Towards a Full-Reference Quality Assessment for Color Images Using Directional Statistics

Article *in* IEEE Transactions on Image Processing · July 2015

DOI: 10.1109/TIP.2015.2456419 · Source: PubMed

2 authors:

Dohyoung Lee
Qualcomm Canada
14 PUBLICATIONS **63** CITATIONS
SEE PROFILE

Konstantinos Plataniotis
University of Toronto
560 PUBLICATIONS **12,346** CITATIONS
SEE PROFILE

Towards a Full-Reference Quality Assessment for Color Images Using Directional Statistics

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

***Abstract*—This paper presents a novel computational model for quantifying the perceptual quality of color images consistently with subjective evaluations. The proposed full-reference color metric, namely, a directional statistics-based color similarity index, is designed to consistently perform well over commonly encountered chromatic and achromatic distortions. In order to accurately predict the visual quality of color images, we make use of local color descriptors extracted from three perceptual color channels: 1) hue; 2) chroma; and 3) lightness. In particular, directional statistical tools are employed to properly process hue data by considering their periodicities. Moreover, two weighting mechanisms are exploited to accurately combine locally measured comparison scores into a final score. Extensive experimentation performed on large-scale databases indicates that the proposed metric is effective across a wide range of chromatic and achromatic distortions, making it better suited for the evaluation and optimization of color image processing algorithms.**

***Index Terms*—Color image, perceptual image quality, hue, directional data, directional statistics, chromatic distortion.**

I. Introduction

THROUGHOUT the last few decades, we have witnessed rapid growth of color image contents and imaging devices in various sectors of multimedia communication systems, replacing their conventional grayscale counterparts. With such transition, perceptual assessment of color image quality has become essential in the development and optimization of visual data processing algorithms. For instance, reliable judgement of visual quality plays an important role in improving end users' quality of experience in color image reproduction systems. In order to be adequately reproduced across various imaging systems, e.g. monitors, printers, and handheld devices, color images should be processed according to the color gamut of a target device. A mechanism called color gamut mapping converts each out-of-gamut color in image data into the closest reproducible color. This process often introduces visual artifacts, resulting in degradation of the reproduced image quality. Ideally, gamut mapping should be carried out in a way that visual distortion in the reproduction is minimized. Hence, visual quality prediction models can be embedded into gamut mapping systems to ensure faithful reproduction of color image data over different platforms.

In practice, assessment of image quality can be achieved in two ways, subjective judgements and objective metrics[1]. Since the human visual system (HVS) is the ultimate receiver of visual information, the most reliable way to measure image quality is to conduct subjective judgement [1]. However, subjective assessment is infeasible on many occasions since: i) conducting psycho-visual tests following standard protocols is laborious; and ii) subjective tests are not applicable for practical automated systems. Therefore, objective metrics, capable of approximating subjective opinion of an average human observer, e.g. Mean Opinion Score (MOS), have been considered as practical alternatives to subjective judgements.

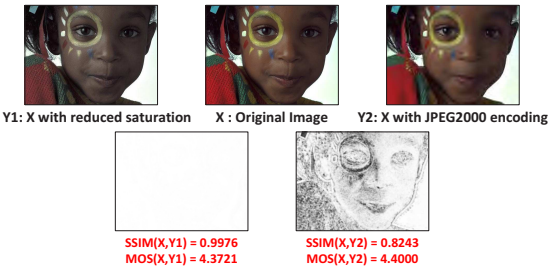


Fig. 1. SSIM scores and SSIM index maps for two images of nearly the same level of subjective ratings, exhibiting chromatic and achromatic distortions.

Numerous objective metrics have been proposed to estimate perceived quality in visual data. Recent studies [1]–[3] demonstrated that some representative methods, such as the Structural SIMilarity (SSIM) index family [4]–[8], yield accurate results on publicly available databases, e.g. LIVE DB [9] and CSIQ DB [10]. However, most commonly used metrics are designed to rely on geometric features in the grayscale domain either because of the relative insignificance of color information in their intended applications or because of computational efficiency. They are not successful in dealing with color images exhibiting chromatic deviations, e.g. images generated from color gamut mapping or preferred memory color reproduction, since they overlook the contribution of chromatic information in visual quality. Fig. 1 demonstrates that quality prediction exclusively based on grayscale information using the SSIM index could result in an overly optimistic score since it underestimates the visual degradation caused by the reduction of saturation in the color image.

In this paper, we introduce a full-reference (FR) color image quality metric, called the Directional Statistics based Color Similarity Index (DSCSI), which can properly handle image data exhibiting both achromatic and chromatic distortions. FR metrics are intended for the evaluation of image data in off-line scenarios where both original and distorted images are available. For example, at the off-line prototyping or

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

[1]Although the term "metric" used throughout this article may not satisfy the mathematical properties of metric, we use it adopting a common practice in the field of image quality assessment.

optimization stage of color image processing systems, FR quality metrics become useful since algorithm developers can have access to both original and distorted data.

The proposed metric provides an accurate prediction of perceived image quality by extracting local image descriptors from both chromatic and achromatic components of input color signals. In particular, perceptually significant hue information is processed with *directional statistics* to properly take into account its angular nature. The concept of applying directional statistics on hue data in order to formulate color descriptors is initially introduced in [11]. However, the relevance of such directional color descriptors to visual perception has been largely unexplored so far; in this paper, we highlight the potential of hue descriptors inferred by directional statistics as perceptual cues for visual quality prediction. The main contributions of this article are as follows:

1) We derive hue similarity measures for accurate quality prediction based on two complementary directional statistics: a circular mean and a circular variance. They are designed to quantify the perceived quality degradation caused by changes in large-scale image structures and changes in edge details of hue channels. With respect to our previous work in [12], we redefine the relationship between the perceived distortion and the deviation of circular mean in hue component as described in (4) in order to beneficially exploit the perceptual uniformity of the deployed color space, i.e. the angular difference in hue data correlates well with the perceived quality degradation. The proposed hue measures and other similarity measures considered in DSCSI provide complementary information for quality perception and, together, improve the accuracy of the metric.
2) We exploit two spatial weighting schemes in order to accurately combine the locally measured color similarity scores into a combined quality score: i) a weighting scheme that considers the significance of hue depending on its chroma value; and ii) a weighting scheme based on a Minkowski framework that distinguishes strong local deviations from weaker ones. These weighting schemes effectively increase the correlation between the proposed metric and the visual perception.
3) Throughout a systematic validation, we show that the proposed metric correlates well with subjective ratings not only for images with chromatic distortions, but also for images with generic distortions, e.g. lossy compression, demonstrating that it can be used for general-purpose in a broad range of color imaging applications.

The rest of this paper is organized as follows. An overview of representative color image quality metrics is presented in Section II. Section III demonstrates the proposed DSCSI metric in detail. Section IV is devoted to experimental results and conclusions are drawn in Section V.

II. Background

Perceptual assessment of image quality plays a significant role in a wide range of image processing applications, e.g. compression, enhancement, and reproduction. In recent years, major efforts have been made to replace subjective judgements with computational metrics that closely agree with HVS characteristics. Existing objective metrics can be classified into three groups depending on the accessibility of the original image to compare with: i) full-reference (FR): the original image is provided along with the distorted image whose quality is to be assessed; ii) reduced-reference (RR): partial information of the original image (e.g. extracted features) is available; and iii) no-reference (NR): the original image is completely unknown. In this paper, our discussion is confined to FR metrics. One of the simplest computational tools to measure image distortion is the Mean Squared Error (MSE). Although the MSE has a clear physical meaning [21], its performance is often suboptimal since the squared magnitude of pixel difference typically has a poor correlation with visual perception [22]. A more elaborate approach that attempts to incorporate structural information in visual quality assessment, the SSIM index [4], has received considerable attention from the research community because of its higher correlation with perception and simple mathematical formulation. It initially estimates the pixel-by-pixel similarities between two corresponding local regions by examining three complementary statistics—mean, variance, and correlation—then derives the global similarity score by taking an average of local measurements. Several extensions, such as the Multi-Scale SSIM (MS-SSIM) [5] and gradient-based SSIM [6]–[8] have been introduced for enhanced prediction performance.

However, the aforementioned metrics are not intended for color images exhibiting chromatic deviations since they disregard the contribution of chromatic information in visual quality. In principle, color image quality can be estimated by applying a grayscale metric on individual RGB channels and combining channel scores afterwards. Since such naive extension is suboptimal, more elaborate approaches exploiting properties of color perception have been proposed (Table I). The most fundamental tools to estimate color distortion are the Euclidean distance in uniform color systems recommended by International Commission on Illumination (CIE): CIELAB and CIELUV. Thereafter, CIE formulated more advanced difference equations, e.g. CIE94, and CIEDE2000 [13]. Traditionally the perceived distortion between color images are determined by computing CIE equations on a pixel basis and then examining summary statistics, e.g. mean or median. Although they are useful to compare uniformly colored patches, there is no guarantee that they work well with complex real-world image data.

Because of its accuracy and compact formulation, the SSIM index has been extended to incorporate color information [14]–[18], [20]. For example, the Universal Color Image Fidelity (UCIF) [14] estimates SSIM indices for each color channel of $l\alpha\beta$ color space and derives a final similarity score by taking the weighted sum of channel indices. Similarly, the Video SSIM (VSSIM)[2] [15] extracts SSIM indices on individual YCbCr channels and combines channel scores linearly. The authors of the UCIF and the VSSIM noted that larger weighting

[2]In this paper, we only consider the intra-frame similarity measurement of the VSSIM and disregard the inter-frame assessment, i.e. the perceptual tuning based on motion vectors, since we are dealing with images.

TABLE I
LIST OF THE REPRESENTATIVE FULL-REFERENCE COLOR IMAGE QUALITY METRICS IN LITERATURE

Metric	Basis Color Space	Incorporated Features	Performance Highlight on Large-scale Databases
CIE color difference equations (CIE76, CIE94, CIEDE2000) [13]	CIELAB	pixel value of L,a^*,b^*	N/A
UCIF (Universal Color Image Fidelity) [14]	l$\alpha\beta$	local statistics (mean, variance, and covariance) of each color channel	N/A
VSSIM (Video Structural SIMilarity) [15]	YCbCr	local statistics (mean, variance, and covariance) of each color channel	SRCC = 0.812 on VQEG Phase I DB
FSIMc (Feature SIMilarity - color) [16]	YIQ	phase congruency and gradient magnitude of luminance Y, pixel values of chrominance I and Q	a) SRCC = 0.8840, PLCC = 0.8762 on TID2008 DB b) SRCC = 0.9645, PLCC = 0.9613 on LIVE DB c) SRCC = 0.9310, PLCC = 0.9192 on CSIQ DB
SHSIM (Structure and Hue SIMilarity) [17]	HSI	local statistics (mean, variance, and covariance) of luminance, local mean of hue	PLCC = 0.9732 on LIVE DB
PSIM (Picture SIMilarity) [18]	HSY	cumulative distribution of each color channel, local statistics (mean, variance, and covariance) of brightness Y	N/A
JID (Just Intolerable Difference) [19]	CIELAB	structural image distortion, non-structural image distortion	N/A
CID (Color Image Difference) [20]	LAB2000HL	local mean of hue and chroma, local statistics (mean, variance, and covariance) of lightness	a) Hitrate = 0.681 on gamut mapping DB with single-scale CID b) SRCC = 0.798 on TID2008 DB with multi-scale CID

should be given to achromatic channels than chromatic ones because of its significance in visual perception. More recently, Zhang et al. [16] extended the Feature SIMilarity (FSIM) by exploiting I,Q chromatic channels in YIQ color space and demonstrated its state of the art prediction performance on large-scale image databases. The aforementioned color metrics are practical in the sense that they exploit opponent color spaces with which most of the image data stream are encoded.

More intuitive color metrics [17], [18], [20] make use of cylindrical color spaces and rely on color attributes that correlate well with human perception, e.g. hue and chroma. The most representative metric in this line of research is the Color Image Difference (CID) [20], which compares images in perceptually uniform space, namely LAB2000HL [23], using three color attributes: hue, chroma, and lightness. The authors noted that it outperforms the SSIM and the MS-SSIM on large-scale color gamut mapping databases, demonstrating its high correlation with human observers in predicting perceived color difference. Although existing metrics based on cylindrical color spaces have shown successful prediction performance, one general limitation is their numerical instability caused by treating directional hue data as linear data, without considering their periodicities. We believe appropriate handling of highly informative hue data will improve the prediction accuracy of color metrics and the aim of this paper is to provide a systematic framework to quantify the visual quality by properly handling directional hue data in an effective way.

III. PROPOSED COLOR IMAGE QUALITY METRIC

In this section, we introduce a full-reference metric to quantify the perceived visual quality of color image data, called the Directional Statistics based Color Similarity Index (DSCSI). The proposed metric is general-purpose, in the sense that it consistently performs well over commonly encountered chromatic and achromatic distortions. The inputs to the metric are two RGB images with identical spatial resolution, $\mathbf{X}$ and $\mathbf{Y}$, denoting the original and the distorted images, respectively. They are assumed to be consistent in bit depth and properly aligned. The output quality score, denoted as $\mathcal{Q}(\mathbf{X},\mathbf{Y})$, lies in the range of [0,1], where the best quality value 1 is achieved when both images are identical.

The workflow of the proposed metric is illustrated in Fig. 2. Initially, $\mathbf{X}$ and $\mathbf{Y}$ are translated to a cylindrical color space where image data are normalized based on viewing distance and decomposed into hue/chroma/lightness channels (Section III-A). In order to account for spatially varying distortions, image similarities are initially estimated in local regions and combined afterwards. Assuming that the quality degradation of image data is caused by a combination of various factors, the DSCSI derives six similarity measures, where each of them retrieves independent indicators for visual quality perception on a pixel basis (Section III-B). Subsequently, pixel-wise similarity scores are combined into a final quality score by applying a spatial pooling for individual measures, followed by a component pooling (Section III-C).

A. Color Space Conversion

To better approximate the color perception, original RGB images are transformed to a color space that is more compatible with human intuition. Among various perceptual color models, e.g. HSV and CIELAB, we make use of the S-CIELAB (spatial extension of CIELAB) [24] since: i) it is a perceptually uniform color model[3], which is appropriate for defining a simple[4] yet precise measure of color distortions; and ii) it incorporates a filtering operation based on a contrast sensitivity function (CSF) into CIELAB channels in order to mimic spatial sensitivities of the HVS. This pre-processor essentially normalizes visual data based on the viewing distance by selectively attenuating imperceptible image details. The importance of CSF filtering in quality prediction performance will be discussed further in Section IV-D.

After RGB to S-CIELAB transformation, each pixel of $\mathbf{X}$ contains three color components: lightness L^*, red-green a^*,

[3]A color space $\mathbb{C} \subset \mathbb{R}^3$ is perceptually uniform if $\|\mathbf{c}_1 - \mathbf{c}_2\|_2 = \mathbf{v}(\mathbf{c}_1, \mathbf{c}_2)$ for $\mathbf{c}_i \in \mathbb{C}$ where $\|\cdot\|_2$ is Euclidean distance in $\mathbb{C}$, and $\mathbf{v}(\mathbf{c}_1, \mathbf{c}_2)$ is the perceived color difference of two color vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ [23].

[4]Simple in the sense that perceived quality Q can be modeled in terms of the Euclidean distance in channel values.

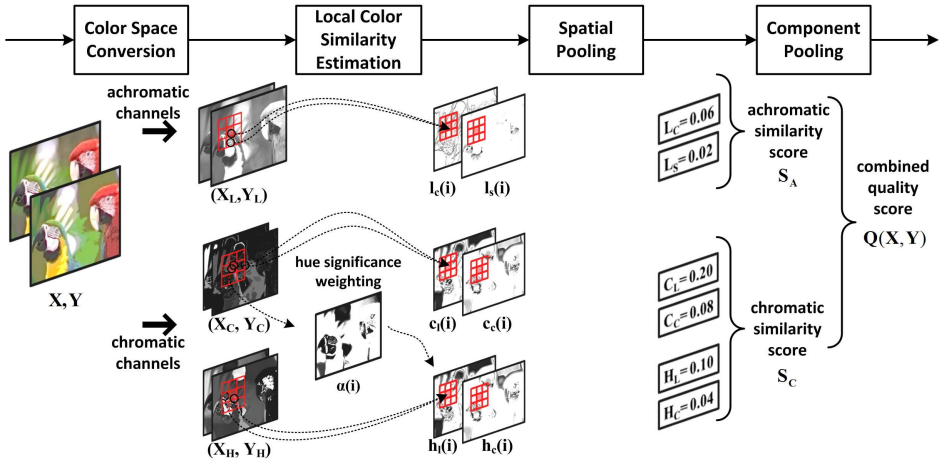


Fig. 2. Illustration of the DSCSI workflow for given two RGB color images $\mathbf{X}$ and $\mathbf{Y}$ (Terminologies are explained in Section III-B and III-C).

and blue-yellow b^*, i.e. $\mathbf{X}(\mathbf{i}) = [X_L(\mathbf{i}), X_a(\mathbf{i}), X_b(\mathbf{i})]^T$, where $\mathbf{i} = (i_1, i_2)$ represents the pixel coordinate. It is widely known that cylindrical color models come closer to how humans discriminate colors since they specify colors with perceptual attributes, e.g. hue, chroma, and lightness [25]. Therefore, we extract the cylindrical representation of the S-CIELAB by deriving hue and chroma components from two chromaticities a^* and b^* as follows:

$$\begin{aligned} X_H(\mathbf{i}) &= \arctan\left[X_b(\mathbf{i}) / X_a(\mathbf{i})\right] \\ X_C(\mathbf{i}) &= \sqrt{X_a(\mathbf{i})^2 + X_b(\mathbf{i})^2} \end{aligned} \tag{1}$$

B. Derivation of Local Color Similarity Measures

The proposed metric models the perceived image quality as a combination of six similarity measures extracted from hue, chroma, and lightness channels. They are designed to have complementary effects so that the combined score correlates well with visual perception across various types of impairments. The importance of each measure may vary depending on operation scenarios. For instance, hue-based measures are primary indicators of visual artifacts when they are used to evaluate color constancy algorithms since discounting illumination color from the color image often leads to hue shift. On the other hand, chroma-based measures are important in the evaluation of gamut mapping algorithms since a majority of them involve clipping/scaling of chroma values while preserving other properties, e.g. hue [26].

Adopting the SSIM framework [4], individual similarity measures process a $k \times k$ window and represent its central pixel by statistical descriptors which convey discriminating information of local color appearance. By comparing local descriptors of two corresponding windows $\mathbf{x}_i$ and $\mathbf{y}_i$ (denoting the local windows of $\mathbf{X}$ and $\mathbf{Y}$, centered at pixel coordinate $\mathbf{i}$), the pixel-wise similarity score is derived. The window moves pixel-by-pixel from the top-left corner of the images to their bottom-right corner. We define the size of a window to $k = 7$ for all similarity measures, which provides a good balance between processing speed and prediction accuracy. Thus, a set of pixels within the local window centered at coordinate (i_1, i_2) can be represented as $\zeta = \{i_1 - 3, \cdots, i_1 + 3\} \times \{i_2 - 3, \cdots, i_2 + 3\}$. In the rest of this section, we develop similarity measures from ones based on directional hue component to ones based on linear color components, e.g. chroma and lightness.

1) Hue-based Similarity Measures: To compare hue channels of two images, denoted as $\mathbf{X}_H$ and $\mathbf{Y}_H$, we derive two measures based on directional statistics: a circular mean and a circular variance (See Appendix A for the computation of directional statistics). Note that by characterizing each hue pixel with its circular mean, larger-scale (or smoothed) hue image structures are mainly retained, while by representing each pixel with its circular variance, edge details in hue images are extracted.

First, we compute the weighted circular mean $\overline{\theta}_H$ and the weighted circular variance v_H for each pixel by taking a $k \times k$ local window. For example, two descriptors for image $\mathbf{X}$ are estimated as:

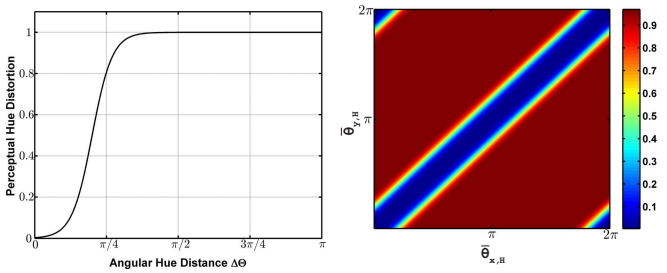
$$\begin{aligned} \overline{\theta}_{\mathbf{x_i},H} &= \arctan\left[\frac{\sum_{\mathbf{i}\in\zeta} w(\mathbf{i}) \sin X_H(\mathbf{i})}{\sum_{\mathbf{i}\in\zeta} w(\mathbf{i}) \cos X_H(\mathbf{i})}\right] \\ v_{\mathbf{x_i},H} &= 1 - \frac{\sqrt{\sum_{\mathbf{i}\in\zeta}[w(\mathbf{i}) \cos X_H(\mathbf{i})]^2 + \sum_{\mathbf{i}\in\zeta}[w(\mathbf{i}) \sin X_H(\mathbf{i})]^2}}{|\zeta|} \end{aligned} \tag{2}$$

where the ζ is a set of pixel coordinates within the window, and the operator $|\cdot|$ indicates the cardinality of a set. The weighting factors $\{w(\mathbf{i}) | \mathbf{i} \in \zeta\}$ are introduced to assign more importance on central pixels for computation of descriptors. We use a $k \times k$ circular-symmetric Gaussian function to define the weight at pixel coordinate (i_1, i_2) around the window center (p_1, p_2) as follows:

$$w(i_1, i_2) = \frac{1}{2\pi\sigma_w} \exp\left[\frac{-\left((i_1 - p_1)^2 + (i_2 - p_2)^2\right)}{2\sigma_w^2}\right] \tag{3}$$

with the standard deviation $\sigma_w = 1$. This Gaussian weighting is used since it effectively reduces the undesirable blocking effect in the resultant similarity map [4].

The first hue-based measure, which compares two circular means $\overline{\theta}_{\mathbf{x_i},H}$ and $\overline{\theta}_{\mathbf{y_i},H}$ at pixel coordinate $\mathbf{i}$, namely, the *hue*



(a) (b)

Fig. 3. (a) Perceptual hue mapping function in (5) with $h_0 = 0.2\pi$ and $\tau = 0.35$ (empirically determined in Section IV-B), (b) Contour plot of the $\mathbf{h}_l$ for different average hue values of two compared windows.

mean similarity $\mathbf{h}_l$ is defined by:

$$\mathbf{h}_l(\mathbf{i}) = \left[1 - f\Big(\pi - \big|\pi - \big|\overline{\theta}_{\mathbf{x}_\mathbf{i},H} - \overline{\theta}_{\mathbf{y}_\mathbf{i},H}\big|\big|\Big)\right]\alpha(\mathbf{i}) \tag{4}$$

where $f(\cdot)$ is the perceptual tuning function and $\alpha(\mathbf{i})$ is the pixel-wise weighting function based on hue significance. The *perceptual tuning function* $f(\cdot) : [0,\pi] \rightarrow [0,1]$ defines the nonlinear relationship between the circular mean difference in degree (i.e. $\Delta\theta_H = \pi - |\pi - |\overline{\theta}_{\mathbf{x},H} - \overline{\theta}_{\mathbf{y},H}||$) and the perceived hue distortion. We define $f(\cdot)$ using a sigmoid function as follows:

$$f(\Delta\theta_H) = 0.5 + 0.5\tanh\left(\frac{\Delta\theta_H - h_0}{\tau \cdot h_0}\right) \tag{5}$$

where $\tau > 0$ is the curvature parameter (the higher the τ, the smoother the curve), and h_0 is the switch parameter, i.e. $f(\Delta\theta_H) = 0.5$ when $\Delta\theta_H = h_0$. A sigmoid curve (See Fig. 3a) is selected since it is a commonly used psychometric function which describes the relationship between level of a physical stimulus and subjective response [27]. The perceptual response f is saturated to the maximum value for large $\Delta\theta_H$ value. This is because quantifying hue distortion is only sensible between similar hue groups, as human observers are unable to make a reasonable estimation for a pair of hues taken from far apart in the hue wheel. The slow growth rate of the curve for insignificant $\Delta\theta_H$ corresponds to stimulus below the detection threshold.

Fig. 3b presents a contour plot of $\mathbf{h}_l$ for different values of $\overline{\theta}_{\mathbf{x},H}$ and $\overline{\theta}_{\mathbf{y},H}$.[5] As can be seen, the mathematical form in (4) takes into account the circular property of hue by ensuring the periodicity of 2π.

The *hue significance function* $\alpha(\mathbf{i}) \in [0,1]$ in (4) facilitates a pixel-wise weighting for hue-based measures based on hue reliability. For a color pixel near the achromatic axis, i.e. achromatic pixel, its hue value has no significance as small perturbations in pixel value may result in dramatic hue changes [25]. Intuitively, for a pixel coordinate $\mathbf{i}$, if either $\mathbf{X}(\mathbf{i})$ or $\mathbf{Y}(\mathbf{i})$ has a low chroma value, this pixel will have minimal influence on the evaluation of hue distortion. Therefore, $\alpha(\mathbf{i})$ is a function of $c_m(\mathbf{i}) = \min(X_C(\mathbf{i}), Y_C(\mathbf{i}))$. Similarly to (5), we exploit a sigmoid function to define the weighting [28]:

$$\alpha(\mathbf{i}) = 0.5 + 0.5\tanh\left(\frac{c_m(\mathbf{i}) - c_0}{l \cdot c_0}\right) \tag{6}$$

where l is the curvature parameter and c_0 is the parameter that controls the point for which the significance level is equal to 0.5. The parameters are set to $c_0 = 10$ and $l = 0.25$ for smooth weight variation between achromatic and chromatic pixels. Essentially, the spatial weighting using $\alpha(\mathbf{i})$ allows the metric to increase the relative importance of chroma- and lightness-based measures for achromatic pixels.

The second hue measure, the *hue dispersion similarity* $\mathbf{h}_c$ complements the $\mathbf{h}_l$ by comparing the circular variances of two corresponding windows as follows:

$$\mathbf{h}_c(\mathbf{i}) = \left[\frac{2v_{\mathbf{x}_\mathbf{i},H} \cdot v_{\mathbf{y}_\mathbf{i},H} + K_H}{v^2_{\mathbf{x}_\mathbf{i},H} + v^2_{\mathbf{y}_\mathbf{i},H} + K_H}\right]\alpha(\mathbf{i}) \tag{7}$$

where $K_H > 0$ is a constant that prevents the denominator of $\mathbf{h}_c$ from being zero. The mathematical form of (7) is commonly used to derive the similarity between real numbers and its output lies between 0 and 1 [4]. The parameter K_H is experimentally determined in Section IV-B based on the dynamic range of circular variance v_H. The hue significance weighting is exploited to derive $\mathbf{h}_c$ since local hue variance is only meaningful for non-achromatic pixels.

2) Chroma-based Similarity Measures: To quantify the visual distortion in chroma components, we exploit two chroma descriptors: a weighted mean μ_C and a weighted standard deviation σ_C. The rationale behind the use of two chroma descriptors is similar to the argument of complementarity made for hue descriptors; that μ_C allows for the comparison of smoothed chroma images, while σ_C allows for the comparison of edge details (or contrast) in chroma images. Unlike hue descriptors, they are computed using linear statistics. For example, both descriptors for image $\mathbf{X}$ within the window centered at $\mathbf{i}$ are estimated as:

$$\begin{aligned}\mu_{\mathbf{x}_\mathbf{i},C} &= \sum_{\mathbf{i}\in\zeta} w(\mathbf{i}) X_C(\mathbf{i}) \\ \sigma_{\mathbf{x}_\mathbf{i},C} &= \left[\sum_{\mathbf{i}\in\zeta} w(\mathbf{i})\left(X_C(\mathbf{i}) - \mu_{\mathbf{x}_\mathbf{i},C}\right)^2\right]^{(1/2)}\end{aligned} \tag{8}$$

where the ζ is a set of coordinates within the local window and the weighting factors $w(\mathbf{i})$ are defined as (3).

The *chroma mean similarity* $\mathbf{c}_l$ compares chroma means, $\mu_{\mathbf{x}_\mathbf{i},C}$ and $\mu_{\mathbf{y}_\mathbf{i},C}$ from corresponding windows at $\mathbf{i}$ as follows:

$$\mathbf{c}_l(\mathbf{i}) = \frac{1}{K_{C1}(\mu_{\mathbf{x}_\mathbf{i},C} - \mu_{\mathbf{y}_\mathbf{i},C})^2 + 1} \tag{9}$$

where K_{C1} is the free parameter to control the relative importance of chroma mean difference in perceived quality. The mathematical form of (9) is adopted from the CID [20], which represents the similarity between two real numbers given in perceptually uniform color domain.

The second comparison measure for chroma components, the *chroma contrast similarity* $\mathbf{c}_c$, is adopted from the contrast term of the SSIM [4]:

$$\mathbf{c}_c(\mathbf{i}) = \frac{2\sigma_{\mathbf{x}_\mathbf{i},C} \cdot \sigma_{\mathbf{y}_\mathbf{i},C} + K_{C2}}{\sigma^2_{\mathbf{x}_\mathbf{i},C} + \sigma^2_{\mathbf{y}_\mathbf{i},C} + K_{C2}} \tag{10}$$

[5]The pixel-wise weighting function α is disregarded in the computation of $\mathbf{h}_l$ for this plot by setting it to 1.

where K_{C2} is the positive stabilizing constant.[6] The values for two free parameters, K_{C1} and K_{C2}, are determined experimentally in Section IV-B.

3) Lightness-based Similarity Measures: In order to quantify the visual distortion in the achromatic component, we consider two dominant factors: loss of correlation and local contrast distortion, adopting the underlying principle of the SSIM [4], [30]. The reasons behind this are the high prediction accuracy of the SSIM towards a wide range of achromatic distortion types and its widespread popularity in practical applications. The derivation of two factors involves computation of local statistics from lightness channels $\mathbf{X}_L$ and $\mathbf{Y}_L$, the local standard deviation and the covariance between two corresponding patches of lightness channels. They can be estimated by:

$$\sigma_{\mathbf{x_i},L} = \left[\sum_{\mathbf{i}\in\zeta} w(\mathbf{i})\left(X_L(\mathbf{i}) - \mu_{\mathbf{x_i},L}\right)^2\right]^{(1/2)} \tag{11}$$

$$\sigma_{\mathbf{xy_i},L} = \sum_{\mathbf{i}\in\zeta} w(\mathbf{i})\left(X_L(\mathbf{i}) - \mu_{\mathbf{x_i},L}\right)\left(Y_L(\mathbf{i}) - \mu_{\mathbf{y_i},L}\right) \tag{12}$$

where $\sigma_{\mathbf{x_i},L}$ and $\sigma_{\mathbf{xy_i},L}$ denote the weighted standard deviations of $\mathbf{x_i}$ and the weighted covariance of $\mathbf{x_i}$ and $\mathbf{y_i}$, respectively. The local lightness mean $\mu_{\mathbf{x_i},L}$ in (11) is estimated by $\mu_{\mathbf{x_i},L} = \sum_{\mathbf{i}\in\zeta} w(\mathbf{i}) X_L(\mathbf{i})$.

Now we derive two lightness-based measures that reflect discriminating achromatic characteristics, the *lightness contrast similarity* $\mathbf{l}_c$ and the *lightness structural similarity*[7] $\mathbf{l}_s$, where each represents the contrast distortion and the loss of linear correlation in lightness channel as follows:

$$\mathbf{l}_c(\mathbf{i}) = \frac{2\sigma_{\mathbf{x_i},L}\sigma_{\mathbf{y_i},L} + K_{L1}}{\sigma^2_{\mathbf{x_i},L} + \sigma^2_{\mathbf{y_i},L} + K_{L1}} \tag{13}$$

$$\mathbf{l}_s(\mathbf{i}) = \frac{|\sigma_{\mathbf{xy_i},L}| + K_{L2}}{\sigma_{\mathbf{x_i},L}\sigma_{\mathbf{y_i},L} + K_{L2}} \tag{14}$$

The K_{L1} and K_{L2} are positive stabilizing constants that are determined experimentally in Section IV-B. Please note that local lightness means, i.e. $\mu_{\mathbf{x_i},L}$ and $\mu_{\mathbf{y_i},L}$, are not explicitly compared, since the HVS is less sensitive to absolute lightness variations than the aforementioned two factors [1], [30].

C. Integration for Combined Quality Metric

In order to derive a single aggregated score representing the overall perceived quality, we need to combine locally estimated similarities, $\mathbf{h}_l$, $\mathbf{h}_c$, $\mathbf{c}_l$, $\mathbf{c}_c$, $\mathbf{l}_c$, and $\mathbf{l}_s$ from the previous stage. In the proposed metric, we initially perform a spatial pooling of each comparison map, followed by a component pooling of each channel score into a final value. This pooling sequence is inherited from [20], since it yields better prediction performance than the pooling in reverse order. To combine local measurements, we exploit a generalized Minkowski framework [31] since it devises a simple approach to assign different weights on the spatial location depending on the degree of local distortions. In other words, the pixel-wise hue mean similarity $\mathbf{h}_l(\mathbf{i})$ is converted into the global hue mean similarity score $\mathbf{H}_l \in [0,1]$, as follows:

$$\mathbf{H}_l = 1 - \left[\frac{1}{|\Omega|}\sum_{\mathbf{i}\in\Omega}\left[1 - \mathbf{h}_l(\mathbf{i})\right]^p\right]^{(1/p)} \tag{15}$$

where Ω denotes a set of coordinates for the entire image. For $p = 1$, (15) is equivalent to the arithmetic mean pooling of local scores [4] and as p increases, more emphasis is given to highly distorted regions. Intuitively, it makes sense as visual attention is attracted to few higher deviations more than many lower ones [31]. For instance, global chrominance shift caused by the scene illumination change will be less perceptible than strong localized memory color deviation. The calculation of scores for other measures, i.e. $\mathbf{H}_c$, $\mathbf{C}_l$, $\mathbf{C}_c$, $\mathbf{L}_c$, and $\mathbf{L}_s$, can be done analogously as in (15) by replacing $\mathbf{h}_l(\mathbf{i})$ with corresponding measures. From them, we can evaluate two scores representing the degree of similarity for chromatic and achromatic components, namely the *chromatic similarity* and *achromatic similarity* scores, as follows:

$$\begin{aligned}\mathbf{S}_C &= \mathbf{H}_l \cdot \mathbf{H}_c \cdot \mathbf{C}_l \cdot \mathbf{C}_c \\ \mathbf{S}_A &= \mathbf{L}_c \cdot \mathbf{L}_s\end{aligned} \tag{16}$$

In (16), individual measures are combined in a multiplicative manner as they are assumed to be relatively independent of each other. Finally, the overall DSCSI score can be obtained by combining two aforementioned scores as follows:

$$\mathcal{Q}(\mathbf{X},\mathbf{Y}) = \mathbf{S}_A \cdot (\mathbf{S}_C)^{\lambda} \tag{17}$$

where the nonnegative parameter λ adjusts the significance of chromatic components in evaluation of overall quality. The λ is introduced since sensitivities of the HVS for variations in chromatic and achromatic components are not equal. $\lambda = 0$ completely ignores the contribution of chromatic components in image quality prediction; as we increase λ, the significance of chromatic components increases. In this paper, λ is fixed at 0.8 by default, which provides consistently stable performance over various distortion types and image contents (See Fig. 9 and Fig. 10 in Section IV-E). This value is consistent with the visual mechanism that the HVS is generally more sensitive to achromatic information than to chromatic information [32].

IV. Experimental Results and Discussion

In this section, we present experimental results to compare the prediction accuracy, robustness, and efficiency of the DSCSI against other representative image quality metrics.

A. Databases and Evaluation Criteria

We incorporate three publicly available databases in our validation, i.e. TID2013 DB [33], LIVE DB [9], and CSIQ DB [10] (Table II), since they are the most commonly used collections in image quality assessment research covering a wide range of commonly encountered distortions in real-world applications. They are annotated with subjective ratings, i.e. MOS or DMOS, facilitating convenient benchmarking of

[6]Please note that the Improved CID (iCID) metric [29] also makes use of this term to measure contrast variations of chroma components.

[7]Note that an absolute value of $\sigma_{\mathbf{xy_i},L}$ is taken to prevent the lightness structural similarity $\mathbf{l}_s(\mathbf{i})$ being negative, and to ensure $\mathcal{Q}(\mathbf{X},\mathbf{Y}) \in [0,1]$.

TABLE II
DESCRIPTION OF EVALUATION DATABASES (N_R/N_T: NUMBER OF ORIGINAL/TEST IMAGES; F: IMAGE FORMAT; R: IMAGE RESOLUTION)

Database	N_R/N_T	F	R	Types of Distortions
TID2013 DB	25 / 3000	8 bit/CH BMP	512x384	24 types of distortions: additive white Gaussian noise (AWGN); AWGN with color; additive Gaussian spatially correlated; masked noise; high frequency; impulse noise; quantization; Gaussian blur; image denoising; JPEG lossy; JPEG2000 lossy; JPEG transmission; JPEG2000 transmission; non-essentricity pattern; local blockwise of different intensity; mean shift; contrast change; change of color saturation; multiplicative Gaussian; comfort noise; lossy compression of noisy images; image color quantization with dither; chromatic aberrations; sparse sampling and reconstruction
LIVE DB	29 / 779	8 bit/CH BMP	480x720 to 768x512	5 types of distortions: AWGN; Gaussian blur; JPEG lossy; JPEG2000 lossy; JPEG2000 transmission;
CSIQ DB	30 / 866	8 bit/CH PNG	512x512	6 types of distortions: AWGN; AWGN with color; Gaussian blur; JPEG lossy; JPEG2000 lossy; contrast change;

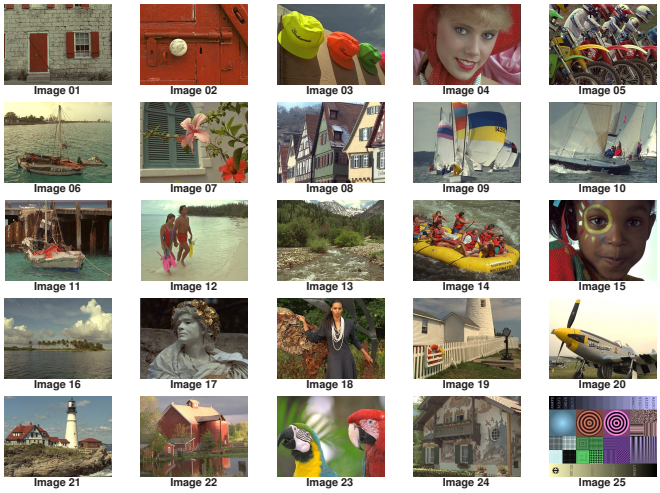


Fig. 4. Set of source images from TID2013 DB [33].

the proposed metric with other algorithms. Test (distorted) images in the aforementioned databases are derived from a set of source images that reflects adequate diversity in color complexity and edge/texture details, ranging from pictures of humans to natural scenes to man-made objects. Some visual examples of source images from TID2013 DB are illustrated in Fig. 4.

TID2013 DB is among the largest databases according to the number of test images and distortion types. In particular, it is useful for our analysis since it contains six types of chromatic distortions: additive white Gaussian noise (AWGN) in color, quantization noise, JPEG compression, color saturation change, color quantization noise with dither, and chromatic aberrations.

In order to evaluate whether a metric is statistically consistent with human perception, validation is performed by comparing metric scores with provided subjective ratings. The following evaluation criteria for image quality metrics are employed: i) Pearson linear correlation coefficient (PLCC), ii) Spearman rank-order correlation coefficient (SRCC), iii) Kendall rank-order correlation coefficient (KRCC).

The PLCC is a correlation coefficient used to measure the prediction accuracy of a metric, i.e. the ability of predicting the subjective ratings with low error. For N pairs of metric and subjective scores (p_i, s_i), PLCC is given by:

$$PLCC = \frac{\sum_{i=1}^{N}(p_i - \overline{p})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(p_i - \overline{p})^2}\sqrt{\sum_{i=1}^{N}(s_i - \overline{s})^2}} \tag{18}$$

where $\overline{p}$ and $\overline{s}$ denote the means of metric scores and subjective scores, respectively. Typically, PLCC is measured after performing logistic regression using the logistic function f :

$$\tilde{p} = f(p) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(p - \beta_3))}\right] + \beta_4 p + \beta_5 \tag{19}$$

where p is the metric score, $\tilde{p}$ is the mapped output, and $\{\beta_1, \cdots, \beta_5\}$ are the parameters to be fitted by minimizing the sum of squared differences between $\tilde{p}$ and the subjective score. This function removes non-linearity of the subjective rating process [9]. The SRCC measures the prediction monotonicity, i.e. the degree to which a metric agrees with the rank of the subjective ratings, and is defined as:

$$SRCC = 1 - \frac{6\sum_{i=1}^{N}[R(p_i) - R(s_i)]^2}{N(N^2 - 1)} \tag{20}$$

where $R(p_i)$ and $R(s_i)$ represent ranks of the prediction score p_i and the subjective score s_i, respectively, and N is the number of score sets. The KRCC is also a correlation coefficient to measure the prediction monotonicity, defined as:

$$KRCC = \frac{N_c - N_d}{N(N-1)/2} \tag{21}$$

where N_c and N_d denote the numbers of concordant and discordant pairs in scores, respectively.[8] Unlike PLCC, both SRCC and KRCC only take into account the rank of the score and neglect the relative distance between scores. All three coefficients vary between -1 and 1, and the fidelity of an objective metric is considered high if they are close to unity.

B. Configuration of Default Parameters

The DSCSI metric contains several parameters to be adjusted in order to accurately represent the HVS. The best way to determine parameters is to conduct psychophysical experiments and optimize them accordingly. However, we found that it is difficult to collect sufficient data to facilitate reliable estimation of free parameters. Instead, we fit the parameters in order to optimize the metric performance on existing databases.

[8]Let p_i and s_i $(i = 1, \cdots, N)$ are scores obtained from objective metrics and subjective tests, respectively. Then, for $i \neq j$, if $sign(p_j - p_i) = sign(s_j - s_i)$, the pair is called concordant, otherwise the pair is called discordant.

TABLE III
DEFAULT PARAMETERS OF THE DSCSI METRIC

	K_H	K_{C1}	K_{C2}	K_{L1},K_{L2}	h_0	τ
Equation index	(7)	(9)	(10)	(13),(14)	(5)	(5)
Default value	0.0008	0.0008	16	0.8	0.2π	0.35

1) Optimization of DSCSI Parameters: The values of parameters K_H, K_{C1}, K_{C2}, K_{L1}, and K_{L2} of the DSCSI are determined based on the threshold of perception and the dynamic range of corresponding terms [4]. Several assumptions are made for tuning of the parameters. First, we assume $K_H < K_{L1} < K_{C2}$, which are used to quantify contrast distortion of hue, lightness, and chroma components, respectively. The reason behind this relationship is that the main factor to determine their values is the dynamic range of the corresponding terms [4]. K_H is the smallest among the three since the hue variance descriptor always lies within the range of $[0, 1]$. In general, for natural images, we found that the dynamic range of chroma standard deviation σ_C is larger than the one of lightness standard deviation σ_L. Thus, the assumption $K_H < K_{L1} < K_{C2}$ holds. In addition, we assume $K_{L1} = K_{L2}$ for simplicity.

A subset of TID2013 DB is used to train the free parameters of the DSCSI. Inheriting the experimental setup from [16], we use eight reference images and associated 1000 distorted images in TID2013 DB as training data. The parameters are optimized by maximizing SRCC on the training set[9]. We believe that training using TID2013 DB will allow us to extend the applicability of the metric to more general applications as it covers a wide range of distortion types. As a result, the optimal parameters for the DSCSI are determined as in Table III. In the rest of this paper, the DSCSI is calibrated using the default parameters provided as in Table III.

2) Configuration of Viewing Condition Parameters: Apart from the parameters discussed in the previous section, the CSF filter of S-CIELAB (discussed in Section III-A) is configured according to the viewing condition of evaluation data. In this work, we adopted the CSF implementation of spatial domain, which uses two-dimensional separable convolution kernels. For details of CSF implementation, we refer the reader to [34].

Based on given experimental conditions, the effective visual resolution r is computed in pixels per degree of visual angle (p/deg) for individual databases as follows [35]:

$$r = dv \tan\left(\frac{\pi}{180}\right) \tag{22}$$

where v denotes the viewing distance in the units of distance, and d denotes the display resolution in pixels per unit distance. Table IV summarizes the computed viewing distance parameters for individual databases.

C. Significance of Each Similarity Measure of DSCSI

In this section, the single and joint contributions of individual similarity measures (i.e. $\mathbf{H}_l$, $\mathbf{H}_c$, $\mathbf{C}_l$, $\mathbf{C}_c$, $\mathbf{L}_c$, and $\mathbf{L}_s$) on the performance of quality prediction are discussed. In this experiment, each measure is either used or unused in the computation of a final quality value $\mathcal{Q}(\mathbf{X}, \mathbf{Y})$ exploiting (16) and (17). Unused individual measure is set to 1, so it does not affect the quality score. This results in $2^6 = 64$ ways to derive a quality score from six measures. Fig. 5 shows the SRCC performance between subjective scores and metric values with various combinations of measures against evaluation datasets. Note that TID2013 DB is decomposed into two subsets, one containing images of chromatic distortions and the other containing images of achromatic distortions to examine how the metric behaves on either distortion type. In the rest of this paper, we refer the former as TID2013(C) and the latter as TID2013(NC). We make the following major observations:

- For the LIVE and CSIQ databases, the most effective measure for quality prediction is the lightness contrast $\mathbf{L}_c$. This result is in accordance with [20], where the CID is examined on a set of gamut mapped images. In fact, solely $\mathbf{L}_c$ provides nearly the maximum achievable prediction performance for TID2013(NC), LIVE DB, and CSIQ DB. For TID2013 DB and TID2013(C), the chroma contrast $\mathbf{C}_c$ turns out to be the most informative measure, while $\mathbf{L}_c$ is inferior to $\mathbf{H}_l$ and $\mathbf{C}_l$, implying relatively increased importance of chromatic measures.
- For the LIVE and CSIQ databases, the achromatic similarity, $\mathbf{S}_A = \mathbf{L}_c\mathbf{L}_s$, yields superior prediction accuracy to the chromatic similarity, $\mathbf{S}_C = \mathbf{H}_l\mathbf{H}_c\mathbf{C}_l\mathbf{C}_c$. It is expected since both databases lack images exhibiting chromatic distortions. For TID2013(C), $\mathbf{S}_C$ performs substantially better than $\mathbf{S}_A$. This fact again underlines the significance of the proposed chromatic measures in dealing with image artifacts that involve chromatic components.
- In general, for the LIVE and CSIQ databases, $\mathbf{L}_c\mathbf{L}_s$ yields nearly the maximum achievable SRCC, and adding more chromatic measures do not provide significant complementary effect and even affects adversely in some cases. However, complementary effects of chromatic measures are more apparent for TID2013 DB. One possible explanation for an adverse effect of some chromatic measures on LIVE and CSIQ databases is that the default value of λ is not optimal for these two databases. Such undesirable negative effects can be minimized by decreasing the λ parameter in (17).
- Fig. 5 justifies the combinational use of all six measures for accurate quality prediction, since aggregating them yields the best SRCC performance on TID2013 DB, which contains the most distortion types, while maintaining reasonable performance on LIVE and CSIQ

TABLE IV
VIEWING CONDITION PARAMETERS FOR EVALUATION DATABASES

Database	Display characteristic	Display resolution (d)	Viewing distance (v)	Visual resolution (r)
TID2013	19 inch screen with 1152x864 screen resolution	75.79 pixels per inch (ppi)	Not specified. Assumed to be 2.5 times screen heights	36.7 p/deg
LIVE	21 inch screen with 1024x768	60.95 ppi	2-2.5 times screen heights	30.2 p/deg
CSIQ	24 inch screen with 1920x1200	94.34 ppi	70 cm	45.4 p/deg

[9]Please note that a set of free parameters optimized using PLCC is very close to the one obtained using SRCC.

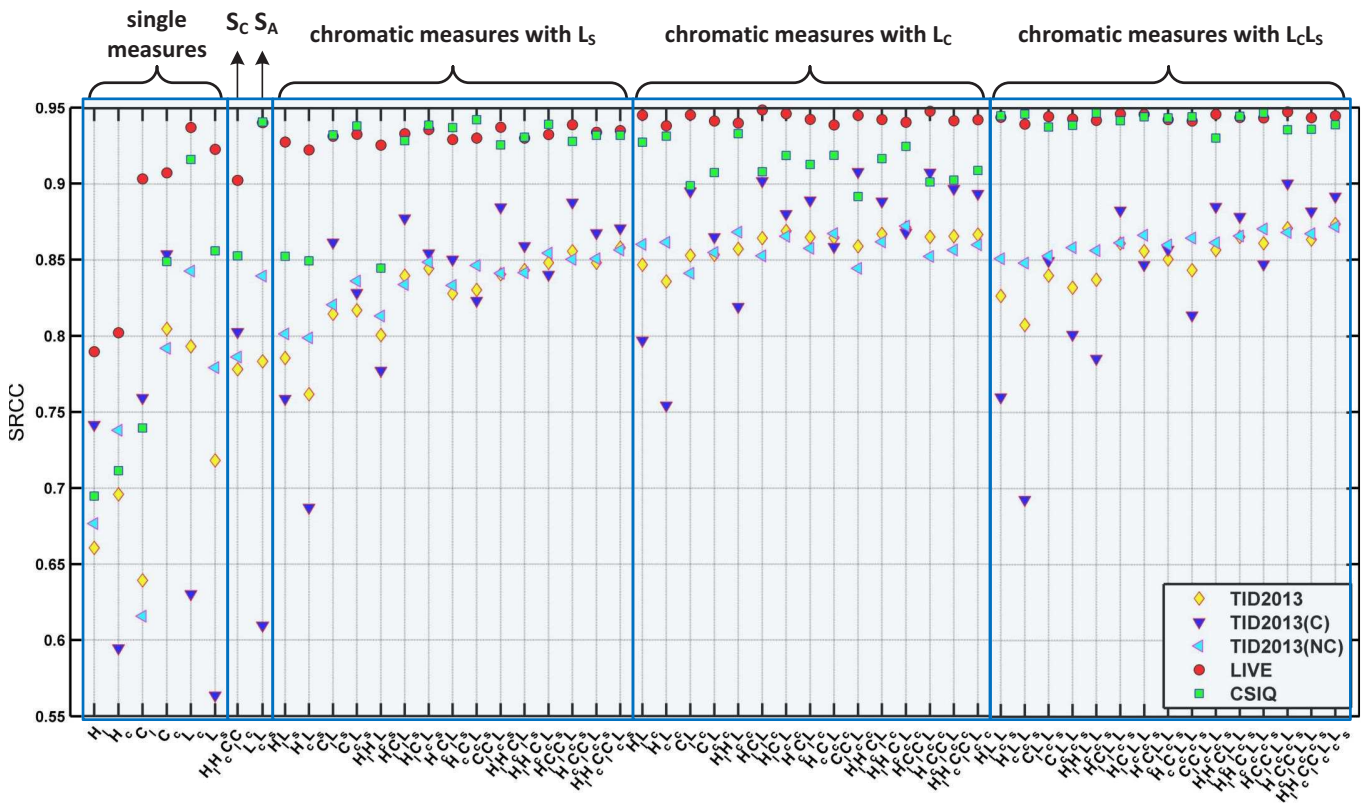


Fig. 5. Performance of the DSCSI with various measure combinations. TID2013(C) denotes the subset of TID2013 DB containing images of chromatic distortions, while TID2013(NC) denotes the subset of achromatic distortions. SRCC closer to 1 indicates the more accurate predictor of subjective assessments.

databases, i.e. the SRCC values for $\mathbf{H}_l\mathbf{H}_c\mathbf{C}_l\mathbf{C}_c\mathbf{L}_c\mathbf{L}_s$ are not smaller than 99% of the highest possible SRCC for both databases. It indicates that local color descriptors used in the DSCSI provide complementary information for quality perception and, together, improve the accuracy of prediction model.

D. Significance of Contrast Sensitivity Filtering of DSCSI

This section briefly discusses the importance of viewing distance normalization step in the proposed metric. In Table V, the prediction accuracy of DSCSI is reported on validation databases with and without the CSF filtering embedded in S-CIELAB[10]. A simple pre-scaling routine proposed by Wang et al. [36][11] is applied for input images when the CSF filtering is disabled. Table V shows that the prediction accuracy of DSCSI is consistently improved for all three databases by applying the CSF filtering. Significant performance gain reported on TID2013(C) indicates that CSF filtering is essential, especially for handling chromatic distortion types. Overall, viewing distance normalization by incorporating CSF filtering is a necessary pre-processing step to ensure the optimal performance of the DSCSI, and it justifies the use of S-CIELAB as a basis color space.

TABLE V
SRCC PERFORMANCE OF THE PROPOSED METRIC ON BENCHMARK DATABASES WITH AND WITHOUT THE CSF FILTERING

Database	DSCSI (CSF enabled)	DSCSI (CSF disabled)
TID2013 DB	0.8738	0.7510
TID2013(C)	0.8919	0.6076
TID2013(NC)	0.8722	0.7963
LIVE DB	0.9450	0.9021
CSIQ DB	0.9392	0.8763

E. Combined Metric Performance on Benchmark Databases

1) Overall Prediction Accuracy: This section compares the prediction performance of the proposed DSCSI with other representative metrics in the literature. Metrics in comparison are: i) the SSIM [4], ii) the MS-SSIM [5], iii) the intra-frame quality assessment of Video SSIM (VSSIM) [15], iv) the UCIF [14], v) the color extension of the FSIM (FSIMc) [16], vi) the CID [20], vii) the Euclidean distance in CIELAB (CIEDE), and viii) the CIEDE2000 difference equation (DE2000) [13]. Implementations of most metrics are publicly available except for the VSSIM and the UCIF which are self-implemented. To guarantee a fair comparison, we use the default parameters configured by the original authors for each metric.

[10]Essentially, disabling the CSF filtering implies that the proposed metric operates in CIELAB color space.

[11]It applies $F \times F$ averaging and down-sampling by F sequentially on the input image. The scaling factor F is determined by $F = \max(1, round(N/256))$, where N is the image height. This routine is widely used in the practical implementation of existing image quality metrics, such as the SSIM and the FSIM.

TABLE VI
PERFORMANCE OF METRICS ON BENCHMARK DATABASES (TID2013(C) INDICATES THE SUBSET OF TID2013 DB CONTAINING IMAGES OF CHROMATIC DISTORTIONS). TWO BEST METRICS ARE BOLDED. THE VALUE CLOSER TO 1 INDICATES THE MORE ACCURATE PREDICTOR OF SUBJECTIVE RATINGS

	TID2013(C)			**TID2013 DB**			**LIVE DB**			**CSIQ DB**		
Metrics	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
SSIM	0.7882	0.5509	0.4303	0.7895	0.7417	0.5588	0.9449	0.9479	0.7963	0.8612	0.8755	0.6900
MSSSIM	0.7952	0.5556	0.4376	0.8329	0.7859	0.6047	**0.9489**	**0.9513**	**0.8044**	0.8991	0.9132	0.7386
VSSIM	0.7632	0.6178	0.4564	0.8032	0.7752	0.5825	0.9342	0.9371	0.7749	0.9021	0.9170	0.7376
UCIF	0.7303	0.6724	0.4794	0.7019	0.6235	0.4511	0.8859	0.8695	0.6761	0.8304	0.8128	0.6117
FSIMc	0.8449	0.7751	0.5918	**0.8769**	0.8510	**0.6665**	**0.9613**	**0.9645**	**0.8363**	**0.9191**	**0.9309**	**0.7684**
CID	0.8638	0.8517	0.6504	0.8044	0.7879	0.5867	0.9279	0.9270	0.7571	0.9079	0.9116	0.7283
CIEDE	0.6429	0.6577	0.4591	0.4246	0.3919	0.2620	0.7639	0.7562	0.5568	0.7142	0.7271	0.5140
DE2000	0.6317	0.6392	0.4461	0.4237	0.3861	0.2584	0.7551	0.7483	0.5493	0.7162	0.7255	0.5114
DSCSI(h0p1)	0.8825	0.8720	0.6741	0.8214	0.8167	0.6200	0.9171	0.9208	0.7467	0.9000	0.9109	0.7284
DSCSI(h0p2)	0.8662	0.8558	0.6551	0.8548	**0.8572**	0.6577	0.9171	0.9163	0.7378	0.9088	0.9090	0.7269
DSCSI(h1p1)	**0.9010**	**0.8875**	**0.7029**	0.8288	0.8246	0.6304	0.9301	0.9368	0.7751	0.9083	0.9246	0.7501
DSCSI(h1p2)	**0.9111**	**0.8919**	**0.7130**	**0.8765**	**0.8738**	**0.6839**	0.9384	0.9450	0.7916	**0.9297**	**0.9392**	**0.7738**

TABLE VII
STATISTICAL SIGNIFICANCE TEST ON EVALUATION DATABASES. "1" AT LOCATION (i, j) INDICATES METHOD i IS SIGNIFICANTLY BETTER THAN j AT 95% CONFIDENCE LEVEL. "-1" INDICATES THE OPPOSITE AND "0" INDICATES THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN TWO METHODS

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
01. SSIM	0	0	0	1	-1	-1	1	1	-1	-1	-1	-1
02. MSSSIM	0	0	0	1	-1	-1	1	1	-1	-1	-1	-1
03. VSSIM	0	0	0	0	-1	-1	1	1	-1	-1	-1	-1
04. UCIF	-1	-1	0	0	-1	-1	1	1	-1	-1	-1	-1
05. FSIMc	1	1	1	1	0	0	1	1	-1	0	-1	-1
06. CID	1	1	1	1	0	0	1	1	0	0	-1	-1
07. CIEDE	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
08. DE2000	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
09. DSCSI(h0p1)	1	1	1	1	1	0	1	1	0	0	-1	-1
10. DSCSI(h0p2)	1	1	1	1	0	0	1	1	0	0	-1	-1
11. DSCSI(h1p1)	1	1	1	1	1	1	1	1	1	1	0	0
12. DSCSI(h1p2)	1	1	1	1	1	1	1	1	1	1	0	0

(a) TID2013(C)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
01. SSIM	0	-1	0	1	-1	0	1	1	-1	-1	-1	-1
02. MSSSIM	1	0	1	1	-1	1	1	1	0	-1	0	-1
03. VSSIM	0	-1	0	1	-1	0	1	1	-1	-1	-1	-1
04. UCIF	-1	-1	-1	0	-1	-1	1	1	-1	-1	-1	-1
05. FSIMc	1	1	1	1	0	1	1	1	1	1	1	0
06. CID	0	-1	0	1	-1	0	1	1	-1	-1	-1	-1
07. CIEDE	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
08. DE2000	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
09. DSCSI(h0p1)	1	0	1	1	-1	1	1	1	0	-1	0	-1
10. DSCSI(h0p2)	1	1	1	1	-1	1	1	1	1	0	1	-1
11. DSCSI(h1p1)	1	0	1	1	-1	1	1	1	0	-1	0	-1
12. DSCSI(h1p2)	1	1	1	1	0	1	1	1	1	1	1	0

(b) TID2013 DB

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
01. SSIM	0	0	1	1	-1	1	1	1	1	1	1	0
02. MSSSIM	0	0	1	1	-1	1	1	1	1	1	1	1
03. VSSIM	-1	-1	0	1	-1	0	1	1	1	1	0	0
04. UCIF	-1	-1	-1	0	-1	-1	1	1	-1	-1	-1	-1
05. FSIMc	1	1	1	1	0	1	1	1	1	1	1	1
06. CID	-1	-1	0	1	-1	0	1	1	0	0	0	-1
07. CIEDE	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
08. DE2000	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
09. DSCSI(h0p1)	-1	-1	-1	1	-1	0	1	1	0	0	-1	-1
10. DSCSI(h0p2)	-1	-1	-1	1	-1	0	1	1	0	0	-1	-1
11. DSCSI(h1p1)	-1	-1	0	1	-1	0	1	1	1	1	0	0
12. DSCSI(h1p2)	0	-1	0	1	-1	1	1	1	1	1	0	0

(c) LIVE DB

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
01. SSIM	0	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1
02. MSSSIM	1	0	0	1	-1	0	1	1	0	0	0	-1
03. VSSIM	1	0	0	1	-1	0	1	1	0	0	0	-1
04. UCIF	-1	-1	-1	0	-1	-1	1	1	-1	-1	-1	-1
05. FSIMc	1	1	1	1	0	0	1	1	1	0	0	-1
06. CID	1	0	0	1	0	0	1	1	0	0	0	-1
07. CIEDE	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
08. DE2000	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
09. DSCSI(h0p1)	1	0	0	1	-1	0	1	1	0	0	0	-1
10. DSCSI(h0p2)	1	0	0	1	0	0	1	1	0	0	0	-1
11. DSCSI(h1p1)	1	0	0	1	0	0	1	1	0	0	0	-1
12. DSCSI(h1p2)	1	1	1	1	1	1	1	1	1	1	1	0

(d) CSIQ DB

As a basis for comparisons, two representative grayscale metrics, the SSIM and MS-SSIM, are applied on luminance component Y of input images, which are computed by $Y = 0.299R + 0.587G + 0.114B$ (ITU-R BT.601). It should be noted that although VSSIM is originally developed for video data, its intra-frame quality assessment mechanism (i.e. an estimation of SSIM indices for individual YCbCr color channels of input images, followed by a linear combination of channel indices into a final quality score) constitutes a practical color image quality metric. Therefore, the intra-frame assessment of VSSIM is considered in our comparison. The VSSIM and UCIF are implemented with an ad-hoc pre-scaling routine proposed by Wang et al. [36] for their optimal performance. The FSIMc and CID are considered in the analysis because of their leading performance against conventional distortions and color gamut mapping distortions, respectively. In addition, we include CIE equations in our analysis since they are well-known standards for industrial applications. For CIE equations, pixel-wise color differences $\Delta E(\mathbf{i})$ are initially computed then the overall difference is obtained by taking $\Delta\overline{E} = \frac{1}{|\Omega|}\sum_{\mathbf{i}\in\Omega}\Delta E(\mathbf{i})$, where Ω denotes a set of pixel coordinates for entire image.

Table VI allows us to compare metrics on the entire set of images from three databases. To examine how metrics behave particularly on chromatic distortions, we separately considered the subset of TID2013 DB containing images of chromatic distortions. The performance of the DSCSI is reported with four different configurations: i) by either disabling or enabling the hue significance weighting $\alpha(\mathbf{i})$ (denoted as $h0$ and $h1$); ii) by setting the spatial pooling parameter p in (15), with either 1 or 2 (denoted as $p1$ and $p2$)[12]; in order to demonstrate the impact of spatial weighting schemes on the prediction accuracy. It is evident from Table VI that the DSCSI consistently performs better with the hue significance weighting than without it. Such consistent improvements are strong indications of the reliability of the hue significance weighting $\alpha(\mathbf{i})$.

[12]We only consider up to $p = 2$ as the prediction accuracy starts to degrade with $p > 2$.

Considering the parameter p, it is found that the DSCSI($p2$) outperforms the DSCSI($p1$) across three databases in terms of the prediction accuracy when the hue significance weighting is enabled. Some contradictory results reported with $h0$, e.g. the DSCSI($h0p2$) performs worse than the DSCSI($h0p1$) on LIVE DB, are caused by overestimation of hue distortion detected on local regions of achromatic pixels. In general, the DSCSI($h1p2$) maintains great performance on generic distortions as it performs the best on CSIQ DB and the entire TID2013 DB. On the LIVE DB, which is composed mainly of images exhibiting achromatic distortions, the DSCSI is outperformed by the FSIMc and two grayscale metrics.

We also perform a statistical significance test to demonstrate whether the difference between metric performance is significant or not. Assuming that prediction errors are distributed according to Gaussian distribution, a two-tailed F-test [9] is conducted[13]. Let v_A and v_B be the sample variances of the prediction errors from metrics A and B, respectively, and the test statistic F is defined as $F = v_A/v_B$. The metric A has significantly greater errors than the metric B if $F > F_{upper}$, while A has significantly lower errors than B for $F < F_{lower}$, where F_{upper} and F_{lower} are two critical values selected based on the number of distorted images and the confidence level. If $F_{lower} < F < F_{upper}$, we accept the null hypothesis that the errors from both metrics are statistically indistinguishable.

Table VII indicates that the DSCSI($h1p2$) is either the best or tied for the best performer in a statistical sense at 95% confidence level, with the exception of the MS-SSIM and FSIMc on LIVE DB. The suboptimal performance of the DSCSI on LIVE DB is contributed to the negative effect of some chromatic measures as illustrated in Fig. 5. Using CIE equations significantly decreases the correlation with the human observers in all databases. It is understandable as they compare images pixel-by-pixel without considering the spatial context between pixels. Simple extensions of the SSIM into color channels, e.g. the VSSIM and UCIF, generally offer limited gain compared to the SSIM and are outperformed by more elaborate metrics, i.e. the CID and FSIMc. Table VII also shows that the DSCSI with the hue significance weighting, i.e. $h1p1$ and $h1p2$, consistently outperforms or performs equally accurate as the CID across databases regardless of p values.

The performance of the DSCSI($h1p2$) is further compared with the CID using a visual example in Fig. 6. In this example, a source image from TID2013 DB is degraded with three different types of distortions, and for each distorted image, we derive objective scores as well as comparison maps using both metrics. Please note that the comparison maps for both metrics are acquired by a pixel-wise combination of comparison measures as suggested by the SSIM index. Fig. 6 indicates that the ranking of the DSCSI($h1p2$) is more consistent with the MOS scores, as the CID underestimates the visual degradation caused by chromatic aberration on the sample distorted images. Given that two metrics are derived from a similar underlying principle, i.e. the HVS is sensitive to hue, chroma, and lightness distortions, there are two possible

[13]If prediction errors are not Gaussian distributed, the significance test becomes inconclusive [10]. A Jarque-Bera (JB) test can be performed to examine whether the samples are Gaussian distributed or not.

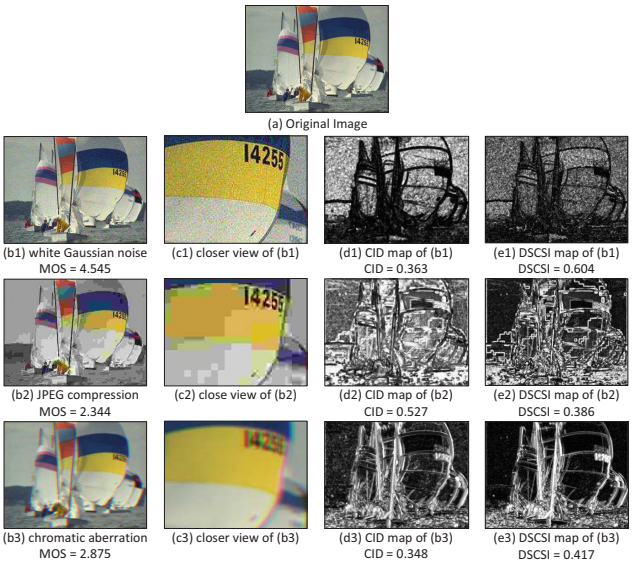


Fig. 6. Performance comparison between the CID and the DSCSI against three distortion types from TID2013 DB. Higher CID score corresponds to lower MOS as it measures the perceived differences. Brighter gray level in the map indicates higher distortion.

explanations for the relative robustness of the DSCSI (shown statistically in Table VI and visually in Fig. 6): i) the DSCSI properly handles discriminative hue information by properly taking into account the periodicity of hue; and ii) the CID is optimized mainly for gamut mapping distortion, while the DSCSI is optimized for more generic distortions.

Fig. 7 shows scatter plots of some metrics of interest on TID2013 DB. The more concentrated distribution of points near the fitted curve is, the closer the metric complies with subjective ratings. A large number of points in the vicinity of $x = 1$ for the SSIM and MS-SSIM points out that grayscale metrics tends to underestimate distortions in chromatic components (example shown in Fig. 1). Comparably poor correlation between CIE equations and visual perception is again shown in Fig. 7 as points deviate considerably from the curve. On the other hand, objective scores predicted by the DSCSI correlate more consistently with the subjective ratings for both classes of distortions, without generating significant outliers. Fig. 7 also demonstrates that the DSCSI, especially with $p = 2$, results in a wider range of scores between 0 and 1 than other metrics, exhibiting almost a linear relationship with the MOS. Such linearity of a metric relative to subjective rating is desirable for straightforward interpretation of metric scores [37].

2) Individual Distortion Analysis: To further examine the robustness of the proposed metric, the performance on individual distortion type is shown in Table VIII, where the three best metrics are bolded for each type. This analysis is meaningful since a good general-purpose metric should provide stable prediction accuracy across different types of image distortions. For performance indicator, we use the SRCC as it is not influenced by the fitting. Note that the hue significance weighting is enabled by default for the DSCSI in this section. The DSCSI performs generally well on conventional distortions, e.g. image compression, AWGN. In particular, its effectiveness on the

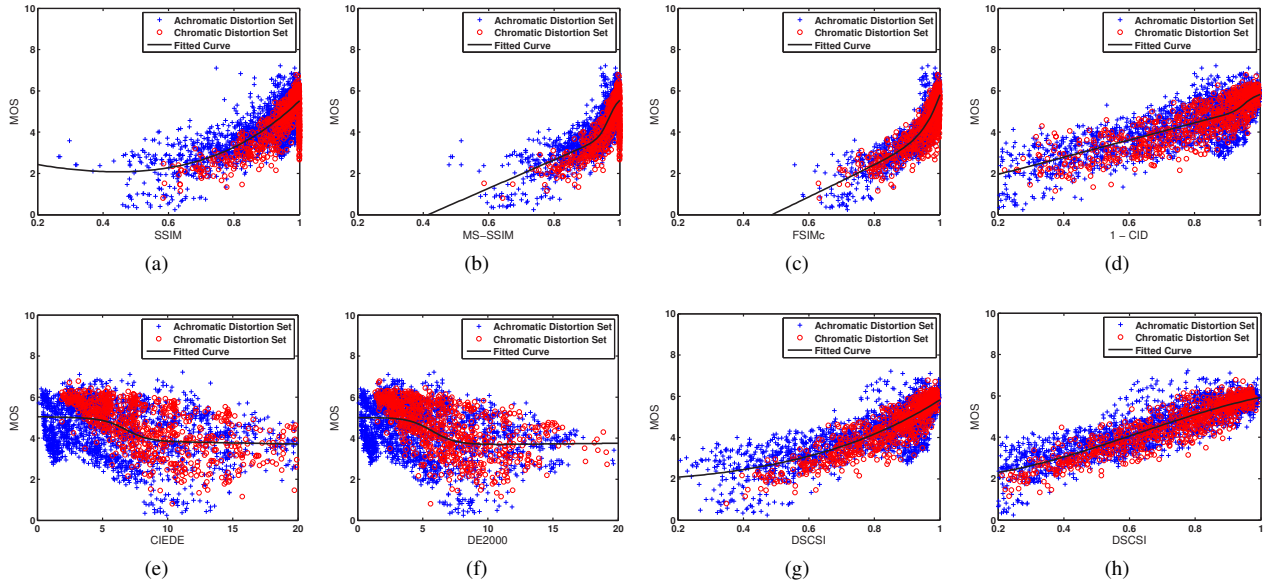


Fig. 7. Scatter plots of subjective MOS versus predicted scores on TID2013 DB. Chromatic and achromatic distortions are represented by different symbols. (a) SSIM. (b) MS-SSIM. (c) FSIMc. (d) CID. (e) CIEDE. (f) DE2000. (g) DSCSI(h1p1). (h) DSCSI(h1p2).

TABLE VIII
SRCC VALUE OF METRICS FOR INDIVIDUAL DISTORTION TYPE. CHROMATIC DISTORTION TYPES OF TID2013 DB ARE INDICATED WITH (C). THE LAST ROW SHOWS THE NUMBER OF TIMES SPECIFIC METRIC YIELDS SRCC OF LARGER THAN 0.95, I.E. HIGHLY CORRELATED WITH SUBJECTIVE SCORES

Database	Distortion	SSIM	MSSSIM	VSSIM	UCIF	FSIMc	CID	CIEDE	DE2000	DSCSI(p1)	DSCSI(p2)
TID2013	Awgn	0.8671	0.8646	0.9092	0.7409	0.9101	0.7854	**0.9226**	**0.9125**	0.8917	**0.9275**
	Awgn-color (C)	0.7726	0.7730	0.8426	0.6661	**0.8537**	0.7384	**0.8748**	**0.8592**	0.8222	0.8530
	Spatially correlated noise	0.8515	0.8544	0.9015	0.7387	0.8900	0.7878	**0.9222**	**0.9088**	0.8968	**0.9318**
	Masked noise	0.7767	0.8073	**0.8365**	0.6621	**0.8094**	0.7901	**0.8193**	0.7899	0.8074	0.7834
	High frequency noise	0.8634	0.8604	0.8911	0.7957	**0.9040**	0.8290	**0.9349**	**0.9344**	0.8840	0.9031
	Impulse noise	0.7503	0.7629	0.8246	0.6554	0.8251	0.7951	**0.9220**	**0.9266**	**0.8538**	0.8476
	Quantization noise (C)	0.8657	0.8706	**0.8750**	0.7619	**0.8807**	0.8267	0.8555	0.8140	0.8745	**0.9043**
	Gaussian blur	**0.9668**	**0.9673**	**0.9671**	0.9017	0.9551	0.9474	0.9151	0.9171	0.9597	0.9541
	Denoising	0.9254	0.9268	**0.9365**	0.8606	**0.9330**	0.8818	0.9057	0.8718	0.9237	**0.9378**
	Jpeg compression (C)	0.9200	0.9265	**0.9401**	0.8581	0.9339	0.8993	0.9365	**0.9411**	0.9374	**0.9400**
	Jpeg2000 compression	0.9468	0.9504	**0.9550**	0.8783	**0.9589**	0.9329	0.9192	0.9214	0.9518	**0.9571**
	Jpeg-trans-errors	0.8493	0.8475	**0.8805**	0.8148	0.8610	**0.8888**	0.7413	0.7873	0.8783	**0.8826**
	Jpeg2000-trans-errors	0.8828	0.8889	**0.8915**	0.7539	**0.8919**	0.8292	0.8210	0.8371	0.8895	**0.9052**
	Non-essentricity pattern	0.7821	0.7968	0.7892	0.7774	0.7937	**0.8032**	0.7899	0.7671	**0.8013**	**0.8200**
	Block distortion	0.5720	0.4801	**0.5849**	0.5582	0.5532	**0.5849**	0.3861	0.5641	**0.6148**	0.5700
	Mean shift	0.7752	**0.7906**	0.7876	0.7437	0.7487	**0.8051**	**0.7917**	0.7827	0.7685	0.7445
	Contrast change	0.3775	**0.4634**	**0.4618**	0.3438	**0.4679**	0.4506	0.3907	0.4577	0.4319	0.3548
	Color saturation change (C)	-0.4141	-0.4099	**0.8351**	0.7909	**0.8359**	**0.8243**	0.6944	0.8151	0.7217	0.8072
	Multiplicative Gaussian	0.7803	0.7786	0.8437	0.6472	0.8569	0.6631	**0.8918**	**0.8675**	0.8103	**0.8598**
	Comfort noise	0.8566	0.8528	0.8697	0.8116	**0.9135**	0.8238	**0.9126**	0.8821	0.8834	**0.9116**
	Lossy compression	0.9057	0.9068	0.9199	0.8583	**0.9485**	0.8731	**0.9534**	0.9284	0.9316	**0.9624**
	Color quantization (C)	0.8542	0.8555	0.8689	0.7761	**0.8815**	0.8499	0.8636	0.8611	**0.8863**	**0.9117**
	Chromatic aberration (C)	0.8775	**0.8784**	**0.9115**	0.8518	**0.8925**	0.8360	0.7023	0.7026	0.8592	0.8352
	Sparse sampling	0.9461	0.9483	**0.9507**	0.8969	**0.9576**	0.9350	0.9117	0.8950	0.9452	**0.9581**
LIVE	Jpeg2000 compression	0.9614	**0.9627**	0.9577	0.9347	**0.9724**	0.9547	0.8203	0.8197	0.9608	**0.9671**
	Jpeg compression	**0.9764**	**0.9815**	0.9711	0.9526	**0.9840**	0.9715	0.9105	0.9164	0.9699	0.9714
	Awgn	0.9694	0.9733	**0.9819**	0.9283	0.9716	0.9534	**0.9845**	**0.9797**	0.9737	0.9793
	Gaussian blur	0.9517	0.9542	0.9489	0.9626	**0.9708**	**0.9691**	0.7701	0.7758	0.9539	**0.9648**
	Fast fading	0.9556	0.9471	0.9534	**0.9565**	0.9519	0.9497	0.7776	0.7787	**0.9603**	**0.9595**
CSIQ	Awgn	0.8974	0.9471	**0.9636**	0.7530	0.9360	0.8914	0.9382	0.9188	**0.9589**	**0.9692**
	Jpeg compression	0.9543	**0.9631**	0.9516	0.9253	**0.9662**	**0.9556**	0.9194	0.9094	0.9521	0.9543
	Jpeg2000 compression	0.9605	**0.9682**	0.9548	0.8953	**0.9703**	0.9517	0.9159	0.9134	0.9565	**0.9703**
	Gaussian pink noise	0.8924	0.9330	**0.9574**	0.7701	0.9369	0.9419	0.9462	0.9392	**0.9585**	**0.9541**
	Gaussian blur	0.9608	**0.9711**	0.9582	0.9239	**0.9728**	**0.9680**	0.9216	0.9191	0.9575	0.9669
	Contrast change	0.7925	**0.9528**	0.9172	0.8714	**0.9439**	**0.9443**	0.8957	0.9291	0.9260	0.9136
	#(SRCC $>$ 0.95)	9	10	12	3	11	7	2	1	12	14

JPEG2000 compression implies that the DSCSI effectively handles coding artifacts, e.g. washed-out color (by chroma sub-sampling), ringing, and blurring. However, the DSCSI yields limited accuracy on the contrast change set in TID2013 DB. It is because this set contains both contrast-enhanced and contrast-reduced images and the DSCSI is not always successful in handling contrast-enhanced images[14].

Interestingly, metrics performing well on many single-type

[14]Often, contrast enhancement in image data results in an improvement of perceptual quality. Thus, larger MOS scores are given for some highly contrast-enhanced images in TID2013 DB, and all metrics in the comparison yield suboptimal performance on this set (i.e. $SRCC < 0.5$).

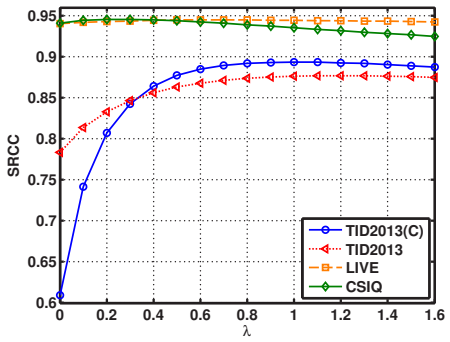


Fig. 8. Variation of SRCC for different λ values of the DSCSI($h1p2$).

distortions are not necessarily effective across a wide range of distortion types. For example, the performance of CIEDE is highly dependent on distortions types, as it performs quite well on specific types, e.g. AWGN and mean shift, but it yields limited accuracy on others, leading to overall limited prediction performance on the entire database (See Table VI). On the other hand, although the DSCSI is not always ranked as the best performer in individual chromatic distortions, its consistently stable performance across them makes it the best quality metric for handling various chromatic distortions as shown in Table VI. Overall, the DSCSI is effective when image distortions are caused by a single factor since it shows the highest number of occurrences that its SRCC is above 0.95.

The DSCSI can be tailored for a specific distortion by tuning its relevant parameters. For instance, making the metric more sensitive to a small change in chromatic contrast leads to an increased prediction accuracy on chromatic aberration since the main artifacts induced by it are unnaturally colored (or blurred) edges. Thus, SRCC performance of DSCSI($p2$) on the chromatic aberration set of TID2013 DB can be enhanced from 0.8352 to 0.8617 by adjusting K_H and K_{C2} parameters to 0.5×10^{-5} and 0.1, respectively, from their default values, which make the hue dispersion measure $\mathbf{H}_c$ and the chroma contrast measure $\mathbf{C}_c$ more responsive to edge distortions.

3) Parameter Sensitivity: Recall from Section III-C that the relative significance of chromatic/achromatic components in the DSCSI is controlled by the λ parameter in (17). In Fig. 8, we present the prediction performance of the DSCSI($h1p2$) depending on the selection of λ. The following conclusions can be drawn from Fig. 8:

- The prediction accuracy deteriorates substantially as we decrease λ to zero on TID2013(C). This is expected because the contribution of chromatic components to the quality perception is underestimated with small λ values.
- The prediction accuracy is relatively insensitive to the choice of λ for the LIVE and CSIQ databases, implying that achromatic measures play a primary role in quality prediction. On the other hand, the prediction accuracy is highly dependent on λ for TID2013 DB. Although the optimal λ varies across databases, when λ is in the range $[0.7, 0.9]$, SRCC values are very close to (no smaller than 99% of) the highest possible values for all three databases. Hence, we confirm that $\lambda = 0.8$ is a reasonable database-independent value.

In addition, Fig. 9 demonstrates SRCC performance of the DSCSI on various distortion types as a function of λ. This analysis provides a guideline for tuning the λ parameter when the metric is applied for a predefined type of distortion. It shows that, in general, DSCSI maintains reasonable performance against individual distortions when $\lambda = 0.8$. However, for certain distortion types, e.g. block distortion and saturation change, the λ may need to be adjusted since the prediction accuracy is highly dependent on its value. Lastly, the image content dependency of DSCSI performance is investigated in Fig. 10 by varying λ values on 25 individual source images of TID2013 DB. The SRCC curves for all source images exhibit nearly similar patterns except for "*image 25*". Please note that "*image 25*" is distinguished from the other 24 natural scene images in TID2013 DB since it contains an artificially generated scene of complex texture and edge details. Nevertheless, regardless of natural or artificial origin, the default λ offers a good generalization ability for a broad range of image contents since it yields sufficiently high correlation between metric and subjective scores (i.e. SRCC > 0.8) for all source images.

4) Complexity Analysis: Despite the fact DSCSI is not developed under a real-time operation constraint, its computational complexity is analyzed in terms of MATLAB execution time in order to demonstrate that it does not incur significant computational overhead to achieve high prediction accuracy. Experimental results are obtained on a 2.90-GHz Intel Core i7 CPU with 4 GB RAM running the Windows 7 operating system. Please note that the measurements are rough estimates as our implementation is not optimized. Table IX shows that the processing delay of the DSCSI lies between two state-of-the-art color metrics, i.e. FSIMc and CID.

Table X summarizes the percentage of time devoted to individual stages in DSCSI. We note that the main computation part (including estimation of local similarity measures and the combined quality score) occupies less than 30% of the time while the CSF filtering occupies more than 55%. The CSF filtering is a computationally intensive process as it involves decomposition of input color signal into opponent color channels, and application of independent spatial filters for individual color channels. We expect that the speed of our metric can be significantly improved by having a more efficient implementation of the CSF filtering using C/C++ compilable code. Such improvement will greatly improve the applicability of DSCSI in practical systems.

V. Conclusion

We introduce a full-reference color image quality metric, namely the DSCSI, capable of handling both chromatic and achromatic distortions originated from various imaging applications. The DSCSI extracts various kinds of local color descriptors from a perceptually uniform color space, which convey complementary information for quality perception and, together, improve the accuracy of the quality prediction model. Specifically, directional descriptors are exploited for hue data to take into account their periodicities. To achieve a good correlation between predicted scores and human judgements,

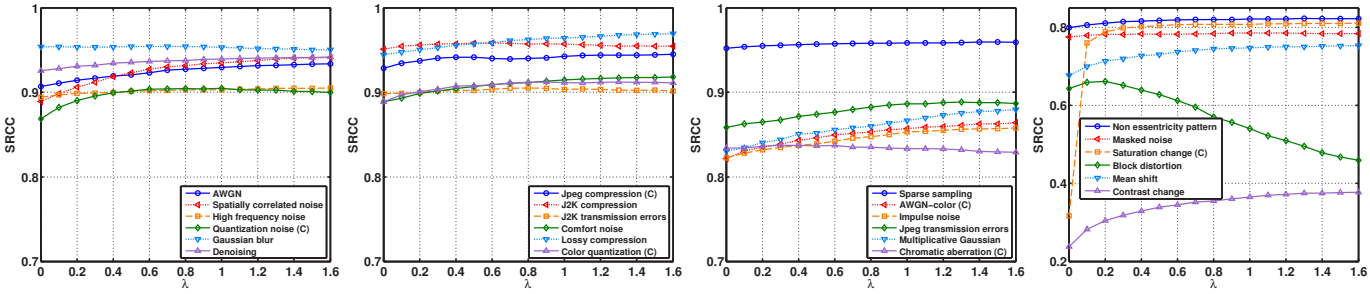


Fig. 9. Variation of SRCC as a function of λ on various distortion types in TID2013 DB. Chromatic distortion types are indicated with (C).

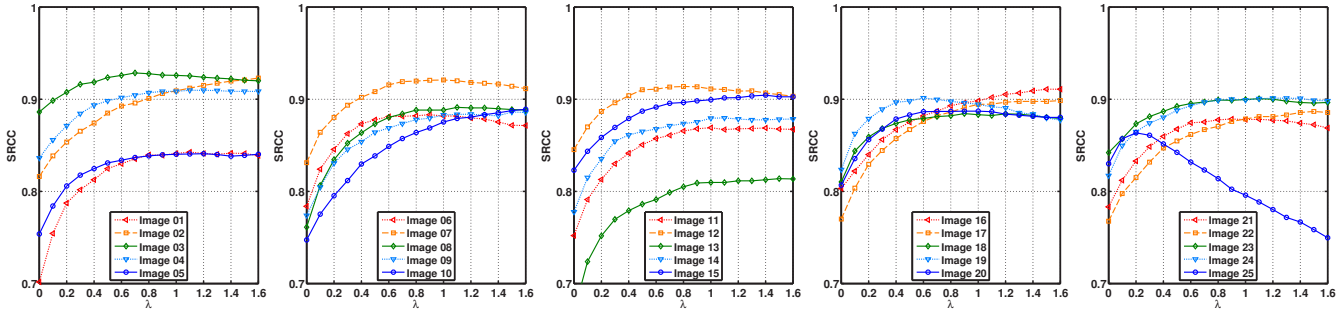


Fig. 10. Variation of SRCC as a function of λ for 25 individual source images in TID2013 DB. Indices for images are provided in Fig. 4.

TABLE IX
PROCESSING DELAY IN SECONDS PER IMAGE (512×384) FOR IMAGE QUALITY METRICS

Metric	SSIM	MSSSIM	VSSIM	UCIF	FSIMc	CID	CIEDE	DE2000	DSCSI
Processing Time (sec)	0.0361	0.1089	0.0600	0.0634	0.3109	0.5672	0.1868	0.2726	0.4659

TABLE X
PERCENTAGE OF TIME DEDICATED FOR INDIVIDUAL STAGES IN DSCSI

Processing Stage	**Percentage of Time**
1. RGB to S-CIELAB conversion	70.8
1-(a). CSF filtering	55.9
1-(b). filtered output to S-CIELAB	14.9
2. Derivation of local similarity measures	25.1
3. Derivation of final quality score	4.1

we also incorporate a weighting scheme that takes into account the significance of hue, as well as a spatial pooling based on HVS characteristics. Extensive validation performed on large-scale databases demonstrates that the DSCSI is competitive with state of the art metrics in terms of prediction accuracy and computational complexity. Experimental results indicate that the DSCSI is widely applicable for off-line prototyping and optimization of color image processing systems because of its high correlation with perception against both chromatic and achromatic distortions.

There are several issues which require further investigation to improve the effectiveness and applicability of the proposed metric. The prediction accuracy of the proposed scheme can be further enhanced by incorporating a more advanced pooling mechanism that closely simulates HVS mechanism, e.g. a color saliency model. As another direction of future research, the proposed general-purpose metric can be tailored to specific target applications, e.g. compression or preferred color reproduction systems, by adjusting free parameters based on domain-specific knowledge. Such application-specific metrics can be incorporated as objective functions for optimizing existing image processing solutions.

APPENDIX A
DIRECTIONAL STATISTICS FOR HUE DATA

Although we generally deal with data in the linear domain in most of the image analysis, there exist many scenarios that involve angular observations, e.g. hue angles in cylindrical color spaces. Conventional linear statistics are not suitable to handle them since they do not take into account their periodicities. Driven by this shortcoming, statistical tools to analyze directional data have been introduced over the last few decades [38], [39]. In this section, a number of basic quantities are presented which describe important features of the population from which hue data have been sampled [11].

We consider the case of N hue observations $\theta_1, \ldots, \theta_N$, where $\theta_i \in [0, 2\pi)$. A measure for angular data that is analogous to the linear sample mean is the *sample mean direction*. Initially, sample hue data are transformed to unit vectors in a two-dimensional plane, represented by sample points $[\cos\theta_i, \sin\theta_i]^T$ on a unit circle. Then we obtain the

resultant vector of N unit vectors from the origin by summing them component wise:

$$\mathbf{R} = \Big(\sum_{i=1}^{N} \cos\theta_i, \sum_{i=1}^{N} \sin\theta_i \Big) = \Big(C_N, S_N \Big) \tag{23}$$

The sample mean direction $\overline{\theta}$, average angle of the hue samples, can be computed from the direction of the resultant vector $\mathbf{R}$ as follows:

$$\overline{\theta} = \arctan(S_N / C_N) \tag{24}$$

The weighted sample mean direction can be computed simply by replacing (23) with

$$\mathbf{R} = \Big(\sum_{i=1}^{N} w_i \cos\theta_i, \sum_{i=1}^{N} w_i \sin\theta_i \Big) = \Big(C_N^w, S_N^w \Big) \tag{25}$$

and replacing C_N and S_N in (24) with C_N^w and S_N^w, where $\{w_i\}_{i=1}^N$ are the weighting factors such that $\sum_{i=1}^N w_i = 1$.

The *sample circular variance* $V \in [0,1]$ is a measure of dispersion of hue samples which is defined as:

$$V = 1 - \overline{R} \tag{26}$$

where $\overline{R}$ is the mean length of the resultant vector given by:

$$\overline{R} = \frac{\sqrt{C_N^2 + S_N^2}}{N} \tag{27}$$

Similar to the variance of linear data, it becomes close to zero if hue samples are closely clustered around the mean. However $V = 1$ does not imply a maximally dispersed distribution (i.e. uniform dispersion around the circle). It should be noted that V is a reliable indicator of dispersion only if hue samples exhibit a highly concentrated unimodal distribution.

REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual Visual Quality Metrics: A Survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297 – 312, 2011.

[2] D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," *ISRN Signal Processing*, vol. 2013, 2013, Art. ID 905685.

[3] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, June 2011.

[4] 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.

[5] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst. and Comput.*, vol. 2, 2003, pp. 9–13.

[6] A. Liu, W. Lin, and M. Narwaria, "Image Quality Assessment Based on Gradient Similarity," *IEEE Trans. Image. Process.*, vol. 21, no. 4, pp. 1500–1512, 2012.

[7] J. Zhu and N. Wang, "Image Quality Assessment by Visual Gradient Similarity," *IEEE Trans. Image. Process.*, vol. 21, no. 3, pp. 919–933, 2012.

[8] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient Magnitude Similarity Deviation : A Highly Efficient Perceptual Image Quality Index," *IEEE Trans. Image. Process.*, vol. 23, no. 2, pp. 684–695, 2014.

[9] 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, 2006.

[10] 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, pp. 1–21, Jan. 2010.

[11] A. Hanbury, "Circular Statistics Applied to Colour Images," in *8th Computer Vision Winter Workshop*, 2003.

[12] D. Lee and K. Plataniotis, "Towards a novel perceptual color difference metric using circular processing of hue components," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2014, pp. 166–170.

[13] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.*, vol. 26, no. 5, pp. 340–350, 2001.

[14] A. Toet and M. P. Lucassen, "A New Universal Colour Image Fidelity Metric," *Displays*, vol. 24, no. 4-5, pp. 197 – 207, 2003.

[15] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121 – 132, 2004.

[16] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM : A Feature Similarity Index for Image Quality Assessment," *IEEE Trans. Image. Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.

[17] Y. Shi, Y. Ding, R. Zhang, and J. Li, "Structure and hue similarity for color image quality assessment," in *Proc. Int. Conf. Electron. Comput. Technol.*, Feb 2009, pp. 329–333.

[18] H. Kikuchi, S. Kataoka, S. Muramatsu, and H. Huttunen, "Color-tone similarity of digital images," in *Proc. IEEE Int. Conf. Image Process.*, Sept 2013, pp. 393–397.

[19] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation," in *Proc. SPIE*, vol. 7527, 2010, pp. 75 271L–75 271L–9.

[20] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, "Image Difference Prediction: From Grayscale to Color," *IEEE Trans. Image. Process.*, vol. 22, no. 2, pp. 435–446, Feb. 2013.

[21] Z. Wang and A. C. Bovik, "Mean Squared Error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, 2009.

[22] S. Karunasekera and N. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image. Process.*, vol. 4, no. 6, pp. 713–724, 1995.

[23] I. Lissner and P. Urban, "Toward a Unified Color Space for Perception Based Image Processing," *IEEE Trans. Image. Process.*, vol. 21, no. 3, pp. 1153–1168, 2012.

[24] X. Zhang and B. A. Wandell, "A Spatial Extension of CIELAB for Digital Color-Image Reproduction," *J. Soc. Inf. Display*, vol. 5, no. 1, pp. 61–63, 1997.

[25] K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*. Springer-Verlag New York, Inc., 2000.

[26] F. Dugay, I. Farup, and J. Y. Hardeberg, "Perceptual evaluation of color gamut mapping algorithms," *Color Res. Appl.*, vol. 33, no. 6, pp. 470–476, 2008.

[27] F. Wichmann and N. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Perception & Psychophysics*, vol. 63, no. 8, pp. 1293–1313, 2001.

[28] E. Aptoula and S. Lefèvre, "On the morphological processing of hue," *Image Vis. Comput.*, vol. 27, no. 9, pp. 1394–1401, Aug. 2009.

[29] 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, March 2014.

[30] D. Rouse and S. Hemami, "Understanding and simplifying the structural similarity metric," in *Proc. IEEE Int. Conf. Image Process.*, Oct 2008, pp. 1188–1191.

[31] Z. Wang and Q. Li, "Information Content Weighting for Perceptual Image Quality Assessment," *IEEE Trans. Image. Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[32] A. B. Watson, Ed., *Digital Images and Human Vision*. MIT Press, 1993.

[33] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color Image Database TID2013 : Peculiarities and Preliminary Results," in *4th Europian Workshop on Visual Information Processing*, 2013.

[34] G. M. Johnson and M. D. Fairchild, "A Top Down Description of S-CIELAB and CIEDE2000," *Color Res. Appl.*, vol. 28, no. 6, pp. 425–435, 2003.

[35] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of Wavelet Quantization Noise," *IEEE Trans. Image. Process.*, vol. 6, no. 8, pp. 1164–1175, 1997.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "SSIM Index for Image Quality Assessment," Accessed: Aug 15, 2014. [Online]. Available: https://ece.uwaterloo.ca/~z70wang/research/ssim/

[37] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video," *IEEE Trans. Image. Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[38] K. Mardia, *Statistics of Directional Data*, ser. Probability and mathematical statistics. Academic Press, 1972.

[39] N. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.

Dohyoung Lee (S'14) received the B.Eng. degree in electrical engineering from the 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 (NSERC) Postgraduate Scholarship for 2013-2015.

Konstantinos N. Plataniotis (S'93-M'95-SM'03-F'12) is 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 are statistical signal processing, knowledge and digital media design, multimedia systems, biometrics, image and signal 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). Dr. Plataniotis 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 has received the IEEE Canada Engineering Educator Award for contributions to engineering education and inspirational guidance of graduate students.