Application of the CIE color spaces for the digital image quality assessment

Krzysztof Okarma[*]

Chair of Signal Processing and Multimedia Engineering,
West Pomeranian University of Technology,
ul. 26. Kwietnia 10, 71-126 Szczecin, Poland.

Abstract – The digital image quality assessment is one of the most relevant aspects of contemporary digital image processing. A rapid development of some modern quality assessment techniques in recent years has caused the introduction of some new metrics, much better correlated with the Human Visual System (HVS) than the traditional ones such as the Mean Squared Error (MSE) or PSNR (Peak Signal to Noise Ratio). One of the most popular modern image quality assessment techniques is the usage of the Structural Similarity index (SSIM) defined in 2004. Unfortunately, even some modern image quality metrics are usually defined for the grayscale images so the colour information is often ignored. A typical classical approach to the quality assessment of the color images is the use of the Normalized Color Difference calculated in the CIE L*a*b* ob CIE Lu'v' colour space but its correlation with the human assessment is rather poor. In the paper the analysis of the influence of using the color spaces recommended by the CIE on the results of the digital image quality assessment using some modern metrics is performed. All the results have been calculated for the widely known LIVE database (release 2) containing the Differential Mean Opinion Scores (DMOS) for nearly 1000 color images with five types of distortions: JPEG compression, JPEG2000 compression, Gaussian blur, white noise and transmission over the simulated fast fading Rayleigh channel typical of wireless transmissions. As the final result the comparison of the correlations between the DMOS values and the SSIM metric calculated for various colour spaces recommended by the CIE is presented.

[*]okarma@zut.edu.pl

1   Introduction

Applications of the digital image quality assessment in the field of contemporary image processing and analysis are mainly related to the verification of the results of using newly developed image processing (e.g. filtering) algorithms as well as some new lossy compression methods. In order to validate the usefulness of such compression algorithm, it is crucial to somehow measure the loss of the visual quality of such compressed image in comparison to the results of using some existing techniques. Obviously, a new compression scheme introducing many visible artifacts with only a little smaller file size would not be a desirable solution, similarly to any computationally demanding filtering method leading to an insignificant increase of the image quality.

Reliable image quality assessment is especially relevant in the domain of colour image processing. The first reason is the specific character of many colour image processing methods, which are not the straightforward extensions of their grayscale equivalents. A good example can be the median filtration, which should be treated as the vector operation for multichannel images, so a number of various algorithms for the impulse noise suppression in colour images have been proposed. On the other hand, many researchers still use some classical methods of image quality assessment [1], often based on the Mean Squared Error (MSE) and usually developed for grayscale images. Considering the traditional image quality metrics, only the Normalized Color Difference (NCD) can be distinguished as devoted to the assessment of the colour images. Assuming the usage of the CIE L*a*b* colour space (Lu'v' denoted also as the u'v'L can be also used) it is defined as:

$$NCD = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} \sqrt{(L_{i,j} - L'_{i,j})^2 + (a_{i,j} - a'_{i,j})^2 (b_{i,j} - b'_{i,j})^2}}{\sum_{i=1}^{M} \sum_{j=1}^{M} \sqrt{L_{i,j}^2 + a_{i,j}^2 + b_{i,j}^2}} \tag{1}$$

where L,a,b denote CIE L*a*b* channels of the original image and L',a',b' stand for the same for the distorted one. The conversion from RGB to CIE XYZ and then to CIE L*a*b* (or Lu'v') colour space is based on nonlinear functions depending also on the white point [12].

2   Structural similarity for colour images

The most attractive approach to image quality assessment is based on the objective methods as the final result is computed automatically by the computer system. Such approach can be successfully implemented e.g. for optimization purposes, especially using the scalar metrics, where the quality of the image can be described using a single numerical value only.

Another group of objective quality assessment methods consists of vector measures expressing the image quality as the set of numbers usually related to various types of distortions. Such metrics are often based on some complicated numerical approximations of the Human Visual System (HVS), so they are rarely used due to their computational complexity.

The alternative method is using the human interaction which leads to the development of some subjective methods. A typical technique is based on the Mean Opinion Scores (MOS) calculated on the basis of analysis of scores specified by a number of human observers. Such procedure can be very helpful for the development of some new objective measures but cannot substitute them because of the time consuming interaction with observers.

A well known database of the images contaminated by five types of distortions with the subjective quality scores is so called LIVE Image Quality Database [2]. Release 2 contains 982 images with five types of distortions: JPEG and JPEG2000 compression, white Gaussian noise, Gaussian blurred ones and the JPEG2000 compressed images transmitted over the simulated fast fading Rayleigh channel with bit errors typical of the wireless transmission. Each image has been the subject of the quality evaluation by several observers seven times. Taking into account the presence of 29 original images, 203 of the scores are related to the images without any distortions. In this paper they are removed from the analysis because of their influence on the final results and the fact that the values of the objective metrics are easy to predict (no errors, maximum similarity).

Taking into account the possibilities of using the DMOS (Differential MOS) values present in the LIVE database for the comparison and optimization purposes in relation to the objective image quality metrics, some colour-related modifications of the state-of-the-art metric known as the Structural Similarity (SSIM) index [3] are proposed in the paper.

The SSIM metric comes from the previously introduced Universal Image Quality Index (UIQI) [4] defined as

$$Q = \frac{4\mu_x\mu_y\sigma_{xy}}{(\mu_x^2 + \mu_y^2)(\sigma_x^2 + \sigma_y^2)} \tag{2}$$

where

$$\mu_x = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{ij}, \quad \mu_y = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_{ij}, \tag{3}$$

$$\sigma_x^2 = \frac{1}{N^2-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_{ij}-\mu_x)^2, \quad \sigma_y^2 = \frac{1}{N^2-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(y_{ij}-\mu_y)^2, \tag{4}$$

$$\sigma_{xy} = \frac{1}{N^2-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_{ij}-\mu_x)(y_{ij}-\mu_y). \tag{5}$$

assuming that x and y denote the respective pixels of original and distorted images.

Considering the possibility of obtaining unstable results because of the division by the values close to zero for flat and/or dark fragments of the image, the additional stability constants have been introduced as the extension of the metric into the Structural Similarity index. Their only role is the prevention against division by zero so their values should be relatively small in order to avoid any noticeable changes of the results for the other fragments of the image. The default values of both coefficients proposed in the paper [3] are equal to $C1 = 0.01 \cdot L$ and $C2 = 0.03 \cdot L$

where L denotes the dynamic range (number of luminance levels for the grayscale images) and the final formula can be expressed in the following form:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{6}$$

Both metrics are applied using the sliding window approach so the quality map of the image is obtained as the result. The overall quality index is calculated as the mean value of the whole map. Another difference between both metrics is the type and size of the window: 8×8 rectangular window for the UIQI and 11×11 pixels Gaussian one for the SSIM. Nevertheless, both methods are sensitive to the three common types of distortions: the loss of correlation, luminance distortion and the loss of contrast.

Regardless of much better correlation of the SSIM index with the subjective evaluation in comparison to the classical metrics, its definition does not comply with the color information since it is defined for the single channel. Independent application of the metric for three RGB channels usually leads to varying results so the proper choice of the color model seems to be an important element of the color image quality assessment process.

In this paper the application of the colour models recommended by the CIE (*Commission Internationale de l'Eclairage*) [5] is investigated. The color models used in the experiments are supported by MATLAB's Image Processing Toolbox [6]:

- **XYZ** (the original, 1931 CIE colour space specification),
- **xyY** (CIE specification that provides normalized chromaticity values; the capital Y value represents luminance and is the same as in XYZ),
- **uvL (uvY)** (CIE specification that attempts to make the chromaticity plane more visually uniform, where L is the luminance and is the same as Y in XYZ),
- **u'v'L (u'v'Y)** (CIE specification in which u and v are rescaled to improve uniformity),
- **L*a*b*** (CIE specification that attempts to make the luminance scale more perceptually uniform, where L* is a nonlinear scaling of L, normalized to a reference white point),
- **L*ch** (CIE specification where c is chroma and h is hue; these values are the polar coordinate conversion of a* and b* in L*a*b*).

3   Details of the experiments and results

In order to verify the usefulness of the CIE colour models, the SSIM metric has been applied for each of the channels in all colour models. All the calculations have been performed for all images present in the LIVE database excluding the original ones. Finally, the Pearson linear correlation coefficients between the DMOS and each modified SSIM metric have been computed for each type of distortion as well as for the whole database (excluding original images). The best correlation results have been obtained for the luminance channels, the detailed results are presented in Table 1.

The illustration of the correlation relevance is presented in the scatter plots (Fig. 1–4) obtained for the whole LIVE database excluding the original images for the CIE L*a*b* channels and the NCD metric (much wider distribution of respective points in comparison to the usage of the CIE L* luminance can be easily observed).

Table 1. Linear correlation coefficients between DMOS and the SSIM values obtained for the CIE colour models in comparison to the NCD.

| Colour model	XYZ	xyY	uvY	u'v'Y	L*a*b*	L*ch	NCD
	X	x	u	u'	L*	L*	
JPEG	0.8809	0.5814	0.4942	0.4942	0.8948	0.8948	0.6747
JPEG2000	0.8468	0.5383	0.4871	0.4871	0.8482	0.8482	0.6597
White noise	0.9649	0.9273	0.9372	0.9372	0.9606	0.9606	0.8462
Gaussian blur	0.8174	0.5856	0.4648	0.4648	0.8511	0.8511	0.5853
Fast fading	0.8760	0.5587	0.4706	0.4706	0.9024	0.9024	0.7444
All	**0.6920**	0.3174	0.2880	0.2880	**0.7266**	**0.7266**	0.5213
	Y	y	v	v'	a*	c	
JPEG	0.8787	0.4318	0.3709	0.3968	0.4688	0.6631	
JPEG2000	0.8428	0.4296	0.4350	0.4352	0.4646	0.7859	
White noise	0.9599	0.9062	0.9506	0.9420	0.9134	0.9181	
Gaussian blur	0.8196	0.4260	0.2930	0.3478	0.5907	0.7068	
Fast fading	0.8780	0.4499	0.3206	0.3584	0.4828	0.5813	
All	0.6747	0.2751	0.2905	0.2806	0.2453	0.4375	
	Z	Y	Y	Y	b*	h	
JPEG	0.8325	0.8787	0.8787	0.8787	0.5610	0.3808	
JPEG2000	0.8214	0.8428	0.8428	0.8428	0.5082	0.5276	
White noise	0.9568	0.9599	0.9599	0.9599	0.9172	0.7702	
Gaussian blur	0.7668	0.8196	0.8196	0.8196	0.6565	0.5969	
Fast fading	0.8355	0.8780	0.8780	0.8780	0.5196	0.5067	
All	0.6177	0.6747	0.6747	0.6747	0.2735	0.3246	
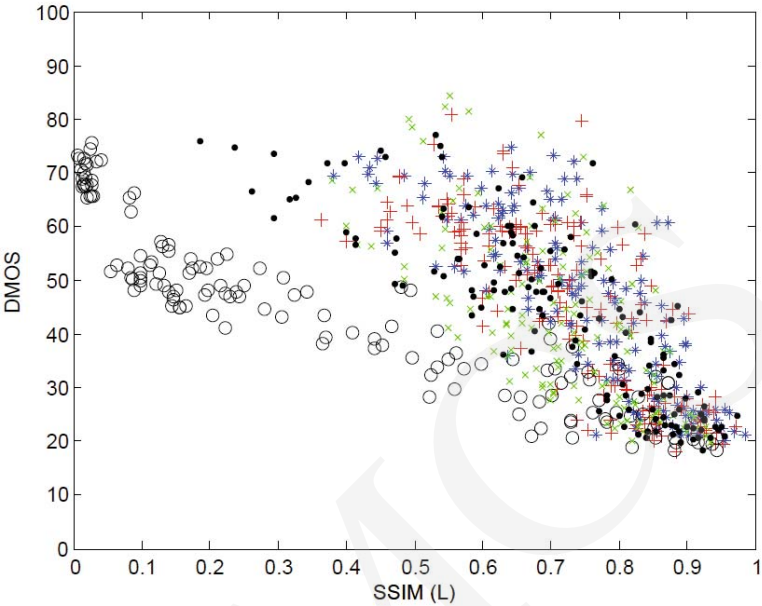
Fig. 1. Scatter plot of the DMOS vs. the SSIM values for the CIE L*a*b* luminance channel.
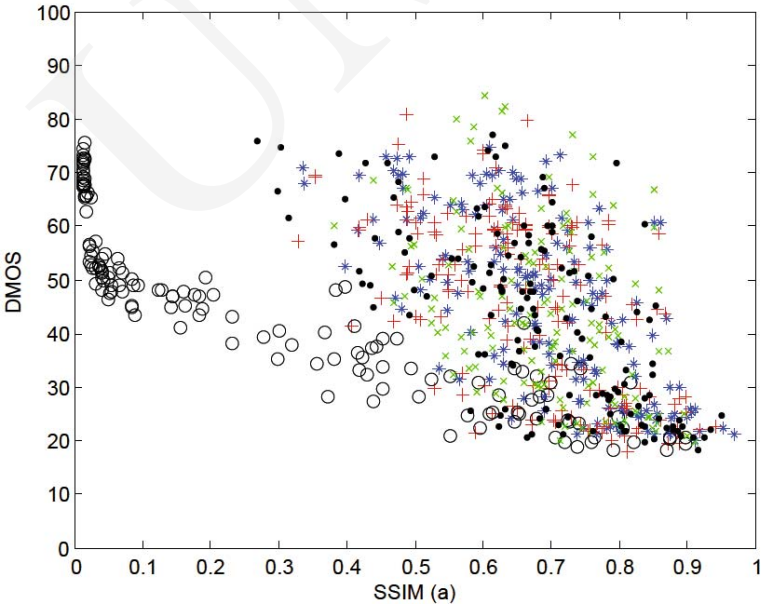

Fig. 2. Scatter plot of the DMOS vs. the SSIM values for the a* channel.

The obtained correlation results are relatively small because of the linearity assumption. Taking into account the nonlinear character of the Human Visual System, the nonlinear
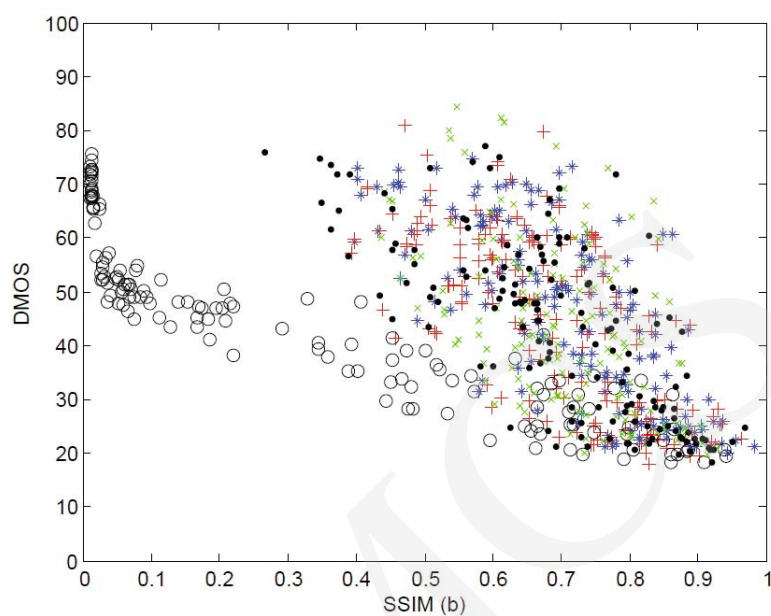
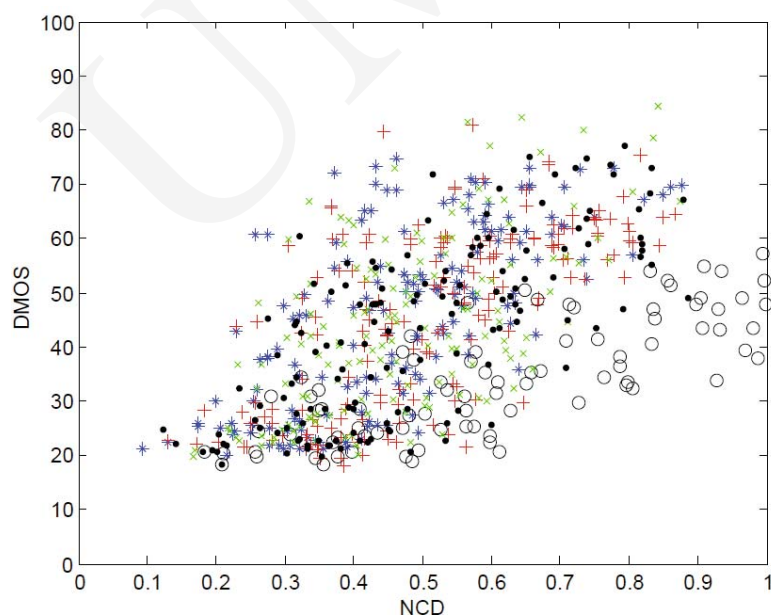Fig. 3. Scatter plot of the DMOS vs. the SSIM values for the b* channel.

Fig. 4. Scatter plot of the DMOS vs. the NCD values [7].

combination of channels is proposed based on the SSIM index extended by the use of the exponents for each channel. The weighted SSIM index can be expressed as

$$WSSIM = (SSIM_{channel1})^A \cdot (SSIM_{channel2})^B \cdot (SSIM_{channel3})^C, \qquad (7)$$

The exponents related to the chrominance channels obtained after the optimization are equal to zero, so the most relevant channels are the luminance ones. The best results can be obtained for the combination of two luminance channels from two different colour models as shown in Table 2. The respective scatter plot obtained for the proposed metric is shown in Fig. 5.
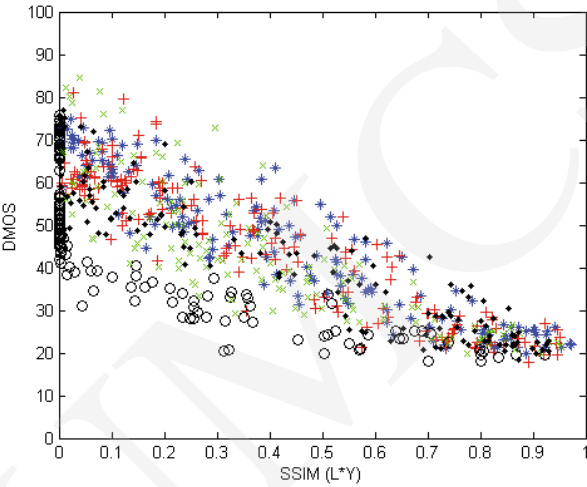

Fig. 5. Scatter plot of the DMOS vs. the proposed metric.

Table 2. Pearson correlation coefficients between the DMOS and the modified SSIM values for the optimized exponents using the CIE luminance channels.

Channel	L*	Y	L*Y
exponent	4.88	6.07	4.33 / 0.67
JPEG	0.9300	0.9317	0.9328
JPEG2000	0.9218	0.9191	0.9234
White noise	0.8188	0.7953	0.8191
Gaussian blur	0.8615	0.8433	0.8609
Fast fading	0.9340	0.9268	0.9342
All	**0.8491**	**0.8377**	**0.8493**

4   Conclusions

Image quality assessment of the colour images is not the simple extension of the methods known for the grayscale images. Nevertheless, the proposed metric can be an impulse for

further research as the obtained results illustrate the relevance of the problem. The experiments were limited by the number of images and scores present in the LIVE database as well as the limited number of distortion types. Finally,the obtained results are the compromise solution because one can notice that some types of distortions can influence on the overall correlation much more than other ones.

Analyzing the scatter plot shown in Fig. 5, one can observe that the price for the good correlation of the four types of distortions is the reduced correlation of the fifth type (white Gaussian noise). The reason is different character of that type of distortion and the problem is typical of all universal approaches to the image quality assessment.

Future work should provide even better correlation with the DMOS values with the possibility of utilizing other colour spaces. To this end future research should utilize the nonlinear combination of such channels as well as possible combination of various window types and sizes used during the SSIM calculations for different channels.

References

[1] Eskicioglu A., Fisher P., Chen S., Image quality measures and their performance, IEEE Transactions on Communications 43(12) (1995): 2959–2965.

[2] Sheikh H. R., Wang Z., Cormack L., Bovik A. C., LIVE Image Quality Assessment Database (release 2). http://live.ece.utexas.edu/research/quality.

[3] Wang Z., Bovik A., Sheikh H., Simoncelli E., Image quality assessment: from error measurement to structural Similarity, IEEE Transactions on Image Processing 13(4) (2004): 600–612.

[4] Wang, Z., Bovik, A., A universal image quality index, IEEE Signal Processing Letters 9(3) (2002): 81–84.

[5] CIE Recommendations on Uniform Color Spaces, Color-difference Equations and Psychometric Color Terms, Supplement 2 to Publication CIE 15, Colorimetry (E-1.3.1) 1971 (Bureau Central de la CIE, Paris, 1978).

[6] Matlab Image Processing Toolbox 6.2 Documentation, MathWorksTM.

[7] Okarma, K., Colour image quality assessment using structural similarity index and singular value decomposition, Lecture Notes in Computer Science (Proc. Int. Conf. Computer Vision and Graphics ICCVG2008) (in printing), (Springer Verlag, 2009).