Dual Frame Motion Compensation with Uneven Quality Assignment

Vijay Chellappa, Pamela C. Cosman and Geoffrey M. Voelker
Dept. of Electrical and Computer Engineering, Dept. of Computer Science and Engineering
University of California, San Diego La Jolla, CA 92093-0407
E-mail:{vchellap,pcosman}@ucsd.edu,voelker@cs.ucsd.edu

Abstract

Video codecs that use motion compensation have shown PSNR gains from the use of multiple frame prediction, in which more than one past reference frame is available for motion estimation. In dual frame motion compensation, one short-term reference frame and one long-term reference frame are available for prediction. In this paper, we propose a dual frame motion compensation technique that allocates bits unevenly among frames to periodically create a high-quality frame that serves as the long-term reference frame for some time. By modifying an MPEG-4 encoder to use this technique on a set of video sequences, we show that it outperforms a normal dual frame motion compensation scheme in which the long-term reference frames are regular frames that are not allocated any extra rate.

1 Introduction

In this paper we describe the design, implementation, and evaluation of a dual frame motion compensation technique that allocates bits unevenly among frames to periodically create a high-quality frame that serves as the long-term reference frame. Contemporary hybrid video codecs use motion-compensated prediction to efficiently encode a raw input video stream. For each block in the current frame to be encoded, the encoder searches in the reference frame (usually the immediate past frame) to find the best match block for it. The best match block is often called the prediction of the current block. The difference between the current block and its prediction from the reference frame is compressed and transmitted, along with the displacement (motion) vector that describes the location of the best match block relative to the current block. Called *inter* coding, this is the basic approach found in the video coding standards MPEG, MPEG-2, MPEG-4 [1], H.263 [2] and the latest and state-of-the-art H.264/AVC [3].

In multiple frame prediction, more than one past frame is used in the search for the best match block. At the cost of extra memory storage and extra complexity for searching, multiple frame prediction has been shown to provide a clear advantage in compression performance. One of the earliest papers describing the use of multiple reference frames

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

is [4]. Since the complexity of the encoder increases, methods to reduce complexity were proposed. These included searching a smaller windowed area from the previous frames, and using pixel decimation. Even with sub-optimal search strategies, bit rate reductions of about 4% could be realized.

In [5], multiple frame motion compensation was proposed as a more robust and efficient scheme to combat wireless packet losses and errors. As in [4], a coding advantage was shown especially for "talking head" video sequences with static backgrounds such as the News or Foreman sequences. The multiframe approach had lower error probability than the single frame approach. To control the variable time delay in transmitting motion vectors for multiple frames, [6] used rate-constrained motion estimation. The variable time delay problem in transmitting the motion vector bits becomes acute when the long-term memory is large. The efficiency of motion compensated prediction improved when using up to 50 past frames.

In [7], the multiple frame scenario was examined in a low bit rate environment for the H.263 standard. The encoder used two frames for predictive coding, a short-term reference frame and a long-term reference frame. We refer to this as dual frame motion compensation. The short-term reference frame was found to be useful for local motion and the long-term reference frame was useful for still regions of the image. Dual frame motion compensation was also the subject of [8], which used optimized inter/intra frame switching within a rate-distortion framework.

In our preliminary work [11], we considered dual-frame motion compensation in the context of a rate-switching network. The "Always Best Connected" network is a service that provides a seamless transition among various connections at different rates [12], always connecting the user using the best connection available at the moment. In this context, we found that, when switching from a high bandwidth connection (e.g., Ethernet at 10Mbps) to a low-bandwidth connection (e.g., GPRS at 16kbps), the codec might benefit from retaining the last frame from the high-quality connection and using it as a reference frame in a multiple frame prediction scheme [11].

Given this result, we reasoned that, in the absence of any rate switches or alternative connections, a user with a low bandwidth connection at 16kbps might still benefit from periodically creating a higher quality frame and then retaining this frame for some time as a long-term reference frame in a multiple reference frame scheme. In this paper, first we explore the dual frame technique and the use of higher quality long-term frames to improve video quality in Section 2. Then we demonstrate the benefits of our dual frame technique in an MPEG-4 encoder in Section 3. Finally, we conclude in Section 4.

2 Dual Frame Motion Compensation

In this section, we discuss the basic use of dual frame motion compensation. While encoding frame n, the encoder and decoder both maintain two reference frames in memory. The short-term reference frame is frame $n-1$. The long-term reference frame is, say, frame $n-k$, where k may be variable but is always greater than 1. Each macroblock can be encoded in one of three coding modes: intra coding, inter coding using the short-term buffer (inter-ST-coding), and inter coding using the long-term buffer (inter-LT-coding). Figure 1

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

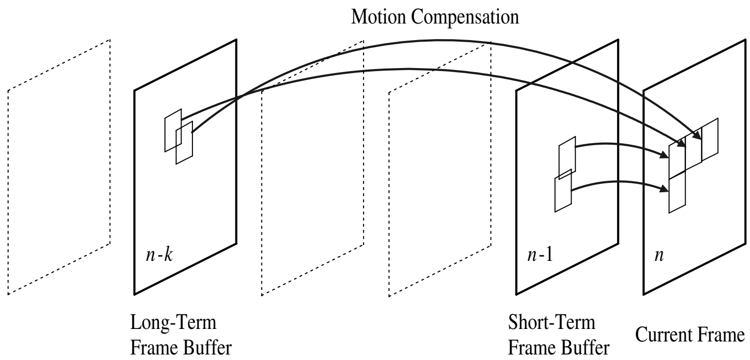


Figure 1: Dual Frame Buffer Motion Compensation.

illustrates this technique. Each time the encoder chooses inter coding, the encoded bit stream would require an additional bit to inform the decoder which one of the two reference frames was used by the current block.

In [9], two different approaches for choosing the long-term reference frame were described. In one approach, called *jump updating*, the long-term reference frame varies from as recent as frame $n-2$ to as old as frame $n-N-1$, where N is the jump update parameter. When encoding frame n, if the long-term reference frame is $n-N-1$, then, when the encoder moves on to encoding frame $n+1$, the short-term reference frame will slide forward by one to frame n, and the long-term reference frame will jump forward by N to frame $n-1$. The long-term reference frame will then remain static for N frames, and then jump forward again.

A second approach, called *continuous updating*, entails continuously updating the long-term frame buffer so that it contains a frame with a fixed temporal distance from the current buffer. Therefore, the long-term frame buffer always contains the $n-D$ frame for each frame n. D was called the continuous update parameter.

In [9], continuous updating was found to outperform jump updating for most sequences. However, both continuous updating and jump updating were viewed as a process of *choosing* a long-term reference frame from among a set of recently encoded frames, which are all largely equivalent and differ primarily in their temporal distance relative to the current frame being encoded.

In our approach we use jump updating, where a long-term reference frame is taken every N frames. In contrast to [9], we view the problem as one of *creating* rather than choosing a long-term reference frame, that is, allocating extra bits to a frame so that it will have high-quality. This long-term high-quality reference frame can then serve as a source of high-quality matching blocks for some time to come.

2.1 Scheduler

The use of a high-quality reference frame requires extra bandwidth/delay to transmit the frame. The allocation of extra bits can come from a scheduler. For example, we assume that there are multiple users utilizing the system resources which are allocated by a sched-

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

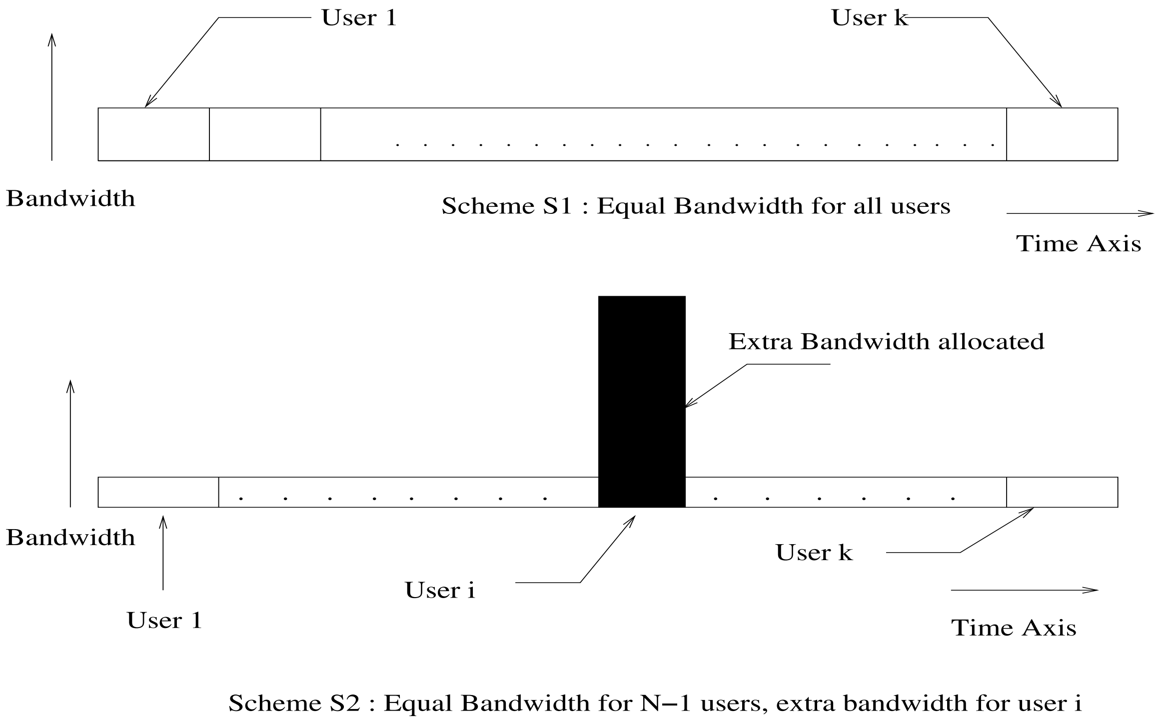


Figure 2: Schematic representation of the scheduler

uler. The system we consider is a common wireless network medium like an High Data Rate (HDR) system [13] which has rates of 400kbps. Scheduler S_1 divides the available bandwidth B from the network equally among the k users. Each user has a bandwidth of B/k. Scheduler S_2 reserves some portion b of the total bandwidth, and divides the remainder among all users. The reserved portion b will be allocated to each user in turn, cycling round-robin among the users. Each user has a bandwidth $(B - b)/k$ during $k - 1$ time slots, and has a bandwidth $b + (B - b)/k$ during one time slot. Figure 2 schematically represents the two schedulers, where the bandwidth is represented by the height and the horizontal axis is time. It is uniform for S_1, but high for User i in S_2. This extra bandwidth is allocated to the other $k-1$ users in the system at different times. The average bandwidth over all users remains the same at any point of time under both schedulers S_1 and S_2. In this paper, we compare the performance of the dual frame encoder using schedulers S_1 and S_2. The extra periodic bandwidth alloted by S_2 is used for creating a high-quality frame which is used as the long-term frame in the dual frame motion compensation scheme.

Although in this paper we assume that extra bandwidth comes from a group scheduler, our dual frame scheme with periodic high-quality frames does not require this. A single user with fixed average bandwidth can use extra bits for one frame (with fewer bits for the frames before and after) if some amount of extra delay can be tolerated in the video. The extra bits for the frame simply translate into more time for the transmission of that frame. This trade-off of compression performance and delay is the subject of future work. In the present work, we assume that no extra delay is incurred for the high-quality frames because extra bandwidth is periodically provided. However, we constrain the total average bandwidth to be equal to the bandwidth of the evenly allocated S_1 system.

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

2.2 Choice of Reference frame

Our modified MPEG4 encoder has three modes of coding from which to choose: intra coding, inter coding using the short-term reference frame, and inter coding using the long-term reference frame. For each macroblock (MB), we choose among them by first choosing between intra and inter coding as follows. We compute the distortion d_{min} between the best match MB in the short term or long-term frame and the current MB to be coded. We compute the standard deviation σ of the current MB. Intra coding is chosen if $\sigma < d_{min} - 512$, otherwise inter coding is chosen. The choice between inter coding from the short-term and long-term high-quality frames is done on the basis of the distortion between the motion compensated MB and the current MB to be encoded. The reference frame which yields a lower distortion is chosen.

Our MPEG-4 encoder treats each frame in the video sequence as a single object. The rate control method employed by the encoder is that described in [10]. It consists of three stages: pre-encoding, encoding, and post-encoding. The model parameters used in rate control are determined by using the Quantization Parameter (QP) and texture bits. The pre-encoding stage calculates the target bits for the current frame based upon the buffer fullness and the number of bits used by the previous frame. The Quantization Parameter is then estimated using the model parameters which were determined in the most recent post-encoding stage.

3 Results

In this section, we evaluate the implementation of our dual frame motion compensation with high-quality long-term frames using five video sequences: News, Foreman, Akiyo, Claire and Container. Each sequence had 300 frames. The News sequence consists of periodic background changes with news readers in the foreground. Foreman shows a man talking in a relatively static background with a scene change towards the end. The Container sequence depicts a ship moving slowly into the ocean. The Akiyo sequence is also a news reader sequence, but there are no background changes. Claire shows a woman using sign language against a static background. The format of all sequences is QCIF and the frame rate is 10 frames/second.

We evaluate a scenario where all users use a common shared access medium such as an HDR wireless network with rates of approximately 400kbps. We evaluated the impact of our encoder on a single user stream in the context of a system of 20 users. With 20 users, each user stream has baseline bandwidths ranging from 16.16 kbps to 16.64 kbps and periodically receives extra bandwidths ranging from 9.6 kbps for the Foreman sequence to 28.8 kbps for Claire.

We evaluate each sequence using four different versions of the encoder. "Regular Quality without Update" uses a dual frame buffer with a single long-term frame with Scheduler S_1. The single long-term frame is the first frame of the sequence, and it remains the long-term reference frame for the entire sequence. The results for this version are essentially the same as for a conventional single reference frame encoder that has rather even quality. "Regular Quality with Update" uses a dual frame buffer that periodically updates its long-

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

term frame again using Scheduler S_1. "High Quality" uses Scheduler S_2 to periodically create high-quality frames. The "High Quality" mode of encoding uses the dual buffer coding scheme using Scheduler S_2 wherein extra bits allocated to a particular user are used to create high-quality frames, but the long-term frame *is not updated*. Gains in PSNR come solely because of the propagation of high-quality by the mechanism of short-term predictive coding from the point of high quality frame creation. The results for "High Quality" are essentially the same as for a conventional single reference frame encoder that has pulses of high-quality. On the other hand, "High quality with Update" uses Scheduler S_2 to periodically create high-quality frames that the dual buffer encoder uses as the long-term reference frame. In this scheme, the long-term frame of the dual buffer encoder is updated to the newly created high-quality frame from the extra bits allocated by the scheduler S_2. Again, the average rate used by all the video sequences using either scheduler S_1 or S_2 is kept constant.

Figure 3 compares the performance of the various encoders on a variety of video sequences. It shows the results of the PSNR in dB versus period of updating for the News, Container, Foreman, Akiyo and Claire sequences. The amount of extra bandwidth required for the high quality is the same for a given video sequence with each jump update parameter but differs with video sequences because of different content. Each plot shows the results of simulating a video sequence with the four different versions of the encoder described above. Extra bandwidth was allocated to the user using scheduler S_2 for the encoder versions "High Quality" and "High Quality with Update" every N frames (which is the period of high-quality). This extra bandwidth is deducted from the other $N-1$ frames such that the average bandwidth is equal to the average bandwidth by using scheduler S_1 for encoder versions "Regular Quality" and "Regular Quality with Update". For example, to compute the average PSNR over 300 frames using "High Quality" with a refresh period of 20, we simulated the News sequence by adding, for every 20^{th} frame, 60% extra bandwidth (that is 60% more bits than the maximum number of bits used by any one frame in the most recent group of regular frames).

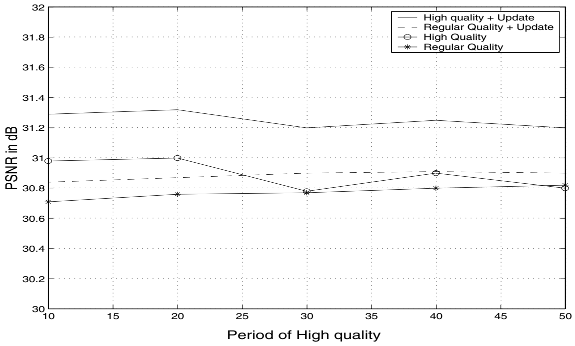
As can be seen from the plots in Figure 3, "High Quality with Update" performs significantly better than any of the other encoders. For example in the case of the Claire sequence, the average PSNR of "High Quality with Update" is better than "High Quality" by 0.3-0.4 dB and this in turn is better than "Regular Quality without Update" by another 0.2 dB.

"High Quality" without update generally outperforms "Regular Quality with Update" because the sequences are fairly static. When a sequence is static, propagation of high-quality from the high-quality frames by the mechanism of short-term predictive coding can be significant. For example, in the case of the Claire sequence, "High Quality" without updating provides around 0.3 dB gain over "Regular Quality with Update." Akiyo is also quite static, and the gains from "High Quality" alone are high. For static sequences, updating the dual frame buffer is less important than pulsing the quality.

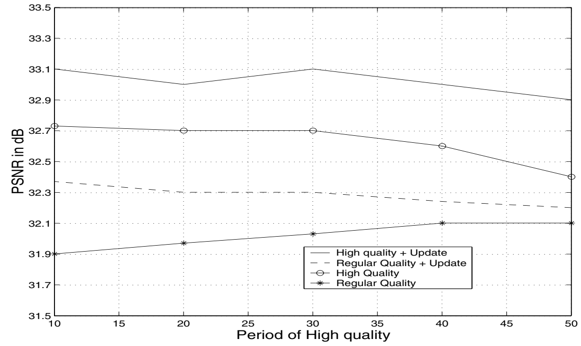
Finally the "Regular Quality" coding performs the worst as it has no updating of the long-term reference frame, and has a fairly even quality distribution among frames. It is about 0.6-0.7 dB worse for Claire than "High Quality with Update".

The Container sequence performs the best among the sequences we tried. This is because the container sequence gains most from "High Quality with Update", it consists of sequence consists of a ship moving slowly out of the picture. This relative motion between

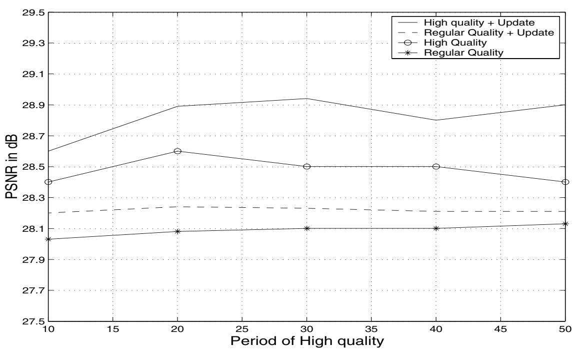
Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE



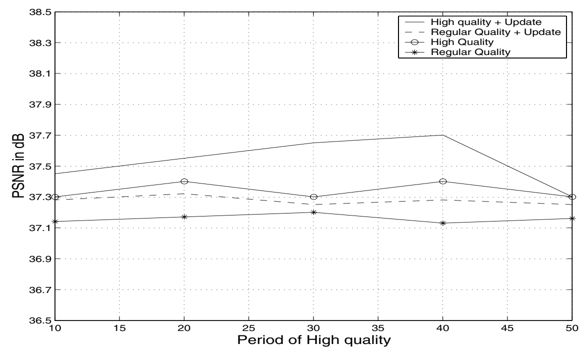
(a) News with 60% extra bandwidth



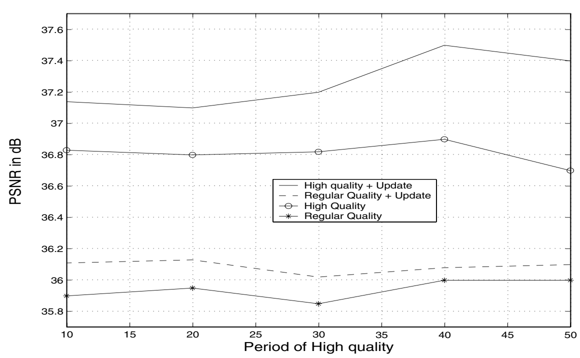
(b) Container with 45% extra bandwidth



(c) Foreman with 20% extra bandwidth



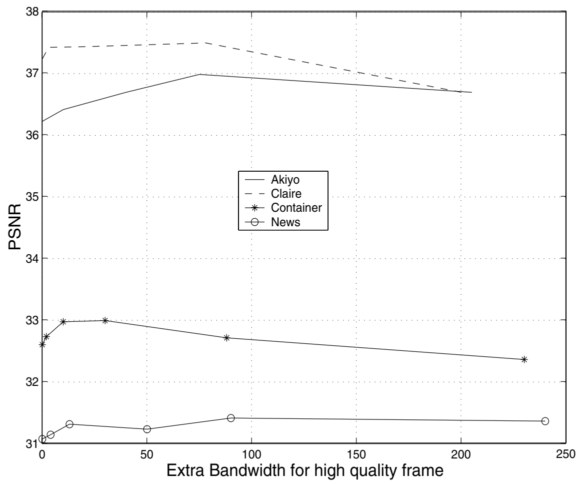
(d) Claire with 80% extra bandwidth



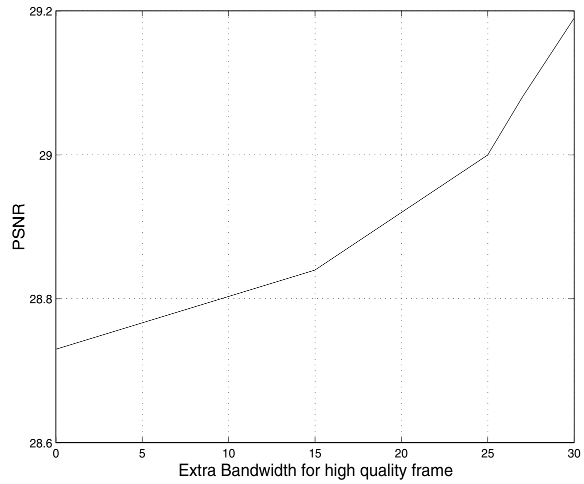
(e) Akiyo with 55% extra bandwidth

Figure 3: Plot of PSNR versus the period for updating the dual frame buffer. There is varying extra bandwidth allocated for different video sequences. This is because the number of extra bits allocated for the high quality frame depends upon the number of bits that can be stolen from the other frames as well as the original number of bits the frame requires for regular quality and both these factors are different for different video sequences.

1068-0314/04 $ 20.00 © 2004 IEEE



(a) All sequences except Foreman



(b) Foreman

Figure 4: Plot of PSNR versus amount of high-quality for a jump update parameter of 50. Foreman has a separate figure because the extra bandwidth allocation for Foreman is distinctly different. This is the result of extra motion vector coding required for Foreman for the same quantization parameter values as compared to other sequences.

the ship and the camera implies that portions of the ship are no longer present in the current and future frames, therefore, updating the long term frame would reflect the future trend more accurately and give performance gains. From a perceptual point of view, the Container and Akiyo sequences have noticeable improvements, but the perceptual gains for the other sequences are not manifestly evident.

We note that the curves for "Regular Quality" without Update do depend slightly on the update period even though no updating is being done for them. This is because of slight variations in the average bit rates for different periods. For example, coding with update in the "Regular" mode with a target rate of 16 kbps could produce an actual average bit rate of 15.9 kbps. When encoding the versions without update, we match this bit rate for a fair comparison and determine average PSNR. Since the bit rate changes slightly for different period simulations, the average PSNR versus period changes slightly even if we are encoding with no update.

The "High Quality" without update coding version also depends on the jump update period, as creation of the high-quality frame ensures the propagation of high-quality through short-term inter coding even though there is no long-term *frame update*.

In the next set of experiments, we study the effect of allocating more bits for the high quality frame on the average PSNR of the entire video sequence. We simulate the performance of the encoder version "High Quality with Update" by varying the amount of high-quality injected, and measuring the effects on PSNR. There is a limit to the amount of quality that can be given to the high-quality frame, since we cannot violate the rate constraint. That is, a very high-quality frame may consume so many extra bits that it is not possible to achieve an average bit rate of 16kbps. Without violating the rate constraint,

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

there is a trade-off in which giving too much rate to the high-quality frame will starve the other frames to the point that the average PSNR for the sequence declines.

Figure 4 shows the plots of PSNR for a "High Quality with Update" encoder versus the amount of high-quality injected into the frame for a jump update parameter of 50. The Foreman sequence has been separated from the other video sequences because the amount of extra bandwidth allocated to the high-quality frame is distinctly different from the other frames. This is because in the Foreman sequence, more bits are required for coding the motion vectors and the texture bits for the same quantization parameter as compared to the other video sequences due to more relative motion. Therefore, this implies that fewer bits can be assigned to the high-quality frame.

From these graphs, we see that the choice of how much extra bandwidth to give the high-quality frames can affect the final PSNR by up to 0.7 dB. Providing more quality to the high-quality frames beyond a certain point will cause the PSNR for the sequence to decline, and, additionally, it may cause the encoder to violate the rate constraint. There is an optimal choice of allocating extra bandwidth to the high-quality frame and this would be different for different video sequences, the subject of which is left for future work.

4 Conclusions

In this paper we describe the design, implementation, and evaluation of a dual frame motion compensation technique that allocates bits unevenly among frames to periodically create a high-quality frame that serves as the long-term reference frame. We showed that our proposed dual frame buffer for low bandwidth networks gave good PSNR gains over a standard dual frame buffer using a particular type of scheduler. We found that on average there was a 0.6 dB gain in the average PSNR by using "High Quality with Update" as compared to "Regular Quality with Update". We also showed that the average PSNR was somewhat insensitive to the period of high-quality frame creation. The variation in average PSNR for periods within the range 10-50 was usually 0.1-0.2 dB (although sometimes as high as 0.4 dB). Varying the quality given to the high-quality frame by changing the quantization parameter used to encode it significantly affected the overall PSNR (up to 0.7dB for the Claire sequence). We found that there is an optimum allocation of bandwidth for the high quality frame. The optimal selection of a fixed updating period and a fixed amount of extra bandwidth is a subject for future research. Further improvements might arise from not fixing these parameters, but rather dynamically changing these quantities during the encoding.

Acknowledgment: This research was sponsored by the California Institute for Telecommunications and Information Technology, by the CoRe program of the State of California, and by Ericsson Wireless Communications Incorporated.

References

[1] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[2] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 849–865, Nov. 1998.

[3] T. Wiegand, "Joint final committee draft for joint video specification H.264," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-D157, July 2002.

[4] M. Gothe and J. Vaisey, "Improving Motion Compensation Using Multiple Temporal Frames," *IEEE Pac. Rim Conf. on Comm., Comp. and Sig. Proc.,* v.1, pp. 157-160, 1993.

[5] M. Budagavi and J.D. Gibson,"Multiframe Video Coding for Improved Performance Over Wireless Channels," *IEEE Trans. Image Proc.,* vol. 10, pp. 252-265, Feb. 2001.

[6] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction," *IEEE Trans. Circand Syst. for Video Tech.*, vol. 9, pp. 70-84, Feb. 1999.

[7] T. Fukuhara, K. Asai, T. Murakamai,"Very Low Bit-Rate Video Coding with Block Partitioning and Adaptive Selection of Two Time-Differential Frame Memories," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 3, pp. 212-220, Feb. 1997.

[8] A. Leontaris and P. Cosman, "Video compression with intra/inter mode switching and a dual frame buffer," *Data Compression Conference 2003*, pp. 63-72, 2003.

[9] A. Leontaris and P. Cosman, "Dual Frame Video Encoding with Feedback," *Asilomar Conference on Signals, Systems, and Computers* November 2003.

[10] A. Vetro, H. Sun, Y. Wang, "MPEG-4 Rate Control for Multiple Video Objects," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 186-199, Feb. 1999.

[11] V. Chellappa, P. Cosman and G. Voelker, "Dual Frame Motion Compensation for a Rate Switching Network", *Asilomar Conf. on Signals, Systems, and Comp.* Nov. 2003.

[12] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49-55, Feb. 2003.

[13] A. Jalali, R. Padovani and R. Pankaj, "Data Throughput of CDMA-HDR : A High Efficiency High Data Rate Personal Communication Wireless System," *IEEE 51st Vehicular Technology Conference Proceedings*, vol. 3, 2000, pp. 1854-1858

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE

IEEE COMPUTER SOCIETY