

Image Quality Assessment and Color Difference

Dogancan Temel and Ghassan AlRegib
Center for Signal and Information Processing (CSIP)
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia, USA
{cantemel,alregib}@gatech.edu

Abstract—An average healthy person does not perceive the world in just black and white. Moreover, the perceived world is not composed of pixels and through vision humans perceive structures. However, the acquisition and display systems discretize the world. Therefore, we need to consider pixels, structures and colors to model the quality of experience. Quality assessment methods use the pixel-wise and structural metrics whereas color science approaches use the patch-based color differences. In this work, we combine these approaches by extending CIEDE2000 formula with perceptual color difference to assess image quality. We examine how perceptual color difference-based metric (P_{CDM}) performs compared to PSNR, CIEDE2000, SSIM, MS-SSIM and CW-SSIM on the LIVE database. In terms of linear correlation, P_{CDM} obtains compatible results under white noise (97.9%), Jpeg (95.9%) and Jp2k (95.6%) with an overall correlation of 92.7%. We also show that P_{CDM} captures color-based artifacts that can not be captured by structure-based metrics.

Index Terms—color-difference, perceptual quality, objective quality metrics, color artifacts

I. INTRODUCTION

The phrase "quality of experience" in the image processing literature indicates the perceived quality of images. Therefore, perception matters as much as the fidelity for consumer electronics applications. Images are commonly analyzed in terms of pixels and structures in the image processing literature. However, color science literature mostly focus on large patches that are functions of visual fields. In our view, in order to model the full quality of experience, we need to consider the literature of both image processing and color science. The proposed approach contributes to the literature by utilizing the color label differences in the quality of experience estimation.

Objective quality metrics are used in the image processing literature to estimate the quality of experience or to quantify distortions. Pixel-wise fidelity metrics focus on the exact differences between pixels of the images. As an example, root-mean-square error and peak signal-to-noise ratio (PSNR) are commonly used in the literature because of their simplicity. Instead of calculating the pixel-wise fidelity, structural fidelity of images are also used to estimate the quality. SSIM calculates the local statistics of images over a single scale whereas MS-SSIM follows a multi-scale approach to calculate SSIM over different resolutions using Laplacian pyramid as described in [1]. CW-SSIM [2] also follows a multi-scale approach but instead of calculating the local statistics in the spatial domain, wavelet coefficients are used. Most of these quality metrics use

the luminance components or grayscale images and neglect the color channels.

In contrast to the luminance-based image quality measures, color information is commonly used in the color science literature to detect the differences between similar color tones [3]. The International Commission on Illumination (CIE) is responsible for the international coordination of lighting related technical standards including color difference. CIEDE2000 color difference equation was developed by the CIE technical committee and it is one of the state of the art metrics in the color science literature as described in [4], [5]. Color differences and similarities can be used as descriptors of the images. The authors in [6] propose learning color names from real-world images, which can be used for object recognition and image classification as described in [7]. Color naming descriptors are also used in image classification in terms of the aesthetics quality of the images as explained in [8]. Moreover, color naming descriptors are used in [9] to perform color-based edge detection.

Color difference formulas are commonly used in tone matching for color reproduction. The authors in [10] use a color difference-based metric to predict texture visibility of printer halftone patterns. The way color difference is used in [10] can be considered as a transition in the application field of color difference equations from basic tone matching to textured image comparisons. In [11], the authors discuss the connections between image quality, difference and appearance. Even color image quality is discussed in these approaches, authors consider the problem from the point of color science and leave the discussion limited without fully studying the objective quality metrics and their performance under different kinds of distortions such as compression and communication errors. The authors in [12] take the difference equations one step further and describe a calibration process for the color difference equations under experimental conditions as well as the usage of CIEDE2000 as an image quality metric. However, a very fundamental characteristic of the difference equations is overlooked. In principle, CIEDE2000 is designed for tone matching between similar colors that are bounded by, at most, medium differences and not utilized for significant tone differences.

In this paper, we augment the range of CIEDE2000 formula with perceptual color difference as in [9] to generalize the method proposed by the authors in [12]. In Section II, we de-

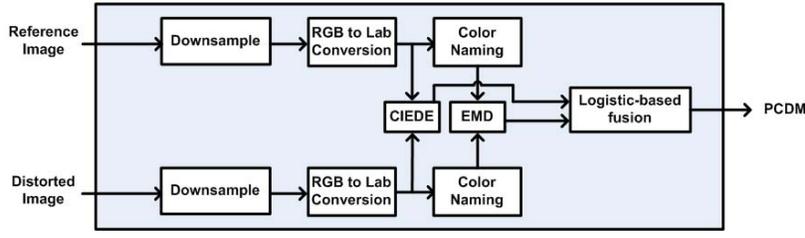


Fig. 1. Perceptual Color Difference-based Metric Pipeline

scribe the main blocks in the proposed quality metric pipeline. We discuss the experimental setup, results and observations in Section III and conclude our discussion in Section IV.

II. COLOR-BASED IMAGE QUALITY ASSESSMENT

The pipeline of the proposed image-quality assessment method is given in Fig. 1. Since human visual system is less sensitive to the color compared to the structure, color-based image quality can be calculated over the smoothed version of the image. Images are downsampled using the defaults of bicubic interpolation and anti-aliasing. Sampling rate is set to 0.05 after exhaustively simulating the range from 0.02 to 0.25. The effect of downsampling is explained in Section III-B. After downsampling, the RGB images are converted to the Lab domain to represent pixels in a perceptually correlated color space. Color descriptors are calculated for each pixel as it is described in Section II-A. We explain the usage of the Earth Mover’s Distance (EMD), the CIEDE2000 formula and the logistic-based fusion function in Section II-B.

A. Color Naming

Linguistic color names are used to label image pixels with perceived color classes. The authors in [6] introduced color naming as a 11 – D image descriptor. Each dimension in the descriptor corresponds to the probability of that pixel to be perceived as one of the 11 basic colors. These basic color names are: black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. In order to determine the pixel values for the finite color vocabulary, Google Image is used to obtain the training set, which also includes the wrongly labeled images. Color names are learned from the noisy data using the variants of the probabilistic latent semantic analysis model as explained in [6]. In addition to 11 – D color descriptors, we have also experimented 25 – D and 50 – D descriptors that span a wider color range. However, we have not observed a significant increase in the estimation accuracy.

B. Color Difference

The CIEDE2000 color difference equation is designed to calculate the difference between similar colors with low level variations. In the proposed metric, we keep the display and viewing related parameters constant to make the metric independent of acquisition and display configurations. In order to limit the calculation of CIEDE2000 to low-level differences, a threshold is set for color difference values and the result is divided by the same threshold to normalize the difference to be between 0 and 1. S-CIELab [13] color difference equation is

not used in the proposed metric to eliminate parameter tuning including but not limited to spatial and color calibration.

The Earth Mover’s Distance (EMD) is designed to calculate the difference between two distributions. The basic idea behind EMD is to calculate the minimal cost that is required to transfer one distribution into the other [14]. In [9], EMD is calculated between two 11 – D color naming descriptors where the flow between each color label probability is calculated to obtain the cost. Instead of using the uniform distance, the flow between color labels is scaled according to the perceived color distance. We can calculate this distance by using the joint distribution of basic color terms in the Lab color cube as explained in [6]. The visualization of the perceived distance is depicted in Fig. 2.

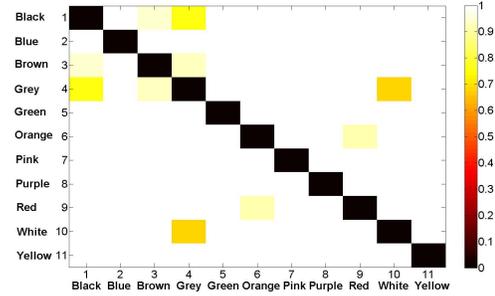


Fig. 2. Perceived distance between basic color categories.

EMD considers all flow scenarios from the source to the target color descriptor to minimize the total cost. We can formulate the EMD expression as in Eq. 1 where i is the index of the color label in the reference color descriptor and j is the index of the compared color descriptor. Flow from the i_{th} color probability in the reference to the j_{th} color probability in the compared descriptor is represented by f_{ij} and the perceived distance between the color terms in the dictionary is shown with d_{ij} .

$$EMD = \min_{f_{i,j}} \left\{ \sum_{i=1}^{11} \sum_{j=1}^{11} d_{i,j} f_{i,j} \right\} \quad (1)$$

The authors in [9] combined CIEDE2000 with EMD [14] using a logistic function to obtain a perceptually correlated difference function as formulated in Eq. 2. The difference function is used to detect edges. S_i is the Lab value and P_i is the color naming descriptor corresponding to a single pixel of the image indexed with i . α is set to 0.5 to equivalently combine CIEDE2000 and EMD and z is set to 10 as in [9]. The parameters in the logistic-based fusion are selected independent from the tested image database.

$$\frac{1}{1 + e^{-z((\alpha)CIEDE(S_1, S_2) + (1-\alpha)EMD(P_1, P_2)) - \frac{1}{2}}} \quad (2)$$

In the proposed pipeline, we assess the perceived quality of the images by extending the range of the color difference equation. CIEDE2000 is unreliable when the color difference is more than 7 CIELAB. Therefore, we threshold the color difference equation for high level differences and just use the transportation distance between the color descriptors that are inherently used for images significantly different from each other. The scope of the proposed metric is to introduce the perceptual color differences into image quality estimation. As a future work, other distance metrics can also be used to quantify the difference between color descriptors.

III. EXPERIMENTAL EVALUATION

A. Experimental Setup

In this paper, we use the release 2 version of the LIVE image database for the validation of the proposed metric. The resolution of most of the images in the database is 768x512 and images are interpolated (bicubic) to 1024x768 for subjective tests. There are 29 reference images and 779 distorted images. JPEG, JPEG2000 (Jp2k), White Noise (Wn), Gaussian Blur (Gblur) and Simulated Fast Fading Rayleigh Channel errors (FF) are the main sources of degradation in the image database. A more detailed information related to LIVE image database can be found in [15]. As it is explained in [15], the non-linear regression formulated in Eq. 3 is applied to each objective quality metric to make a fair comparison.

$$S = \beta_1 \left(\frac{1}{1 - \frac{1}{2 + \exp(\beta_2(S_0 - \beta_3))}} \right) + \beta_4 S_0 + \beta_5 \quad (3)$$

PCDM is used to refer to the proposed perceptual color difference-based metric in the rest of the paper. We compare the performance of PCDM with PSNR, CIEDE2000, SSIM, MS-SSIM and CW-SSIM, which were described in Section I.

B. Results

The scatter plots of PCDM under different distortion types are given in Fig. 3. Solid lines correspond to the ideal scenario where quality estimation is equivalent to the average subjective scores. Dashed lines are located one standard deviation away from the solid line and dotted lines are two standard deviations away. Linear correlation coefficient (CC) and root-mean-square error (RMSE) are calculated between the quality metric estimates and the difference mean opinion scores (DMOS) after non-linear regression. High CC values mean that the relation between the estimates and the subjective scores are highly linear. Low RMSE values indicate the proximity of the estimates to the DMOS values.

PCDM accurately estimates the subjective results under White Noise with a CC of 0.979 and RMSE of 5.08 as shown in Fig. 3(d). In the cases of Jpeg and Jp2k, nearly all of the estimates are in the distance of one standard deviation with CC values around 0.95 and RMSE values ranging from 7.10 to 7.49 as given in Fig. 3(b)-(c). However, when the images are distorted with Gaussian Blur and Fast Fading, some of the PCDM-based estimates exceed the one standard deviation boundary with CC values less than 0.9 and RMSE values

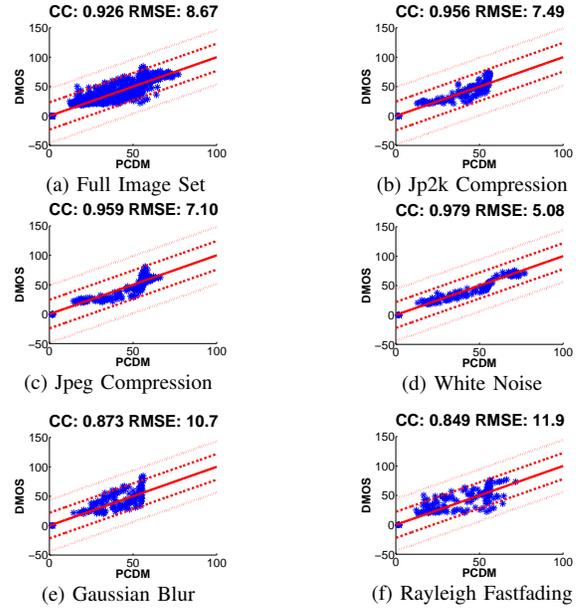


Fig. 3. Scatter plots of DMOS versus PCDM after non-linear regression.

higher than 10.0 as shown in Fig. 3(e)-(f). The performance comparison between PCDM and other objective quality metrics are given in Table I.

PCDM, CIEDE2000, SSIM and MS-SSIM are close to each other in terms of accurately estimating the subjective results under Jp2k, Jpeg and Wn distortions. However, both of the color-based metrics perform poorly compared to SSIM and MS-SSIM under FF and Gblur. PCDM performs better than CW-SSIM under all types of distortions and it performs better than PSNR under all of the distortion types except FF. Extending the range of CIEDE2000 with perceptual color difference leads to a decrease of 0.35 in the RMSE value and an increase of 0.07 in the CC value.

TABLE I
PERFORMANCE OF THE OBJECTIVE QUALITY METRICS

Metrics	Jp2k	Jpeg	Wn	Gblur	FF	All
Pearson CC (Linear)						
PSNR	0.923	0.913	0.945	0.843	0.887	0.898
CIEDE2000	0.954	0.956	0.981	0.892	0.850	0.920
SSIM	0.963	0.957	0.976	0.940	0.956	0.945
MS-SSIM	0.962	0.961	0.977	0.943	0.948	0.946
CW-SSIM	0.926	0.927	0.949	0.768	0.835	0.872
PCDM	0.956	0.959	0.979	0.873	0.849	0.927
RMSE						
PSNR	9.92	10.10	8.34	11.80	10.22	10.12
CIEDE2000	7.61	7.79	5.64	11.33	11.92	9.02
SSIM	7.11	7.74	8.65	7.54	6.45	7.52
MS-SSIM	7.12	7.30	8.38	7.38	7.04	7.43
CW-SSIM	9.75	9.30	9.24	14.45	13.62	10.87
PCDM	7.49	7.10	5.08	10.73	11.96	8.67

The authors in [15] examine quality versus distortion parameter distributions for the LIVE database. In the case of Fast Fading, data points are distributed all around the scatter plot with a high variation. Because there is not a high correlation between the distortion parameters and DMOS, it is expected that PCDM would not be a good model to estimate Fast Fading. However, there is a high correlation between the distortion

parameters and DMOS in Gaussian Blur and hence PCDM is expected to model the Gaussian Blur distortions.

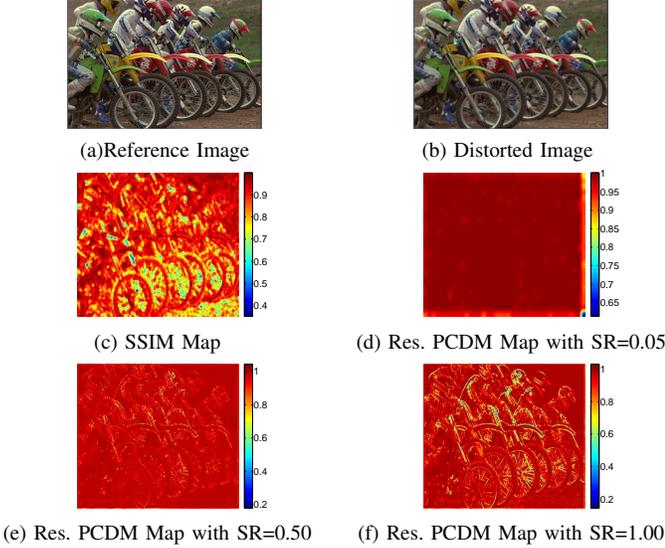


Fig. 4. Reference image, distorted image and corresponding distortion maps. SR: sampling rate, Res:Residual

We randomly selected the bikes image from the LIVE database to visualize the distortion maps. However, the same observations are also valid for other images. Reference and distorted images are shown in Fig. 4. Circular-symmetric 2-D Gaussian kernel with a standard deviation of 1.3 is used to degrade the reference image. In general, overall quality of the images is calculated by taking the average over the distortion maps. However, in Fig. 4, we directly show the distortion maps calculated by SSIM and PCDM to examine how the metrics perform. Residual of PCDM is plotted for a fair comparison since high PCDM corresponds to low quality and vice versa.

SSIM detects the observable degradations which are mostly around the textured regions as shown in Fig. 4(c). We examine the effect of downsampling by varying the sampling rate in PCDM model from 0.05 to 0.50 and then to 1.00. A sampling rate of 0.05 results in oversampling of the image and PCDM does not work well as shown in Fig. 4(d). When we increase the sampling rate to 0.5 and to 1.0, PCDM captures the degradations around the textured regions as shown in Fig. 4(e)-(f). We can obtain higher CC and lower RMSE with a higher sampling rate. However, the proposed method becomes infeasible because of the time-complexity. We summarize the changes in the average residual PCDM and the execution time per single image under varying sampling rate in Table II. We used an Intel(R) Core(TM) i7-3770 CPU @3.50 GHz with 32.0 GB ram.

TABLE II
PCDM UNDER VARYING SAMPLING RATE

PCDM Sampling Rate	Avg. Res. PCDM	Execution Time (sec)
0.05	0.98	1.3
0.50	0.96	120.5
1.00	0.93	479.5

Quality metrics such as SSIM, MS-SSIM and CW-SSIM estimate the perceived quality based on the structural cues by neglecting *chroma* information. However, PCDM utilizes

both *luma* and *chroma*. In order to examine how metrics perform under structural and color artifacts, Jpeg compressed image with a bpp of 0.208 is used. In the visualization, compression artifacts are preferred over Gaussian Blur and White Noise because compression corresponds to a more realistic degradation scenario. The distorted image is converted from the RGB color space to the YCbCr color space. We replace the chrominance channels (Cb and Cr) in the distorted image with the reference to obtain *Intensity* distortion as given in Fig. 5(a) and we replace the distorted luminance (Y) channel with the error-free channel in the reference to obtain the *Chroma* distortion as given in Fig. 5(b). *Chroma* distortion results in loss of color information around some connected regions such as the soil surface and the grass region. We also observe tonal changes over some objects and small regions. Whereas, *Intensity* distortion leads to blockiness artifacts all over the image especially around textured regions. SSIM detects the blockiness in *Intensity* distortion whereas PCDM does not capture the distortions with a sampling rate of 0.05. However, PCDM detects color losses at the background region and the tonal changes around foreground objects in the *Chroma* distortion, which are overlooked by SSIM.

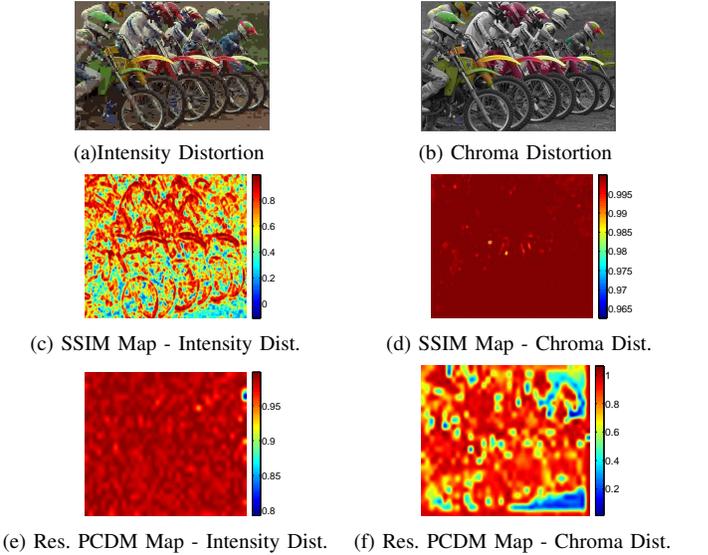


Fig. 5. Intensity and chroma distortion with corresponding distortion maps

IV. CONCLUSION

In this paper, we extended the CIEDE2000 formula using a perceptual color difference metric to estimate the subjective quality of images. We have shown that perceptual color difference metric results highly correlate with DMOS and it performs better than the pixel-wise fidelity metrics and the CIEDE2000 formula in terms of correlation and root-mean-square error. However, structural metrics perform better than color-based metric under Fast Fading and Gaussian Blur due to the oversampling in PCDM . When the sampling ratio is increased, PCDM performs better but it also increases the time-complexity significantly. We started combining color and structure based metrics to estimate the quality of experience for the end user and the hybrid metric already leads to promising results in LIVE and TID2013 image databases.

REFERENCES

- [1] Z. Wang, E. P. Simoncelli and A. C. Bovik, 'Multi-Scale Structural Similarity for Image Quality Assessment', *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003.
- [2] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, 'Complex Wavelet Structural Similarity: A New Image Similarity Index', *IEEE Transactions on Image Processing*, vol. 18, no. 11, Nov. 2009.
- [3] G. Wyszecki and W. S. Stiles, 'Color Science: Concepts and Methods, Quantitative Data and Formulae', *Wiley*, 1982.
- [4] M. Luo, G. Cui and B. Rigg, 'The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000', *Color Research & Application*, 2001.
- [5] G. Sharma, W. Wu and E. Dalal, 'The CIEDE2000 Color-Difference Formula: Implementation notes, Supplementary Test Data and Mathematical Observations', *Color Research & Application*, 2005.
- [6] J. Van De Wijer, C. Schmid, J. Verbeek and D. Larlus, 'Learning Color Names for Real World Applications', *IEEE Transactions in Image Processing*, 2009 2000.
- [7] J. van de Weijer and C. Schmid, 'Applying color names to image description', *IEEE International Conference on Image Processing*, Sep. 2007.
- [8] D. Temel and G. AlRegib, 'A Comparative Study of Computational Aesthetics', *IEEE International Conferernce in Image Processing*, 2014.
- [9] O. Pele and M. Werman, 'Improving Perceptual Color Difference using Basic Color Terms', *CoRR*, 2012.
- [10] X. Zhang and D.A. Silverstein and J.E. Farrell, and B.A. Wandell, 'Color image quality metric S-CIELAB and its application to halftone texture visibility', *IEEE International Computer Conference*, 1997.
- [11] G.M. Johnson and M.D. Fairchild, 'Measuring images: differences, quality, and appearance', *SPIE Human Vision and Electronic Imaging*, 2003.
- [12] Y. Yang and J. Ming and N. Yu, 'Color Image Quality Assessment Based on CIEDE2000', *Advances in Multimedia*, 2012.
- [13] X. Zhang and B. A. Wandell, 'A Spatial Extension of CIELAB for Digital Color Image Reproduction', *SID Journal*, 2012.
- [14] Y. Rubner, C. Tomasi and L. J. Guibas, 'The Earth Mover's Distance as a Metric for Image Retrieval', *International Journal of Computer Vision*, 2000.
- [15] H. R. Sheikh, M. F. Sabir and A. C. Bovik, 'A Statistical Evaluation of Recent Full Reference Quality Assessment Algorithms', *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.