

Adaptive Foveation of MPEG Video

T H Reeves

Department of Systems Design Engineering
University of Waterloo
Ontario, Canada, N2L 3G1
threeves@monet.uwaterloo.ca

J A Robinson

Faculty of Engineering and Applied Science
Memorial University of Newfoundland
Newfoundland, Canada, A1B 3X5
john@enr.mun.ca

ABSTRACT

We report experiments on the ‘foveation’ of MPEG-standard video by a network-based rate adaptor. The rate adaptor is controlled by input from receiving users, and constrained by the capacity of each receiver’s channel. It generates versions of the input video which trade off spatial and temporal resolution in different regions of the picture. While there are many possible ways to make the tradeoffs, users have only one or two dimensions of control (in addition to two dimensions of position for the fovea centre). We therefore propose constraints for mapping 2D user control input into a fovea/periphery bit budget, and provide design curves for spatial resolution control.

KEYWORDS: Compression/Decompression Analysis, Communications/Networking/VOD, Logical/Conceptual Manipulation of Video, Foveated Coding

1 INTRODUCTION

Multipoint multimedia communication relies on point-to-point channels to support sharing of information. There are various strategies for dealing with differences in the channels’ capacities. For asynchronous sharing, users with lower data rate connections simply have to wait longer to see the same information as those with higher data rate connections. For synchronous applications like conferencing, however, realtime interaction between all participants is required. Dynamic media like audio and video must be distributed continuously, with little delay.

Multirate media coding achieves this without constraining the global information throughput to that of the slowest link. For example, video can be simultaneously distributed at a variety of data rates, matched to receivers’ channels. Those with higher capacity channels

will enjoy higher quality pictures. Scalable and layered coding algorithms are therefore the subject of much current research. In this paper we are concerned with recoding MPEG-2 [1] video for multirate distribution, with the aim of minimizing recoding computation cost. In order to give end-users maximum control over their received images, we introduce a flexible ‘foveation’ strategy for data reduction. Adaptive foveation is a means of providing dynamic quality of service (QoS); that is, dynamic control of the temporal and spatial parameters characterizing a video sequence [2].

Figure 1 shows our model for multipoint video distribution over heterogeneous channels. The rate adaptor may be either network-based as shown, or collocated with the video source, but the essential elements of the model are that the receiving channels have, in general, different capacities, and the users at the receivers have different requirements of the video. The data rate output from the video source is assumed to be sufficiently high to support high-resolution full-frame moving pictures, while the distribution channels to the receivers afford only lower rates. The rate adaptor therefore has the responsibility of discarding some of the information from the source to match the MPEG stream to the channels.

MPEG-2’s scalable profiles were designed to allow partial reconstruction of the original video stream by low bandwidth receivers in heterogeneous transmission environments [1]. Unfortunately, these scalable profiles can provide only a very limited number of different layers [3]. Networks which contain many disparate elements (e.g. a variety of physical network links, protocols, and receiving terminals) require that video streams be available on a continuum of bit rates. This issue is addressed in [4]. In [4] however, the image contents are not prioritized, so the spatial resolution of the recoded bitstream is regionally invariant within the images.

In our model we assume low bandwidth return channels from each receiver to the rate adaptor/distributor. Based on control data transmitted over these channels, which prioritize the video signal in some way, the rate adaptor adjusts the video stream for each receiver. The approach taken here is ‘foveation’: the input from the

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

ACM Multimedia 96, Boston MA USA
© 1996 ACM 0-89791-871-1/96/11 ..\$3.50

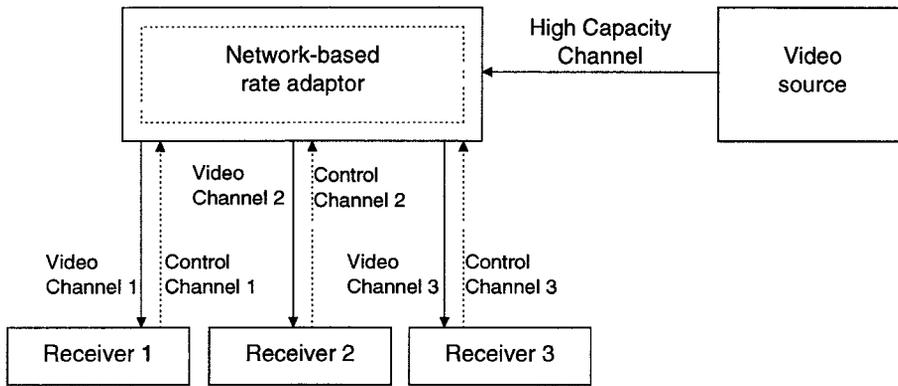


Figure 1: Multipoint video distribution system

receiving users describes a particular fovea or region of interest in the frame, and the rate adaptor then gives priority to information in that region.

In order to accommodate higher spatial and/or temporal resolution within the fovea (which necessarily increases the bit rate), the rate adaptor must implement strategies to reduce the bit rate outside the fovea. At the same time, the rate adaptor must minimise the delay in transforming the input video source bitstream into individualised bitstreams for each receiver. The consequence of these two requirements is that, whenever possible, the rate adaptor must prefer to discard information over the more time-consuming option of recoding the original input bitstream (e.g. recalculating motion vectors or requantising DCT coefficients).

Examples of application areas which would benefit from adaptive foveation include remote surveillance, tele-education, and videoconferencing. In security surveillance, where the user is monitoring one or more remote scenes, they may wish to give special attention to a particular region in a scene in which there appears to be suspicious activity. The system rate adaptor would be responsible for prioritizing the video bitstream to ensure that this foveated region is presented with high spatial and/or temporal resolution.

The use of remote classrooms in tele-education also presents an opportunity for prioritised video. During a discussion with a student in the classroom audience, the instructor may wish to focus on that particular student. As well, remote business boardrooms, and other videoconferencing situations involving multiple participants, would benefit from user-prioritized video. Furthermore, prioritized video is appropriate in multicast environments where transmission modes include wired and wireless channels [4].

For a given coding method (in our case MPEG-2), re-

coding can reduce the bit rate by either:

- (a) Reducing temporal resolution (frame rate).
- (b) Reducing spatial resolution (typically by discarding high frequency transform coefficients).

With foveation, the picture content is divided into two classes:

- (1) High-resolution fovea or foreground region, which is compact and convex (typically a disk or a square.) This is centred on user-selected coordinates, and moves in response to user input. (The fovea is also referred to as a 'region of interest', or ROI.)
- (2) Lower-resolution periphery or background, being the remainder of the picture outside the fovea.

The use of bit rate reduction methods (a) and (b), independently applied in the two regions (1) and (2), gives four dimensions on which to trade-off picture quality against bit rate. With the fovea size (radius) being a further dimension of control, there is a very large space of possible picture appearances and bit budgets. We are concerned with reducing the dimensionality of the space of possibilities, so that user control can be simple and natural.

Section 2 of this paper describes the user interface we are developing for receivers in this model, showing how one- or two-dimensional priority information, plus 2D coordinates, can be input by receiving users. Specifically, then, the problem is mapping the five dimensions of bit allocation (fovea size, foveal temporal resolution, foveal spatial resolution, peripheral temporal resolution, peripheral spatial resolution) onto two dimensions of control.

Section 3 discusses the constraints and assumptions we make in order to solve the problem. In particular, we argue that temporal resolution gives only coarse control over the bit rate, and that the chosen peripheral temporal resolution must be selected based on picture content. We therefore provide the user with relatively few control points in the temporal dimensions, and limit the dynamic variability of temporal resolution.

In section 4, for a particular important control point in temporal resolution space, we show the spatial resolution tradeoffs at various data rates, and provide design curves that map the two dimensions of spatial control to a single dimension.

In our work we have constrained all recorded streams to be true MPEG. Although the receivers are assumed to have a mechanism for user input (see section 2), they are running standard MPEG decoders. Furthermore the source is an MPEG stream. The consequence of this is that it is sometimes necessary to transmit null information where a custom system could adopt a more economical coding. Finally, therefore, in Section 5, we analyze the MPEG overhead of foveation at different data rates.

Our implementation of the network-based rate adaptor is based on modifications to the MPEG Software Simulation Group's publicly available MPEG-2 codec [5].

2 User Control of Prioritised Video

Research into prioritized video has traditionally focused on videoconferencing applications. Girod [6] posited that the distribution of bits used to code video sequences, for example video telephony, could be adjusted depending on the viewer's focus of attention. Videoconferencing sequences, especially the 'head and shoulders' type, are particularly amenable to prioritized region segmentation. Faces are natural high-priority areas because of the increased movement and detail they present (as well as for obvious semantic reasons). The background is in the periphery of the viewer's vision, where the mammalian vision system is less acute. Considerable research has been devoted to prioritized coding in videoconferencing applications [7]–[11]. There is an interest in tracking the viewer's gaze to determine the viewer's region of interest [6, 11]. In [11], the intended application is a videoconferencing sequence with multiple participants. The system tracks the spectator's line-of-gaze, based on the position and inclination of the spectator's head, and devotes more resources to coding the participant at whom the line-of-gaze is directed.

Any system using eye tracking, or a conventional 2-D input device such as a mouse, can provide fovea centre coordinates. We are investigating user interfaces which provide richer information, so that users have more control over the foveation tradeoffs. These use 3D gloves

or 3D tracking plus gesture recognition to provide 3+n dimensional input.

We envisage that the user-controlled prioritized video system will benefit from such a rich degree of interactivity. The user will control the position and frame rate of the fovea by pointing at the region of interest within the image. The fovea/periphery tradeoff will be determined by the distance between the user's hand and the video screen. For instance, to observe a specific region of the video screen in high resolution, the user will position their finger very close to the screen in the location of this region. As the user moves their hand away from the screen, the resolution within the fovea will decrease, with a corresponding increase in the periphery resolution.

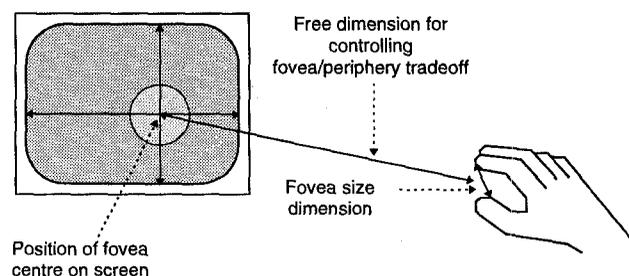


Figure 2: User interface dimensions of control

User control of the size and shape of the fovea could be provided by other measurements afforded by the 3D mouse or data glove. Any two of the yaw/pitch/roll measurements from the 3D mouse could determine the height and width of a rectangular fovea. However, we do not favour use of orientation dimensions. Instead, we envisage the user indicating with thumb and forefinger the desired size of the fovea. Figure 2 shows the four dimensions of control in our user interface. Because two dimensions map naturally and directly to the fovea centre's position, and we have chosen to map one directly to fovea size, only one dimension (the distance of the hand from the screen) is available for controlling the fovea/periphery tradeoff. The remainder of this paper explains how we map between this dimension and the resolutions in the picture.

3 Temporal Resolution

Realtime video over restricted capacity transmission systems is conventionally accomplished by temporal sub-sampling. Videophones typically use 10 frames/s. The unrealistic movement evident at this rate can impede natural interaction between participants, and is perhaps one reason that such systems have not found widespread acceptance amongst consumers. Indeed, in their

study of user perception of temporal resolution reductions in a variety of video sequences [2], Apteker et al. conclude that slower frame rates are less acceptable in sequences with slow, limited motion (e.g. snooker, stand-up comedy), compared to high motion content sequences (e.g. sporting highlights, music videos). Furthermore, the fact that television traditionally has been interlaced suggests that most video consumers prefer to accept some spatial resolution degradation in favour of maintaining temporal resolution. As a consequence of these considerations, in our system we have constrained the temporal resolution in the fovea to its original frame rate. Therefore we constrain our investigations to one hyperplane in the foveal temporal resolution dimension.

Since the image outside of the fovea (the periphery) has been designated by the user as having low priority, temporal subsampling in this region is worth considering. Using a PAL video rate sequence (25 frames/s), we investigated the consequences on bit rate of three peripheral temporal resolutions: every other I frame, every I frame, and every I and P frame. The GOP structure of the sequence has one I frame for every 3 P frames and 8 B frames; i.e. I B B P B B P B B P B B I. Consequently, the three temporal resolutions correspond to (approximately) one periphery refresh per second, two periphery refreshes per second, and eight periphery refreshes per second, respectively. Although the every-I-and-P frame scheme represents a considerable increase in peripheral frame rate over the two other schemes, we included it in our investigation because of its similarity to conventional videophone frame rates.

Because the rate adaptor must produce a legitimate MPEG bitstream, there remains a cost associated with the discarded peripheral frames. In order to maintain the full frame rate in the fovea, the 'discarded' peripheral frames are actually recoded in such a way that they are 'transparent'. The result is that the last non-discarded I or P frame endures in the periphery until the next non-discarded I or P frame is encountered. (In these 'transparent' frames, the slices in the discarded peripheral frames consist primarily of skipped macroblocks, as well as a few motion-compensated macroblocks with zero motion vectors. In terms of operations, the recoding required is negligible.) The cost per slice of discarded peripheral frame in our test sequences was approximately 72 bits. (This value varies with the information in the slice header and the number of skipped macroblocks.)

The images in the 'class' test sequence are 320x240 pel, colour pictures, and the bit rate required to support the original MPEG-2 file is 1 Mbit/s. This test sequence has many of the characteristics which we would expect to observe in our intended applications. The camera's view of the scene is fairly constant (no panning or zooming),

and there is limited, localized motion in different areas of the foreground and midground.

The graphs in figure 3 represent, for different fovea sizes, the average bit rate required to present the fovea in the 'class' sequence in real-time with a temporally subsampled periphery. The fovea begins in the middle of the image with dimensions 0x0 macroblocks. The fovea remains square as its size is increased to 15x15 macroblocks. The periphery has been spatially subsampled by limiting the number of DCT coefficients used to encode peripheral macroblocks.

(Since we expect that foveation techniques will be used to provide realtime video over low data rate lines, and not for archiving high-quality video, the editing aids provided by the GOP header are unnecessary. Throughout our tests we have discarded the GOP headers in the interest of reducing the bit rate. These headers are typically about 60 bits in size.)

The bit rates for no fovea with only the dc coefficients in the periphery (which obviously constitutes the entire image in this case) represent the lower bounds on the achievable bit rates for each temporal subsampling scheme. Since the bit rate of the original MPEG sequence is 1 Mbit/s, the every-I-and-P-frame subsampling scheme at 165 kbits/s offers worse than 6:1 compression of the original MPEG sequence. The every-I-frame subsampling scheme achieves 14:1 compression, and the every-other-I-frame scheme achieves 20:1 compression. We found these compression ratios to be consistent across several different video sequences coded with the same GOP structure. Figure 3(d) clearly shows that even relatively low resolution (320x240) MPEG video, temporally subsampled at 8 frames/s, is not practical for low bit rate systems.

The bit rate results above, and the associated quality of the video pictures, obviously depend on the amount and degree of motion in the periphery. When evaluating the repercussions of temporal subsampling in the periphery, we are concerned with the following issues:

- Fine temporal subsampling reduces the bit rate savings benefits of adaptive foveation.
- The temporal subsampling rate required to provide the user with sufficient information about activity in the periphery depends on picture content. One of the primary concepts behind adaptive foveation is that the user has the ability to alter the size and location of the fovea in response to activity in the image of interest to them. Obviously the temporal resolution in the periphery must not be so coarse as to eliminate motion which may be of interest to the user.
- The probability that coarse temporal subsampling will

resolutions are necessary to further reduce the bit rate. In our test sequences with the periphery subsampled to 2 frames/s, the DCT coefficients used to code intra-macroblocks and the error residue for motion-compensated macroblocks account for 70% – 80% of the total bit rate. Therefore, we consider controlling spatial resolutions in the image by:

- limiting the number of DCT coefficients used in representing the periphery,
- limiting the number of DCT coefficients used in intra-macroblocks and error residue macroblocks in the fovea.

We intend that the free dimension in the user interface (the distance of the hand from the screen) be used to control the trade-off between these spatial resolutions.

In addition to the foveated videoconferencing research mentioned earlier, methods for compromising spatial resolution in the periphery for the benefit of foveal spatial resolution have been investigated in [12]–[16]. All of the methods use reduced representations of the periphery (e.g. through subsampling, scaling, coarse pattern matching, etc.) to maintain image fidelity in the fovea(e).

Transform based coding systems [10, 11, 14, 16] limit the number of transform coefficients in the periphery, either by quantising them more coarsely, or by simply eliminating them. These strategies are appealing because the loss of resolution in the periphery, in favour of maintaining spatial and temporal resolution in the fovea, is analogous to the human visual system (HVS). (For detailed analysis of the repercussions of the spatial characteristics of the HVS on foveated coding, see [12, 13]. For investigations of the temporal characteristics of the HVS, see [6, 17].)

The number of DCT coefficients used in the fovea can also be restricted in order to reduce a sequence’s average bit rate. The number of coefficients retained for foveated intra-macroblocks should be at least as great as the number of coefficients retained for peripheral intra-macroblocks, because coarser resolution in the fovea than in the periphery defeats the purpose of foveation.

The role that the number of error residue coefficients plays in the fovea spatial resolution is less clear. Motion predicted macroblocks have the same spatial resolution as the macroblocks from which they are predicted. This could be considered sufficient reason for discarding the error residues entirely from the fovea. However, if a motion predicted macroblock relies on an area outside of the fovea for its values, its spatial resolution will match that of the periphery instead of

the fovea. Furthermore, the appearance of two superimposed macroblocks, resulting from a bidirectionally predicted macroblock without error compensation, can be more visually disturbing than the spatial resolution would suggest. Therefore we examined two strategies for handling the DCT coefficients representing error residue macroblocks in the fovea. In the first strategy, we limit the number of error residue coefficients to the same number as the intra-block coefficients. In the second strategy, we dispense with the error residues entirely.

Because of the processing time involved, we wish to avoid recoding DCT coefficients using a different quantisation level. Instead, the rate adaptor first decodes the requested number of coefficients from the input bitstream. (Because the coefficients are run-length encoded, more coefficients than requested may be decoded.) The rate adaptor must then introduce an ‘End of Block’ code into the modified bitstream which is transmitted to the user’s terminal. The EOB code is either 2 or 4 bits. To eliminate the error residues, we remove the coded block pattern flag (‘macroblock_pattern’ [1]) from the macroblock headers, and exclude all error blocks from the recoded bitstream.

Our preliminary tests on spatial resolution tradeoffs have focused on two sequences which are representative of the picture content for which the adaptive foveation system is intended. The question in which we are interested is “If we increase the number of DCT coefficients in the fovea by n , by how many must we reduce the number of DCT coefficients in the periphery?”. The tests are intended to establish the feasibility of creating design curves that answer this question, leaving control of the subjective video quality tradeoffs at the discretion of the user. Using the free dimension in the user interface, the user will control the discrepancy in resolution between the periphery and fovea. The rate adaptor will use the design curves to trade off the two spatial resolutions while maintaining a constant bit rate.

The pictures in figure 4 are single frames from our test sequences. The number of DCT coefficient representing peripheral blocks is limited to 5, while the fovea is at full spatial resolution. In both sequences, the camera is not steady (due to mechanical vibration), so the peripheral motion vectors in the P and B frames are often nonzero.

The peripheral/foveal spatial resolution tests used fixed window sizes, encompassing a plausible region of interest in each sequence. In each of the ‘class’ and ‘shoplifter’ sequences, the fovea’s area is 6x6 macroblocks. The frame rates are 25 frames/s and 30 frames/s, respectively; the sequence durations are 11 seconds, and 10 seconds.

Design curves to tradeoff the spatial resolutions are rep-



'Class' sequence

Image size: 320x240 pixels

The fovea is 6x6 macroblocks, and is centred on the figure of the lecturer at the blackboard. The movement in the sequence consists of the lecturer drawing on the blackboard, students shifting about in their desks, and one student entering through the door on the left and seating herself in front of the lecturer.



'Shoplifter' sequence

Image size: 320x240 pixels

The fovea is 6x6 macroblocks. It is centred on the figure in the foreground. During the sequence, the figure lifts an item from the shelf and places it in his pocket. In the background, people wander across the aisle.

Figure 4: Test sequences

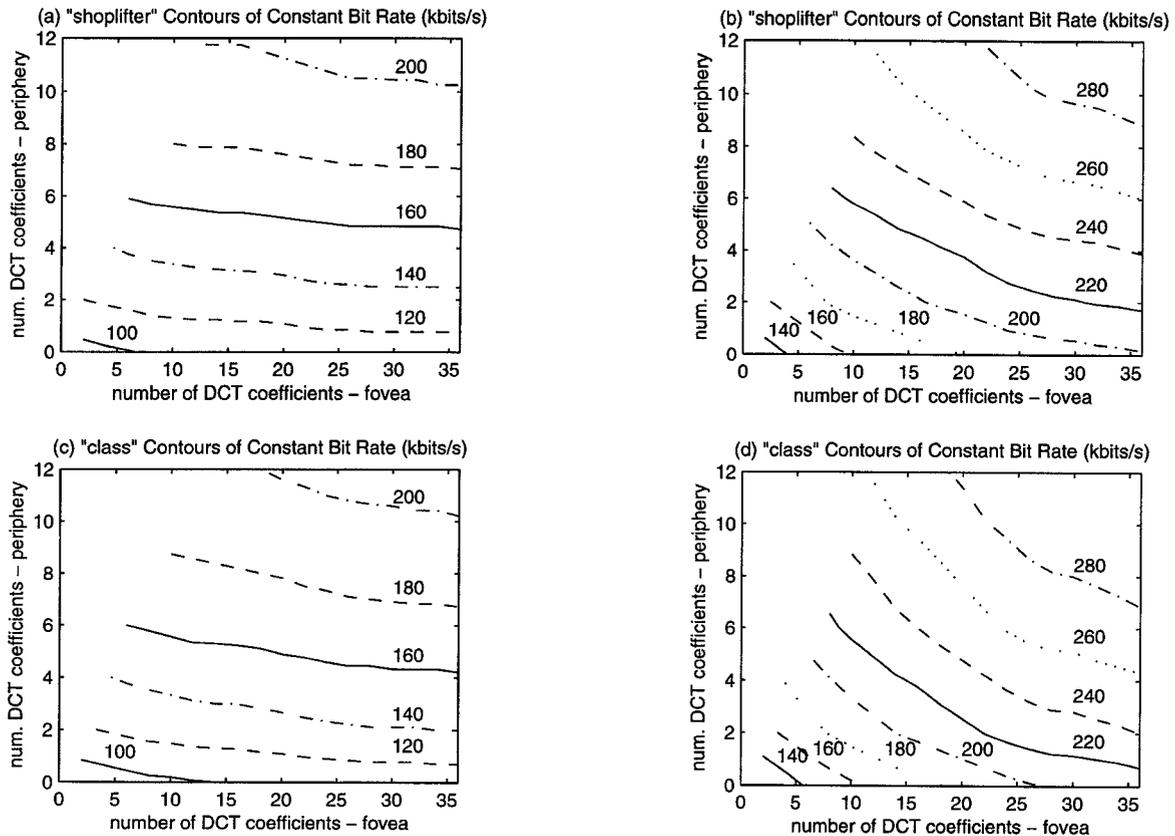
resented by contours of constant bit rate in the foveal spatial resolution/peripheral spatial resolution plane. Iso-bit rate contour graphs are presented in figure 5 for the 'class' and 'shoplifter' sequences. The graphs on the left result from discarding the error compensation information from motion predicted macroblocks in the fovea. For the graphs on the right, the DCT coefficients representing the error compensation information were limited to the same number as the DCT coefficients used to represent intra-macroblocks in the fovea.

If the free dimension in the user interface only provided control points in one spatial resolution dimension, the rate adaptor would be responsible for the complex task of estimating the tradeoff required in the other spatial dimension to maintain a constant bit rate. Since the rate adaptor is constrained in terms of the computational load it can accept without endangering the real-time delivery of video to its receivers, this scheme is infeasible. Our solution is to use the contours of constant bit rate as design curves to be used by the rate adaptor. The design curves lie in the foveal spatial resolution/peripheral spatial resolution plane. These design curves function as lookup tables, where each lookup ta-

ble is associated with a specific bit rate. The current lookup table is chosen by the rate adaptor based on the fixed peripheral temporal resolution, the current fovea size, and the capacity of the receiver's channel.

The distance between the user's hand and the screen provides one control point. This point can be considered as an index into the lookup, or a mapping onto the design curve. At the control point on the design curve, the rate adaptor reads off two values: the peripheral spatial resolution and the foveal spatial resolution. The rate adaptor is guaranteed that this combination of spatial resolutions is within the carrying capacity of the receiver's channel. The user is free to interactively adapt the image appearance subject to their individual preferences (e.g. amount of peripheral detail versus foveal detail).

The design curves are very dependent on the picture's physical characteristics (e.g. size, chroma subsampling, etc.) and its content (e.g. amount and extent of motion). We intend to investigate the possibility that the design curves for sequences used in same application are sufficiently similar to allow the sharing of common design curves. Ideally, the rate adaptor would be trained



Results without error residues

Results with error residues

(Motion predicted macroblocks contain no error compensation.)

(Error compensation blocks have the same spatial resolution as intra-blocks.)

Figure 5: Spatial resolution tradeoffs — MPEG bitstream

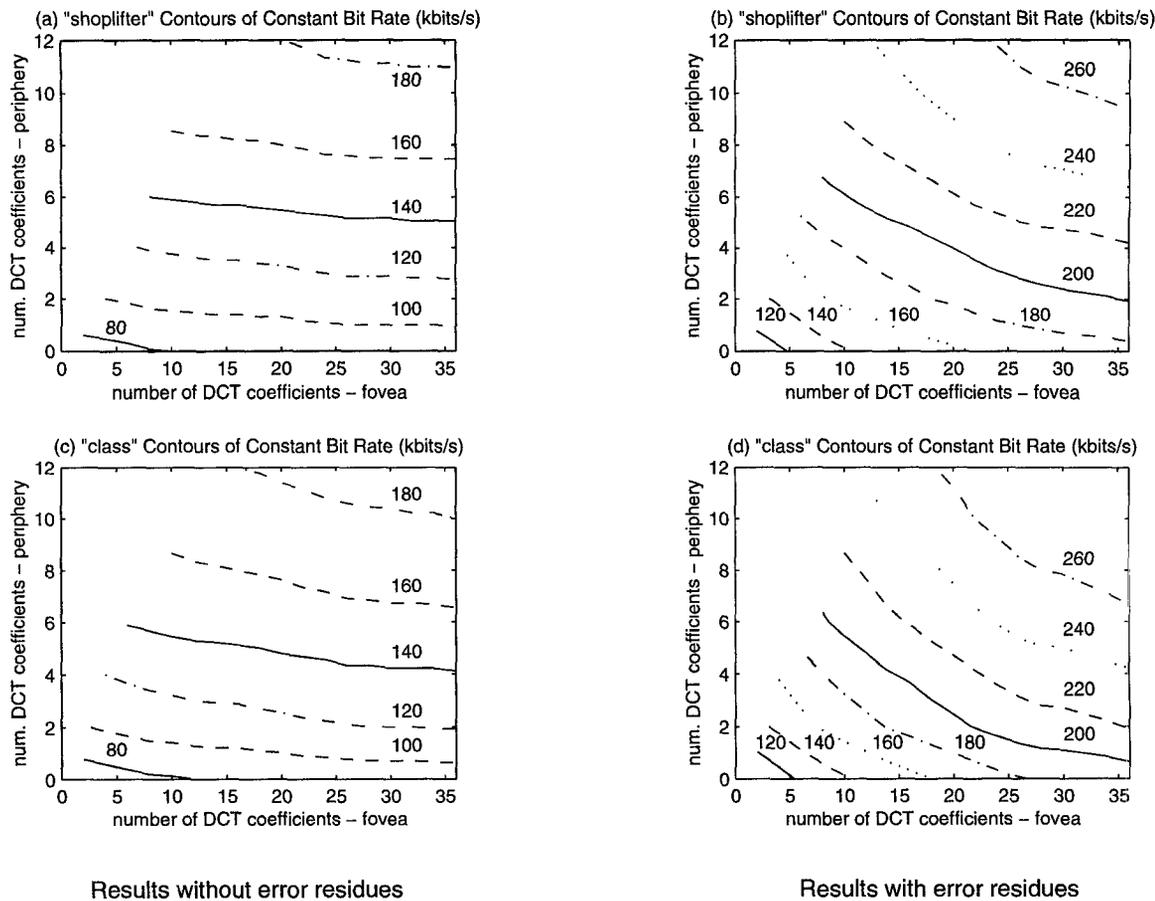
for a specific application to create custom design curves. For instance, for a remote tele-education system having a fixed classroom with a relatively constant number of students, the rate adaptor could be trained at the beginning of the term, and therefore provide adaptive foveated video at reliable bit rates throughout the term.

5 MPEG Overhead of Foveation

All of the implementations of bit reduction strategies have been developed with the constraint that the modified bitstream must conform to the MPEG-2 standard. We have investigated the cost of including information ('null' information) in the bitstream whose only purpose is to maintain MPEG conformance, for example the 'transparent' slices in discarded peripheral frames. We allow that MPEG headers above the slice level (except for the GOP header) serve a purpose beyond conformance. Slice headers in I frames and in the fovea are also considered necessary.

The graphs in figure 6 show the bit rates that result from discarding the null information, and therefore counting just the 'essential' bits. The contours shown are for the same bit rates as in the true MPEG bitstream tests above. Although we have compromised the MPEG legitimacy of the bitstream, the graphs for each sequence have the same general shapes as their true MPEG counterparts above. This shows that the cost of the MPEG overhead is constant as we move along the design curves. The relative tradeoffs between the two spatial resolutions appear to be independent of the MPEG constraints on the bitstream.

Figure 7 shows the cost of the true MPEG bitstream compared to the essential bit stream as a function of fovea size. The 'shoplifter' test sequence was used with full spatial resolution in the fovea. For the graph on the left, the peripheral spatial resolution was limited to the dc coefficient; for the graph on the right, the periphery



(Motion predicted macroblocks contain no error compensation.)

(Error compensation blocks have the same spatial resolution as intra-blocks.)

Figure 6: Spatial resolution tradeoffs — 'Essential' bitstream

was at full spatial resolution.

The cost of MPEG is inversely related to the fovea size because the number of transparent slices and zero-motion predicted macroblocks required increases as the fovea size decreases. For the relative fovea and periphery sizes shown here, the cost of the MPEG overhead is 14–30 kbits/s. For low bit rate foveated sequences around 200 kbits/s, this overhead constitutes between 7% and 15% of the total bit rate. Note that we have been generous in classifying much of the MPEG-2 syntax as essential. For example, Sun and Zdepski [18] conclude that missing quantisation matrix information in picture headers causes only slight degradations in the quality of reconstructed sequences.

These graphs also illustrate the advantages of relaxing the MPEG conformance constraint. The design curves from the true MPEG streams have been shifted along the peripheral spatial resolution and foveal spatial res-

olution axes. The consequence is an improvement in image quality at the same bit rate as the true MPEG stream.

6 Conclusions

We presented a network-based rate adaptor which delivers individually prioritized video streams to multiple, independent clients. Low data rate transmission for realtime video delivery over restricted-capacity transmission media is accomplished by adaptive foveation of the image sequence. We have identified five main dimensions of control (two temporal and three spatial) over the foveated bitstream. For the sake of the quality of video in the fovea, we constrain the foveal temporal resolution to the original frame rate. The amount of temporal subsampling in the periphery is preselected by the user, based on the picture content.

Using a subset of the many dimensions offered by a versatile user interface device, such as the 3D mouse,

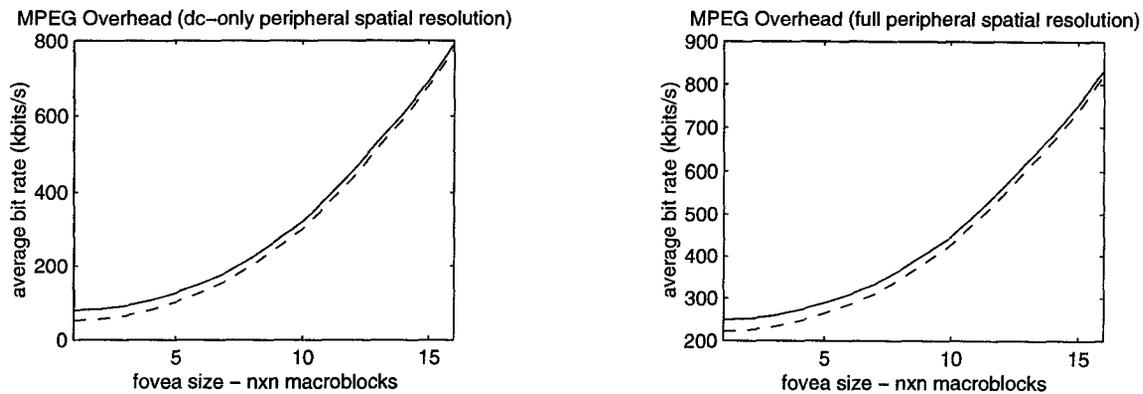


Figure 7: Cost of MPEG overhead versus fovea size

we allow the user simple and intuitive control over the fovea size and the spatial resolution discrepancy between the periphery and the fovea. We presented examples of curves of constant bit rate in the foveal spatial resolution/peripheral spatial resolution plane. The rate adaptor maintains a constant bit rate by trading off the peripheral and the foveal spatial resolutions according to these design curves. We found that by sacrificing the MPEG conformance of the bitstream, the design curves are shifted along the spatial resolution axes towards finer resolution. The consequence of this shift is better video image quality for the user. However, the shapes of the design curves are relatively unaltered from their true MPEG bitstream form, indicating that the cost of the MPEG overhead is constant for constant bit rate spatial resolution tradeoffs.

REFERENCES

1. ISO/IEC/JTC1 Draft International Standard 13818 / ITU-T Draft Rec. H.262, May 1994.
2. R.T. Apteker, J.A. Fisher, V.S. Kisimov, H. Neishlos, "Distributed multimedia: user perception and dynamic QoS", *SPIE*, Vol. 2188, 1994, pp. 226-234.
3. S-F. Chang, A. Eleftheriadis, D. Anastassiou, "Development of Columbia's video on demand testbed", *Signal Processing: Image Communication*, Vol. 8, April 1996, pp. 191-207.
4. A. Eleftheriadis, D. Anastassiou, "Constrained and general dynamic rate shaping of compressed digital video", *IEEE International Conference on Image Processing*, Vol. 3, 1995, pp. 396-399.
5. MPEG Software Simulation Group; C code for MPEG-2 codec available by anonymous ftp from ftp.mpeg.org/pub/mpeg/mssg.
6. B. Girod, "Eye Movements and Coding of Video Sequences", *SPIE Visual Communications and Image Processing*, Vol. 1001, 1988, pp. 398-405.
7. A. Basu, A. Sullivan, K.J. Wiebe, "Variable resolution teleconferencing", *IEEE Systems, Man, and Cybernetics Conference Proceedings*, 1993, pp. 170-175.
8. A. Basu, K.J. Wiebe, "Videoconferencing using Spatially Varying Sensing with Multiple and Moving Foveae", *Proceedings of the IEEE International Conference on Pattern Recognition*, Vol. 3, 1994, pp. 30-34.
9. A. Eleftheriadis, A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates", *Signal Processing: Image Communication*, Vol. 7, No. 3, Sept. 1995, pp. 231-248.
10. A. Eleftheriadis, A. Jacquin, "Low bit rate model-assisted H.261-compatible coding of video", *IEEE International Conference on Image Processing*, 1995, pp. 418-421.
11. L.C. De Silva, K. Aizawa, M.Hatori, "Use of Steerable Viewing Window (SVW) to improve the visual sensation in face to face teleconferencing", *ICASSP Proceedings*, Vol. 5, 1994, pp. 421-424.
12. A.T. Duchowski, B.H. McCormick, "Pre-attentive considerations for gaze-contingent image processing", *SPIE Human Vision, Visual Processing, and Digital Display VI*, Vol. 2411, 1995, pp. 128-139.
13. A.T. Duchowski, B.H. McCormick, "Simple multiresolution approach for representing multiple regions of interest (ROIs)", *SPIE Visual Communications and Image Processing*, Vol. 2501, 1995, pp. 175-186.

14. E. Nguyen, C. Labit, J-M. Odobez, "A ROI approach for hybrid image sequence coding", *IEEE International Conference on Image Processing*, 1994, pp. 245–249.
15. P.L. Silsbee, A.C. Bovik, D. Chen, "Visual Pattern Image Sequence Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 3, No. 4, August 1993, pp. 291–301.
16. A. Vlaicu, S. Lungu, N. Crisan, S. Persa, "New compression techniques for storage and transmission of 2d and 3d medical images", *SPIE Advanced Image and Video Communications and Storage Technologies*, Vol. 2451, pp. 370–377.
17. J.H.D.M. Westerink, C. Teunissen, "Perceived sharpness in moving images", *SPIE Human Vision and Electronic Imaging: Models, Methods, and Applications*, Vol. 1249, 1990, pp. 78–87.
18. J. Sun, J. Zdepski, "Error Concealment Strategy for Picture-header Loss in MPEG Compressed Video", *SPIE High-Speed Networking and Multimedia Computing*, Vol. 2188, 1994, pp. 145–152.