Impairment Metrics for MC/DPCM/DCT Encoded Digital Video

H. R. Wu[1], Z. Yu[2], S. Winkler[3] and T. Chen[4]

[1] School of Computer Science and Software Engineering, Monash University, VIC 3800, Australia
[2] Motorola Australian Research Centre, Sydney, NSW 2019, Australia
[3] Genimedia, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland
[4] Sarnoff Corporation, Princeton, NJ 08543-5300, U.S.A.

ABSTRACT

Two impairment metrics presented in this paper are for quantification of blocking artifacts and ringing artifacts, respectively, in digitally coded monochrome video sequences. They are based on a multichannel vision model which has been parameterised using the subjective quality assessment data recently provided by the VQEG (Video Quality Experts Group). Segmentation algorithms are used to identify regions dominated by blocking and ringing, respectively, and perceptual distortions in these regions are summed up to form perceptual distortion metrics. As an example, a perceptual blocking distortion metric (PBDM) is presented which is based on a simplified distortion detection model. Subjective and objective tests have been conducted, and the results show a strong correlation between the objective blocking ratings and the mean opinion scores on blocking artifacts.

1. INTRODUCTION

International standardisation activities have resulted in a series of international video and associated audio coding standards, and have led to a proliferation of applications in video communications, digital television, multimedia computing, etc. Various video coding standards have adopted a hybrid of motion compensated temporal differential pulse code modulation and the block discrete cosine transform (hybrid MC/DPCM/DCT) algorithm [1]. Whilst the coding algorithm exploits statistical and psychovisual redundancies of input video sequences to achieve a low bit rate, it will cause visible coding distortions in reconstructed sequences due to its lossy coding nature. In order to evaluate, monitor and improve the coding system performance, it is imperative to develop quantitative digital video quality/impairment metrics. It is well known that the traditional objective measures such as the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) do not always correlate well with the perceived digital video picture quality [2], nevertheless they are still used frequently to evaluate the quality of digital video.

People have long been investigating objective quality assessment methods. The first Human Visual System (HVS) based quantitative measure of video quality, was proposed by Lukas and Budrikis in 1982 [3]. Recently this has become a very active research topic as reflected by a number of quality metrics proposed [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. The current VQEG activities represent international standardisation efforts towards an objective video quality metric, with delegations from ITU-T Study Groups 9 and 12 and ITU-R Study Group 11 [17].

Existing quality metrics can be categorised mainly into two groups: vision model based and feature extraction based. In general, an overall index to the quality of a given video sequence is assigned by each of the metrics.

Advances in vision research have provided crucial information on the structure and the working mechanism of the human vision system, which have been adopted to design quality metrics [8], [9], [11], [12]. Most current psychovisual quality metrics share the commonality of being based on multichannel vision models [18]. With these metrics a perceptual distortion map can be generated for every spatial location.

Digital video coding distortions have been well understood and classified [19], [20], such as blocking, ringing, blurring, etc. For many applications of digital video, it is highly desirable that we are able not only to give the overall distortion measure, but also to describe the type of distortions as well as the quality degradation caused by each type of distortion, so that we can measure the system performance and make improvements accordingly.

A number of researchers have already addressed blocking impairment metrics, but mainly for still images [21], [22], [23]. A blocking impairment metric for video sequences is proposed in [24], which only acts as a building block for a single-ended quality metric, rather than a stand-alone blocking impairment metric, and is not based on vision models. In general, current achievements in vision research and quality metrics have not yet been reflected in quantifying blocking artifacts for digital video. So far, there has been little work done on ringing impairment metrics for digital video.

2. FORMULATION OF VIDEO IMPAIRMENT METRICS BASED ON A VISION MODEL

Modelling human vision has long been a challenging research task [25], [26]. In this section, our aim is to devise a distortion detection model on which digital video impairment metrics can be built. The main challenge here is to find whether two images

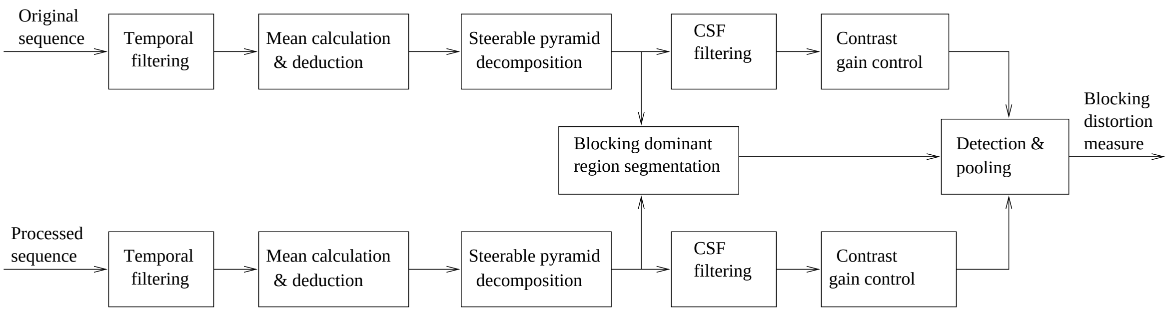


Fig. 1. Block diagram of the perceptual blocking distortion measure.

are similar or dissimilar. Therefore, most of our attention is focused on the pattern sensitivity aspect of human vision.

The impairment metrics investigated in this paper are based on a model firstly introduced by Teo and Heeger [5] and later extended by van den Branden Lambrecht [8] and Winkler [9].

A number of simplifications have been introduced to reduce the computation complexity of the implementation of the model, including a first-order IIR temporal filter to model sustained channel, calculation of mean pixel value, subtraction of the mean pixel value from each pixel, steerable pyramid decomposition [27], CSF filtering and contrast gain control.

The model described in this paper is designed for monochrome video sequences and the parameterisation was carried out using the VQEG subjective test data.

Segmentation algorithms are used to identify blocking and ringing dominant regions, respectively, based on information obtained after spatio-temporal decomposition. For instance, horizontal or vertical edges can be identified as having a similar waveform to that of the step response of a highpass filter after the decomposition. Blocking regions are segmented out by removing edges not related to blocking such as edges in the original sequence, too short to be considered as block-edge, etc. Strong ringing artifacts occur along the high-contrast edges of objects, which will be reflected as strong oscillations in the highpass channel of the original sequence. The segmentation algorithm generates ringing region map based on this observation.

The detection and pooling are carried out after contrast sensitivity function filtering and contrast gain control operations, only in the blocking/ringing artifact dominant regions.

3. PERCEPTUAL BLOCKING DISTORTION METRIC

Blocking artifacts are defined as discontinuities found across block boundaries [19]. A perceptual blocking distortion metric (PBDM) is devised in this section based on the above mentioned distortion detection model to calculate perceptual distortions, as shown in Fig. 1. The blocking distortion d can be converted to the Objective Blocking Rating (OBR) with the following empirical formula:

$$OBR = 5 - d^{0.6} \tag{1}$$

where d is the summed distortion which is averaged by the number of frames, and the OBR is on a scale of 1 to 5. The exponent (0.6) is determined experimentally.

Both subjective and objective tests have been conducted and the correlations between the subjective and objective data have been calculated to evaluate the performance of the proposed blocking impairment metric compared with that of the PSNR.

The test scenes were selected from two sources: an ANSI T1A1 data set and the VQEG data set. Bit rates were selected so that the generated sequences covered a full range of impairment. The video encoder used is a software simulator of the MPEG-2 Test Model 5 (TM5) [30]. The subjective tests followed the Double-Stimulus Impairment Scale Variant II (DSIS-II) method defined in the ITU-R BT.500-9 [4].

A number of evaluation metrics were used to measure the performance of the PBDM as an estimator of video blocking artifacts in a variety of applications. The Spearman rank order correlation coefficient is related to the prediction monotonicity of the objective model, and the Pearson linear correlation coefficient is related to the prediction accuracy of the model [28].

Table I presents the evaluation results of the metrics. The 95% confidence bounds of the Pearson-Logistic metric have also been calculated using the method described in Ref. [31]. Fig. 2 shows the scatter plot of the OBR versus the MOS with the logistic fit. For comparison, the PSNR results are also reported in the table and the figure. As shown experimentally, a very good agreement between the MOS and the OBR has been achieved. Although the PSNR performs well in the VQEG test [17], where the dominant distortion is blurring, it is un-

TABLE I
CORRELATION COEFFICIENTS.

Evaluation metric	Proposed blocking distortion metric			PSNR		
	Correlation	Upper bound	Lower bound	Correlation	Upper bound	Lower bound
Spearman	0.937			0.508		
Pearson-Logistic	0.961	0.982	0.918	0.489	0.726	0.149

suitable for measuring blocking artifacts.

4. CONCLUSIONS

In this paper, two impairment metrics for MC/DPCM/DCT encoded video based on a model of the human visual system were presented. Investigations have been conducted to simplify and refine the distortion detection model. As an example, the performance of the proposed perceptual blocking distortion metric is evaluated through experiments, which shows high correlations with subjective test results.

5. ACKNOWLEDGEMENTS

The work was supported by the Australian Research Council under Grant A49927209. The authors thank C. J. van den Branden Lambrecht of the EMC Media Group for providing the NVFM software, S. Wolf of the NTIA/ITS for offering the T1A1 data set and the VQEG for sharing the video sequences and the subjective test data. The authors also wish to thank following colleagues and organisation in Monash University: D. Tan for his assistance on the data analysis, M. Gillies for his assistance on the subjective test, the Centre for Telecommunications and Information Engineering (CTIE) for the support on test equipment, and the participation of the assessors in subjective tests.

REFERENCES

[1] K. R. Rao and J. J. Huang, *Techniques and Standards for Image, Video and Audio Coding*, Prentice Hall, 1996.

[2] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision*, A. B. Watson, Cambridge, MA:MIT Press, 1993, pp. 207-220.

[3] F. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. COM-30, pp. 1679-1692, Jul. 1982.

[4] ITU-R BT. 500-9, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU, Geneva, Switzerland, 1998.

[5] P. T. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. First IEEE International Conference on Image Processing*, vol. 2, pp. 982-986, Nov. 1994.

[6] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Cambridge, MA:MIT Press, 1993, pp. 163-178.

[7] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in Digital images and human vision, A. B. Watson, Cambridge, MA:MIT Press, 1993, pp. 179-206.

[8] C. J. van den Branden Lambrecht, *Perceptual Models and Architectures for Video Coding Applications*, Ph.D. Thesis, EPFL, 1996.

[9] S. Winkler, "A Perceptual Distortion Metric for Digital Color Video," in *Human Vision and Electronic Imaging, Proc. SPIE*, vol. 3644, pp. 175-184, 1999.

[10] KPN Research, "Objective Measurement of Video Quality," *VQEG Doc. vqeg012*, Feb. 1997, available from *ftp.its.bldrdoc.gov*.

[11] Sarnoff Corp., "Sarnoff JND Vision Model Algorithm Description and Testing," *VQEG Doc. jrg003*, Aug. 1997, available from *ftp.its.bldrdoc.gov*.

[12] A. B. Watson, "Toward a perceptual video quality metric," in *Human Vision and Electronic Imaging III, Proc. SPIE*, vol. 3299, pp. 139-147, 1998.

[13] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," in *Proc. SPIE International Symposium on Voice, Video, and Data Communications*, Boston, MA, Sept. 1999.

[14] T. Yamashita, M. Kameda and M. Miyahara, "An objective picture quality scale for video images (PQSvideo) - definition of distortion factors," in *Visual Communications and Image Processing 2000*, King N. Ngan, Thomas Sikora, Ming-Ting Sun, SPIE vol. 4067, pp. 801-811, 2000.

[15] P. Bretillon et. al, "Quality meter and digital television applications," in *Visual Communications and Image Processing 2000*, King N. Ngan, Thomas Sikora, Ming-Ting Sun, SPIE vol. 4067, pp. 780-790, 2000.

[16] N. Damera-Venkata et. al, "Image quality assessment based on a degradation model," *IEEE Trans. Image Proc.*, vol. 9, no. 4, pp. 636-650, 2000.

[17] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," Mar. 2000, available from *ftp.its.bldrdoc.gov*.

[18] Z. Yu and H. R. Wu, "Human visual system based objective digital video quality metrics," in *Proc. Int. Conf. Signal Processing*, vol. II, pp.1088-1095, Aug. 2000.

[19] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, pp. 247-278, 1998.

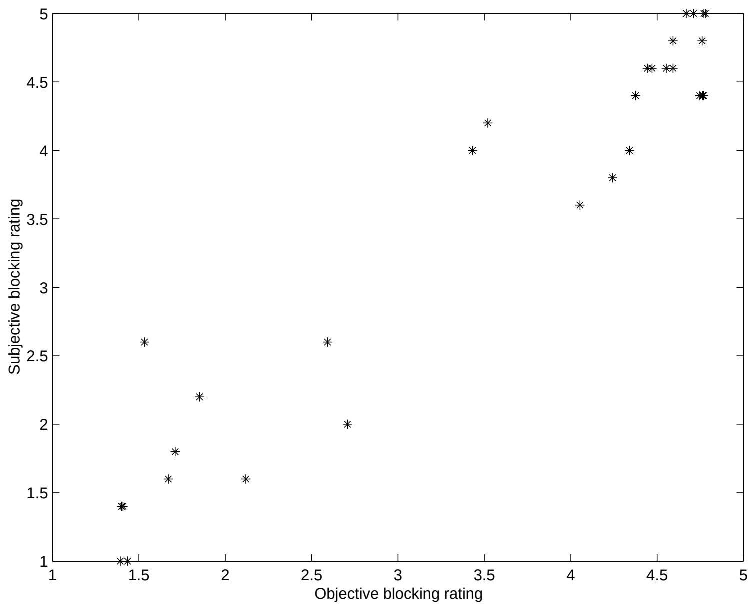
[20] ANSI T1.801.02-1995, *Digital Transport of Video Teleconferencing/Video Telephony Signals - Performance Terms, Definitions, and Examples.* American National Standard for Telecommunication, ANSI, 1995.

[21] M. Miyahara and K. Kotani, "Block distortion in orthogonal transform coding - analysis, minimization, and distortion measure," *IEEE Trans. Commun.*, vol. COM-33, no.1, pp. 90-96, Jan. 1985.

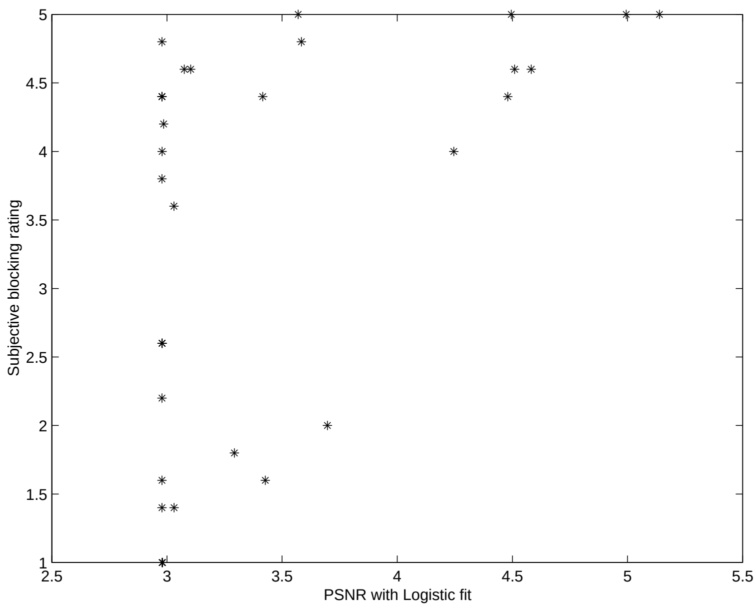
[22] S. A. Karunadekera and N. K. Kingsbury, "A distortion measure for blocking artifacts on images based on human visual sensitivity," *IEEE Trans. Image Processing*, vol. 4, pp. 713-724, June 1995.

[23] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric (GBIM) for video coding," *IEEE Signal Processing Letters*, vol. 4, no. 11, pp. 317-320, Nov. 1997.

[24] J. E. Caviedes, A. Drouot, A. Gesnot and L. Rouvellou, "Impairment metrics for digital video and their role in objective quality assessment," in *Visual Communications and Image Processing 2000*, King N. Ngan,



(a) Proposed objective blocking rating



(b) PSNR after the logistic fit

Fig. 2. Scatter plot of objective rating versus subjective rating with the logistic fit.

Thomas Sikora, Ming-Ting Sun, SPIE vol. 4067, pp. 791-800, 2000.

[25] B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.

[26] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Processing*, vol. 78, no. 2, pp. 231-252, 1999.

[27] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE 2nd Int'l Conf. On Image Processing*, Washington DC, Oct. 1995.

[28] Co-chair Objective Testgroup VQEG, KPN Research, "Evaluation of New Methods for Objective Testing of Video Quality: Objective Test Plan," Sept. 1998, available from *ftp.its.bldrdoc.gov*.

[29] A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Am. A*, vol. 14, pp. 2379-2391, 1997.

[30] MPEG Software Simulation Group, mpeg2encode/mpeg2decode version 1.1, Jun. 1994, available from *ftp.netcom.com*.

[31] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Iowa State University Press, 1989.