ERROR CONCEALMENT USING A DVC APPROACH FOR VIDEO STREAMING APPLICATIONS

R. Bernardini[†], *M. Fumagalli*[*], *M. Naccari*[‡], *R. Rinaldo*[†], *M. Tagliasacchi*[‡], *S. Tubaro*[‡], *P. Zontone*[†]

[†] **DIEGM**
Università degli studi di Udine
Via Delle Scienze, 208 - 33100 Udine, Italy

[‡]**Dip. di Elettronica e Informazione**
Politecnico di Milano
Via Ponzio, 34/5 20133 - Milano, Italy

[*]**Cefriel**
Politecnico di Milano
Via R. Fucini 2 20133 Milano, Italy

ABSTRACT

In this paper we consider an error resilience scheme, based on a Distributed Video Coding (DVC) paradigm, for the transmission of coded video over an error prone channel. In the proposed scheme, an auxiliary stream which contains generalized parity bits is generated for each coded frame. At the decoder side, error concealed decoded frames are used as side information to feed a Wyner-Ziv turbo decoder. We use an extended version of the Recursive Optimal per-Pixel Estimate (ROPE) algorithm to establish how many parity bits should be sent to the turbo decoder in order to correct the decoded and concealed frame. To validate the proposed scheme, tests with video sequences and realistic loss patterns are reported.

1. INTRODUCTION

In a video broadcasting transmission scenario, encoded bitstreams are sent to multiple receivers over a packet switched network; when the network does not provide any QoS, the encoded data is protected adding some redundancy bits (parity bits) to ensure a correct decoding even if some packets are lost during transmission. This method is known as forward error protection (FEP) and exploits techniques provided by research in the channel coding field. In this approach, it is important to find a criterion to establish a protection degree for coded data, in order to tune the bitrate relative to parity bits with respect to loss rate. Looking for this criterion would require to estimate the distortion of the decoded frames at the encoder side.

The general framework proposed in [1] considers an MPEG coded video bitstream which is sent over an error prone channel with little or no protection; an auxiliary bitstream, generated using Wyner-Ziv coding, is sent for error resilience. At the decoder side, the error concealed decoded MPEG frame becomes the side information for the Wyner-Ziv decoder which further enhances the quality of the concealed frame. The auxiliary Wyner-Ziv stream is generated by computing parity bits of a Reed-Solomon code, where the systematic data consists of a downsampled, coarsely quantized version of the original sequence, together with mode decisions and motion vectors. This scheme works with a fixed rate allocated at the encoder, which does not depend on the actual distortion induced by the channel loss.

If the packet loss rate (PLR) is known, together with the error concealment technique used at the decoder, the ROPE algorithm [2] allows the distortion estimation of a decoded frame without a direct comparison between the original and the decoded signal. In other words, it can estimate, at the encoder side, the expected distortion at the decoder. In its former version [2], the ROPE algorithm performs this estimation considering integer pixel precision for motion compensation in the video coder. Since half pixel precision is widely adopted by standard coders, in this work we use the half pixel precision version of ROPE presented in [3].

In this paper, we propose a forward error correcting coding scheme that employs an auxiliary redundant stream encoded according to a Wyner-Ziv approach. Unlike [1], we use turbo codes to compute the parity bits of the auxiliary stream. We work in the transform domain, by protecting the most significant bit-planes of DCT coefficients. In order to allocate the appropriate number of parity bits for each DCT frequency band, we use a modified version of ROPE, that works in the DCT domain [4]. The information provided by ROPE is also used to determine which frames are more likely to suffer from drift. Therefore, the auxiliary redundant data is sent only for a subset of the frames. We also show how prior information that can be obtained at the decoder, based on the actual error pattern, can be used to efficiently help turbo decoding.

The remainder of the paper is organized as follows: Section 2 presents a brief overview of rate estimation techniques, Section 3 presents the principles of the proposed scheme, including the rate estimator at the encoder. Section 4 reports experimental results with video sequences and compares the proposed scheme with one exploiting feedback from the channel. Finally, Section 5 discusses future and ongoing work.

2. PRELIMINARY REMARKS

To evaluate the number of Wyner-Ziv parity bits which should be transmitted into the auxiliary bitstream, we estimate the expected frame distortion in the DCT domain. Such distortion estimation is computed by means of the ROPE algorithm [2], which requires knowledge of the transmission channel packet loss rate and of the error concealment strategy used at the decoder. In its first and basic formulation, the ROPE algorithm estimates the expected distortion in the pixel domain and considers video coding with integer precision motion vectors (MVs). As mentioned before, integer precision MVs may lead to unacceptable coding performance. In this work, we use the extension of the ROPE algorithm to half-pixel precision as proposed in [3]. To calculate the expected frame distortion at the decoder in the DCT domain, we use the EDDD[1] algorithm [4]. This method requires in input the expected frame distortion in the pixel domain calculated by the ROPE algorithm, and outputs the expected distortion of the coefficients in the transform domain.

[1]Expected Distortion of Decoded DCT-coefficients.

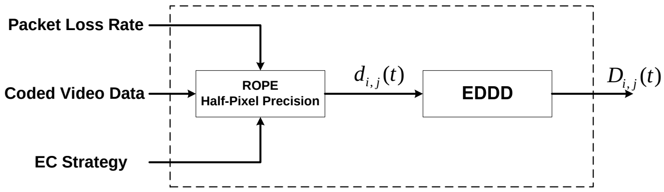


Fig. 1. *Rate Estimator module for the expected distortion in DCT domain.*

Figure 1 shows the *Rate Estimator* module and highlights its inputs and its output, i.e., the expected channel induced distortion $D_{i,j}(t)$ for each DCT coefficient, defined as:

$$D_{i,j}(t) = E[(\hat{X}_{i,j}(t) - \tilde{X}_{i,j}(t))^2] \tag{1}$$

where $i = 0,\ldots,K-1$ denotes the block index within a frame and $j = 0,\ldots,J$ is the DCT coefficient index within a block. $\hat{X}_{i,j}(t)$ represents the reconstructed DCT coefficient at the encoder at time t, while $\tilde{X}_{i,j}(t)$ is the co-located coefficient reconstructed at the decoder, after the error concealment.

We want to emphasize that the estimated value of $D_{i,j}(t)$ represents the channel induced distortion only, and it is obtained as follows. Let c, q denote, respectively, transmission and quantization errors, supposed uncorrelated. For each (i, j) DCT coefficient we can write, at the encoder:

$$\hat{X}_{i,j}(t) = X_{i,j}(t) + q_{i,j}(t), \tag{2}$$

where $X_{i,j}(t)$ is the original DCT coefficient. At the decoder, we get:

$$\tilde{X}_{i,j}(t) = \hat{X}_{i,j}(t) + c_{i,j}(t) \tag{3}$$

The total distortion at the decoder is given by

$$\begin{aligned} E[(X_{i,j}(t) - \tilde{X}_{i,j}(t))^2] &= E[(q_{i,j}(t))^2] + E[(\hat{X}_{i,j}(t) - \tilde{X}_{i,j}(t))^2] \\ &= E[(q_{i,j}(t))^2] + D_{i,j}(t) \end{aligned} \tag{4}$$

where the l.h.s in equation (4) is unknown and represents the variance of drift, provided by EDDD (for details, see [4]). The first term on the r.h.s. can be approximated by knowing the quantizer step size Δ as $E[(q_{i,j}(t))^2] = \Delta^2/12$, by assuming uniformly distributed quantization noise.

The distortion $D_{i,j}(t)$ is then used to compute the amount of transmitted parity bits, to correct the tth error concealed decoded frame $\tilde{X}(t)$ into a "cleaner" version $X'(t)$, as we shall see later.

3. THE PROPOSED SCHEME

The proposed scheme is depicted in Fig. 2. The input video signal is independently coded with a standard motion-compensated predictive (MCP) encoder and a Wyner-Ziv encoder. The generated bitstreams are transmitted over the error-prone channel characterized by a packet loss probability p_l. At the receiver side, the decoder decodes the primary bitstream and performs error concealment. We use a motion-compensated temporal concealment that can be briefly summarized as follows:

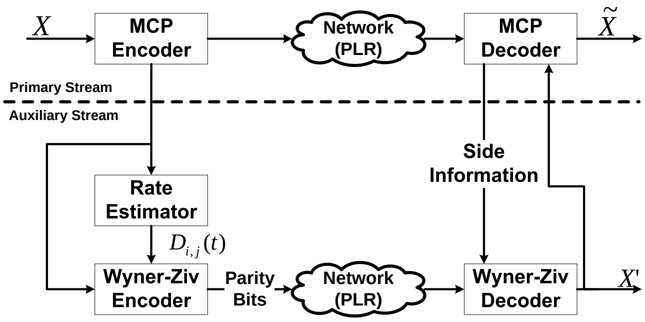


Fig. 2. *Overall of the proposed scheme.*

- If a macroblock (MB) is lost (this happens with probability p_l), then it is replaced with the one in the previous frame pointed by the motion vectors of the MB above the one under consideration;
- If the MB above the one being concealed is lost too (this happens with probability p_l^2), then the current macroblock is replaced with the homologous in the reference frame.

The Wyner-Ziv encoder is similar to the one described in [5], with the difference that in [5] the side information is generated at the decoder by motion-compensated interpolation, while here the side information is the concealed reconstructed frame at the decoder $\tilde{X}(t)$. To prevent mismatch between the encoder and decoder, we use a locally decoded version of the compressed video $\hat{X}(t)$ as input to the Wyner-Ziv encoder, rather than the original video sequence $X(t)$.

At the encoder, Wyner-Ziv redundancy bits for frame $\tilde{X}(t)$ are generated only if the expected distortion at the frame level is above a pre-determined threshold, i.e., only if

$$D(t) = \frac{1}{KJ} \sum_{i=0}^{K-1} \sum_{j=0}^{J-1} D_{i,j}(t) > \tau \tag{5}$$

Otherwise, no redundancy bits are transmitted. This allows us to concentrate the bit budget on those frames that are more likely to be affected by drift.

To generate the Wyner-Ziv bitstream, DCT is applied to the frame $\hat{X}(t)$ of the locally decoded sequence; the transform coefficients are grouped together to form coefficients subbands; each subband is then quantized using a dead-zone uniform quantizer and, for each subband, the corresponding bitplanes are independently coded using a turbo encoder. We transmit the parity bits relative to all the DCT coefficients of each 8×8 block. In the proposed scheme, we use the expected distortion that EDDD computes to adaptively estimate the number of parity bits that are required by the Wyner-Ziv decoder. As a matter of fact, we expect that, if the distortion calculated by ROPE increases, the rate required by the decoder for exact reconstruction increases too. As explained in Section 2, at the encoder it is possible to estimate the expected distortion for each DCT coefficient, denoted by $D_{i,j}(t)$. Since encoding is performed by grouping together DCT coefficients belonging to the same subband j, we first compute the average distortion for each subband as

$$\sigma_j^2(t) = \frac{1}{K} \sum_{i=0}^{K-1} D_{i,j}(t) \tag{6}$$

Then, we need to obtain for each bit-plane b of each subband j, the expected cross-over probability $p_j^b(t)$. To this end, we assume a Laplacian distribution for the error $n_j(t) = \hat{X}_j(t) - \tilde{X}_j(t)$.

$$f_{n_j(t)}(n) = \frac{\alpha_j(t)}{2} e^{-\alpha_j(t)|n|} \quad (7)$$

The distribution parameter $\alpha_j(t)$ is derived by the expected distortion $\sigma_j^2(t)$ as $\alpha_j^2(t) = 2/\sigma_j^2(t)$. The joint knowledge of the Laplacian distribution and of the quantizer used allows us to estimate the cross-over probability $p_j^b(t)$ for each bit-plane. Therefore, we can calculate $H(p_j^b(t)) = -p_j^b(t) \cdot \log_2(p_j^b(t)) - (1 - p_j^b(t)) \cdot \log_2(1 - p_j^b(t))$ and using this value, we have estimated the bits $R_j^b(t) = H(p_j^b(t))$ that an ideal receiver needs to recover the original bits. Note that $H(p_j^b(t))$ is the amount of redundancy required by a code for a symmetric binary channel achieving channel capacity.

The encoder transmits about $R_j^b(t)$ parity bits[2] for each bit-plane and subband, until the auxiliary stream bit budget $R_{WZ}(t)$ is exhausted. First, the parity bits of the most significant bit-plane of all J DCT subbands are sent. Then, encoding proceeds with the remaining bit-planes. The encoder also transmits the average distortion computed by the *rate estimator* module, i.e., $\sigma_j^2(t)$, since it is exploited by the turbo decoder, as explained below. We notice that, as it has been analyzed in [6], by fixing the rate at the encoder we obtain a significant improvement: the transmission rate for a single bitplane never exceeds 1. In fact, if the estimated rate is is greater than 1, it is convenient to transmit the uncoded bits.

At the receiver side, the Wyner-Ziv decoder takes the concealed frame $\tilde{X}(t)$ and transforms it using a block-based DCT. The transform coefficients are then grouped together to form the side information subbands $\tilde{X}_j(t)$. The received parity bits are used to "correct" each subband $\tilde{X}_j(t)$ into $\hat{X}'_j(t)$. Turbo decoding is applied at the bit-plane level, starting from the most significant bit-plane of each subband. In the proposed system, the turbo decoder exploits the received distortions estimated at the encoder $\sigma_j^2(t)$ in order to tune the correlation statistics between the source $\hat{X}_j(t)$ and the side information $\tilde{X}_j(t)$. Therefore, adaptivity is guaranteed both at subband and frame level.

In addition, we propose to exploit the knowledge of the actual error pattern, which is available at the decoder only. In fact, the decoder knows exactly which slices are lost, and can perform an error tracking algorithm in order to determine which blocks might be affected by errors. Apart from the blocks belonging to lost slices, also those blocks that depend from previously corrupted blocks are flagged as "potentially noisy". The error tracker produces, for each frame, a binary map that indicates if the reconstructed block at the decoder might differ from the corresponding one at the encoder. The algorithm is similar to the one presented in [7], with the important difference that in our case error tracking is performed at the decoder only. The turbo decoder can take advantage from this information by adaptively setting the conditional probability to 1 for those coefficients that are certainly correct. This means that the turbo decoder totally trusts the side information in these cases.

After turbo decoding, some residual errors might occur if the number of received parity bits is not sufficient for the exact reconstruction of every bit-plane. In this case, we do not consider the error-prone bit-plane in the reconstruction process; only the bits of the correctly decoded bit-planes are regrouped and the ML reconstruction is finally applied (see [5]). Error detection at the turbo decoder is performed by monitoring the average of the modulus of the LAPP (log a-posteriori likelihood ratio) as described in [8].

Finally, once Wyner-Ziv decoding is successfully completed, the reconstructed frame $X'(t)$ is copied into the buffer of the MCP decoder, to serve as reference frame at time $t+1$. This way the amount of drift propagated to successive frames is reduced.

4. EXPERIMENTAL RESULTS

In our simulations, the input video signal is compressed using an H.263+ video coder with half-pixel precision motion estimation. The compressed bitstream is transmitted using the primary channel. The auxiliary bitstream carries Wyner-Ziv redundancy bits that are used at the decoder to correct the bit-planes of the DCT coefficients of the side-information, i.e., of the reconstructed frames after concealment.

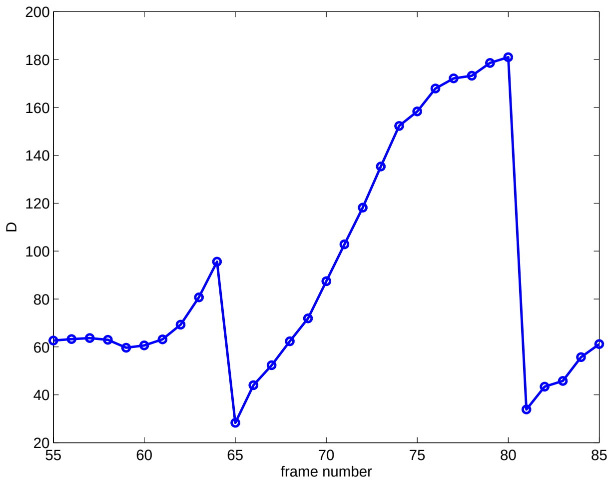
For the first set of simulations, we consider 30 frames of the QCIF *Foreman* sequence. The *Foreman* sequence is coded at 30 fps, QP=4, and GOP size equal to 16. Each slice comprises one row of 16×16 macroblocks, with 9 slices per frame. For transmission, we send one slice per packet, and we assume a packet loss probability $p_l = 10\%$, with independent losses.

As mentioned above, we do not send Wyner-Ziv redundancy bits when the expected distortion estimated by ROPE is below a certain threshold, which represents the minimum acceptable quality at the decoder. Note that such a threshold could be computed by the encoder on the basis of the average expected distortion for past frames. In our experiments, the encoder does not send Wyner-Ziv bits when $D(t) < \tau$, where τ is empirically determined for each sequence. When Wyner-Ziv bits are transmitted, the number of bits is set to $\beta(p_j^b)H(p_j^b(t))$, where $\beta(p_j^b)$ ranges from 2 to 4 depending on $p_j^b(t)$. The values $\beta(p_j^b)$ are set on the basis of operational curves of the Turbo coder performance for a binary symmetric channel with cross-over probability $p_j^b(t)$.

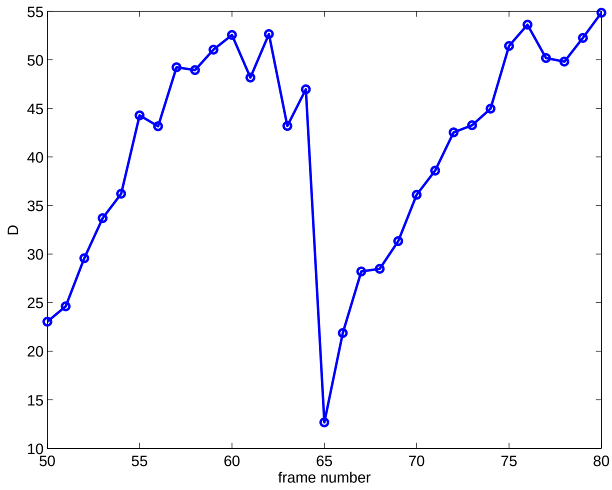
In the following, we will compare the proposed scheme with one using the H.263+ intra-macroblock refresh coding option, where error resilience is obtained by coding, in each frame, a certain of macroblocks in intra-mode. The proportion is chosen to spend roughly the same additional bitrate required by the Wyner-Ziv redundancy stream. Figure 3a shows the ROPE expected distortion for frames 55-85 of the *Foreman* sequence. For this sequence, τ is equal to 80 (equivalent to an expected PSNR equal to 30 dB).

Figure 4 and Figure 5 show the PSNR performance of the proposed scheme and the required rate. The results are obtained as the average of 30 different channel simulations. In this experiment, the main H.263+ stream has an average bitrate $R = 568$ kbps. Additional 63 kbps (equivalent to approximately 11% of the original rate) are spent for the auxiliary Wyner-Ziv coded stream, for a total rate $R = 631$ kpbs.

[2] A higher rate is typically required to take into account the suboptimality of the channel code and model inaccuracy.



(a) *Foreman*



(b) *Carphone*

Fig. 3. Expected distortion given by ROPE for *Foreman* and *Carphone* sequences

Figure 4 also shows the PSNR for H.263+, when the intra macroblock refresh rate has been set equal to 6 to achieve approximately the same rate ($R = 624$ kbps) as the proposed scheme.

It is apparent from the figures that the proposed scheme has comparable or better performance than the scheme using the intra macroblock refresh procedure. To this respect, note that the intra macroblock refresh procedure has to be used at coding time. In our scheme, instead, we are allowed to start from a precoded sequence (the most common situation for video transmission) and add Wyner-Ziv bits for protection.

Similar results can be seen in Figure 6 and Figure 7, where we consider 30 frames of the *Carphone* sequence. In this case, the main H.263+ stream is encoded at a rate $R = 294$ kbps. Additional 89 kbps are added as Wyner-Ziv parity bits to a achieve an overall rate equal to 386 kbps. When intra macroblock refresh is used, the refresh rate is set to 8. The ROPE expected distortion for the *Carphone* sequence is depicted in Figure 3b. In this case, we do not send parity bits when the $D(t)$ is below 35.

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a video coding scheme where channel errors are corrected using an auxiliary bitstream with Wyner-Ziv parity bits. The amount of parity bits is decided on the basis of an estimate, computed at the encoder-side, of the expected distortion at the decoder. The scheme requires no feedback and compares well with previous schemes presented in the literature. Further research is required to compare the proposed scheme with other error protection procedures, e.g., packet-based FEC protection.

REFERENCES

[1] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod, “Systematic lossy forward error protection for error-resilient digital video broadcasting - a Wyner-Ziv coding approach,” *IEEE International Conference on Image Processing, Singapore*, 2004.

[2] R. Zhang, S. L. Regunathan, , and K. Rose, “Video coding with optimal inter/intra mode switching for packet loss resilience,” *IEEE Journal on Selected Areas in Communications*, vol. 18 (no 6), pp. 966–76, 2000.

[3] V. Bocca, M. Fumagalli, R. Lancini, and S. Tubaro, “Accurate estimate of the decoded video quality: extension of ROPE algorithm to half-pixel precision,” *Picture coding Symposium (PCS)*, pp. 15–17, Dicember 2004.

[4] M. Fumagalli, M. Tagliasacchi, and S. Tubaro, “Expected distortion of video decoded DCT-coefficients in video streaming over unreliable channel,” *International Workshop on Very Low Bitrate Video Coding*, vol. 02, pp. 1–4, 2005.

[5] A. Aaron, S. Rane, E. Setton, and B. Girod, “Transform-domain Wyner-Ziv codec for video,” *Proc. Visual Communications and Image Processing, VCIP-2004, San Jose, CA*, Jan. 2004.

[6] D. P. Varodayan, “Wyner-Ziv coding of still images with rate estimation at the encoder,” *IEEE*, vol. 1, p. 1, 2002.

[7] B. Girod and N. Farber, “Feedback-based error control for mobile video transmission,” *Proceedings of the IEEE*, vol. 87, pp. 1707–1723, year=1997, month=Oct.

[8] F. Zhai and I. Fair, “Techniques for early stopping and error detection in turbo decoding,” vol. 51, pp. 1617–1623, Oct. 2003.

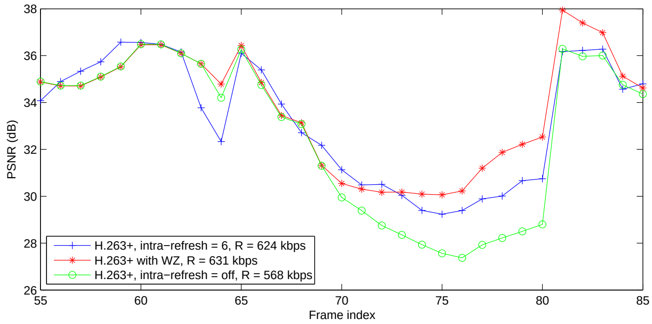


Fig. 4. *PSNR performance for 30 frames of the* Foreman *sequence.*

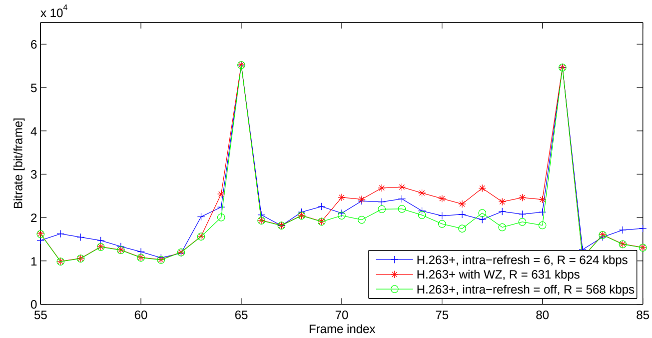


Fig. 5. *Frame based rate allocation for the* Foreman *sequence.*

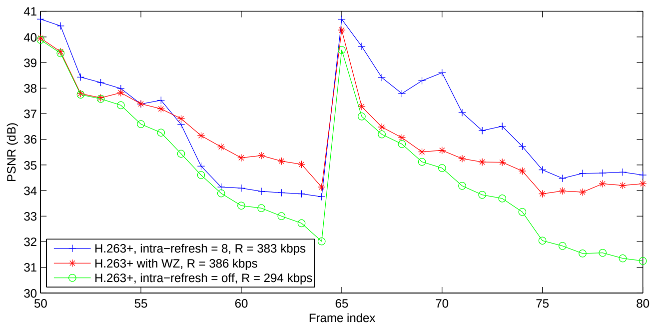


Fig. 6. *PSNR for 30 frames of the* Carphone *sequence.*

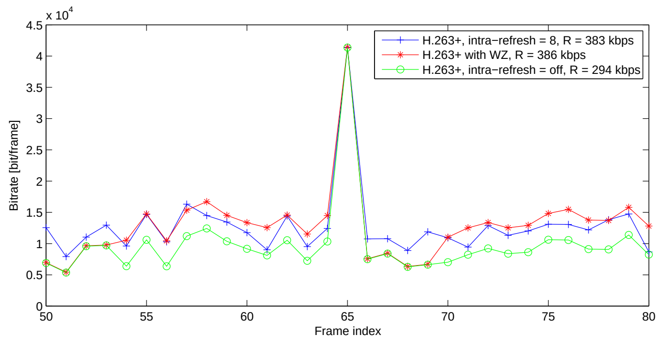


Fig. 7. *Frame based rate allocation for the* Carphone *sequence.*