RATE-DISTORTION OPTIMIZED DELIVERY OF JPEG2000 COMPRESSED VIDEO WITH HIERARCHICAL MOTION SIDE INFORMATION

Aous Thabit Naman and David Taubman

University of New South Wales,
School of Electrical Engineering and Telecommunications,
Sydney, Australia.

ABSTRACT

Streaming video as a sequence of JPEG2000 images provides the scalability, flexibility, and accessibility at a wide range of bit-rates that is lacking from the current motion-compensated predictive video coding standards; however, streaming this sequence requires considerably more bandwidth. The authors have recently proposed a novel approach that reduces the required bandwidth; this approach uses motion compensation and conditional replenishment of the JPEG2000 code-blocks, aided by server-optimized selection of these code-blocks. This work extends the previous work to the case of hierarchical arrangement of frames, similar to the hierarchical B-frames of the SVC scalable video coding extension of the H.264/AVC standard. We employ a Lagrangian-style rate-distortion optimization procedure to the server transmission problem and compare the performance to that of streaming individual frames and also to that of predictive video coding. The proposed approach can serve a diverse range of client requirements and can adapt immediately to interactive changes in client interests, such as forward or backward playback and zooming into individual frames. This paper introduces the concepts, formulates the optimization problem, proposes a solution, and compares the performance to alternate strategies.

Index Terms— Teleconferencing, video signal processing, image coding, image communication.

1. INTRODUCTION

Scalable video is a highly sought after solution to the problems of accommodating the varying needs of different clients and to adapt to changing conditions of the network. Existing video standards such as MPEG-1 through MPEG-4 and H.261 through H.264 offer at best limited scalability. As such, scalable video compression has received great attention in the last few years with very promising results [1] [2] [3]. To improve the available scalability options in existing standards, a scalable video coding (SVC) extension to H.264/AVC has been recently approved within the ISO working group known as MPEG; however, this still imposes restrictions on the video structure, which limits video accessibility for streaming applications. For example, if the client is interested in a high quality version of one frame, a large number of other frames might need to be reconstructed.

Recently, the authors have presented a new approach for serving scalable video [4] [5]. This approach utilities JPEG2000 to independently compress the individual frames and provide for quality and spatial resolution scalability. To reduce inter-frame redundancy the approach relies on motion compensation and optimized selection of JPEG2000 code-blocks. These goals are achieved through the cooperation of loosely coupled server and client policies. The server policy dynamically determines which code-blocks to send, while the client policy determines how best to make use of the data which is received from the server, possibly in conjunction with some motion model.

In [4], a realistic implementation of the proposed approach is examined; however, that work is limited to the special case in which redundancy is exploited only within disjoint frame pairs. In [5], the server optimization policy is extended to a sliding window of sequential frames, with the potential to exploit the redundancy between any two frames within the window. In this paper, we turn our attention to optimizing a hierarchical group of frames, similar to the hierarchical B-frames dyadic prediction structure used by the SVC extension of H.264/AVC. The results we present here are idealistic, in the sense that the client is assumed to have all the information required to make optimal decisions regarding the data it receives from the server. This allows us to focus only on the server policy in this paper.

A recent work by Devaux et al. [6] investigates a problem similar to the one in this paper; however, that work deals only with sequential prediction and does not employ motion-compensation. Another recent work by Cheung and Ortega [7] is similar to our work in attempting to enable flexible video delivery by dealing with the motion model and the residual distortion as side data; unlike the approach proposed here, theirs is based on distributed video coding.

The approach proposed in this paper is suited to a diverse range of applications, of which we briefly mention some here; the interested reader can also refer to [4] [5]. For clients with limited processing power, the server can limit the streamed video to a low resolution, or to the desired region of interest, without recoding the video. In fact, the server can easily support clients which are unable to perform motion compensation. For lossy transmission environments, the server does not need to retransmit lost packets, instead it can adapt by adjusting its delivery policy for future frames alone. In this paper, we choose to examine a scenario in which the client may be interested in viewing a particular frame from the video, in addition to browsing the streamed video as a whole. Such a scenario is common in video surveillance browsing applications and also interesting for video editing. We compare the performance of motion compensated predictive video coding with that of our proposed rate-distortion optimized delivery system of intra-coded video frames for both the single-frame case and the streamed video case.

The remainder of the paper is structured as follows. Sections 2 and 3 elaborate on our client and server policies, explaining their theoretical aspects. Section 4 provides experimental results. Finally, Section 5 states our conclusions.

2. CLIENT POLICY

Frames are arranged in a dyadic hierarchical structure with temporal decimation levels $T_0, T_1, \ldots T_K$. Figure 1 shows this structure for the case of $K = 3$. Each frame belongs to one or more temporal decimation level T_k depending on its position. Frames at temporal level T_K are not predicted from any other frame (intra-frames). At temporal levels with $k < K$, the code-blocks can either come from the frame itself or else they are obtained from motion compensating the two nearby frames as shown in Figure 1. We write $f^k_{\rightarrow n}$ for the

predicted frame given by

$$f^k_{\rightarrow n} = \frac{1}{2}\left(\overrightarrow{\mathcal{W}}_{n-p\rightarrow n}\left(f^k_{n-p}\right) + \overrightarrow{\mathcal{W}}_{n+p\rightarrow n}\left(f^k_{n+p}\right)\right)$$
$$= \frac{1}{2}\left(f^k_{n-p\rightarrow n} + f^k_{n+p\rightarrow n}\right) \quad (1)$$

where $\overrightarrow{\mathcal{W}}_{a\rightarrow b}$ is the motion compensation operator mapping f_a to f_b and $p = 2^k$ is the distance between a frame and the two adjacent frames at the k^{th} temporal level. Other choices are possible and in fact are often used in the literature; however, we choose not to consider them for the present study.

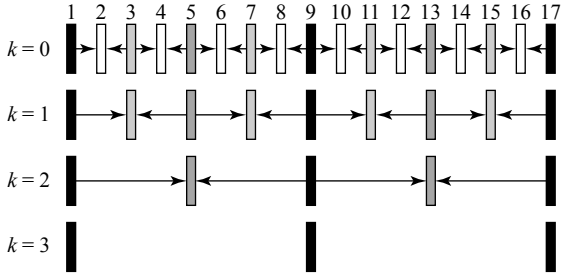

Fig. 1. Two groups of pictures in the dyadic hierarchical structure. The arrows show prediction directions and the numbers at the top are frame indices.

For each code-block C^β_n of each frame f_n the client receives some number of quality layers q^β_n, possibly 0. We write $\tilde{C}^\beta_n$ for the dequantized samples which can be recovered from these layers and $\tilde{D}^\beta_n = \|\tilde{C}^\beta_n - C^\beta_n\|^2$ for the corresponding distortion. For each code-block, the client chooses the sub-band samples in $C^\beta_{\rightarrow n}$ obtained from the two-dimensional discrete wavelet transform (2D-DWT) of frame $f_{\rightarrow n}$ if the corresponding distortion error $D^\beta_{\rightarrow n} = \|C^\beta_{\rightarrow n} - C^\beta_n\|^2$ is smaller than $\tilde{D}^\beta_n$; otherwise $\tilde{C}^\beta_n$ is selected. Thus, the error in C^β_n is given by

$$D^\beta_n = \min\left\{\tilde{D}^\beta_n, D^\beta_{\rightarrow n}\right\} \quad (2)$$

For simplicity, we assume that the client would always make the right decisions regarding the code-block selection; in other words, it has all the necessary information to make these decisions. This is unrealistic as the client usually receives only quantized information that are sub-optimal.

3. SERVER POLICY

Optimization is performed at epochs; where each epoch corresponds to the time of one group of pictures $\mathcal{G}$. Longer or shorter epochs are possible; however, we do not pursue them here. Each epoch $\mathcal{G}$ comprises $2^K + 1$ frames; where the last frame of epoch $\mathcal{G}_s$ is the first frame of epoch $\mathcal{G}_{s+1}$. We choose to jointly allocate $L_{\max}$ bytes for each $\mathcal{G}$. It can be seen that frames in temporal level T_K have two chances of receiving data. Using an additive model, the distortion in one $\mathcal{G}$ is given by

$$D = \sum_{n\in\mathcal{G}_s}\sum_{\beta\in f_n} D^\beta_n \quad (3)$$

The minimization of D subject to length constraint $L_{\max}$ can be (approximately) recast as the minimization of a family of Lagrangian functionals,

$$J_\lambda = \sum_{n\in\mathcal{G}_s}\sum_{\beta\in f_n}\left(D^\beta_n + \lambda\cdot|q^\beta_n|\right) \quad (4)$$

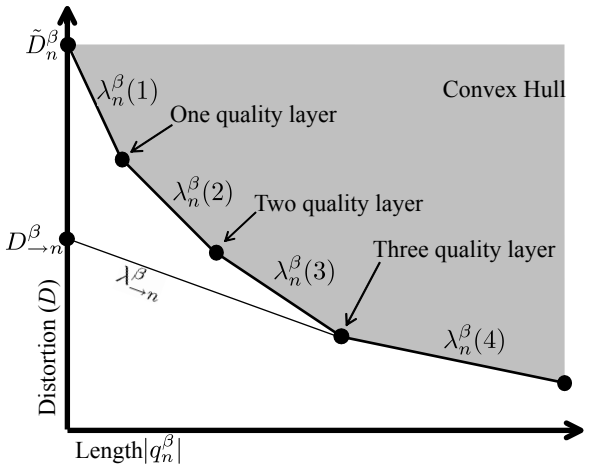

Fig. 2. A typical rate-distortion curve for a code-block C^β_n showing the quality layers and $D^\beta_{\rightarrow n}$ when $D^\beta_{\rightarrow n} < \tilde{D}^\beta_n(0)$.

where $|q^\beta_n|$ denotes the number of bytes in q^β_n layers of C^β_n and the Lagrangian parameter λ is adjusted until the solution which minimizes J_λ satisfies the length constraint.

Direct minimization of equation (4) appears difficult, since each D^β_n, except for frames in T_K, depends in a complicated way on other frames whose contribution is also being optimized. Also, the choice of predicted code-blocks for frames at temporal levels $k < K$ affect the importance of their reference code-blocks. A near optimal solution, however, is possible.

The optimization procedure is split into two passes. In PASS1, the number of quality layers q^β_n is determined for each code-block C^β_n, in a manner which will be explained below. In PASS2, these reconstruction decisions are used to produce importance weights for all the code-blocks that reflect their contribution to the overall reconstructed video sequence quality. Multiple iterations of PASS1 followed by PASS2 are possible, and the last iteration can skip PASS2.

For PASS1, the frames in level T_K are independent of any other frames and q^β_n can be easily determined for any value of λ, as explained next. Once these are determined, the distortions D^β_n for all the frames in level T_{K-1} are known and q^β_n for these frames can be determined as well. This process is repeated until all the quality layers are decided.

Figure 2 shows a typical rate-distortion curve for a code-block C^β_n. Each circle in the figure represents one quality layer. Also shown in the figure is the distortion $D^\beta_{\rightarrow n}$. Note that the distortion-length characteristic, $\tilde{D}^\beta_n$ vs. $|q^\beta_n|$, is typically convex by construction in JPEG2000 code-blocks [8].

When $D^\beta_{\rightarrow n} \geqslant \tilde{D}^\beta_n$ or in the absence of any previous data, such as for frames in level T_K, the optimal choice for q^β_n is given by

$$q^\beta_n = \max\left\{q \mid \lambda^\beta_n(q) > \lambda\right\} \quad (5)$$

where $\lambda^\beta_n(q) = (\tilde{D}^\beta_n(q-1) - \tilde{D}^\beta_n(q))/(|q| - |q-1|)$ is the distortion length slope associated with layer q.

For frames in level T_k, $k < K$, if $D^\beta_{\rightarrow n} < \tilde{D}^\beta_n(0)$ the effective distortion associated with the choice $q^\beta_n = 0$ reduces to $D^\beta_n = D^\beta_{\rightarrow n}$. Figure 2 shows the impact of this option on the effective distortion-length convex hull. The new maximum distortion-length slope is

written as $\lambda^{\beta}_{\rightarrow n}$ and the optimal choice for q^{β}_{n} becomes

$$q^{\beta}_{n} = \begin{cases} 0 & \text{if } \lambda > \lambda^{\beta}_{\rightarrow n} \\ \max\left\{q \mid \lambda^{\beta}_{n}(q) > \lambda\right\} & \text{if } \lambda \leqslant \lambda^{\beta}_{\rightarrow n} \end{cases} \quad (6)$$

We turn our attention now to PASS2. Distortion in frames at any level T_k propagates into frames in levels $T_j,\ j < k$, through motion compensation. Therefore, the significance of distortion in the samples of any code-block depends on how many other samples they affect. This affects the proportion of the overall length budget which should be allocated to each code-block.

In [4], we show that distortion in sub-band samples of a motion-compensated frame can be split into two components: quantization distortion $D^{\mathrm{Q},\beta}_{\rightarrow n}$, due to the distorted reference frames used for motion compensation; and motion distortion $D^{\mathrm{M},\beta}_{\rightarrow n}$, due to motion modelling difficulties. That is,

$$D^{\beta}_{\rightarrow n} \approx D^{\mathrm{Q},\beta}_{\rightarrow n} + D^{\mathrm{M},\beta}_{\rightarrow n} \quad (7)$$

For (4), $D^{\mathrm{M},\beta}_{\rightarrow n}$ is a constant and can be safely ignored for this minimization problem. The distortion in $f^{k}_{\rightarrow n}$ of (1) comes from both $f^{k}_{n-p\rightarrow n}$ and $f^{k}_{n+p\rightarrow n}$; however, here we focus on one side as it can be easily extended to two or more sources. For brevity, we use f_r for the reference frame and we use $\overrightarrow{\mathcal{W}}$ for $\overrightarrow{\mathcal{W}}_{r\rightarrow n}$. The error in frame f_r, can be expressed in term of the errors at each location $\mathbf{k}$ in each of its sub-bands b, as

$$\delta f_r = \sum_b \sum_{\mathbf{k}} \delta B^{b}_{r}[\mathbf{k}] \cdot S^{b}_{\mathbf{k}}$$

where $S^{b}_{\mathbf{k}}$ denotes the relevant synthesis vectors (images). The error at the $\mathbf{p}^{\text{th}}$ location in the predicted sub-band b' of f_n, due to quantization in the $\mathbf{k}^{\text{th}}$ location of sub-band b in f_r is then given by

$$\delta B^{\mathrm{Q},b'}_{r\rightarrow n}[\mathbf{p}] = \frac{1}{2} \cdot \sum_b \sum_{\mathbf{k}} \delta B^{b}_{r}[\mathbf{k}] \cdot \left\langle \overrightarrow{\mathcal{W}}(S^{b}_{\mathbf{k}}), A^{b'}_{\mathbf{p}} \right\rangle$$

where $A^{b'}_{\mathbf{p}}$ is the analysis function for sub-band b' at location $\mathbf{p}$. Assuming that quantization errors in the sub-bands are approximately uncorrelated, the distortion power for some region $\mathcal{R}'$ in sub-band b' can then be approximated by

$$\sum_{\mathbf{p}\in\mathcal{R}'} \left|\delta B^{\mathrm{Q},b'}_{r\rightarrow n}[\mathbf{p}]\right|^2 \approx \frac{1}{4} \cdot \sum_b \underbrace{\sum_{\mathbf{p}\in\mathcal{R}'} \sum_{\mathbf{k}} \left|\delta B^{b}_{r}[\mathbf{k}]\right|^2 \cdot \left\langle \overrightarrow{\mathcal{W}}(S^{b}_{\mathbf{k}}), A^{b'}_{\mathbf{p}} \right\rangle^2}_{D^{b\rightarrow b'}_{\mathcal{R}'}}$$

A result of the decay in the finite support operators $\overrightarrow{\mathcal{W}}(S^{b}_{\mathbf{k}})$ and $A^{b'}_{\mathbf{p}}$ is that $D^{b\rightarrow b'}_{\mathcal{R}'}$ depends mainly on the distortion contributions $\delta B^{b}_{r}[\mathbf{k}]$ inside and around the region $\mathcal{R}$, the projection of $\mathcal{R}'$ in sub-band b. If $\mathcal{R}'$ is small enough such that the distortion around it can be approximated by a single average quantization noise power, we have

$$D^{b\rightarrow b'}_{\mathcal{R}'} \approx \frac{D^{b}_{\mathcal{R}}}{|\mathcal{R}|} \cdot \sum_{\mathbf{p}\in\mathcal{R}'} \underbrace{\sum_{\mathbf{k}} \left\langle \overrightarrow{\mathcal{W}}(S^{b}_{\mathbf{k}}), A^{b'}_{\mathbf{p}} \right\rangle^2}_{W^{b\rightarrow b'}_{\mathbf{p}}}$$

Here, $W^{b\rightarrow b'}_{\mathbf{p}}$ represents a weight for the contribution of the average quantization noise power around $\mathcal{R}$ in sub-band b to the distortion at location $\mathbf{p}$ in sub-band b'. Denoting the average quantization noise power $D^{b}_{\mathcal{R}}/\,|\mathcal{R}|$ around $\mathcal{R}$ by $\bar{D}^{b}_{r}[\mathbf{k}]$ gives

$$\left|\delta B^{\mathrm{Q},b'}_{r\rightarrow n}[\mathbf{p}]\right|^2 \approx \frac{1}{4} \cdot \sum_b \bar{D}^{b}_{r}\left[\overleftarrow{\mathcal{W}}_{b'\rightarrow b}(\mathbf{p})\right] \cdot W^{b\rightarrow b'}_{\mathbf{p}}$$

where $\overleftarrow{\mathcal{W}}_{b'\rightarrow b}$ is the reverse motion mapping, which takes a location from sub-band b' in f_n to a location in sub-band b of f_r. Equivalently, the predicted code-block distortions propagating from D^{β}_{r} are given by

$$D^{\mathrm{Q},\beta'}_{r\rightarrow n} \approx \frac{1}{4} \cdot \sum_{\mathbf{p}\in C^{\beta'}_{n}} \sum_{\beta \ni \overleftarrow{\mathcal{W}}_{b'\rightarrow b}(\mathbf{p}) \in C^{\beta}_{r}} D^{\beta}_{r} \frac{W^{b\rightarrow b'}_{\mathbf{p}}}{\left|C^{\beta}_{r}\right|}$$

So the impact of D^{β}_{r} on $D^{\beta'}_{r\rightarrow n}$ can be expressed as

$$D^{\beta}_{r} \cdot \underbrace{\frac{1}{4} \cdot \sum_{\beta' \ni \lambda > \lambda^{\beta'}_{\rightarrow n}} \sum_{\substack{\mathbf{p}\in C^{\beta'}_{n}, \\ \overleftarrow{\mathcal{W}}_{b'\rightarrow b}(\mathbf{p}) \in C^{\beta}_{r}}} \frac{W^{b\rightarrow b'}_{\mathbf{p}}}{\left|C^{\beta}_{r}\right|}}_{w^{\beta}_{r}}$$

Although the last equation looks complicated, the interpretation is simple. w^{β}_{r} is a state variable associated with each code-block in f_r. For a given value of λ, we identify all code-blocks β' in the predicted frame such that in (2) we have $D^{\beta}_{n} = D^{\beta}_{\rightarrow n}$, meaning that they will be assigned no bits. For each location in each such code-block, we find the corresponding location in each sub-band of f_r and add $W^{b'\rightarrow b}_{\mathbf{p}}/\left|C^{\beta}_{r}\right|$ to the corresponding state variable, w^{β}_{r}. In practice, we actually perform this state mapping process on a coarse grid of 4×4 samples, using pre-computed values for $W^{b'\rightarrow b}_{\mathbf{p}}$ which are stored at the server on this grid.

These weights modify the rate-distortion optimization calculations. Specifically, equation (6) becomes

$$q^{\beta}_{n} = \begin{cases} 0 & \text{if } \lambda > \lambda^{*,\beta}_{\rightarrow n} \\ \max\left\{q \mid (1 + w^{\beta}_{n}) \cdot \lambda^{\beta}_{n}(q) > \lambda\right\} & \text{if } \lambda \leqslant \lambda^{*,\beta}_{\rightarrow n} \end{cases} \quad (8)$$

where $\lambda^{*,\beta}_{\rightarrow n} = (1 + w^{\beta}_{n}) \cdot \lambda^{\beta}_{\rightarrow n}$.

4. EXPERIMENTAL RESULTS

The results presented here are for the MPEG test sequences, "crew" and "harbour". Only the Y component of the first 49 frames of each sequence is used. The frames have a resolution of 704×576 with a bit depth of 8 bits. Those frames are compressed into the JPEG2000 format using Kakadu[1], employing five levels of wavelet decomposition, 20 quality layers, and a code-block size of 32×32. Motion compensation employs a block-based motion model that uses a grid spacing of 4 pixels and is applied only to the highest resolution. Motion is estimated to $1/8$ of a pixel accuracy by first employing $1/2$ pixel spline interpolation; followed by bilinear interpolation to $1/8$ of a pixel. The frame rate is 30 frames/second and the rates given here are only for the encoded sub-band samples; they exclude any headers, motion vectors, and signaling to the client, etc.

The proposed method is compared against two other methods. The first method, identified as "INTRA", individually optimizes each frame separately, such that the rate constraint is met. The second method, identified as "PRED", jointly optimizes the whole 49-frame sequence for a given rate. In this case, frames at T_K are kept as intra-frames; while all the $k < K$ temporal level frames are stored as residues obtained from $f_n - f_{\rightarrow n}$ of (1). The optimization problem for the "PRED" method is cast as the minimization of

$$J_{\lambda} = \sum_n \sum_{\beta\in f_n} \left((1 + w_n) \cdot D^{\beta}_{n} + \lambda \cdot |q^{\beta}_{n}|\right) \quad (9)$$

[1] http://www.kakadusoftware.com/, Kakadu software, version 5.2.4.

subject to the length constraints. w_n is the weight that reflects the contribution of frame f_n to the overall reconstructed video quality. Finding these weights is a tedious task, so we give only their values here for $K = 3$. Frames at levels $T_0 \ldots T_3$ have weights of 0.0, 0.5, 1.75, and 2.375, respectively. It is worth noting that the first and the final frames in the test sequence have a lower weight of 1.1875, since they contribute less compared to other intra-frames in the middle of the sequence. The final method, identified as "OPTIMAL", is the method presented in this paper. Three iterations of PASS1-PASS2 are used since we found that iterations beyond the second iteration produce negligible improvements. For the "PRED" and "OPTIMAL" methods, K is set equal to 3.

Figures 3 and 4 show the average MSE expressed in terms of PSNR at various bit rates for the "crew" and "harbour" sequences, respectively. It can be seen that at low bit rates "PRED" outperforms "OPTIMAL" by between 0.2 and 0.5 dB, while at mid to high bit rates "PRED" is better by perhaps 2 dB. At low to mid bit rates both "OPTIMAL" and "PRED" perform better than "INTRA"; however, the difference between these methods reduces at high rates.

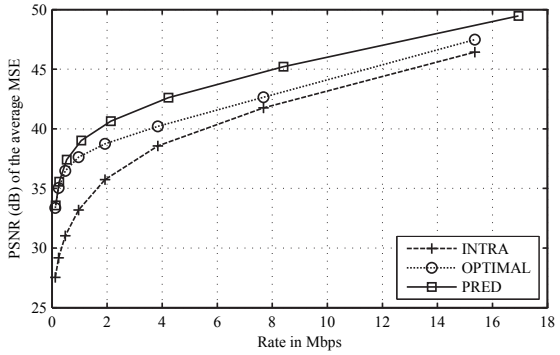

Fig. 3. PSNR of the average MSE for the first 49 frames of the "crew" sequence at various bit rates.

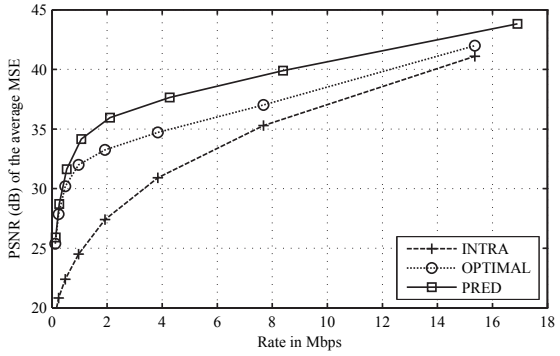

Fig. 4. PSNR of the average MSE for the first 49 frames of the "harbour" sequence at various bit rates.

The second set of results considers the benefits of the proposed approach for a client interested in seeing only one frame. Figure 5 shows the PSNR of frame 10, a frame from temporal level T_0 of the "harbour" sequence, at various bit-rates for the OPTIMAL/INTRA and PRED methods. Note that our proposed approach reduces to INTRA for the case in which only one frame is requested by an interactive client; moreover, half of all the frames in the sequence are in level T_0 only. For the PRED method, the effective rate shown in the figure is the sum of the rates for frames 9, 10, 11, 13, and 17 as all these frames contribute to frame 10, as shown in Figure 1. It can be seen that the savings are enormous.

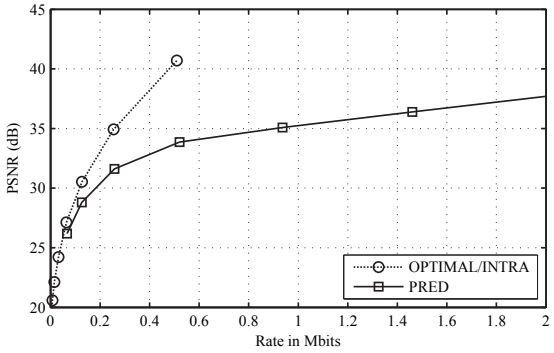

Fig. 5. PSNR of frame 10 of the "harbour" sequence at various bit-rates for the OPTIMAL/INTRA and PRED methods.

5. CONCLUSION

The approach presented in this paper shows very promising results and addresses problems in video accessibility and scalability. The price paid for this increased flexibility of the proposed delivery method depends on the bit rate; it varies between 0.2 dB at low bit-rates to around 2 dB at high rates. The hierarchical arrangement of frames provides good exploitation of temporal redundancies for both the proposed method and predictive compression schemes. However, the proposed paradigm considerably outperforms the delivery of predictively compressed video in the case where a client wishes to browse individual frames of the video during playback. More research is needed to enable a fully realistic implementation of the method, particularly for the client policy.

6. REFERENCES

[1] N. Mehrseresht and D. Taubman, "An efficient content-adaptive motion compensated 3D-DWT with enhanced spatial and temporal scalability," *IEEE Trans. Image Proc.*, pp. 1397–1412, June 2006.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[3] J.-R. Ohm, "Advances in scalable video coding," *Proc. of the IEEE*, vol. 93, January 2005.

[4] A.T. Naman and D. Taubman, "A novel paradigm for optimized scalable video transmission based on JPEG2000 with motion," *Proc. IEEE Int. Conf. Image Proc. 2007*, Septmeber 2007.

[5] A.T. Naman and D. Taubman, "Optimized scalable video transmission based on conditional replenishment of JPEG2000 code-blocks with motion compensation," *MV '07: Proceedings of the International Workshop on Workshop on Mobile Video*, pp. 43–48, Septmeber 2007.

[6] F.-O. Devaux, J. Meessen, C. Parisot, J.-F. Delaigle, B. Macq, and C. De Vleeschouwer, "A flexible video transmission system based on JPEG2000 conditional replenishment with multiple references," *Proc. IEEE Int. Conf. Acoust. Speech and Sig. Proc.*, April 2007.

[7] N.-M. Cheung and A. Ortega, "Flexible video decoding: A distributed source coding approach," *IEEE 9th Workshop on Multimedia Signal Processing, 2007, MMSP 2007.*, pp. 103–106, 1-3 Oct. 2007.

[8] D.S. Taubman and M.W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Boston, 2002.