Quality Assessment from Grayscale to Color Images

Mohammed A. Hassan^{1,*}, Mazen S. Bashraheel²

¹Department of Information Systems, Seiyun Community College, Seiyun, Yemen

²Department of Computer Engineering, Jordan University of Science and Technology, Irbid, Jordan

Copyright ©2017 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Color images reveal more meaningful information to the human observers rather than grayscale ones. Regardless of the advantages of the existing well-known objective image quality metrics, one of the common and major limitations of these metrics is that they evaluate the quality of grayscale images only and don't make use of color information. In this paper we propose an improved method for image quality assessment that adds a color comparison to the criteria of the well-known Structural Similarity index (SSIM). We validated the proposed metric on the TID2013 image database and the experimental results show that the quality scores given by the proposed metric are more consistent with HVS than other state-of-art metrics.

Keywords Color images, Image Quality Assessment, Structural Similarity Index (SSIM), SCIELAB

1 Introduction

Image quality assessment is an important tool in image processing systems, such as signal acquisition, synthesis, enhancement, watermarking, compression, transmission, storage, retrieval, reconstruction, authentication, display and printing. Image quality assessment methods can be classified into two categories: subjective and objective. The subjective image quality assessment methods are accurate in estimating the visual quality of an image because they are carried out by human subjects but involve a costly process which requires a large number of observers and takes significant time. On the other hand the objective image quality assessment methods are computer based methods that can automatically predict the perceived image quality. Hence the objective image quality assessment methods gained more popularity although they do not necessarily correlate well with the quality as perceived by humans [1, 2].

The objective image quality assessment methods can be employed for the following functions: *monitoring* image quality in real-time applications, *benchmarking* image-

processing algorithms and for *embedding* in into image-processing and transmission systems for optimization purposes [3].

Objective image quality assessment methods also may be classified into full reference, reduced reference, and no-reference methods based on the availability of the reference image. Full reference image quality assessment requires complete information about the reference image; and partial information about the reference image is required for the reduced reference image quality assessment; while no information about the reference image is needed in no-reference image quality assessment. This paper focuses on the full reference image quality assessment methods for color images where both the original and the test images are available.

Over the years a large number of objective metrics have been proposed to assess image quality. These include many algorithms that incorporate perceptual models [4, 5, 6, 7, 8, 9, 10] and distortion models [11, 12]. Miyahara [10] proposed a Picture Quality Scale (PQS) based on three distortion factors: the amount, location, and structure of error. Wang and Bovik [13] proposed a new universal image quality index (UQI) and its improved form the Structural Similarity (SSIM) index [14] by modeling the image distortion as a combination of loss of luminance, contrast, and correlation. More recent algorithms based on the successful Structural Similarity (SSIM) index that incorporate saliency [15], compression specificity [16], multi-scales [17], amount and local information [18], wavelet-domain processing [19], and color [20]. Information fidelity criterion (IFC) [3] and visual information fidelity (VIF) [21] both are based on information-theory in which the distorted image is modeled as a sequence of passing the reference images through distortion channels and quantify the visual quality as a mutual information between the test image and the reference image. Shnayderman [22] explored the feasibility of Singular Value Decomposition (SVD) for quality measurement. In [23] a two staged wavelet based visual signal to noise ratio (VSNR) was proposed based on the low-level and the mid-level properties of human vision.

Consider the original image in Fig. 1 that is distorted by different noises from TID2013 image database[24]. Images

in Fig. 1b and Fig. 1d are distorted by two color-related noises, quantization noise and additive noise in color components respectively. Image in Fig. 1c is distorted by additive Gaussian noise which is non color-related noise. It is clear that Structural Similarity index (SSIM) fails to assess the perceptual quality of these color-related noises as it is a grayscale index. Images in Fig. 1c and Fig. 1d are almost of the same perceptual quality but having drastically different SSIM scores as 0.579 and 0.830 respectively. Image in Fig. 1b has the worst perceptual quality among all distorted images but is given the highest SSIM score 0.850. This necessitates considering color information while assessing the perceptual quality of color images.

Figure 1. Different distortion types from TID2013 image database [24]. (a) Original image, (b) Quantization noise distorted image with SSIM=0.850, (c) Additive Gaussian noise distorted image with SSIM=0.579, (d) Additive noise in color components distorted image with SSIM=0.830.

Based on these considerations, a novel image quality assessment method, CSSIM, has been designed and published [25]. Due to limited space in Conference Proceedings, results of comparisons between CSSIM and the existing image quality assessment methods have not been addressed and analyzed in [25].

The remainder of this paper is organized as follows. In Section 2 the grayscale SSIM is simply introduced. The proposed method is described in Section 3. Section 4 discusses the experimental results and discussion. Finally, section 5 draws the conclusion.

2 The Grayscale Index

Structural Similarity (SSIM) index [26] was proposed based on the assumption that the human visual system is highly adapted to extract structural information from visual scenes. Structural similarity index consists of three local comparison

functions namely luminance comparison, contrast comparison, and structure comparison between two images x and y :

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (1)$$

$$C(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2)$$

$$S(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (3)$$

Where μ_x and μ_y are the sample means of the images x and y respectively, σ_x and σ_y are the sample standard deviations of x and y , and σ_{xy} is the sample correlation coefficient between x and y . The constants C_1 , C_2 , and C_3 are used to stabilize the algorithm when the denominators approach to zero. These statistics are calculated within a local window.

Then the general form of the SSIM index, as shown in Figure 2, is given by combining the three comparison functions:

$$SSIM = l(x, y)^\alpha \cdot C(x, y)^\beta \cdot S(x, y)^\gamma \quad (4)$$

where α , β , and γ are parameters which define the relative importance of the three components. Usually, $\alpha = \beta = \gamma = 1$ and $C_2 = 2C_3$ yielding:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

3 The Color Image Quality Metric

The common RGB color space used for representing images is not consistent with the human visual system (HVS) in which the Euclidean distances between different colors in the color space does not approximately correspond to perceived color differences by HVS. On the other hand The CIE-LAB color space proposed by Commission Internationale de l'Eclairage (CIE) [27] is consistent with the human visual system and is considered perceptually uniform and referred to as uniform color space. This color space was created based on data from color matching and discrimination experiments using large uniform colored areas. With the development of the color image technology, many applications have been developed to process real images. However, most real images are made up of many small region and texture structures rather than of large uniform areas, this makes the application of the CIELAB color space for calculating color differences in images is unsatisfactory. Zhang and Wandell [28] proposed the S-CIELAB, which is an improvement and a spatial extension of CIELAB by adding a color separation and spatial filtering procedure in order to simulate the spatial blurring by the human visual system. The detailed steps and parameters setting are as follow:

The implementation of S-CIELAB includes transforming the input color image data, specified in terms of the CIE XYZ tristimulus values, into an opponent-colors space by the following linear transformation:

Figure 2. Diagram of the Structural Similarity (SSIM) Measurement System [26].

$$\begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} = \begin{bmatrix} 0.279 & 0.72 & -0.107 \\ -0.449 & 0.29 & -0.077 \\ 0.086 & -0.59 & 0.501 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (6)$$

where Q_1 , Q_2 , and Q_3 represent the three opponent colors planes: luminance, red-green and blue-yellow respectively.

Each of these opponent-colors planes Q_1 , Q_2 and Q_3 are then filtered with a specific spatial kernel which is determined by spatial response of human visual system and the values of parameters are obtained by vision physiological and perception experiments. The calculation formula for the kernels is given as following:

$$f = k \sum_i w_i E_i, \quad E_i = k_i \exp \left[-(\hat{x}^2 + \hat{y}^2) / \sigma_i^2 \right] \quad (7)$$

where k and k_i are selected as normalization factors: k_i normalizes E_i such that E_i sums to one and k normalizes each color plane such that the kernel f sums to one. The parameters w_i and σ_i are the values for weight and spread of the filtering kernels which are different for the three color planes. $\hat{x}$ and $\hat{y}$ represent 2D kernel's width and height in pixels. i varies from 1 to 3 for Luminance plane (Q_1) and from 1 to 2 for red-green (Q_2) and blue-yellow (Q_3) as shown in Table 1. More details of the spatial filtering can be found in [28, 29, 30, 31, 32].

Then, the filtered data in each plane is transformed back to CIE XYZ tristimulus values using the following formula:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.6204 & -1.8704 & -0.1553 \\ 1.3661 & 0.9316 & 0.4339 \\ 1.5013 & 1.4176 & 2.5331 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \quad (8)$$

Finally, this representation is transformed to CIELAB space as following:

Table 1. weight and spread of the kernels for different color plane

Color Plane	i	w_i	σ_i
Luminance	1	0.921	0.0283
	2	0.105	0.133
	3	-0.108	4.336
Red-green	1	0.531	0.0392
	2	0.330	0.494
Blue-yellow	1	0.488	0.0536
	2	0.371	0.386

$$L^* = \begin{cases} 116(Y/Y_0)^{1/3} - 16, & (Y/Y_0 > 0.008856) \\ 903.3(Y/Y_0)^{1/3}, & (Y/Y_0 \leq 0.008856) \end{cases}$$

$$a^* = 500[f(X/X_0) - f(Y/Y_0)],$$

$$b^* = 200[f(Y/Y_0) - f(Z/Z_0)] \quad (9)$$

where $f(t) = \begin{cases} t^{1/3}, & (t > 0.008856) \\ 7.787t + 16/116, & (t \leq 0.008856) \end{cases}$, and X_0 , Y_0 , and Z_0 are the tristimulus values of the reference white (D_{65}).

The new index is proposed by extending the grayscale metric (4) to include color information. This is done by modeling any image distortion as a combination of four local comparison functions namely luminance comparison, contrast comparison, structure comparison, and color comparison. The proposed quality metric is defined as:

$$CSSIM = l(x, y) \cdot C(x, y) \cdot S(x, y) \cdot C_r(x, y) \quad (10)$$

where $l(x, y)$, $C(x, y)$, and $S(x, y)$ are the luminance, contrast, and structure comparisons between the original and distorted images. $C_r(x, y)$ is the color comparison which is defined as:

$$C_r(x, y) = 1 - \frac{1}{k} \Delta E(x, y) \quad (11)$$

where $\Delta E(x, y)$ is the perceptual color difference between the original and the distorted images in the S-CIELAB color space. k is a weighting constant.

To calculate the perceptual difference between the original image and the distorted one, the ΔE value is computed as follows:

$$\Delta E(x, y) = \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2} \quad (12)$$

In order to choose the best value of the weighting constant k , we randomly selected one of the distortion types from the TID2013 image database. For various values of k , the value that maximized the Pearson Linear Correlation Coefficient (PLCC) was chosen as the value of k . Fig. 3 that shows the Pearson Linear Correlation Coefficient (PLCC) as a function of k for $CSSIM(x, y)$ and consequently 45 was chosen as the best value for k .

Figure 3. Pearson Linear Correlation Coefficient (PLCC) as a function of k for $CSSIM(x, y)$.

The CSSIM metric is computed by sliding 11×11 window over the entire image as [26]. At each image coordinate, the CSSIM metric is calculated yielding CSSIM metric map that can be spatially pooled resulting a single descriptor of the image objective quality.

4 RESULTS AND DISCUSSION

In this section, the performance of the proposed image quality metric, CSSIM, in terms of the ability of predicting the subjective ratings is analyzed. The proposed quality metric and the other image quality metrics used in this analysis were applied to the set of images from Tampere Image Database (TID2013) [24]. This database is the most recent and largest database so far available that includes more images and more distortion types for verification of full reference

quality metrics. The TID2013 database contains a 3000 test images obtained from 25 reference images, 24 types of distortions for each reference image, and 5 levels for each type of distortion. Mean opinion scores (MOS) for the new database have been collected by performing 985 subjective experiments with observers from five countries (Finland, France, Italy, Ukraine, and USA).

Figure 4 shows the scatter plots for the scores given by the proposed CSSIM metric and the original SSIM against the MOS (Mean Opinion Scores) from TID2013 image database for different distortion types. To accurately reflect the MOS, scatter plots should be close to the line. It is easy to know that the proposed CSSIM metric is better close to the line than SSIM. So, there is clear improvement in the performance of CSSIM compared to SSIM.

For numerical comparison, two criteria were used to validate the performance of the proposed CSSIM metric along with several well-known objective image quality metrics. These criteria characterize two attributes related to the prediction of each image quality metric [33]: Pearson's linear correlation coefficient (PLCC) as a measure of Prediction linearity, and Spearman's rank order correlation coefficient (SRCC) as a measure of Prediction Monotonicity.

The image quality metrics used in this analysis are: noise quality measure (NQM) metric [11], weighted signal to noise ratio (WSNR) [34], visual signal to noise ratio (VSNR) [23], information fidelity criterion (IFC) [3], visual information fidelity (VIF) [21], feature similarity (FSIM) metric [35], structural similarity index (SSIM) [14] and multiscale structural similarity index (MSSIM) [17]. These metrics are commonly used and their implementations are publicly available on the Internet.

For overall performance comparison of image quality metrics, Table 2 and Table 3 show the performance comparison of the Pearson's linear correlation coefficients (PLCC) and the Spearman's rank order correlation coefficients (SROCC) respectively of the nine quality metrics for each distortion type in the TID2013 image database. It is clear from those Tables that the proposed CSSIM metric greatly improves the performance of the classical SSIM for almost all distortion types. This improvement reaches more than 100% in change of color saturation distortion type and around 40% in contrast change distortion. It also can be noticed that the proposed CSSIM metric outperforms the other metrics for many types of distortion and comparable for the rest.

Practically, researchers showed that it is reasonable to study MOS estimated for all types of distortions as well as for particular subsets [36, 37]. A subset is usually formed by researchers depending upon an application and it may include one or several types of distortions. Table 4 shows subsets of distortions for TID2013 image database used for verification of image quality metrics (distortions that belong to a given subset are marked by +).

The subset "Noise" contains different types of noise and distortions in conventional image processing; the subset "Actual" relates to types of distortions most common in practice of image processing including compression, the subset "Sim-

Figure 4. Scatter plots for the scores given by the proposed CSSIM metric (left) and the original SSIM metric (right) against the MOS (Mean Opinion Scores) from TID2013 image database for different distortion types: (a) Additive Gaussian noise, (b) High frequency noise, (c) Image denoising, (d) JPEG compression, (e) JPEG2000 transmission errors, (f) Change of color saturation, (g) Multiplicative Gaussian noise, (h) Comfort noise, (i) Lossy compression of noisy images, (j) Image color quantization with dither, (k) Chromatic aberrations, (l) Sparse sampling and reconstruction.

Table 2. Pearson's Linear Correlation Coefficients (PLCC) of the scores given by different image quality assessment methods against MOS from the TID2013 image database.

Type of distortion	NQM	WSNR	VSNR	IFC	VIF	FSIM	MSSIM	SSIM	CSSIM
Additive Gaussian noise	0.789	0.878	0.814	0.634	0.902	0.825	0.793	0.828	0.905
Additive noise in color components	0.706	0.846	0.739	0.516	0.866	0.816	0.764	0.781	0.893
Spatially correlated noise	0.770	0.854	0.787	0.661	0.879	0.804	0.770	0.807	0.905
Masked noise	0.670	0.640	0.743	0.726	0.879	0.797	0.796	0.817	0.843
High frequency noise	0.896	0.933	0.879	0.717	0.946	0.848	0.835	0.873	0.936
Impulse noise	0.764	0.899	0.723	0.601	0.849	0.696	0.658	0.751	0.825
Quantization noise	0.806	0.881	0.829	0.116	0.781	0.800	0.771	0.767	0.875
Gaussian blur	0.886	0.948	0.934	0.895	0.953	0.891	0.862	0.924	0.876
Image denoising	0.943	0.948	0.898	0.701	0.892	0.924	0.910	0.914	0.937
JPEG compression	0.905	0.930	0.904	0.784	0.933	0.920	0.924	0.918	0.959
JPEG2000 compression	0.942	0.954	0.937	0.789	0.919	0.933	0.917	0.895	0.954
JPEG transmission errors	0.770	0.743	0.658	0.773	0.900	0.870	0.856	0.854	0.859
JPEG2000 transmission errors	0.802	0.874	0.813	0.737	0.870	0.797	0.788	0.825	0.875
Non eccentricity pattern noise	0.741	0.745	0.646	0.537	0.769	0.790	0.741	0.742	0.747
Local block-wise distortions of dif intensity	0.010	-0.013	0.235	0.237	0.502	0.528	0.458	0.610	0.585
Mean shift (intensity shift)	0.588	0.738	NaN	0.607	0.683	0.690	0.678	0.738	0.789
Contrast change	0.511	0.479	0.387	-0.248	0.875	0.692	0.727	0.429	0.603
Change of color saturation	0.342	0.046	0.059	-0.320	-0.040	0.178	-0.369	-0.390	0.678
Multiplicative Gaussian noise	0.755	0.841	0.750	0.593	0.857	0.787	0.735	0.756	0.880
Comfort noise	0.891	0.923	0.871	0.731	0.899	0.916	0.814	0.839	0.925
Lossy compression of noisy images	0.911	0.948	0.865	0.745	0.904	0.932	0.894	0.910	0.951
Image color quantization with dither	0.849	0.899	0.859	0.311	0.860	0.791	0.776	0.771	0.906
Chromatic aberrations	0.914	0.895	0.857	0.816	0.944	0.941	0.910	0.947	0.956
Sparse sampling and reconstruction	0.953	0.952	0.896	0.786	0.907	0.935	0.922	0.890	0.952

Table 3. Spearman's Rank Order Correlation Coefficient (SRCC) of the scores given by different image quality assessment methods against MOS from the TID2013 image database.

Type of distortion	NQM	WSNR	VSNR	IFC	VIF	FSIM	MSSIM	SSIM	CSSIM
Additive Gaussian noise	0.803	0.880	0.828	0.663	0.899	0.902	0.866	0.854	0.934
Additive noise in color components	0.713	0.850	0.731	0.544	0.831	0.823	0.773	0.780	0.869
Spatially correlated noise	0.782	0.859	0.801	0.661	0.883	0.879	0.854	0.862	0.930
Masked noise	0.668	0.639	0.701	0.743	0.848	0.800	0.809	0.810	0.809
High frequency noise	0.861	0.912	0.846	0.745	0.897	0.898	0.861	0.847	0.930
Impulse noise	0.783	0.904	0.737	0.640	0.854	0.811	0.763	0.799	0.890
Quantization noise	0.814	0.889	0.836	0.632	0.786	0.872	0.866	0.806	0.913
Gaussian blur	0.894	0.949	0.947	0.910	0.965	0.957	0.970	0.963	0.948
Image denoising	0.917	0.939	0.908	0.781	0.893	0.931	0.924	0.911	0.954
JPEG compression	0.869	0.927	0.901	0.835	0.919	0.930	0.928	0.910	0.951
JPEG2000 compression	0.923	0.945	0.926	0.908	0.952	0.958	0.948	0.905	0.953
JPEG transmission errors	0.722	0.733	0.791	0.750	0.842	0.845	0.850	0.818	0.850
JPEG2000 transmission errors	0.806	0.878	0.841	0.778	0.877	0.895	0.883	0.870	0.928
Non eccentricity pattern noise	0.742	0.770	0.665	0.573	0.772	0.789	0.794	0.759	0.790
Local block-wise distortions of dif intensity	0.005	-0.045	0.178	0.240	0.530	0.555	0.468	0.618	0.594
Mean shift (intensity shift)	0.600	0.773	0.538	0.550	0.631	0.729	0.793	0.777	0.783
Contrast change	0.460	0.420	0.333	-0.272	0.840	0.472	0.465	0.347	0.477
Change of color saturation	0.354	-0.023	0.038	-0.369	-0.162	0.189	-0.357	-0.357	0.721
Multiplicative Gaussian noise	0.761	0.856	0.765	0.623	0.846	0.854	0.784	0.776	0.911
Comfort noise	0.872	0.919	0.868	0.818	0.895	0.913	0.858	0.818	0.928
Lossy compression of noisy images	0.902	0.944	0.882	0.818	0.921	0.948	0.908	0.911	0.955
Image color quantization with dither	0.852	0.895	0.866	0.608	0.846	0.877	0.858	0.790	0.913
Chromatic aberrations	0.812	0.859	0.865	0.843	0.886	0.874	0.883	0.889	0.875
Sparse sampling and reconstruction	0.946	0.945	0.934	0.889	0.936	0.957	0.948	0.903	0.958

Table 4. Distortion types and considered subsets of TID2013 [24].

Type of distortion	Noise	Actual	Simple	Exotic	New	Color
Additive Gaussian noise	+	+	+	-	-	-
Additive noise in color components	+	-	-	-	-	+
Spatially correlated noise	+	+	-	-	-	-
Masked noise	+	+	-	-	-	-
High frequency noise	+	+	-	-	-	-
Impulse noise	+	+	-	-	-	-
Quantization noise	+	-	-	-	-	+
Gaussian blur	+	+	+	-	-	-
Image denoising	+	+	-	-	-	-
JPEG compression	-	+	+	-	-	+
JPEG2000 compression	-	+	-	-	-	-
JPEG transmission errors	-	-	-	+	-	-
JPEG2000 transmission errors	-	-	-	+	-	-
Non eccentricity pattern noise	-	-	-	+	-	-
Local block-wise distortions of different intensity	-	-	-	+	-	-
Mean shift (intensity shift)	-	-	-	+	-	-
Contrast change	-	-	-	+	-	-
Change of color saturation	-	-	-	-	+	+
Multiplicative Gaussian noise	+	+	-	-	+	-
Comfort noise	-	-	-	+	+	-
Lossy compression of noisy images	+	+	-	-	+	-
Image color quantization with dither	-	-	-	-	+	+
Chromatic aberrations	-	-	-	+	+	+
Sparse sampling and reconstruction	-	-	-	+	+	-

Table 5. Distortion Group-wise Spearman's Rank Order Correlation Coefficient (SRCC) of the scores given by different image quality assessment methods against MOS from the TID2013 image database.

Distortion Group	NQM	WSNR	VSNR	IFC	VIFP	FSIM	MSSIM	SSIM	CSSIM
Noise	0.836	0.88	0.869	0.7218	0.784	0.897	0.873	0.757	0.885
Actual	0.857	0.897	0.882	0.7608	0.815	0.911	0.887	0.788	0.899
Simple	0.875	0.933	0.912	0.7792	0.897	0.949	0.905	0.837	0.882
Exotic	0.589	0.423	0.706	0.3610	0.557	0.844	0.841	0.632	0.404
New	0.625	0.646	0.589	0.5444	0.589	0.649	0.631	0.579	0.891
Color	0.538	0.555	0.512	0.4449	0.506	0.565	0.566	0.505	0.892

ple” includes only three standard types of distortions; the subset ”Exotic” corresponds to distortions that happen not often but are among the ”most difficult” for visual quality metrics. The subset ”new” is for new types of distortions introduced to the old version of the image database TID2013 [36], the subset ”color” to distortion types that are in one or another manner connected with changes of color content [24]. Table 5 shows the performance comparison of the Spearman’s rank order correlation coefficients (SROCC) for the considered metrics for subsets of distortion types from TID2013 image database. As expected, for the subset ”color” the proposed CSSIM metric performs well compared with the other metrics as it fully considers color information in the process of image quality assessment. The situation is similar for the subset ”new”. For the the subsets ”Noise”, ”Actual” and ”Simple” the performance of the proposed CSSIM metric is comparable with the other metrics. In turn, the subset ”Exotic” causes problems for most of the metrics. Only two metrics (FSIM, MSSIM) have SROCC with MOS above % 0.80.

5 CONCLUSIONS

In this paper, we present an improvement to the well-known Structural Similarity index (SSIM), which follows the HVS’s characteristic that color images reveal more meaningful information to the human observers rather than grayscale ones. This may be the primary reason that the proposed CSSIM metric has greatly improved the performance of SSIM and also more consistent with human perception of image quality assessment, especially for the color-related distortions.

REFERENCES

- [1] S. Ouni, M. Chambah, M. Herbin, and E. Zagrouba. Are existing procedures enough? image and video quality assessment: Review of subjective and objective metrics. In *Electronic Imaging, Image Quality and System Performance*, San Jose, USA, 2008. SPIE.
- [2] Z. Wang, A. C. Bovik, and L. Lu. Why is image quality assessment so difficult? In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 3313–3316, 2002.
- [3] H. R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.
- [4] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.
- [5] S.A. Karunasekera and N.G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE Transactions on Image Processing*, 4(6):713–724, 1995.
- [6] N.B. Mill. A visual model weighted cosine transform for image compression and quality assessment. *Transactions on Communications*, 33(6):551–557, 1985.
- [7] J.A. Saghri. Image quality measure based on a human visual system model. *Optical Engineering*, 28(7):813–818, 1989.
- [8] A.P. Bradley. A wavelet visible difference predictor. *IEEE Transactions on Image Processing*, 8:717–730, 1999.
- [9] M.P. Eckert and A.P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Processing*, 70(3):177–200, 1998.
- [10] M. Miyahara, K. Kotani, and R. Algazi. Objective picture quality scale (pqs) for image coding. *IEEE Transactions on Communications*, 46(9):1215–1226, 1996.
- [11] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 9:636–650, 2000.
- [12] A. Toet and M. Lucassen. A new universal colour image fidelity metric. *Displays*, 24:197–204, 2003.
- [13] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.
- [14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [15] A.K. Moorthy and A.C. Bovik. Visual importance pooling for image quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):193–200, 2009.
- [16] A.C. Brooks, X.N. Zhao, and T.N. Pappas. Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. *IEEE Transactions on Image Processing*, 17(8):1–12, 2008.
- [17] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multi-scale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems, and Computers*, pages 1398–1402, 2003.
- [18] X.B. Gao, T. Wang, and J. Li. A content-based image quality metric. In *Proceedings of LNAI*, volume 3642, pages 231–240, 2005.
- [19] Z. Wang and E. Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 573–576, 2005.

- [20] M. Hassan and C. Bhagvati. Structural similarity measure for color images. *International Journal of Computer Applications*, 43(14):7–12, 2012.
- [21] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.
- [22] A. Shnayderman, A. Gusev, and A.M. Eskicioglu. An SVD-based gray-scale image quality measure for local and global assessment. *IEEE Transaction on Image Processing*, 15(2):422–429, 2006.
- [23] D.M. Chandler and S. S. Hemami. VSNR: A wavelet base visual signal-to-noise ratio for natural images. *IEEE Transaction on Image Processing*, 16(9):2284–2298, 2007.
- [24] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, Jan 2015.
- [25] M. Hassan and M. Bashraheel. Color-based structural similarity image quality assessment. In *8th International Conference on Information Technology, Jordan*, pages 691–696, May 2017.
- [26] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Transaction on Image Processing*, 13(4):600–612, 2004.
- [27] CIE: Recommendations on uniform color spaces, color difference equations, psychometric color terms. Publication CIE 15 (E.-1.3.1), Supplement No.2, Bureau Central de la CIE, Vienna, 1971.
- [28] X. Zhang and B. A. Wandell. A spatial extension of cielab for digital color image reproduction. *SID*, 5(1):61–63, 1997.
- [29] Xuemei Zhang, Joyce E Farrell, Brian A Wandell, et al. Applications of a spatial extension to cielab. In *Proc. SPIE*, volume 3025, pages 154–158, 1997.
- [30] Xuemei Zhang and Brian A Wandell. Color image fidelity metrics evaluated using image distortion maps. *Signal processing*, 70(3):201–214, 1998.
- [31] Xuemei Zhang, D Amnon Silverstein, Joyce E Farrell, and Brian A Wandell. Color image quality metric s-cielab and its application on halftone texture visibility. In *Compcon'97. Proceedings, IEEE*, pages 44–48. IEEE, 1997.
- [32] Allen B Poirson and Brian A Wandell. Appearance of colored patterns: pattern-color separability. *JOSA A*, 10(12):2458–2470, 1993.
- [33] BT Series. Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT*, pages 500–13, 2012.
- [34] T. Mitsa and K. Varkur. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. In *ICASSP*, pages 301–304, 1993.
- [35] L. Zhang, X. Mou, and D. Zhang. FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(5):2378–2386, 2011.
- [36] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.
- [37] N. Ponomarenko, O. Ieremeiev, V. Lukin, L. Jin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al. A new color image database tid2013: Innovations and results. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 402–413. Springer, 2013.