Transactions Letters

Dual Frame Motion Compensation With Uneven Quality Assignment

Vijay Chellappa, Pamela C. Cosman, and Geoffrey M. Voelker

***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 explore using dual frame motion compensation in two contexts. We first show that using a single fixed long-term reference frame in the context of a rate switching network can enhance video quality. Next, by periodically creating high-quality long-term reference frames, we show that the performance is superior to a standard dual frame technique that has the same average rate but no high-quality frames.**

***Index Terms*—Dual frame buffer, long-term reference frame, motion compensation.**

I. INTRODUCTION

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 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 is compressed and transmitted, along with the 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 and the state-of-the-art H.264/AVC [2].

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 provides an advantage in compression performance [3]–[5]. Multiple frame prediction has been standardized in H.264/AVC [2]. Dual frame coding, depicted in Fig. 1, is a special case of multiple frame prediction in which only two reference frames are used, the *short-term reference* (STR) frame (the immediate past reference frame) and the *long-term reference* (LTR) frame [6]–[9]. For encoding frame n, the encoder uses the previous frame $n-1$ as the STR and the frame $n-k$ as the LTR frame.

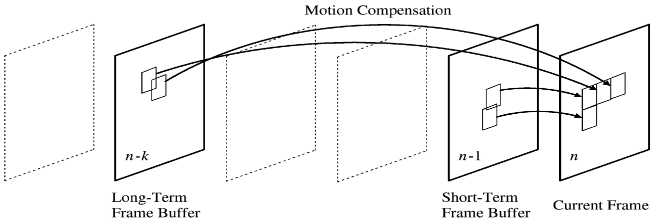


Fig. 1. Dual frame motion compensation.

In this paper, we explore the advantages of using uneven quality levels for the different reference frames. We consider two particular scenarios: a single fixed high-quality LTR frame, and a series of periodic high-quality LTR frames. The rest of this paper is organized as follows. Section II describes the performance of dual reference coding with a single fixed high-quality frame, an approach that benefits networks with large bandwidth variations. Section III describes the performance of dual reference coding with periodic high-quality frames, an approach that benefits low-bandwidth networks. We conclude in Section IV.

II. DUAL REFERENCE CODING WITH A SINGLE HIGH-QUALITY FRAME

We first apply the dual frame approach to wireless network settings that experience significant transitions in network capacity, such as those in a cognitive radio scenario or by network handoffs while using services like Always Best Connected (ABC) [12]. By using dual frame encoding in this context, the system can improve the quality of frames transmitted immediately after a drop in capacity, smoothing the abrupt and severe transition.

With the widespread use of wireless access networks and the deployment of 3G services, connectivity to the Internet with mobile devices such as PDAs and laptops needs to be maintained as users roam from one network to another. The ABC service supports various wireless networks: WLAN (11 Mbps–1 Gbps), HDR (High Data Rate, 500 kpbs), 1xRTT (150 kbps), and GPRS (10–30 kbps). ABC probes for network conditions and availability, seamlessly upgrading a user to a better network when it becomes available, and transitioning as gracefully as possible to a lower quality network if the current one becomes unavailable. The ABC network provides explicit support for notifying applications, service provider networks, and application servers about significant changes in network conditions, such as transitions from one access network to another [12]. When a network handoff occurs from high to low bandwidth, the ABC network

Manuscript received March 5, 2005; revised June 4, 2007. This research was sponsored by Ericsson, Inc., and by a UC Discovery Grant. Portions of this work appeared previously at the Asilomar Conference on Signals, Systems, and Computing, November 2003 and at the 2004 IEEE Data Compression Conference. This paper was recommended by Associate Editor W. Zhuz.

V. Chellappa and P. C. Cosman are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: v_chellappa@yahoo.com; pcosman@ucsd.edu).

G. M. Voelker is with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: voelker@cs.ucsd.edu).

Digital Object Identifier 10.1109/TCSVT.2008.916575

can use a signalling protocol to proactively notify the video encoder of the change, preventing the encoder from overwhelming the low bandwidth network with frames encoded for high rates.

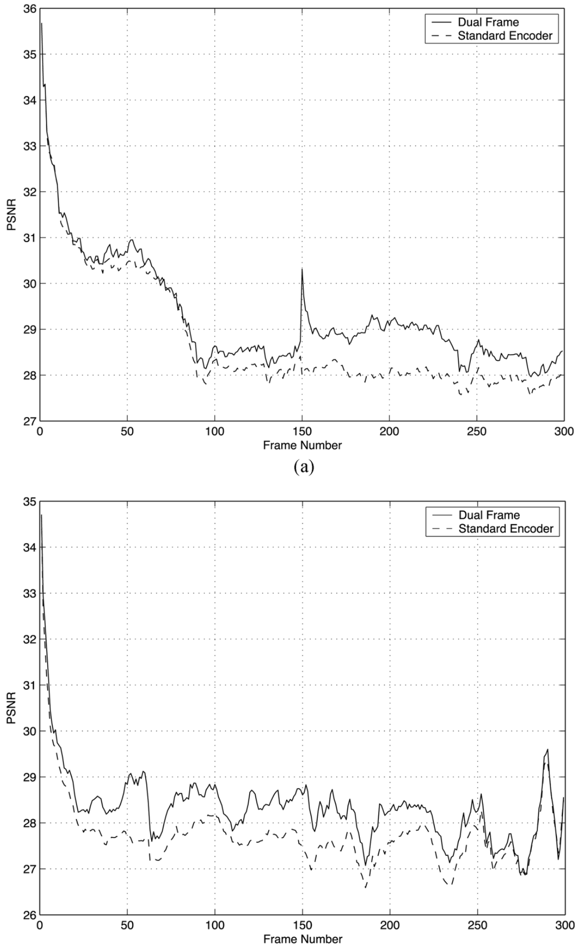
We used simulations to quantify the benefits of our dual frame approach for smoothing transitions from a high-bandwidth to a low-bandwidth situation. We assume that the network provides a timely delivery of packets with minimal loss. To counter the huge swings in bandwidth, we assume that the quantization parameter (QP) for each frame can be varied over its full range (1–31). We switch to a lower rate stream by adjusting the QP to achieve the desired bit rate. When a video application undergoes a substantial drop in rate, we assign the LTR frame to be the last frame coded just prior to the rate drop.

To evaluate the effectiveness of the dual frame buffer technique, we simulated it by modifying the standard MPEG-4 coder. The bit stream is standard compliant with the exception that, when inter coding is used, a single bit is added to the MB header to indicate use of the STR or the LTR frame. Note that, with H.264/AVC, this dual frame coding would be fully standard compliant. We used an intra coding period of 100 frames (lowering the intra coding period results in higher bit rates that exceed the bit rates available for a GPRS system). The bit stream consists of I- and P-frames only.

For each MB not in an I-frame, the encoder must choose among intra coding, inter-short-term coding, and inter-long-term coding. The encoder first chooses between inter-short-term coding and inter-long-term coding by computing the mean squared error (MSE) between the best match MB in the STR or LTR, and the current MB to be encoded. It selects the reference frame that yields the lower distortion $d_{\min}$. The encoder computes the standard deviation σ of the current MB and chooses between intra and inter coding as follows: if $\sigma < d_{\min} - 512$, intra coding is used, otherwise it chooses inter coding. This method of inter/intra choice is similar to the approach described in [10, p. 178].

Each frame is treated as a single object for the MPEG-4 encoder. Rate control is employed on each frame as in [11], consisting of pre-encoding, encoding and post-encoding. The model parameters used in the rate control are determined by the texture bits and the QP. We linearly combine the bits used in the previous frame and the expected bits per frame (which is the ratio of the remaining bits to the remaining frames) with weights of 0.05 (called past percentage) and 0.95. This is the initial value of target bits in the pre-encoding stage. The buffer size is 64 000 bits. The target bits for the next frame is refined from the remaining buffer level to ensure that there is no buffer underflow or overflow. The post-encoding stage then estimates the QP using the model parameters determined in the most recent post-encoding stage. Details of this scheme can be found in [11].

We used the News, Container, and Foreman QCIF sequences, each with 300 frames, 30 frames per second (fps), YUV 4:2:0 format. The foreground of News is fairly static, and its background shows a revolving dancer who returns to the starting position and then repeats. We expect dual frame coding to be effective for the static foreground for some time after the rate drop occurs, as well as for the repeating background. Foreman shows a man talking in a fairly static background with a scene change towards the end. The Container sequence depicts a ship moving slowly in the ocean with no significant background change. To investigate the effects of switching to different low bandwidth networks, we simulated switching from 1 Mbps to low bandwidth networks ranging from 10 (GPRS) to 150 kbps (1xRTT CDMA). We encoded each sequence using our dual frame coder as well as with a conventional MPEG-4 coder for comparison.



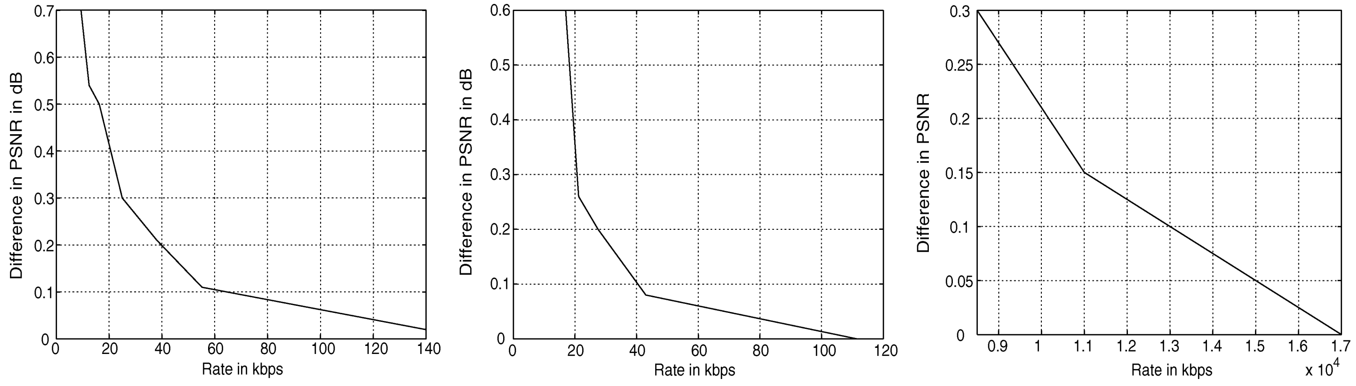
(a)

(b)

Fig. 2. PSNR in decibels versus frame number. (a) News sequence. (b) Foreman sequence.

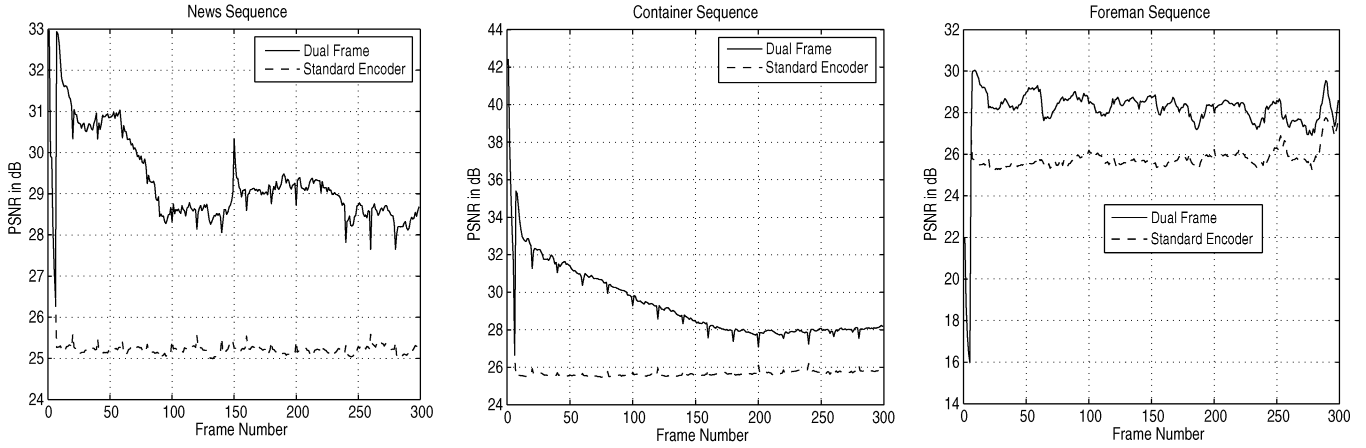
A. *Results for Rate-Switching Networks*

We assume a signalling protocol with no significant latency in the ABC network, allowing the encoder to adjust its coding rate without loss when there is a network switch. Later we examine the effects of signalling delays that result in loss. Fig. 2 shows the PSNR in decibels for each decoded frame for both the MPEG-4 standard and the dual frame encoder. The high bit rate for the plots in Fig. 2 is 1 Mbps, and the low bit rate is 16 kbps. The News sequence shows gains primarily after the 150th frame because, when the dancer returns to the starting position, the background closely matches the LTR background, causing an upward spike in peak signal-to-noise ratio (PSNR) at frame 150. This PSNR gain propagates to subsequent frames. The Foreman sequence initially shows gains of around 1 dB in PSNR. The gain disappears completely towards the end of the sequence (at frame 268) because of a scene change. The Container sequence also shows that the value of the LTR diminishes over time.



(a) (b) (c)

Fig. 3. Average difference in PSNR in decibels. (a) News sequence. (b) Foreman sequence. (c) Container sequence.



(a) (b) (c)

Fig. 4. News, container and foreman sequence when the transition from a high to low bandwidth network results in five lost frames. (a) News transition. (b) Container transition. (c) Foreman transition.

Fig. 3 plots the average difference in PSNR between the dual frame coder and the standard MPEG-4 coder versus the bit rate of the low rate connection. In this plot, we first compute the MSE, convert this to PSNR, and then average the PSNR over the 300 frames in the sequence. The PSNR gap diminishes as the low rate connection increases in rate. Only the first frame of the sequence is transmitted at high quality, assumed to be part of the high-rate (1 Mbps) portion of the transmission. The remaining frames (2–300) are at the low bit rate. As expected, the PSNR gains in these plots decrease as the rate of the low bandwidth connection increases. The difference in quality for most of the sequences was not perceptible except for the News sequence. This is because the first LTR frame and the 150th frame are very similar, yielding a visible quality improvement.

If the response of the signalling protocol is delayed, the encoder does not immediately know to switch to a low bit rate when there is a bandwidth transition. In such a case, we will inevitably lose some video frames. When the encoder discovers the bandwidth transition, it will transmit an intra coded frame to re-establish synchronization between the encoder and decoder. The high quality long term frame is still known to both the encoder and the decoder, which is the last frame to be transmitted without error.

To examine the effects of delay in the signalling protocol, we simulated the News, Container, and Foreman sequences when a network transition results in lost frames. In each case we assumed that, due to the transition delay, five frames are lost in the process. The decoder conceals each of the five lost frames by replacing them with the last frame that was losslessly received. The sixth frame is intra coded to ensure that the encoder and decoder are resynchronized. These results show that, with lost frames, the dual frame scheme significantly outperforms the standard encoder (see Fig. 4). For the News sequence, for example, the average PSNR using the dual frame encoder is 29.34 dB compared with 26.37 dB using the standard video encoder, yielding a PSNR gain of close to 3 dB.

We note that the performance improvement using dual frame coding over regular coding is substantially larger when a delay in signalling is assumed, with concomitant loss of frames, compared to the performance improvement achieved during a smooth transition to low bandwidth. Following the loss of frames, both versions of the coder must create an I-frame to re-establish synchronization, and this I-frame will be of fairly low quality, because it is produced during a low-rate connection. The dual-frame coder, which still retains as reference the last frame which was transmitted before the transition, can benefit from the high quality of that frame for motion compensated prediction, even though, because of the assumed loss of five frames, the reference starts out being at least six frames old.

From our results, we conclude that using a dual frame encoder with a static high-quality LTR improves video quality for a short period of time. Seeing that gains with a high-quality frame were initially substantial, but diminished rapidly over time, we next considered the possibility of a periodic refreshment of the high-quality LTR.

III. DUAL FRAME CODING WITH PERIODIC HIGH-QUALITY FRAMES

Given the results with the rate-switching network scenario, we reasoned that, in the absence of any rate switches or alternative connections, a user with a low-bandwidth connection at 16 kbps might still benefit from periodically creating a higher quality frame and then retaining this frame for some time as an LTR.

Creating a high-quality reference frame in a low bandwidth situation requires extra bits, so either extra bandwidth, or extra time, will be required to transmit that frame. There are several practical scenarios for transmitting the high quality frame using extra bandwidth. In one scenario, extra bandwidth can be allocated by a scheduler. Instead of dividing the available bandwidth B equally among the k users, a scheduler that is trying to accommodate high-quality frames could reserve some portion b of the total bandwidth, and divide the remainder among all users. The reserved portion will be allocated to each user in round-robin turn. Each user has a bandwidth $(B-b)/k$ during $k-1$ time slots, and has $b+(B-b)/k$ during one time slot. The total for all users remains B at any point of time. The extra periodic bandwidth allotted can be used for creating a high-quality LTR.

A second plausible scenario for transmitting periodic or occasional high quality frames using extra bandwidth is with a cognitive radio (e.g., [16]). If some portion of spectrum is sensed to be temporarily unused, the cognitive radio can exploit it, perhaps through a short-term rental of bandwidth. In a third scenario, the video encoder will simply take more time to transmit a high-quality frame, thereby incurring delay. The subsequent frames, transmitted at a low rate, will allow the decoder to "catch up" in its delay. From the point of view of the final viewer, this variable delay will not be noticeable since it would be encompassed within the overall delay budget. For example, suppose a user initiates a streaming application running at 30 fps. The application may send one second's worth of data (30 frames) which are buffered in compressed form at the receiver before display begins. After 1 second, the decoder begins displaying frames at a constant rate of 30 fps. If a high quality frame is sent at some point which requires many bits (say, five times the average number of bits for a frame), the decoder input buffer will then temporarily contain fewer than the usual number of frames. The encoder follows the high quality frame with some "regular" frames requiring few bits each, so the decoder recovers to again have roughly one second's worth of frames available to display. This process is transparent to the viewer, who sees frames display at a regular 30 fps. The underlying variable delay incurred in these frames is not apparent since it is smoothed out by the buffering made possible by an initial short delay in the display start.

We note that, in this scenario, the size of the video encoder output buffer (and of the video decoder input buffer) for a single user has to be large enough to accomodate the extra bits required for the high quality frames over and above that needed for rate control and to smooth network variations. The increase in video quality therefore comes at the cost of additional memory and of slight extra starting delay.

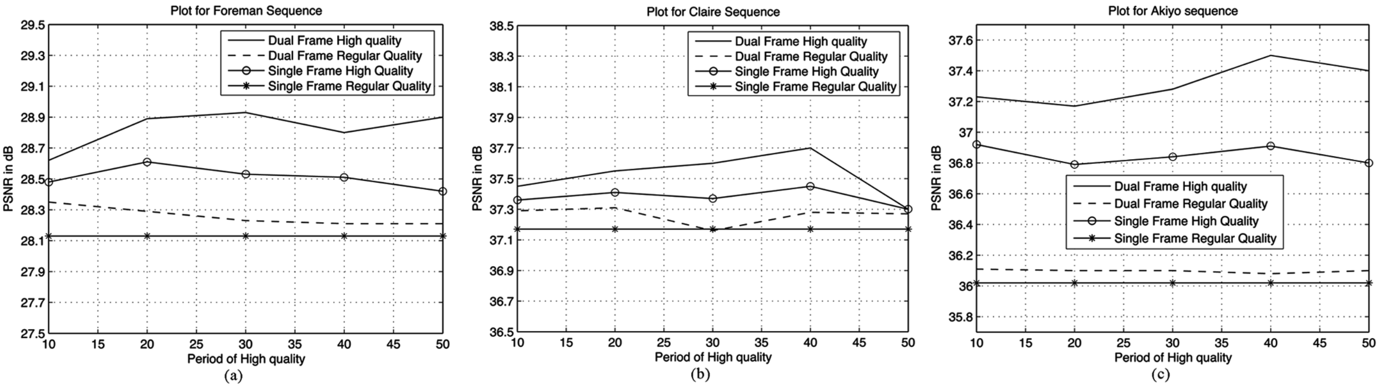
To evaluate the effectiveness of dual frame coding, we modified MPEG-4 to implement four encoders. The average rates used by the video sequences using any of these four coders are the same. SF-RQ (Single Frame Regular Quality) uses the standard encoder with a single reference frame. Regular quality means that there is no special assignment of higher quality to any frames. With regular quality there may still be variation in the actual number of bits used to describe different frames, because of the usual mechanisms of rate control, amount of motion, complexity of the frame, and so forth. DF-RQ (Dual Frame Regular Quality) uses a dual frame buffer that periodically updates its LTR, without special assignment of high quality to any frames.

SF-HQ (Single Frame High Quality) assumes that some mechanism is used to accommodate extra bits in a periodic fashion to allow the creation of high-quality frames. SF-HQ uses standard MPEG-4. 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. DF-HQ (Dual Frame High Quality) periodically creates high-quality frames, and the dual buffer encoder uses them as the LTRs. Here, if a high-quality frame is created out of frame n of the sequence, then it will be used as the STR when frame $n+1$ is encoded, and it will be used as the LTR starting with frame $n+2$. In doing so, it will replace the previous LTR in the LTR frame buffer. We refer to the time between updates to the high-quality frame as the refresh period characterized by a jump update parameter.

A. *Results for Periodic High-Quality*

In this section, we evaluate the scheme using five video sequences: News, Foreman, Akiyo, Claire, and Container. All users use a common shared access medium with a rate of approximately 400 kbps. We evaluate 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 16.16–16.64 kbps and periodically receives extra bandwidths ranging from 9.6 kbps for the Foreman sequence to 28.8 kbps for Claire.

Fig. 5 shows the results of the PSNR in decibels versus the period of updating. 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 bits were allocated every N frames for the encoder versions SF-HQ and DF-HQ, so N 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 for the regular quality encoder versions. For example, to compute the average PSNR over 300 frames using



(a) (b) (c)

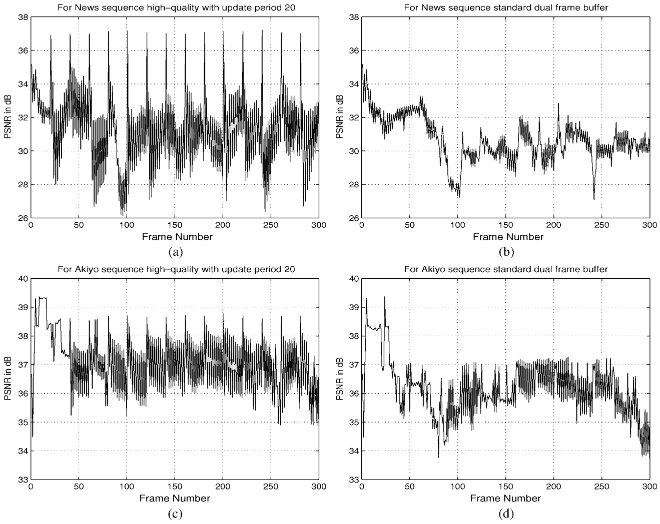
Fig. 5. PSNR versus the period for updating the dual frame buffer. Note that the extra bits allocated for the high-quality frame depends upon the number of bits taken from the other frames, as well as the original number of bits the frame requires for regular quality. Both these factors are different for the various video sequences. (a) Foreman with 20% extra bits. (b) Claire with 80% extra bits. (c) Akiyo with 55% extra bits.

SF-HQ with a refresh period of 20, we simulated the Claire sequence by adding 80% extra bits for every 20th frame.

As can be seen in Fig. 5, DF-HQ performs significantly better than any of the other encoders. For example, for Claire, the average PSNR of DF-HQ is better than SF-HQ by 0.3–0.4 dB and this in turn is better than SF-RQ by another 0.2 dB. SF-HQ generally outperforms DF-RQ 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 Claire, SF-HQ provides around 0.3 dB gain over DF-RQ. Akiyo is also quite static, and the gains from SF-HQ alone are substantial. For static sequences, updating the dual frame buffer is less important than pulsing the quality. Finally, the SF-RQ coding performs the worst among the four versions of the video coder. It is about 0.6-0.7 dB worse than DF-HQ for Claire.

In reporting these results, we note that the high-quality frames that are periodically created can, in principle, unfairly weight the average PSNR. As an extreme example, if one of the high-quality frames were created *losslessly* using the extra bandwidth, that frame would have infinite PSNR. As a result, even if the other frames are starved of bits, the average PSNR would be infinite. To avoid this effect, we compute the average PSNR without including the high-quality frames in the calculation. The results are similar to Fig. 5 and demonstrate that even without taking into account the "high-quality" frames, PSNR improvements for the video sequences are significant. The "high quality" created periodically is transferred to the other "regular quality" frames, thus improving the overall quality of the video.

One side-effect of using periodic high-quality frames to improve overall quality is that these high-quality frames could introduce perceptible and objectionable pulses in quality over time. To explore the extent of this effect, we examined both PSNR over time and the overall variation in PSNR for our video sequences. The graphs in Fig. 6 show the variation of PSNR over time by plotting PSNR in decibels versus frame number. We show results for News and Akiyo because they had the largest and smallest differences in standard deviation between the DF-HQ and DF-RQ encoders. In evaluating the sequences visually, we found that the pulsing is not perceptible. We speculate that this is in part because our chosen levels of pulsing extra bits are not extreme, and in part because high-quality propagates into subsequent frames, so the pulsing is visually smoothed out.



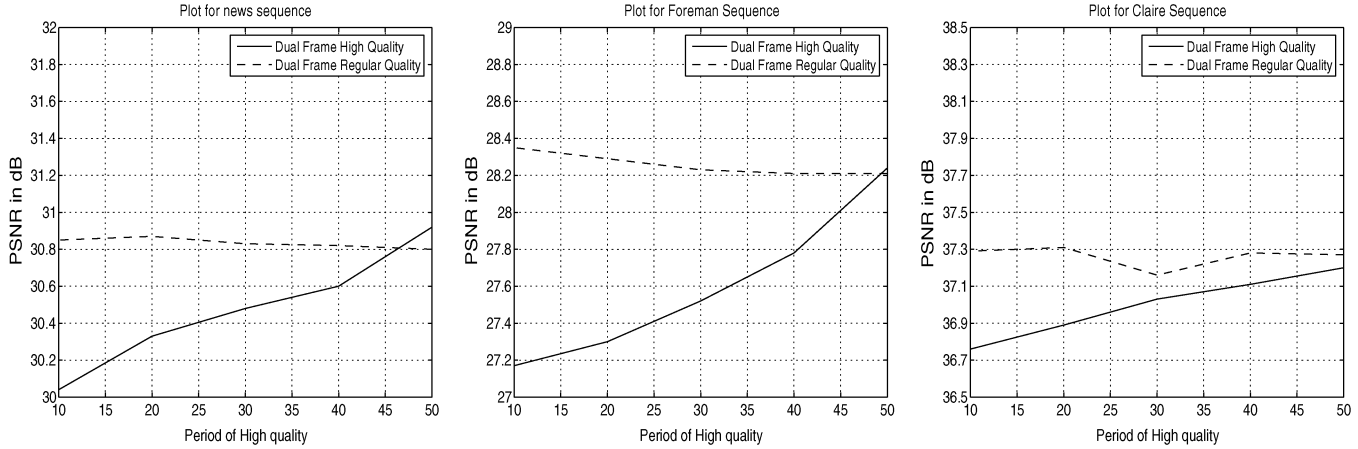
(a) (b) (c) (d)

Fig. 6. Plots of PSNR versus the frame number. (a) News using "dual frame high quality." (b) News using "dual frame regular quality." (c) Akiyo using "dual frame high quality." (d) Akiyo using "dual frame regular quality."

TABLE I
STATISTICS OF PSNR FOR DF-HQ AND DF-RQ VIDEO ENCODERS

Seq.	Encoder	Min dB	Max dB	Average dB	Deviation
News	DF-RQ	27.08	35.18	30.76	1.336
News	DF-HQ	26.13	37.21	31.3	2.39
Container	DF-RQ	28.69	35.43	32.13	0.9068
Container	DF-HQ	27.94	36.95	33.2	1.76
Foreman	DF-RQ	24.54	34.6	28.13	1.31
Foreman	DF-HQ	24.45	34.61	28.89	1.356
Akiyo	DF-RQ	33.61	39.3	36.02	1.01
Akiyo	DF-HQ	34.47	39.36	37.17	1.0774
Claire	DF-RQ	34.62	39.9	37.17	0.9
Claire	DF-HQ	33.97	41.08	37.55	1.4644

As another perspective, Table I presents the overall variation in PSNR by showing the minimum value, maximum value, the mean and the standard deviation in decibels for each of the video



(a) (b) (c)

Fig. 7. Plots of PSNR versus the period for updating the dual frame buffer. The high-quality frame is always the I-frame for the sequence. (a) News with 60% extra bits. (b) Foreman with 20% extra bits.(c) Claire with 80% extra bits.

TABLE II
PERCENTAGE OF MOTION VECTORS THAT POINT TO THE STR AND LTR FRAMES FOR JUMP UPDATE PARAMETERS OF 20 AND 50

	Short-term %		Long-term %	
Sequence	Period of 20	Period of 50	Period of 20	Period of 50
News	35.5	44.75	64.5	55.25
Container	52.17	63.81	47.83	36.19
Foreman	66.52	70	33.48	30
Akiyo	26.27	36.72	73.73	63.28
Claire	63.34	72.98	36.66	27.02

streams. For the comparison, we encode each of the streams using DF-HQ and DF-RQ. From these results, we see that the standard deviations are comparable when using periodic high-quality frames and periodic regular-quality frames.

Rate control strategies are generally used where the number of bits for the I-frame is increased at the expense of the P-frames. By making I-frames the high-quality frames, we conducted another set of experiments to see the effect on average PSNR. Fig. 7 shows the results of the simulation for three of the sequences. We found that, in general, the average PSNR is worse than DF-RQ. I-frames have lower compression efficiency than P-frames. Because I-frames require many bits to encode, not many bits are left for the remaining “regular” frames and they are hence starved of bits. Because of this, the average PSNR of the sequences decreases as the frequency of I-frames increases. This effect is illustrated in the graphs. Overall, as the frequency of I-frames reduces simultaneously with the high-quality update period, the average PSNR improves for all sequences.

Table II shows the percentage of motion vectors that point to the STR and LTR frames for jump update parameters of 20 and 50. The percentage of MVs pointing to the LTR frame is higher for static sequences, such as Akiyo. The Foreman sequence has a low fraction of MVs pointing to the LTR frame because it is a high motion sequence. The same pattern appears for a refresh period of 50, but the absolute fraction of MVs pointing to the LTR frame for each sequence is lower than that for a refresh period of 20.

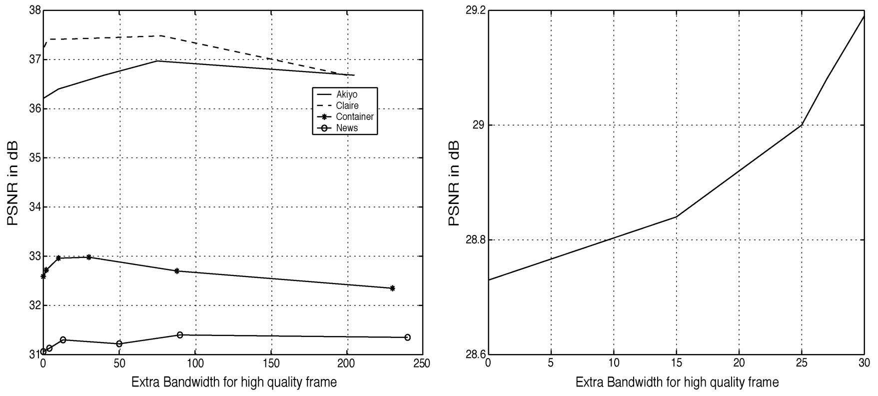
In the next set of experiments, we study the effect of allocating more or fewer bits for the high quality frame on the average PSNR of the entire video sequence. We simulate the performance of the DF-HQ encoder by varying the amount of high quality injected, and measuring the effects on PSNR. There is a limit to the amount of quality that can be given to the high-quality frame since we cannot violate the overall rate constraint for the entire sequence. 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 16 kbps. Without violating the rate constraint, there is a tradeoff 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.

For a DF-HQ encoder, Fig. 8 shows the plots of PSNR versus the amount of high-quality injected into the frame for a fixed jump update parameter of 50 frames. We separate Foreman from the other sequences because the amount of extra bandwidth allocated to the high-quality frame is distinctly different compared to the other sequences. Because of the higher motion content of Foreman, 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. As a result, 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. Dynamically determining this optimum for a given video sequence is an open problem. From a perceptual point of view, the Container and Akiyo sequences have noticeable improvement, but the perceptual gains for the other sequences are not as apparent.

B. Multiple Frame Motion Compensation

So far we have studied the benefits of using a dual frame encoder. A natural extension to this approach is to use additional



(a) (b)

Fig. 8. Plots of PSNR versus the percentage extra bit rate given to the high quality frames. The jump update parameter is 50. (a) All sequences except Foreman. (b) Foreman.

reference frames. As a final experiment, we explore the effects of a third reference frame on PSNR gains. We add an extra LTR to the encoder in memory so that each MB was predicted using three reference frames: the immediate past frame and two LTR frames. We explored allocating the extra bandwidth both equally and unequally to the two LTR frames. We simulated the three reference frame video encoder for the News, Foreman and Container sequences. Although the triple frame buffer outperforms the dual buffer for all three sequences, we observed only small gains of 0.1–0.2 dB in PSNR on average. As one might expect, the performance gap between dual frame and triple frame coding is much smaller than that between single frame and dual frame coding. Adding more reference frames beyond just dual frame encoding yields sharply diminishing returns, albeit with an essentially linear increase in encoder complexity for each additional reference frame being added.

IV. CONCLUSION

Scalable video coding (SVC), such as fine grain scalablity coding in MPEG-4 [17], has also been proposed as being useful for networks with fluctuating rates, it is worth discussing how SVC relates to the research on pulsed quality video described in this paper. SVC has many varieties. With a resolution-scalable or spatially scalable bit stream, decoding a portion of the compressed stream can yield a smaller sized image. Likewise, with a temporally scalable bit stream, decoding a portion of the compression stream can yield video at a lower frame rate. Quality-scalable or SNR-scalable coding usually involves a base layer, which contains the most important information such as coding modes and motion vectors, and an enhancement layer, which may contain the DCT residuals as a progressively encoded (embedded) bit stream.

Resolution-scalable streams are useful particularly when the video encoder does not know what the display capabilities of the end device are. Likewise, temporally scalable streams are useful particularly when the video encoder does not know the complexity capabilities of the end device, or what the bandwidth of the transmission will be. An encoder can pre-encode a video and store it on a server. An end device with limited display capability can download a reduced-size version, and a device with limited computational power or bandwidth might choose to download a version at reduced frame-rate. Rather than storing multiple versions of the video on the server, the encoder can pre-encode one stream, which has embedded within it all the sub-streams necessary for serving out to these different users. Likewise, quality scalability is also aimed at unknown fluctuating bandwidth.

In contrast, our work on periodic creation of high-quality frames is aimed at situations where the fluctuations in bandwidth are learned (or known in advance) and reacted to in real-time by the encoder. For example, we assume a scenario such as a central scheduler, which provides extra bandwidth on a known round-robin basis. We also consider a fixed-bandwidth situation where the user creates high-quality frames by using up some of the delay budget for those frames. Our scenarios also included cognitive radio, where we assume that extra bandwidth can be rented by the fraction of a second. In all these cases, the bandwidth is assumed to be known, and the encoder can take advantage of it at the moment. For the rate-switching network as well, we also assume that the encoder knows about the bandwidth change when it occurs (or shortly thereafter).

In this paper, we describe the design, implementation and evaluation of a dual frame video encoder using a high-quality LTR frame. The high-quality LTR frame can be retained from a high bandwidth network or can be periodically created. The main contribution of our work has been showing that pulsing the quality of a dual frame coder outperforms both a pulsed-quality single frame coder and a regular (nonpulsed) quality dual frame coder.

For a rate switching network scenario, we find that retaining the last high-quality frame from the high-bandwidth connection as the LTR in the dual frame encoder resulted in better video quality for a few hundred frames. Such benefits can smooth the transition from high-bandwidth to low-bandwidth situations for roaming users.

Periodic high-quality LTR creation resulted in 0.6 dB better performance than with a regular dual frame encoder. The average PSNR is not strongly dependent on the period of high-quality frame creation except for the Container sequence. In general, panning sequences would have strong correlation between PSNR improvement and refresh period for the high-quality frames. We found that there is a tradeoff between the amount of high-quality injected and the average PSNR of the sequence. We speculate that an adaptive method of choosing the quantization parameter controlling the amount of quality of the high-quality frame would further improve quality. In addition, adaptive choice of the temporal position of the high-quality frames might substantially outperform the fixed periodicity of the high-quality frames used in the current work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their many helpful comments.

REFERENCES

[1] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[2] 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, Jul. 2002.

[3] M. Gothe and J. Vaisey, "Improving motion compensation using multiple temporal frames," in *IEEE Pac. Rim Conf. Commun., Comp. Signal Process.*, 1993, vol. 1, pp. 157–160.

[4] N. Vasconcelos and A. Lippman, "Library-based image coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Apr. 1994, vol. v, pp. 489–492.

[5] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 70–84, Feb. 1999.

[6] T. Fukuhara, K. Asai, and T. Murakamai, "Very low bit-rate video coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 212–220, Feb. 1997.

[7] C.-S. Kim, R.-C. Kim, and S.-U. Lee, "Robust transmission of video sequence using double-vector motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 1011–1021, Sep. 2001.

[8] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 252–265, Feb. 2001.

[9] A. Leontaris and P. Cosman, "Video compression with intra/inter mode switching and a dual frame buffer," in *Proc. IEEE Data Compression Conf. 2003*, 2003, pp. 63–72.

[10] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*. Englewood Cliffs, NJ: Prentice-Hall.

[11] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 186–199, Feb. 1999.

[12] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Commun.*, vol. 10, no. 1, pp. 49–55, Feb. 2003.

[13] V. Chellappa, P. Cosman, and G. Voelker, "Dual frame motion compensation for a rate switching network," in *Proc. Asilomar Conf. Signals, Syst., Comp.*, Nov. 2003, pp. 338–354.

[14] V. Chellappa, P. Cosman, and G. Voelker, "Dual frame motion compensation with uneven quality assignment," in *Proc. IEEE Data Compression Conf.*, 2004, pp. 895–902.

[15] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 70–84, Feb. 1999.

[16] J. Mitola, III and G. Q. Maguire, Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, pp. 13–18, Aug. 1999.

[17] W. Li, "Overview of fine granularity scalability in the MPEG-4 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301–317, Mar. 2001.