Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System

Christian J. van den Branden Lambrecht and Olivier Verscheure*
Signal Processing Laboratory
Swiss federal Institute of Technology
CH-1015 Lausanne Switzerland
vdb@lts.de.epfl.ch

ABSTRACT

This paper addresses the problem of quality estimation of digitally coded video sequences. The topic is of great interest since many products in digital video are about to be released and it is thus important to have robust methodologies for testing and performance evaluation of such devices. The inherent problem is that human vision has to be taken into account in order to assess the quality of a sequence with a good correlation with human judgment. It is well known that the commonly used metric, the signal-to-noise ratio is not correlated with human vision.

A metric for the assessment of video coding quality is presented. It is based on a multi-channel model of human spatio-temporal vision that has been parameterized for video coding applications by psychophysical experiments. The visual mechanisms of vision are simulated by a spatio-temporal filter bank. The decomposition is then used to account for phenomena as contrast sensitivity and masking. Once the amount of distortions actually perceived is known, quality estimation can be assessed at various levels. The described metric is able to rate the overall quality of the decoded video sequence as well as the rendition of important features of the sequence such as contours or textures.

Keywords— **Quality Assessment, Quality Metric, Vision Science, Test, Coding Quality, MPEG, H.263**

1 INTRODUCTION

Quality assessment of digitally coded pictures and image sequences nowadays becomes a very important issue and receives a large attention from the signal processing community. The point is that a new generation of products is about to reach the consumer market. It is based on digital representation of image and video and addresses applications such as television, personal video communications and multimedia. However, the testing procedure of analog equipment cannot be used for such devices and new methodologies have to be designed. Few works did address the problem. For now, testing equipment consists in pattern generators[1] or bitstream verifiers.[2] However, a lot of work remains to do in objective quality assessment. The generic metric currently used is the signal-to-noise ratio (SNR), which is known to be uncorrelated with the human visual system and cannot be

*now with Telecommunications Laboratory, Swiss Federal Institute of Technology,CH-1015 Lausanne Switzerland

trusted for this purpose. Some work has been carried out in the field of still-picture quality evaluation,[3–5] however no metric really existed for moving pictures.

This work uses a complete spatio-temporal model[6] of the human visual system (HVS). The latter is used as a front end to several proposed quality metrics. A multi-measurement scheme is introduced into which the distortions of several features of the scene are evaluated. The paper is divided as follows: the human visual system is presented in Sec. 2 and an efficient way of modeling it is the subject of Sec. 3. The metrics are described in Sec. 4 and Sec. 5 presents some simulations. Finally, Sec. 6 concludes the paper.

2 THE HUMAN VISUAL SYSTEM

The level of description that is of interest is at the cortex level. The formalism is that of psychophysics where, the human visual system is modeled as a system characterized by a transfer function. The model is based on the following considerations: electro-physiological experiments showed that cells of the primary visual cortex are tuned to bands in spatial frequency and orientation.[7] Such data have been confirmed by psychophysical experiments, showing the presence of several mechanisms of vision that divide the frequency plane into frequency and orientation bands.[8] Temporal perception has been studied as well and there is considerable evidence for the existence of two different mechanisms of temporal vision. The first mechanism is sensitive to low temporal frequencies and is called *sustained* mechanism, whereas the second, termed *transient* mechanism, is band pass. Such concept will be further developed in Sec. 3. Perception is mainly governed by two key concepts that are *contrast sensitivity* and *masking*. The first phenomenon accounts for the perception of single wavelength, or stimuli. The second quantizes the interactions between several stimuli.

It is well known in the video community that the human eye is less sensitive to higher spatial frequency than to lower and this fact has been used in the design of video equipments. Indeed, the human eye has a spatial frequency response that is bandpass with a peak frequency around 4 cycles per degree (cpd). This has been quantized as the *contrast sensitivity*. More precisely, a signal is detected by the eye only if its contrast is greater than some threshold defined as the *detection threshold*. The detection threshold varies as a function of spatial frequency. The sensitivity is defined as the inverse of the detection threshold and is thus a function of spatial frequency. The term *contrast sensitivity function* (CSF) is usually used to denote this function. A typical CSF is illustrated in Fig. 1.

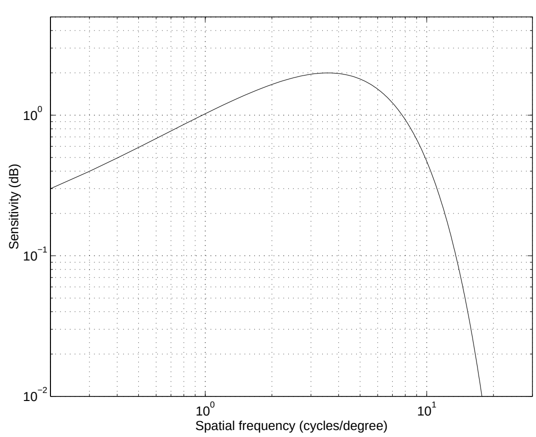

Figure 1: Illustration of the sensitivity of the human eye as a function of spatial frequency.

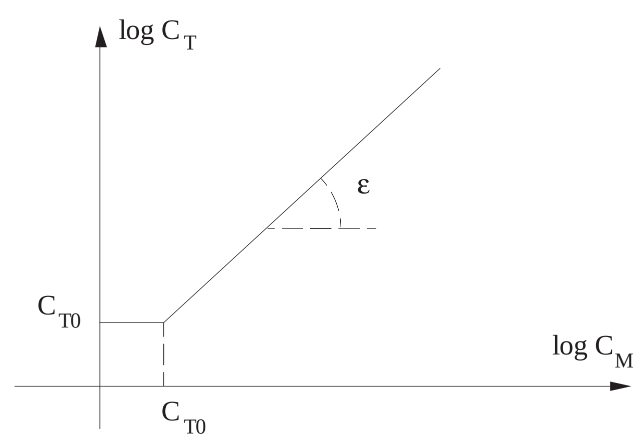

Figure 2: Model of the masking phenomenon.

More generally, the eye sensitivity varies as a function of spatial frequency, orientation and temporal frequency, defining a contrast sensitivity surface. It is commonly admitted that the sensitivity is larger to horizontal and vertical patterns than to diagonal ones and more to moving objects than to still ones.

The other dominant characteristic of human vision is its response to the combination of several signals. A stimulus will be perceived differently as a function of the background onto which it lies. Both signals (the foreground and the background) will indeed interfere and their relative perception will be modified. In actuality, the detection threshold of the foreground will be modified as a function of the contrast of the background. This is known as *masking*. The phenomenon is a key aspect in engineering since it will define the quality of rendition of a scene: any distortion introduced by a physical device can be thought of as a stimulus onto a background (the original scene). The perception of the distortion can be (according to its nature) either strongly visible, attenuated or completely masked by the scene.

Masking has been intensively studied. It has been shown that its effect is maximum when the stimulus and the masker are closely coupled (in terms of orientation, spatial and temporal frequency) and decreases rapidly as the distance between the considered signals increases in the spectral domain. A common model of masking is a non linear transducer as illustrated in Fig. 2. Consider two stimuli, the "signal" and the "masker" that are closely coupled. Let C_{T0} denote the detection threshold of the signal measured in the absence of masker (as obtained from the CSF), C_T the detection threshold of the signal in the presence of a masker and C_M the contrast of the masker. Three regions can be identified:

- At low values of C_M, the detection threshold remains constant, that is $C_T = C_{T0}$.
- As C_M gets closer to C_{T0}, the detection threshold slightly decreases, and presents a dipper (this phenomenon is neglected in Fig. 2).
- Finally, as C_M increases, C_T increases as a power of the contrast masker. This function is linear in a log-log graph and its slope is denoted ε.

The actual detection threshold, C_T is computed on this basis as:

$$C_T = \begin{cases} C_{T0} & \text{if } C_M < C_{T0} \\ C_{T0}\left(\frac{C_M}{C_{T0}}\right)^{\varepsilon} & \text{otherwise,} \end{cases}$$

where ε is termed *slope of the masking function* as illustrated in Fig. 2.

3 MULTI-CHANNEL MODEL OF HUMAN VISION

Both the physiological and psychophysical experiments carried out on perception gave evidence of the bandpass nature of the cortical cells' response in the spectral domain and of the well defined structure of the peak sensitivity positions and the bandwidths of the cells' response. The human brain seems therefore to possess a collection of separate mechanisms,[9] each being more sensitive to a portion of the frequency domain. This suggests a filter bank approach to the modeling of vision. The filter bank is then the approximation of the various mechanisms of vision and decomposes the visual data in a collection of signals that are band-limited in orientation, spatial frequency and temporal frequency. Such band-limited signals are termed *channels*. The structure of the mechanisms has been actively studied by means of psychophysical experiments.

The profile of the cortical cells is extremely similar to Gabor function[9] and some authors associate this to an optimal encoding that would be performed by the brain (since Gabor patches are the functions that are the most compact both in spatial or temporal and frequency domain). The Gabor shape can be a problem for some applications since the resulting filter bank will not guarantee perfect reconstruction, however it is possible to approximate such filter bank using suboptimal solutions.[5]

The spatial frequency domain is divided in four to eight bands in a logarithmic partition and there exists about the same number of orientation bands, although the division is linear. The temporal frequency axis seems to be covered by two to three channels.[10–12] Many studies give controversial opinions onto the existence of the third temporal frequency mechanism (that would only exist at very low spatial frequency). Recent studies tend to confirm the common concept of the existence of only two temporal mechanisms (transient and sustained).[13] In this work, the spatial filter bank is made of 17 filters as illustrated in Fig. 3.[5] All filters have a Gabor profile. One of the filters is isotropic and centered around the spatial frequency of 0. Its bandwidth is 2 cpd. The frequency plane is then divided in radial frequency and orientation. Four orientation bands have been chosen, at 0, $\pi/4$, $\pi/2$ and $3\pi/4$ radians. There are four frequency bands as well, dividing the frequency axis according to an octave band division. The four filters respectively have peak frequencies of 2, 4, 8 and 16 cpd and respective bandwidth of one octave. The bank is illustrated in Fig. 3. The temporal filter bank simulates the sustained and transient mechanisms of human vision. It has two filters and is illustrated in Fig. 4.

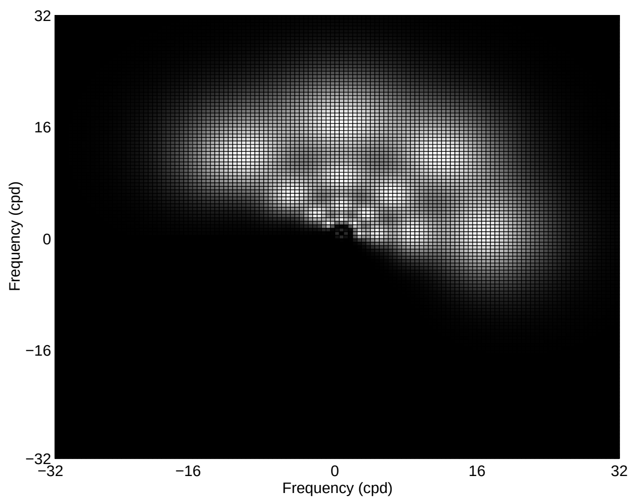

Figure 3: The spatial filter bank, featuring 17 filters (5 spatial frequencies and 4 orientations). The magnitude of the frequency response of the filters are plotted on the frequency plane. The lowest frequency filter is isotropic.

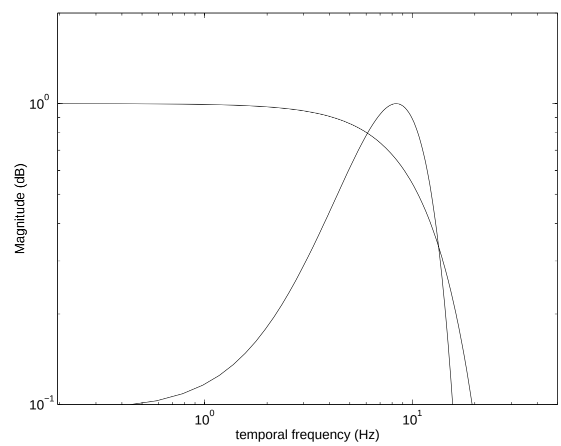

Figure 4: The temporal filter bank accounting for two mechanisms: one low pass (the sustained mechanism) and one band pass (the transient mechanism). The frequency response of the filters is plotted as a function of temporal frequency.

It has been previously stated that masking of one stimulus by another is higher the more the stimuli are coupled and decreases as the distance between the stimuli (in terms of localization in the frequency domain) increases. The consequence of this is that channels are considered to be independent one from the other. It is known that it is not exactly the case and some researchers now model inter-channel masking.[14,15] In this model, complexity has been restricted to intra-channel masking only.

Another important issue is separability. Separability can be considered at two different stages of the model: the contrast sensitivity function and the filter bank. The spatio-temporal contrast sensitivity function is clearly non-separable as pointed out by many studies.[10,11] For example, the perceptual temporal characteristics of moving objects are dependent on their spatial properties: the temporal dependence is band pass at low spatial frequencies and low pass at high spatial frequencies.

As far as spatial vision is concerned, it has been pointed out that the mechanisms of vision are tuned in frequency and orientation. They thus have a polar structure, which makes the filter bank not separable in Cartesian spatial coordinates. For the temporal aspect, two schools of thoughts exist to explain the interaction between spatial and temporal perception. According to the first one, this would be due to a dependence of the peak sensitivity of the temporal filters with spatial frequency. This is referred to as the sensitivity scaling hypothesis. The other school relates this dependency to a spatio-temporal covariation in the temporal properties across the population of filters, which constitute the covariation hypothesis. The sensitivity scaling hypothesis means that a filter bank separable in the temporal and spatial domains can be used. The covariance hypothesis would yield a more complex filter bank. The recent study performed by Hess and Snowden[11] confirms results of previous studies, promoting the sensitivity scaling hypothesis. This result has been assumed for the present work and the variation in the filter gains is modeled at the level of the CSF that is a non separable function of spatial and temporal frequencies.

The model has been parameterized by means of psychophysics.[6,16] The goal of the experiments was to measure the CSF. The psychophysical experiments have been performed to study specifically perception of coding noise. This has been done in the following way: five subjects took part to experiments where they have been asked to assess the visibility of stimuli. White noise filtered by a perceptual channel has been used as stimuli as it will represent a signal close to coding noise filtered by the same channel. The experiment was a two alternatives forced choice discrimination task.[17] The level of the stimuli were adaptively decided on the fly by a modified PEST procedure.[18] Details of the experiment are reported in another work.[16] They resulted in the estimation of the spatio-temporal CSF.

4 METRICS

This section describes the metrics. It explains how using the above knowledge about the human visual system, a model suitable for image processing purposes is built. The general structure of the model is presented in Fig. 5 and the various building blocks are described hereafter.

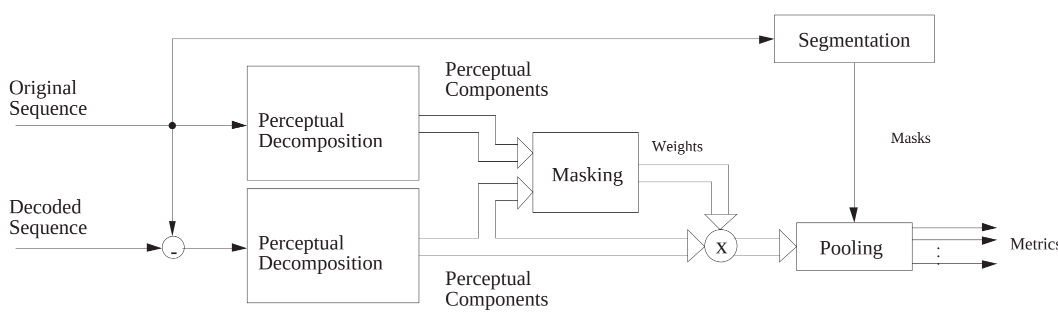

Figure 5: Block diagram of the quality metrics. The thick lines represent a set of perceptual components. The thin lines represent sequences.

The steps involved in the computation of the metric are the following: first of all, a coarse segmentation of the original sequence is computed. The original and coding error sequences are decomposed into perceptual components by the filter bank. Contrast sensitivity and masking are considered in each of the channels. This is done by computing, pixel by pixel, the actual detection threshold, considering that the original sequence will be a masker to the distortion. The masking strategy then consists of dividing the filtered error signal by the detection threshold. This expresses the data in *units above threshold* also referred to as *just noticeable differences* (jnd's). Finally, the data is pooled over the channels to compute the distortion measures.

The segmentation that is used is fairly simple. It segments the sequence into uniform areas, contours and textures.[19] Basically it proceeds as follows: the input images are parsed and processed block by block. The variance of each elementary block is computed, as well as the variance in the horizontal, vertical and diagonal directions. This data will be used to assess the nature of the center pixel of the block. If the global variance is below some threshold, the pixel is considered as belonging to a uniform area. If the variance is above the threshold and more or less isotropic, the pixel is set as texture. Otherwise, i.e. if there is one direction of very low variance compared to the others, the pixel is considered to be lying on a contour.

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 level of perception, which is termed *pooling*. This work proposes a *multi-measurement* scheme, into which not only a global quality measure is computed for a sequence but some detailed metrics as well. This section presents a global quality metric as well as more detailed metrics that measure rendition of three basic components of images: uniform areas, contours and textures. Pooling will be slightly different for the global metric than for the detailed metrics.

Human observers are never looking at the whole picture at the same time but rather at regions. This is due to both the focus of attention and the viewing distance. To take those facts into account, the global metric will be 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. 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 fovea. 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. (1):

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

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 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.[10] This metric is termed *moving pictures quality metric* (MPQM).

Pooling for the detailed metrics is slightly different. Once the masked error signal is known, a Minkowski summation is directly computed region by region (for textures, uniform areas and contours). This directly constitutes the distortion measure for each of the features.

The distortion E computed in Eq. (1) has to be expressed on some known scale to be significant. Two solutions are proposed. As engineers are used to work with decibels (dB's), the distortion could be expressed on a logarithmic scale. The metric, that by analogy to the work of Comes,[5] can be named MPSNR (masked peak signal-to-noise ratio), is then computed as:

$$MPSNR = 10 \ \log_{10} \frac{255^2}{E^2} \ .$$

This scale does not have the exact same meaning as the known dB's, hence we refer to it as "visual decibels" (vdB's). Another solution is the use of a quality scale that has often been used for subjective testing in the engineering community.[20,21] It is a scale from 1 to 5, as described in Tbl. 1.

The quality rating on this scale is obtained using the normalized conversion[21] described in Eq.(2):

$$Q = \frac{5}{1 + N \ E} , \tag{2}$$

where Q is the quality rating and E is the measured distortion. N is a normalization constant. This constant is usually chosen so that a known reference distortion maps to a known quality rating. In the case of a perceptual

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Table 1: Quality rating on a 1 to 5 scale.

metric as MPQM, the choice of N has to be made on the basis of the vision model. N has been estimated in the following way. Assume a sequence that only has an error of one jnd in only one channel, at a single pixel of a single block. This is the smallest error that could theoretically be perceived. Hence the quality rating of that sequence should be extremely close to the highest quality level. We considered that such an error would yield a quality rating of 4.99 and solved for N in Eq. (2). The value of N then turns out to be:

$$N = 0.623.$$

5 SIMULATIONS

This section presents quality assessment on some compressed video sequences. It presents results on two main applications: high bitrate broadcasting for which the considered coder is MPEG-2[22] and very low bitrate communications, using the recommendation H.263.[23]

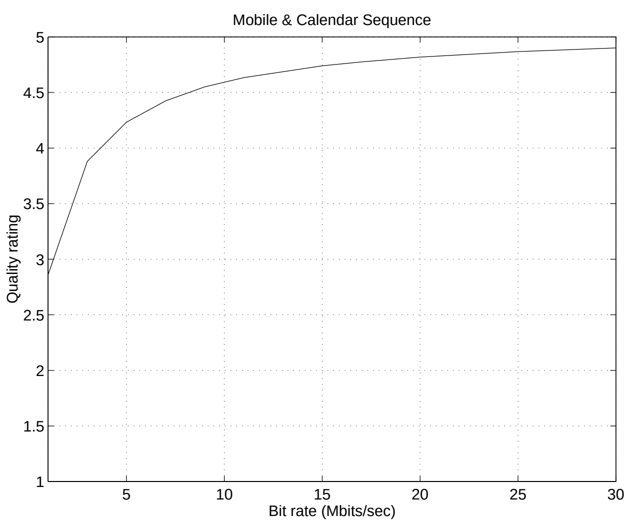

Figure 6: MPQM quality assessment of MPEG-2 video for the Mobile & Calendar sequence as a function of the bit rate.

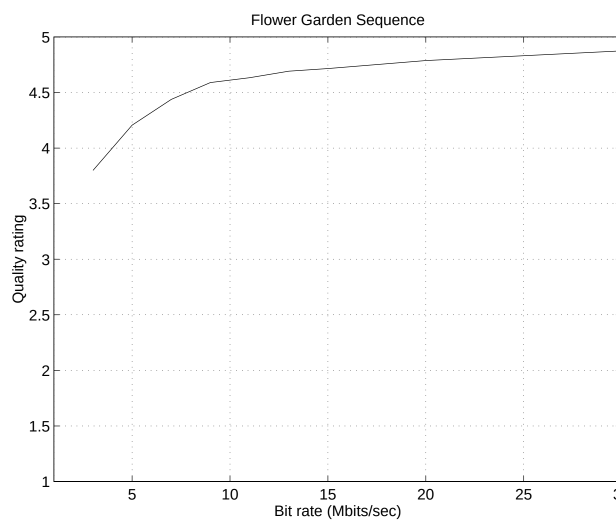

Figure 7: MPQM quality assessment of MPEG-2 video for the Flower Garden sequence as a function of the bit rate.

5.1 Characterization of MPEG-2

Three common sequences have been used for simulations. They are *Mobile & Calendar*, *Flower Garden* and *Basket Ball*. The sequences have been encoded with a software simulator of the test model 5 of MPEG-2.[24] The sequences have been encoded 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 window 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. Coding has been performed on the range of bit rates that MPEG-2 typically addresses.

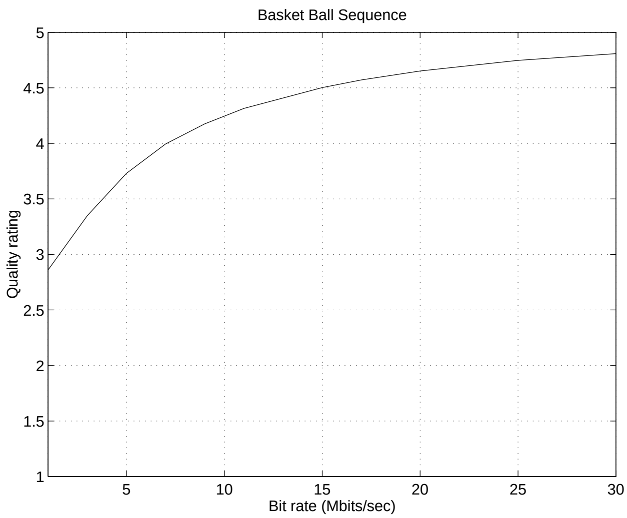

Figure 8: MPQM quality assessment of MPEG-2 video for the Basket Ball sequence as a function of the bit rate.

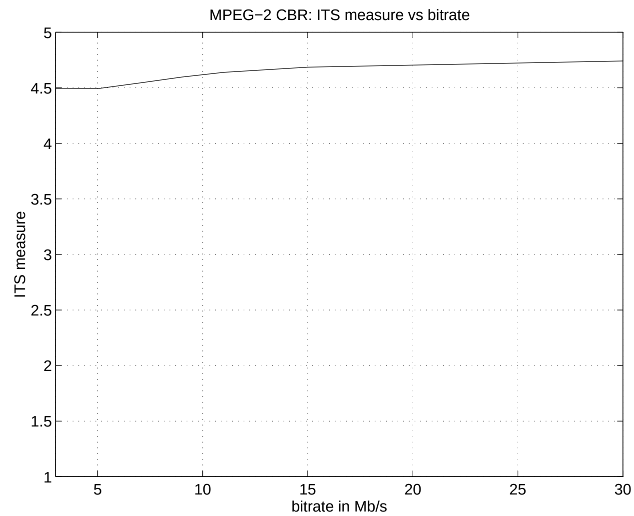

Figure 9: $\hat{s}$ quality assessment of MPEG-2 video for the Basket Ball sequence as a function of the bit rate.

Figures 6, 7 and 8 present the quality assessment by MPQM for the three sequences. The important result that can be extracted from the graphs is the saturation in quality at high bitrates, meaning that, at some point, increasing the bitrate is a waste of bandwidth since the end user does not really perceive an improvement in quality. It can also be seen that Basket Ball is the sequence that is most difficult to encode among the three. Its quality saturates much more slowly than the others. The values obtained with MPQM are very consistent with collected subjective data.[20] As it is shown in other works,[25] MPQM performs much better than other metrics developed for video. This is illustrated by Fig. 9, showing quality assessment for Basket Ball by a metric known as $\hat{s}$.[26] Clearly, $\hat{s}$ has a dynamic that is much too small and over-estimate quality in the lower range of MPEG-2 bit rates.

5.2 Characterization of H.263

Simulations are now presented in a very low bit rate framework. Coding quality of the recommendation H.263[23] was studied with a software simulator.[27] Encoding has been performed in variable bit rate mode, without the PB-frames and syntax-based arithmetic coding options but with unrestricted motion vectors and overlapped motion compensation. Two sequences have been used for the experiment. One is the *Carphone* sequence, and the other is the *LTS* sequence.[28] Comparison of the MPSNR and PSNR metrics are shown in Fig. 10 and 11 for the Carphone sequence and in Fig. 12 and 13 for the LTS sequence. The saturation effect in perceived quality can again be observed. The latter is however completely missed by the PSNR, since it does not consider any aspect of vision.

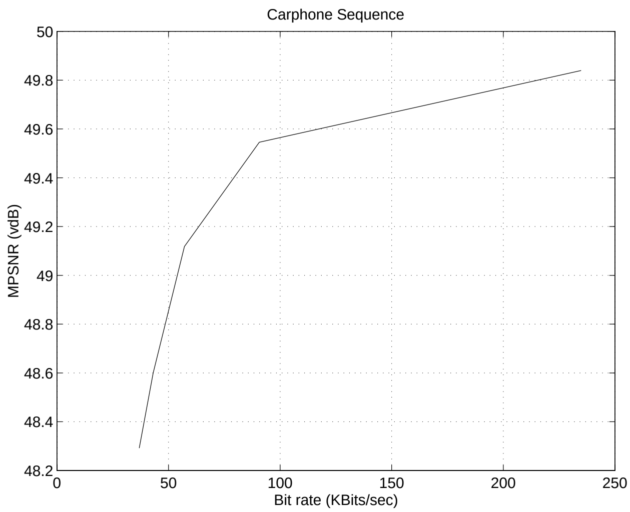

Figure 10: MPSNR quality assessment for the Carphone sequence as a function of the bitrate.

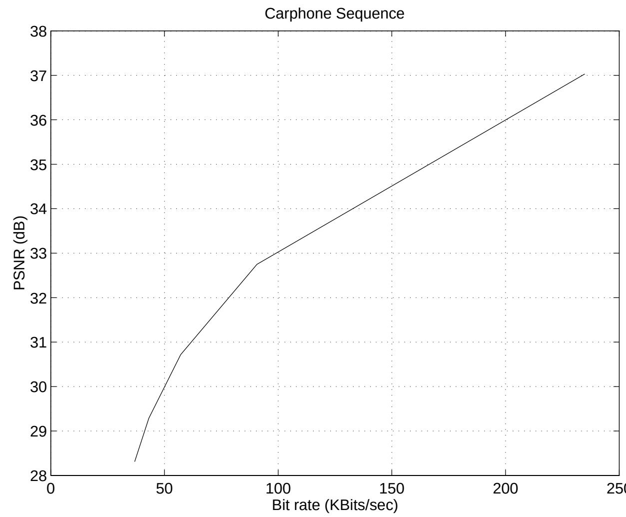

Figure 11: PSNR quality assessment for the Carphone sequence as a function of the bitrate.

Finally, as an illustration, the MPQM rating for both sequences is shown in Fig. 14 and 15. The aspect of the curves are identical to the MPSNR curves which clearly shows that the saturation observed in the graph is measured by the metric and is not a side effect of the mapping function in Eq. (2). The quality rating of the sequences remains pretty low whatever the bitrate is since the estimation has been performed on the very critical part of both sequences.

5.3 Detailed Metrics

The work on testing methodology[1] for digital codecs proposes synthetic test patterns. One of the patterns, namely a test sequence meant to study the rendition of edges, is now used to illustrate the detailed metrics. The sequence consists in a rotating square. It has been used as input to the MPEG-2 encoder. Input format was CIF and the bitrate was 700 Kbits/sec. Figure 16 presents quality curves for the various features of the sequence, along with the global quality curve obtained with MPQM. The measured features are contours, textures and uniform areas. All curves are close to and consistent with the global quality curve. However, it can be seen that contour rendition is highly variable (the test sequence was meant to fully test this effect by presenting edges in various orientations). Uniform areas have a behavior that is very similar to the general quality curve. Eventually, textures areas have the lowest quality (due to block DCT coding) and exhibit a drop in quality as time goes on. This effect is due to the level of occupancy of the video buffer verifier of the coder. As the later gets full, quantization becomes coarser and textures are very sensitive to this effect.

6 CONCLUSION

This paper presented quality metrics for digitally coded video. They are based on a model of the human visual system. The model accounts for spatial and temporal aspects of human vision and their interaction. It has a multi-channel structure and predicts the response from the neurons of the primary visual cortex. The metrics then consist of a pooling of such prediction data to account for higher level of cognition. A general quality metric for video, termed MPQM (moving pictures quality metric) is proposed. It showed to be consistent with subjective

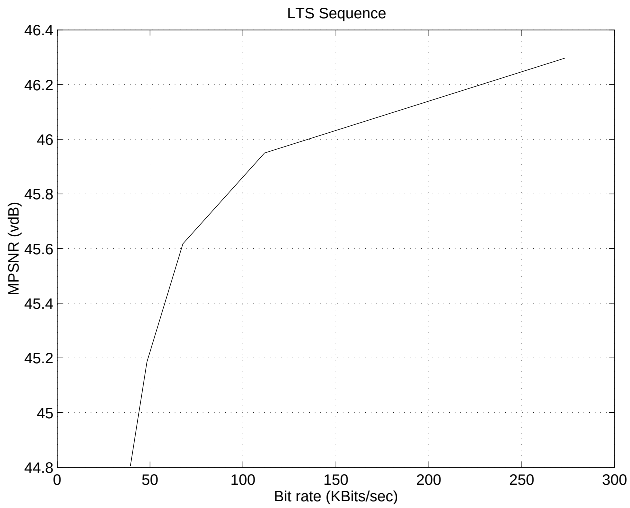

Figure 12: MPSNR quality assessment for the LTS sequence as a function of the bitrate.

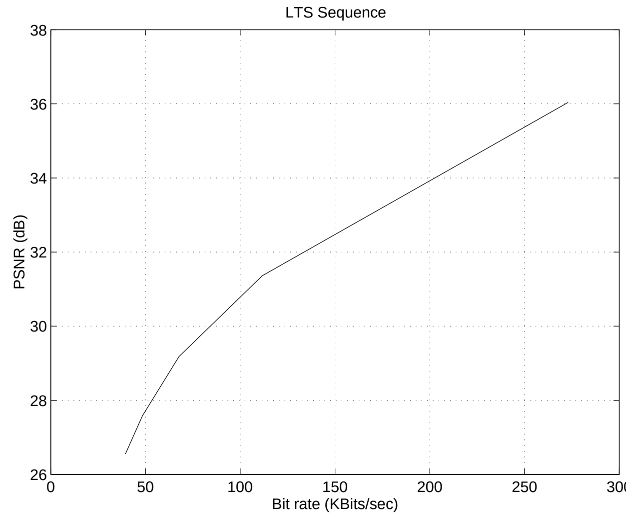

Figure 13: PSNR quality assessment for the LTS sequence as a function of the bitrate.

rating of coded video. Other metrics are proposed. They are meant to evaluate the rendition of basic features in a video sequence, namely contours, textures and uniform areas.

7 REFERENCES

[1] Christian J. van den Branden Lambrecht, Vasudev Bhaskaran, Al Kovalick, and Murat Kunt. "Automatically Assessing MPEG Coding Fidelity". *IEEE Design and Test Magazine*, pp. 28–33, winter 1995.

[2] P.H. Meehan, R.N. Hurst, M.A. Isnardi, and P.K. Shah. "MPEG Compliance Bitstream Design". In *Proceedings of the Consumer Electronics Society of the IEEE*, pp. 316–319, Chicago, IL, June 7-9 1995.

[3] J. L. Mannos and D. J. Sakrison. "The effects of a visual fidelity criterion on the encoding of images". *IEEE Transactions on Information Theory*, Vol. IT-20, No. 4, pp. 525–536, 1974.

[4] J. A. Saghri, P. S. Cheatham, and A. Habibi. "Image quality measure based on a human visual system model". *Optical Engineering*, Vol. 28, No. 7, pp. 813–818, 1989.

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

[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*, Atlanta, GA, May 7-10 1996. accepted for publication.

[7] R. L. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1988.

[8] J. G. Daugman. "Spatial Visual Channels in the Fourier Plane". *Vision Research*, Vol. 24, pp. 891–910, 1984.

[9] J. G. Daugman. "Two-Dimensional Spectral Analysis of the of Cortical Receptive Field Profiles". *Vision Research*, Vol. 20, pp. 847–856, 1980.

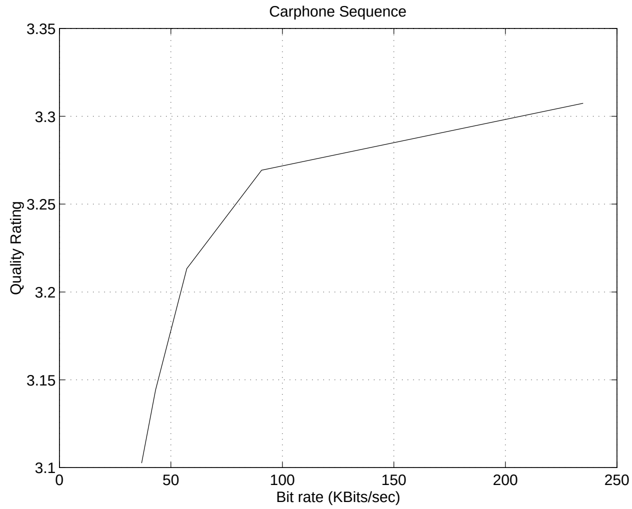

Figure 14: Quality rating for the Carphone sequence as a function of the bitrate.

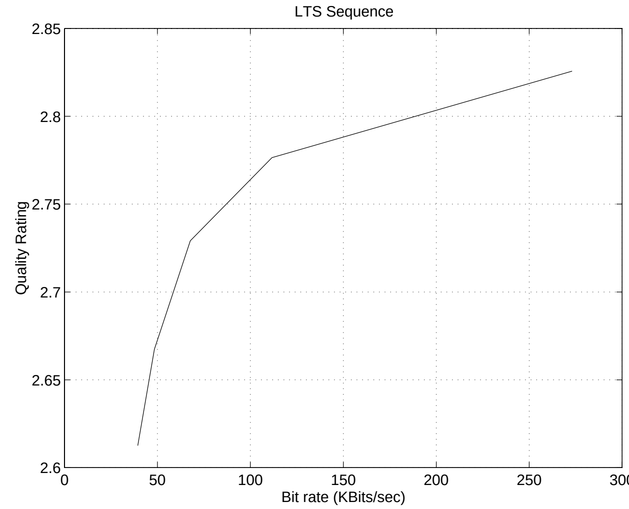

Figure 15: Quality rating for the LTS sequence as a function of the bitrate.

[10] Andrew B. Watson. *Handbook of Perception and Human Performance*, Vol. 1, Sensory Processes and Perception, Chapter 6, Temporal Sensitivity. John Wiley, 1986.

[11] R. F. Hess and R. J. Snowden. "Temporal Properties of Human Visual Filters: Number, Shapes and Spatial Covariation". *Vision Research*, Vol. 32, No. 1, pp. 47–59, 1992.

[12] Christina A. Burbek and D. H. Kelly. "Spatiotemporal Characteristics of Visual Mechanisms: Excitatory-Inhibitory Model". *Journal of the Optical Society of America*, Vol. 70, No. 9, pp. 1121–1126, September 1980.

[13] S. T. Hammett and A. T. Smith. "Two Temporal Channels or Three? A Re-evaluation". *Vision Research*, Vol. 32, No. 2, pp. 285–291, 1992.

[14] J. A. Solomon and A. B. Watson. "Spatial and Spatial Frequency Spreads of Masking: Measurements and a Contrast-Gain-Control Model". on http://vision.arc.nasa.gov, 1995.

[15] 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.

[16] Christian J. van den Branden Lambrecht and Murat Kunt. "Characterization of Human Visual Sensitivity for Video Imaging Applications". *Signal Processing*. submitted paper.

[17] J. C. Falmagne. *Handbook of Perception and Human Performance*, Vol. 1, Sensory Processes and Perception, Chapter 1, Psychophysical Measurement and Theory. John Wiley, 1986.

[18] J. L. Hall. "Hybrid Adaptive Procedure for Estimation of Psychometric Functions". *Journal of the Acoustical Society of America*, Vol. 69, No. 6, pp. 1763–1769, June 1981.

[19] O. Egger, W. Li, and M. Kunt. "High Compression Image Coding Using an Adaptive Morphological Subband Decomposition". *Proceedings of the IEEE,* Special Issue on Advances in Image and Video Compression, Vol. 83, No. 2, pp. 272–287, February 1995.

[20] 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.

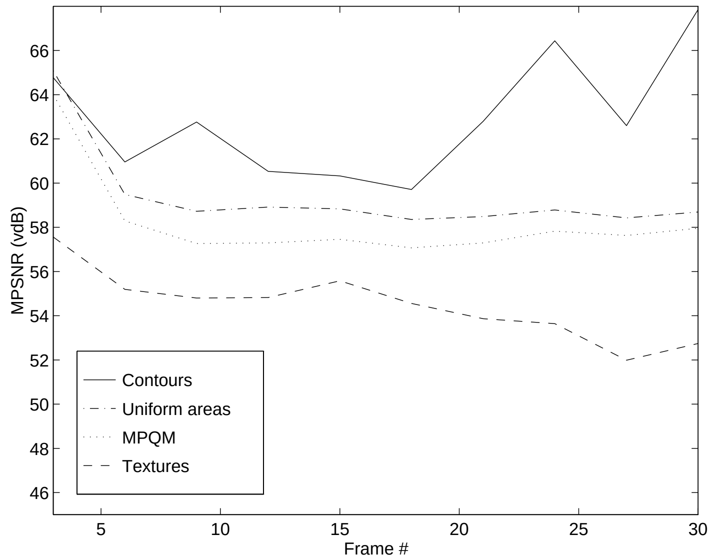

Figure 16: Detailed metrics for the edge rendition synthetic test sequence. Dotted line is the MPQM, solid line is contour rendition, dashed line texture rendition and dot-dashed line uniform areas.

[21] S. Comes and B. Macq. "Human Visual Quality Criterion". *SPIE Visual Communications and Image Processing*, Vol. 1360, pp. 2–7, 1990.

[22] ISO/IEC JTc1/SC29/WG11 13818. "Generic Coding if Moving Pictures and Associated Audio; Part 2: Video". Technical report, International Standardization Organization/International Electrotechnical Commission, Geneva, Switzerland, 1995.

[23] ITU-T Study Group 15. "ITU-T Draft H.263, Line Transmission of Non-Telephone Signals — Video Coding for Low Bitrate Communications". Technical report, July 1995.

[24] Chadd Fogg. "mpeg2encode/mpeg2decode version 1.1. available via anonymous ftp at netcom.com, June 1994. MPEG Software Simulation Group.

[25] Andrea Basso, İsmail Dalgıç, Fouad A. Tobagi, and Christian J. van den Branden Lambrecht. "Study of MPEG-2 Coding Performance based on a Perceptual Quality Metric". In *Proceedings of the Picture Coding Symposium*, Melbourne, Australia, 1996. accepted for publication.

[26] Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson, Stephen D. Voran, and Stephen Wolf. "An objective video 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, February 1993.

[27] Karl Olav Lillevold. "H.263 compatible Video Encoder-Decoder version 1.3. available on http://www.nta.no/brukere/DVC/, July 1995. Digital Video Coding at Telenor Research.

[28] A. Nicoulin. "The LTS/EPFL video sequence for very low bit rate coding". Technical Report 11, Swiss Federal Institute of Technology-LTS, September 1994.