

ADAPTIVE FRAME AND QP SELECTION FOR TEMPORALLY SUPER-RESOLVED FULL-EXPOSURE-TIME VIDEO

Mihoko Shimano ^o, Gene Cheung [#], Imari Sato [#]

^o University of Tokyo, PRESTO, JST, [#] National Institute of Informatics

ABSTRACT

In order to allow sufficient amount of light into the image sensor, videos captured in poor lighting conditions typically have low frame rate and frame exposure time equals to inter-frame period—commonly called full exposure time (FET). FET low-frame-rate videos are common in situations where lighting cannot be improved a priori due to practical (e.g., large physical distance between camera and captured objects) or economical (e.g., long duration of night-time surveillance) reasons. Previous computer vision work has shown that content at a desired higher frame rate can be recovered (to some degree of precision) from the captured FET video using self-similarity-based temporal super-resolution.

For a network streaming scenario, where a client receives a FET video stream from a server and plays back in real-time, the following practical question remains, however: what is the most suitable representation of the captured FET video at encoder, given that a video at higher frame rate must be constructed at the decoder at low complexity? In this paper, we present an adaptive frame and quantization parameter (QP) selection strategy, where, for a given targeted rate-distortion (RD) tradeoff, FET video frames at appropriate temporal resolutions and QP are selected for encoding using standard H.264 tools at encoder. At the decoder, temporal super-resolution is performed at low complexity on the decoded frames to synthesize the desired high frame rate video for display in real-time. We formulate the selection of individual FET frames at different temporal resolutions and QP as a shortest path problem to minimize Lagrangian cost of the encoded sequence. Then, we propose a computation-efficient algorithm based on monotonicity in predictor's temporal resolution and QP to find the shortest path. Experiments show that our strategy outperforms alternative naïve non-adaptive approaches by up to 1.3dB at the same bitrate.

Index Terms— Video compression, temporal super-resolution

1. INTRODUCTION

Suitable exposure time (and subsequently frame rate) for a captured video is heavily influenced by the lighting condition of the scene of interest. If the lighting condition is poor, then exposure time must be long to permit sufficient amount of light into the photographic film or image sensor in order to avoid undesirable underexposure effects. Obviously, exposure time cannot be longer than inter-frame period in a video. Hence, a longer required exposure time can lead to lower video frame rate, which is usually not desirable.

Quite often, lighting condition cannot be improved a priori due to practical or economical constraints. For example, the scene of interest can be too far from the camera to physically insert light sources before capture. Another example is a night-time surveillance / observation situation, where prolonged illumination in a large area will lead to unacceptably high cost, or disturbance to the captured animals in their nocturnal habitat. Thus, it is often unavoidable to han-

dle captured videos with full exposure time (FET) and lower frame rate than desirable, where the exposure time of each frame equals the inter-frame period of the captured video. See Fig. 1 for an illustration.

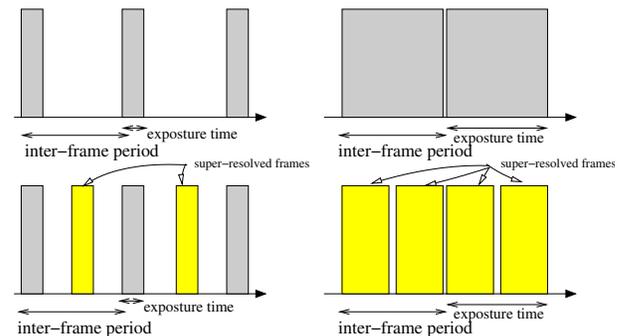


Fig. 1. Illustration of temporal super-resolution for: a) sub-exposure-time video; and, b) full-exposure-time (FET) video. Super-resolved frames are colored yellow.

Given the low frame rate in a typical FET video is not desirable, to synthesize video at a required higher frame rate, previous computer vision work [1] has proposed *temporal super-resolution* using *self-similarity* (TSR-SS) to exploit the FET property. The key observation is that the same motion blur patterns due to long exposure time often reappear at multiple timescales. By identifying motion blur patches in one FET video frame (using Maximum A Posteriori (MAP) estimate [1]), two corresponding frames at twice the frame rate can be constructed by halving the blurs at appropriate spatial locations. [1] has shown that, for FET videos, TSR-SS outperforms previous temporal super-resolution methods like optical flow [2] by up to 6dB in PSNR of the reconstructed high-frame-rate video.

For a typical network streaming scenario, where a client receives a FET video streaming from a server and plays back in real-time, however, the following practical question remains: what is the most suitable representation of the captured FET video at encoder, given that a video at higher frame rate must be constructed at decoder at low complexity? One obvious solution is to perform temporal super-resolution at encoder at high complexity to up-sample captured FET video to high-frame-rate content prior to compression. However, this means compression must be performed in higher-than-capture frame rate, expanding large number of coding bits. An alternative is to encode low-frame-rate captured FET video at encoder, relying on decoder to perform temporal super-resolution on compressed frames to recover required high-frame-rate content. However, both the decoder stringent complexity constraint (limiting the size of the neighborhood of frames used to construct a super-resolved frame) and the resulting quantization noise due to compression can severely affect the quality of the super-resolved video frames.

In this paper, we present instead an adaptive compression strat-

egy for FET video, where FET video frames are selectively encoded at *appropriate temporal resolutions and quantization parameters* (QP) to optimize rate-distortion (RD) performance. The key idea is to encode at low temporal resolution (and possibly at coarse QP) frames that can be easily synthesized via super-resolution at decoder to save bits, and to encode at high temporal resolution (and possibly at fine QP) frames that are difficult to synthesize after compression to preserve quality. We formulate the selection of FET frames at different temporal resolutions and QP as a shortest path problem in a directed acyclic graph (DAG). We then develop a fast shortest-path search algorithm based on assumption on monotonicity in predictor's temporal resolution and QP. Experiments show that our adaptive strategy outperforms alternative naïve non-adaptive approaches by up to 1.3dB at the same bitrate.

The outline of the paper is as follows. After a brief discussion on related work in Section 2, we overview TSR-SS in Section 3. We formulate our RD optimization problem in Section 4, and detail our shortest-path algorithm in Section 5. We present results and conclusion in Section 6 and 7, respectively.

2. RELATED WORK

Frame interpolation in time was studied in the context of loss concealment for network video streaming [3]. While similar in motivation to our temporal super-resolution problem, such methods typically rely on motion compensation or optical flow analysis [2], which do not perform well for FET video with motion blurs at low frame rates. By comparison, our proposed TSR-SS [1] has shown superior super-resolution performance for FET video.

Our problem of selecting FET frames at the “best” temporal resolutions and QPs for video encoding belongs to a family of RD-optimizing dependent bit allocation problems. In [4], bit allocation for dependent frames was studied for motion-compensated video coding. [5] later studied the problem of selecting QPs and skipped frames¹ in video to minimize resulting distortion subject to a rate constraint. Recently, [6] studied the problem of selecting available texture and depth maps of multiview images for differential encoding at appropriate QPs to minimize synthesized view distortions at decoder subject to a rate constraint. Our current work differs in two respects. First, our unique problem setting on FET video coding demands a selection of frames at different temporal resolutions to optimize RD performance, which has not been previously studied. Second, our fast solution search is developed partly based on monotonicity in predictor's temporal resolution, which is also new.

In our earlier work [7], we studied the selection of FET frames at fixed QP to optimize RD performance. Our current work is a non-trivial extension of [7], where: i) QP for each selected FET frame is also optimally selected in a 3D trellis; and ii) decoder complexity is also considered so that it can be traded off with RD performance.

3. SELF SIMILARITY FOR TEMPORAL SUPER-RESOLUTION

In natural images, self-similar textures tend to recur spatially many times inside an image, both within the same scale as well as across different scales. In the same way, self-similarity exists in the spatio-temporal domain of videos, for example, if the objects follow the similar trajectories with constant but different velocities. On the basis of this observation, previously we proposed a method for constructing temporally super-resolved video from a single FET video

¹Skipped frames can be construed as one form of temporal resolution. However, unlike a skipped frame, in our FET context a lower resolution FET frame is a pixel-by-pixel average of two higher resolution FET frames.

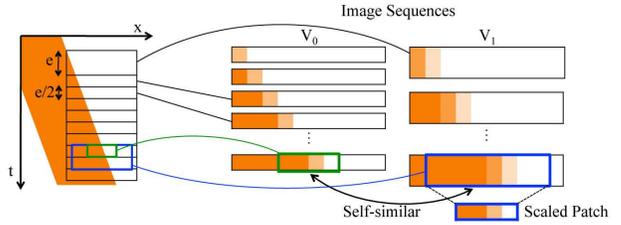


Fig. 2. Self-similarity. A x - t plane for single scan line, and 1-D image sequences with different temporal resolutions.

by exploiting self-similarity, *i.e.*, a self-similar appearance that represents integrated motion of objects during each exposure time of videos with different temporal resolutions [1].

For simplicity, let us consider self-similarity in the case of 1D image sequences of a uniformly moving object. Fig. 2 illustrates two FET image sequences: V_0 has exposure time (also inter-frame period) $e/2$, half of that of V_1 , e . Consider, for example, a small 1D image patch of V_0 with exposure time $e/2$. In a patch of V_1 captured with exposure time e at the same position in x - t plane, the same object moves twice the distance. If the spatial size of the patch of V_1 is twice that of the patch of V_0 , the patch of V_0 becomes similar to the corresponding patch of V_1 because the object moves twice the distance during the exposure time of V_1 . This self-similar relationship can be extended to a 2D-image patch. Utilizing this self-similar relationship between such different temporal resolution frames which are created from the original captured frames, TSR-SS can create a high-frame-rate video from the scaled self-similar patches.

4. PROBLEM FORMULATION

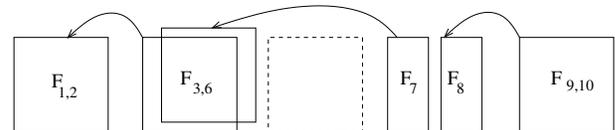


Fig. 3. Example of FET frame selection. $F_{(1,2)}$ and $F_{(9,10)}$ are MR (captured) frames. LR frame $F_{(3,6)}$ is average of MR frames $F_{(3,4)}$ and $F_{(5,6)}$. SR F_7 and F_8 are synthesized from MR frame $F_{(7,8)}$.

We formulate our problem of selecting FET frames at appropriate temporal resolutions and QPs as a combinatorial optimization. We first overview the degrees of freedom in frame selection. Denote by $\mathcal{F}^M = \{F_{(1,2)}, F_{(3,4)}, \dots, F_{(2N-1,2N)}\}$ the N captured FET (medium temporal resolution or MR) frames. Denote by $\mathcal{F}^S = \{F_1, F_2, \dots, F_{2N}\}$ the $2N$ *super-resolved* (SR) frames. In addition, let averages of neighboring captured frames (LR frames) be $\mathcal{F}^L = \{F_{(1,4)}, F_{(3,6)}, \dots, F_{(2N-3,2N)}\}$, where $F_{(i-3,i)}$ is the pixel-by-pixel average of MR frames $F_{(i-3,i-2)}$ and $F_{(i-1,i)}$. Each MR frame $F_{(i-1,i)} \in \mathcal{F}^M$ can be super-resolved via TSR-SS into two corresponding SR frames, F_{i-1} and F_i . It can also be combined with MR frame $F_{(i-3,i-2)}$ into a one corresponding LR frame $F_{(i-3,i)}$. (In turn, LR frame $F_{(i-3,i)}$ can be super-resolved into MR frames $F_{(i-3,i-2)}$ and $F_{(i-1,i)}$.) The goal is to construct best quality SR frames F_i 's at decoder using subset of \mathcal{F}^S , \mathcal{F}^M and \mathcal{F}^L encoded using a fixed bit budget. Coding SR frames at encoder gives high quality but requires many bits, while coding LR frames at encoder gives poor constructed SR frame quality at decoder but requires few bits. Fig. 3 shows one example of a frame combination.

4.1. DAG Representation of Frame Selection

We construct a three-dimensional (3D) directed acyclic graph (DAG) to represent selections of frames and corresponding QPs as follows.

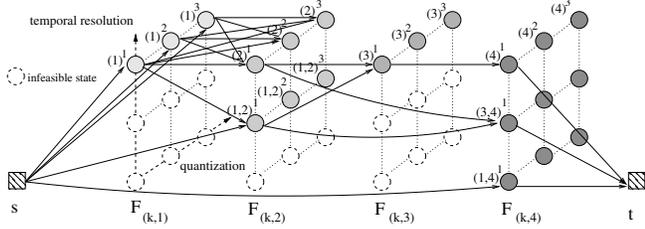


Fig. 4. DAG example for 4 planes, and start and end node s, t .

For each triple of SR frame F_i , MR frame $F_{(i-1,i)}$ and LR frame $F_{(i-3,i)}$, we create a *plane* i of states, as shown in Fig. 4. These frames correspond to content of F_i at different temporal resolutions. Three state rows of each plane i correspond to F_i , $F_{(i-1,i)}$ and $F_{(i-3,i)}$, respectively. Each state x represents the selection of frame $F_{f(x)}$ coded at QP $q(x)$. Some rows contain infeasible states; e.g., states in second and third row of plane 1 are not feasible since frame $F_{0,1}$ and $F_{-2,1}$ do not exist. In addition, we create start and end nodes s and t at the left and right end of the DAG.

We draw an edge $e_{x \rightarrow y}$ from state x to y to represent the case when frame $F_{f(y)}$ is differentially coded at QP $q(y)$ using $F_{f(x)}$, coded at QP $q(x)$, as predictor. Essentially, any feasible state in plane i can be predictor for states x 's, where $f(x)$ is either $i+1$ or $(i+1, k)$, $\forall k$. Start node s has edges to states of candidate first frames of the sequence (I-frame)—frames F_1 , $F_{(1,2)}$ and $F_{(1,4)}$. End node t has edges stemming from states in the last plane, corresponding to candidate last frames F_{2N} , $F_{(2N-1,2N)}$ and $F_{(2N-3,2N)}$.

Because a motion-compensated frame has one predictor frame, which in turn has one predictor frame, all the way back to the first frame that is encoded as an I-frame, a *path* \mathbf{p} from s to t through the constructed DAG represents a selection of frames and corresponding QPs for encoding of the sequence. If we can assign cost to the edges representing rate and distortion costs, then we can formulate the frame selection problem as a minimum cost path problem. Towards that goal, we first discuss rate and distortion model.

4.2. Rate Model

For a given path $\mathbf{p} = \{s, p_1, p_2, \dots, p_{L_p}, t\}$, where L_p is the length of path \mathbf{p} minus s and t , we can write the encoding rate $R(\mathbf{p})$ of the coded sequence described by path \mathbf{p} as sum of rates of individual traversed states p_l 's, each denoting selected frame and corresponding QP:

$$R(\mathbf{p}) = \sum_{l=1}^{L_p} r^c(p_l, p_{l-1}) \quad (1)$$

where $p_0 = s$, and the code rate $r^c(p_1, p_0)$ of the first frame depends only on the selection of the first frame $F_{f(p_1)}$ and its QP $q(p_1)$. The code rate $r^c(p_l, p_{l-1})$ of the l -th frame depends only on the selection of l -th frame, $F_{f(p_l)}$, l -th frame's QP, $q(p_l)$, its predictor $F_{f(p_{l-1})}$, and its predictor's QP, $q(p_{l-1})$. This Markovian rate dependency model has been validated to be sufficiently accurate and useful [6].

4.3. Distortion Model

We can similarly write the resulting distortion $D(\mathbf{p})$ of the super-resolved sequence at decoder, given encoded frame selection path \mathbf{p} , as a simple sum:

$$D(\mathbf{p}) = \sum_{l=1}^{L_p} d(p_l, \mathbf{p}) \quad (2)$$

where $d(p_l, \mathbf{p})$ is the distortion corresponding to coded frame $F_{f(p_l)}$. If coded frame $F_{f(p_l)}$ is a SR frame (constructed at encoder),

then distortion $d(p_l, \mathbf{p})$ is simply the coded distortion $d^c(p_l, p_{l-1})$ that depends on the selection of its predictor state p_{l-1} only. If $F_{f(p_l)}$ is a MR frame or a LR frame, then we must account for distortions of all SR frames *to be constructed at decoder* of which $F_{f(p_l)}$ contains content. Let $(a(p_l), b(p_l))$ be the range of SR frames of which frame $F_{f(p_l)}$ averaged over, given $F_{f(p_l)}$ is a MR frame $F_{(i-1,i)}$ or an LR frame $F_{(i-3,i)}$. We can write distortion $d(p_l, \mathbf{p})$ of coded frame $F_{f(p_l)}$ as:

$$d(p_l, \mathbf{p}) = \begin{cases} d^c(p_l, p_{l-1}) & \text{if } F_{f(p_l)} \in \mathcal{F}^S \\ \sum_{k=a(p_l)}^{b(p_l)} d^s(k, \mathbf{p}_{\lfloor k/W \rfloor W+1}^{\lfloor k/W \rfloor W}) & \text{o.w.} \end{cases} \quad (3)$$

where $d^s(k, \mathbf{p}_{\lfloor k/W \rfloor W+1}^{\lfloor k/W \rfloor W})$ is the distortion of the SR frame F_k at decoder given selected frame path \mathbf{p} . TSR-SS at decoder uses a non-overlapping *construction window* of W SR frames, $\lfloor k/W \rfloor W + 1, \dots, \lfloor k/W \rfloor W$ that contains SR frame k to detect self-similarity. Larger W means larger SR construction complexity at decoder.

4.4. Shortest Path in DAG

Having defined rate cost $R(\mathbf{p})$ and distortion cost $D(\mathbf{p})$ for a given frame selection path \mathbf{p} , we can formalize our path selection problem as a Lagrangian optimization problem:

$$\min_{\mathbf{p} \in \mathcal{P}} \Theta(\mathbf{p}) = D(\mathbf{p}) + \lambda R(\mathbf{p}) \quad (4)$$

where \mathcal{P} is the set of feasible paths from s to t , and λ is the Lagrangian multiplier. We can alternatively write (4) as a sum of individual Lagrangian costs $\theta(p_l, \mathbf{p})$ of coded frame F_{p_l} :

$$\begin{aligned} \Theta(\mathbf{p}) &= \sum_{l=1}^{L_p} \theta(p_l, \mathbf{p}) \\ \theta(p_l, \mathbf{p}) &= d(p_l, \mathbf{p}) + \lambda r^c(p_l, p_{l-1}) \end{aligned} \quad (5)$$

By assigning each individual Lagrangian cost $\theta(p_l, \mathbf{p})$ to edge e_{p_{l-1}, p_l} of path \mathbf{p} , we see now that a shortest cost path in the DAG, where the cost of the path \mathbf{p} is the sum of individual edge costs $c(e_{p_{l-1}, p_l})$'s, corresponds to the optimal frame selection in (5). We discuss how we find this shortest path in the next section.

5. FAST SHORTEST PATH ALGORITHM

There are two complications when trying to solve the shortest path problem in (5). The first is that Lagrangian cost $\theta(p_l, \mathbf{p})$ of edge e_{p_{l-1}, p_l} depends on the *entire path* \mathbf{p} rather than just the two end nodes of the edge, p_{l-1} and p_l . The second is that even if edge cost $c(e_{p_{l-1}, p_l})$ can be determined solely as function of p_{l-1} and p_l , finding the shortest path for a large number of nodes and edges can be computationally expensive. We address both issues in order.

5.1. Iterative Shortest Path Procedure

To address the first concern, we propose an iterative procedure where, in each iteration, a shortest path (SP) in the DAG is found given *fixed* edge costs; i.e. each cost $c(e_{x \rightarrow y})$ of edge $e_{x \rightarrow y}$ is fixed given nodes x and y , independent of other nodes. Edge costs are adjusted after each iteration. The procedure terminates when costs of paths of two consecutive iterations differ by less than δ . Details of the procedure is shown in *Iterative SP Procedure*.

We explain the procedure's rationale as follows. Step 2 initialize each edge $e_{x \rightarrow y}$ to be the smallest $\theta(y, \mathbf{p}^i)$ possible, since all other nodes besides x and y are SR nodes, resulting in the smallest synthesized distortions. This provides the procedure an opportunity to find a path away from initial path of all SR nodes. Subsequently, for each discovered SP \mathbf{p} , each edge cost of SP will only increase

Iterative SP Procedure

- 1: Initialize previous path \mathbf{p}' to be path with all SR nodes.
 - 2: Initialize $c(e_{x \rightarrow y})$ to be $\theta(y, \mathbf{p}^i)$, where path \mathbf{p}^i contains x and y , and all other nodes are SR nodes.
 - 3: Find SP \mathbf{p} given fixed edge costs.
 - 4: $c(e_{p_{l-1} \rightarrow p_l}) \leftarrow \max\{\theta(p_l, \mathbf{p}), c(e_{p_{l-1} \rightarrow p_l})\}$.
 - 5: **if** $|\Theta(\mathbf{p}') - \Theta(\mathbf{p})| > \delta$ **then**
 - 6: $\mathbf{p}' \leftarrow \mathbf{p}$. Goto step 3.
 - 7: **end if**
-

according to selected nodes in \mathbf{p} . This means procedure is guaranteed to converge. Moreover, the calculated cost of the converged SP is an upper-bound of the Lagrangian cost of the true SP due to max operation in step 4.

5.2. Monotonicity in Predictor's Temporal Resolution and QP

For fixed edge costs, we can speed up the SP search with assumptions of *monotonicity in predictor's temporal resolution and QP*. The first observation is that finer temporal resolution of a frame, say $F_{(i-i,i)}$, is in general a better predictor for *any* future frame than coarser temporal resolution of the same frame, say $F_{(i-3,i)}$. So if the local Lagrangian cost of choosing the finer temporal resolution version is already less than the coarser resolution, then the coarser resolution version is globally sub-optimal.

In more rigorous details, let $\psi(x)$ be the shortest sub-path from s to state x . We state the above observation formally as follows:

Lemma 1 *If $\psi(x) < \psi(y)$, where $F_{f(x)}$ and $F_{f(y)}$ are fine and coarse temporal resolution of the same frame at same QP, i.e., $q(x) = q(y)$, then state y is globally sub-optimal.*

See [7] for a proof. Similarly, we can construct the following lemma concerning monotonicity in predictor's QP:

Lemma 2 *If $\psi(x) < \psi(y)$, where both state x and y are of the same frame, i.e., $f(x) = f(y)$, and state x has finer QP than y , i.e., $q(x) < q(y)$, then state y is globally sub-optimal.*

One can easily construct a proof similar to one in [7] for this lemma. The two lemmas can be used in combination to speed up the SP search as follows. Shortest sub-paths to states are updated by calculating edges stemming from states in plane i for increasing i . At each plane i , check first if two states x and y in plane i satisfies the two conditions: i) $\psi(x) < \psi(y)$, and ii) $F_{f(x)}$ is a finer or same temporal resolution of $F_{f(y)}$ and QP $q(x) \leq q(y)$. If so, state y is sub-optimal and it can be pruned from DAG right away.

6. EXPERIMENTATION

To test the performance of our proposed compression scheme (adapt), we first shot two 300-frame 384×216 FET video sequence *intrsct3* and *bear* as data for MR frame set \mathcal{F}^M . We then created SR frame set \mathcal{F}^S at encoder using window of 150 frames for each SR frame via TSR-SS, and \mathcal{F}^L by averaging neighboring MR frames. Assuming a Group of Pictures (GOP) (N) size of 30 captured frames, we then encoded combinations of \mathcal{F}^M , \mathcal{F}^S and \mathcal{F}^L at different combinations of QPs using H.264 to calculate encoding rates r^c 's for all frames, and distortions d^c 's for SR frames \mathcal{F}^S . For distortions of MR and LR frames, \mathcal{F}^M and \mathcal{F}^L , we assume a window of 30 frames is used to super-resolved each SR frame at the decoder. We varied multiplier value λ when finding the best frame combinations of frames and QPs for encoding via our iterative SP procedure to obtain different RD performance points.

In Fig. 5(a), we compare our scheme to two simple schemes for *intrsct3*: i) HR up-sampled the entire FET video via TSR-SS

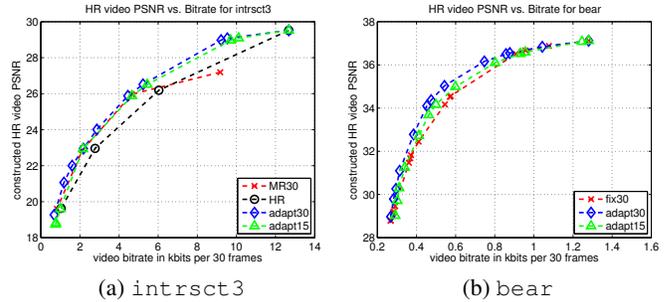


Fig. 5. PSNR Comparison between different encoding schemes. x -axis is kbits per 30 frames, and y -axis is video quality in PSNR.

before encoding, and ii) MR encoded captured FET frames as is, then performed TSR-SS at decoder to super-resolve SR frames. QP was adjusted for HR and MR to obtain different RD tradeoffs. We observe that for decoder complexity $W = 30$, adapt outperformed HR and MR at all bitrates: by up to 1.3dB and 1.8 in PSNR. The gain is particularly pronounced at middle bitrate, where neither MR frames only in MR nor SR frames only in HR were able to perform well. This shows that adaptive selection of FET frames at appropriate temporal resolutions and QPs is important.

When the window used for TSR-SS to super-resolve SR frames at decoder is reduced to $W = 15$, we see a slight drop in performance for adapt. Nonetheless, adapt15 still outperformed HR and MR30, showing that our scheme is able to adapt well to more stringent decoder complexity constraint.

Fig. 5(b) uses *bear* to compare our proposed adapt to fix in [7], where temporal resolutions were adaptively selected but QP was fixed. We see that adapt outperformed fix by up to 1.1dB.

7. CONCLUSION

In this paper, we propose an adaptive encoding strategy to select full-exposure-time (FET) frames at different temporal resolutions and QPs to optimize RD performance. Results show our adaptive scheme outperformed naïve schemes by up to 1.3dB in PSNR.

8. REFERENCES

- [1] M. Shimano, T. Okabe, I. Sato, and Y. Sato, "Video temporal super-resolution based on self-similarity," in *The Tenth Asian Conference on Computer Vision*, Queenstown, New Zealand, November 2010.
- [2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. ECCV 2004 (LNCS3024)*, 2004, pp. 25–36.
- [3] J. Zhai, K. Yu, J. Li, and S. Li, "A low complexity motion compensated frame interpolation method," in *IEEE International Symposium on Circuits and Systems*, Kobe, Japan, May 2005.
- [4] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," in *IEEE Transactions on Image Processing*, September 1994, vol. 3, no. 5.
- [5] S. Liu and C.C. J. Kuo, "Joint temporal-spatial bit allocation for video coding with dependency," in *IEEE Transactions on Circuits and Systems for Video Technology*, January 2005, vol. 15, no. 1, pp. 15–26.
- [6] G. Cheung and V. Velisavljević, "Efficient bit allocation for multiview image coding & view synthesis," in *IEEE International Conference on Image Processing*, Hong Kong, September 2010.
- [7] M. Shimano, G. Cheung, and I. Sato, "Compression using self-similarity-based temporal super-resolution for full-exposure-time video," in (accepted to) *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011.