SELECTIVE USE OF MODEL-BASED CODING FOR LARGE MOVING OBJECTS

Don Pearson
Department of Electronic Systems Engineering
University of Essex
Colchester CO4 3SQ, UK

Tel: +44 1206 872865; Fax: +44 1206 872900
Email: dep@essex.ac.uk

ABSTRACT

Measurements using a continuous quality recording method have revealed the extent of quality variations that occur in MPEG2 pictures at low bit rates. Large moving objects in particular can give rise to particularly severe troughs in quality. The complementary characteristics of model-based coding are examined with a view to a synthesis of the two methods in a switched coder, with possible increased overall coding efficiency.

1 INTRODUCTION

One of the lessons to be learnt from the history of picture coding (and indeed of technology in general), is that techniques which are developed for one purpose often end up being used for another. Run-length coding was originally developed in an attempt to compress pictures with a continuous grey scale, but found its real and most effective application in the coding of bi-level documents for facsimile. DPCM and transform coding were once regarded as competitors, but now work together quite happily, taking complementary roles in modern hybrid interframe coders such as MPEG1/2 and H.261/3. Just a few years ago, conditional replenishment was seen as applicable only to videoconference pictures; broadcasters regarded them as too unreliable in terms of the high quality standards traditionally adopted in this field. Now this form of coding is being widely implemented as the basis for digital television broadcasting.

In this contribution I should like to take a fresh look at Model-Based Coding (MBC) [1] - [5]. MBC has been viewed largely as a technique suitable for coding low-resolution head-and-shoulders video (CIF or QCIF) into bit rates in the region 5-64 kbit/s, with applications to videoconferencing and the videophone. With fall-back modes to handle failures of tracking, etc., it may well be successful in such applications. But its real strengths could possibly lie in the coding of large moving objects, where current MPEG-type coding is at its weakest. In that respect the two could form a well-matched complementary pair in a switched hybrid coder. No coding method works well for all types of object or scene; each has its strong and weak points. In constructing a house, we do not build it all of concrete, or all of bricks, or all of wood, or all of metal; we use these materials selectively according to their strengths. In making a coder for video that will code a wide range of objects and types of movement, we need to utilise more than one coding method with characteristics that dovetail together.

2 PROBLEMS WITH MPEG1/2 AND H.261/3

Current interframe block-based MPEG1/2 or H.261/3 methods have problems dealing with large moving objects. Such objects are divided up into a large number of blocks, each if which is separately coded. In CBR mode, buffer overflow can occur when these objects move rapidly.

If the coder is operated in VBR mode, it is possible to gain a better appreciation of the interaction between object movement and coding method. Using this approach, it has been found that higher bit rates tend to be generated for larger objects; the effect of doubling the linear dimension of the object (quadrupling its area) is to increase the bit rate by about a factor of about 2 in the case of untextured objects and about 4 for textured objects [6], [7]. This is probably because for untextured objects the main contribution to the bit rate is made by nonzero interframe differences at the borders of the object as it moves; if the circumference is doubled, the bit rate roughly doubles. In the case of textured objects the bit rate increase is roughly proportional to the area increase.

The type of visual distortion that occurs with buffer feedback in CBR mode is illustrated in Fig. 1, which is a part of a single frame taken from a video sequence *Exam Conditions*, coded using MPEG-2 at 2 Mbit/s [8]. The coded image of the moving object, in this case a runner, is blocky and of low quality, whereas the non-moving background is generally of higher quality.

3 TEMPORAL VARIATION OF QUALITY

In recent times it has been shown to be possible to record moment-by-moment variations in subjective video quality [9], [10]. This can be done by giving subjects a slider or T-bar control linked to a recording computer. The subjects move the slider up and down according to their impression of the picture quality. Reasonable consistency between subjects and laboratories has been observed and the technique has recently been standardised by the ITU-R [11] as a Single Stimulus Continuous Quality Evaluation (SSCQE) method.

When blockiness is very objectionable, subjects move the slider towards the low-quality end of the range. When there is little motion or small-area motion causing low-level visible distortion, they push the slider up towards the high-quality end of the scale. Thus as the objects and movement in the video sequence change, so a trace of the variation in quality is obtained, as illustrated in Fig. 2. The graph shown is an average of the recordings of 24 non-expert subjects for the whole 20 minutes of the sequence *Exam Conditions*, MPEG-2 coded video at 2 Mbit/s; the vertical axis represents quality on a range from 0 to 1, this being the full extent of travel of the slider.

Over the period of 20 minutes of the test, it is possible to identify peaks and troughs in the quality. One of the worst region of troughs is right at the start. This part consisted of a number of runners in a sprint race. The photograph in Fig. 1 is taken from this section of the video, and shows one of the runners at the finishing tape.

The SSCQE method is relatively new and only a limited amount of investigation has taken place into the types of scene for which MPEG-2 tends to deliver poor quality. From the measurement taken thus far, however, large rapidly-moving objects are observed to result in poor coded picture quality at low bit rates. There may be other circumstances e.g. large numbers of small moving objects, for which this is also true.

A practical consequence is that bit rates for broadcast television tend to be set primarily on the basis of the worst-occurring scenes, among them being large moving objects. As the bit rate is progressively raised above 2 Mbit/s, so a situation is reached where most of the time there is no visible distortion in the picture. During these periods the coding is unnecessarily efficient, because the coding distortion is reduced to sub-threshold levels. If a means could be found to reduce the negative quality peaks, then the bit rate overall could be lowered.

4 CHARACTERISTICS OF MBC

In [12], [13] we showed that it was possible to switch between MBC and H.261 to give better coding efficiency and/or quality. We studied the switching patterns for various types of moving object, and found that for large rigid objects with simple translational or rotational motion, MBC was able to code more efficiently than H.261. The reason for this is not difficult to understand; in block-based coding, every moving block in every frame requires the specification of a motion vector; with a large object, there may be many blocks within its boundaries. MBC initially requires the shape of a rigid object to be transmitted, but once that is done only one motion vector per frame is required. The larger the object (in terms of blocks), the larger the saving in bits. It is consequently true that the advantage of MBC increases rapidly with higher image resolution; applications to broadcast and super high definition video are potentially attractive [2].

Changes in the object shape will require additional bits to be transmitted, but the evidence is that this does not usually occur rapidly, so that the bit rate is low. If a particular video sequence has numerous small, flexible objects which do not remain in the view of the camera for very long, then the parameterisation of the video into shape, motion and residual texture may not yield an advantage over conventional techniques.

A further characteristic of MBC is that in difficult tracking conditions it may position a moving object inexactly in a frame, but the object itself is unlikely to be distorted as in Fig. 1. This may create a better subjective impression.

5 MODEL FORMATION TIME

The available evidence as to how MBC will perform in terms of bit-rate or quality variations on longer video sequences is currently sparse.

When there are changes in scene content which require new models to be formed, a latency factor is probable [5]. Studies of model formation using a single camera have indicated that crude modelling can be completed within a few frames, but more complex modelling may take longer. If a stereo camera is used, then potentially the 3D object shape can be captured immediately; however, the computational procedures required for deriving 3D shape and covering them with a wire frame can also involve a considerable period of time, this being again dependent on the accuracy of modelling required [14].

Following an object or scene change, therefore, it appears that fall-back to a less efficient coding mode will be necessary for a substantial number of frames. The likelihood is that extended peaks in bit rate (VBR) or troughs in quality (CBR) will occur in such circumstances. Nevertheless (see Fig. 3) the quality troughs in MPEG2 can last a considerable time, so that MBC may be capable of developing even complex models (such as that of a moving human figure) to alleviate this situation. If pre-formed models can be quickly selected from a model ensemble, this could be even better. These solutions are currently being investigated.

6 CONCLUSIONS

Current hybrid interframe coding methods such as MPEG1/2 suffer from quality degradations for certain types of movement, particularly rapid large-object movement. The characteristics of model-based coding are roughly opposite to this, namely that it appears to work well with large moving objects. It has been shown that switching between the two types of coder is possible; hence possibilities exist for improving the performance of MPEG coders in this way. Further work remains to be done on latency aspects of model formation and/or the use of pre-formed models.

7 ACKNOWLEDGEMENT

The author would like to acknowledge with thanks the help of Richard Aldridge and David Hands in obtaining Figs. 1 and 2.

8 REFERENCES

[1] H. Li, A. Lundmark and R. Forchheimer, “Image Sequence Coding at Very Low Bit Rates: A Review”, *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 589-609, 1994.

[2] D. Pearson, “Developments in Model-based Video Coding”, *Proc. IEEE*, vol. 83, no. 6, pp. 892-906, 1995.

[3] H. G. Musmann, “A layered coding system for very low bit rate video coding”, *Signal Processing: Image Communication*, vol. 7, nos. 4-6, pp. 267-278, 1995.

[4] M. Kaneko, “Very Low Bitrate Video Coding - Recent and Current Researches in Japan”, in *Proc. International Workshop on Coding Techniques for Very Low Bit-rate Video*, paper E-1, Tokyo, Japan, 8 - 10 November 1995.

[5] D. Pearson, “Model-based Coding: Past and Future”, *Proc. International Picture Coding Symposium*, Melbourne, Australia, pp. 1-6, 13-15 March 1996.

[6] D.E. Pearson, J. Xiong and V. Seferidis, “A motion-related model for packet video”, *Proc. Fourth International Workshop on Packet Video*, paper D1, Kyoto, Japan, 29-30 August 1991.

[7] A. Robinson, D. Pearson and J. Xiong, “The influence of scene content on bit-rate variation in ATM video”, *IEE Colloquium on Coding for Packet Video and Speech Transmission*, London, Digest No. 1992/199, pp. 7/1 - 7/5, 9 November 1992.

[8] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands and D.E. Pearson, “Measurement of scene-dependent quality variations in digitally-coded television pictures”, *IEE Proceedings on Vision, Image and Signal Processing*, vol. 142, no. 3, pp. 149-154, 1995.

[9] R. Hamberg and H. de Ridder, “Continuous assessment of perceptual image quality”, *J. Opt. Soc. Am.*, vol. 12, no. 12, pp. 2573-2577, December 1995.

[10] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands and D. Pearson, “Subjective Assessment of Time-varying Coding Distortions”, *Proc. International Picture Coding Symposium*, Melbourne, Australia, pp. 269-274, 13-15 March 1996.

[11] Draft Revision of Recommendation ITU-R BT.500-7 to introduce a new method for Single Stimulus Continuous Quality Evaluation (SSCQE), 29 March 1996.

[12] M. F. Chowdhury, A.F. Clark, A.C. Downton, E. Morimatsu and D. E. Pearson, “A Switched Model-based Coder for Video Signals”, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 4, no. 3, pp. 216-217, 1994.

[13] M. F. Chowdhury, A.F. Clark, A.C. Downton, and D. E. Pearson, “Performance of a Switched Model-based Coder”, in *Proc. International Workshop on Coding Techniques for Very Low Bit-rate Video*, paper 3.4, Colchester, UK, 7-8 April 1994.

[14] D.V. Papadimitriou and T.J. Dennis, “Three-dimensional parameter estimation from stereo image sequences for model-based image coding”, *Signal Processing: Image Communication*, vol. 7, nos. 4-6, pp. 471-487, 1995.

Fig. 1 Section of a single frame from the sequence Exam Conditions, MPEG2 coded at 2Mbit/s

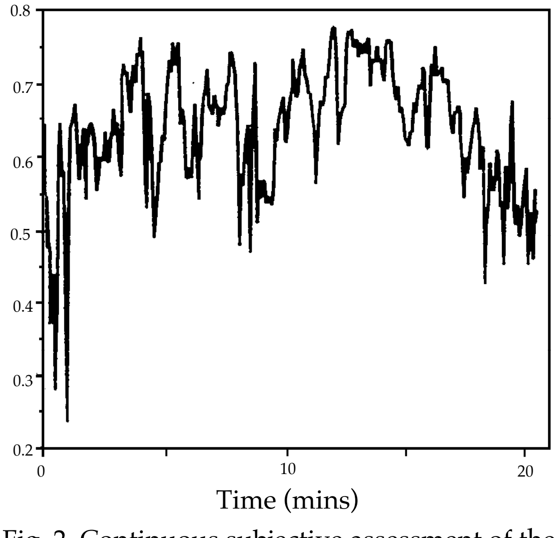

Fig. 2 Continuous subjective assessment of the sequence *Exam Conditions* using the SSCQE method

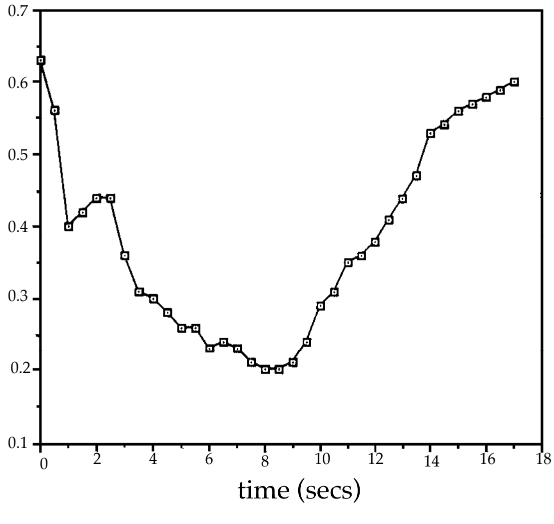

Fig. 3 Time enlargement of the worst quality trough in Fig. 2