VISUAL QUALITY ASSESSMENT USING A CONTRAST GAIN CONTROL MODEL

Stefan Winkler
Signal Processing Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
http://ltswww.epfl.ch/~winkler/
Stefan.Winkler@epfl.ch

Abstract - Much of the work on visual quality assessment has been devoted to gray-level images; metrics taking into account color information and the temporal component are still relatively rare. This paper presents a quality metric for color video which is based on a recent contrast gain control model of the human visual system. It is used to assess the quality of MPEG-coded sequences and exhibits a behavior that is consistent with subjective ratings.

INTRODUCTION

Visual quality assessment is becoming a growing concern for the video industry [9]. The sophisticated compression of digitally coded visual information and the resulting artefacts have made the evaluation of the actual visibility of distortions in images and video increasingly difficult. Video engineers usually have to rely on subjective viewing tests in order to obtain reliable quality ratings. Such tests have been standardized in ITU-R Recommendation 500 [3] and have been used for many years. While they undoubtedly represent the benchmark for visual quality measurements, these tests are complex and time-consuming, hence expensive and highly impractical or not feasible at all. Consequently, researchers often turn to very basic error measures such as root mean squared error (RMSE) or peak signal-to-noise ratio (PSNR) as alternatives, suggesting that they would be equally valid. However, these simple error measures operate solely on a pixel-by-pixel basis and neglect the much more complex behavior of the human visual system.

These problems prompted the recent interest in tools for automatic visual quality assessment. Some exploit a-priori knowledge about particular artefacts that are introduced by certain compression or transmission methods, while others are more generally based on models of the human visual system [9]. The metric presented here belongs to the latter category; as such it is independent of specific types of distortions in the video. This paper gives an overview of the underlying vision model and its integration into a quality metric for color video, and presents some new results.

VISION MODEL

The proposed quality metric models color perception, pattern sensitivity, and the multi-channel organization of the human visual system. As a matter of fact, these aspects are closely interrelated. Pattern sensitivity a.k.a. masking is of particular importance in the context of visual quality assessment, because it explains the dependence of artefact visibility on the local image background. Analyses of the responses of neurons in the primary visual cortex showed that this phenomenon can be accounted for by a process of *contrast gain control* [2,5,6]. The metric presented here combines such a contrast gain control model with a theory of color perception and adds the dimension of time for the analysis of video [7, 8].

The structure of the proposed quality metric is outlined in Figure 1. It takes two inputs, a reference sequence and its processed version, which go through identical stages of the vision model until their differences can be combined in a perceptually meaningful way.

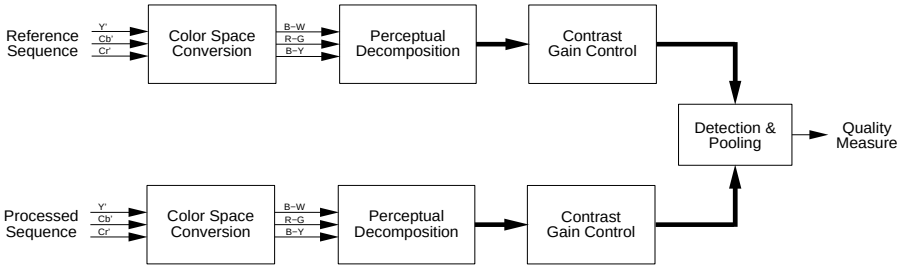

Figure 1: Block diagram of the proposed quality metric.

The first stage of the model converts the input sequence into a series of progressive frames and performs the conversion from the typical Y'Cb'Cr' color space of component video to an opponent-colors space. This color space is inspired by the processing taking place in the human visual system, which presumably uses separate channels for black-white (B-W), red-green (R-G), and blue-yellow (B-Y) information. Specifically, the model relies on a pattern-color separable opponent-colors space introduced recently [4].

The second stage models the perceptual decomposition of visual information into a number of channels with different frequencies and orientations. This decomposition is carried out first in the temporal and then in the spatial domain; the corresponding filters are shown in Figure 2. The temporal filter bank approximates the low-pass and the band-pass mechanism of vision by means of IIR filters. They have been designed so as to introduce as little delay as possible, because in some quality metric applications a short response time is crucial. We have achieved delays of about 150 ms while still maintaining a satisfactory approximation accuracy. The spatial filter bank further decomposes the frames into bands with different spatial frequencies

and orientations. In the present implementation, the spatial filters have octave bandwidth and octave spacing; four orientation bands are used. Finally, each of the channels resulting from this perceptual decomposition is weighted in accordance with the contrast sensitivity of the human visual system.

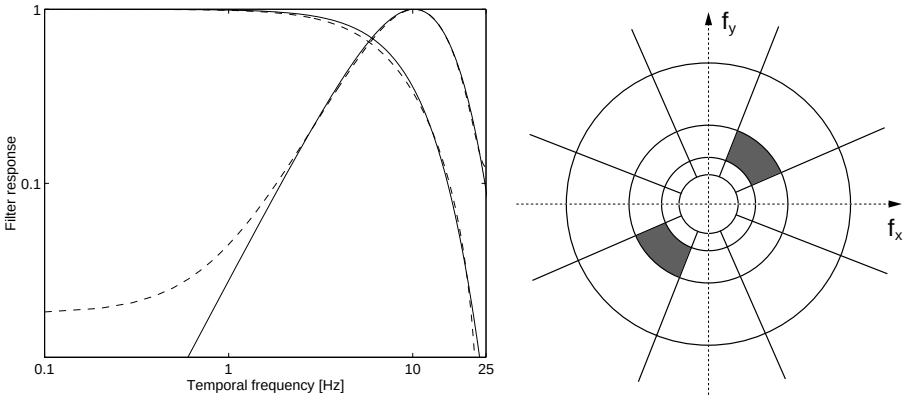

Figure 2: Illustration of the temporal filters (left) and the approximate partitioning of the spatial frequency plane (right) employed in the metric.

The third stage implements a contrast gain control model for pattern masking. In the primary visual cortex, contrast gain control serves as a mechanism to keep neural responses within their permissible range while retaining the information conveyed by the activity of the ensemble of neurons. Each neuron is modeled by an accelerating nonlinearity in the excitatory path, and is inhibited by a pool of responses of other neurons via divisive normalization. Masking occurs through the inhibitory effect of this normalizing pool [6]. Within the vision model, the coefficients from all the channels of the perceptual decomposition take the place of neuronal responses and are subjected to this process of contrast gain control.

The final stage performs detection and pooling over channels by computing a Minkowski sum of the difference between the sensor outputs from the reference sequence and the sensor outputs from the processed sequence, yielding a measure for the amount of perceptual distortion.

RESULTS

Parameter Adjustment

The vision model contains several parameters that have to be adjusted in order to accurately represent the human visual system. This adjustment is carried out by fitting the output of the metric to data from psychophysical experiments on contrast sensitivity and pattern masking [7, 8]. Contrast sensitivity is modeled by setting the weights of the filters in the perceptual

decomposition in such a way that the ensemble produces the empirical spatio-temporal contrast sensitivity curve for the three opponent-color channels. The parameters of the contrast gain control stage are determined by fitting the metric's responses to masked gratings to data from masking experiments.

The overall quality of the fits to these empirical data is very good and close to the difference between measurements from different observers. In fact, most of the effects found in the psychophysical experiments are captured by the model.

Distortion Maps

In order to demonstrate how the proposed quality metric works, I encoded the *Basketball* sequence with the MPEG-2 encoder of the MPEG Software Simulation Group.[1] This sequence contains significant spatial detail, a considerable amount of fast motion (the players in the foreground), and slow camera panning, which makes it an interesting test sequence for a spatio-temporal vision model. Figure 3 shows a sample frame and the pertinent distortion maps for each color channel produced by the proposed quality metric. These distortion maps already include temporal aspects of the distortions. As can be seen, most of the visible artefacts appear in the B-W channel around the silhouettes of the players. The distortions in the R-G and B-Y channels show a distinct block structure. This is due to the subsampling in the implementation of the spatial decomposition and shows that the model correctly emphasizes low-frequency distortions in these channels. Note that the distortions in the chromatic channels are much smaller than in the B-W channel, but the maps were normalized to reveal more spatial detail.

Quality Ratings

Since the metric intrinsically measures the distortion, i.e. the visual difference between the two input sequences, its output has to be converted to a quantity that can be related to subjective quality ratings more easily. This is achieved on the basis of a correlation plot of subjective ratings and the raw distortion measurements of the metric. In general, this plot will approximate a straight line, but at the high and low ends of the scale some compression may occur due to the limited discrimination capability of subjects at these ends. In order to take this behavior into account, the output of the metric undergoes a nonlinearity of the form

$$Q = a_0 + a_1 d + a_2 d^2 + a_3 d^3,$$

where d is the raw distortion output from above, and Q is the final quality prediction of the metric. The coefficients a_i are computed from fitting this function to subjective ratings.

[1] This software is available at http://www.mpeg.org/MPEG/MSSG/.

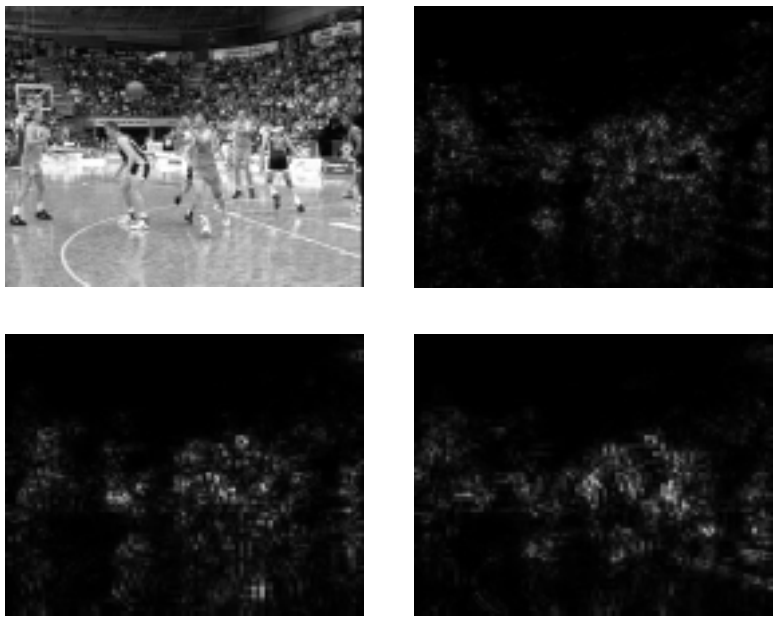

Figure 3: Distortion maps of the sample frame (top left) for the B-W channel (top right), the R-G channel (bottom left) and the B-Y channel (bottom right). The images were normalized; actually the visible distortions in the chromatic channels are much smaller.

The performance of a video quality assessment system is best demonstrated by a statistical analysis of the correlation of its predictions with subjective ratings of a large set of test sequences. However, the inherent problem with such a procedure is the limited availability of subjectively rated sequences. As a matter of fact, there is an ongoing effort to collect such data for distribution-class video within the work of the Video Quality Experts Group (VQEG).[2] As the work of this group is not yet concluded, data from [1] is used here for a preliminary validation of the proposed video quality metric. Figure 4 compares some subjective ratings to the metric's predictions for the *Basketball* sequence encoded at various bitrates using the MPEG-2 encoder mentioned above. Although it has to be noted that the subjective data are rather noisy, it can be seen that the predictions of the proposed quality metric are consistent with the data. The quality grows monotonically with the bitrate, and the saturation of quality that can be expected at higher bitrates is evident. Future work will include a more detailed analysis of the metric's performance using the data from the VQEG effort.

[2] Consult the official VQEG web site http://www.crc.ca/vqeg/ for more information.

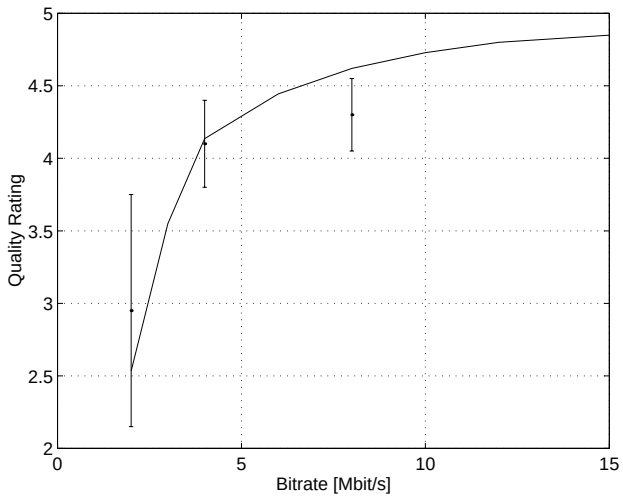

Figure 4: Comparison of subjective ratings from [1] with the predictions of the proposed quality metric for the *Basketball* sequence at various bitrates.

References

[1] M. Ardito et al.: "Compression and quality." in *Proc. HDTV Workshop*, 547–559, Turin, Italy, 1994.

[2] J. M. Foley: "Human luminance pattern-vision mechanisms: Masking experiments require a new model." *J. Opt. Soc. Am. A* **11**(6), 1710–1719, 1994.

[3] ITU-R Recommendation BT.500-7: "Methodology for the subjective assessment of the quality of television pictures." ITU, Geneva, Switzerland, 1995.

[4] A. B. Poirson, B. A. Wandell: "Pattern-color separable pathways predict sensitivity to simple colored patterns." *Vision Res.* **36**(4), 515–526, 1996.

[5] P. C. Teo, D. J. Heeger: "Perceptual image distortion." in *Proc. SPIE*, vol. 2179, 127–141, San Jose, CA, 1994.

[6] A. B. Watson, J. A. Solomon: "Model of visual contrast gain control and pattern masking." *J. Opt. Soc. Am. A* **14**(9), 2379–2391, 1997.

[7] S. Winkler: "A perceptual distortion metric for digital color images." in *Proc. ICIP*, vol. 3, 399–403, Chicago, IL, 1998.

[8] S. Winkler: "A perceptual distortion metric for digital color video." in *Proc. SPIE*, vol. 3644, San Jose, CA, 1999.

[9] S. Winkler: "Issues in vision modeling for perceptual video quality assessment." *Signal Processing* **78**(2), 1999.