Perceptual Quality Metric for Digitally Coded Color Images

Christian J. van den Branden Lambrecht
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
vdb@lts.de.epfl.ch
http://ltswww.epfl.ch/~vdb/

Joyce E. Farrell
Imaging Technology Department
Hewlett-Packard Laboratories
1501 Page Mill Road MS 1U20
Palo Alto, CA 94304
Farrell@hpl.hp.com

ABSTRACT

In this paper, a computational metric that incorporates many aspects of human vision and color perception to predict the quality of color coded images is presented. The proposed distortion measure is built on opponent-colors theory and on a multi-channel model of spatial vision. The metric has been validated by psychophysical data on 400 images and two human observers.

1 INTRODUCTION

Image quality assessment is an important and yet unsolved problem in image processing. For years, engineers have been designing and testing devices using the peak signal-to-noise ratio (PSNR) despite the fact that it is rarely correlated with human judgment. The inherent problem with the PSNR is that it does not take into account viewing conditions and spatial pattern sensitivity. The most recently published computational metrics for image quality assessment have incorporated vision models for spatial pattern sensitivity and visual masking [1–3, 10]. This paper extends one of these metrics [1] to include color by proposing a comprehensive color image quality metric for digitally encoded images.

The metric is based on the opponent-colors theory to model color perception and on a multi-channel model of spatial vision to account for pattern sensitivity. Visual masking is incorporated in the computation to eventually yield a distortion measure that predicts image quality better than the PSNR. The metric has been tested on a set of 400 color images for which psychophysical data had been collected with two human observers. The paper is structured as follows: the metric is described in Sec. 2. Section 3 presents the psychophysical material that has been used to validate the metric and compares the performance of the proposed distortion measure with the mean square error. Sec. 4 summarizes the results and makes some final conclusions.

2 THE METRIC

The computational steps of the metric are illustrated in Fig. 1 and can be described as so: consider two images in some standard color space, RGB for example. One picture is an original and the other is a distorted version of the original. The color image quality metric is used to predict the ability of people to detect the difference between the original and the distorted image.

The first step in the computation of the metric is to transform the RGB values in the digital images into RGB values that are linear with luminance, i.e. linear RGB values. The second step converts the linear RGB values into coordinate values in an opponent-colors space. The three coordinates of this color-opponent space correspond to luminance (B/W), red-green (R/G), and blue-yellow (B/Y) channels which are thought to mediate color perception [11].

In the next step of the computation, each color component of the original and error images (B/W, R/G, and B/Y) is analyzed by a filter bank created by Gabor functions tuned in spatial frequency and orientation. The portion of the frequency plane covered by a filter is denoted "channel". Such modeling is fairly common [1, 13] and accounts for the fact that the cortex analyzes visual information in this way.

In the fourth step of the computation, a detection threshold is computed for contrast based on contrast sensitivity and masking phenomenon. This accounts for both detection of single stimulus and inter-stimuli interferences in a channel. The error signal in each channel is multiplied by the inverse of that detection threshold, resulting in a quantity expressed in *units above threshold* also referred to as *just noticeable differences* or jnd's. Once the filtered error signal has been masked, a distortion metric is computed by pooling the error over the channels. Like PSNR, the measure can then be expressed on a logarithmic scale. It has been termed Color Masked Signal to Noise Ratio (CMPSNR)

Some of the building blocks of the metric are now described with more details.

2.1 Opponent-Colors Space

The opponent-colors theory defines a color space for which the principal coordinates are perceptually orthogonal. The three coordinates of this space correspond to luminance (B/W), red-green (R/G), and blue-yellow

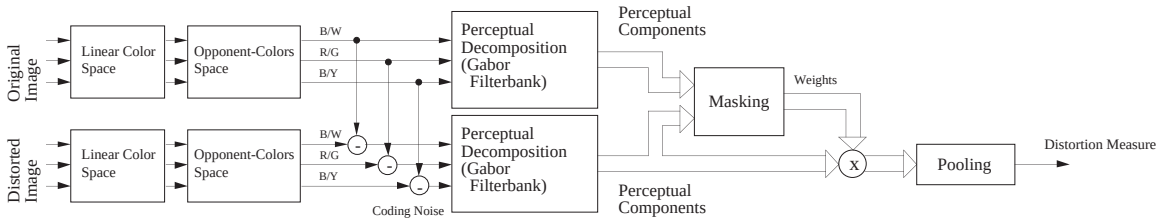

Figure 1: General block diagram of the color quality metric.

(B/Y) channels. The opponent-colors space chosen here has been developed by Poirson and Wandell who measured color appearance and derived a pattern-color separable model [7, 8, 11]. The sensitivity of the three components as a function of the wavelength are depicted in Fig. 2.

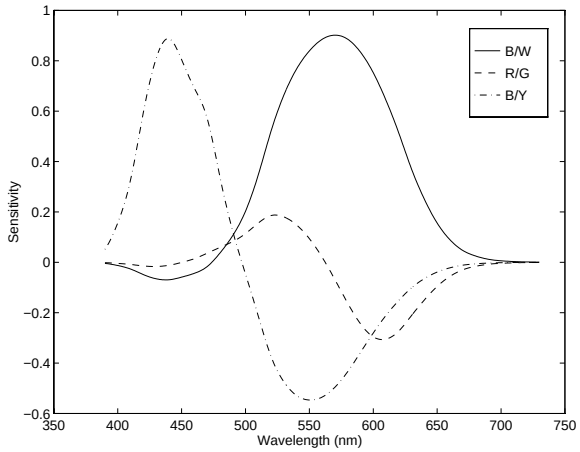

Figure 2: Spectral sensitivity of the Wandell-Poirson pattern-separable opponent-colors space. The solid line is the luminance channel (termed B/W), the dashed line is the red-green channel (termed R/G) and the dot-dashed line is the blue-yellow channel (termed B/Y).

2.2 Channel Decomposition

The images are decomposed by a filterbank in a multiresolution scheme. The filterbank being used here decomposes the image into frequency and orientation bands. One of the filters, centered on the null frequency, is isotropic and has a bandwidth of 2 cycles per degree (cpd). The other filters split the image into 4 frequency bands, centered in 2, 4, 8 and 16 cpd, each being one octave wide and 4 orientation bands (centered 0, $\pi/4$, $\pi/2$ and $3\pi/4$) with a bandwidth of $\pi/4$. This makes a total of 17 filters. The filterbank is illustrated in Fig. 3. Only the B/W signal is decomposed into 17 components. Due to the much lower sensitivity for the red-green and blue-yellow wavelengths, the corresponding components are decomposed using only 9 filters, stopping the frequency decomposition at 8 cpd [13].

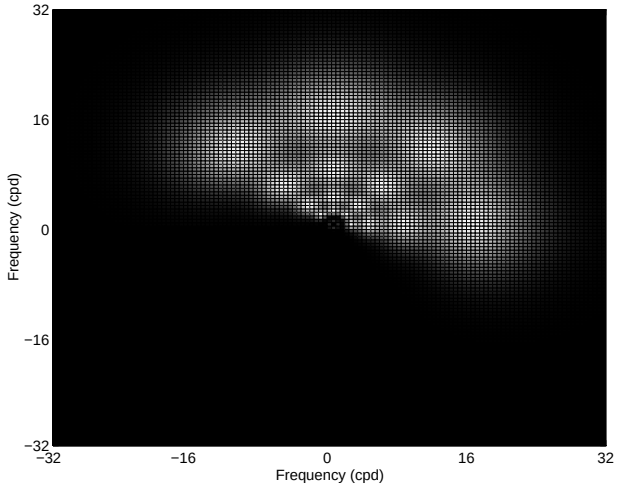

Figure 3: The spatial filter bank, featuring 17 filters (5 spatial frequencies and 4 orientations). The magnitude of the frequency response of the filters are plotted on the frequency plane. The lowest frequency filter is isotropic.

2.3 Pattern Sensitivity

Once the images have been decomposed into perceptual channels, pattern sensitivity of the coding noise can be assessed. Two phenomena are modeled here. The first is contrast sensitivity and the second is visual masking. Human sensitivity to contrast is dependent on frequency and this defines a contrast sensitivity function that permits to predict the visibility of a single stimulus. Masking accounts for the interferences between stimuli: part of the error will be masked by the background (i.e. the original image) and this will influence the perception of the noise. The model of masking that is being used here is the non-linear transducer [5] depicted in Fig. 4. The detection threshold C_T for a stimulus can then be computed as a function of the detection threshold of that stimulus in the absence of a masker, C_{T_0} (i.e. as given by the contrast sensitivity function) and the contrast of

the masker C_M. The relationship is given by Eq. (1):

$$C_T = \begin{cases} C_{T0} & \text{if } C_M < C_{T0} \\ C_{T0}\left(\frac{C_M}{C_{T0}}\right)^{\varepsilon} & \text{if } C_M \geq C_{T0} \end{cases} . \quad (1)$$

The masked signal is then divided by C_T so as to be expressed in units above threshold.

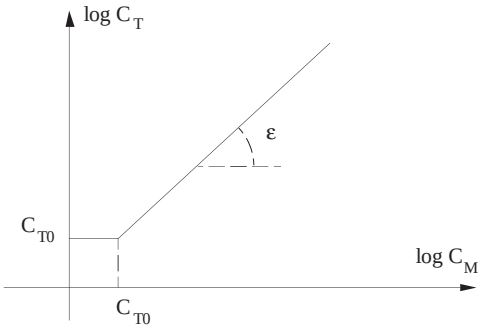

Figure 4: Non linear transducer model of masking.

2.4 Summation

Masking the output of the perceptual channels predicts the response from the cells of area V1 of the cortex. The data has then to be gathered together to account for higher levels of perception, which is termed *pooling*. A first phenomenon that one should take into account is that an observer concentrates on regions rather than on a complete image. This is due to both the focus of attention and the viewing distance. To incorporate this into the metric, the image is divided into blocks. The dimension of the block is chosen so that a block covers two degrees of visual angle, which is the dimension of the fovea (and thus of the region onto which the attention is focused). The distortion measure is computed for each block by pooling the error over the channels. Basically, the magnitude of the channels' output are combined by Minkowski summation with a higher exponent to weight the higher distortions more. The actual computation of the distortion E for a given block is computed according to Eq. (2):

$$E = \left(\frac{1}{N} \sum_{c=1}^{N} \left(\frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |e[x,y,c]| \right)^{\beta} \right)^{\frac{1}{\beta}} , \quad (2)$$

where $e[x,y,c]$ is the masked error signal at position (x,y) in the current block and in the channel c; N_x and N_y are the horizontal and vertical dimensions of the blocks; N is the number of channels. The exponent of the Minkowski summation is β and has a value of 4, which is close to probability summation. This value has shown to be a best choice [3].

3 VALIDATION

The metric has been validated by comparing its ability to predict psychophysical data collected from two subjects who participated in a discrimination experiment for an original color image and 400 distorted versions of it. The original image features a boy surrounded by colorful balls. The distorted images have been generated in the following way: The original RGB image was converted to the YUV color space and compressed using the JPEG algorithm [4]. The coding noise has been determined by subtracting the original to the distorted image. The test images were created by adding scaled version of the coding noise to the original image. Twenty classes of images have been created this way, each of them characterized by a distortion in one direction (for example in U or Y+U or Y-V). Each class is subdivided into twenty images, each image being characterized by a scale factor for the noise.

Psychophysical data was collected for two subjects and for each class of images by a QUEST procedure [12]. In the experiments, the subjects were presented the original image with two test images. One of the test image was the original and the other a distorted image. The subjects' task was to identify the distorted image. The percentage of correct answers has been measured for each image. Such data can be modeled by the psychometric function [9]:

$$P(C) = \frac{1}{2}\left(1 + \frac{1}{1 + e^{\frac{-(x-M)}{S}}}\right), \quad (3)$$

where $P(C)$ is the probability of correct answer, x the strength of the stimulus, M the midpoint and S the spread of the psychometric function. Two curves have been fit to each class of data using for one curve the CMPSNR values and for the other the MSE values. The fitting procedure uses the output of the metric as values for the variable x in Eq. (3) and estimates the parameters M and S by a maximum likelihood estimation [6]. Such a result is depicted, for one block of trial, in Fig. 5. The solid line is the fit obtained with the perceptual metric and the dashed line is the fit obtained with the mean square error. It can be seen that the quality of the fit is better with the proposed metric.

Figure 6 presents a measure of performance of both metrics for all blocks of trial. The psychometric function defined in Eq. (3) has been fit to each block of trial and the likelihood of the fit has been measured. The higher the likelihood value, the better the fit is. Figure 6 presents the likelihood values as a function of the class of images. It can be seen that the fit obtained with the CMPSNR is always better than the fit obtained with the MSE. The variance of the likelihood values for the CMPSNR is much smaller than for the MSE, indicating a behavior of the CMPSNR that is more consistent with the subjective data.

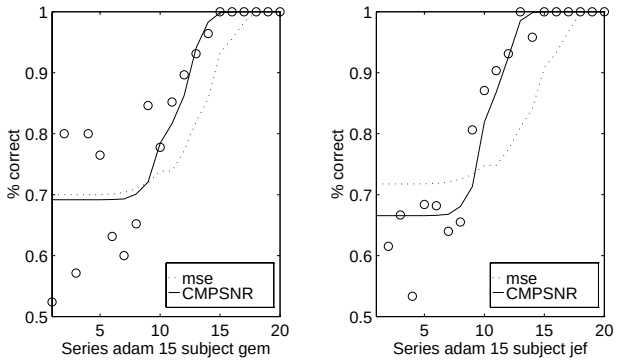

Figure 5: Fit, for a particular class, of the CMPSNR and MSE values to the subjective data with a model of the experiment.

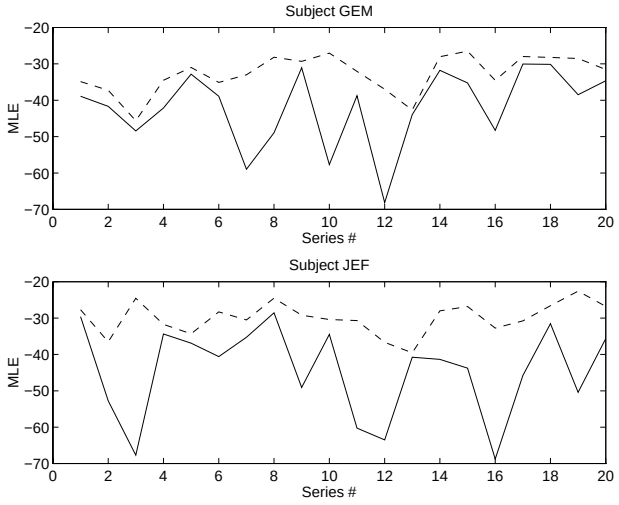

Figure 6: Graph of the likelihood of the fit as a function of the block of trial.

4 CONCLUSION

In this paper, a metric for the assessment of color still picture fidelity has been presented. The metric is based on a multi-channel model of human spatial vision and incorporates modeling of color perception, channel decomposition, contrast sensitivity and visual masking. Color perception is modeled by mapping the data into a perceptual color space, termed the opponent-colors space. The three coordinates of this color-opponent space correspond to luminance (B/W), red-green (R/G), and blue-yellow (B/Y) channels which are thought to mediate color perception. The metric has been tested on a collection of 400 JPEG-distorted images and proved to be able to predict image fidelity better than a common criterion such as the mean square error.

References

[1] Serge Comes, *Les traitements perceptifs d'images numerisées*, PhD thesis, Université Catholique de Louvain, 1995.

[2] Scott Daly, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity", *in Proceedings of SPIE*, Vol. 1616, pp. 2–15, 1992.

[3] Huib de Ridder, "Minkowski-Metrics as a Combination Rule for Digital-Image-Coding Impairments", *SPIE Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, pp. 16–26, 1992.

[4] ISO/IEC JTC1/SC2/WG10 Joint Picture Experts Group, "JPEG technical specification, revision 8", Technical Report JPEG-8-R8, 1990.

[5] Gordon E. Legge and John M. Foley, "Contrast Masking in Human Vision", *Journal of the Optical Society of America*, Vol. 70, No. 12, pp. 1458–1471, December 1980.

[6] J. M. Mendel, *Lessons in Digital Estimation Theory*, Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[7] Allen B. Poirson and Brian A. Wandell, "Appearance of Colored Patterns: Pattern-Color Separability", *Journal of the Optical Society of America*, Vol. 10, No. 12, pp. 2458–2470, December 1993.

[8] Allen B. Poirson and Brian A. Wandell, "Pattern-Color Separable Pathways Predict Sensitivity to Simple Colored Patterns", *Vision Research*, 1995, in press.

[9] M. M. Taylor and C. Douglas Creelman, "PEST: Efficient Estimates on Probability Functions", *Journal of the Acoustical Society of America*, Vol. 41, No. 4, pp. 782–787, April 1967.

[10] Patrick C. Teo and David J. Heeger, "Perceptual Image Distortion", *in Proceedings of the International Conference on Image Processing*, pp. 982–986, Austin, TX, November 1994.

[11] Brian A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.

[12] A. B. Watson and D. G. Pelli, "QUEST: A Bayesian Adaptive Psychometric Method", *Perception and Psychophysics*, Vol. 33, No. 2, pp. 113–120, 1983.

[13] Andrew B. Watson, "Perceptual-Component Architecture for Digital Video", *Journal of the Optical Society of America*, Vol. 7, No. 10, pp. 1943–1954, October 1990.