16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

A NOVEL TECHNIQUE FOR PRACTICAL IMPLEMENTATION OF PIXEL-DOMAIN WYNER-ZIV VIDEO CODING IN MULTI-USER SYSTEMS

Charles Yaacoub[1,2], Joumana Farah[1], and Béatrice Pesquet-Popescu[2]

[1]Faculty of Sciences and Computer Engineering, Holy-Spirit University of Kaslik, Jounieh, Lebanon
{charlesyaacoub, joumanafarah} @ usek.edu.lb
[2]Signal and Image Processing Dept., Ecole Nationale Supérieure des Télécommunications, Paris, France
beatrice.pesquet @ enst.fr

ABSTRACT

In this paper, we present a novel rate allocation technique with a frame dropping mechanism for multi-user distributed video coding systems. Using an analytical approach based on entropy calculations, the quantization parameters are dynamically varied and the available system bandwidth unequally distributed among multiple users without the need for a permanent feedback channel. These calculations take into account the actual amount of movement in the transmitted video as well as the transmission channel conditions of the different users. A significant gain in the system performance is noticed, compared to the case with equal allocation of channel resources and constant quantization parameters.

1. INTRODUCTION

During the last decade, distributed source coding [1-3] has gained a great deal of attention especially in the area of video applications. In traditional video coding techniques such as MPEG or H.26x, motion estimation is performed at the encoder side, yielding very complex inter-frame encoders. The aim of distributed video coding (DVC) is to permit low-complexity encoding for small power-limited and memory-limited devices such as camera-equipped mobile phones or wireless video sensors, by moving the computation burden from the encoder side to the decoder. The idea emerged essentially from the Slepian-Wolf theorem [4] stated as follows: given two statistically dependent sources X and Y, with Y being separately compressed to its entropy limit H(Y), X can be transmitted at a rate very close to the conditional entropy H(X|Y), provided that Y is perfectly recovered at the receiver as side information. The application of this concept to lossy source coding is known as the Wyner-Ziv coding [5]. The most popular pixel-domain Wyner-Ziv video codec was first introduced by Girod et al. in [2]. A subset of frames, known as key frames, is compressed using traditional intra-coding techniques, and used to generate the side information at the decoder. One or more frames following each key frame, known as the Wyner-Ziv (WZ) frames, are then compressed by appropriate puncturing of a turbo-encoder output. One of the main drawbacks of this system is the use of a feedback channel (FC) [6] to allow flexible rate control and ensure successful decoding of WZ frames. Since several decoding runs are required to successfully recover a WZ frame, the FC imposes instantaneous decoding in the receiver. Additionally, in a multi-user system, the application of WZ coding becomes impractical because of the difficulty of implementing appropriate rate allocation algorithms, especially when the system is under harsh rate and delay constraints. For all these reasons, the introduction of new techniques for estimating the necessary bitrate to successfully decode each WZ frame becomes crucial. In fact, the problem of the return channel in DVC has rarely been targeted in the literature. Simple techniques that allow the removal of the feedback channel were recently proposed by Artigas and Torres [7] and Morbée et al. [8]. These techniques mainly rely on performance tables built by running offline simulations. These tables are used by the encoder to predict the compression level of each particular frame. However, these studies do not take into account the rate constraints in limited bandwidth applications. Besides, the influence of the channel impairments on the proposed rate allocation techniques is not considered. Moreover, both techniques are designed for a single-user scenario; their implementation in a multi-user application would cause unoptimal distribution of the channel resources between the different users. Furthermore, a fixed quantization parameter (e.g. a fixed number of quantization levels) is used for the compression of a given sequence at a target bitrate.

In a previous work [9], we presented an analytical approach for estimating the compression limits of a pixel-domain Wyner-Ziv video coding system with a transmission over error-prone channels, without the need for a feedback channel. Simulation results showed that the theoretical bounds can be used in a broadcasting system to predict the compression level for each frame with a minor loss in the decoding PSNR, compared to the classical feedback-based coding system. In [10], we introduced an adaptive rate allocation (ARA) technique for Wyner-Ziv video coding in a multi-user application. In this paper, we show the advantage of using a dynamic quantization algorithm and a frame dropping mechanism, along with our ARA technique. Dynamic quantization and rate allocation are performed depending on the amount of movement in each captured video and on the transmission channel conditions of the different users. When the analytical estimations show that successful decoding of a given WZ frame will not be possible in the receiver, because of very high motion and/or bad channel conditions, the WZ frame is dropped at the transmitter in order to avoid unnecessary channel use.

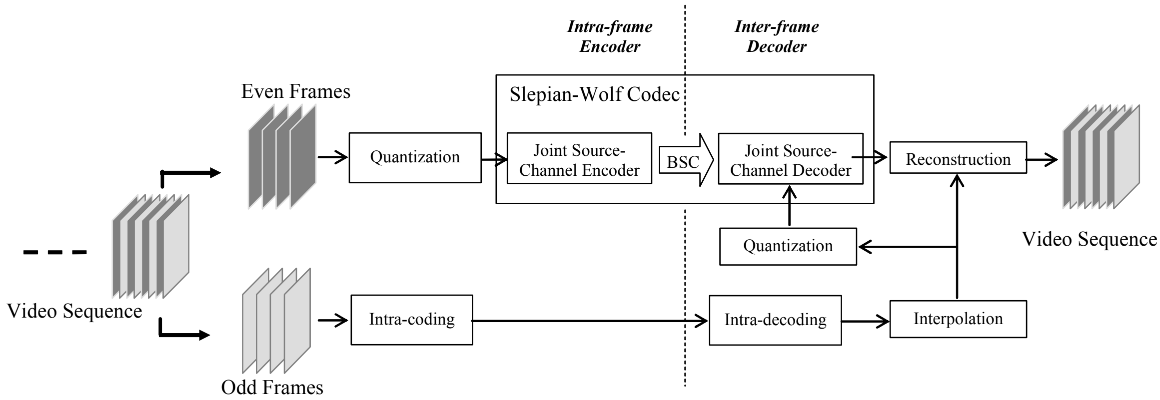

Figure 1. Block-diagram of the pixel-domain distributed video coding system.

This paper is organized as follows: we begin, in Section 2, by a brief description of the WZ video codec and the calculation of the theoretical compression bound. In Section 3, we present our proposed algorithm and its application in a network of multiple wireless users. Simulation results are presented and discussed in Section 4.

2. SYSTEM DESCRIPTION AND THEORETICAL COMPRESSION BOUND

The distributed video coding system considered in this study can be represented by the block-diagram in Fig.1. Odd frames are compressed using traditional intra-coding techniques and are assumed to be perfectly recovered at the receiver. Side information of a particular even frame is generated by motion-compensated interpolation of the two adjacent odd frames, with symmetric motion vectors [2].

As for the compression of the even frames, it starts by a uniform scalar quantization to obtain an M-bit representation of the eight-bit pixels, $M \in \{1,2,4\}$. The source-channel encoder consists of a parallel concatenation of two 16-state quadri-binary convolutional encoders separated by an internal interleaver and resulting in a minimum global coding rate of 2/3. The generator polynomials in octal notation are $(23, 35, 31, 37, 27)_8$ from [11]. At the encoder output, systematic information is discarded, while parity information is punctured and transmitted to the decoder.

In the joint source-channel decoder, the conditional probabilities in the turbo-decoding process depend on the residual signal statistics between the even frames and the side information on one hand, and on the channel conditions on the other [3]. The reconstruction block is used to recover an eight-bit version of the even frame using the available side information [2].

In case of error-free transmission, the compression bound of the WZ codec in Fig.1 is calculated using the conditional entropy H(X|Y) defined as:

$$H(X \mid Y) = -\sum_{i=0}^{2^M-1} \sum_{j=0}^{2^M-1} P(Y=j)\, g\left[P(X=i \mid Y=j)\right], \quad (1)$$

where X represents the WZ frame, Y represents the interpolated side information, M the quantization parameter representing the number of quantization bits per pixel, and $g(x) = x \log_2(x)$. The statistics of the residual error between the side information and the WZ frame are modeled by a Laplacian distribution [2] with parameter α: $P(X-Y=d) = c(\alpha/2)e^{-\alpha|d|}$, where c is a normalization factor such that $\sum_d P(X-Y=d) = 1$.

In [9], we have shown that Eq.1 can be expressed as:

$$H(X \mid Y) = \frac{-1}{2^M} \sum_{i=0}^{2^M-1} \sum_{j=0}^{2^M-1} g\left[c \frac{\alpha}{2} \frac{2^M e^{-\alpha |d_{i-j}|}}{L_{d_{i-j}}} \right], \quad (2)$$

where $L_{d_{i-j}} = 2^M - |i-j|$ represents the number of possible couples (i, j) that yield the difference i - j = d, and $d_{i-j} = 2^M(i-j)$ is the difference between two quantized pixel values. The parameter α can be approximately estimated at the receiver side using the available odd frames. It can also be estimated by the encoder and transmitted as side information to the receiver.

Let $H_f(M)$ be the theoretical lower compression bound for frame *f* transmitted in the absence of noise expressed as in Eq.2. In this study, we model the transmission channel between the mobile user and the base station by a Binary Symmetric Channel (BSC) with a crossover probability *p*. Any other channel model can be used in our system by mapping it to a BSC [9] using an equivalence of the stability functions, as detailed in [12]. Based on [1], the overall theoretical compression bound becomes:

$$H_f'(M) = H_f(M) / \mathscr{C}(p), \quad (3)$$

where $\mathscr{C}(p)$ is the capacity of the BSC.

3. RATE ALLOCATION WITH DYNAMIC QUANTIZATION AND FRAME DROPPING

Consider a system of N mobile users sharing the same wireless medium to transmit data to a base station at a total bitrate of R bits per second (bps). Instead of assigning R/N bps for each user, the base station first determines the lower compression bound for each frame (Eq.3) based on its content (parameter α) and on the user transmission conditions (channel crossover probability *p*). Since the optimal quantization parameter has not been determined yet, Eq.3 is calculated for

M=8, assuming 8-bit raw video data before compression. Then, in a proportionally fair attribution, the user n is assigned the rate:

$$R_{f,n} = \frac{H'_{f,n}(8)}{\sum_{n=1}^{N} H'_{f,n}(8)} R, \tag{4}$$

where $H'_{f,n}(8)$ represents Eq.3 calculated for frame f at node n with M=8.

Let $\rho_{f,n}$ be the compression rate for frame f at node n defined as the ratio of the number of output bits (after compression) over the number of input bits (before compression), and $A_{f,n} = M_{f,n} \times \rho_{f,n}$ the average number of bits per pixel. $A_{f,n}$ is related to $R_{f,n}$ by:

$$A_{f,n} = \frac{R_{f,n}}{(m \cdot n \cdot q)}, \tag{5}$$

where (m, n) represents the dimensions of a given frame and q the WZ frame rate (in frames per second).

In our joint source-channel codec, $0 \le \rho_{f,n} \le ½$; $\rho_{f,n} = 0$ when no parity bits are sent and $\rho_{f,n} = ½$ when all parity bits are transmitted. Once $A_{f,n}$ has been determined for each user, the base station needs to determine, for each frame f at every mobile node n, the couple ($M_{f,n}$, $\rho_{f,n}$) that yields the best video output after reconstruction.

After a thorough analysis of the system performance observed for different values of M, we noticed that in most cases, for a given target bitrate, choosing the lowest allowable value of M yields the best video quality at the decoder output. Indeed, by reducing the number of quantization levels, the system is able to transmit a greater amount of parity bits to protect the quantized bitstream from channel errors, especially when p increases. However, in some cases, the assigned bitrate is sufficient enough to permit efficient error protection when a greater value of M is selected; thus, a better reconstructed output is obtained. In all cases, we noticed that the system behavior for different configurations of the couples ($M_{f,n}$, $\rho_{f,n}$) is directly related to the ratio between $A_{f,n}$ and the theoretical compression bound defined as:

$$C_{f,n}(M_{f,n}) = \frac{A_{f,n}}{H'_{f,n}(M_{f,n})}. \tag{6}$$

Therefore, we define the thresholds T_1, T_2 and T_4 which indicate the average value of the ratio $C_{f,n}(M_{f,n})$ that permits a correct decoding of a transmitted frame for M=1, 2 and 4 respectively. These thresholds are determined experimentally by observing the system performance for different values of the ratio $C_{f,n}(M_{f,n})$ as will be detailed later. Our proposed algorithm then proceeds as follows (Fig.2):

Step 1: Initially, set $M_{f,n}$ to the lowest value that permits to reach $A_{f,n}$. This will allow for the maximum error protection for a given $A_{f,n}$.

Step 2: Calculate $C_{f,n}(1)$, $C_{f,n}(2)$ and $C_{f,n}(4)$.

Step 3: If $M_{f,n}$=4 and $C_{f,n}(4) \le T_4$, set $M_{f,n}$=2 and $\rho_{f,n} = ½$. In other words, if $A_{f,n}$ could not be reached for $M_{f,n}$<4 (i.e. $A_{f,n}$>1) and the amount of error-protection transmitted with $M_{f,n}$=4 does not yield an acceptable decoding error rate, set $M_{f,n}$ to the next lower value and transmit all parity bits.

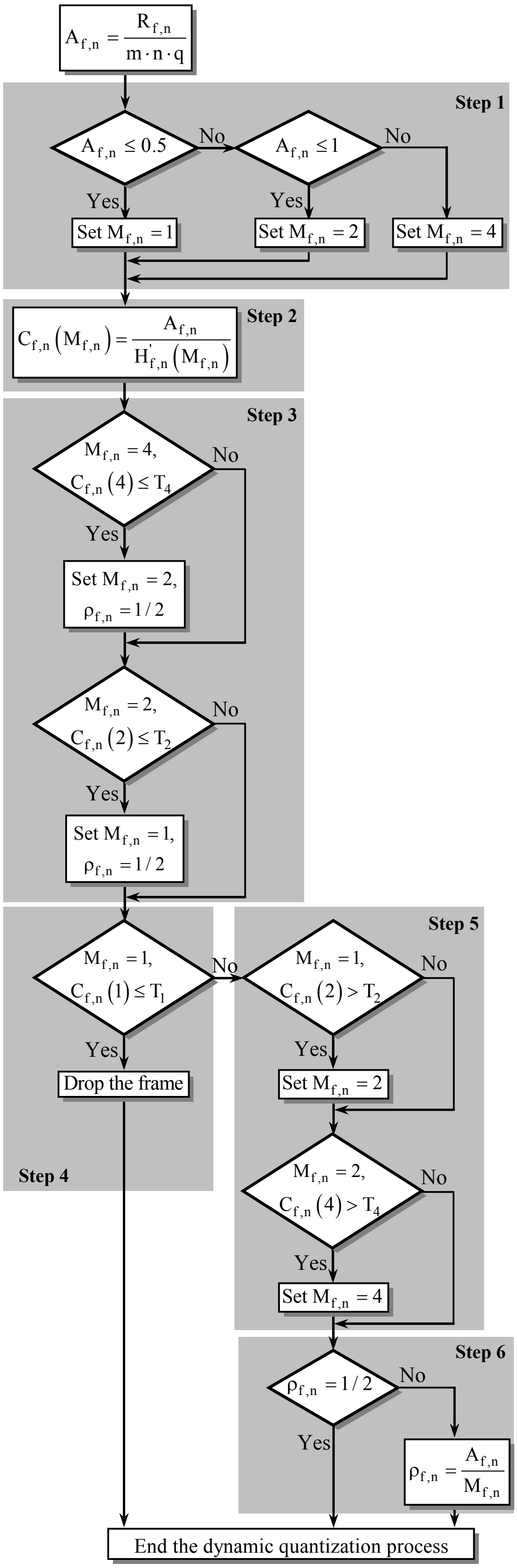

Figure 2. Dynamic quantization algorithm with a frame dropping mechanism.

In this case, the given frame is transmitted at a rate lower than the target bitrate since the target $A_{f,n}$ could not be reached exactly. Similarly, if $M_{f,n}$=2 and $C_{f,n}(2) \leq T_2$, set $M_{f,n}$=1 and $\rho_{f,n} = ½$.

Step 4: If $M_{f,n}$=1 and $C_{f,n}(1) \leq T_1$, drop the frame (set $M_{f,n}$=0 and $\rho_{f,n} = 0$). In this case, the amount of transmitted bits will not permit efficient decoding, even if the number of quantization levels is reduced to its minimum. At the decoder, the dropped frame is replaced by the corresponding side information.

Step 5: If $M_{f,n}$=1 and $C_{f,n}(2) > T_2$, set $M_{f,n}$=2. In other words, if $A_{f,n}$ is reachable with $M_{f,n}$=1 and 2, and it is possible to send a sufficient amount of parity bits to correctly decode the frame with $M_{f,n}$=2, set $M_{f,n}$=2 since it yields a better reconstructed output. Similarly, if $M_{f,n}$=2 and $C_{f,n}(4) > T_4$, set $M_{f,n}$=4.

Step 6: If the frame was not dropped and $\rho_{f,n}$ was not already set to ½, set $\rho_{f,n} = A_{f,n} / M_{f,n}$.

Step 7: Transmit the couple ($M_{f,n}$, $\rho_{f,n}$) to the corresponding user.

As it can be seen from the proposed algorithm, control information is sent only once from the base station to the users. Moreover, instantaneous decoding at the receiver is no more required, and only one decoding run is performed for each frame. As a result, all the disadvantages related to the return channel in traditional Wyner-Ziv applications are avoided.

4. SIMULATION RESULTS

In our simulation setup, we consider a set of three mobile users (N=3) capturing different scenes and transmitting the resulting video to a central base station. These scenes are assumed to be the Foreman, Carphone, and Mother-Daughter QCIF video sequences sampled at a rate of 15 WZ frames per second (fps). In our simulations, we consider the first 100 frames of each sequence repeated 50 times. The varying nature of the wireless channel is modeled by a uniform random variation of the channel crossover probability p between 0.001 and 0.02, independently for each user. The side information is generated by motion-compensated interpolation with symmetric motion vectors as described in [2], and the receiver is assumed to be able to determine the value of the Laplacian parameter α with the closest fit to the data.

After assigning a different transmission rate for each user using Eq.4, we need to determine the thresholds T_1, T_2 and T_4 for the system to proceed with the dynamic quantization algorithm as explained in the previous section. In Fig.3, we show the Bit-Error-Rate (BER) obtained after source-channel decoding as a function of the ratio $C_{f,n}$ defined in Eq.6. For each value of M, the transmission of 22500 frames from different video sequences was simulated with variable channel conditions. In general, a BER near 10^{-3} is desired for an acceptable performance [2]. For this reason, we chose T_1=6 and T_2=4. When M=4, a higher BER can be tolerated since the two least significant bits in the quantized pixels have a lower importance than the two most significant bits. Therefore, we chose T_4=2.4.

In Fig.4, we can first see the rate-distortion (R-D) curves obtained for a traditional (TRD) system where all users are

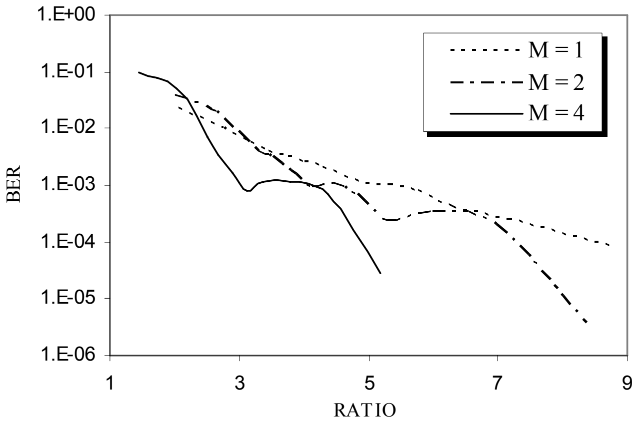

Figure 3. BER as a function of the ratio in Eq.6 for M=1, 2, and 4.

assigned an equal bandwidth. The quantization parameter M is fixed and is the same for all users. The results are presented in terms of the Peak Signal to Noise Ratio (PSNR) averaged over the three video scenes as a function of the total WZ bitrate occupied by all the users. We can also see the results obtained with our adaptive rate allocation (ARA) technique, but with a constant quantization parameter using Eq.4 and Eq.5. We can observe that when the rate regions overlap for different values of M, the best performance is obtained when the lowest value of M is used. For example, at 500 kbps, the use of a 1-bit quantizer yields a gain in performance of nearly 1.5 dB compared to the case with a 2-bit quantizer. A similar effect is noticed at 1 Mbps where the system with a 2-bit quantizer outperforms the one with a 4-bit quantizer by 1.5 dB.

By applying dynamic quantization, in addition to adaptive rate allocation, the problem of overlapping rate regions is avoided. When no frame dropping (ARAQ) is performed (step 4 in the dynamic quantization algorithm is skipped), very close performances to the ARA technique are obtained at low rates while an important enhancement is noticed at medium and high rates. In fact, at very low rates, the available bandwidth to be allocated for the different users is barely sufficient to protect the transmitted bit streams. This leads to a performance loss of 0.5 dB at a rate near 200 kbps, whereas a gain of 1.5 dB is observed with the frame dropping mechanism (ARAQ-D). When the bitrate increases, less frames are dropped. Starting from 600 kbps, the R-D curves are the same with or without frame dropping. On the other hand, at 500 kbps, our proposed algorithm yields similar performance compared to the TRD system with M=1, but a gain of 1.4 dB is observed towards the case where M=2. At 1 Mbps, the gain reaches 0.6 dB compared to M=2 and 1.5 dB to M=4. Furthermore, it is important to note that when M is fixed (TRD and ARA), the transmission bitrate for the WZ codec is limited to a narrow range. For example, a traditional system with N=3 cannot transmit at rates greater than 570 kbps when M=1, and 1140 kbps when M=2. Our proposed algorithm allows a wider range of transmission rates and with an optimized decoding quality.

In Table 1, we show the individual user PSNR ranges obtained with both TRD and ARAQ-D systems. We also show the bitrate ranges corresponding to the ARAQ-D technique.

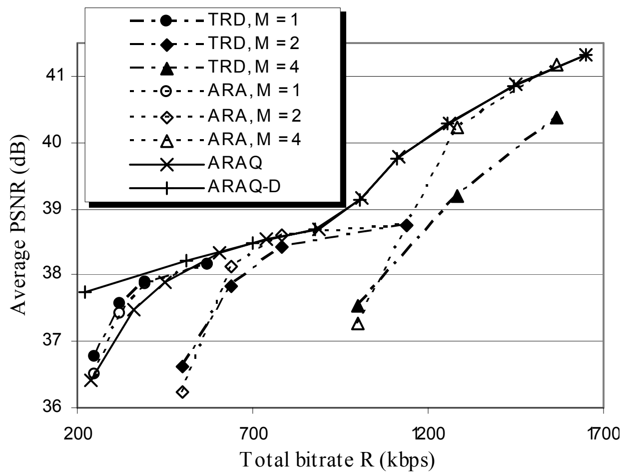

Figure 4. Rate-Distortion curves for different rate allocation and quantization techniques.

The high values of the PSNR obtained with Mother-Daughter are due to the very low motion in this sequence, whereas the two other sequences are characterized by average and high motion. Our proposed algorithm allocates the lowest rates for videos with low motion and/or experiencing good channel conditions, while the highest rates are assigned to users capturing high-motion scenes and/or suffering from a bad channel. Hence, as it can be seen from Table 1, Mother-Daughter is assigned the smallest range of bitrates, whereas in a traditional system, all users operate at the same bitrate (80-550 kbps). Our flexible allocation technique allows for the assignment of higher bitrates to the two other sequences which results in an improvement in the overall system performance. Indeed, a performance gain of 1.1 dB to 1.7 dB is observed for Carphone, 1 dB for Foreman, and 0.1 dB to 0.3 dB for Mother-Daughter.

TABLE I. INDIVIDUAL PSNR AND BITRATE RANGES

Sequence	*TRD PSNR (dB)*	*ARAQ-D PSNR (dB)*	*ARAQ-D Bitrate (kbps)*
Carphone	34.2 - 38	35.3 - 39.7	119.1 - 613.9
Foreman	35.4 - 38.9	36.4 - 39.9	51.9 - 608
Mother-Daughter	40.6 - 44.2	41.4 - 44.3	54 - 426.4
	Average: 36.7 - 40.4	Average: 37.7 - 41.3	Total: 224.9 - 1648.3

In our simulations, key frames were assumed to be perfectly recovered at the receiver. The case where key frames are subject to degradations due to lossy source coding or channel impairments can be easily taken into account in our study by modifying the entropy calculations in Section 3 accordingly. However, it should be noted that, in this case, all studied systems (TRD, ARA, ARAQ, and ARAQ-D) would be subject to a similar performance degradation. As a result, the performance analysis presented earlier would still hold.

5. CONCLUSION

In this paper, we introduced a novel technique for dynamic rate allocation with variable quantization and a frame dropping mechanism in distributed multi-user Wyner-Ziv video coding systems. Using entropy calculations, the available system bandwidth is unequally distributed among several mobile users transmitting data to a central base station without the need for continuous feedback information to the encoders. The gain in the average system performance can reach 1.5 dB compared to a traditional system where all users have equal bandwidth and a fixed quantization parameter.

REFERENCES

[1] A. Aaron and B. Girod, "Compression with Side Information Using Turbo Codes", *Data Compression Conference*, pp. 252- 261, April 2002.

[2] A. Aaron, R. Zhang, and B.Girod, "Wyner-Ziv Coding of Motion Video", 36th Asilomar Conf. Signals, Systems and Computers, pp. 240-244, Nov. 2002.

[3] J. Farah, C. Yaacoub, N. Rachkidy, and F. Marx, "Binary and non-Binary Turbo Codes for the Compression of Correlated Sources Transmitted through Error-Prone Channels", 4th International Symposium on Turbo Codes and Related Topics, Germany, April 2006.

[4] D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. Information Theory*, Vol. IT-19, pp. 471-480, July 1973.

[5] D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Trans. Information Theory*, Vol. IT-22, pp. 1-10, Jan. 1976.

[6] C. Brites, J. Ascenso and F. Pereira, "Feedback Channel in Pixel Domain Wyner-Ziv Video Coding: Myths and Realities", *14th European Signal Processing Conference*, Italy, September 2006.

[7] X. Artigas and L. Torres, "Improved signal reconstruction and return channel suppression in Distributed Video Coding systems", *47th International Symposium ELMAR-2005*, Croatia, June 2005.

[8] M. Morbee, J. Prades-Nebot, A. Pizurica and W. Philips, "Rate Allocation Algorithm For Pixel-Domain Distributed Video Coding Without Feedback Channel", *32nd International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, April 2007.

[9] J. Farah, C. Yaacoub, F. Marx, and B. Pesquet-Popescu, "Performance Analysis of a Distributed Video Coding System - Application to Broadcasting over an Error-Prone Channel", *15th European Signal Processing Conference*, Poland, Sept. 2007.

[10] C. Yaacoub, J. Farah, and B. Pesquet-Popescu, "Joint Source-Channel Wyner-Ziv Coding in Wireless Video Sensor Networks", *7th IEEE International Symposium on Signal Processing and Information Technology*, Egypt, Dec. 2007.

[11] D. Divsalar and F. Pollara, "Multiple Turbo Codes", Military Communications Conference, pp. 279-285, 1995.

[12] S.-Y. Chung, "On the Construction of Some Capacity-Approaching Coding Schemes", Ph.D. thesis, Massachusetts Institute of Technology, 2000.