Evaluating color difference measures in images

Conference Paper · June 2016

DOI: 10.1109/QoMEX.2016.7498922

Benhur Ortiz Jaramillo
Cassini Technologies BV

Asli Kumcu
Ghent University

Evaluating color difference measures in images

B. Ortiz-Jaramillo, A. Kumcu and W. Philips
TELIN-IPI-iMinds, Ghent University
Email: {bortiz, akumcu, philips}@telin.ugent.be

***Abstract*—The most well known and widely used method for comparing two homogeneous color samples is the CIEDE2000 color difference formula because of its strong agreement with human perception. However, the formula is unreliable when applied over images and its spatial extensions have shown little improvement compared with the original formula. Hence, researchers have proposed many methods intending to measure color differences (CDs) in natural scene color images. However, these existing methods have not yet been rigorously compared.**
Therefore, in this work we review and evaluate CD measures with the purpose of answering the question *to what extent do state-of-the-art CD measures agree with human perception of CDs in images?* To answer the question, we have reviewed and evaluated eight state-of-the-art CD measures on a public image quality database. We found that the CIEDE2000, its spatial extension and the just noticeable CD measure perform well in computing CDs in images distorted by black level shift and color quantization algorithms (correlation higher than 0.8). However, none of the tested CD measures perform well on identifying CDs for the variety of color related distortions tested in this work, e.g., most of the tested CD measures showed a correlation lower than 0.65.

***Keywords*—color appearance, color difference assessment, image difference assessment.**

I. INTRODUCTION

Traditionally, color differences (CDs) between images have been assessed by first computing pixel wise differences with the CIEDE2000 formula [1] and then assigning statistics such as mean, median, and maximum of the differences as overall CD [2], [3]. While the CIEDE2000 agrees very well with humans when measuring differences of homogeneous color samples [4], the obtained differences may be unreliable when aggregated over images [2], [3], [5]. Hence, the assessment of CD in images is still an active area in the research of color science and imaging technology because of its wide range of applications such as color correction [6], color quantization [7], color image similarity and retrieval [8], gamut mapping [9], among others. However, currently there is not a standard procedure for performing such a task [2], [3], [10], [11].

The CD measures in the state-of-the-art are often tested on databases containing multiple distortions in addition to the color related distortions, with a few testing samples or that are not publicly available. This makes the experimental results of little significance for benchmarking CD measures. For instance, Ming et.al. [12] and He et.al. [13] used the LIVE database phase 1 and 2 [14], respectively, for evaluating the performance of their proposed color quality measures.[1] These databases are composed of a set of images exhibiting different types of distortions such as additive Gaussian noise, JPEG compression, and blurring, among others. The performance of the measures are often reported as the average performance over all the distortion types. Thus, the performance specific to the color related distortions has not yet been rigorously explored.

To the best of the authors' knowledge, only a few works in the literature explicitly study the color-related aspect of image difference assessment. Also, they test only on private databases or on a few CD measures and cannot therefore be used for benchmarking new CD measures. For instance, in [15] three CD measures (including the CIEDE2000) were tested on a non publicly available database (6 source images) with the purpose of identifying the best performing method for evaluating gamut mapping algorithms. Hardeberg et. al. [2] have evaluated a general purpose image quality measure (SSIM [16]), the CIEDE2000, two spatial extensions of the CIEDE2000 and a measure based in hue angle with the purpose of testing gamut mapping algorithms in 6 source images. Kivinen et. al. [17] evaluated and compared three CD formulas (including the CIEDE2000) and three spatial extensions of the CIEDE2000 in a non publicly available database (8 source images) for identifying the measures with high agreement with perceived CDs. In [18] three CD formulas (including the CIEDE2000) and two spatial extensions of the CIEDE2000 were tested in a non publicly available database (6 source images) for identifying perceptibility thresholds in color changes.

In summary, there are a few works addressing the problem of reviewing and testing CD measures in images, and the existing ones are very limited regarding the number of test samples and CD measures. In contrast with [2], [15], this work takes into account four types of color related distortions from a public database. Additionally, the 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. Furthermore, unlike [2], [15], [17], [18], the CD measures included in this work are dedicated to measure color distortions making them more suitable to be used on color related applications [6], [7], [9].

The aim of this work is to review and evaluate the state-of-the-art in CD measures using a publicly available database. First, we conduct a short review of the color science for evaluating CDs, where eight state-of-the-art CD measures are studied. Then, we evaluate the performance of the CD measures and elaborate on the cases where the CD measures fail. The measures are tested on a publicly available database (25 source images and 4 different color related distortions). Additionally, we introduce a methodology to analyse the performance of the CD measures in function of the images' color content with the purpose of identifying the types of content on which the measures fail. Our results show that the CIEDE2000 [19], its spatial extension [5] and the just noticeable CD measure [20]

[1] It is called color quality measure because it uses color information in the assessment but it is not a dedicated CD measure. "Dedicated" means that the CD measures have been designed exclusively to measure CDs and not other distortion types.

are the best performing measures in the tested set. Overall, these measures perform well in computing CDs in images affected by quantization noise or intensity shift distortions (correlation with human scores higher than 0.8). However, the results of our experiments also reveal a poor correlation between subjective scores and the 8 measures tested in this work on images affected by change in color saturation and contrast change (correlation with human scores lower than 0.65). The color content analysis suggests that the higher the color content of the source image (number of colors in the image) the lower the performance of the CD measure.

This work is organized as follows. In Section II, current approaches dealing with CD assessment of images are discussed. The experimental setup and the proposed validation methodology are described in Section III. Thereafter, in Section IV, we present and discuss the results obtained in this study. Finally, in Section V, we draw conclusions and propose future work.

II. Background

The study of color appearance models is based on the existence of a space with three principal components that, when mixed together, can produce all other colors. For example, the RGB color space is the most well known and widely used additive color appearance model, i.e., any color can be represented as a linear combination of its components [21]. Although it is a very intuitive and simple color representation, it is not perceptually uniform[2] [21]. Hence, many color appearance models which achieve good perceptual uniformity in homogeneous color samples have been proposed in the literature. For brevity, we do not define or describe these color appearance models[3] and instead concentrate on eight measures specifically designed for computing CDs in images, summarized in Table I.

The CIEDE2000 formula [19] was designed from experimental color data using the outcome of psychovisual studies with trained and untrained observers in regards to judging CDs. The obtained formula, termed ΔE_{00}, includes correction terms that describe the effects of lightness, chroma, and hue in visual perception [3]. The spatial extension of the CIEDE2000, termed $\Delta E_{00}^{\mathrm{S}}$, uses an opponent color space and an approximation of the contrast sensitivity function to mimic contrast masking in human vision [5]. The image CD measure based on image appearance models, termed ΔE^{I}, includes attributes to account for changes in overall luminance, background luminance, surround viewing conditions and contrast masking [24]. The just noticeable CD measure, termed ΔE^{J}, considers the visibility threshold of each pixel as a function of chroma, local luminance gradient and background uniformity [20]. Note that ΔE_{00}, $\Delta E_{00}^{\mathrm{S}}$, ΔE^{I} and ΔE^{J} use the CIELAB color space during the CD assessment.

"Colorfulness" is a color attribute first proposed by Hasler and Susstrunk [10] with the purpose of measuring the color intensity (chromatic level) of natural scene color images. The CD, termed ΔC^{H}, is usually estimated using a non-linear combination of global statistics of color components in an opponent color space [10]. The color extension of the Structural Similarity Index Measure (termed cssim), proposed by Toet and Lucassen [25], is computed as a non-linear combination of the SSIM indices computed on three uncorrelated color components (one luminance and two opponent color components). The chroma spread and extreme indexes quantify the changes in the spread of the distribution of two-dimensional color samples and severe localized color impairments, respectively. The CD measure, termed Ch, is computed as a weighted summation of these two statistics estimated over small image patches [26]. The adaptive spatio-chromatic image difference, termed ΔE^{A}, is a CD measure based on an adaptive signal decomposition method in which the adaptive functions are chosen in RGB color space to capture differences in luminance, hue, and saturation [11].

Table I summarizes[4] the studied dedicated CD measures. In summary, the reviewed CD measures consider five color appearance models (CIELAB, opponent color space, $\ell\alpha\beta$, $\mathrm{YC_BC_R}$ and RGB).[5] Particularly, four of the eight measures are computed in the CIELAB color space. Six out of the eight tested measures do have a spatial processing. That is, neighbour pixels are taken into account in each step of the CD computation. However, the overall CD for six out of the eight tested methods (ΔE_{00}, $\Delta E_{00}^{\mathrm{S}}$, cssim, ΔE^{I}, ΔE^{J} and ΔE^{A}) is computed as the average of individual pixel differences. Further information can be found in the related publications listed in Table I.

TABLE I. CD MEASURES EVALUATED IN THIS WORK

CD measure	Symbol	Appearance model	Type	Spatial processing
CIEDE2000 [19]	ΔE_{00}	CIELAB [22]	Full reference	No
Spatial extension ΔE_{00} [5]	$\Delta E_{00}^{\mathrm{S}}$	CIELAB [22]	Full reference	Yes
Colorfulness [10]	ΔC^{H}	Opponent color space [10]	Reduced reference	No
Color extension of the SSIM index [25]	cssim	$\ell\alpha\beta$ [27]	Full reference	Yes
Chroma spread and extreme [26]	Ch	$\mathrm{YC_BC_R}$ [21]	Reduced reference	Yes
Image CD measure based on image appearance models [24]	ΔE^{I}	CIELAB [22]	Full reference	Yes
Just noticeable CD measure [20]	ΔE^{J}	CIELAB [22]	Full reference	Yes
Adaptive spatio-chromatic image difference [11]	ΔE^{A}	RGB [21]	Full reference	Yes

Symbol is the notation for referring to a specific CD measure. Type refers to the availability of the source sample (full or partial/reduced information). Spatial processing is whether or not neighbour pixels are taken into account in the CD measure.

III. Experimental setup

A. Test data

The test data was selected such that the most common applications of the color-related aspect of image difference assessment are included. Particularly, the following applications were considered: color correction, color quantization, color matching, gamut mapping and multiview imaging systems. The

[2]Perceptual uniformity means that the Euclidean distance between two colors in the RGB space does match the actual perceived CD [21].

[3]Further information about color appearance models for computing CDs in homogeneous color samples can be found in [21]–[23]

[4]The algorithms will be available in the web page of the first author after the publication of this work [http://telin.ugent.be/~bortiz/color]

[5]cf. Sharma Chapter 1 [21] for detailed information about these color appearance models.

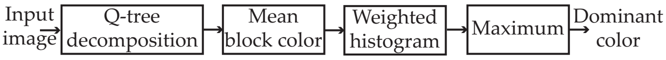

Fig. 1. Dominant color extraction.

test dataset was obtained from the publicly available Tampere Image Database (TID2013 [http://www.ponomarenko.info/tid2013.htm]). TID2013 is a database intended for evaluating image quality measures. TID2013 has subjective scores, in terms of Mean Opinion Score (MOS), for comparing the performance between quality measures [28]. TID2013 contains 25 reference images and 3000 distorted images (25 reference images $\times$ 24 types of distortions $\times$ 5 levels of distortion).

The following 4 color related distortion types were selected from the 24 types available in TID2013: quantization noise, mean intensity shift, contrast change, and change of color saturation. We selected this subset of distortions because they encompass the most important applications of the color-related aspect of image difference assessment. For instance, quantization noise is closely related to color quantization applications. Thus, a CD measure performing well on this type of distortion could be further used with the purpose of improving color quantization algorithms to achieve quantization steps with the minimum overall perceived CD. Intensity shift, contrast change and change in color saturation are distortions produced by color matching, color correction and gamut mapping algorithms as well as by multiview imaging systems [6], [29], [30]. The remaining 20 distortions were not used because they possess spatial distortions which impact the quality of the image much more strongly than CDs. 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 quality of the image much more strongly than the CDs.

B. Evaluation methodology

First we explore the performance for the whole dataset with the purpose of evaluating the tested CD measures across different color content and distortion types. We compute the Pearson Correlation Coefficient (PCC) and the Spearman's Rank Order Correlation Coefficient (SROCC) between the subjective scores and the values computed by the tested CD measures. Thereafter, we split the dataset in subsets according to the distortion type for evaluating each use case separately. We compute the PCC and the SROCC between the subjective scores and the CD measure values for each distortion type.

Afterwards, we split the dataset according to the source content, i.e., we compute the PCC and the SROCC between the CD measure and the MOS over all distortions and distortion levels for each source image. We do this with the purpose of identifying those CD measures performing very well accross color content. Also, we use the CIEDE2000 as a benchmark to find significant differences with respect to the more advanced methods tested in this work. We use CIEDE2000 as benchmark because it is the most well-know CD formula to date and it has shown better performance than other reported formulas at least for computing CDs in homogeneous color samples [31]. Specifically, we use pairwise comparisons as discussed by Garcia et. al. [32]. Here, the related samples are the performances (PCCs and SROCCs) of the CD measures. The

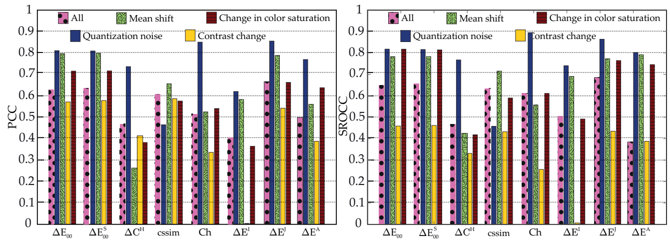

Fig. 2. Performance of the considered CD measures evaluated on the TID2013 database for all data and per individual color related distortion. Performance is given in terms of PCC.

objective of the pairwise comparison is to determine if we may conclude from the data (PCCs and SROCCs between CD measures and subjective scores) that there are statistically significant differences in terms of the performance between the benchmark and the other tested CD measures.

Finally, we propose a novel methodology to analyse the performance of the CD measures in function of the dominant color of the source image. Here, the dominant color was computed using the method in Figure 1. In the Q-tree decomposition [33], every block is further split up if any pixel color deviates from the average by more than $3\Delta E_{00}$ units which corresponds to a medium perceived CD [34]. We divide the image because the analysis of individual pixels does not provide good results on extracting the dominant color [35]. The mean block color computes the average of each node from the resulting Q-tree decomposition. The weighted color histogram is build on the a^*b^* coordinates of the CIELAB space. The weights of the histogram are computed as $\exp\left(-\frac{1}{N}\right)$ where N is the number of pixels in the resulting Q-tree node. These weights are used to give more importance to those blocks with bigger area coverage [36]. Finally, the dominant color is selected as the argument maximum of the obtained weighted color histogram.

IV. Results and Discussion

We start this Section by presenting and discussing the results of the experiments described in Section III. After that, in Section IV-C we analyse the performance of the best performing CD measure in function of the color content of the source image. Since the SROCC values of Figures 2 and 3 show a similar trend than the PCC values, the following analysis applies for both the PCC and the SROCC values.

A. Evaluation of the tested CD measures

Figure 2 shows the PCC and SROCC computed between the 8 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. For the whole dataset, the worst performance is achieved by ΔE^{I} followed by ΔC^{H}, ΔE^{A} and Ch (PCC equal to 0.4, 0.46, 0.5 and 0.52, respectively). For instance, it has been already highlighted in the state-of-the-art that RGB (ΔE^{A}) and $\mathrm{YC_BC_R}$ (Ch) are limited in terms of correlation with human perception [21]. That is, the results indicate that the poor performance of these CD measures can be due to the color space they use. ΔC^{H} uses global descriptive statistics for comparing colors between images. This is insufficient to describe CDs because it can be possible to have perceived

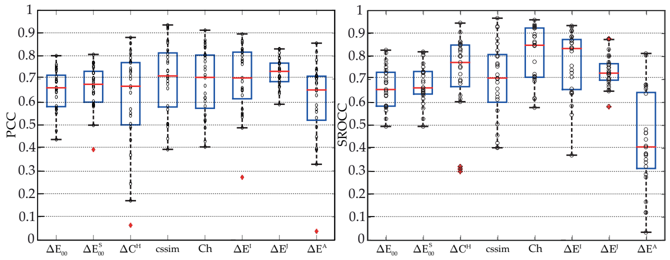

Fig. 3. Box plot of the performance of the considered CD measures computed on the TID2013 databases per individual source image.

CDs without affecting the global descriptive statistics. Finally, the poor performance of ΔE^{I} may be due to the fact that the measure includes complex spatial interactions such as perception of contrast, graininess, and sharpness. However, it is well known that humans perceive better CD in flat areas than in complex structures [36]. In any case, the highest PCC in the selected test conditions is equal to 0.66. This value is achieved by ΔE^{J} which in overall it is a poor correlation with subjective scores (only 45% of the subjective scores variability is accounted by the CD measure).

These results suggest that none of the state-of-the-art CD measures tested in this work is able to handle multiple color related distortion types. Therefore, we split the TID2013 database in subsets according to the type of distortion to further explore the failure scenarios of the tested measures. Figure 2 shows the PCC and the SROCC appraised on the TID2013 database for each individual distortion type. In this experiment, the worst performance is achieved once more by ΔC^{H} and ΔE^{I} (PCC lower than 0.6 excluding the quantization noise subset where the PCC is higher than 0.6). Note that the tested CD measures achieve the highest performance in the quantization noise and the mean intensity shift subsets (PCC higher than 0.78 between subjective scores and ΔE_{00}, $\Delta E_{00}^{\mathrm{S}}$ as well as ΔE^{J}), i.e., more than the 61% of the subjective scores variability is accounted by the CD measures. This suggests that the evaluation of quantization noise and intensity shift distortions are simpler than the evaluation of the other color related distortions such as, contrast change and/or change of color saturation.

For the images affected by contrast changes, all of the methods yield a PCC lower than 0.6. As such, these methods fail on this material and it is necessary to find an alternative mechanism for measuring such differences. For instance, in [37], the authors have proposed a method which accounts for contrast changes, and they found PCCs higher than 0.9 between this method and subjective scores on the same subset. That measure assesses contrast changes by comparing the contrast ratio according to Webers law computed over local image patches on the luminance channel. This suggests that, even though contrast changes also produce CDs, the differences affect mainly the amount of perceived details, rather than perceived color changes. Therefore, we do not further explore this subset of images because they cannot be assessed by using the CD measures tested in this work.

B. *Evaluation of the tested CD measures per source*

Figure 3 shows the performance of the considered CD measures computed on the TID2013 database per individual source image. The box plot represents the PCC between the CD measure and the MOS over all distortions and distortion levels for each source image. Thus, it shows the variability of the agreement between the CD measure and the subjective scores for different image content. Although the PCC per individual source image overestimates the global agreement with subjective scores, it is a good method for identifying those measures with very low performance and/or high variability of the performance across image content. For instance, the box plot of ΔC^{H} is characterized by a large box and whiskers, i.e., the agreement between the CD measure and the subjective scores is highly dependent on image content of the source image. Similarly, cssim, Ch, ΔE^{I} and ΔE^{A} are characterized by large boxes and whiskers making these measures inadequate for computing CDs because they are highly dependent on the content of the source image.

ΔE^{J} achieves better performance than the previously mention CD measures in the tested data. This may be due to the operational color space (cf. Table I) and the introduction of weights to give higher importance to CDs located on flat areas. Note that, the methods using the CIELAB color space are the best performing methods. That is, although this color space was designed specifically for evaluating CDs in homogeneous color samples, the color space is still useful in computing CDs in images because they are based on psychovisual findings. For instance, the CIELAB color space tries to match perceptual uniformity, i.e., the Euclidean distance between two colors in CIELAB space is very close to the actual perceived CD. Additionally, its luminance component closely matches human perception of lightness [21].

The data of Figure 3 also shows that there are not significant differences in terms of PCC with subjective scores when spatial processing based on filtering is applied prior to computing pixel-wise the CIEDE2000. Specifically, no statistically significant differences exist between ΔE_{00} and $\Delta E_{00}^{\mathrm{S}}$ (p-value = 0.205). This is mainly because the spatial processing is based on filtering (band pass filtering simulating contrast masking as proposed by Zhang and Wandell [5]), relying on the computation of pixel-wise differences between the images, and the use of the average as the overall CD. Additionally, the only CD measure that presents significant differences with the CIEDE2000 is ΔE^{J} (p-value lower than 0.05). That is, ΔE^{J} performs statistically better than ΔE_{00} and $\Delta E_{00}^{\mathrm{S}}$ in the tested data. We further analyse the performance of this CD measure in function of the color content of the source image. We use ΔE^{J} as an example of how the proposed analysis can be used to identify possible content related failure scenarios and therefore how the measure can be improved by taking into account the results of such an analysis.

C. *Performance of the CD measures in function of the color content*

Figure 4 shows the performance of ΔE^{J} in function of the dominant color of the source image. Note that the size of the marker is scaled by the PCC value. The markers in the plot are a^*b^* coordinates of the dominant color of each source image. The a^*b^* coordinates can be associated to the color displayed on the color chart of Figure 4(right side). We examine the images where a poor correlation is achieved (PCC<0.7): 3, 4, 12, 23, 24 and 25 (see [28] for illustration of the source images).

Although image i03 has similar dominant color component as

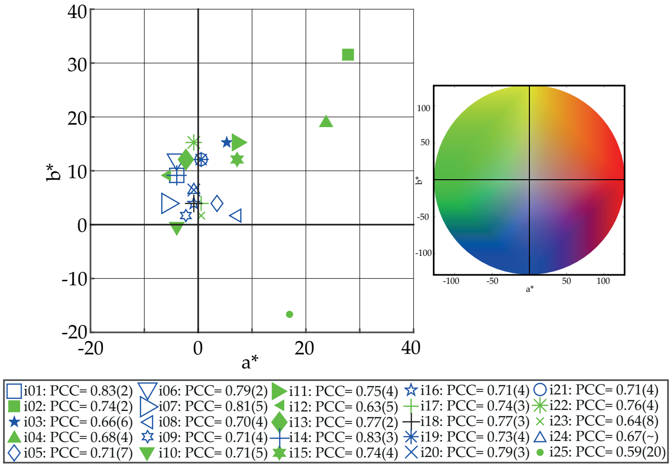

Fig. 4. Performance of ΔE^{J} in function of the dominant color of the source image (color is represented using the ab components of the CIELAB color space, right side color chart). Each marker represents the dominant color of one of the 25 source images in the TID2013 database. The PCC values are shown in the legends together with the number of dominant colors within brackets. Also note that the size of the marker is scaled by the PCC value.

images i11 and i15, the performance achieved in such an image is significatively lower. The three images possess a brown like dominant component. However, unlike images i11 and i15 (67% and 62% of the image is covered by the dominant color, respectively), image i03 possesses other colors where humans are more susceptible to CDs covering a big image area (yellow 17%, green 8%, orange 12% and blue 10% of the image). Similarly, image i12 has a dominant color (gray like) similar to the one of images i01 and i14 (87% and 67% of image coverage, respectively). However, image i12 has multiple colors covering a significant image area (grey 36%, blue 20%, yellow 15%, red 4%, green 8% and pink 3% of the image). It is possible to conclude the same for image i23 (red 17%, yellow 25%, blue 4%, green 24%, gray, black and white 30% of the image). This suggests that the performance of the CD measure is strongly related to the number of colors presented in the source image and their image coverage (see number of representative colors next to the PCC in the legends of Figure 4). That is, the higher the number of colors covering a significant area in the image the lower the performance.

For the case of images i02 and i04, the dominant color is in the red range. Although in this range the CIELAB color space agrees very well with human perception [31], image i04 is a face of a women and it is well known that humans are very susceptible to CDs in skin tones [2]. Also, the same magnitude of CD in skin colors can produce reddish, yellowish or greenish effects making the measure inaccurate in such a region. Even though image i24 has the same dominant color as image i20, image i24 is characterized by highly textured areas, making it difficult to even identify multiple color components. Additionally, the color histogram of the image i24 is located almost at the origin of a^*b^* plane, i.e., the image is mostly represented by shades of gray. ΔE^{J} achieves the lowest performance in image i25. However, we believe that the results in such an image are not interesting because it is a test pattern with very high frequency content and around 20 colors covering a significant image area which is not found in natural scenes making it irrelevant from the application point of view.

Previous analysis suggests that when a small number of colors covering a significant image area are presented in the source image, ΔE^{J} performs relatively well (at least 64% of the subjective scores variance is accounted by the CD measure). However, the number of colors is not the only factor. For instance, noteworthy is that since ΔE^{J} differences are based on the Euclidean distance between colors in the CIELAB color space, the performance on skin-like colors is very poor because such a distance does not include correction factors like those used on ΔE_{00} formula. For example, ΔE_{00} was designed to include not only lightness, chroma, and hue weighting functions, but also an interactive term between chroma and hue differences for improving the performance for blue colors and a scaling factor for the $a*$ component for improving the performance for gray colors [19]. That is, some of the issues of ΔE^{J} could be solved by using the ΔE_{00} formula instead of the Euclidean distance.

V. Conclusions

This paper has reviewed and evaluated CD measures in the color-related aspect of image difference assessment. We tested eight state-of-the-art CD measures on selected data from the TID2013 database. We selected the data such that the following applications are included: color correction, color quantization, color matching, gamut mapping and multiview imaging systems. The results of our experiments reveal a poor correlation between subjective scores and the measures tested in this work for the whole dataset (PCC<0.65). Nonetheless, the ΔE_{00}, $\Delta E_{00}^{\mathrm{S}}$ and ΔE^{J} perform relatively well in computing CDs in images with quantization noise or intensity shift distortions (PCC>0.8). This means that these CD measures can be used in multiview imaging systems (for example, it can be used to assess CDs in a multiview imaging system where color inconsistencies between views are produced by different illumination conditions producing a different black level for each view) and in color quantization algorithms as an objective function for selecting the quantization step with the minimum CD impact.

Additionally, the results suggest that the CIELAB color space is the best performing color appearance model. Thus, we suggest that further development on CD measures in images should be designed by using this color appearance model. Also note that no statistically significant differences in terms of PCC were found when spatial preprocessing based on filtering is applied before computing pixel-wise CDs (specifically comparing ΔE_{00} and $\Delta E_{00}^{\mathrm{S}}$). This is important because it shows that many CD measures for images are designed using an ineffective mechanisms for computing CDs, i.e., the computation of pixel-wise differences after pre-processing based on filtering. However, it is well known that humans perceive better CD in flat areas than in complex structures [36]. Thus, it will be more desirable to measure CDs in homogeneous patches and then combine them into an overall CD.

Since the color content analysis is a novel procedure for studying the performance of CD measures, the study of other mechanisms to improve the proposed analysis in Section IV-C remains as future work. For instance, it would be interesting to analyse not only one color but instead to analyse the PCC in function of the number of colors and its location in the a^*b^*

plane. Furthermore, the evaluation of other CD measures in the state-of-the-art with the purpose of including more spatial processing methods and color appearance models remains as future work. Also, since the TID2013 database is not entirely suitable for color research because the images have assigned a quality score (MOS) instead of a perceived CD, it would be more desirable to build a dedicate CD assessment database including only color related distortions and scores assigned by using perceived CDs. Finally, a more comprehensive analysis with additional data (TID2013 is limited in color related distortions such as color shifts, gamut mapping, among others) with the purpose of evaluating the variability due to content and distortion types, should be conducted.

ACKNOWLEDGMENT

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

REFERENCES

[1] "Colorimetry - part 6: CIEDE2000 colour difference formula," https://www.iso.org/obp/ui/#iso:std:63731:en, 2014.

[2] J. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," *Coloration Technology*, vol. 124, pp. 243 – 253, 2008.

[3] Y. Yang, J. Ming, and N. Yu, "Color image quality assessment based on CIEDE2000," *Advances in Multimedia*, pp. 1 – 6, 2012.

[4] S. Millward, "Color difference equations and their assessment," *Test Targets*, vol. 9, pp. 19 – 26, 2009.

[5] X. Zhang and B. Wandell, "A spatial extension of CIELAB for digital color-image reproduction," *Journal of the Society for Information Display*, vol. 5, pp. 61 – 63, 1997.

[6] S. Fezza, M. Larabi, and K. Faraoun, "Feature-based color correction of multiview video for coding and rendering enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 1486 – 1498, 2014.

[7] L. Brun and A. Tremeau, *Digital Color Imaging Handbook*. CRC press, 2002, ch. Color quantization.

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

[9] J. Morovic, *Color Gamut Mapping*. John Wiley & Sons, Inc., 2008, ch. Desired Color Reproduction Properties and their Evaluation.

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

[11] U. Rajashekar, Z. Wang, and 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.

[12] Y. Ming, L. Huijuan, G. Yingchun, and 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.

[13] L. He, X. Gao, W. Lu, X. Li, and D. Tao, "Image quality assessment based on S-CIELAB model," *Signal, Image and Video Processing*, vol. 5, pp. 283 – 290, 2011.

[14] H. Sheikh, L. C. Z. Wang, and A. Bovik, "LIVE image quality database," http://live.ece.utexas.edu/research/quality/subjective.htm, 2004.

[15] E. Bando, J. Hardeberg, and 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.

[16] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 3, pp. 81 – 84, 2002.

[17] H. Kivinen, M. Nuutinen, and 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.

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

[19] M. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research and Application*, vol. 26, pp. 340 – 350, 2001.

[20] C. Chou and 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.

[21] G. Sharma, *Digital Color Imaging Handbook*. CRC press, 2002, ch. Color fundamentals for digital imaging.

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

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

[24] 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.

[25] A. Toet and M. Lucassen, "A new universal colour image fidelity metric," *Displays*, vol. 24, pp. 197 – 207, 2003.

[26] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE transactions on broadcasting*, vol. 50, pp. 312–322, 2004.

[27] D. Ruderman, T. Cronin, and C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, pp. 2036 – 2045, 1998.

[28] N. P. et.al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57 – 77, 2015.

[29] Z. Baranczuk, P. Zolliker, and J. Giesen, "Image-individualized gamut mapping algorithms," *Journal of Imaging Science and Technology*, vol. 54, pp. 030 201 1 – 7, 2010.

[30] X. Wei and 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.

[31] M. Habekost, "Which color differencing equation should be used?" *International Circular of Graphic Education and Research*, vol. 6, pp. 20 – 33, 2013.

[32] S. Garcia, A. Fernandez, J. Luengo, and F. Herrara, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Journal of Information Sciences*, vol. 180, p. 20442064, 2010.

[33] A. Klinger and C. Dyer, "Experiments on picture representation using regular decomposition," *Computer Graphics and Image Processing*, vol. 5, pp. 68 – 105, 1976.

[34] "Knowledge base - color management," http://help.efi.com/fieryxf/KnowledgeBase/, 2013.

[35] P. Rodrigues, L. Lopes, and M. Fonseca, "Getting a glimpse of your pictures using dominant colors," in *Proc. of the National Conference in Computer Graphics*, 2009, pp. 1 – 2.

[36] Y. Deng, C. Kenney, M. Moore, and 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.

[37] B. Ortiz-Jaramillo, A. Kumcu, L. Platisa, and 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. 89 – 94.