ADAPTIVE RATE CONTROL FOR WYNER-ZIV VIDEO CODING

Ghazaleh Esmaili, Pamela Cosman

University of California, San Diego
La Jolla, CA, 92093-047
gesmaili@ucsd.edu, pcosman@ucsd.edu

ABSTRACT

In Wyner-Ziv video coding architectures, the available bit budget to each GOP is shared between key frames and Wyner-Ziv frames. In this work, we first propose a model to express the relationship between quantization step size of key and WZ frames based on their motion activity. Then we apply this model to propose an adaptive algorithm adjusting the quantization step size of key and WZ frames to achieve and maintain a target bit rate. We evaluate the rate distortion performance of the proposed method and compare to a common method in the literature.

Index Terms— Wyner-Ziv video coding, distributed source coding, rate control

1. INTRODUCTION

Wyner-Ziv video coding which is founded on the Slepian-Wolf [1] and Wyner-Ziv [2] theorems is a promising solution for applications which require simple and low cost encoding. In this approach, the complexity is largely shifted from the encoder to the decoder by encoding individual frames independently (intraframe encoding) but decoding them conditionally (interframe decoding) [3]. The algorithm proposed by Aaron et al. in [4] which requires feedback and is based on Turbo coding became the basis for considerable further research. In most WZ codecs, no bit rate control is considered, and one of the main challenges is efficiently allocating the available bit budget between key frames and WZ frames. Here, predicting the number of bits required to reliably decode WZ frames is difficult since the side information is not available at the encoder. With more accurate side information, fewer bits are required to encode the WZ frame. The accuracy of the side information is affected by the quality of key frames. In [5], the impact of the key frame quality on the WZ frame coding and on the overall video codec RD performance was investigated. Their experimental results showed that increasing the quality of key frames results in improving the overall performance until a peak point after which the overall quality degrades.

In most existing WZ video codecs, without considering any bit rate constraint, the quantization parameters (QPs) of key frames and WZ frames are selected offline by exhaustive search to provide maximum coding efficiency with similar quality for key and WZ frames. The offline exhaustive search approach is not viable for online applications. Also, in a video sequence there are usually different scenes with different content and motion characteristics which are treated uniformly by this method. In [6], a quality control algorithm without any bit constraint was proposed in which the QP of key and WZ frames is dynamically adjusted to provide constant quality for both key and WZ frames. In this approach, the quantization level of each frequency band of the WZ frame needs to be obtained through an iterative loop which increases the complexity of the encoder. An RD performance loss of about 0.4 to 1.0 dB compared to the WZ coding solution without quality control was reported. In [7], a rate control algorithm for pixel-domain Wyner-Ziv video coding was proposed which estimated the rate and distortion of each video frame as a function of the coding mode and the QP. In [8], based on the motion activity between adjacent key frames, a table was suggested to select QPs of key and WZ frames for six different quality levels. No solution was suggested for an arbitrary target quality. In [9], to obtain similar quality for Intra and WZ frames, the relevant parameters are controlled jointly: QP for the key frames and the quantization for the WZ frames. In this work we first propose a method to efficiently distribute the bit budget between key and WZ frames by modeling the relationship between quantization step size of a WZ frame and its neighboring key frames. We next apply this model to propose an adaptive algorithm to meet and maintain a target bit rate by dynamically adjusting the quantization parameters of key and WZ frames based on the residual energy between the WZ frame and the estimation of the side information at the encoder. We also evaluate the objective quality of our proposed method compared with the Discover method [10] where quantization parameters are predefined (offline exhaustive search).

The paper is organized as follows: In Section 2, transform domain WZ coding is explained. In Section 3, our method of finding the relationship between the quantization step size of key and WZ frames for efficient bit budget distribution is explained in detail. Our adaptive rate control algorithm is described in Section 4, and its objective performance is evaluated in Section 5. Section 6 concludes the paper.

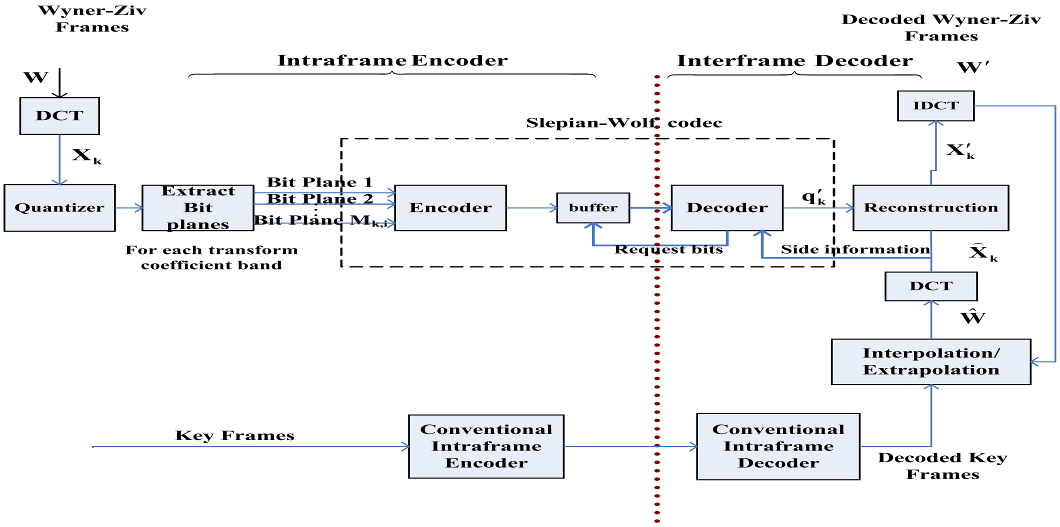


Fig. 1. Transform domain Wyner-Ziv video codec

2. TRANSFORM DOMAIN WYNER-ZIV CODING

The transform domain Wyner-Ziv (TDWZ) video codec architecture proposed in [4] is adopted in this work. As depicted in Fig. 1, key frames are encoded and decoded by a conventional intraframe codec (H.264 in this work). The frames between them (Wyner-Ziv frames) are also encoded independently of any other frame, but their decoding makes use of other frames. In the following, the term decoder refers to the entire interframe decoder of Fig. 1 whereas the term Slepian-Wolf decoder refers to the decoder module inside the Slepian-Wolf codec.

At the encoder, a blockwise 8×8 discrete cosine transform (DCT) is applied on Wyner-Ziv frames as in [11]. If there are N blocks in the image, X_k (for $k = 1$ to 64) is a vector of length N obtained by grouping together the k^{th} DCT coefficients from all blocks. All the coefficients are quantized using uniform scalar quantizers. A quantization matrix (QM) is used to provide finer quantization to more perceivable frequency components. Here we adopt the well known QM from the JPEG standard. We use QS_{WZ} to denote the quantization step size of a WZ frame, and it is used to quantize the DCT coefficients of WZ frames as follows $Q(a_{i,j}) = round\,(\frac{a_{i,j}}{QS_{WZ} \times c_{i,j}})$, where $Q(a_{i,j})$ and $a_{i,j}$ are the quantized and unquantized coefficients at position (i, j), respectively. $c_{i,j}$ is the element of the QM at position (i, j). The coefficients of X_k are quantized to form a vector of quantized symbols, q_k. That is, q_k is the vector of quantization step indices for the elements of X_k. After representing the quantized values in binary form, bit plane vectors $M_{k,i}$ ($i = 1$ to I_k) are extracted, where I_k is the maximum number of bit planes for frequency band k, is calculated by:

$$I_k = \begin{cases} \lfloor log_2|v_k|_{max} + 1 \rfloor & \text{if } k = 1 \\ \lfloor log_2|v_k|_{max} + 1 \rfloor + 1 & \text{otherwise} \end{cases} \qquad (1)$$

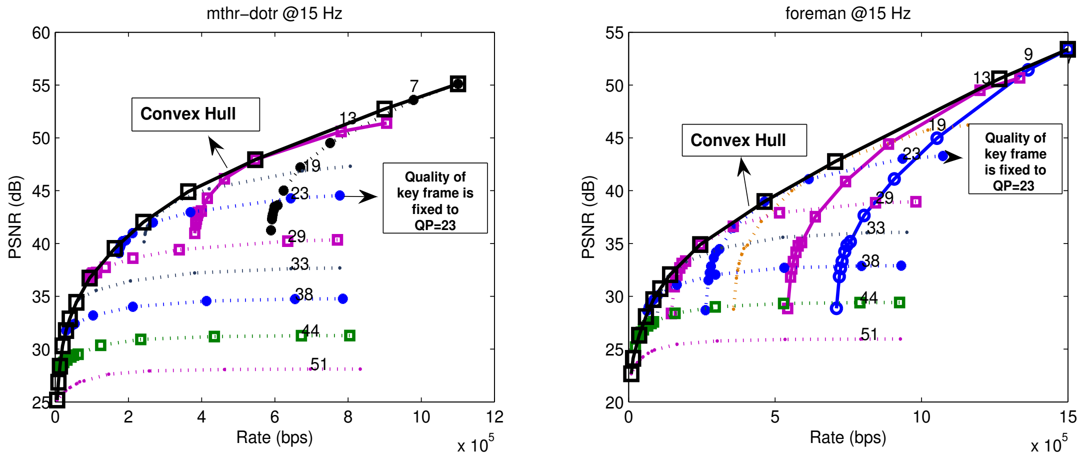
Here $|v_k|_{max}$ is the highest absolute value within frequency band k. The encoder lets the decoder know the maximum number of bit planes for each frequency band within a frame. Each bit-plane vector then enters the Slepian-Wolf (Turbo or LDPCA) encoder. The parity bits (or accumulated syndrome bits) generated by the Turbo (or LDPCA) encoder are stored in the buffer and sent in chunks (upon decoder request through the feedback channel) until a desired bit error rate (in our case, $< 10^{-3}$) is met.

At the decoder, $\hat{W}$ is the estimate of W (Wyner-Ziv frame) which is generated by applying extrapolation or interpolation techniques on decoded key frames. For a GOP of size 2, a motion compensation frame interpolation (MCFI) technique as explained in [3] is applied on previous and next key frames to estimate the Wyner-Ziv frame in between. A blockwise 8×8 DCT is applied on $\hat{W}$ to produce $\hat{X}$. $\hat{X}_k$, the side information corresponding to X_k, is generated by grouping the transform coefficients of $\hat{X}$. When all the bit-planes are decoded, the bits are regrouped to form a vector of reconstructed quantized symbols, $\acute{q}_k$. At the end, the reconstructed coefficient band X'_k is calculated as $E(X_k|\acute{q}_k, \hat{X}_k)$.

The Slepian-Wolf decoder and reconstruction block assume a Laplacian distribution to model the statistical dependency between X_k and $\hat{X}_k$. The distribution of d can be approximated as $f(d) = \frac{\alpha}{2}e^{-\alpha|d|}$, where d denotes the difference between corresponding elements of X_k and $\hat{X}_k$. A different α parameter is assigned for each frequency band and is estimated by the method proposed in [3].

3. DEPENDENCE BETWEEN KEY AND WZ FRAME QUALITY

In Wyner-Ziv video coding, the bit budget for each GOP is shared between key and WZ frames. To propose a rate control algorithm, an efficient method to distribute the bit budget



(a) (b)

Fig. 2. PSNR vs. rate of WZ codec at a given quality of Key frames for different quality of WZ frames

for GOP between key and WZ frames should be investigated. key frames are intra encoded and decoded by a conventional video codec. Therefore the rate distortion performance of key frames is independent of other frames. MCFI methods are applied on key frames to generate side information. Since side information is used in both the decoding and reconstruction blocks, the rate and distortion of WZ frames strongly depend on the quantization step size of the key frames and MCFI success. MCFI methods provide better estimation where the motion is smooth and translational. Therefore MCFI methods are usually more successful for low motion sequences than high motion ones. So, the compression efficiency of WZ frames is affected by their motion activity. In this section we will investigate the relationship between the quantization step size of WZ frames and key frames in order to efficiently distribute the bit budget between key and WZ frames. As a first step, the impact of the quality of key frames is studied for sequences with different motion characteristics. We used $Claire$, $Mother$-$daughter$, $Foreman$, and $Soccer$ QCIF (176×144) sequences at 15fps which are different in content and motion characteristics.

In Fig. 2 (a)-(b), the x axes show the average rate of all frames and the y axes show the average PSNR of all frames. Average PSNR was calculated by first computing average MSE across all frames and then converting the final result to PSNR. The QP of key frames, QP_K is fixed for each curve. There are several points on each curve representing different quantization step sizes of WZ frames. In H.264, QP values range from 0 to 51 and it is possible to calculate the equivalent quantization step size (Qstep) for each value of QP. As QP increases, Qstep increases; in fact, Qstep doubles for every increase of 6 in QP. In this work, the Qstep set for key frames is $\{1.375$, 1.75, 2.25, 2.75, 4.5, 5.5, 7, 9, 11, 18, 22, 28, 40, 52, 72, 104, $224\}$ corresponding to the QP set $\{7$, 9, 11, 13, 17, 19, 21, 23, 25, 29, 31, 33, 36, 38, 41, 44, 47, $51\}$. The Qstep set for WZ frames is $\{0.03$, 0.05, 0.195, 0.5, 1.47, 3.55, 4, 5, 7, 9, $12\}$. Each rate distortion (RD) point in these figures corresponds to a certain Qstep vector $QS = \{QS_K, QS_{WZ}\}$ for key frames and WZ frames. For each sequence, the convex hull of all these RD points gives the optimum rate distortion curve. To find the best relationship between the Qstep of WZ and key frames, we study the set of QS vectors corresponding to RD points forming the convex hull of each sequence. Empirically, these sets of QS vectors for $Foreman$ and $Soccer$ which are relatively high motion sequences are the same. The ones for $Claire$ and $Mother$-$daughter$ which are relatively low motion sequences are also the same. So, we define two classes, low and high motion activity. We study our experimental results for each class separately to find the relationship between quantization step size of key and WZ frames. The proposed method to classify each GOP into low or high motion activity class is explained later in Section 4.3.

As shown in Fig. 2 (a)-(b) at low rates, the convex hull is obtained by connecting the very first point of consecutive curves from the left side. These points belong to the cases where no bits are sent for WZ frames, and only side information generated by key frames is used to reconstruct them at the decoder. This indicates that the best bit allocation at low rates is allocating the total bit budget to key frames. It should be noted that the range of what is considered low rate is different based on the motion activity of the sequence.

3.1. High motion activity

MCFI methods become less successful where there is high motion activity. Therefore the error between a WZ frame and its corresponding side information grows and as a result more bits need to be sent. The QS vectors corresponding to RD points forming the convex hull for relatively high motion sequences ($Foreman$ and $Soccer$)are $(1.75, 0.03)$, $(2.25, 0.05)$, $(7, 0.195)$, $(11, 0.5)$, $(40, 4)$, $(52, 5)$, $(72, 7)$, $(104, 12)$. For these QS vectors, QS_{WZ} is plotted vs. QS_K in Fig. 3. We use a curve fitting technique to estimate QS_{WZ} as a polynomial function of QS_K. We divide these points into two separate sets, $S_1 = \{\ (1.75, 0.03), (2.25, 0.05), (7, 0.195), (11, 0.5), (40, 4),\ \}$ and $S_2 = \{(40,4),(52,5),(72,7),(104,12)\}$, to have a more reliable estimation. A least squares fitting technique was used to find the best quadratic polynomial fit for each sample set, which for set S_1 is $f(x) = 0.0019\,x^2 + 0.00242\,x - 0.0246$, and for set S_2 is $f(x) = 9.5 \times 10^{-4}\,x^2 - 0.0124\,x + 3.01$. Fig. 3 (a) and (b) show data of sets S_1 and S_2 and their polynomial fits.

3.2. Low motion activity

The QS vectors corresponding to RD points forming the convex hull for the relatively low motion ($Claire$ and $Mother$-$daughter$) sequences are $(1.375, 0.03)$, $(1.75, 0.05)$, $(2.75, 0.195)$, $(4.5, 0.5)$, $(7, 1.47)$, $(11, 3.55)$, $(18, 10)$. The best polynomial fit of degree 2 for these points is $f(x) = a_1x^2 + a_2x + a_3$, where a_1, a_2 and a_3 are equal to 3.2×10^{-2}, -3.17×10^{-2} and 1.8×10^{-2}, respectively. Fig. 3 (c) shows data of set S_3 and its polynomial fit.

4. RATE CONTROL

In the previous section we found a relationship between the Qstep size of key and WZ frames in order to efficiently distribute the bit budget between key and WZ frames for low and high motion activity cases. We use this result to define an algorithm to first select Qstep sizes for key frames based on the target rate, and then for WZ frames based on their motion activities and on the Qstep size of the corresponding key frames. The bit budget for each GOP (B_{GOP}) is calculated by $B_{GOP} = R_T \times \frac{N}{f}$ where R_T , N and f are the target rate, GOP size and frame rate, respectively. In this work we consider a GOP of size 2 which is commonly used in the context of WZ video coding. Each GOP consists of a WZ frame and the next adjacent key frame.

4.1. The relation between rate and Qstep size for key frames

For key frames, the source distortion is closely related to the quantization error which is controlled by the quantization step size. The relation between rate and quantization is usually derived based on a rate-distortion (R-D) model. In order to keep the encoder low complexity, in this work we use the simple R-D model as $B_K = \frac{A}{QS_K}$, where A is a constant, QS_K is the quantization step size and B_K is the number of coded bits for the frame. In practice, we start coding the first and third frames of the sequence which are key frames with some initial QP. A gets updated as $A = B_K \times QS_K$ every time a key frame is coded. This is used to calculate QS_K for the next GOP. In this work, we set the initial QP based on the target average number of bits per pixel (bpp), calculated by $bpp = \frac{\frac{B_{GOP}}{N}}{W \times H}$ where B_{GOP}, N, W and H are the bit budget for each GOP, GOP size, width and height of each frame, respectively. The initial QP is set as follows:

$$Initial\ QP = \begin{cases} 45 & \text{if } 0 < bpp \leq 0.3 \\ 35 & \text{if } 0.3 < bpp \leq 0.7 \\ 30 & \text{if } 0.7 < bpp \leq 1.5 \\ 25 & \text{if } bpp > 1.5 \end{cases} \qquad (2)$$

As shown in Equation (2), threshold values are set to 0.3, 0.7 and 1.5. These values are selected based on RD curves in Fig. 2 (a)-(b) to avoid being far away from the target rate in the beginning.

4.2. Choosing Qstep size for key frames

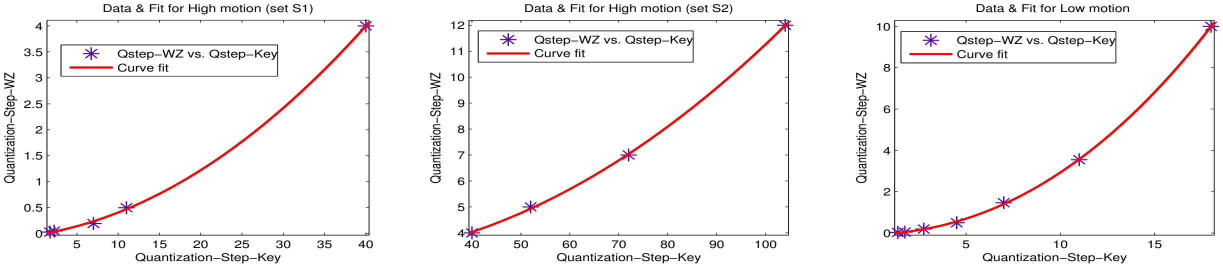
To choose Qstep of key frames for each GOP, we need to find the relationship between the number of bits for each GOP and the corresponding QS_K. Then, given a target number bits, we can determine the Qstep. We use $\widehat{B}_{GOP}$ to denote the estimated number of bits for the next GOP. $\widehat{B}_{GOP} = \widehat{B}_K + \widehat{B}_{WZ}$ where $\widehat{B}_K$ and $\widehat{B}_{WZ}$ are the estimated numbers of bits for corresponding key and WZ frames. As explained before, the number of bits for a key frame can be estimated by $\widehat{B}_K = \frac{A}{QS_K}$ where A is calculated from the previously coded key frame. We define $c = \frac{B_{WZ}}{B_K}$ which gets updated every time a GOP is coded. This is used to estimate the number of bits for the WZ frame of the next GOP as $\widehat{B}_{WZ} = c \times \widehat{B}_K$. Therefore the number of bits for each GOP is estimated by:

$$\begin{aligned} \widehat{B}_{GOP} &= \widehat{B}_{WZ} + \widehat{B}_K \\ &= c \times \widehat{B}_K + \widehat{B}_K \\ &= (c+1) \times \widehat{B}_K \\ &= (c+1) \times \frac{A}{QS_K} \end{aligned} \qquad (3)$$

by setting this equal to the bit budget of the GOP, we calculate QS_K as $QS_K = (c+1) \times \frac{A}{B_{GOP}}$. Since in the H.264 standard, only 51 different values can be used for Qstep, QS_K is set to the one closest to the calculated value. After a GOP is encoded, the actual number of bits used should be subtracted from the total bit budget to obtain the new bit budget. When the motion activity of the next GOP is very different from that of the current GOP, this would result in an abrupt change in Qstep and PSNR between GOPs which can be subjectively annoying. To limit this sharp change, our algorithm does not allow the QP_K difference between GOPs to exceed 2.

4.3. Motion activity classification

To find the Qstep size for WZ frames we first need to determine the motion activity class of the WZ frame. For this purpose we need to estimate the side information at the encoder. The previous or next adjacent key frames or the average of both can be used at the encoder as a rough estimation of side information. From these three candidates, the one that minimizes the total absolute difference D between the candidate



(a) (b) (c)

Fig. 3. Polynomial fit to data points of high and low motion convex hull

side information and the WZ frame is selected as the encoder side information. The candidate that minimizes D is called SI_e and its corresponding D is called D_{min}. D_{min} is calculated for each frame of our four sample sequences.

Table 1 shows the minimum, maximum and average values of D_{min} over all frames for different sequences. To divide frames into low and high motion activity classes, a threshold value must be defined. We considered $Claire$ and $Mother-Daughter$ to be low motion sequences throughout, and so chose 7.6×10^4 which is the highest value of D_{min} over all frames of these two sequences to be the threshold value. A frame is classified as low motion if $D_{min} \leq 7.6 \times 10^4$ and as high motion otherwise. Based on this threshold value, 100% of WZ frames of $Claire$ and $Mother - Daughter$ and 0% and 16% of WZ frames of $Soccer$ and $Foreman$ are considered to be low motion.

4.4. Choosing Qstep size for WZ frames

Once the QS_K is determined, QS_{WZ} which is the quantization step size of the WZ frame can be calculated. As a first step, D_{min} which estimates the motion activity of the WZ frame should be calculated as in Section 4.3. Then the WZ frame is classified as low or high motion using Equation (4). Based on the results of Section III-A and B, QS_W is calculated as follows:

For a low motion activity WZ frame:

$$QS_{WZ} = \begin{cases} 0.032QS_K^2 - 0.03QS_K + 0.02 & \text{if } QS_K \leq 18 \\ \infty & \text{otherwise} \end{cases} \tag{4}$$

$QS_{WZ} = \infty$ means that all bit budget goes for the key frame and no bits are sent for the WZ frame. For a high motion WZ frame:

$$QS_{WZ} = \begin{cases} 0.002QS_K^2 + 0.02QS_K - 0.02 & \text{if } QS_K \leq 40 \\ 0.001QS_K^2 - 0.01QS_K + 3 & \text{if } 40 < QS_K \leq 104 \\ \infty & \text{otherwise} \end{cases} \tag{5}$$

5. RATE CONTROL SIMULATION RESULTS

The test sequences are $Foreman$, $Hall-Monitor$, $Coastguard$ and $Soccer$ QCIF (176 × 144) at 15 fps. Fig. 4 (a)-(b) compares the rate-distortion performance of the WZ video codec applying our proposed rate control algorithm to the one applying the method proposed in [10] (Discover) for the test sequences. As we can see, our proposed rate control algorithm which automatically selects the Quantization step size of key and WZ frames based on motion activity provides better or equal performance to Discover where the quantization parameters of key and WZ frames are selected offline. For $Hall-Monitor$, the performance gain is up to 0.8 dB, whereas for the other test sequences the gain is smaller or negligible. Note that the main purpose of our work is to accomplish on-line simple rate control to achieve a given target rate. The fact that our method also gives a slight performance improvement is an extra benefit.

Our success in achieving the target rate after coding each GOP is monitored by calculating $R_{GOP}(n)$ and shown in Fig. 5 (a)-(b). $R_{GOP}(n)$ is the actual rate achieved after coding the n^{th} GOP and is calculated by $R_{GOP}(n) = \frac{\sum_{i=1}^{2*n+1} B_F(i)}{2*n+1} \times f$, where $B_F(i)$ is the number of coded bits for frame i, n is the GOP number and f is the frame rate which is 15 for our test sequences. As we can see, our proposed method is capable of achieving the target rate quickly. For our test sequences the achieved rate is at most 10% different from the target rate after a maximum of 15 GOPs.

6. CONCLUSION

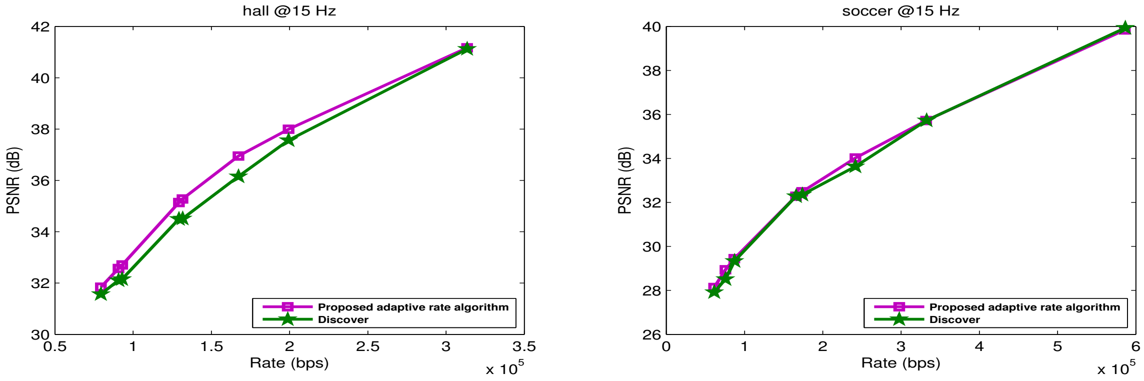
We investigated the relationship between quantization step sizes of key and WZ frames in a GOP of size 2 for WZ video coding in order to efficiently distribute the bit budget between key and WZ frames. The result was applied to propose a rate control algorithm which dynamically adjusts the quantization parameters to achieve a certain rate. In this method, the quantization step sizes of key and WZ frames was automatically adjusted based on the target rate and motion activity of the WZ frame. In our approach, GOPs are differentiated in selecting Qstep of key and WZ frames based on their motion characteristics. Simulation results showed the proposed method achieves better (up to 0.8 dB) or equal rate distortion performance to [10] where there is no rate control and all quantization parameters are set offline for each individual sequence which is therefore applicable only for archival video. In contrast, our approach would be valid for low delay video coding and limited bandwidth or file size applications as all coding parameters are determined automatically on the fly with a bit rate constraint. The experimental results showed that the proposed method is successful to meet the target rate after a few GOPs.

7. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

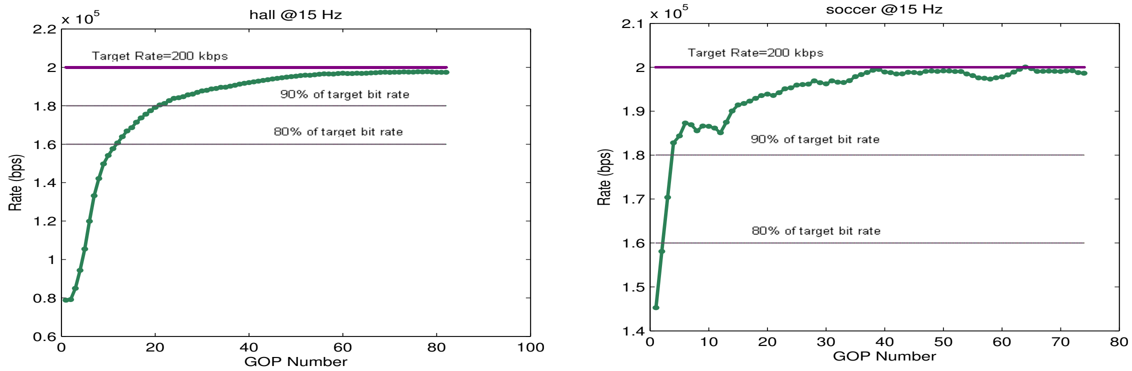
Table 1. Minimum, Maximum and Average of D_{min} over all frames of four sequences

D_{min}	$Minimum$	$Maximum$	$Average$
Soccer	1.54×10^5	7.63×10^5	3.71×10^5
Foreman	0.24×10^5	6.13×10^5	2.02×10^5
Mother-Daughter	0.16×10^5	0.76×10^5	0.37×10^5
Claire	0.11×10^5	0.46×10^5	0.22×10^5



(a) (b)

Fig. 4. Rate-distortion performance of WZ video codec applying our proposed rate control algorithm and Discover method



(a) (b)

Fig. 5. The achieved bit rate at each GOP for 15 fps sequences

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[3] G. Esmaili and P. Cosman, "Wyner-Ziv video coding with classified correlation noise estimation and key frame coding mode selection," *IEEE Trans. on Image Processing*, vol. 20, pp. 2463–2474, Sep. 2011.

[4] A. Aaron, S. Rane, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *VCIP*, vol. 5308, pp. 520–528, Jan. 2004.

[5] J. D. Areia, F. Pereira, and W.A.C. Fernando, "Impact of the key frames quality on the overall Wyner-Ziv video coding performance," *50th International Symposium ELMAR-2008*, 2008.

[6] S. Sofke, F. Pereira, and E. Muller, "Dynamic quality control for transform domain wyner-ziv video coding," *EURASIP Journal on Image and Video Processing*, 2009.

[7] A. Roca, M. Morbee, J. Prades-Nebot, and E. J. Delp, "Rate control algorithm for pixel-domain Wyner-Ziv video coding," *VCIP*, vol. 6822, 2008.

[8] U. Cirac, M. Dalai, and R. Leonardi, "Adaptive key frame rate allocation for distributed video coding," *PCS*, November 2007.

[9] Mariusz Jakubowski, Joo Ascenso, and Grzegorz Pastuszak, "Constant bitrate control for a distributed video coding system," *SIGMAP*, pp. 131–138, 2008.

[10] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," .

[11] J. Zhang, H. Li, Q. Liu, and C. W. Chen, "A transform domain classification based Wyner-Ziv video codec," *ICME*, pp. 144–147, July 2007.