

Objective Quality Assessment for Color-to-Gray Image Conversion

Kede Ma, *Student Member, IEEE*, Tiesong Zhao, *Member, IEEE*, Kai Zeng, and Zhou Wang, *Fellow, IEEE*

Abstract—Color-to-gray (C2G) image conversion is the process of transforming a color image into a grayscale one. Despite its wide usage in real-world applications, little work has been dedicated to compare the performance of C2G conversion algorithms. Subjective evaluation is reliable but is also inconvenient and time consuming. Here, we make one of the first attempts to develop an objective quality model that automatically predicts the perceived quality of C2G converted images. Inspired by the philosophy of the structural similarity index, we propose a C2G structural similarity (C2G-SSIM) index, which evaluates the luminance, contrast, and structure similarities between the reference color image and the C2G converted image. The three components are then combined depending on image type to yield an overall quality measure. Experimental results show that the proposed C2G-SSIM index has close agreement with subjective rankings and significantly outperforms existing objective quality metrics for C2G conversion. To explore the potentials of C2G-SSIM, we further demonstrate its use in two applications: 1) automatic parameter tuning for C2G conversion algorithms and 2) adaptive fusion of C2G converted images.

Index Terms—Image quality assessment, color-to-gray conversion, perceptual image processing, structural similarity.

I. INTRODUCTION

COLOR-TO-GRAY (C2G) image conversion [1], also referred to as decolorization, has been widely used in real-world applications including black-and-white printing of color images, aesthetic digital black-and-white photography, and preprocessing in image processing and machine vision systems. Since color is fundamentally a multi-dimensional phenomenon described by the perceptual attributes of luminance, chroma and hue [2], C2G image conversion, which pursues a 1D representation of the color image, inevitably causes information loss. The goal of C2G conversion is to preserve as much visually meaningful information about the reference color images as possible, while simultaneously produce perceptually natural and pleasing grayscale images.

In the literature, many C2G conversion algorithms have been proposed [1], [3]–[15]. With multiple C2G conversion

algorithms available, one would be interested in knowing which algorithm produces the best quality grayscale image. Subjective evaluation has been employed as the most straightforward image quality assessment (IQA) method. In [16], Čadík conducted two subjective experiments to evaluate the performance of C2G conversion algorithms, where a two-alternative forced choice approach was adopted to assess the accuracy and preference of pairs of images generated by 7 C2G algorithms from 24 reference color images. However, the subjective evaluation method is time consuming, expensive, and most importantly, cannot be incorporated into automated systems to monitor image quality and to optimize image processing algorithms. Therefore, objective quality assessment of C2G images is highly desirable. In 2008, Kuhn et al. developed a root weighted mean square to capture the preservation of color differences in the grayscale image [9]. This measure has not been tested on (or calibrated against) subjective data. A similar color contrast preserving ratio is proposed in [11] and later on evolves into the so called E-score by combining color content fidelity ratio [14]. Although E-score provides the most promising results so far, it cannot make adequate quality predictions of C2G images. Note that conventional full-reference approaches [17] such as mean squared error (MSE) [18] and structural similarity (SSIM) [19] are not applicable in this scenario, because the reference and distorted images do not have the same dimension. Applying reduced-reference and no-reference measures is also conceptually inappropriate because the source image is fully available that contains even more information than the test image [20].

To address this problem, we make one of the first attempts to develop an objective IQA model that evaluates the quality of a C2G image using its corresponding color image as reference. Our work is primarily inspired by the philosophy of SSIM which assumes that human visual perception is highly adapted for extracting structural information from its viewing field [19]. As in SSIM, our model, named C2G-SSIM, consists of three components that measure luminance, contrast and structure similarities, respectively. The three components are then integrated into an overall quality measure based on the type of content (photographic or synthetic) in the image. Validations on the subjective database [16] show good correlations between the subjective rankings and predictions of C2G-SSIM, and superiority over existing objective models for C2G images. Furthermore, we use two examples—automatic parameter tuning of C2G conversion algorithms and adaptive fusion of C2G images—to demonstrate the potential applications of C2G-SSIM.

Manuscript received November 1, 2014; revised March 29, 2015 and June 12, 2015; accepted July 7, 2015. Date of publication July 22, 2015; date of current version September 18, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stefan Winkler.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: k29ma@uwaterloo.ca; ztiesong@uwaterloo.ca; kzeng@uwaterloo.ca; zhouwang@ieec.org).

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.2015.2460015

II. RELATED WORK

A. Existing Color-to-Gray Algorithms

Most existing C2G conversion algorithms seek to preserve color distinctions of the input color image in the corresponding grayscale image with some additional constraints, such as global consistency and grayscale preservation. As one of the first attempts, Bala and Eschbach introduced high frequency chrominance information into the luminance channel so as to preserve distinctions between adjacent colors [1]. This algorithm tends to produce artificial edges in the C2G image. Rasche et al. incorporated contrast preservation and luminance consistency into a linear programming problem, where the difference between two gray values is proportional to that between the corresponding color values [3]. Gooth et al. transformed the C2G problem into a quadratic optimization one by quantifying the preservation of color differences between two distinct points in the grayscale image [4]. By using predominant component analysis, Grundland and Dodgson computed prevailing chromatic contrasts along the predominant chromatic axis and used them to compensate the luminance channel [5]. A coloroid system-based [21] C2G conversion algorithm is proposed in [6], where color and luminance contrasts form a gradient field and enhancement are achieved by reducing the inconsistency of the field. Smith et al. developed a two-step approach that first globally assigns gray values incorporating the Helmholtz-Kohlrausch color appearance effect [22] and then locally enhances the grayscale values to reproduce the original contrast [8]. The algorithm performs the best based on Čadík's subjective experiment [16]. A mass-spring system is introduced in [9] to perform C2G conversion. Kim's method adopts a nonlinear global mapping to perform robust decolorization [7]. Song et al. incorporated spatial consistency, structure information and color channel perception priority into a probabilistic graphic model and optimized the model as an integral minimization problem [10]. Lu's method attempts to maximally preserve the original color contrast by minimizing a bimodal Gaussian function [11], [14]. However, contrast preservation or enhancement does not necessarily lead to perceptual quality improvement, but may produce some unnatural images due to luminance inconsistency [11]. Song et al. [12] and Zhou et al. [13] independently revisited a simple C2G conversion model that linearly combines RGB channels. The weights are determined based on predefined contrast preservation and saliency preservation measures. More recently, Eynard et al. assumed that if a color transformed image preserves the structural information of the original image, the respective Laplacians are jointly diagonalizable or equivalently commutative. Using Laplacians commutativity as the criterion, they minimized it with respect to the parameters of a color transformation to achieve optimal structure preservation [15].

B. The SSIM Index

Suppose \mathbf{x}' and \mathbf{y}' are local image patches taken from the same location of two images being compared, the local SSIM index computes three components: the luminance

similarity $l(\mathbf{x}', \mathbf{y}')$, contrast similarity $c(\mathbf{x}', \mathbf{y}')$ and structure similarity $s(\mathbf{x}', \mathbf{y}')$

$$l(\mathbf{x}', \mathbf{y}') = \frac{2\mu_{x'}\mu_{y'} + C_1}{\mu_{x'}^2 + \mu_{y'}^2 + C_1}, \quad (1)$$

$$c(\mathbf{x}', \mathbf{y}') = \frac{2\sigma_{x'}\sigma_{y'} + C_2}{\sigma_{x'}^2 + \sigma_{y'}^2 + C_2}, \quad (2)$$

$$s(\mathbf{x}', \mathbf{y}') = \frac{\sigma_{x'y'} + C_3}{2\sigma_{x'}\sigma_{y'} + C_3}, \quad (3)$$

where μ , σ and $\sigma_{x'y'}$ denote the mean, standard deviation (std) and covariance of the image patches, respectively [19]. C_1 , C_2 and C_3 are small positive constants to avoid instability, when the denominators are close to 0. Finally, the three measures are combined to yield the SSIM index

$$\text{SSIM}(\mathbf{x}', \mathbf{y}') = l(\mathbf{x}', \mathbf{y}')^\alpha \cdot c(\mathbf{x}', \mathbf{y}')^\beta \cdot s(\mathbf{x}', \mathbf{y}')^\gamma, \quad (4)$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components, respectively. By setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, the simplified SSIM index that is widely used in practice is given by

$$\text{SSIM}(\mathbf{x}', \mathbf{y}') = \frac{(2\mu_{x'}\mu_{y'} + C_1)(2\sigma_{x'y'} + C_2)}{(\mu_{x'}^2 + \mu_{y'}^2 + C_1)(\sigma_{x'}^2 + \sigma_{y'}^2 + C_2)}. \quad (5)$$

It is widely recognized that SSIM is better correlated with the human visual system (HVS) than MSE [17], [18], [23] and has a number of desirable mathematical properties [24] for optimization purposes [23], [25].

C. Other Atypical IQA Problems

Typical full-reference IQA problems make the following assumptions on the reference and test images: the reference image should be pristine or distortion-free (1); both images should have the same spatial resolution (2), the same dynamic range (3) and the same number of color channels (4); and there is one reference image (5). An IQA problem that does not satisfy at least one of the above assumptions is atypical. For example, quality assessment of contrast-enhanced and dehazed images allows the test image to have better perceived quality than that of the reference image [26]–[28]; quality assessment for image interpolation and super-resolution make use of reference images of different spatial resolutions from test images [29], [30]; quality assessment of high dynamic range image tone mapping algorithms deals with images of different dynamic ranges [31]–[33]; quality assessment of image fusion algorithms uses a sequence of images as reference [34]–[37]. The current paper aims to solve a different atypical IQA problem, where the reference and the test images have different numbers of color channels. Each atypical IQA problem casts certain new challenges. While similar design principles may be used in all of them, a general solution is not possible, and more focus is necessary to put on the specific domain problems and novel solutions to tackle them.

III. THE C2G-SSIM INDEX

The diagram of the proposed C2G-SSIM index is shown in Fig. 1. First, we transform both the reference color image

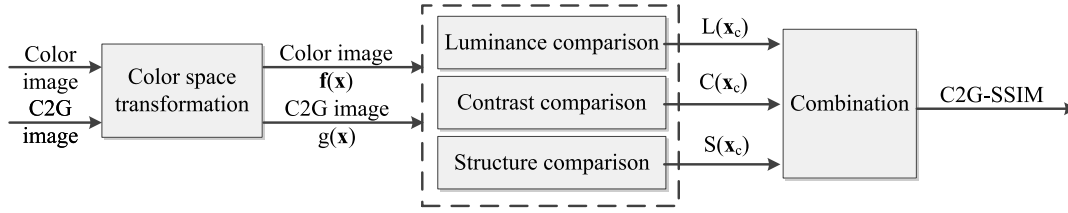


Fig. 1. Three-stage structure of C2G-SSIM.

and the test C2G image into a color space, where the color representation is better matched to the HVS. Next, we measure luminance, contrast and structure distortions to capture perceived quality changes introduced by C2G conversion. Finally, we combine the above three measurements into an overall quality measure based on the type of image content.

A. Color Space Transformation

To capture the perceived quality loss during C2G conversion, we desire to work in a color space of perceptual uniformity, where the Euclidean distance between two color points is proportional to the perceived color difference, denoted by ΔE . In the commonly used RGB color space, the color components are highly correlated, and the structural information may be over/underestimated [2]. Unfortunately, no perfectly uniform color space has been discovered yet [38]. Some approximations have been proposed, including CIELAB [39], CIECAM02 [40] and LAB2000HL [41], where the perceptual uniformity generally holds for small color differences (SCDs). Considering both the computational complexity and the effectiveness, we choose CIELAB as our working color space. For a C2G image, its luminance value can also be transformed into the achromatic axis in CIELAB space. Therefore, we use the absolute luminance difference to represent the color difference in a C2G image.

It needs to be aware that it is meaningful to predict perceptual color differences in CIELAB space only for certain range of SCDs. At the high end ($\Delta E > 15$), it makes little sense to rely on CIELAB distances to differentiate large color differences (LCDs) [42]. For instance, the HVS is typically certain about the difference between $\Delta E = 3$ and $\Delta E = 4$, whereas has major difficulties in differentiating $\Delta E = 33$ and $\Delta E = 34$. On the other hand, at the low extreme ($\Delta E < 2.3$), the HVS cannot perceive the color differences [43]. As a result, when ΔE is lower than a just-noticeable difference (JND) level, the differences in ΔE value do not have any perceptual meaning.

B. Similarity Measurement

Let \mathbf{x} represent the spatial image coordinate, and $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$ denote the color and C2G images, respectively. At any particular spatial location \mathbf{x} , $\mathbf{f}(\mathbf{x})$ is a 3-vector and $g(\mathbf{x})$ is a scalar. As in the SSIM approach, we start with image similarity assessment at each spatial location. A useful approach to accomplish this is to define a geometric proximity function centered at any given spatial location \mathbf{x}_c . The proximity function is denoted by $p(\mathbf{x}, \mathbf{x}_c)$. A special case

is patch-based method, which corresponds to the case that $p(\mathbf{x}, \mathbf{x}_c)$ is a 2D box function centered at \mathbf{x}_c . But in general, $p(\mathbf{x}, \mathbf{x}_c)$ may take many other forms that provide smoother transitions at block boundaries [44]–[47]. Here we adopt a radially symmetric Gaussian function centered at \mathbf{x}_c

$$p(\mathbf{x}, \mathbf{x}_c) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_c\|^2}{2\sigma_p^2}\right), \quad (6)$$

where σ_p is the geometric spread determining the size of the Gaussian function.

To compare $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$ at \mathbf{x}_c , we follow the idea of SSIM by combining three distinct similarity measures of luminance, contrast and structure. Specifically, the luminance measure $L(\mathbf{x}_c)$ assesses the local luminance consistency between $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$; the contrast measure $C(\mathbf{x}_c)$ indicates the local contrast similarity between $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$; and the structure measure $S(\mathbf{x}_c)$ evaluates the local structure similarity between $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$. By combining the three relatively independent components, we define the overall quality measure at \mathbf{x}_c as

$$q(\mathbf{x}_c) = \mathcal{F}(L(\mathbf{x}_c), C(\mathbf{x}_c), S(\mathbf{x}_c)), \quad (7)$$

where $\mathcal{F}(\cdot)$ is a combination function that monotonically increases with the three components such that any loss in luminance, contrast or structure results in degradation of the overall quality. The three similarity components are described as follows.

1) *Luminance Similarity*: We first extract the luminance components $l_f(\mathbf{x})$ and $l_g(\mathbf{x})$ of $\mathbf{f}(\mathbf{x})$ and $g(\mathbf{x})$, respectively. Assuming continuous signals, the weighted mean luminance of the color image is defined as:

$$u_f(\mathbf{x}_c) = k_p^{-1}(\mathbf{x}_c) \int l_f(\mathbf{x}) p(\mathbf{x}, \mathbf{x}_c) d\mathbf{x}, \quad (8)$$

where $k_p(\mathbf{x}_c)$ is a normalizing term

$$k_p(\mathbf{x}_c) = \int p(\mathbf{x}, \mathbf{x}_c) d\mathbf{x}. \quad (9)$$

Computationally, when the same proximity function is applied to all spatial locations, k_p is a constant and Eq. (8) can be implemented by a low-pass filter. Furthermore, if the filter is radially symmetric, $p(\mathbf{x}, \mathbf{x}_c)$ is only a function of the vector difference $\mathbf{x} - \mathbf{x}_c$.

The mean luminance $u_g(\mathbf{x}_c)$ of the C2G image is defined similarly. Based on the comparison of $u_f(\mathbf{x}_c)$ and $u_g(\mathbf{x}_c)$ and taking the form of SSIM [19], we define the luminance measure as

$$L(\mathbf{x}_c) = \frac{2u_f(\mathbf{x}_c)u_g(\mathbf{x}_c) + C_1}{u_f(\mathbf{x}_c)^2 + u_g(\mathbf{x}_c)^2 + C_1}, \quad (10)$$

where C_1 is a small positive stabilizing constant.

2) *Contrast Similarity*: In order to evaluate the local color contrast at spatial location \mathbf{x}_c , we compute its weighted mean color difference from its surroundings, which takes continuous form

$$d_f(\mathbf{x}_c) = k_p^{-1}(\mathbf{x}_c) \int \phi(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|) p(\mathbf{x}, \mathbf{x}_c) d\mathbf{x}, \quad (11)$$

where $\phi(\cdot)$ is a nonlinear mapping (detailed discussion will be given later) applied to the Euclidean distance $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|$ (the perceived color difference ΔE). By replacing $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|$ with $|g(\mathbf{x}) - g(\mathbf{x}_c)|$ in Eq. (11), we can also compute the mean gray tone difference $d_g(\mathbf{x}_c)$ of the C2G image. In the following text, we use ΔE to replace $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|$ and $|g(\mathbf{x}) - g(\mathbf{x}_c)|$ for simplicity.

The reasons to apply $\phi(\cdot)$ on top of ΔE are threefold. First, the HVS can hardly perceive ΔE less than 2.3 in CIELAB space, which corresponds to a JND [39]. Second, the perceived LCD is poorly approximated by ΔE in CIELAB or any other color spaces [40], [41], which achieve reasonable perceptual uniformity at SCDs only. Third, the HVS has difficulty in comparing two pairs of LCDs, especially when their magnitudes approximate each other [42]. A useful design of $\phi(\cdot)$ is to let it saturate at both low and high ends, and normalize the middle range to be between 0 and 1, such that SCDs below JND are considered insignificant and mapped to 0, and all significant LCDs are mapped to 1. Ideally, $\phi(\cdot)$ should be a monotonically increasing function with both the low end and the high end asymptotically approaching 0 and 1, respectively. Following typical psychometric functions used to describe visual sensitivity of contrast [48], we choose a cumulative normal distribution function to define this nonlinear mapping

$$\phi(\Delta E) = \frac{1}{\sqrt{2\pi}\sigma_\phi} \int_{-\infty}^{\Delta E} \exp\left[-\frac{(\omega - \mu_\phi)^2}{2\sigma_\phi^2}\right] d\omega, \quad (12)$$

where μ_ϕ and σ_ϕ are the mean and std of the normal distribution. Practically, we use two points $\phi(2.3) = 0.05$ and $\phi(15) = 0.95$ to determine the curve, where $\Delta E = 2.3$ is the JND in CIELAB space [39] and $\Delta E = 15$ represents the soft threshold of LCD to approximate the minimum sample pair of LCD in [42]. With these two control points, it is easy to find $\mu_\phi = 11.15$ and $\sigma_\phi = 5.38$. A visual demonstration of $\phi(\cdot)$ is shown in Fig. 2.

The contrast measure $C(\mathbf{x}_c)$ is defined as a function of $d_f(\mathbf{x}_c)$ and $d_g(\mathbf{x}_c)$. Following the form used in SSIM [19], we define the contrast measure as

$$C(\mathbf{x}_c) = \frac{2d_f(\mathbf{x}_c)d_g(\mathbf{x}_c) + C_2}{d_f(\mathbf{x}_c)^2 + d_g(\mathbf{x}_c)^2 + C_2}, \quad (13)$$

where C_2 is a small positive constant to avoid instability when the denominator is close to zero.

3) *Structure Similarity*: The structure measure takes a similar form as in SSIM [19]

$$S(\mathbf{x}_c) = \frac{\sigma_{fg}(\mathbf{x}_c) + C_3}{\sigma_f(\mathbf{x}_c)\sigma_g(\mathbf{x}_c) + C_3}, \quad (14)$$

where C_3 is also a small positive constant introduced in both the denominator and numerator. $\sigma_f(\mathbf{x}_c)$, $\sigma_g(\mathbf{x}_c)$ and $\sigma_{fg}(\mathbf{x}_c)$ are stds of $\phi(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|)$, $\phi(|g(\mathbf{x}) - g(\mathbf{x}_c)|)$ and cross

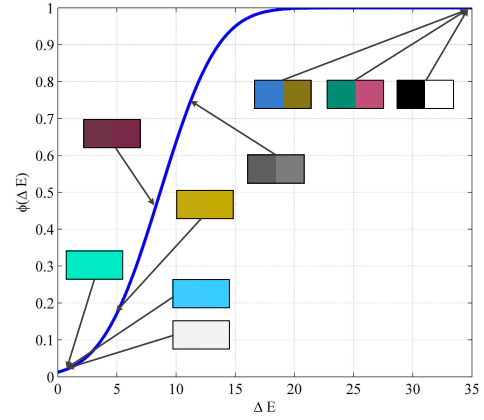


Fig. 2. Illustration of the nonlinear mapping $\phi(\cdot)$ with different color differences.

correlation between $\phi(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|)$ and $\phi(|g(\mathbf{x}) - g(\mathbf{x}_c)|)$, respectively. In continuous form, $\sigma_f^2(\mathbf{x}_c)$ and $\sigma_{fg}(\mathbf{x}_c)$ are evaluated by

$$\sigma_f^2(\mathbf{x}_c) = k_p^{-1}(\mathbf{x}_c) \int \left[\phi(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|) - d_f(\mathbf{x}_c) \right]^2 p(\mathbf{x}, \mathbf{x}_c) d\mathbf{x} \quad (15)$$

and

$$\sigma_{fg}(\mathbf{x}_c) = k_p^{-1}(\mathbf{x}_c) \int \left\{ \left[\phi(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_c)\|) - d_f(\mathbf{x}_c) \right] \cdot \left[\phi(|g(\mathbf{x}) - g(\mathbf{x}_c)|) - d_g(\mathbf{x}_c) \right] \right\} p(\mathbf{x}, \mathbf{x}_c) d\mathbf{x}, \quad (16)$$

respectively. $\sigma_g^2(\mathbf{x}_c)$ can be obtained in a similar way using Eq. (15).

C. Overall Quality Measure

The luminance measure $L(\mathbf{x}_c)$, contrast measure $C(\mathbf{x}_c)$ and structure measure $S(\mathbf{x}_c)$ describe three different aspects of the perceptual quality of the C2G image. $L(\mathbf{x}_c)$ quantifies the luminance consistency, whose importance in assessing the quality of C2G images varies according to the nature of image source, while $C(\mathbf{x}_c)$ and $S(\mathbf{x}_c)$ are more related to structural detail preservation of the C2G conversion. Specifically, for photographic images (PI) of natural scenes, human observers have strong prior knowledge about the luminance information. Whether such information is maintained in C2G images is well reflected by the luminance measure $L(\mathbf{x}_c)$. On the other hand, for synthetic images (SI) generated via computer graphics, human observers have little prior knowledge about the luminance of the synthetic objects, and thus $L(\mathbf{x}_c)$ is less relevant. To justify the above intuition, we carried out an informal test. Specifically, we asked human subjects to score C2G images and then asked them how they had evaluated the quality degradation of the image being tested. We found that for PI, one of the most common answers was that the luminance of certain parts of the image does not look right, but this was almost never the case for SI. This suggests that the subjects were using distinct strategies to make the judgement

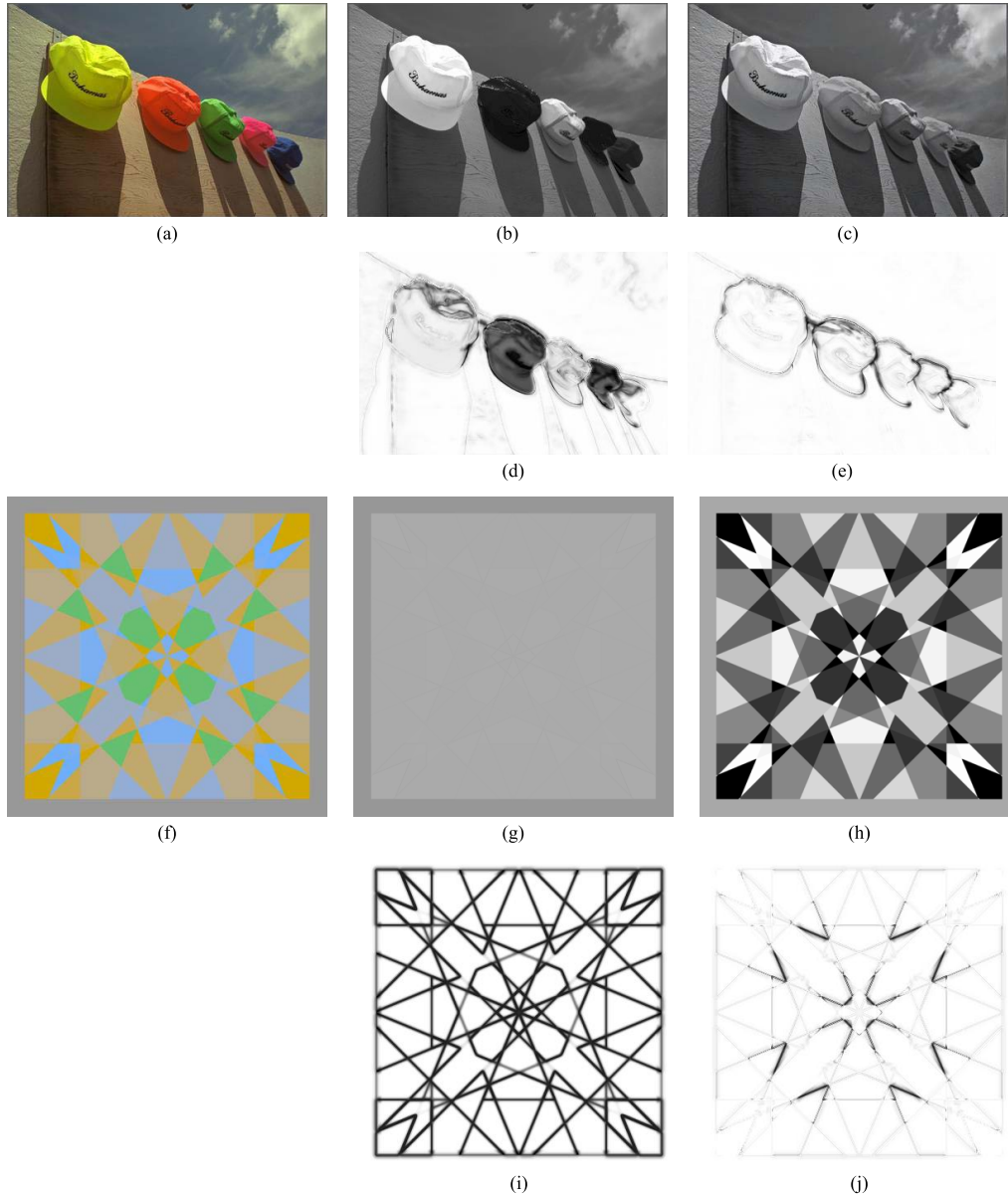


Fig. 3. C2G images and their C2G-SSIM index maps. (a) and (f) are reference color images. (b), (c), (g) and (h) are C2G images created by the methods in [3] and [8], CIEY and [11], respectively. (d), (e), (i) and (j) are the corresponding C2G-SSIM maps of (b), (c), (g) and (h), respectively. In all C2G-SSIM index maps, brighter indicates better quality.

about the two types of images. This can also be observed in the subjective data of [16] in Table VI, where the correlations between luminance similarity and perceptual quality are sharply different for PI and SI: one is highly relevant and the other has very low correlation. The above observations motivate us to construct an overall C2G-SSIM index that allows for flexible combinations of the three components:

$$q(\mathbf{x}_c) = L(\mathbf{x}_c)^\alpha \cdot C(\mathbf{x}_c)^\beta \cdot S(\mathbf{x}_c)^\gamma, \quad (17)$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are user-defined control parameters to adjust the relative importance of the three components similar to SSIM [19]. Specifically, in order to simplify the expression, we set $\beta = \gamma = 1$ and leave only one free parameter α , which typically takes a value on the interval of [0, 1]. In our current implementation, $\alpha = 1$ is used

when the input color image is PI so that all three components are equally weighted, and $\alpha = 0$ is applied when the input image is SI so that only contrast and structure measures are under consideration. Furthermore, users may adjust α between 0 and 1 to account for in-between cases, for example, cartoon pictures of natural scenes.

The local comparison is applied using a sliding window across the entire image, resulting in a quality map indicating how the luminance consistency and structural detail are preserved at each spatial location. This local computation is meaningful with regard to visual perception of image quality for two reasons. First, image contrast and structure are spatially nonstationary. Second, at one time instance the HVS can only perceive a local area in the image with high resolution [49]. A visual demonstration is shown in Fig. 3,

where the brightness indicates the magnitude of local C2G-SSIM value. As can be seen, the quality maps well reflect spatial variations of the perceived image quality of different C2G images. Specifically, the C2G image in Fig. 3(c) shows better luminance consistency than the C2G image in Fig. 3(b), where the luminance of the hats are severely altered, as clearly indicated by the C2G-SSIM maps. Moreover, even stronger penalty (marked as black pixels in the C2G-SSIM map) is given to the letter regions in the front parts of the second and the fourth hats, where the structural details are gone. For synthetic image, the C2G image in Fig 3(g) created by simply extracting the luminance channel of CIEXYZ color space fails to perverse the structural color pattern in Fig 3(f). This structural distortion is well reflected in its C2G-SSIM map in Fig 3(i). By contrast, the C2G image created by Lu's method [11] is much better in preserving the contrast and structure in the color image. Therefore, the C2G-SSIM map in Fig 3(j) is generally bright, indicating problems at only a few color edges.

In practice, one usually desires a single score for the overall quality of the entire image. A single C2G-SSIM score can be obtained by taking the average of the C2G-SSIM map:

$$Q(\mathbf{f}, g) = \frac{\int q(\mathbf{x}_c) d\mathbf{x}_c}{\int d\mathbf{x}_c}. \quad (18)$$

Since the maximum of $q(\mathbf{x}_c)$ is 1, Q is also upper bounded by 1. Throughout this paper, we use a circularly symmetric Gaussian sliding window with size $W = 15$ and std $\sigma_p = \lfloor \frac{W}{6} \rfloor = 2$, which allows the window to cover approximately 3 stds of the Gaussian profile, as suggested in [50]. We set $C_1 = 10$, $C_2 = 0.1$ and $C_3 = 0.01$, respectively. Empirically, we find that the overall performance of C2G-SSIM is robust to variations of these parameters.

IV. EXPERIMENTAL RESULTS

A. Subjective C2G IQA Database [16]

For the purpose of perceptual evaluation of C2G images, Čadík [16] created a subjective C2G IQA database that includes 24 images generated by state-of-the-art C2G conversion algorithms. A two-alternatives forced choice methodology was adopted in the experiment, where the subjects were asked to select the more favorable C2G image from a pair of images. Two subjective experiments were conducted, including (1) *accuracy*, in which the two C2G images were shown at the left and right sides with the corresponding reference color image in the middle, and (2) *preference*, in which the subjects gave opinions without any reference. To the best of our knowledge, so far this is the only publicly available subjective database dedicated to C2G conversions.

A total of 24 high-quality color images were decolorized by 7 C2G conversions, which led to $24 \times 7 = 168$ C2G images. Default parameter settings were adopted for all C2G algorithms without any tuning or adaption for better quality. A total of 119 subjects (59 for *accuracy* and 60 for *preference*) were recruited in the experiment which ended up with 20328 observations of pair-wise comparisons. For each observation, the selected C2G image by a subject was given a score

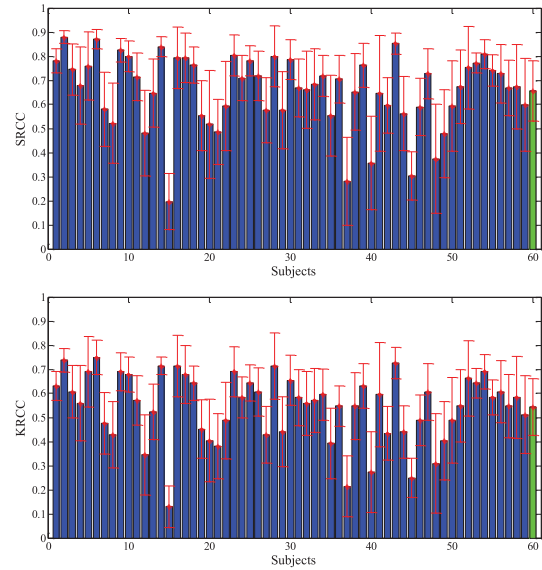


Fig. 4. Mean and std of SRCC and KRCC values between individual subject and average subject rankings in the *accuracy* test. The rightmost column gives the average performance of all subjects.

of 1, and the other a score of 0. As a result, a 7×7 frequency matrix was created for each subject, along with all color and C2G images in the database. In addition, the database included the standard score (z-score) which were converted from the frequency matrix using Thurstone's Law of Comparative Judgments, Case V [51].

Based on the subjective scores, it is useful to analyze and observe the behavior of all subjects for each image set, which consists of a color image and its corresponding C2G images. The comparison is based on Spearman's rank-order correlation coefficient (SRCC) and Kendall's rank-order correlation coefficient (KRCC) [52]. The better C2G IQA model would need to have larger SRCC and KRCC values with respect to subjective test results.

For each image set, the rankings given to each image are averaged over all subjects. Considering these average ranking scores as the "ground truth", the performance of each individual subject can be observed by computing SRCC and KRCC between their ranking scores with the "ground truth" for each image set. Furthermore, the overall performance of the subject can be evaluated by the average SRCC and KRCC values over all 24 image sets.

The mean and std of SRCC and KRCC values for each individual subject in the *accuracy* test are shown in Fig. 4, while those in the *preference* test are given in Fig. 5. It can be observed that there is a considerable agreement between different subjects on ranking the quality of C2G images in both tests. The degree of agreement is significantly higher for *preference* than *accuracy*, because the stds of SRCC and KRCC for *preference* are much lower. In both Fig. 4 and Fig. 5, the average performance across all individual subjects is also given in the rightmost column, which provides a general idea about the behaviors of an "average subject" and also supplies a very useful baseline for the evaluation of objective quality assessment models. Table I summarizes the results.

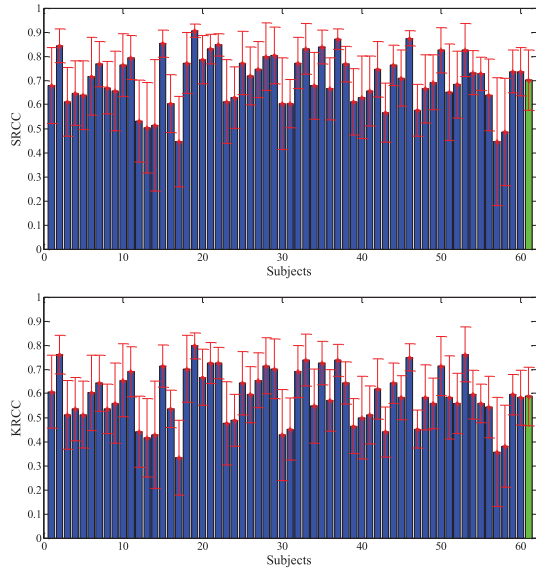


Fig. 5. Mean and std of SRCC and KRCC values between individual subject and average subject rankings in the *preference* test. The rightmost column gives the average performance of all subjects.

TABLE I
PERFORMANCE OF AN AVERAGE SUBJECT IN *ACCURACY*
AND *PREFERENCE* TESTS

	SRCC		KRCC	
	Mean	Std	Mean	Std
<i>accuracy</i>	0.6577	0.2492	0.5449	0.2367
<i>preference</i>	0.7033	0.1241	0.5898	0.1218

TABLE II
CATEGORIZATION OF REFERENCE COLOR IMAGES IN [16]

Category	Image No.							α
PI	1	3	4	9	10	11	13	1
	14	15	16	19	22	23	24	
SI	2	5	6	7	8	12	17	0
	18	20	21					

B. Validation of C2G-SSIM

Based on the above database, the performance of the C2G-SSIM index can be evaluated through the comparison between objective quality values and subjective ranking scores in terms of SRCC and KRCC. The 24 reference color images in the database can be divided into PI and SI categories, where Table II lists the members of each category. The image indices are in the same order as in [16]. Generally, an input color image can be easily classified into one of the two categories. In our implementation, α in Eq. (17) is set to be 1 and 0 for PI and SI, respectively.

We compare C2G-SSIM with existing metrics RWMS [9], CCPR, CCFR and E-score [14]. We implemented the exact version of RWMS without k-means algorithm for color quantization by ourselves and obtained the codes of CCPR, CCFR and E-score from the authors' website [53]. All three measures in [14] include one common parameter τ , which ranges between 1 and 40, and needs to be hand-picked by users. Here, we report the mean E-score values averaging over τ from 1 to 40. Moreover, the behavior of an average subject

as discussed in Section IV-A for each color image provides a useful benchmark to evaluate the relative performance of the objective quality metrics. The comparison results for the *accuracy* and *preference* tests are listed in Table III and Table IV, respectively. The average SRCC and KRCC values of an average subject, and all objective metrics for each image category and the overall database are marked with bold. It can be seen that on PI subset, the proposed C2G-SSIM index achieves better performance than an average subject for *accuracy* test and also significantly outperforms existing objective metrics [9], [14]. For *preference* test, C2G-SSIM is comparable to an average subject. A visual example is shown in Fig. 6, where images are displayed in ascending order from left to right in terms of C2G-SSIM.

To ascertain that the improvement of the proposed model is statistically significant, we carried out a statistical significance analysis by following the approach introduced in [54]. First, a nonlinear regression function is applied to map the objective quality scores to predict the subjective Z-scores. We observe that the prediction residuals all have zero-mean, and thus the model with lower variance is generally considered better than the one with higher variance. We conduct a hypothesis testing using F-statistics. The test statistic is the ratio of variances. The null hypothesis is that the prediction residuals from one quality model come from the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another model. After comparing every possible pairs of objective models, the results are summarized in Table V, where a symbol "1" means the row model performs significantly better than the column model, a symbol "0" means the opposite, and a symbol "-" indicates that the row and column models are statistically indistinguishable. Each entry in the table includes two characters, which correspond to the *accuracy* and *preference* tests [16], respectively. We have also added a random guess procedure as the benchmark, whose scores are randomly sampled from an independent Gaussian process with zero mean and unit std. It can be observed that the proposed model is statistically better than random guess, RWMS and E-score.

C. Discussion

1) *Automatic Selection of α* : To fully automate C2G-SSIM, we suggest a simple yet efficient feature to classify images into photographic and synthetic groups. Our motivation is that synthetic images often contain a small number of dominant colors, some of which are isoluminant. As a result, the histogram of synthetic images in their luminance channels tend to be highly compact, resulting in a small entropy. By contrast, the entropy of a photographic image is typically larger due to a more spread histogram. Based on the fact that the maximum possible entropy of an 8-bit image is 8, we set the classification threshold to be 4, and determine α by

$$\alpha = \begin{cases} 1 & \text{if } T \geq 4 \\ 0 & \text{if } T < 4, \end{cases} \quad (19)$$

where T stands for the entropy of the luminance channel of a test image. We test this simple method on

TABLE III
PERFORMANCE COMPARISON OF C2G-SSIM WITH EXISTING METRICS ON [16] FOR ACCURACY TEST

Image set		SRCC				KRCC			
		RWMS	E-score	Subject	C2G-SSIM	RWMS	E-score	Subject	C2G-SSIM
PI	Image1	-0.1071	-0.5357	0.5697	0.7143	0.0476	-0.2381	0.4512	0.6190
	Image3	0.4286	0.3214	0.5204	0.6071	0.3333	0.1429	0.4150	0.4286
	Image4	-0.2500	0.0000	0.6667	0.8214	-0.1429	-0.0476	0.5601	0.7143
	Image9	0.3929	0.3929	0.4323	0.7500	0.2381	0.3333	0.3484	0.6190
	Image10	0.5000	0.5357	0.4906	0.6429	0.3333	0.4286	0.3784	0.4286
	Image11	0.5000	0.5714	0.4787	0.9643	0.3333	0.4286	0.3868	0.9048
	Image13	-0.5357	-0.2857	0.6015	0.4286	-0.4286	-0.2381	0.4987	0.4286
	Image14	-0.1071	-0.3571	0.6147	0.9643	-0.1429	-0.3333	0.5288	0.9048
	Image15	0.6071	0.6429	0.5376	0.9643	0.5238	0.4286	0.4286	0.9048
	Image16	0.4286	0.7500	0.5969	0.8571	0.3333	0.6190	0.4739	0.7143
	Image19	0.1071	0.5357	0.6429	0.9286	0.1429	0.2381	0.5038	0.8095
	Image22	0.6071	0.3929	0.7538	0.5714	0.5238	0.2381	0.6642	0.3333
	Image23	0.2500	0.0714	0.7194	0.9286	0.1429	0.0476	0.6100	0.8095
	Image24	0.6071	0.5357	0.6523	0.8214	0.5238	0.4286	0.5188	0.6190
PI Average		0.2449	0.2551	0.5912	0.7832	0.1973	0.1769	0.4833	0.6599
SI	Image2	0.0357	0.8571	0.8853	0.5714	0.0476	0.7143	0.8045	0.4286
	Image5	0.2143	0.8214	0.8010	0.8929	0.1429	0.6190	0.6689	0.8095
	Image6	0.5714	0.9643	0.7801	0.9286	0.4286	0.9048	0.6541	0.8095
	Image7	0.6429	0.0714	0.5752	0.7500	0.4286	0.0476	0.4586	0.5238
	Image8	0.3571	0.8214	0.8402	0.8571	0.3333	0.6190	0.7043	0.7143
	Image12	0.2143	0.7143	0.8327	0.8571	0.1429	0.6190	0.7193	0.7143
	Image17	0.2857	0.2143	0.6616	0.3929	0.1429	0.0476	0.5465	0.3333
	Image18	0.1071	0.1786	0.5697	0.9286	0.1429	0.0476	0.4286	0.8095
	Image20	0.5357	0.6071	0.8233	0.7500	0.4286	0.5238	0.7043	0.6190
	Image21	-0.2143	0.6786	0.7379	0.8214	-0.1429	0.5238	0.6217	0.7143
	SI Average		0.2750	0.5929	0.7507	0.7750	0.2095	0.4667	0.6311
Overall		0.2574	0.3958	0.6577	0.7798	0.2024	0.2976	0.5449	0.6548

TABLE IV
PERFORMANCE COMPARISON OF C2G-SSIM WITH EXISTING METRICS ON [16] FOR PREFERENCE TEST

Image set		SRCC				KRCC			
		RWMS	E-score	Subject	C2G-SSIM	RWMS	E-score	Subject	C2G-SSIM
PI	Image1	0.4286	-0.1071	0.6143	0.6786	0.3333	0.0476	0.5143	0.5238
	Image3	0.0714	0.1429	0.4982	0.7143	0.0476	0.0476	0.3810	0.5238
	Image4	-0.2500	0.0000	0.8750	0.8214	-0.1429	-0.0476	0.7714	0.7143
	Image9	0.4643	0.5000	0.5771	0.7500	0.4286	0.3333	0.4837	0.6190
	Image10	0.4286	0.4643	0.7870	0.7500	0.2381	0.3333	0.6824	0.5238
	Image11	0.2857	0.2143	0.5977	0.7500	0.2381	0.1429	0.4586	0.6190
	Image13	0.1786	0.3214	0.4388	0.6071	0.2381	0.2381	0.3333	0.3333
	Image14	-0.0357	0.1429	0.5017	0.6071	0.0476	0.2381	0.4014	0.5238
	Image15	0.5000	0.5357	0.5561	0.8571	0.3333	0.4286	0.4603	0.7143
	Image16	0.1429	0.6786	0.6933	0.5357	0.0476	0.5238	0.5605	0.4283
	Image19	0.2143	0.6429	0.7619	1.0000	0.1429	0.4286	0.6281	1.0000
	Image22	0.5357	0.3929	0.7519	0.7143	0.3333	0.2381	0.6441	0.5238
	Image23	0.4643	0.1429	0.7179	1.0000	0.3333	0.0476	0.6286	1.0000
	Image24	0.8571	0.7500	0.5969	0.4643	0.7143	0.6190	0.4649	0.4286
PI Average		0.3061	0.3444	0.6406	0.7321	0.2381	0.2585	0.5295	0.6054
SI	Image2	0.2143	0.9286	0.9492	0.6429	0.1429	0.8095	0.8947	0.5238
	Image5	0.4286	0.6071	0.8321	0.8214	0.3333	0.4286	0.7048	0.6190
	Image6	0.3571	0.8929	0.8553	0.8571	0.3333	0.8095	0.7243	0.7143
	Image7	0.6786	0.2143	0.7279	0.8214	0.5238	0.1429	0.5964	0.6190
	Image8	0.2857	0.7500	0.8797	0.8214	0.2381	0.5238	0.7744	0.6190
	Image12	0.5000	0.5714	0.8384	0.8214	0.3333	0.4286	0.6916	0.7143
	Image17	-0.0357	0.0714	0.7161	0.2143	-0.0476	0.0476	0.6000	0.1429
	Image18	0.4643	0.2500	0.5018	0.8571	0.2381	0.1429	0.3857	0.7143
	Image20	0.4643	0.5714	0.8095	0.7143	0.3333	0.4286	0.6780	0.5238
	Image21	-0.0714	0.7857	0.8008	0.9286	-0.0476	0.6190	0.6942	0.8095
	SI Average		0.3286	0.5643	0.7911	0.7500	0.2381	0.4381	0.6744
Overall		0.3155	0.4360	0.7033	0.7396	0.2381	0.3333	0.5898	0.6032

Čadík's database [16] and obtain a 91.70% classification accuracy. We also test the feature on a recently published database named COLOR250 [14], which contains 250 images with 200 photographic and 50 synthetic ones. We obtain a 97.20% classification accuracy, which again verifies the effectiveness of this approach.

2) *Individual Contributions*: To further investigate the individual contributions of the luminance, contrast and structure components in C2G-SSIM, the SRCC values between the subjective ranking scores and any combination of the three components are given in [16, Table VI]. It can be observed that 1) the luminance and structure measures alone are relatively

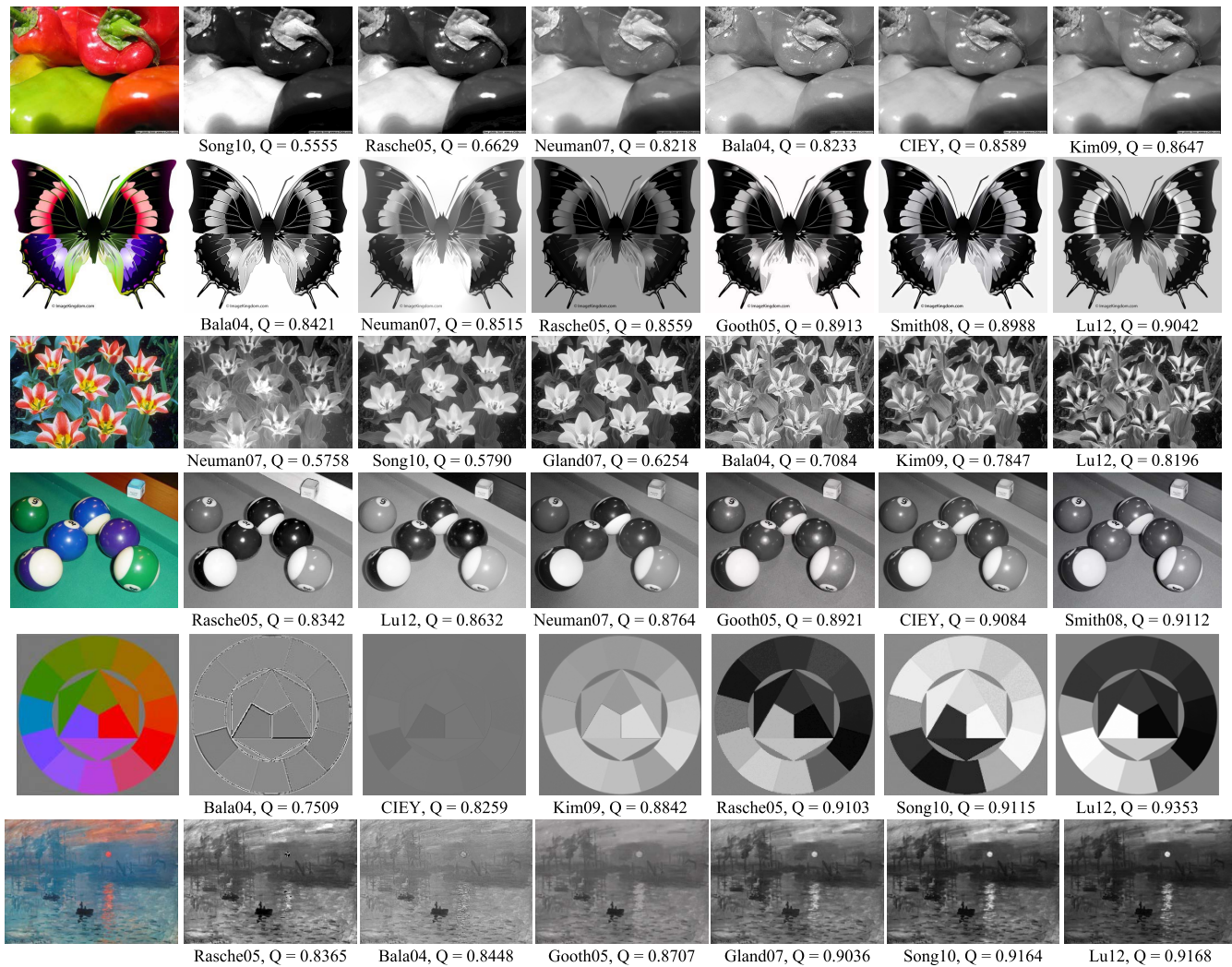


Fig. 6. Visual demonstration of C2G-SSIM. Bala04, Rasche05, Gooth05, Gland07, Smith08, Kim09, Song10 and Lu12 are algorithms from [1], [3], [4], [5], [8], [7], [10], and [11], respectively. CIEY is the luminance channel of CIEXYZ color space.

TABLE V

STATISTICAL SIGNIFICANCE MATRIX BASED ON QUALITY PREDICTION RESIDUALS. A SYMBOL “1” MEANS THAT THE PERFORMANCE OF THE ROW MODEL IS STATISTICALLY BETTER THAN THAT OF THE COLUMN MODEL, A SYMBOL “0” MEANS THAT THE ROW MODEL IS STATISTICALLY WORSE, AND A SYMBOL “-” MEANS THAT THE ROW AND COLUMN MODELS ARE STATISTICALLY INDISTINGUISHABLE

	Random guess	RWMS	E-score	C2G-SSIM
Random guess	--	11	00	00
RWMS	00	--	00	00
E-score	11	11	--	00
C2G-SSIM	11	11	11	--

TABLE VI

CONTRIBUTIONS OF INDIVIDUAL COMPONENTS AND THEIR COMBINATIONS IN TERMS OF SRCC ON [16]

Components	accuracy		preference	
	PI	SI	PI	SI
Luminance	0.5485	-0.1214	0.4770	-0.1464
Contrast	0.1633	0.3821	0.1735	0.3107
Structure	0.6862	0.4750	0.6939	0.5036
Luminance&Contrast	0.5026	—	0.3980	—
Luminance&Structure	0.7372	—	0.6811	—
Contrast&Structure	0.6276	0.7750	0.6250	0.7500
All	0.7832	—	0.7321	—

better predictors of the perceptual quality of the PI subset; 2) contrast measure plays a more important role for the SI subset; 3) by combining all three components together, the overall quality prediction is improved. This suggests that the three components are all useful and complementary to each other.

3) *Robustness Against Window Size W*: In order to investigate the robustness of C2G-SSIM to variations of the Gaussian window size W , we test its SRCC and KRCC performance variations as functions of W on *accuracy* and *preference* tests. The results are plotted in Fig. 7, from which we have two useful findings. First, the performance of C2G-SSIM is robust to variations of W . Second, medium window sizes have slightly better performance than that of small or large sizes.

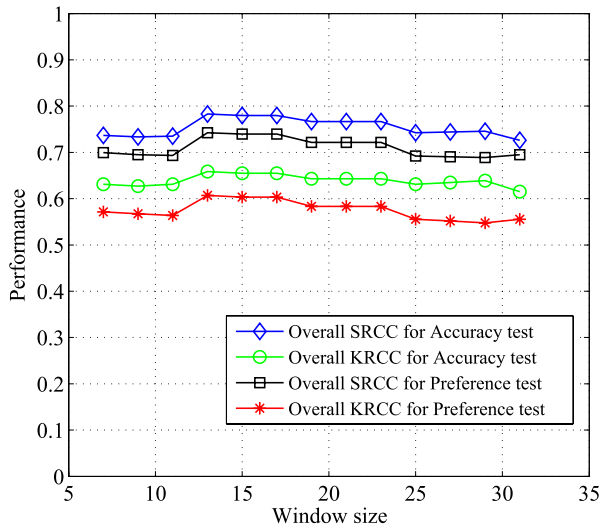


Fig. 7. Performance of C2G-SSIM versus window size.

This makes sense because medium window sizes achieve a good compromise between covering a sufficient neighboring region surrounding the center pixel and excluding faraway pixels that are less correlated with the center pixel.

4) *Computational Complexity*: The computational complexity of C2G-SSIM increases linearly with the number of pixels in the image. Our unoptimized MATLAB implementation takes around 9 seconds for a 225×312 image on a computer with an Intel(R) Core(TM) i5-3320M CPU at 2.60 GHz.

V. POTENTIAL APPLICATIONS OF C2G-SSIM

An objective quality assessment model can not only serve as a benchmark of image processing algorithms and systems. More importantly, an effective objective quality model may guide the development of novel image processing algorithms. In this section, we provide two examples to demonstrate the potential applications of the C2G-SSIM index.

A. Parameter Tuning of C2G Conversion Algorithms

Many state-of-the-art C2G conversion algorithms are parametric. Parameters in those algorithms can be roughly categorized into two types: one type controls the influence of the chromatic contrast on the luminance component [4], [7], [8]; the other enables certain flexibility of the implementation [5], [11]. These parameters are typically user-specified, but it is often a challenging task to find the best parameters without manually testing many options. The reason is that the performance of these C2G conversion algorithms are often highly parameter-sensitive, such that different parameter settings could lead to drastically different results. An objective IQA model, such as C2G-SSIM, provides a useful tool to automatically pick the best parameters without involving human intervention. To demonstrate this potential, we adopt C2G-SSIM to tune the parameter of the C2G conversion algorithm in [11], which includes one parameter σ to adjust the shape of the cost function. There is no suggested span of σ in [11] and the default value is given

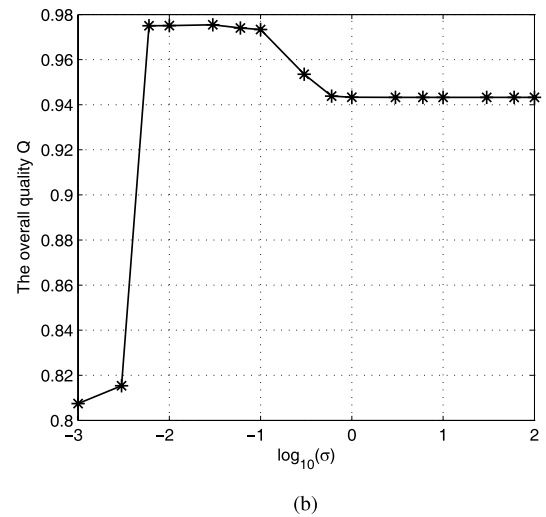
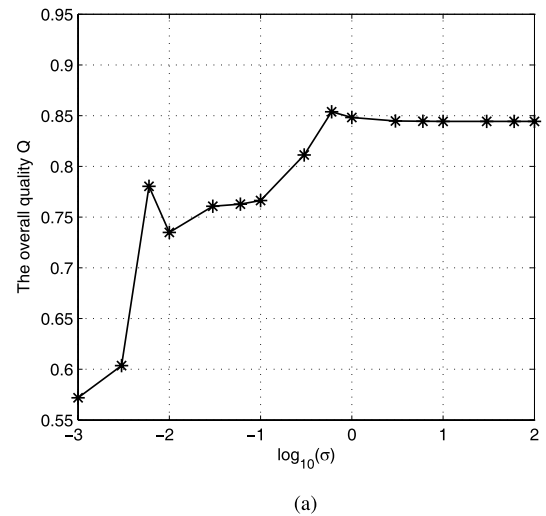


Fig. 8. The Q values versus the parameter σ in [11] for (a) “Goods” and (b) “Image6” images. The optimal values of σ for (a) and (b) are around 0.6 and 0.03 with the corresponding Q values 0.8539 and 0.9755, respectively.

by $\sigma = 0.02$. In our experiment, we test a wide range of σ values between $[0.001, 100]$ and pick the best one in terms of C2G-SSIM.

Fig. 8 depicts the overall quality value Q under different σ values for photographic image “Goods” and synthetic image “Image6” (image No. in [16]). An important observation is that the optimal values for the two images are different: $\sigma = 0.6$ for “Goods” and $\sigma = 0.03$ for “Image6”, respectively. These results demonstrate that the default empirical value $\sigma = 0.02$ in [11] has major difficulty in adapting to different image content. A few representative C2G images and their corresponding quality maps with different σ values are shown in Fig. 9 and Fig. 10, respectively. It can be clearly seen that for photographic image “Goods”, the best σ results in good balance between preserving the structural information and producing consistent luminance; for synthetic image “Image6”, the best σ leads to excellent preservation of structural details and contrast enhancement for better visibility. All of these are well captured by C2G-SSIM.

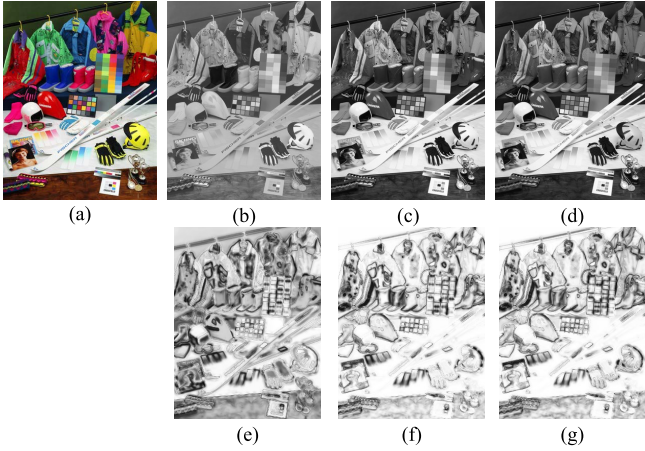


Fig. 9. C2G images created with different σ values in [11] and their corresponding C2G-SSIM quality maps. (a) Reference color image “Goods”. (b) $\sigma = 0.01$ and $Q = 0.7349$. (c) $\sigma = 0.6$ and $Q = 0.8539$. (d) $\sigma = 30$ and $Q = 0.8443$. (e)-(g) are the corresponding quality maps of (b)-(d).

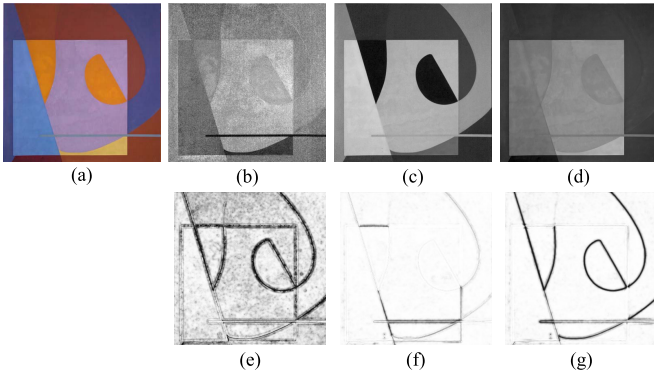


Fig. 10. C2G images created with different σ values in [11] and their corresponding C2G-SSIM quality maps. (a) Reference color image “Image6”. (b) $\sigma = 0.001$ and $Q = 0.8074$. (c) $\sigma = 0.03$ and $Q = 0.9755$. (d) $\sigma = 10$ and $Q = 0.9432$. (e)-(g) are the corresponding quality maps of (b)-(d).

B. Adaptive Fusion of C2G Images

The analysis of individual images reveals that no existing single C2G conversion algorithm produces universally good results for all test images. For a single C2G image, the best conversion may also vary if different regions of the image have substantially different types of content. In addition, the merit of different C2G conversion algorithms may complement one another. For instance, trivial C2G conversion algorithms such as MATLAB *rgb2gray* function typically retain the luminance channel, which gives more emphasis on the luminance consistency of the image at the risk of losing the distinction between two spatially adjacent colors of similar luminance. Some recent advanced C2G conversion algorithms [10], [11], however, aim at maximally preserving the original color contrast while ignoring the luminance component of the image. This motivates us to employ image fusion algorithms to integrate multiple C2G images, where C2G-SSIM could play an important role in such an adaptive fusion process.

With multiple candidate images available, a natural and widely used framework to fuse them is weighted average, where the weights are determined by the quality of each image.

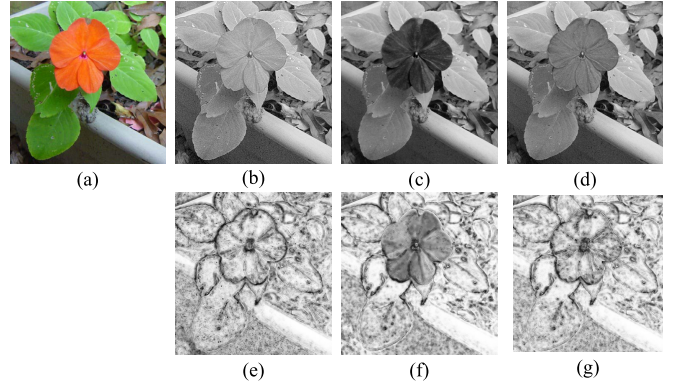


Fig. 11. Adaptive fusion of C2G images. (a) Reference color image “Flower”. (b) C2G image generated by [1], $Q = 0.7208$. (c) C2G image generated by [5], $Q = 0.7426$. (d) Fused image using C2G-SSIM, $Q = 0.7807$. (e)-(g) are the corresponding quality maps of (b)-(d).

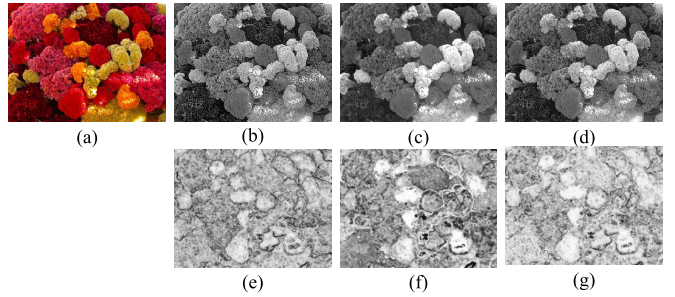


Fig. 12. Adaptive fusion of C2G images. (a) Reference color image “Flowers”. (b) C2G image generated by [6], $Q = 0.6543$. (c) C2G image generated by [1], $Q = 0.6673$. (d) Fused image using C2G-SSIM, $Q = 0.7585$. (e)-(g) are the corresponding quality maps of (b)-(d).

C2G-SSIM provides an ideal fit to this framework because it produces a spatial quality map that allows for spatially adaptive weight assignment. Specifically, assume $g_i(\mathbf{x})$ is the i th C2G image to be fused with its C2G-SSIM quality map $q_i(\mathbf{x})$, then a fused image is created by

$$g_F(\mathbf{x}) = \frac{\sum_{i=1}^N \max\{q_i(\mathbf{x}), c\} g_i(\mathbf{x})}{\sum_{i=1}^N \max\{q_i(\mathbf{x}), c\}}, \quad (20)$$

where c is a small positive integer (e.g., 10^{-6}) to avoid instability at the denominator.

Fig. 11 and Fig. 12 illustrate two examples of photographic images “Flower” and “Flowers”, respectively. In Fig. 11, two C2G images are shown, where (b) gives more luminance consistent appearance while (c) better preserves the color contrast. Careful comparison between the fused image and the two input images shows that the fused image achieves a better balance between structural preservation and luminance consistency, leading to better perceptual quality. Similarly, the perceptual quality of the fused image in Fig. 12 is improved upon (b) and (c) by C2G-SSIM weighted fusion.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we develop an objective IQA model, namely C2G-SSIM, to assess the perceptual quality of C2G images

using the original color image as reference. C2G-SSIM evaluates luminance, contrast and structure similarities between the reference color image and the C2G image. Image type dependent combination is then applied to yield an overall quality measure. The proposed C2G-SSIM index compares favorably against an average subject based on the database in [16] and significantly outperforms existing objective quality metrics for C2G conversion. Moreover, we use two examples to demonstrate the potential applications of C2G-SSIM.

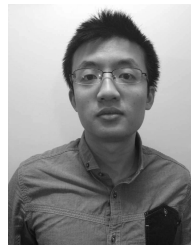
We consider C2G-SSIM as one of the initial attempts of C2G IQA, based on which future work may improve the performance in the following aspects:

- The current contrast and structure measures are based on color difference undergoing a nonlinear mapping. Since the response of the HVS to color difference of complex visual stimuli remains an active research topic, more advanced and accurate estimates of color image contrast and structure may be further investigated.
- Simple averaging is currently adopted to pool the C2G-SSIM map into a single score. Advanced pooling strategies taking into account visual attention is worth investigating to improve the quality evaluation performance.
- As has been shown, for different images the impacts of the luminance, contrast and structure measures may be different depending on the image content. As a result, a parameter α needs to be predetermined between the range of $[0, 1]$. Currently, this is done by assuming a known image type and associating each image type with a fixed α value, or by estimating a binary choice of α based on the proposed entropy feature. The binary estimation of α has its limitations, because in practice there may be mixed-class images. A continuous value of α is worth exploring in future research, which could result in possibly an averaging effect of the PI and SI cases, and might give a better prediction of the MOS of a mixed-class test image. However, as discussed in Section III, human subjects tend to use distinct strategies to make the judgement about PI and SI. As a result, a continuous value of α could give a score at the average point of a bi-modal distribution and may not reflect the opinion of a typical human subject. In the future, localized approaches for the classifications of PI and SI, and human behaviors on assessing C2G images need to be further investigated.

REFERENCES

- [1] R. Bala and R. Eschbach, "Spatial color-to-grayscale transform preserving chrominance edge information," in *Proc. 12th Color Imag. Conf., Color Sci. Eng. Syst., Technol., Appl.*, 2004, pp. 82–86.
- [2] G. Wyszecki and W. S. Stiles, *Color Science*. New York, NY, USA: Wiley, 1982.
- [3] K. Rasche, R. Geist, and J. Westall, "Re-coloring images for gamuts of lower dimension," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 423–432, 2005.
- [4] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch, "Color2Gray: Saliency-preserving color removal," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 634–639, 2005.
- [5] M. Grundland and N. A. Dodgson, "Decolorize: Fast, contrast enhancing, color to grayscale conversion," *Pattern Recognit.*, vol. 40, no. 11, pp. 2891–2896, 2007.
- [6] L. Neumann, M. Čadík, and A. Nemcsics, "An efficient perception-based adaptive color to gray transformation," in *Proc. 3rd Eurograph. Conf. Comput. Aesthetics Graph., Visualizat., Imag.*, 2007, pp. 73–80.
- [7] Y. Kim, C. Jang, J. Demouth, and S. Lee, "Robust color-to-gray via nonlinear global mapping," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. ID 161.
- [8] K. Smith, P.-E. Landes, J. Thollot, and K. Myszkowski, "Apparent greyscale: A simple and fast conversion to perceptually accurate images and video," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 193–200, 2008.
- [9] G. R. Kuhn, M. M. Oliveira, and L. A. F. Fernandes, "An improved contrast enhancing approach for color-to-grayscale mappings," *Vis. Comput.*, vol. 24, nos. 7–9, pp. 505–514, 2008.
- [10] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Sep. 2010.
- [11] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization," in *Proc. IEEE Int. Conf. Comput. Photogr.*, Apr. 2012, pp. 1–7.
- [12] Y. Song, L. Bao, X. Xu, and Q. Yang, "Decolorization: Is rgb2gray() out?" in *Proc. ACM SIGGRAPH Asia Tech. Briefs*, 2013, Art. ID 15.
- [13] M. Zhou, B. Sheng, and L. Ma, "Saliency preserving decolorization," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.
- [14] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization with perception-based quality metrics," *Int. J. Comput. Vis.*, vol. 110, no. 2, pp. 222–239, 2014.
- [15] D. Eynard, A. Kovnatsky, and M. M. Bronstein, "Laplacian colormaps: A framework for structure-preserving color transformations," *Comput. Graph. Forum*, vol. 33, no. 2, pp. 215–224, 2014.
- [16] M. Čadík, "Perceptual evaluation of color-to-grayscale image conversions," *Comput. Graph. Forum*, vol. 27, no. 7, pp. 1745–1754, 2008.
- [17] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment* (Synthesis Lectures on Image, Video, and Multimedia Processing), vol. 2. San Rafael, CA, USA: Morgan & Claypool, 2006, no. 1, pp. 1–156.
- [18] 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, Jan. 2009.
- [19] 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.
- [20] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.
- [21] A. Nemcsics, "Recent experiments investigating the harmony interval based color space of the Coloroid color system," in *Proc. Int. Soc. Opt. Photon. 9th Congr. Int. Color Assoc.*, 2002, pp. 865–868.
- [22] Y. Nayatani, "Simple estimation methods for the Helmholtz–Kohlrausch effect," *Color Res. Appl.*, vol. 22, no. 6, pp. 385–401, 1997.
- [23] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *J. Vis.*, vol. 8, no. 12, p. 8, 2008.
- [24] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.
- [25] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.
- [26] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.
- [27] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3003–3010.
- [28] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015.
- [29] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment for image super-resolution: A natural scene statistics approach," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1481–1484.
- [30] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment of interpolated natural images," *IEEE Trans. Image Process.*, to be published.

- [31] 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.
- [32] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image tone mapping by optimizing tone mapped image quality index," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.
- [33] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
- [34] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganriere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.
- [35] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *Proc. 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 7–12.
- [36] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [37] R. Hassen, Z. Wang, and M. M. A. Salama, "Objective quality assessment for multiexposure multifocus image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2712–2724, Sep. 2015.
- [38] D. B. Judd, "Ideal color space: Curvature of color space and its implications for industrial color tolerances," *Palette*, vol. 29, no. 21–28, pp. 4–25, 1968.
- [39] G. Sharma and R. Bala, Eds., *Digital Color Imaging Handbook*. Boca Raton, FL, USA: CRC Press, 2010.
- [40] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, "The CIECAM02 color appearance model," in *Proc. Soc. Imag. Sci. Technol. 10th Color Imag. Conf.*, 2002, pp. 23–27.
- [41] 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, Mar. 2012.
- [42] H. Wang, G. Cui, M. R. Luo, and H. Xu, "Evaluation of colour-difference formulae for different colour-difference magnitudes," *Color Res. Appl.*, vol. 37, no. 5, pp. 316–325, 2012.
- [43] J. Hardeberg, *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches*. Boca Raton, FL, USA: Universal-Publishers, 2001.
- [44] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th IEEE Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
- [45] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
- [46] A. C. Bovik, *Handbook of Image and Video Processing*. New York, NY, USA: Academic, 2010.
- [47] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Jan. 2013.
- [48] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, vol. 72. Bellingham, WA, USA: SPIE, 1999.
- [49] W. S. Geisler and M. S. Banks, "Visual performance," in *Handbook of Optics*, vol. 1, 2nd ed. New York, NY, USA: McGraw-Hill, 1995.
- [50] 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.
- [51] L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, 1927.
- [52] VQEG. (Apr. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: <http://www.vqeg.org>
- [53] *Contrast Preserving Decolorization*. [Online]. Available: <http://www.cse.cuhk.edu.hk/leojia/projects/decolorization/index.htm>, accessed Jun. 1, 2015.
- [54] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.



Kede Ma (S'13) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interest lies in perceptual image processing.



Tiesong Zhao (S'08–M'12) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. From 2011 to 2012, he was a Research Associate with the Department of Computer Science, City University of Hong Kong. Then until 2013, he served as a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. He is currently a Research Scientist with the Department of Computer Science and Engineering, State University of New York at Buffalo, NY, USA. His research interests include image/video processing, visual quality assessment, video coding, and transmission.



Kai Zeng received the B.E. and M.A.Sc. degrees in electrical engineering from Xidian University, Xi'an, China, in 2009, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2013, where he is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering. His research interests include computational video and image pattern analysis, multimedia communications, and image and video processing (coding, denoising, analysis, and representation), with an emphasis on image and video quality assessment and corresponding applications. He was a recipient of the IEEE Signal Processing Society student travel grant at the 2010 and 2012 IEEE International Conference on Image Processing, and the prestigious 2013 Chinese Government Award for Outstanding Students Abroad.



Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications, and biomedical signal processing. He has over 100 publications in these fields with over 25 000 citations (Google Scholar). He is a member of the IEEE Multimedia Signal Processing Technical Committee (2013–2015). He served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), *Pattern Recognition* (2006–present), and the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009), the *EURASIP Journal of Image and Video Processing* (2009–2010), and *Signal, Image and Video Processing* (2011–2013). He was a recipient of the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Best Magazine Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as a Senior Author).