Evaluation of color differences in natural scene color images

B. Ortiz-Jaramillo, A. Kumcu, L. Platisa and W. Philips

Authors are with imec - IPI, Department of Telecommunications and Information Processing at Ghent University

Abstract

Since there is a wide range of applications requiring image color difference (CD) assessment (e.g. color quantization, color mapping), a number of CD measures for images have been proposed. However, the performance evaluation of such measures often suffers from the following major flaws: (1) test images contain primarily spatial- (e.g. blur) rather than color-specific distortions (e.g. quantization noise), (2) there are too few test images (lack of variability in color content), and (3) test images are not publicly available (difficult to reproduce and compare). Accordingly, the performance of CD measures reported in the state-of-the-art is ambiguous and therefore inconclusive to be used for any specific color-related application.

In this work, we review a total of twenty four state-of-the-art CD measures. Then, based on the findings of our review, we propose a novel method to compute CDs in natural scene color images. We have tested our measure as well as the state-of-the-art measures on three color related distortions from a publicly available database (mean shift, change in color saturation and quantization noise). Our experimental results show that the correlation between the subjective scores and the proposed measure exceeds 85% which is better than the other twenty four CD measures tested in this work (for illustration the best performing state-of-the-art CD measures achieve correlations with humans lower than

*Corresponding author
Email address: `benhur.ortizjaramillo@ugent.be` (B. Ortiz-Jaramillo)

80%).

Keywords: Color appearance, color difference assessment, fidelity assessment

1. Introduction

Nowadays, fidelity assessment of images in terms of color or simply assessment of color differences (CDs) in images has become an active area in the research of color science and imaging technology due to its wide range of applications such as color correction [1, 2], color quantization [3], color mapping [4], color image similarity and retrieval [5]. For instance, in multiview imaging, color correction is used to eliminate color inconsistencies between views. In that application, the fidelity assessment of color corrected images relative to the current view image can be used to select the color correction algorithm that produces the smallest perceived color differences. In color mapping and color quantization algorithms, pixel colors are replaced following certain criteria while they ensure a good correspondence in terms of perceived color between the original image and its reproduction. There, CD assessment can be used to find the appropriate quantization step size and/or range of displayable colors to obtain the reproduction with the minimum perceived CD. Another example is color image similarity and retrieval where all images with color composition similar to the query image are retrieved from a database. Thus, the assessment of CDs between images is very important to identify the images with color content similar to that of the query image.

While many CD measures for natural scene color images have been proposed, there has not yet been any rigorous investigation into the performance comparison of the existing measures [6, 7, 8, 9, 10]. The CD measures in the state-of-the-art are often tested on databases which: (1) contain multiple distortions in combination with the color-related distortions, (2) include few test image samples, and/or (3) are not publicly available but rather kept private. Additionally, the performance of the CD measures is often reported as average performance over all images of a given database. Overall, to the best of our knowledge, there is little research addressing the problem of reviewing and especially testing CD measures and the existing reports are very limited in test samples and/or CD measures. Also, the majority of studies in the state-of-the-art are devoted to evaluating and comparing measures of image quality and not measures of image CD. For instance, in [11], 60 image quality measures (of which 28 based on gray scale image data) were tested on a publicly available database of images. It is important to note that for that dataset the human scores are related to the overall image quality rather than to the overall image differences. Another important aspect of any benchmark image quality database is the type of the image distortions. The database from Ref [11] includes mostly spatial image distortions, e.g., compression artifacts, noise and blur. In this work, we focus on the CD measures; for the readers interested in image quality measures we recommend the references [12, 13, 14, 11, 15, 16].

In order to address the limitations of the current literature, we take into account various types of CD measures and test those using a public image database which addresses specifically color related image alterations. Specifically, our analysis includes 25 source images which leads to more generalizable results compared to the 6 or 8 source images presented in the other related works [17, 7, 18, 19]. The works presented in the Ref [20] and more recently in Ref [21] used more reference images (respectively, 97 and 25) to evaluate color gamut mapping algorithms, yet they considered more image quality measures than dedicated measures of CD. Firstly, we conduct a brief review in color science for evaluating CDs. Thereafter, we evaluate the twenty four state-of-the-art CD measures and discuss their performances as well as investigate the specific cases where the CD measures fail in order to objectively assess the strengths and weaknesses of the tested measures. We made these measures freely available as a plugin on the iFAS [22] software tool.

Additionally, we propose a novel method to compute color differences in natural scene color images based on the findings of the review. We base our measure on the fact that humans assess color differences in natural scene color images by comparing sets of connected pixels or small patches. Those patches are typically characterized for being homogeneous or for possessing an unique texture pattern. Therefore, we use image segmentation based on texture to compute the color differences in the resulting segments. Particularly, we use the Local Binary Patterns as texture descriptor because of its simplicity while being one of the most accurate texture analysis algorithms [23]. To compute the color differences we use the statistics proposed in [24] because they are good measures of the change in the color distribution spread and severe color differences. For computing the intensity differences, we use the well known structural similarity index measure (SSIM) [25]. Finally, the overall color difference is computed as the weighted average of the local differences using as weights the ratio between the number of pixels in the patch and the total number of pixels in the image.

We have tested our measure as well as the state-of-the-art measures on three color related distortions (mean shift, change in color saturation and quantization noise) from one image quality assessment database (TID2013 [26]). We found that the proposed measure is able to accurately predict the color differences typically perceived and reported by a human observer. Particularly, our experimental results show that the correlation between the subjective scores and the proposed measure exceeds 85% which is better than the other twenty four CD measures tested in this work (for illustration the best performing state-of-the-art CD measures achieve correlation with humans lower than 80%).

This work is organized as follows. In Section 2, current approaches dealing with CD assessment in natural scene color images are discussed. The novel methodology is described in Section 3. Thereafter, in Section 4, we present and discuss the results obtained in our experimental study. Finally, we draw conclusions in Section 5.

2. Background

The Commission Internationale de l'Eclairage (CIE) defines color as: "*attribute of visual perception consisting of any combination of chromatic and achromatic content.*" The definition implies that color is an attribute of visual perception, i.e., the study of color is mostly about perception (color appearance) [27]. The study of color appearance seeks to describe the perceptual aspects of human color vision. For instance, the most successful color appearance model (CAM) according to the reports from Refs [28, 29] is the CIELAB. Therefore, most of the CD formulas compute a certain distance measure in the CIELAB color space [30], that is, the color components are expressed in the CIELAB color space at the point of the computation of the specific distance formula, e.g., Mahalanobis, CIEDE2000, among others. Next to the CIELAB, also other CAMs have been proposed in the state-of-the-art such as $\mathrm{YC_BC_R}$ [31], HSI [32], $\ell\alpha\beta$ [33], CIELUV [34], OSA-UCS [35]. Further information about CAMs can be found in [30, 27, 36, 29].

We have explored twenty four color difference measures plus SSIM listed in Table 1. The ID is the identifier used in this work for referring to a specific CD measure. Color space is the color space or appearance model used for computing the CDs. Note that, we only consider here the color space where the actual color differences are computed. SP (Spatial processing) is whether or not neighboring pixels are taken into account in computing the CD measure. Overall CD describes the technique for computing the overall CD measure using the obtained differences.

Overall, we have found eight extensions of the CIEDE2000, four based on statistics of color components, three extensions of the SSIM, one based on discrete cosine transform, three based on weighted average and five based on other color appearance models. The explored measures use 8 CAMs: CIELAB (used by 11 out of 24 measures), 2-component opponent color space (OCC) (1), OSA-UCS (2), $\ell\alpha\beta$ (1), YIQ (1), $\mathrm{YC_BC_R}$ (2), HSI (1), IPT (2), LMN (1), gray scale (1) and RGB (1). For more information about these CAMs, the reader is referred to the original publications listed in Table 1. Note that the CIELAB appearance model is the most popular CAM for computing CDs in natural

[1]PSIM numerical values were obtained from the web page of its authors `https://sites.google.com/site/guke198701/publications`

Table 1: State-of-the-art summary studied in this work.

Measure name	ID	Color space	SP	Overall CD
CIEDE2000 formula [37]	CD00	CIELAB [34]	No	Average of pixel-wise CDs
Spatial-CIELAB [38]	CD01	CIELAB [34]	Yes	Average of pixel-wise CDs
Mahalanobis distance [39]	CD02	CIELAB [34]	No	Average of pixel-wise CDs
Colorfulness [40]	CD03	OCC [6]	No	Difference in global descriptive statistics
Colour image fidelity metric [41]	CD04	$\ell\alpha\beta$ [33]	Yes	Average of SSIM values
Chroma spread and extreme [24]	CD05	YC_BC_R [31]	Yes	Average differences between block-based color features
Histogram intersection [42]	CD06	CIELAB [34]	No	Color histogram intersection
Weighted CIELAB [43]	CD07	CIELAB [34]	Yes	Weighted average of pixel-wise CDs
Image appearance metric [44]	CD08	IPT [44]	Yes	Average of pixel-wise CDs
Just noticeable CD measure [45]	CD09	CIELAB [34]	Yes	Weighted Average of pixel-wise CDs
Chrominance component CD [46]	CD10	HSI [32]	No	Difference in global descriptive statistics
Adaptive image difference [8]	CD11	RGB [30]	Yes	Average of block based CDs
Spatial hue angle metric [47, 48]	CD12	CIELAB [34]	Yes	Weighted average of pixel-wise CDs
Color image difference [49]	CD13	CIELAB [34]	Yes	Average of pixel-wise CDs
Circular processing CD [10]	CD14	CIELAB [34]	Yes	Average of local CDs
OSA-UCS [35]	CD15	OSA-UCS [50]	No	Average of pixel-wise CDs
Spatial-OSA-UCS [51]	CD16	OSA-UCS [50]	Yes	Average of pixel-wise CDs
Spatial colour metric [52]	CD17	CIELAB [34]	Yes	Average of block based CDs
Proposed measure	CD18	YC_BC_R [31]	Yes	Weighted average of patch based CDs
SSIMipt [53]	CD19	IPT [44]	Yes	Average of SSIM values
colorPSNRHMA [54]	CD20	YC_BC_R [31]	Yes	Average difference of DCT coefficients
VSI [55]	CD21	LMN [56]	Yes	Weighted average of color differences
SSIM [25]	CD22	Gray scale	Yes	Average of local statistics
PSIM[1][16]	CD23	YIQ [57]	Yes	Average of color differences
CIEDE76 formula [34]	CD24	CIELAB [34]	No	Average of pixel-wise CDs

scene color images. 7 out of 24 measures do not consider any spatial processing. Finally, irrespective of whether the measure has spatial processing or not, the overall difference in 15 out of the 24 CD measures is computed as the average of the pixel-wise differences.

Traditionally, computing CDs in images has been accomplished by using a CD formula on a pixel-by-pixel basis (some algorithms consider image filtering to simulate the blur property of human eyes) and then examining statistics such as mean, median or maximum. However, subjective evaluation of perceived color differences has shown that, when observing a color image, the observer makes the color sensation from a number of pixels and not a single pixel color [58]. Also, the studies in color enhancement have shown that the perceived color by a human depends on the amount of spatial variation and texture in the scene [59, 60]. That is, two image patches can be perceived by a human as the same color only under the same spatial distribution of pixel color values. Additionally, the experiments carried out in [17, 58, 61] comparing color image differences showed that the observers tend to focus on certain areas of an image, usually, homogeneous areas or areas with the same texture pattern, and give their judgments mainly based on the color difference of those areas.

These findings show that the pixel-wise CDs (even after considering image filtering to simulate the blur property of human eyes) between two images do not represent the CD sensation perceived by a human observer and human observers judge CD in natural scene color images based on the comparison of image patches with similar texture pattern. For instance, the weighted CIELAB [43] is based on the fact that the CDs in larger areas with the same color should be weighted higher compared to those in smaller areas because human eyes tend to be more tolerant towards CDs in smaller areas. Moreover, our methodology agrees with other visual attention models based on saliency maps used in image quality measures such as those presented in [62, 63, 64, 65, 66, 67, 68], where larger homogeneous areas have more influence on the overall quality than highly textured small areas. Note that the tested state-of-the-art CD measures do not consider the texture of the image in the CD computation.

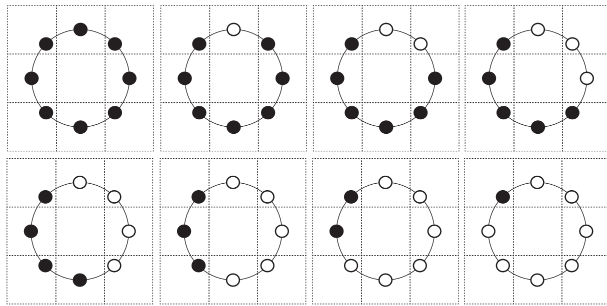

Figure 1: Texture primitives detected by the uLBP. Black points correspond to the binary value 0 while white points to 1.

3. Proposed method

In search for an adequate solution of the problem of computing color differences in natural scene color images, we propose a measure based on the fact that humans assess the differences in image color by comparing small image patches of similar texture. Therefore, we first look for an appropriate method to divide the image in patches with unique texture patterns to later compute the CDs on the obtained patches.

One common way of dividing an image into unique texture patterns is by using the well-known texture descriptors: the Local Binary Patterns (LBP). This method computes relative intensity relations between the pixels in a small neighborhood. See [23] for details about this texture analysis technique. In particular, experimental results over all possible LBP patterns have shown that the subset called "uniform" LBP (uLBP), introduced in [69], covers 90% of all patterns in natural scene images [69, 70]. A LBP pattern is called uniform if the pattern contains at most two $0-1$ or $1-0$ transitions. Figure 1 shows the texture primitives detected by the uLBP. The black points correspond to the binary value 0 and the white points to 1. Note that any other texture primitive can be obtained by rotating or complementing the binary primitives shown in Figure 1.

Figure 2 shows examples of texture primitives computed using the uLBP.

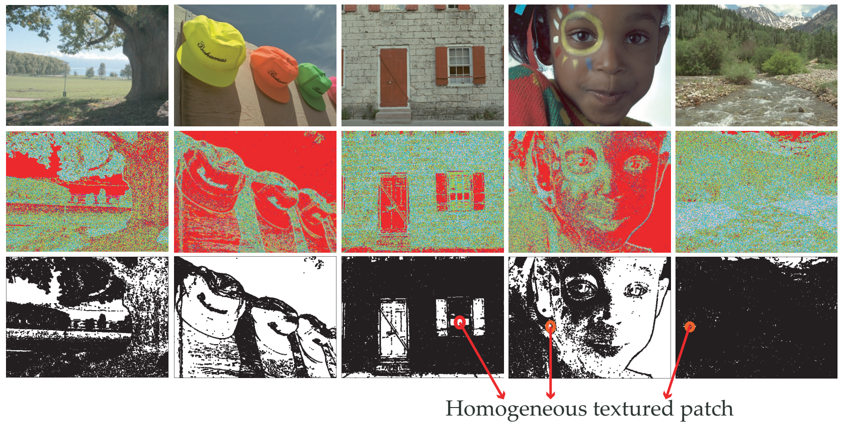

Figure 2: Example of texture primitives detected using uLBP. (top) sample image, (middle) uLBP primitives, (bottom) homogeneous patches for the first (top left corner) texture primitive from Figure 1. The encircled patches are examples of what we call homogeneous textured patches, i.e., a connected set of pixels with unique texture pattern.

In the top we show the sample images while in the middle their corresponding uLBP primitives. In the bottom we show all the textured patches equal to the first texture primitive from Figure 1. The encircled patches in Figure 2 are examples of what we call homogeneous textured patch, a set of connected pixels with an unique uLBP texture pattern.

After dividing the image into a set of unique texture patches using the uLBP descriptors, we are ready to perform the color comparison independently in each homogeneous textured patch. In this case, we can use one of the image CD indices explored in Section 2. Particularly, the statistics used in chroma spread and chroma extreme CD indices proposed by Pinson and Wolf [24] have shown to be good measures of the change of spread in the color distribution and severe color differences, respectively. Accordingly, we propose to measure the CDs in the resulting homogeneous textured patches using the linear combination of the chroma spread and chroma extreme indices because they capture color distribution parameters relevant to the humans [24]. For computing the differences in the intensity channel, we use the well-known structural similarity index measure

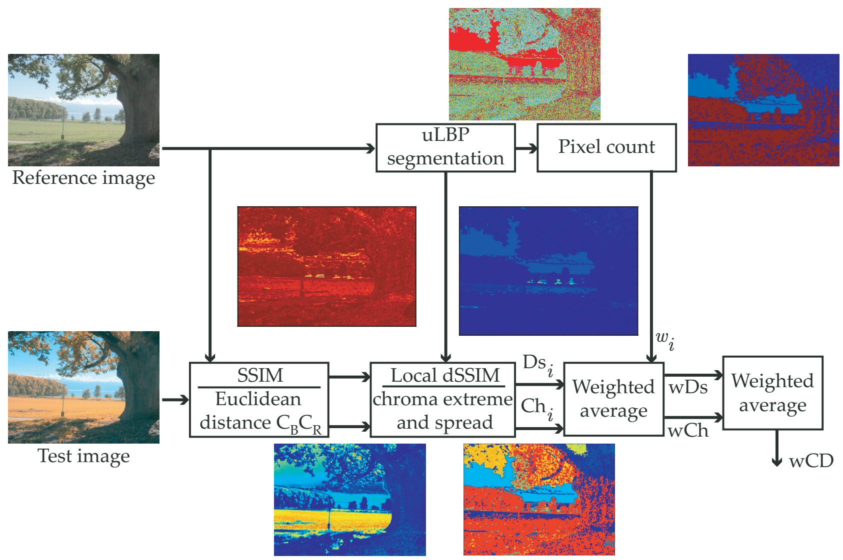

Figure 3: Block diagram of the proposed image CD measure.

(SSIM) [25].

Figure 3 shows the block diagram of the proposed methodology for computing color differences in natural scene color images. The computation of the proposed CD measure is summarized as follows.

1. The Reference and Test images are compared using the Euclidean distance of their corresponding $\mathrm{C_B}$ and $\mathrm{C_R}$ color components as well as using the SSIM between intensity components (Y).

2. The uLBP is computed from the reference image to obtain the set of homogeneous textured patches (uLBP segmentation in Figure 3).

3. In the *Local dSSIM, chroma extreme and spread* block, we compute for each homogeneous textured patch the chroma spread as the standard deviation of the resulting differences and the chroma extreme as the average of the worst 1% and subtract from it the 99% level [24]. Both indices are combined as the chroma spread-extreme index $\mathrm{Ch}_i = 0.0192\mathrm{Ch_s} + 0.0076\mathrm{Ch_e}$, for the ith homogeneous textured patch [24]. The linear combination was

obtained empirically by Pinson and Wolf using training samples from the VQEG FR-TV Phase II database [24]. Similarly, we compute for each homogeneous textured patch the average value of the SSIM after being transformed to dissimilarity, i.e., $\mathrm{Ds}_i = \frac{1-\overline{\mathrm{SSIM}}_i}{2}$, where $\overline{\mathrm{SSIM}}_i$ is the average SSIM of the ith homogeneous textured patch. That is, we compute the local average for each homogeneous textured patch using the obtained dSSIM.

4. The number of pixels in each homogeneous textured patch is count to be used as weights for the spatial pooling. The weights are computed as follows $w_i = \frac{n_i}{NM}$ where n_i is the number of pixels in the ith homogeneous textured patch, N and M are the number of rows and columns of the image, respectively. This assumption agrees with the well-known fact that human eyes tend to be more tolerant towards color difference of smaller image areas [17].

5. The global image color difference is computed as the weighted average of the resulting color differences per patch as

$$\mathrm{wCh} = \sum_{i=1}^{K} w_i \mathrm{Ch}_i,$$

$$\mathrm{wDs} = \sum_{i=1}^{K} w_i \mathrm{Ds}_i,$$

where Ch_i, Ds_i and w_i are the chroma spread-extreme index, the average dissimilarity index and the weight of the ith homogeneous textured patch for K patches, respectively. Note that the number of homogeneous textured patches (K) depends on the image content at hand. For illustration, we have found (from left to right) 4458, 2788, 3658, 3828 and 3652 homogeneous textured patches in the images from Figure 2.

Finally, the proposed global CD (ID: CD18) is computed as the weighted average of the two differences as follows

$$\mathrm{wCD} = \alpha \mathrm{wCh} + \beta \mathrm{wDs}, \tag{1}$$

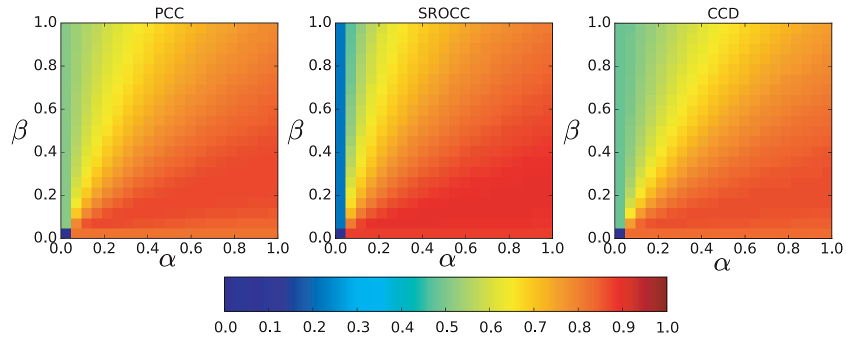

Figure 4: Performance of the proposed CD measure appraised on the test data of TID2013 database in function of the parameters α and β. Performance is given in terms of the PCC, the SROCC and CCD between the resulting CD measure and the corresponding subjective scores.

where α and β are weights that can be adjusted according to the application. In this case, since we are interested in evaluating color differences we give more importance to the color component, i.e., empirically we select the following weights: $\alpha = 0.7$ and $\beta = 0.3$.

Figure 4 shows the correlation between the humans scores in the test data of TID2013 database (see Section 4.2) and the proposed methodology in function of the parameters α and β. The highest correlation is achieved around the region of the selected parameter values ($\alpha = 0.7$ and $\beta = 0.3$). Also note that the performance decreases when a higher weight is assigned to the differences computed in the intensity component of the image. Additionally, this experiment shows that it is possible to further investigate and tune α and β for different applications according to the importance of the differences in the individual color components.

4. Results and Discussion

In this Section we describe the used test images and the performance comparison with the state-of-the-art measures. The performance comparison is made in terms of correlation indices computed between the CD measures and the subjective scores, which are considered as ground truth. The value of 1 indicates high correlation and 0 is no correlation between the tested CD measure and the subjective scores.

The following parameters corresponding to the standard viewing conditions are used in our experiments. The level of ambient illumination is set to low according to the ITU recommendations (4 lux) [71]. The chromaticity of the white displayed on the color monitor was D65 and luminance level of the monitor was around 80 cd/m^2. All settings are suited for sRGB color space. In this work, we have assumed that the distance to the monitor was set to 75 cm [49]. All methods using SSIM measure (including the proposed methodology) are set to the standard parameters [25].

4.1. Evaluation method

We evaluate the CD measures by means of Pearson Coefficient of Correlation (PCC) [72], the Spearman's Rank Order Correlation Coefficient (SROCC) [73] and the Coefficient of Correlation of Distances (CCD) [74] between the subjective/human scores included with the dataset and the values given by the tested CD measures. In these measures, PCC and CCD measure the accuracy or the ability to predict the subjective fidelity scores with low error using linear models and non-linear models, respectively. SROCC measures the monotonicity or the degree to which predictions of the model agree with the magnitudes of subjective quality scores.

Since the PCC, the SROCC and the CCD values obtained in this work lead to analogous conclusions, we only describe our results in terms of the CCD but the analysis applies for all (PCC and SROCC) unless we indicate the opposite. We use the rule of the thumb for interpreting the size of a correlation coefficient [75], i.e., we use the following descriptive scale:

Size of Correlation	Interpretation
0.90 to 1.00	Very strong correlation
0.70 to 0.90	Strong correlation
0.50 to 0.70	Moderate correlation
0.30 to 0.50	Weak correlation
0.00 to 0.30	Very weak correlation

4.2. Test data

In order to carry out a meaningful performance analysis of a CD measure, the test images need to fulfill the minimal requirements: (1) the distortions present in the images are primarily affect color and not spatial properties of the images, and (2) the corresponding subjective quality scores are collected in the scenario which ensures that the human subject is evaluating the difference between two or more images (typically a test image and its corresponding reference image). The main reason for viewing and judging images in pairs is in the fact that the perceived CD highly depends on the appearance of the reference image. Moreover, we have chosen to work with the databases that are publicly available in order to ensure easy and simple data discovery for the readers who may be interested in replicating our experiments and/or comparing or results with other methods.

In this work the test data was selected to include the types of color alterations relevant for the most common applications considering CDs: color correction [1, 2], color quantization [3], color mapping [4], color image similarity and retrieval [5]. The output images in such tasks are typically affected by color modifications such as quantization noise, intensity shift, contrast change, change in color saturation and change in color balance [30, 76, 77, 1]. The considered dataset was obtained from one publicly available image quality database named TID2013 described in the following paragraphs (see [26] for details about this database).

TID2013 provides subjective scores, in terms of Mean Opinion Score (MOS), for comparing the performance between fidelity measures. The TID2013 contains 25 source images and 3000 distorted images (25 source images $\times$ 24 types of distortions × 5 levels of distortions). Source images are obtained from the Kodak Lossless True Color Image Suite. The complete list of the 24 distortions is included next, where the distortions marked in bold produce changes in color [26] 1) additive Gaussian noise, 2) **additive noise in color components**, 3) **spatially correlated noise**, 4) masked noise, 5) high frequency noise, 6) **impulse noise**, 7) **quantization noise**, 8) Gaussian blur, 9) image denoising, 10) **JPEG compression**, 11) JPEG2000 compression, 12) JPEG transmission errors, 13) JPEG2000 transmission errors, 14) non eccentricity pattern noise, 15) local block-wise distortions of different intensity, 16) **mean shift (intensity shift)**, 17) **contrast change**, 18) **change of color saturation**, 19) multiplicative Gaussian noise, 20) comfort noise, 21) lossy compression of noisy images, 22) **image color quantization with dither**, 23) **chromatic aberrations**, 24) sparse sampling and reconstruction.

For our experiments, the following distortion types were selected from the TID2013: quantization noise, mean shift (intensity shift), and change of color saturation. We selected this subset of distortions because they encompass the most important color related distortions in current imaging technologies for natural scene color images. For instance, quantization noise is closely related to color quantization. Intensity shift and change in color saturation are well-known distortions produced by color matching algorithms, color mapping algorithms and multiview imaging systems [76, 77, 1]. The remaining 21 distortions were not used in this work not even those affecting color because they incorporate also spatial distortions which typically impact the quality of the image much more strongly than color alteration. Therefore, the human scores would be then more likely predominantly influenced by the spatial distortions and not the color ones. For instance, we do not use chromatic aberrations and color quantization with dither because even though they have a large influence on color noise, they also produce strong artifacts of spatial nature such as blurring, false edges and/or rainbow edges which impact the "spatial" quality of the image much more strongly than its color alteration. Also, we have shown in previous research that contrast changes are better modeled by using the ratio of intensity

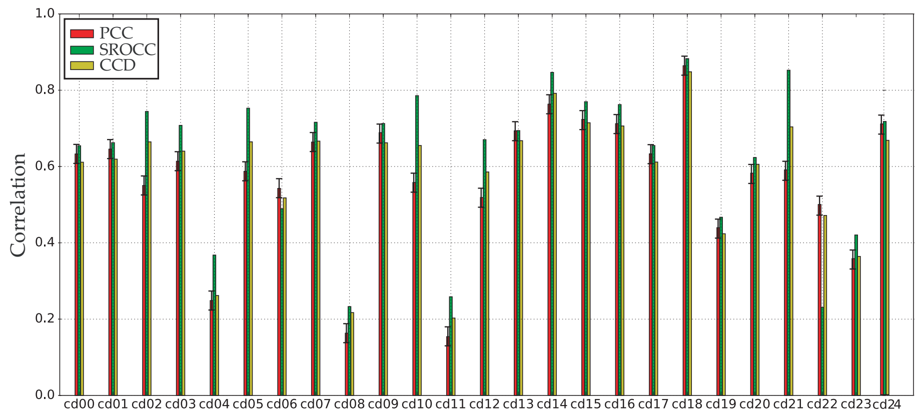

Figure 5: Performance of the considered 25 CD measures (24 existing and the proposed CD18) appraised on the test data of TID2013 database. Performance is given in terms of the PCC, the SROCC and CCD between a given CD measure and the corresponding subjective scores. Error bars are confidence intervals for the PCC values.

values [78, 79, 80]. Therefore, our test data is composed of 25 source images and their corresponding 375 distorted images (25 source images × 3 types of distortions × 5 levels of distortions); thus a total of 400 test images.

The MOS values from TID2013 were collected using a methodology known in psychophysics as two alternative forced choice (2AFC) match to sample [26]. In 2AFC three images are displayed (the reference and two distorted images) and an observer selects one of the two distorted images which they judge as more similar to the reference. That is, human observers are asked to select among two images the image that perceptually differs less from a reference [81]. Thus, the evaluation is made in terms of the presented current stimuli. Since the 2AFC was made within the selected subset of the TID2013, the MOS scores designated to that subset are a measure of the color difference with respect to the reference image perceived by the observers. Therefore, TID2013 allows the individual analysis of certain distortion type or subset of distortion types [26].

4.3. Overall performance of the tested CD measures

Figure 5 shows the PCC, the SROCC and the CCD appraised on the test data of TID2013 database. The best performing CD measures from the state-of-the-art are CD14 (Circular processing CD), CD15 (OSA-UCS), CD16 (Spatial-OSA-UCS), CD21 (VSI) and CD24 (CIEDE76) displaying a strong correlation. However, note that the proposed image CD measure (CD18) outperforms those CD image measures. Table 2 shows the percentage increase of the proposed method compared with the other state-of-the-art measures based on the correlation coefficients shown in Figure 5 after applying the Fisher's z transform. The Fisher's z transform is defined as

$$z' = 0.5 \log \left(\frac{1+r}{1-r} \right),$$

where r is the correlation coefficient. The percentage increase shows that the proposed methodology outperforms all other 24 image CD measures tested in this work.

The worst performance across the three color distortion types is achieved by CD08 (Image appearance metric), CD11 (Adaptive image difference), CD04 (Colour image fidelity metric) displaying a weak correlation. The poor performance of CD08 may be due to the fact that the measure focuses on complex spatial interactions such as perception of contrast, graininess, and sharpness while in fact it should focus on homogeneous textured areas [82]. Although CD11 is an adaptive technique, the CD measure is computed using the RGB color space which is well-known to disagree with human perception of color. CD04 performs better but still the correlation is weak compared with the other tested methods.

We also explore the performance of the tested CD measures on the individual distortion types to assess the strengths and weaknesses of the tested measures. Figures 6, 7 and 8 show the PCC, SROCC and CCD appraised on TID2013 database per individual color distortion type, color saturation, mean shift and quantization noise, respectively. In the quantization noise the best performing are CD20 (colorPSNRHMA), CD24 (CIEDE76) and CD05 (Chroma spread and

Table 2: Percentage increase of the performance appraised on TID2013 of the proposed color difference measure (CD18) compared with the state-of-the-art methods.

	Percentage increase				Percentage increase		
Measure ID	**PCC**	**SROCC**	**CCD**	**Measure ID**	**PCC**	**SROCC**	**CCD**
CD00	52	72	67	CD12	98	66	77
CD01	48	69	64	CD13	33	58	47
CD02	84	41	48	CD14	13	9	10
CD03	59	53	56	CD15	19	26	24
CD04	47	68	70	CD16	25	33	31
CD05	69	38	48	CD17	52	72	67
CD06	87	153	107	CD19	177	174	176
CD07	42	50	47	CD20	96	89	78
CD08	592	470	438	CD21	92	10	42
CD09	34	51	49	CD22	138	490	144
CD10	80	27	51	CD23	249	209	227
CD11	633	411	478	CD24	30	51	49

extreme) followed by CD09 (Just noticeable CD measure) and the proposed methodology CD18 (Figure 8). The proposed methodology shows to be the best performing in the color saturation subset with a strong correlation (correlation between the proposed CD measure and the subjective scores higher than 0.8), see Figure 6. Also, CD18 is one of the best performing methods together with CD13 (Color image difference) and CD24 (CIEDE76) in the mean shift subset (Figure 7).

Figure 9 shows the scatter plots of the proposed color image difference measure (CD18) and the subjective scores of the test data of TID2013 database. Note that the humans consider overall more annoying the color artifact produced by quantization noise (lower MOS) than the change of color saturation but they find overall the color saturation more annoying than mean shift distortion. This is also displayed by our proposed color difference measure (see Figure 9).

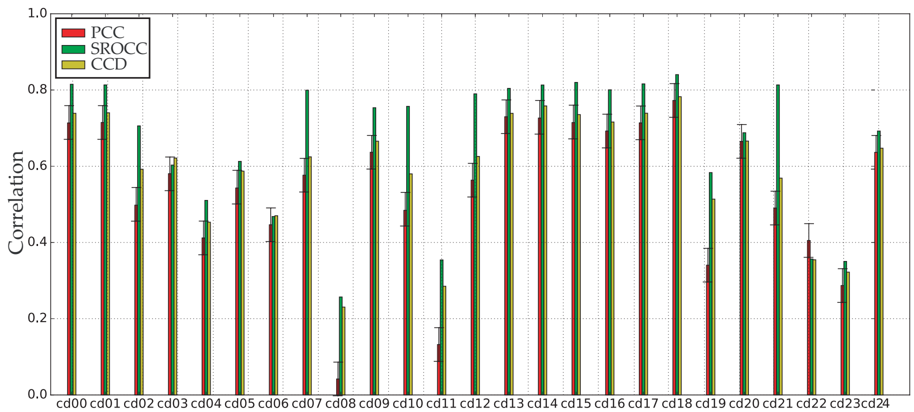

Figure 6: Performance of the considered CD measures appraised on TID2013 color saturation subset. Performance is given in terms of the PCC, the SROCC and CCD between a given CD measure and the corresponding subjective scores. Error bars are confidence intervals for the PCC values.

4.4. Discussion

Note that the good performance of CD05 (Chroma spread and extreme) in the quantization noise subset is partially due to the fact that CD05 compares the color distribution on the YCbCr color space (unlike any other of the considered state-of-the-art methods) and TID2013 quantization noise was processed on the same color space. This suggests that color quantization noise can be evaluated by comparing the color distribution of the images when the comparison is made on the same operational color space where the distorted image was processed. Indeed, since color quantization modifies considerably the distribution of the color histogram in the given color space, a comparison of the distribution in the same space comes forward as an appropriate tool for this type of task. However, CD05 performs poorly in the rest of the test data because the other color related distortions (mean shift and change in color saturation) do not have a considerably impact in the color histogram of the images making CD05 measure ineffective for this type of distortions.

Also note that there are no significant differences between CD00 (CIEDE2000), CD01 (Spatial-CIELAB) and CD17 (Spatial colour metric), i.e., there is a neg-

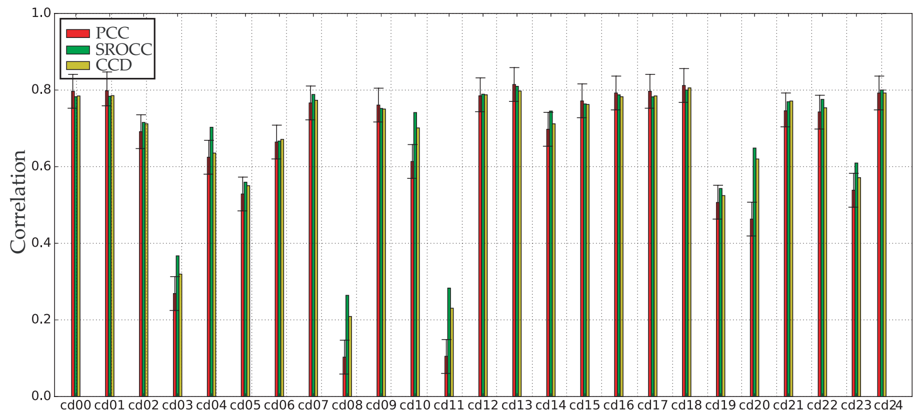

Figure 7: Performance of the considered CD measures appraised on TID2013 mean shift subset. Performance is given in terms of the PCC, the SROCC and CCD between a given CD measure and the corresponding subjective scores. Error bars are confidence intervals for the PCC values.

ligible improvement in terms of PCC, SROCC and CCD with subjective scores when a spatial filtering simulating blur property of human eyes is applied before computation of pixel wise differences (cf. the spatial processing described by [38]). We attribute this behavior to the fact that CDs are perceived easier in large homogeneous areas where there is no contrast masking while CDs in small textured areas with color fluctuations are more difficult to perceive than in large homogeneous areas. Therefore, the spatial processing (band-pass filtering simulating blur property of human eyes as proposed by [38]) displays negligible improvement in our experiments in terms of PCC, SROCC and CCD because the CD formulas are still applied pixel-wise instead of computing region based differences which is more appropriate due to the fact that humans perceive CDs easily in homogeneous textured areas. This is also confirmed by the results shown in Figures 5, 6, 7 and 8 where the proposed methodology (CD18) shows to be the best performing over all subsets of data.

The results show that overall, among all three considered sources of image color distortion, the best performing CD is the proposed methodology CD18 displaying a strong correlation with subjective scores in all test data. CD15

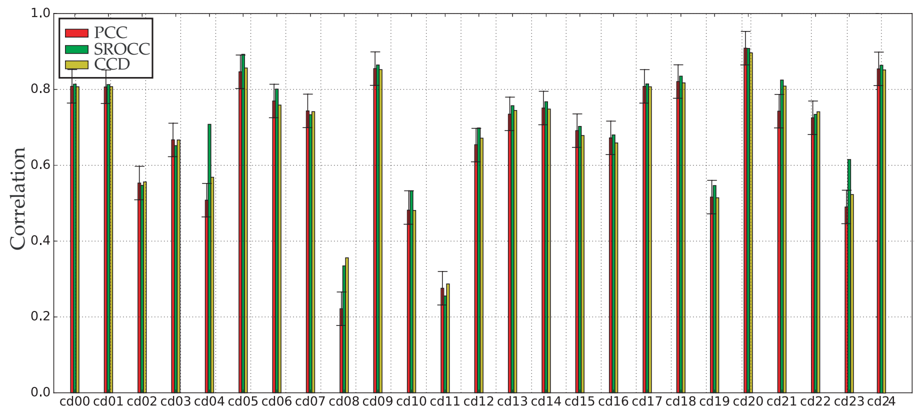

Figure 8: Performance of the considered CD measures appraised on TID2013 quantization noise subset. Performance is given in terms of the PCC, the SROCC and CCD between a given CD measure and the corresponding subjective scores. Error bars are confidence intervals for the PCC values.

(OSA-UCS), CD16 (Spatial-OSA-UCS), CD02 (Mahalanobis distance), CD03 (Colorfulness), CD04 (Colour image fidelity metric), CD05 (Chroma spread and extreme), CD06 (Histogram intersection), CD09 (Just noticeable CD measure) and CD10 (Chrominance component CD) display a moderate correlation with subjective scores. The worst performing methods are CD11 (Adaptive image difference) and CD08 (Image appearance metric) displaying a weak correlation with subjective scores in all test data.

Revising individual color distortions, the previous experiments and results reveal that CD00 (CIEDE2000), CD01 (Spatial-CIELAB), CD05 (Chroma spread and extreme), CD09 (Just noticeable CD measure), CD17 (Spatial colour metric), CD18 (proposed measure), CD20 (colorPSNRHMA) and CD24 (CIEDE76) are the best candidates to be used in color quantization applications displaying a strong correlation with subjective scores in the color quantization subset. Also, the results show that the best candidates to assess images affected by intensity shift are CD18 (proposed method), CD13 (Color image difference) and CD24 (CIEDE76) showing a strong correlation with subjective scores in the mean shift subset. Additionally, the following CD measures are the best candi-

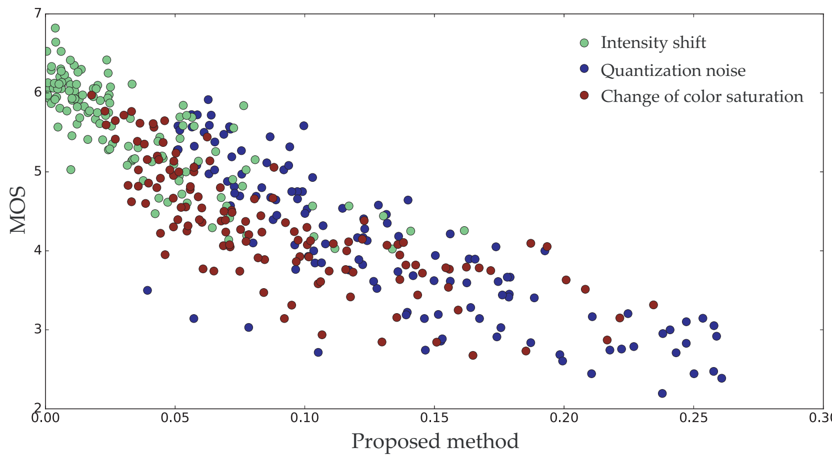

Figure 9: Scatter plots of the proposed color image difference measure (CD18) and the subjective scores of the test data of TID2013 database.

dates for assessing CDs on images affected by change of color saturation: CD00 (CIEDE2000), CD01 (Spatial-CIELAB), CD15 (OSA-UCS), CD14 (Circular processing CD), CD17 (Spatial colour metric), CD18 (proposed method) and CD21 (VSI) displaying a strong correlation with subjective scores (SROCC).

5. Conclusions

This work has reviewed and evaluated CD measures in the natural scene color images. We tested twenty four state-of-the-art CD measures on selected data from one public database. To stimulate further experimentation, we made all the tested methods freely available as a plugin on the iFAS [22] software tool. We selected our test image data such that the following applications are included: color correction, color quantization, color mapping, color image similarity and retrieval. The images in these applications are typically affected by CDs due to quantization noise, intensity shift, contrast change, change in color saturation and change in color balance. Moreover, we have proposed a novel methodology for computing color differences in natural scene color images based on the findings of the state-of-the-art review; the proposed method is named wCD (CD18).

Our experiments show that CD24 (CIEDE76), CD13 (Color image difference) and CD18 (proposed method) achieve a strong correlation with subjective scores in the mean shift subset. In the quantization noise the best performing are the CD20 (colorPSNRHMA), CD24 (CIEDE76), CD05 (Chroma spread and extreme) followed by CD09 (Just noticeable CD measure) and the proposed methodology CD18 displaying a strong correlation with subjective scores. The following CD measures are the best candidates for assessing CDs on images affected by change of color saturation: CD00 (CIEDE2000), CD01 (Spatial-CIELAB), CD15 (OSA-UCS), CD14 (Circular processing CD), CD17 (Spatial colour metric) and CD18 (proposed method) showing a strong correlation with subjective scores. Overall, the proposed methodology CD18 (wCD) is clearly the best performing CD measure tested in this work.

Additionally, we found that relying on descriptive statistics from pixel-wise differences is unreliable for computing color differences typically reported by human observers. The results suggest that there are no significant differences in terms of correlation with subjective scores between CD00 (CIEDE2000), CD01 (Spatial-CIELAB) and CD17 (Spatial colour metric). That is, there is a negligible improvement in terms of correlation with subjective scores when a spatial filtering simulating blur property of human eyes is applied before computation of pixel wise differences. Additionally, considering the fact that humans more easily perceive CD in flat areas than in complex structures, it is more desirable to measure CDs in homogeneous patches (based on image segmentation) and then combine them into an overall CD as the proposed methodology. This is confirmed as well by the performance achieved by the proposed methodology which is based on computation of local differences in homogeneous textured patches.

Future work should further extend the scope of evaluation by including additional publicly available image databases as well as other color related types of distortion (e.g. gamut mapping) with the purpose of validating the results and generalizing the findings of our work. Also, since there is a considerable increase of computer-generated image content [83], the evaluation of the proposed methodology in computer-generated images is proposed as future work.

Acknowledgment

This work was performed within the PANORAMA project ID Number 296104, funded under FP7-JTI (co-funded by grants from Belgium, Italy, France, the Netherlands,the United Kingdom, and the ENIAC Joint Undertaking).

References

[1] S. Fezza, M. Larabi, K. Faraoun, Feature-based color correction of multi-view video for coding and rendering enhancement, IEEE Transactions on Circuits and Systems for Video Technology 24 (2014) 1486 – 1498.

[2] D. Ly, S. Beucher, M. Bilodeau, S. Persa, K. Damstra, R. Pot, J. Rooy, Automatic color correction: region-based approach and performance evaluation using full reference metrics, Journal of Electronic Imaging 24 (2015) 061207 1 – 9.

[3] L. Brun, A. Tremeau, Digital Color Imaging Handbook, CRC press, 2002, Ch. Color quantization.

[4] J. Morovič, Color Gamut Mapping, John Wiley & Sons, Inc., 2008, Ch. Desired Color Reproduction Properties and their Evaluation.

[5] A. Mojsilovic, H. Jianying, E. Soljanin, Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis, IEEE Transactions on Image Processing 11 (2002) 1238 – 1248.

[6] D. Hasler, S. Süsstrunk, Measuring colourfulness in natural images, in: Proc. of the IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging VIII, 2003, pp. 87 – 95.

[7] J. Hardeberg, E. Bando, M. Pedersen, Evaluating colour image difference metrics for gamut-mapped images, Coloration Technology 124 (2008) 243 – 253.

[8] U. Rajashekar, Z. Wang, E. Simoncelli, Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions, in: Proc. of the IEEE International Conference on Image Processing, 2009, pp. 2213 – 2216.

[9] Y. Yang, J. Ming, N. Yu, Color image quality assessment based on CIEDE2000, Advances in Multimedia 2012 (2012) 1 – 6.

[10] D. Lee, E. Rogers, Towards anovel perceptual color difference metric using circular processing of hue components, in: Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2014, pp. 166 – 170.

[11] M. Pedersen, Evaluation of 60 full-reference image quality metrics on the cid: Iq, in: Proc. of the International Conference on Image Processing, 2015, pp. 1588 – 1592.

[12] W. Lin, C. Kuo, Perceptual visual quality metrics: A survey, Journal of Visual Communication and Image Representation 22 (2011) 297 – 312.

[13] M. Pedersen, J. Hardeberg, Full-reference image quality metrics: Classification and evaluation, Foundations and Trends in Computer Graphics and Vision 7 (2012) 1 – 80.

[14] T. Liu, Y. Lin, W. Lin, C. Kuo, Visual quality assessment: recent developments, coding applications and future trends, APSIPA Transactions on Signal and Information Processing 2 (2013) 1 – 20.

[15] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, W. Zhang, Analysis of distortion distribution for pooling in image quality prediction, IEEE Transactions on Broadcasting 62 (2016) 446 – 456.

[16] K. Gu, L. Li, H. Lu, X. Min, W. Lin, A fast reliable image quality predictor by fusing micro- and macro-structures, IEEE Transactions on Industrial Electronics 64 (2017) 3903 – 3912.

[17] E. Bando, J. Hardeberg, D. Connah, Can gamut mapping quality be predicted by colour image difference formulae?, in: Proc. of the IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging X, 2005, pp. 180 – 191.

[18] H. Kivinen, M. Nuutinen, P. Oittinen, Comparison of colour difference methods for natural images, in: Proc. of the Conference on Colour in Graphics, Imaging, and Vision, 2010, pp. 510 – 515.

[19] H. Xu, Q. Tong, R. Gong, Color difference evaluation for digital images, in: Proc. of the Asia Color Association Conference, 2013, pp. 28 – 31.

[20] M. Pedersen, An image difference metric based on simulation of image detail visibility and total variation, in: Proc. of the Color and Imaging Conference, 2014, pp. 37 – 42.

[21] Z. Barańczuk, P. Zolliker, J. Giesen, Image quality measures for evaluating gamut mapping, in: Proc. of the Color and Imaging Conference, 2009, pp. 21 – 26.

[22] B. Ortiz-Jaramillo, L. Platisa, W. Philips, ifas : image fidelity assessment, in: Proceedings of the International Workshop Computational Color Imaging, 2017, pp. 83 – 94.

[23] T. Mäenpää, The local binary pattern approach to texture analysis: Extensions and applications, Ph.D. thesis, University of Oulu (2003).

[24] M. Pinson, S. Wolf, A new standardized method for objectively measuring video quality, IEEE transactions on broadcasting 50 (2004) 312 – 322.

[25] W. Zhou, A. Bovik, H. Sheikh, E. Simoncelli, The ssim index for image quality assessment, `https://ece.uwaterloo.ca/~z70wang/research/ssim/`, last accessed May 2018 (2014).

[26] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C. Jay-Kuo, Image database

TID2013: Peculiarities, results and perspectives, Signal Processing: Image Communication 30 (2015) 57 – 77.

[27] G. Johnson, M. Fairchild, Digital Color Imaging Handbook, CRC press, 2002, Ch. Visual psychophysics and color appearance.

[28] N. Moroney, M. Fairchild, R. Hunt, C. Li, M. Luo, T. Newman, The CIECAM02 color appearance model, in: Proc. of the IS&T/SID Color Imaging Conference, 2002, pp. 23 – 27.

[29] M. Habekost, Which color differencing equation should be used?, International Circular of Graphic Education and Research 6 (2013) 20 – 33.

[30] G. Sharma, Digital Color Imaging Handbook, CRC press, 2002, Ch. Color fundamentals for digital imaging.

[31] Recommendation ITU-R BT.601-5 studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios, `https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-5-199510-S!!PDF-E.pdf`, last accessed May 2018 (1995).

[32] A. Smith, Color gamut transform pairs, Computer Graphics 12 (1978) 12 – 19.

[33] D. Ruderman, T. Cronin, C. Chiao, Statistics of cone responses to natural images: implications for visual coding, Journal of the Optical Society of America A 15 (1998) 2036 – 2045.

[34] CIE, Colorimetry - part 4: CIE 1976 L*a*b* colour space, `https://www.iso.org/obp/ui/#iso:std:iso:11664:-4:ed-1:v1:en`, last accessed May 2018 (1976).

[35] R. Huertas, M. Melgosa, C. Oleari, Performance of a color-difference formula based on OSA-UCS space using small-medium color differences, Journal of the Optical Society of America A 23 (2006) 2077 – 2084.

[36] L. Mandic, S. Grgic, M. Grgic, Comparison of color difference equations, in: Proc. of the International Symposium focused on Multimedia Signal Processing and Communications, 2006, pp. 107 – 110.

[37] M. Luo, G. Cui, B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Research and Application 26 (2001) 340 – 350.

[38] X. Zhang, B. Wandell, A spatial extension of CIELAB for digital color-image reproduction, Journal of the Society for Information Display 5 (1997) 61 – 63.

[39] F. Imai, N. Tsumura, Y. Miyake, Perceptual color difference metric for complex images based on mahalanobis distance, Journal of Electronic Imaging 10 (2001) 385 – 393.

[40] C. Gao, K. Panetta, S. Agaian, No reference color image quality measures, in: Proc. of the IEEE International Conference on Cybernetics, 2013, pp. 243 – 248.

[41] A. Toet, M. Lucassen, A new universal colour image fidelity metric, Displays 24 (2003) 197 – 207.

[42] S. Lee, J. Xin, S. Westland, Evaluation of image similarity by histogram intersection, Color Research and Application 30 (2005) 265 – 274.

[43] G. Hong, M. Luo, A new algorithm for calculating perceived colour difference of images, Imaging Science Journal 54 (2006) 1 – 15.

[44] G. Johnson, Using color appearance in image quality metrics, in: Proc. of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2006, pp. 1 – 4.

[45] C. Chou, K. Liu, A fidelity metric for assessing visual quality of color images, in: Proc. of the International Conference on Computer Communications and Networks, 2007, pp. 1154 – 1159.

[46] Y. Ming, L. Huijuan, G. Yingchun, Z. Dongming, A method for reduced-reference color image quality assessment, in: Proc. of the International Congress on Image and Signal Processing, 2009, pp. 1 – 5.

[47] M. Pedersen, J. Hardeberg, A new spatial hue angle metric for perceptual image difference, in: Proc. of the Computational Color Imaging, 2009, pp. 81 – 90.

[48] M. Pedersen, J. Hardeberg, A new spatial filtering based image difference metric based on hue angle weighting, Journal of Imaging Science and Technology 56 (2012) 50501 1 – 12.

[49] I. Lissner, J. Preiss, P. Urban, M.-S. Lichtenauer, P. Zolliker, Image-difference prediction: From grayscale to color, IEEE Transactions on Image Processing 22 (2013) 435 – 446.

[50] C. Oleari, Color opponencies in the system of the uniform color scales of the optical society of america, Journal of the Optical Society of America A 21 (2004) 677 – 682.

[51] G. Simone, C. Oleari, I. Farup, An alternative color difference formula for computing image difference, in: Proc. of the Gjøvik Color Imaging Symposium, 2009, pp. 8 – 11.

[52] S. Ouni, E. Zagrouba, M. Chambah, M. Herbin, A new spatial colour metric for perceptual comparison, in: Proc. of the International Conference on Computing and e-Systems, 2008, pp. 413 – 428.

[53] N. Bonnier, F. Schmitt, H. Brettel, S. Berche, Evaluation of spatial gamut mapping algorithms, in: Proc. of the Color and Imaging Conference, 2006, pp. 56 – 61.

[54] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, M. Carli, Modified image visual quality metrics for contrast change and mean shift accounting, in: Proc. of the International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, 2011, pp. 305 – 311.

[55] L. Zhang, Y. Shen, H. Li, Vsi: A visual saliency-induced index for perceptual image quality assessment, IEEE Transactions on Image Processing 23 (2014) 4270 – 4281.

[56] J. Geusebroek, R. V. den Boomgaard, A. Smeulders, H. Geerts, Color invariance, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 1338 – 1350.

[57] C. Yang, S. Kwok, Efficient gamut clipping for color image processing using lhs and yiq, Optical Engineering 42 (2003) 701 – 711.

[58] H. Liu, M. Huang, M. Xie, M. Zhu, The effect of image's contents on the visual color difference for digital images, in: Proc. of the International Congress on Image and Signal Processing, 2010, pp. 2356 – 2359.

[59] R. Palma-Amestoy, E. Provenzi, M. Bertalmio, V. Caselles, A perceptually inspired variational framework for color enhancement, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 458 – 474.

[60] M. Bertalmio, V. Caselles, E. Provenzi, Issues about retinex theory and contrast enhancement, International Journal of Computer Vision 83 (2009) 101 – 119.

[61] H. Liu, M. Huang, G. Cui, M. Luo, M. Melgosa, Color-difference evaluation for digital images using a categorical judgment method, Journal Optical Society of America A 30 (2013) 616 – 626.

[62] Z. Wang, X. Shang, Spatial pooling strategies for perceptual image quality assessment, in: Proc. of the IEEE International Conference on Image Processing, 2006, pp. 2945 – 2948.

[63] A. Ninassi, O. L. Meur, P. L. Callet, D. Barba, Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric, in: Proc. of the IEEE International Conference on Image Processing, 2007, pp. 169 – 172.

[64] M. Pedersen, J. Hardeberg, P. Nussbaum, Using gaze information to improve image difference metrics, in: Proc. of the Human Vision and Electronic Imaging XIII, 2008, pp. 680611 1 – 12.

[65] H. Liu, I. Heynderickx, Studying the added value of visual attention in objective image quality metrics based on eye movement data, in: Proc. of the IEEE International Conference on Image Processing, 2009, pp. 3097 – 3100.

[66] A. Moorthy, A. Bovik, Visual importance pooling for image quality assessment, IEEE Journal of Selected Topics in Signal Processing 3 (2009) 193 – 201.

[67] Y. Tong, H. Konik, F. Cheikh, A. Tremeau, Full reference image quality assessment based on saliency map analysis, Journal of Imaging Science and Technology 54 (2010) 30503 1 – 14.

[68] K. Raja, M. Pedersen, Artifact detection in gamut mapped images using saliency, in: Proc. of the Colour and Visual Computing Symposium, 2013, pp. 1 – 6.

[69] M. Topi, O. Timo, P. Matti, S. Maricor, Robust texture classification by subsets of local binary patterns, in: Proceedings of the International Conference on Pattern Recognition, 2000, pp. 935 – 938.

[70] J. Fehr, H. Burkhardt, 3d rotation invariant local binary patterns, in: Proceedings of the International Conference on Pattern Recognition, 2008, pp. 1 – 4.

[71] Recommendation ITU-R BT.2022 general viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays, `https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2022-0-201208-I!!PDF-E.pdf`, last accessed May 2018 (2012).

[72] P. Chen, P. Popovich, Correlation: Parametric and nonparametric measures, Sage Publications, 2002, Ch. The Pearson Product-Moment Correlation.

[73] P. Chen, P. Popovich, Correlation: Parametric and nonparametric measures, Sage Publications, 2002, Ch. Special Cases of Pearson's R.

[74] G. Székely, M. Rizzo, N. Bakirov, Measuring and testing dependence by correlation of distances, The Annals of Statistics 35 (2007) 2769 – 2794.

[75] M. Mukaka, Statistics corner: A guide to appropriate use of correlation coefficient in medical research, Malawi Medical Journal 24 (2012) 69 – 71.

[76] Z. Barańczuk, P. Zolliker, J. Giesen, Image-individualized gamut mapping algorithms, Journal of Imaging Science and Technology 54 (2010) 030201 1 – 7.

[77] X. Wei, J. Mulligan, Performance evaluation of color correction approaches for automatic multi-view image and video stitching, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 263 – 270.

[78] B. Ortiz-Jaramillo, A. Kumcu, L. Platisa, W. Philips, Computing contrast ratio in medical images using local content information, in: Proc. of the Medical Image Perception Conference XVI, 2015, pp. 33 – 33.

[79] B. Ortiz-Jaramillo, A. Kumcu, L. Platisa, W. Philips, Computing contrast ratio in images using local content information, in: Proc. of the Symposium on Signal Processing, Images and Computer Vision, 2015, pp. 1 – 6.

[80] B. Ortiz-Jaramillo, A. Kumcu, L. Platisa, W. Philips, Content-aware contrast ratio measure for images, Journal of Signal Processing: Image Communication 62 (2018) 51 – 63.

[81] F. Kingdom, N. Prins, Psychophysics: A practical introduction, Pearson Prentice Hall, 2010, Ch. Psychometric Functions, pp. 59 – 120.

[82] Y. Deng, C. Kenney, M. Moore, B. Manjunath, Peer group filtering and perceptual color image quantization, in: Proc. of the IEEE International Symposium on Circuits and Systems, 1999, pp. 21 – 24.

[83] K. Gu, J. Qiao, X. Min, G. Yue, W. Lin, D. Thalmann, Evaluating quality of screen content images via structural variation analysis, IEEE Transactions on Visualization and Computer Graphics Early Access (2018) 1 – 13.