A FEEDBACK CONTROL SCHEME FOR LOW LATENCY CONSTANT QUALITY MPEG-2 VIDEO ENCODING *

Andrea Basso^{*,**} İsmail Dalgıç^{*} Fouad A. Tobagi^{*} and
Christian J. van den Branden Lambrecht^{***}

** Department of Electrical Engineering, Stanford University*

*** Telecommunication Laboratory, Swiss Federal Institute of Technology*

**** Signal Processing Laboratory, Swiss Federal Institute of Technology*

ABSTRACT

This paper describes a coding control scheme for MPEG-2 which maintains the perceived video quality constant, and is suitable for real-time and low latency encoding. We have chosen a Proportional Integral Derivative (PID) scheme for the controller similarly to the works presented in^{1,2}, and we adopt the same approach in designing the particular PID feedback function. The main reason for using a PID feedback function is the good tradeoff that it offers between computational complexity, ease of design, and performance. We have chosen to use a new video quality metric called Moving Picture Quality Metric (MPQM). This metric models the human visual system and matches subjective evaluations correctly, outperforming existing quality metrics for video.³ Simulations results will be shown for typical video sequences. A comparison with CBR encoding will be also presented.

1 INTRODUCTION

The concept of constant quality variable bit rate encoding (CQ-VBR) has been introduced for the first time in previous works,^{1,2} where a feedback-controlled encoder is proposed which maintains the perceived video quality constant. This scheme is based on a PID (Proportional, Integral, Derivative) control, and uses a quantitative quality measure developed at the Institute for Telecommunication Sciences.⁴ One of the important reasons for using a feedback-control based scheme is to achieve real-time video encoding with low latency. Another scheme for achieving constant video quality, which is based on iteratively changing the encoder parameters until the desired quality is achieved, is presented in.⁵ The iterative nature of this approach makes it suitable mainly for non-real-time video encoding. The above mentioned schemes focus on H.261 and MPEG-1 encoding algorithms. The objective of this paper is to develop a coding control scheme for MPEG-2 which maintains the perceived video quality constant, and is suitable for real-time and low latency encoding. We have chosen a PID scheme similarly to previous works^{1,2} for the controller, and we follow the same approach in designing the PID feedback function.

^{*}This work was supported in part by NSF under grant NCR-9016032, and by Pacific Bell.

The main reason for using a PID feedback function is the good compromise that it offers between computational complexity, ease of design, and performance. Our approach differs from previous works^{1,2} in several important aspects. First of all, we have chosen to use a new video quality metric called Moving Picture Quality Metric (MPQM), which is more applicable to MPEG-2. This metric models the human visual system and matches subjective judgments correctly, outperforming existing quality metrics for video.³ Secondly, MPEG-2 has more parameters that can be controlled. Although we have for now limited the set of controlled parameters to the quantizer scale (MQANT), we will investigate the effect of using other parameters. The paper is organized as follows. In Section 2 a brief state of the art in video quality assessment and measurement is presented. In Section 3 the Moving Pictures Quality Metric (MPQM) is defined. In Section 4 a review of the fundamentals of the MPEG-2 coding standard will be discussed. In Section 5 the design of the PID controller is presented, and in Section 6 numerical results regarding the quality and traffic for the CQ-VBR MPEG-2 scheme are discussed. Finally, in Section 7 some concluding remarks are given.

2 STATE OF THE ART IN VIDEO QUALITY ASSESSMENT AND MEASUREMENT

The interest for image quality evaluation has been strong since the sixties. Several image quality metrics have been developed, such as the Strehl measure,⁶ based on measurements of the image degradation to which an image is subject to, whenever it passes different real optical systems. A good review of the attempts with bivariate metrics of image quality done during the seventies is presented in.⁷ The mean square error has been retained for its property of being easily analytically tractable. Wilder⁸ did a rather complete evaluation of different mean-square error metrics including power-law, logarithmic, gradient and Laplacian transformations. More recently international committees, such as CCIR and MPEG, have devoted a lot of effort in the study of video quality assessment.⁹⁻¹² We can summarize the recent efforts in image and video quality assessment in two different lines of thought.

Metrics Based On Subjective Data Such metrics are based on linear or non-linear combinations of distortion measures applied to features extracted from images and video. Such features include false edge effect, blocking effect, blurred edges effect and temporal distortions. The coefficients of the linear or non-linear combinations are chosen in order to maximize the correlation with subjective tests. Pioneer on this area is Miyahara¹³ with a method that extracts five features and on this basis computes a simple model which is then tuned to maximize correlation with subjective tests. On the same line F. Lin *et al.*,⁵ Davies¹⁴ and Webster⁴ have developed a non-linear neural network model for image quality assessment. On the same principles a quantitative video quality metric has been designed by Wolf *et al.*¹⁵ at Institute for Telecommunication Sciences (ITS) in Colorado, USA.

Metrics based on Vision Models Such metrics are based on physiological evidences. The metric mimics the human evaluation by modeling closely the initial stages of the human visual system (HVS). Measures based on physiological evidences have been conducted by Mannos and Sakrison, with particular focus on mean square error quality criterions for monochrome images on the basis of a simple model of the human visual system. They are discussed in the Rec.500-3 from CCIR⁹ on subjective assessment of the quality of television pictures. Attempts to predict the response of the cortical neurons have been carried out in studies of perception. De Valois¹⁶ conducted electro-physiological experiments showing that the cells of the primary visual cortex are tuned to bands in spatial frequency and orientation. Physiological experiments conducted by Daugman¹⁷ confirmed the presence of several mechanisms of vision that divide the frequency plane into frequency and orientation bands. Several models have been developed which make use of masking, contrast sensitivity, and a multichannel structure.¹⁸⁻²¹ The research activities of Heeger,²⁰ Daly,¹⁸ Watson,¹⁹ van den Branden,²² and Western *et al.*²³ are the first efforts in the development of metrics based on these concepts.

3 THE VIDEO QUALITY METRIC (MPQM)

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of human vision.^{24,25} A spatio-temporal model of human vision has been developed for the assessment of video coding quality.^{26,27} The model is based on the following properties of human vision:

- The responses of the neurons of the primary visual cortex (called area V1) are band-limited. The human visual system has a collection of mechanisms or detectors (called channels) that mediate perception. A channel is characterized by a localization in spatial frequency, spatial orientation and temporal frequency. Such channels are simulated by a three-dimensional filter bank.
- In a first approximation, those channels can be considered to be independent. Perception can thus be predicted channel by channel without interaction.
- Perception in a channel is characterized by two phenomena: contrast sensitivity and masking. Human sensitivity to contrast is known to be a function of frequency (spatial and temporal) as well as orientation. This leads to the concept of *contrast sensitivity function*, which specifies the threshold of detection for a stimulus as a function of frequency. Masking accounts for inter-stimuli interferences. It is known that the presence of a background stimulus will modify the detection of a foreground stimulus. Masking thus corresponds to a modification of the detection threshold according to the local contrast of the background.

The working model described in²⁶ incorporates the above described considerations of the HVS. The filter bank used in the model decomposes the data according to 5 spatial frequency bands, 4 orientation bands and 2 temporal frequency bands. It has been especially parameterized for the framework of video coding by means of psychophysical experiments.

Such a model predicts the response of the neurons in area V1 of the cortex and thus the *perceived* distortion. This is done by decomposing the data in perceptual channels and predicting perceived stimuli using contrast sensitivity and masking. Thereafter, a distortion measure is computed, accounting for the higher levels of cognition of the brain. At this stage, the metric also accounts for the focus of attention and 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. The spatial dimension corresponds to the 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 M_E for a given block is computed according to Eq. 1:

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

where $\epsilon[x, y, t, c]$ 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 channels and $N_q = N_x N_y N_t$. The exponent of the Minkowski summation is β and has a value of 4.

In this application, the error measure M_E is further mapped onto a quality scale from 1 to 5 according to the following function, relating the error measure to the quality index MPQM:

$$\text{MPQM} = \frac{5}{1 + N M_E},$$

where N ensures a mapping between 1 and 5 as defined by the CCIR.⁹ The meaning of the scale is the following: 5, imperceptible impairment equivalent to an excellent quality, 4 perceptible, but not annoying impairment

equivalent to a good quality, 3 slightly annoying impairment equivalent to a fair quality, 2 annoying impairment equivalent to a poor quality and 1 very annoying impairment equivalent to a bad quality. The free parameter N has been estimated on the basis of the vision model²⁷ and has a value of $N = 0.623$. Further details on how MPQM has been calibrated to this scale can be found in .²²

Results of the application of MPQM to well known test sequences are depicted in Fig. 1. It can be seen that the metric is able to detect the saturation in quality that occurs at high bit rates as observed in subjective quality assessment. At lower bit rates, the metric exhibits a behavior that matches correctly human judgment.¹²

Figure 1: MPQM as a function of the bitrate for MPEG-2 encoding of the Basketball and Mobile & Calendar sequences.

4 BRIEF OVERVIEW OF MPEG-2 CODING STANDARD

MPEG-2 is a standard developed by ISO/IEC/JTC1/SC29/WG11 and known in literature as ISO/IEC 13118.^{28,29} It was formally accepted as international ISO standard in November 1994. It provides a set of methodologies to perform compression as well as the transport and multiplexing of video and audio.

4.1 Coding

The MPEG-2 coding method is based on a DCT and motion compensated intra-frame and inter-frame scheme which operates on sub-parts of the images called macroblocks. The standard defines 3 types of pictures: Intra-coded (I-frames), Predictive coded (P-frames) and Bidirectionally predictive pictures (B-frames). Such frames can be organized in Group of Pictures (GOP). I-frames are coded without motion compensation and prediction and are considered as a reference for the coding of the subsequent P- and B-frames. Five layers are defined in MPEG-2; the block layer, the macroblock layer, the slice layer, the picture layer and the sequence layer.

- **Block layer** MPEG-2 partitions a given input color image in *blocks*. This is smallest coding unit in MPEG-2. It is constituted by a square matrix of 8 by 8 picture elements of the input frame. Pixels can be of three types. Luminance (Y), red chrominance (Cr) or blue chrominance (Cb) pixels.
- **Macroblock layer** Blocks are organized in *macroblocks* that are the smallest unit that the MPEG-2 coding algorithm processes. A macroblock is characterized by 16 by 16 picture elements. In MPEG-2

the chrominance components have half vertical and half horizontal resolution of the luminance component. Thus a macroblock is composed by 4 luminance blocks, one red chrominance and one blue chrominance block. Macroblocks can be intra-coded or inter-coded. Each intra-coded macroblock is processed with the Discrete Cosine Transform (DCT), the zig-zag scanning, adaptive quantization and variable length coding. The inter-coded macroblocks (in P- and B-frames) use motion vectors, which are estimated from a previous picture (P-frames) or both previous and next picture (B-frames). Two motion vectors are present for each inter-coded macroblock in P-frames. Up to 4 macroblock vectors are used for inter-coded macroblocks in B-frames. For inter-coded pictures only the prediction error is coded in intra-mode.

- **Slice layer** An horizontal strip of contiguous macroblocks within a frame is called a *slice*. A slice is the smallest independently codable unit. Coding operations on blocks and macroblocks can be performed only when all the elements of a slice are available. Slices can have different lengths. Typical slice sizes are 11, 22, 44 macroblocks.
- **Picture layer** The picture is the basic display unit. It corresponds to a frame of an input video sequence. The picture size depends on the application and it is 720 by 576 (in Europe) or 720 by 480 (in USA) [ML@MP].
- **Sequence layer** The sequence layer includes all the frames forming a given input sequence in addition to the global characteristics of the sequence such as frame size, format, etc. An MPEG-2 video or audio coder produces as output a coded video or audio *elementary stream* respectively.

4.2 Constant Bit Rate Control

In this paper, we use for the CBR simulations the rate control algorithm proposed in TMN-5,³⁰ a reference implementation of the MPEG-2 coding standard, is articulated in three main steps:

- **Step One: Target Bit Allocation.** An initial *target bit allocation* is estimated on the basis of a priori computed weights. These weights, indicated with X_i, X_p, X_b , reflect the typical coded frame size of I, P, and B pictures. I pictures are usually assigned the largest weight since they have the greatest stability factor in an image sequence and contain the most new information in a sequence. B pictures are assigned the smallest weight since B energy do not propagate into other pictures and are usually more highly correlated with neighboring P and I pictures than P pictures are. The bit target for a frame is based on the frame type, the remaining number of bits left in the Group of Pictures (GOP) allocation, and the immediate statistical history of previously coded pictures. On the basis of X_i, X_p, X_b an initial quantization factor Q is computed. Such value is identical for all the macroblocks.
- **Step Two: Rate control via buffer monitoring.** Before the encoding of the i th macroblock the virtual buffer occupancy is computed by means the following expression:

$$O_i^{p_{type}} = O_0^{p_{type}} + B_{i-1} - \frac{T_{p_{type}} \cdot (i - 1)}{N_{mb}} \quad (2)$$

where $O_0^{p_{type}}$ is the initial buffer occupancy for the picture p_{type} i.e. I, P or B, B_{i-1} is the number of bits necessary for the encoding of all the macroblocks of the picture up-to and not including the macroblock i , N_{mb} is the total number of macroblocks in the picture and $T_{p_{type}}$ is the number of bits allocated a priori for all the macroblocks of the frame p_{type} . The rate control mechanism forces an adjustment of the bit allocation if the difference between the target bits and the actual coded bits for a block of data is significant. If overflows occur (signaled by the virtual buffer verifier) the macroblock quantization step-size is increased. Similarly if underflows occur then the quantization step size is decreased.

- Step Three: Adaptive Quantization. A modulation of the resulting quantization factor is performed on the basis of a macroblock based activity measure applied to the luminance component only. The activity act_i for a macroblock i is chosen as the minimum among its four 8x8 block luminance variances var_{si} :

$$act_i = 1 + \min(var_{si}) \quad (3)$$

The activity measure itself is normalized against the most recently coded picture of the same type (I, P, or B). Let's call it act_i^N . The final quantization scale factor $MQUANT_i$ is obtained by modulation of the quantization factor Q according to the following expression:

$$MQUANT_i = Q * act_i^N \quad (4)$$

5 A CONSTANT PERCEIVED QUALITY FEEDBACK CONTROL SCHEME

Several control strategies are available to the designer of digital control systems. The more classical are proportional derivative and integral controllers. In proportional control the slope of reaction curve can never be adequately fixed. In relation to video, a small slope induces a nearly constant image quality but increases the bit-rate variation. If the slope is high the system tends to overreacts and induces a high variation in image quality. Furthermore the proportional control implies always a steady state error. The proportional integral derivative (PID) control instead allows to impose the steady-state responses to specified signals such as impulse, step and slope with an error equal to zero. Furthermore it guarantees a return to a steady state.

Figure 2: MPEG-2 Constant Quality encoder. Note that the MPEG-2 encoder does not include any rate control algorithm i.e. is an open loop.

Finally, transient responses with a given time-constant can be imposed. The PID controller is thus a good tradeoff between performance and design complexity. The block diagram of the proposed CQ-VBR coder is presented in Fig. 2. The PID controller is of the form:

$$Q(k+1) = Q(0) + K_p c(k) + K_p \sum_{i=1,k} \frac{T}{T_I} c(i) + K_p \frac{T_D}{T} [c(k) - c(k-1)], \quad (5)$$

where $c(k)$ is defined as:

$$c(k) = MPQM(k) - MPQM_{target}, \quad (6)$$

and $Q(k)$ is the quantization factor. The first term of Eq. (5) is proportional to the error and the second and third terms are respectively proportional to the integral and to the derivative of the error. The design variables are K_p, T_I and T_D and the estimation interval w . The Ziegler-Nichols PID tuning using the stability method has been employed to determine these values. The method works as follows. First the control applied is only proportional. The value of gain K_p is increased until continuous oscillations occur. Let K_u be this value and let P_u be the oscillation period. The PID gains are determined by the following expressions: $K_p = 0.6K_u, T_I = P_u/2$, and $T_D = P_u/8$.

In the MPEG-2 coding scheme the frame coding order differs from the frame display order. This is due to the presence of bidirectionally interpolated frames (B-frames), whose decoding depends on future frames. This limits the maximal sampling rate to one third of the frame rate in a *IBBPBBP...* GOP structure. Thus the sampling period T has been chosen at $T = 120\text{ms}$ equivalent to 3 frames time slot. Such interval is equal to the temporal window size of MPQM. In our experiments we obtained $K_u = 16$ and $P_u = 360\text{ms}$ (9 frames).

The CQ-VBR encoder operates as follows. The coded video sequence is sampled with a period of 3 frames and a quality measurement is performed on the 3 frames. The quantization factor Q is then computed according to Eq. 5. Note that all the macroblocks are quantized with the same Q and no further adaptive quantization is performed as in the Constant Rate model presented in subsection 4.2. In the MPEG-2 encoding process the GOP structure *IBBPBBPBBPBBI..* of size of 12 frames has been used.

6 EXPERIMENTAL RESULTS

Video material The video material considered for the experiments includes 105 frames of size 512 by 512 pixels of the *Flower Garden* MPEG-2 test video sequence and 300 frames of the sequence *Ski* of size 512x512. The sequence flower garden is a well known sequence with panning and no scene changes. The sequence ski includes 3 scene changes and fast motion, panning and zooming.

CQ-VBR encoding results In Fig. 3 we plot the the quality rating of MPQM for the sequence Flower Garden. The quality was regulated around a value of 4.3. The average MPQM is 4.23 with a small variance of the order of $1.5 \cdot 10^{-3}$. The minimum MPQM is 4.12 while the maximum is 4.31. Note as expected the bursty behavior of the bitrate shown in Fig. 4 (left side). The ratio between the smallest and the largest frame is 5.5 and peak-to-average frame size ratio of 3.1. The mean bitrate is 5Mb/s . The time evolution of the quantization factor is shown in Fig. 4 (right side). The mean quantization factor is 13.25 while the minimum is 8 and the maximum is 16.

Comparison with CBR encoding We coded the sequence ski with the CQ-VBR and CBR by means of an MPEG-2 encoder operating with the Constant Rate control mechanism described in section 4.2. The chosen rate for the CBR is 3.5Mb/s equal to the mean bitrate of the CQ-VBR simulation. The GOP structure is the same used in the CQ-VBR experiments. The coding has been performed with the software simulator of the test model 5 made available by the MPEG software simulation group.³⁰

In Fig. 5 and Fig. 7 we plot the quality ratings for the CQ-VBR and the CBR respectively. As expected the CBR coding gives worse performance in terms of video quality and temporal video quality variation. A comparison between the values of MPQM for CBR and CQ-VBR is given in Table 1. Note in particular the difference in the values of the variance of the MPQM for the two simulations. The higher difference between the CQ-VBR minimum and the maximum MPQM compared to the flower garden sequence is expected due to the higher complexity and the presence of scene changes in the sequence ski.

Figure 3: Constant Quality (CQ-VBR) encoding for the sequence Flower Garden (105 frames).

Figure 4: CQ-VBR coding of the Flower Garden sequence. Left: bits per frame. Right: MQANT evolution.

	MEAN MPQM	VAR	MIN MPQM	MAX MPQM
CBR	3.12	0.19	2.34	3.76
CQ-VBR	3.39	0.0015	3.84	4.02

Table 1: Comparison between quality ratings of the CBR and CQ-VBR.

Figure 5: Constant Quality (CQ-VBR) encoding for the sequence Ski (300 frames).

Figure 6: CQ-VBR coding of the ski sequence. Left: bits per frame. Right: MQQUANT evolution.

	MEAN FRAME SIZE	VAR	MIN FRAME SIZE	MAX FRAME SIZE
CBR	1.39E5 bits	2.34E9	4.8E4 bits	3.95E5 bits
CQ-VBR	1.41E5 bits	2.2E9	3.5E4 bits	2.55E5 bits

Table 2: Comparison between CBR and CQ-VBR frame sizes.

Figure 7: CBR coding of the ski sequence. Evolution of the quality rating.

In Fig. 6 (left side) and Fig. 8 (left side) the bitrate and the evolution in time of the quantization factor are shown. Note that the `MQUANT` for the CBR simulation is the average of the different `MQUANT` values assigned to the macroblocks of a given frame. A comparison between the values of frame sizes for CBR and CQ-VBR is given in Table 2. Note that the different temporal evolution of the two bitrates in Fig. 6 (left side) and Fig. 8 (left side) is characterized by similar statistical first and second moment.

7 CONCLUSIONS

In this paper we have presented a constant quality variable bit-rate encoder (CQ-VBR) based on MPEG-2 coding scheme. The control method is based on PID feedback control which operates on the quantization scale factor of the MPEG-2 encoder and a new video quality metric based on a spatio-temporal vision model. Coding results for known test sequences have been presented showing the effectiveness of the control scheme. A comparison with MPEG-2 CBR coding has been presented showing that the CQ-VBR scheme is able to provide a significant improvement of the mean image quality for the same mean bitrate compared to the CBR.

8 ACKNOWLEDGMENTS

The authors acknowledge the Centro Ricerche RAI Radio Televisione Italiana for making available the video material used in the experiments.

Figure 8: CBR coding of the ski sequence. Left: bits per frame. Right: MQANT evolution.

9 REFERENCES

- [1] I. Dalgıç and F.A. Tobagi. “Constant quality video encoding”. *ICC-95*, Vol. II, pp. 1255–1261, June 18-22 1995.
- [2] I. Dalgıç and F. A. Tobagi. “A constant quality MPEG-1 video encoding scheme and its traffic characterization”. *Picture Coding Symp.1996*, pp. 105–110, March 13-15 1996.
- [3] 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*, pp. 263–268, Melbourne, Australia, March 13-15 1996. available on http://ltswww.epfl.ch/pub_files/vdb/.
- [4] A. Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf. “An Objective Video Quality Assessment System Based on Human Perception”. *Human Vision, Visual Processing and Digital Display*, Vol. SPIE-1913, pp. 15–26, 1993.
- [5] F. Lin and R. Mersereau. “A constant subjective quality MPEG encoder”. *ICASSP*, pp. 2177–2180, 1995.
- [6] W. K. Pratt. *Digital Image Processing*. Wiley and Sons NY (USA), 1974.
- [7] T. W. Barnard. “A Literature Survey on Image Quality Evaluation”. *The Perkin Elmer Corporation, Engineering Report ER-177*, 1971.
- [8] W. C. Wilder. “Subjectively Relevant Error Criteria for Pictorial data Processing”. *Purdue University, Engineering Report TR-EE 72-34*, Dec. 1972.
- [9] CCIR recomm. 500-3. “Method for the Subjective Assessment of the Quality of Television Pictures”.
- [10] CCIR rep. 1082-1. “Studies toward the Unification of Picture Assessment Methodology”.
- [11] CCIR recomm. 813. “Methods for Objective Quality Assessment in relation to Impairments from Digital Codig of Television Signals”.
- [12] M. Ardito, M. Barbero, M. Stroppiana, and M. Visca. “Compression and Quality”. *HDTV-94*, pp. B-8-2, Oct. 1994.

- [13] M. Miyahara, K. Kotani, and V. R. Algazi. "Objective Picture Quality Scale (pqs) for Image Coding". *SID digest*, pp. 859–862, 1992.
- [14] I. Davies and D. Rose. "Automated Image Quality Assessment". *Human Vision, Visual Processing and Digital Display*, Vol. SPIE-1913, pp. 27–36, 1993.
- [15] S. Wolf. "Features for Automated Quality Assessment of Digitally Transmitted Video". *U.S. dept. of commerce, Nat. Telecomm. and Inf. Adm. report*, pp. 90–264, June 1990.
- [16] R. L. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1988.
- [17] J. G. Daugman. "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 7, pp. 1169–1179, July 1988.
- [18] Scott Daly. "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity". In *Proceedings of SPIE*, Vol. 1616, pp. 2–15, 1992.
- [19] 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.
- [20] 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.
- [21] Serge Comes. *Les traitements perceptifs d'images numérisées*. PhD thesis, Université Catholique de Louvain, 1995.
- [22] Christian J. van den Branden Lambrecht. *Perceptual Models and Architectures for Video Coding Applications*. PhD thesis, Swiss Federal Institute of Technology, CH 1015 Lausanne, Switzerland, 1996. available on http://ltswww.epfl.ch/pub_files/vdb/.
- [23] S. Western, K.L. Lagendijk, and J. Biemond. "Perceptual Image Quality based on a Multiple Channel HVS Model". *ICASSP*, pp. 2351–2354, 1995.
- [24] J. L. Mannon 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.
- [25] Andrew B. Watson. "Efficiency of an Image Code Based on Human Vision". *Journal of the Optical Society of America*, Vol. 4, No. 12, pp. 2401–2417, December 1987.
- [26] 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*, 1996.
- [27] 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 IS&T Symposium on Electronic Imaging: Science and Technology*, editors, *Digital Video Compression: Algorithms and Technologies 1996*, San Jose, CA, January 1996. The Society for Imaging Science and Technology.
- [28] ISO/IEC JTC1/SC29/WG11 13818. "Generic Coding of Moving Pictures and Associated Audio; Part 2: Video". Technical report, International Standardization Organization/International Electrotechnical Commission, Geneva, Switzerland, 1995.
- [29] ISO/IEC JTC1/SC29/WG11 13818. "Generic Coding of Moving Pictures and Associated Audio; Part 3: System". Technical report, International Standardization Organization/International Electrotechnical Commission, Geneva, Switzerland, 1995.
- [30] C. Fogg. "mpeg2encode/mpeg2decode version 1.1, June 1994.