, although they do not require subjective ratings for training. CID2013 DB is disregarded as the types of distortions covered in CID2013 DB are fundamentally different from remaining databases.

Although most examined metrics perform fairly well in this analysis, it is clear from Table VIII that the proposed metric is not database-dependent as the  $V_3$  achieves the highest number of times that it is ranked on top three performers along with the BRISQUE and GM-LOG. Overall, the cross-database validation confirms that the proposed IDEAL is effective in terms of the generalization ability.

#### H. Computational Complexity

Having evaluated the correlation of metric scores with subjective ratings, we now demonstrate that the proposed

TABLE VIII  
SRCC PERFORMANCE OF THE NR IQA METRICS ACROSS EVALUATION DATABASES. THREE BEST METRICS ARE BOLDED FOR EACH CASE

Training DB	Testing DB	BRISQUE	GM-LOG	BLIINDS-II	DIIVINE	C-DIIVINE	NIQE	IL-NIQE	V1	V2	V3
LIVE	CSIQ	<b>0.9052</b>	<b>0.9089</b>	0.8957	0.8248	0.8866	0.8707	0.8801	<b>0.8974</b>	0.8889	0.8828
LIVE	TID2013	0.8910	<b>0.9275</b>	<b>0.9122</b>	0.7876	0.9070	0.8126	0.8770	0.8844	0.9095	<b>0.9116</b>
CSIQ	LIVE	<b>0.9373</b>	<b>0.9467</b>	0.9316	0.8847	0.8937	0.9169	0.9157	<b>0.9361</b>	0.9067	0.9115
CSIQ	TID2013	<b>0.7305</b>	0.7283	0.6718	0.6557	<b>0.7556</b>	0.6981	0.7225	0.7268	0.6790	<b>0.7376</b>
TID2013	LIVE	0.9210	<b>0.9290</b>	0.9166	0.8447	0.8790	0.9169	0.9157	0.9094	<b>0.9289</b>	<b>0.9375</b>
TID2013	CSIQ	0.6652	0.5977	0.6375	0.6463	<b>0.7076</b>	0.6274	<b>0.8143</b>	0.6545	0.6493	<b>0.7072</b>

TABLE IX  
EXECUTION DELAY IN SECONDS PER IMAGE ( $512 \times 384$ ) FOR NR IMAGE QUALITY METRICS

Metric	BRISQUE	GM-LOG	BLIINDS-II	DIIVINE	C-DIIVINE	NIQE	IL-NIQE	IDEAL
Processing Delay (s)	0.080	0.051	56.523	18.543	6.626	0.304	13.057	1.377

TABLE X  
PERCENTAGE OF PROCESSING DELAY ON INDIVIDUAL STAGES

Processing Stage	Percentage of Delay
1. Color descriptor computation	16.25
2. QAF extraction	
(a) $\mathbf{v}_Y$ (AGGD parameter estimation)	2.19
(b) $\mathbf{v}_S$ (GGD parameter estimation)	0.75
(c) $\mathbf{v}_H$ (WCD parameter estimation)	24.92
(d) $\mathbf{v}_O$ (WCD parameter estimation)	27.73
(e) $\mathbf{v}_S$ (WCD parameter estimation)	28.22

metric does not introduce substantial computational overhead to achieve high prediction accuracy. The computational complexity is analyzed in terms of MATLAB execution delay for QAF extraction on a 2.90-GHz Inter Core i7 CPU with 4GM RAM system running Window 7 OS. Table IX indicates that the complexity of the proposed metric lies between the spatial domain-based metrics, e.g., BRISQUE, NIQE and GM-LOG, and the frequency domain-based metrics, e.g., DIIVINE and BLIINDS-II. One can also see that the IDEAL is an efficient color-based metric as it is much faster than an existing color-based solution, i.e., the IL-NIQE. Table X summarizes the proportion of time devoted to each processing stages. The most computationally intensive stages is the estimation of WCD parameters for angular descriptors  $H$ ,  $A$  and  $O$ . Given that the delay is measured using unoptimized MATLAB codes, more efficient C/C++ compilable implementation of the maximum likelihood estimation of WCD parameters will significantly improve the applicability of the proposed metric.

## VI. CONCLUSION

We introduce parametric models capable of describing general characteristics of color information in undistorted natural images. The proposed models characterize the correlation of chromatic information between spatially adjacent pixels using the parameters of distribution models and statistical descriptors. A set of color invariance descriptors is incorporated for the derivation of parametric models to extract fundamental structural information of image scenes which is well correlated with human color perception while less sensitive to viewing geometry and illumination variations than image luminances.

To demonstrate the efficacy of the proposed models, we unify them within a general-purpose NR IQA metric, namely the IDEAL. The proposed metric, to our knowledge, is the first attempt to incorporate color invariance descriptors in NR IQA

and to demonstrate their effectiveness in quantifying visual disturbance caused by image distortions. Comprehensive validation performed on large-scale databases demonstrates that the IDEAL is competitive with the state of the art NR metrics in terms of its prediction accuracy and generalization ability. The superior performance of the proposed metric to existing luminance-only solutions supports the psychological evidence that chromatic information provides complementary information to luminance in IQA. Experimental results indicate that the proposed metric can be used for general-purpose in a broad range of color image processing applications due to its superior accuracy and reasonable computational complexity.

## REFERENCES

- [1] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.
- [2] C.-H. Yang, "Inverted pattern approach to improve image quality of information hiding by LSB substitution," *Pattern Recognit.*, vol. 41, no. 8, pp. 2674–2683, 2008.
- [3] R. Herzog, M. Čadík, T. O. Aydçin, K. I. Kim, K. Myszkowski, and H.-P. Seidel, "NoRM: No-reference image quality metric for realistic image synthesis," *Comput. Graph. Forum*, vol. 31, no. 2, pp. 545–554, 2012.
- [4] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
- [5] J. Preiss, F. Fernandes, and P. Urban, "Color-image quality assessment: From prediction to optimization," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1366–1378, Mar. 2014.
- [6] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.
- [7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [8] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
- [9] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 1918–1927, Nov. 2005.
- [10] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [11] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
- [12] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Process., Image Commun.*, vol. 29, no. 7, pp. 725–747, Aug. 2014.

- [13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [14] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [15] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
- [16] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.
- [17] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
- [18] D. Ghadiyaram and A. C. Bovik, "Scene statistics of authentically distorted images in perceptually relevant color spaces for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3851–3855.
- [19] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.
- [20] D. Lee and K. N. Plataniotis, "Towards a full-reference quality assessment for color images using directional statistics," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3950–3965, Nov. 2015.
- [21] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognit.*, vol. 32, no. 3, pp. 453–464, 1999.
- [22] J. van de Weijer and C. Schmid, "Coloring local feature extraction," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 334–348.
- [23] T. Gevers, A. Gijssenij, J. van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. New York, NY, USA: Wiley, 2012.
- [24] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.
- [25] F. Preucil, "Color hue and ink transfer—Their relation to perfect reproduction," in *Proc. TAGA*, 1953, pp. 102–110.
- [26] J. van de Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, Jan. 2006.
- [27] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 167–192, Jan. 2008.
- [28] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
- [29] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.
- [30] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Color and depth priors in natural images," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2259–2274, Jun. 2013.
- [31] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *J. Opt. Soc. Amer. A*, vol. 15, no. 8, pp. 2036–2045, 1998.
- [32] S. K. Naik and C. A. Murthy, "Hue-preserving color image enhancement without gamut problem," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1591–1598, Dec. 2003.
- [33] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.
- [34] C. C. Koh, J. M. Foley, and S. K. Mitra, "Color preference and perceived color naturalness of digital videos," *Proc. SPIE*, vol. 6057, pp. 60570Q-1–60570Q-11, Feb. 2006.
- [35] A. Gijssenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
- [36] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2281–2284.
- [37] S. Appelle, "Perception and discrimination as a function of stimulus orientation: The 'oblique effect' in man and animals," *Psychol. Bull.*, vol. 78, no. 4, pp. 266–278, 1972.
- [38] J. T. Kent and D. E. Tyler, "Maximum likelihood estimation for the wrapped Cauchy distribution," *J. Appl. Statist.*, vol. 15, no. 2, pp. 247–254, 1988.
- [39] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
- [40] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [41] N. Staelens, D. Deschrijver, E. Vladislavleva, B. Vermeulen, T. Dhaene, and P. Demeester, "Constructing a no-reference H.264/AVC bitstream-based video quality metric using genetic programming-based symbolic regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1322–1333, Aug. 2013.
- [42] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA, USA: MIT Press, 1997, pp. 155–161.
- [43] P. Gastaldo, R. Zunino, and J. Redi, "Supporting visual quality assessment with machine learning," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, pp. 1–15, 2013.
- [44] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep., 2003.
- [45] 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. 3441–3452, Nov. 2006.
- [46] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, Jan. 2010.
- [47] T. Virtanen, M. Nuutilainen, M. Vaahteranoksa, and P. Oittinen, "CID2013: A database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2014.
- [48] "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," Video Quality Experts Group (VQEG), Tech. Rep., 2003.
- [49] A. B. Watson, Ed., *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1993.



**Dohyoung Lee** (S'14) received the B.Eng. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2005, and the M.A.Sc. degree in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 2011, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His research interests include color image/video processing and visual quality assessment. He was a recipient of the Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship from 2013 to 2015.



**Konstantinos N. Plataniotis** (S'93–M'95–SM'03–F'12) is currently a Professor, and Bell Canada Chair in multimedia with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests include statistical signal processing, knowledge and digital media design, multimedia systems, biometrics, image processing, biomedical signal processing, and pattern recognition. Among his publications in these fields are the recent books *WLAN Positioning Systems* (2012) and *Multilinear Subspace Learning: Reduction of Multidimensional Data* (2013). He is a Registered Professional Engineer in Ontario and a fellow of the Engineering Institute of Canada. He has served as the Editor-in-Chief of the *IEEE SIGNAL PROCESSING LETTERS* and the Technical Co-Chair of the 2013 IEEE International Conference in Acoustics, Speech, and Signal Processing. He is the IEEE Signal Processing Society Vice President for membership (2014–2016) and the General Chair of the forthcoming 2018 IEEE International Conference on Image Processing. He was a recipient of the IEEE Canada Engineering Educator Award for contributions to engineering education and inspirational guidance of graduate students.