COLOR IMAGE QUALITY EVALUATION USING GRAYSCALE METRICS IN CIELAB COLOR SPACE

Renata Caminha C. Souza, Lisandro Lovisolo recaminha@gmail.com, lisandro@uerj.br

PROSAICO (Processamento de Sinais, Aplicações Inteligentes e Comunicações) – PEL, University of the State of Rio de Janeiro

ABSTRACT

In many image applications the measurement of visual quality is of special importance. Most of the Full-Reference metrics proposed address the evaluation of grayscale images. This paper investigates an application of two grayscale metrics for the evaluation of color images using CIELAB color space. It also proposes the utilization of an important feature of this space, the possibility to measure the distance between colors, to improve the performance of the metric, i.e. affording a higher correlation between the objective metric and the subjective measurement. The results indicate that the use of color distance with a FR image quality evaluation metric improves its correlation with the subjective scores.

Index Terms— Image quality evaluation, Color images, Full-Reference metrics, SSIM, VIF, Just Noticeable Difference, CIELAB.

1. INTRODUCTION

Digital images exchanged and distributed through communication systems are subject to different types of distortion during acquisition, processing, compression, transmission and reproduction. For example, distortion can be caused by data transfer errors (due to inherent faulty channels as wireless channels), and lossy techniques for image compression.

The most reliable way to assess the image quality is using subjective evaluation, once human observers are the ultimate receivers in most applications. The Mean Opinion Score (MOS), a subjective quality measurement obtained from a number of observers, is a reliable and widely used method for subjective quality evaluation [1]. However, for most applications, this method is inconvenient as it cannot be used in real-time.

To manage this problem, many objective quality assessment algorithms, especially regarding grayscale images, have been investigated [2]-[4]. The closer to the subjective image quality assessment an objective image quality assessment is, the better is the metric. These objective metrics are generally classified into three categories based on the amount of information required from the original image [2]: Full-reference (FR), Noreference (NR) and Reduced-reference (RR) metrics.

Full-reference metrics perform a comparison between the whole reference image (the original one) and the whole distorted image (the processed one), and therefore require the reference to be completely available. No-Reference metrics analyze the processed image all alone without the need of any information from the reference, and always need to make some assumptions about the content of the image or about the existent distortions on it. Reduced-Reference metrics are designed as a tradeoff between FR and NR metrics. They extract attributes from the original image in a way that a comparison to the processed image can be made based on these attributes. [2].

Historically, statistic metrics like MSE (Mean Squared Error) and PSNR (Peak Signal to Noise Ratio) have been largely used, and still are used today because of their simplicity. However, despite of their wide use, their results do not correlate well with the human perception [3].

This article analyses the performance of two FR metrics, originally designed for the evaluation of grayscale images, applied for the evaluation of color images: SSIM (Structural SIMilarity) [5] and VIF (Visual Information Fidelity) [6].

In order to accomplish this goal, the CIELAB color space is used. This specific color space was chosen because of a particular feature: it permits the calculation of the distance between colors as a good approximation of what is perceived by the human vision [7]. The difference threshold of perception between two colors is known as Just Noticeable Difference (JND). This concept is used in order to improve the correlation between objective metrics and subjective scores.

For the tests two image databases with objective scores were used: IVC database [8] constituted by 10 reference images and 120 distorted versions and LIVE database [9], constituted by 29 reference images and 779 distorted versions.

The remainder of this article is organized as follows: Section 2 briefly discusses the grayscale metrics SSIM and VIF used in the experiments. Section 3 describes the CIELAB color space and the JND concept. The details on the experiments are described in Section 4. The results are presented in Section 5 and the conclusions are given in Section 6.

2. GRAYSCALE METRICS

In order to test the use of grayscale metrics in the quality evaluation of color images, two different FR grayscale metrics were chosen, that are SSIM and VIF.

2.1 SSIM Index

The SSIM Index [5] is a FR image quality metric intended to capture the loss of image structure. SSIM was derived by considering hypothetically that one could capture image quality with three aspects of information loss that are complementary to each other: correlation distortion, contrast distortion, and luminance distortion [10].

The basic form of SSIM is computed as follows. Suppose that x and y are patches at the same position from two images that are being compared, the local SSIM index measures three elements in the patch: the similarity -l(x, y) – between the luminances of the patches, the similarity – c(x, y) – between the contrasts of the patches and the similarity – s(x, y) – between the structures of the patches. These similarities are expressed through statistics, computed and combined in a way to produce the local SSIM:

$$S(\mathbf{x}, \mathbf{y}) = l(\mathbf{x}, \mathbf{y}) \cdot c(\mathbf{x}, \mathbf{y}) \cdot s(\mathbf{x}, \mathbf{y})$$

= $\left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right) \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right) \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right).$ (1)

In (1) μ_x and μ_y are the mean from the patches x and y respectively, σ_x and σ_y are their standard deviations, and σ_{xy} is the cross-correlation between the patches x e y after subtracting their means. C_1 , C_2 and C_3 are small positive constants that stabilize each term. Specifically in [11], were chosen $C_1 = (K_1L)^2$ where *L* is the dynamic range for the values of the pixel (255 for 8 bits images), and $K_1 \ll 1$ is a small constant. A similar definition was used for C_2 . Also; it was assumed that $C_3 = C_2/2$, what leads to a specific form of SSIM [11]:

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$
 (2)

2.2 VIF Criterion

The Visual Information Fidelity (VIF) criterion [6] is a perceptual FR image quality metric developed by LIVE team.

The metric is based on the quantification of information shared between the reference and distorted images relative to the information present in the reference image. For that, it uses a Natural Scenes Statistics (NSS) model, an image degradation model and an HVS model, all of them in the wavelet domain [6].

The NSS model used is the Gaussian Scale Mixture (GMS) model, which is a Random Field (RF). The NSS

models each subband in the wavelet decomposition of the image with a separate GSM and is expressed as

$$\mathcal{C} = \mathcal{S} \cdot \mathcal{U} = \{S_i \cdot \vec{U}_i : i \in I\},\tag{3}$$

where $C = \{\vec{C}_i : i \in I\}$ is a GSM, I denotes the set of spatial indices for the RF, $S = \{S_i : i \in I\}$ is an RF of positive scalars, $U = \{\vec{U}_i : i \in I\}$ is a Gaussian vector RF with zero mean and covariance $C_U \cdot \vec{C}_i$ and \vec{U}_i are M dimensional vectors. The subbands are divided in nonoverlapping blocks of M coefficients each, assuming each block to be independent of others.

The distortion model is a signal gain and additive noise model:

$$\mathcal{D} = \mathcal{GC} + \mathcal{V} = \{g_i \cdot \vec{C}_i + \vec{V}_i : i \in I\},\tag{4}$$

where C denotes the RF from a subband in the reference signal, $\mathcal{D} = \{\vec{D}_i : i \in I\}$ denotes the RF from the corresponding subband from the distorted signal, $\mathcal{G} =$ $\{g_i : i \in I\}$ is a deterministic scalar attenuation field and $\mathcal{V} = \{\vec{V}_i : i \in I\}$ is a stationary additive zero-mean Gaussian noise RF with variance $C_V = \sigma_v^2 \mathbf{I}$. The RF \mathcal{V} is white and is independent of S and \mathcal{U} .

To model the HVS, the internal neural model is used, and it is represented by an additive white Gaussian noise model. The neural noise is modeled as the RF $\mathcal{N} = \{\vec{N}_i : i \in I\}$, where \vec{N}_i are zero-mean uncorrelated multivariate Gaussian with the same dimensionality as \vec{C}_i :

$$\mathcal{E} = \mathcal{C} + \mathcal{N} \tag{5}$$

 $\mathcal{F} = \mathcal{D} + \mathcal{N} \tag{6}$

where \mathcal{E} and \mathcal{F} denote the visual signal at the output of the HVS model from the reference and the test images. The covariance of the additive noise is modeled as $C_N = \sigma_n^2 \mathbf{I}$, where σ_n^2 is an HVS model parameter.

With the source and distortion models described, for the calculation of the Visual Information Fidelity let $\vec{C}^N = (\vec{C}_1, \vec{C}_2, ..., \vec{C}_N)$ denote N elements from C and $S^N, \vec{D}^N, \vec{E}^N$ and \vec{F}^N be correspondingly defined. Assuming that the model parameters \mathcal{G}, σ_n^2 and σ_v^2 are known, the conditional mutual information between C, \mathcal{E} (or \mathcal{F}) given \mathcal{S} is analysed.

For the reference image it is analysed $I(\vec{C}^N; \vec{E}^N | S^N = s^N)$, where s^N denotes a realization of S^N . Denoting $I(\vec{C}^N; \vec{E}^N | S^N = s^N)$ as $I(\vec{C}^N; \vec{E}^N | s^N)$, $I(\vec{C}^N; \vec{E}^N | s^N)$ and $I(\vec{C}^N; \vec{F}^N | s^N)$ represent the information that can ideally be extracted from a particular subband in the reference and the test images respectively. The VIF measure is simply the fraction of the reference image information that can be extracted from the test image given by:



Figure 1 – Implementation of color distance together with objective evaluation.

$$\text{VIF} = \frac{\sum_{j \in subbands} I(\vec{C}^{N,j}; \vec{F}^{N,j} | s^{N,j})}{\sum_{j \in subbands} I(\vec{C}^{N,j}; \vec{E}^{N,j} | s^{N,j})},$$
(7)

where the summation occurs over the subbands of interest, and $\vec{C}^{N,j}$ represent N elements of the RF C_j that describes the coefficients from subband *j* and so on.

Grayscale images can be represented by a unique matrix or channel, while color images require three or four channels to be represented, depending on the color space used. However, the choice of an adequate color space is critical in order to produce results well-correlated with the human perception. The next section presents CIELAB, the color space used in the experiments of this work, and presents as well the application of its special feature, the measurement of the differences between colors, in the evaluation of color image quality.

3. CIELAB COLOR SPACE AND JND CALCULATION

The CIELAB color space [12], was defined focusing a measure of the difference between colors that could be perceptually uniform. This color space was established by CIE based on the MacAdam's ellipses theory [13].

The area inside each MacAdam's ellipse defined in the XYZ chromaticity diagram includes all the colors visually identical to the color present in the center of the ellipse [14].

The threshold of the MacAdam's ellipses is known as *just noticeable difference* (JND). The JND concept was brought to CIELAB color space in a way that the Euclidean distance between the coordinates that represents two different colors in this space gives an approximation of the difference perceived by the human vision between the two colors. This distance is also known as DeltaE.

Each one of the three coordinates of the CIELAB color space -L, a and b - represent, respectively the color luminance, the position between red/magenta and green and the position between yellow and blue.

To calculate the Lab coordinates from the RGB color space one does [12]:

$$L^{*} = 116f(Y/Y_{n}) - 16$$

$$a^{*} = 500[f(X/X_{n}) - f(Y/Y_{n})]$$

$$b^{*} = 200[f(Y/Y_{n}) - f(Z/Z_{n})]$$
(8)

where

$$f(t) = \begin{cases} t^{1/3} & \text{for } t > (6/29)^2 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + 4/29 & \text{otherwise.} \end{cases}$$
(9)

In (9) X_n , Y_n , Z_n , are the CIE XYZ tristimulus values of the white reference and *X*,*Y*,*Z* are related to the RGB color space though the following equation [12]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$
(10)

The difference DeltaE between two colors measured in the CIELAB color space is given by:

$$DeltaE = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2}, \quad (11)$$

where (L_1, a_1, b_1) and (L_2, a_2, b_2) are two different colors in the CIELAB color space.

A value for DeltaE that is below a given JND value indicates that the difference between the colors is not perceptible by the human eye [15]. The JND value ranges varies depending on the application and from person to person. For instance, we can cite [7] where there are references for the JND value that vary from 0.38 to 5.6, depending on the application.

In this work we propose to use DeltaE and JND in the image quality evaluation using grayscale metrics in order to improve the correlation with the subjective scores, as described as follows

Proposal to join JND with FR image quality metrics

A way to use DeltaE in the image quality evaluation is shown in the scheme of Figure 1. First, both original and distorted images are converted to CIELAB color space.



Figure 2 - Correlation between the subjective and objective metrics when considering JND.

Then, a map of color differences between the converted images is calculated. From this map, regions where the color difference is below the JND are identified, and in the distorted image these regions are replaced for the corresponding regions in the original image. Finally, the objective metric is calculated between the original and the modified distorted images to produce the objective evaluation.

4. EXPERIMENTS

To evaluate the performance of objective quality measurement metrics, a subjective database is needed. In order to improve the reliability of the results, two databases are used: IVC database [8], containing 10 reference images and 120 distorted versions, and LIVE database [9], containing 29 reference images and 779 distorted versions.

For both image databases, the calculation of the objective metrics considers only component L of the images. The calculation of the color distance considers all components - L, a and b.

In order to find the best JND value, i. e., the value that afford the best correlation between objective metric and subjective measurement, a range of JND values between 0 and 6 is tested in steps of 0.2.

To verify the accuracy of objective metrics a common practice is to compare the objective results to the subjective measurements given by the image databases. For that, the Pearson correlation is calculated between the objective and subjective measurements.

Also, in order to verify the consistence of the results, besides the correlation considering the set of all databases' images, the correlation for subsets of the databases, chosen in a random manner, were calculated.

5. RESULTS

Figure 2 presents the results obtained for the correlation between objective metrics (SSIM and VIF) and subjective ones for the two databases tested (IVC and LIVE) for JND varying from 0 to 6 in steps of 0.2. For that there are four graphics.

At each graphic, there is a thick line and a set of thin lines. The thick line represents the correlation considering the set of all images from the database whereas the thin lines are a mean for cross-validation, and represent the correlation for random subsets of the database, calculated in order to evaluate the consistence of the results.

From the graphics, it can be seen that the correlation peak occurs at a JND value bigger than zero. The correlations obtained for SSIM have their peak at JND=2.8 using IVC database and at JND=1.2 using LIVE database. The VIF also obtained a gain by using JND although small. For the IVC database the correlation peak occurred at JND=2.8, the same JND value obtained for SSIM, and for the LIVE database the peak occurred at JND=0.4. More important than the peaks themselves are the shapes that the correlation curves assume, that follow a consistent pattern. Therefore, the use of the JND concept affords a gain in the correlation between the objective and subjective metrics.

Table 1 presents the improvement in the correlation between each objective metric for each database, compared to the correlation without using the JND concept, i.e. the correlation for JND=0. SSIM has improved 7.220% using IVC database and 0.599% using LIVE database. VIF has showed an improvement of 0.456% using IVC database and of 0.003% using LIVE database.

From these analyses, it can be seen that the use of JND improves the objective evaluation provided by SSIM and VIF.

		SSIM		VIF	
	JND	0	2,8	0	2,8
IVC	Correlation	0,82219	0,88155	0,84240	0,84625

7,220%

1.2

0,86843

0,599%

0

0,91862

0,456%

0.4

0,91864

0,003%

Improvement

regarding JND=0

JND

Correlation

Improvement

regarding JND=0

LIVE

Table 1 - Comparison between correlations using JND for SSIM and VIF

6. CONCLUSIONS

0

0,86326

From this work it can be seen that CIELAB color space is a good choice to be used in the ambit of image quality evaluation using grayscale metrics. It can also be noted that the use of JND concept with FR image quality evaluation metrics can propitiate a gain in the final result.

It was observed that the gain in the correlation was more noticed using IVC database than using LIVE database. A possible reason for that is that both grayscale metrics used in the tests, SSIM and VIF, were developed by the same team that developed LIVE database. Therefore, the parameters in both metrics could have been adjusted to accomplish the best results using this database.

The gain in the correlation was more evident for the SSIM, but also VIF presented a small gain, that is visually more perceived using the IVC database. However, from the results it was observed that gains are attained for both databases. That indicates that the use of JND concept affords an improvement in the calculation of the objective metric.

7. REFERENCES

Z. Wang, H. R. Sheikh, A. C. Bovik, *The Handbook of Video Databases: Design and Applications*, 2nd ed., Boca Raton, FL, USA. CRC Press, 2003, ch. "Objective video quality assessment", pp. 1041-1078.

[2] H. R. Wu and K. R. Rao, Digital Video Image Quality and Perceptual Coding. Boca Raton, FL, United States of America: CRC Press, 2006.

[3] A. Toet, M. P. Lucassen, "A new universal color image fidelity metric," *Displays*, vol. 24, no. 4-5, pp. 197–207, Dec. 2003.

[4] S. Winkler. Digital Video Quality: Vision Models and Metrics. Chichester, England. John Wiley & Sons, 2005.

[5] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[6] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[7] M. Melgosa, M. M. Pérez, A. Yebra, R. Huertas and E. Hita, "Some reflections and recent international recommendations on color-difference evaluation," *Óptica Pura y Aplicada*, vol. 34, 2001.

[8] P. Le Callet and F. Autrusseau. (2005) Subjective quality assessment IRCCyN/IVC database. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/

[9] H. R. Sheikh, et al. LIVE Image Quality Assessment Database Release 2. 2009. [Online]. Available: http://live.ece.utexas.edu/research/quality

- [10] Z. Wang and A. C. Bovik, "Mean Squared Error: Love it or Leave it? – A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no.1, pp. 98–117, Jan. 2009.
- [11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600– 612, Apr. 2004.
- [12] J. Schanda, *Colorimetry: Understanding the CIE System*, Wiley-Interscience, 2007.
- [13] G. Wyszecki and W. S. Stiles, Color Science: Concepts and Methods, quantitative Data and Formulae, 2nd ed., New York, United States of America: John Wiley & Sons, 1982.
- [14] D. L. MacAdam, "Specification of small chromaticity differences in daylight," *Journal of the Optical Society of America*, vol. 33, no. 1, Jan. 1943.
- [15] M. Melgosa, "Color-Difference Formulas," in *Balkan Light* 2008, 2008.