

COLOR MOVING PICTURES QUALITY METRIC

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/>

ABSTRACT

This paper presents a comprehensive quality metric for color moving pictures which is based on a spatio-temporal vision model and on the opponent-colors theory. The metric is used to assess the quality of MPEG compressed video streams and is compared with a grayscale video quality metric developed recently.

Keywords: Quality assessment, vision model, opponent-colors, color perception, MPEG, test

1. INTRODUCTION

There is a growing need and interest for the testing of video communications systems. A key aspect of testing is quality assessment. Metrics for video have been introduced only recently [8, 11]. This paper presents a computational metric for moving pictures that also accounts for color perception, which is rarely addressed. The paper is structured as follows: Sec. 2 presents the spatio-temporal color vision model. Simulation results of the proposed metric are given in Sec. 3 and Sec. 4 concludes the paper.

2. COLOR SPATIO-TEMPORAL VISION MODEL

A spatio-temporal model of human vision has been developed for the framework of video coding and presented in [6, 8]. It models the following aspects of vision:

- Contrast is used to represent the visual information, which makes the visual system adaptive to the local environment.
- The primary visual cortex represents the information at various scales and orientations. Several “mechanisms” or “channels” are thought to mediate visual perception. A channel corresponds to a band in spatial frequency, temporal frequency and orientation.
- Sensitivity to contrast varies with frequency. The *contrast sensitivity function* (CSF) defines the detection threshold for a stimulus as a function of the frequency. In the proposed model, the CSF is modeled by an excitatory-inhibitory formulation to account for the non-separability of spatial and temporal perceptions.
- Destructive interferences between two stimuli are modeled by *visual masking*, which express the modification of

the detection threshold of a stimulus in function of the local contrast of the background.

The model described in [6] models such aspects but can only process luminance information. This paper presents a color extension to the model along with a metric for color moving pictures. The metric is an extension of the moving pictures quality metric introduced in [8]. The block diagram of the computational metric, termed color moving pictures quality metric is illustrated in Fig. 1.

Color perception is difficult to model as it is a “psychological sensation” that depends on many aspects [9]. It is known that humans discriminability in color is poor [9]. The hue cancellation experiment [3] permitted to evidence very important properties of color perception: it has been observed that some pairs of hues can coexist in a single color sensation while others cannot. For example, the combination of red and yellow is perceived as orange and the combination of blue and green is perceived as cyan. On the contrary, a combination of red and green is perceived as two different colors. Schematically, it is believed that the brain uses three different pathways to encode the information, one conveying the luminance signal, another the red and green components and the third one the blue and yellow components. This is the basis of the opponent-colors theory that permits to define the opponent-colors space. The opponent-colors space decorrelates color information as it approximates the representation of the data in the various visual streams of the brain. Each of those streams is independent, hence a separable processing of each color component can be carried out.

The computational metric performs the following operations: it takes as input an original color sequence and a distorted version of the precedent in some standard color space. The data is first linearized then converted to the opponent-colors space. The color-opponent streams of the original and decompressed sequences are subtracted to compute the distortion in each color directions. The distortion and the original sequences are decomposed into perceptual channels to simulate the various detection mechanisms. Pattern sensitivity is then computed for each perceptual component, accounting for contrast sensitivity and masking. This permits to express the distortion relatively to the detection threshold, i.e. in *just noticeable differences* (jnd’s). Finally, the data is pooled over the channels to yield a global distortion measure. The various building

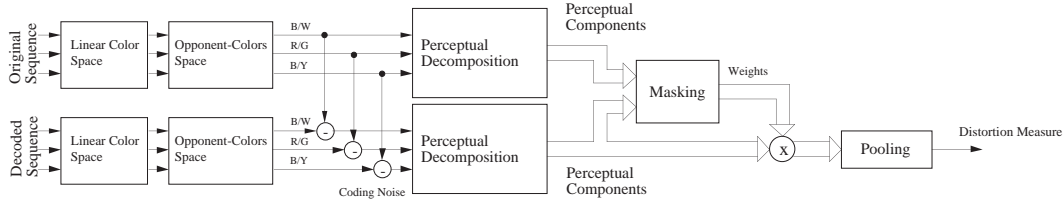


Figure 1: Block diagram of the color moving pictures quality metric.

blocks are now detailed.

2.1. Linearization of the Data

The digital samples that are known from the video sequence are termed the *frame-buffer values*. Those samples are expressed in some color space, usually (Y, U, V) or (Y, Cr, Cb) in video coding. However those values are not linear with the luminance as produced by the display device. The device performs some gamma correction and the luminance of the screen also depends on the phosphor of the screen itself. The first step is to transform the data into a calibrated space to be device-independent and to be able to apply further color space transformation. The exact procedure for display device calibration is explained in [9]. The output of this non-linear block is a compound of three streams that are now in a device-independent space and are linear with luminance.

2.2. Conversion to the Opponent-Colors Space

The next step converts the data to the 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 (B/Y). The opponent-colors space chosen here has been developed by Poirson and Wandell who measured color appearance and derived a pattern-color separable model [5, 9]. The sensitivity of the three components as a function of the wavelength are depicted in Fig. 2. The output of this module are three streams that are the color components in the opponent-colors space. The three pathways are termed luminance (B/W), red-green (R/G) and blue-yellow (B/Y). The B/W component is close to the luminance channel of the (Y, U, V) or (Y, Cr, Cb) spaces.

2.3. Perceptual Decomposition

The channel decomposition of the cortex is simulated by a three dimensional filterbank that is described in [6]. The bank used for the luminance streams features 17 spatial filters. One filter accounts for the low spatial frequency (below 2 cycles per degree (cpd)). The 16 remaining filters decomposes the data into 4 spatial frequency bands and 4 orientation bands. The frequency division is dyadic and the filters are centered at frequencies of 2, 4, 8 and 16 cpd's. The orientation bands are centered around 0, $\pi/4$, $\pi/2$ and

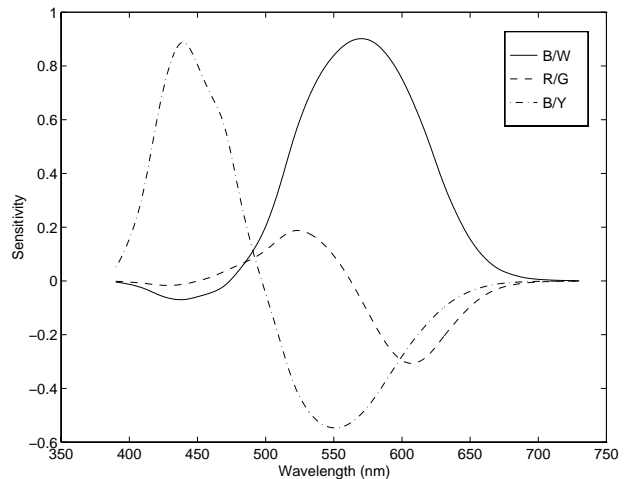


Figure 2: Spectral sensitivity of the Wandell-Poirson pattern-separable opponent-colors space. The solid line is the luminance channel (termed B/W), the dashed line is the red-green channel (termed R/G) and the dot-dashed line is the blue-yellow channel (termed B/Y).

$3\pi/4$. The temporal bank decomposes the data into 2 frequency bands that are termed the *sustained* and *transient* mechanisms.

Human sensitivity to chrominance component is much lower than to luminance information. This is a direct consequence of the characteristics of the photoreceptors mosaic in the retina, namely the spacing of the various types of cones [9]. It follows that a restrained number of channels can be used for the R/G and B/Y pathways. Along the temporal direction, only the sustained mechanism is to be considered, as temporal sensitivity drops very quickly for chrominance. In the spatial domain, the first three frequency bands are used, since sensitivity is very low above 8 cpd. Those considerations are based on the work by Watson [10].

2.4. Masking

The non linear transducer modeling of masking introduced by Legge and Foley [4] is used here. In this model, the detection threshold C_T for a stimulus is 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):

$$C_T = \begin{cases} C_{T_0} & \text{if } C_M < C_{T_0} \\ C_{T_0} \left(\frac{C_M}{C_{T_0}}\right)^\epsilon & \text{if } C_M \geq C_{T_0} \end{cases} \quad (1)$$

The pixels in the perceptual channels of the distortion are then divided by the detection threshold C_T , computed pixel by pixel and channel by channel. In this way, the data are expressed in “units above threshold” or jnd’s.

2.5. Pooling

The above steps yield a prediction of the response from the cells of area V1 of the cortex. The data is then gathered together to yield a single figure and to account for higher levels of perception. This is termed *pooling*. This step is computed as follows. First, it is considered that human observers are not looking at the whole picture at the same time but rather at regions of it. This is due to the focus of attention and the viewing distance. To take those facts into account, the pooling is computed over blocks of the sequence. Such blocks are three-dimensional and their dimensions are chosen as follows: the temporal dimension is chosen to account for persistence of the images on the retina (roughly 100 msec.). The spatial dimension is chosen to consider focus of attention, i.e. the size is computed so that a block covers two degrees of visual angle, which is the dimension of the foveal field. The distortion measure is computed for each block by pooling the error over the channels. Basically, the magnitudes 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):

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

where $e[x, y, t, c]$ is the masked error signal at position (x, y) and time t in the current block and in the channel c ; N_x , N_y and N_t are the horizontal and vertical dimensions of the blocks; N_c is the number of chromatic and achromatic channels. The exponent of the Minkowski summation is β and has a value of 4. In this application, the error measure E is further mapped onto the 1 to 5 quality scale defined by CCIR Rec. 500 [2]. In this scale, 1 is the worst quality and 5 the best. The mapping uses the following function, relating the error measure to the quality index Q :

$$Q = \frac{5}{1 + NE},$$

where N ensures a mapping between 1 and 5. This free parameter has been estimated on the basis of the vision model [8].

3. RESULTS

This subsection presents results of quality assessment on video compressed with the MPEG-2 standard operating in

MP@ML (main profile, main level) and HP@ML (high profile, main level). Two classical test sequences for broadcasting applications have been used for the simulations, *Mobile & Calendar* and *Basket Ball*. The sequences have been encoded with a software simulator of the test model 5 of MPEG-2, supplied by the MPEG Software Simulation Group, as interlaced video, with a constant group of picture structure of 12 frames and 2 B-pictures between every P-picture. The video buffer verifier size was set to its maximum allowed size. The dimension of the search windows for motion estimation was 15 pixels for P-frames, 7 pixels for backward motion estimation in B-frames and 3 pixels for forward motion estimation in B-frames. The coder operates in constant bitrate (CBR) mode. Coding has been performed on the range of bit rates that MPEG-2 typically addresses.

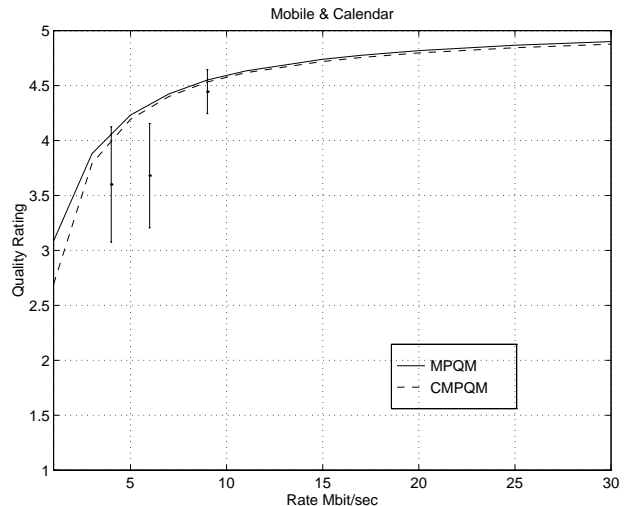


Figure 3: Comparison of CMPQM, MPQM and subjective data for Mobile & Calendar as a function of the bitrate.

Results of CMPQM on Mobile & Calendar and Basket Ball are compared in Fig. 4 and Fig. 3 with the grayscale quality metric, MPQM introduced in [8]. The curves obtained with CMPQM are very similar to those of MPQM. This is due to the low sensitivity of the chromatic pathways. As the chromatic sensitivity is one order of magnitude lower than the achromatic sensitivity, the weight of the chromatic channels is not very significant in the computation of the metric. This has been observed for still pictures too [7]: if the distortion is more or less equally distributed between chromatic and achromatic channels, a distortion measure computed on the achromatic channels can predict the quality of a picture. A distortion measure of the chromatic channels is only significant if there is a much larger distortion in the chromatic pathways than in the achromatic one.

The metrics are also compared with some available subjective data that is represented as error bars in Fig. 4 and Fig. 3. The data has been collected by the research center of RAI, Italy [1] and consists of subjective rating of compressed video by human observers. The data has been collected

according to CCIR Rec. 500-3. The method is a double stimulus continuous quality scale (DSCQS). The subjective data has been adapted to the purpose of this experiment. As both the original and the compressed sequences are given a vote in the DSCQS task, the data has been used as follows: Each result has been normalized with respect to the subjective vote of the original and the distance between the two subjective votes used to deduce an error bar.

Figure 3 presents the curves of the three metrics for the sequence Mobile & Calendar along with the tentative mapping of the subjective data. It can be seen that the subjective data is pretty noisy as the rating at 4 and 6 Mbits/sec. is nearly identical. The metric curves show however a behavior that is consistent with the data. Figure 4 presents the same results for the Basket Ball sequence along with the performance of the ITS metric. This metric is the only alternative metric for moving pictures and is described in [11]. The subjective data seems less noisy in this case. This data show particularly well an increase of perceived quality with the bandwidth in the lower range of bitrates and a saturation effect at higher bitrates, which is captured by both metrics. As pointed out earlier, the MPQM and CMPQM curves are very similar and indeed realize about the same fit. The ITS metric on the contrary is not consistent at all with such data as shown in Fig. 4.

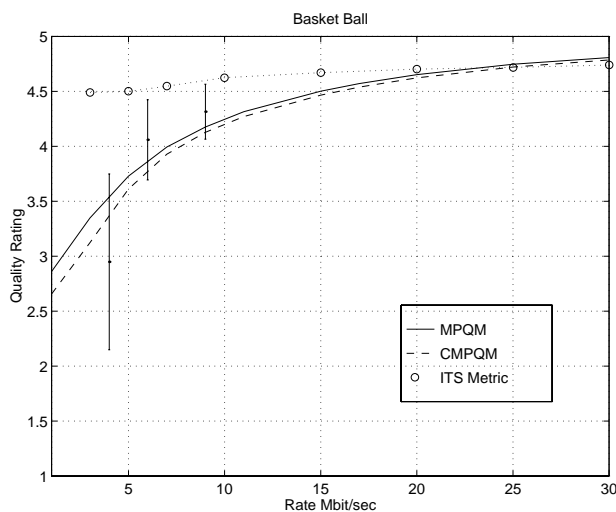


Figure 4: Comparison of CMPQM, MPQM and subjective data for Basket Ball as a function of the bitrate.

A more complete and extensive testing of the metrics remains yet to be performed. The inherent problem with such experiments is that a large number of observers is required and a specific equipment is needed. Moreover, collection of the data takes a huge amount of time. Such experiments could not be performed within the scope of this work.

4. CONCLUSION

This paper presented a quality metric for color moving pictures. The tool is an extension of a metric for grayscale

video introduced recently. The metric is based on a spatio-temporal vision model and color perception is modeled by the opponent-colors theory. Both metrics are compared and exhibit a behavior that is coherent with the subjective data. However, the prediction of both metrics are very similar due to the low sensitivity of the chromatic channels.

5. REFERENCES

- [1] M. Ardito, M. Barbero, M. Stroppiana, and M. Visca, "Compression and Quality", in L. Chiariglione, editor, *Proceedings of the International Workshop on HDTV 94*, pp. B-8-2, Torino, Italia, October, 26-28 1994. Springer-Verlag.
- [2] CCIR, "Method for the Subjective Assessment of the Quality of Television Pictures", *13th Plenary Assembly, Recommendation 500*, Vol. 11, pp. 65-68, 1974.
- [3] L. M. Hurvich and D. Jameson, "An Opponent-Process Theory of Color Vision", *Psychology Review*, Vol. 64, pp. 384-404, 1957.
- [4] 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.
- [5] 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.
- [6] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2293-2296, Atlanta, GA, May 7-10 1996, available on <http://ltswww.epfl.ch/~vdb>.
- [7] Christian J. van den Branden Lambrecht and Joyce E. Farrell, "Perceptual Quality Metric for Digitally Coded Color Images", in *Proceedings of the European Signal Processing Conference*, Trieste, Italy, September 10-13 1996, available on <http://ltswww.epfl.ch/~vdb>.
- [8] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", in *Proceedings of the SPIE*, Vol. 2668, pp. 450-461, San Jose, CA, January 28 - February 2 1996, available on <http://ltswww.epfl.ch/~vdb>.
- [9] Brian A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.
- [10] 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.
- [11] Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson, Stephen D. Voran, and Stephen Wolf, "An objective quality assessment system based on human perception", in *SPIE Human Vision, Visual Processing, and Digital Display IV*, Vol. 1913, pp. 15-26, San Jose, CA, Feb. 1993.