International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011
ISSN: 1793-8201

Region-Based Adaptive Distributed Video Coding

A. Elamin, Varun Jeoti and Samir Belhouari

***Abstract*—the practical application of DVC is referred to as Wyner-Ziv video coding (WZ) where the side information available at the decoder is an estimate of the original frame. The compression is achieved by sending the information that is needed to correct this estimation. Due to the limitation of side information prediction, the predicted frame is expected to have various degree of success along the predicted frame. In this work, we propose partitioning the considered frame into many coding units (region) where each unit is encoded differently. This partitioning is, however, done at the decoder while generating the side-information and the region map is sent over to encoder at very little rate penalty. The partitioning allows allocation of appropriate DVC coding parameters (virtual channel, rate, and quantizer) to each region. The proposed solution preserves the simplicity of the video encoder by performing the frame partitioning process at the decoder during side information generation process. Experimental results show that adaptive DVC parameters selection per region enhances the overall PSNR when compared with frame based.**

***Index Terms*—distributed video coding, side information generation, channel coding, and Region based video coding.**

I. Introduction

Today's digital video coding paradigm represented by the ITU-T and MPEG standards mainly relies on a hybrid of block-based transform and inter-frame predictive coding approaches. In this coding framework, the encoder exploits both the temporal and spatial redundancies present in the video sequence, which is a complex process and it requires a noticeable amount of resources (power and memory). As a result, all standard video encoders have much higher computational complexity than the decoder (typically five to ten times more complex) [1], mainly due to the temporal correlation exploitation tools, especially due to the motion estimation task. Recently new emerging applications such as wireless low-power surveillance and multimedia sensor networks, wireless PC cameras and mobile camera phones, might not afford the encoding complexity. But still require attaining the coding efficiency gained by the motion estimation /compensation algorithm. As a result, the traditional video coding is no longer applicable for these applications. Appropriate video coding paradigm for these applications must have low encoding complexity. Lower encoding complexity could be achieved by moving some of the encoder tasks to the decoder part, particularly the most complex motion estimation process. Two results from information theory paved the way for new video coding known as Distributed Video Coding (DVC) that allows low encoding complexity and approach the efficiency of traditional video coding schemes. These two results known as Slepian-Wolf and Wyner-Ziv [2, 3] suggest that separate encoding - joint decoding system can approach the efficiency of joint encoding-decoding system. In the DVC, the individual frames are encoded independently (intraframe encoding) but decoded conditionally (interframe decoding). The practical application of DVC [4-9] is referred to as Wyner-Ziv video coding (WZ) where an estimate of the original frame "side information" is available at the decoder. The compression is achieved by sending the information that needed to correct this estimation. An error correcting code is used with the assumption that the estimate is a noisy version of the original frame and the rate is the needed amount of the parity bits. A virtual channel is assumed to represent the estimation noise [10]. This paper is organized as follows. First, in section 2, the pixel domain Wyner-Ziv video codec architecture and the most important modules are presented. In section 3, the issues relevant to this architecture are introduced and the suggestions in the extant literature on how to cope with these issues are reviewed. Section 4 introduces the proposed solution to tackle the issues and the gap in the literature. In section 5, the methodology used in evaluating the proposed solution is presented. Section 6, shows the results of several experiments performed in order to evaluate and compare the coding efficiency of the proposed approach and, finally, in section 7, the conclusions and future work.

Manuscript received July 16, 2010

A. Elamin. is with the Electrical Eng, Uinevrsiti Technologi PETRONAS, Perak,31750, Malaysia (phone: 60-175-773599; e-mail: elamina86@yahoo.com).

Varun Jeoti, is with the Electrical Eng, Uinevrsiti Technologi PETRONAS, Perak, 31750, Malaysia. (e-mail: varun_jeoti@petronas.com.my).

Samir Belhouari is with the Electrical Eng, Uinevrsiti Technologi PETRONAS, Perak,31750, Malaysia, (e-mail: brahim_belhaouari @ petronas.com.my).

II. Background, Design's Issues and Related Literature

The stat-of-the-art Pixel Domain Wyner-Ziv (PDWZ) video codec, presented in Figure 1, is based on the popular pixel domain Wyner-Ziv coding architecture proposed in [7]. In a nutshell, the overall coding process is as follows: the video frames are organized into key frames and Wyner-Ziv frames (WZ). Henceforth, this system will be referred to as "frame based" in this paper as it works on frame level. The key frames (KF) are encoded with a conventional intraframe codec with a quality similar to the quality of the WZ frames. Wyner-Ziv frames are encoded pixel by pixel where the

ISSN: 1793-8201

pixels are quantized using a 2M-level uniform scalar quantizer, generating the quantized symbol stream. Over the resulting quantized symbol stream bitplane, extraction is performed and each bitplane is then independently turbo encoded. The turbo encoder encloses two recursive systematic convolutional (RSC) encoders of rate ½ and a pseudo-random interleaver. Each RSC encoder outputs the parity stream and the systematic stream. After turbo encoding a bitplane, the systematic part is discarded and the parity bits are stored in the buffer and transmitted in small amounts upon decoder's request via the feedback channel. At the decoder, the frame interpolation module is used to generate the side information (SI) frame, an estimate of the WZ frame, based on previously decoded frames, X_B and X_F. For a Group of Pictures (GOP) 2, X_B and X_F are the previous and the next temporally adjacent key frames. In longer GOP lengths, such as 4 and 8; the SI frame is generated using both previously decoded key frames and WZ frames, according to the frame interpolation motion estimation/compensation method [1]. Thus, in case the GOP length is 4, the SI frame corresponding to the WZ frame WZ_2 is interpolated from the key frames KF_0 and KF_4, and so on. A similar frame interpolation structure is used for longer GOP lengths. The side information is then used by an iterative turbo decoder to obtain the decoded quantized symbol stream. The turbo decoder is constituted by two soft-input soft-output (SISO) decoders; each SISO decoder is implemented using the Maximum A Posteriori (MAP) algorithm. It is assumed the decoder has ideal error detection capabilities, i.e. the turbo decoder is able to measure in a perfect way the current bitplane error probability P_e. For example, if $P_e > 10^{-3}$, the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the bitplane turbo decoding task is considered successful. The side information is also used in the reconstruction module, together with the decoded quantized symbol stream, to help in the WZ frame reconstruction task. In this section, we review some issues that the frame based system presented in Figure 1 is still plagued with, namely, the accuracy of the correlation noise model, the poor rate control and the non-adaptive quantization.

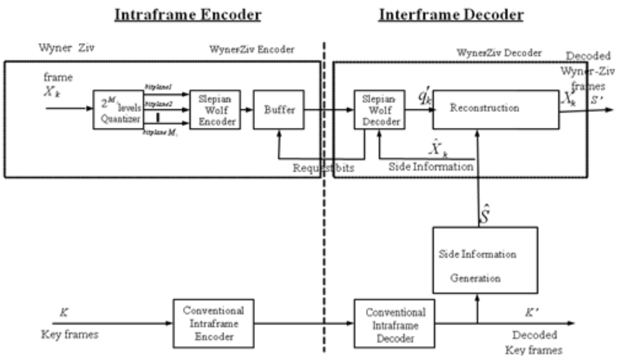

Figure 1 State-of-the-art Pixel Domain Wyner-Ziv video coding [7]

Issue 1 – Inaccurate Correlation Noise Model: Although more sophisticated ME algorithms can be used to generate the side information, still the practical frame prediction is fundamentally faulty due to events like occlusions. Due to the limitation of side information prediction, the predicted frame is expected to have varying degrees of success along the predicted frame [10, 17, 29, 30]. These limitations result in location-specific non-stationary estimation noise, for example, when occlusion occurs. Although this last property is a problem but it can be accommodated within a SW codecwithout extra information about the location-specificity at the cost of a rate penalty. Distributed source coding relies heavily on efficient error correcting codes. The performance of these codes depends greatly on the choice of the noise model that characterizes the virtual dependency channel [10, 31]. The lack of information about the unpredictable occlusion noise can be defeated by the use of interleaver in the Turbo based systems because interleaving eliminates the location specificity of the noise at the cost of some rate penalty. As a result of the spatial variation in the side information, the noise process for the entire frame is represented by a group of uniform PDF models. This group is best described by a mixture of Gaussian or Laplacian to characterize the non-stationary nature of the virtual dependency channel. The hypothesis that the mixture models fit the dependency channel was validated by using the Kolmogorov-Smirnov goodness-of-fit test. Goodness-of-fit tests are used to examine hypothesis that a given data set comes from a model distribution with given parameters. The Kolmogorov-Smirnov (KS) test is a popular goodness-of-fit test. The KS statistic test measures the distance between the empirical CDF and the model CDF which measures the goodness-of-fit. If the empirical CDF is tested against several model CDFs, the model that gives the minimum KS statistic can be taken to be the best fit for the data. Kolmogorov-Smirnov values were obtained as shown in table1, for different video sequences as shown in table 1. For a typical significance level of 5%, those values imply accepting the hypothesis that it follows a Laplacian mixture model. The minimum statistics and the best fit distribution for dependency channel have been indicated in bold.

TABLE 1 KS GOODNESS-OF-FIT TEST

Video sequence	Laplacian	Gaussian	Lap Mixture	Gauss Mixture
Carphone	0.3533	0.5423	**0.0248**	0.0477
Miss America	0.2504	0.7409	**0.0439**	0.0644
Foreman	0.3011	0.6840	**0.0348**	0.0641
Hall Monitor	0.3435	0.5856	**0.0335**	0.0428
Coast Guard	0.3257	0.5645	**0.0218**	0.0427
Akyio	0.3350	0.6340	**0.0391**	0.0656
Mother & daughter	0.3068	0.7530	**0.0506**	0.0655

Issue 2 – Non-adaptive Quantization: The current DVC solutions [5, 7, 14, 23, 32] work on frame level and assume that the quality of the side information is constant across the whole frame. Therefore, it uses one quantizer for the entire frame. Considering the varying motion characteristics and the amount of detail, the estimation of the original frame at the decoder is not the same for all regions of the frame. Based on this fact and utilizing the reconstruction function, one can conclude that the successfully estimated regions with coarser quantizer can achieve similar distortion reduction to the poorly estimated regions with finer quantizer. Therefore to achieve a certain operational characteristics in terms of distortion reduction, the quantizer must consider the quality of the side information. The intuition is that, in reconstructing the coded region, the joint decoder should rely more on side information when the rate is too low. On the other hand, when the rate is high, the decoded quantized symbol of the region becomes morereliable than side information, so the decoder should put more weight on the quantized symbol.

Issue 3 – Poor Rate Control: The feedback channel (FC) plays the role of adapting the bit rate to the changing statistics between the side information and the frame to be encoded. The frame based DVC solutions assume that this changing statistics is constant across the whole frame. Each pixel of WZ frame is scanned row-by-row and then uniformly quantized. The quantized symbols form an input block which is binarized and fed into turbo encoder for coding and transmission. Therefore, the size of input block is assumed as arbitrarily large because it represents all bits from the whole frame or one bit-plane at once [5, 7]. Too large input block will produce significant computation latency during the encoding and decoding process [33]. The system will not be able to provide a timely WZ decoded output due to the vast Slepian-Wolf turbo coding and decoding delay [33]. Frame based DVC systems make use of a rate-compatible-punctured codes (RCPT) codec as the module of the SW codec. The encoder "blindly" sends the parity bits according to the puncturing pattern determined by next lower coding rate. This causes the waste of transmission that some unnecessary parity bits are sent to help decoding some bits which are already correctly decoded [6].

Several works has been proposed for frame-based in order to cope with above drawbacks. In [29], in order to exploit the spatial variability of the correlation noise, the models' parameters are calculated at the block level. The considered block size is equal to the one used in the frame interpolation process (8-by- 8) in order to match the frame interpolation errors more easily. The resulting RD performance is better than that where the noise channel modeling is frame level based. Non-stationary model to characterize the correlation noise is proposed in [30], where two models are used, one to model the non-occluded regions and the other to model the occluded regions. The results show that the performance of the SISO decoder, and therefore of the overall system, can be improved greatly by classifying the decoder-generated side-information into two (or more) reliability classes. It also shows that the channel model should be an accurate representation of the real behavior of the channel; otherwise the decoding performance will heavily degrade. The lengthy block problem has also been addressed in [33, 34] by dividing the WZ frame, *X*, into *M* sub-images X_m , $m = \{0,1,2,3,\ldots..M\}$, and each sub-image is independently encoded using a Turbo-code based Wyner-Ziv encoder. Therefore, the block size of the Turbo encoder decreases to 1/*M* of the frame size. In the proposed system [34], Block-Adaptive wyner-Ziv coding (BAWZC)-based DVC system, the puncturing rate is independently adjusted for each block and the BAWZC decoder has to perform to inform the WZC encoder which blocks need more parity bits. The system proposed in [33], however, restricts the length of SW input to reduce the decoding complexity and the resulting delay. Both works [33, 34] alleviate the problem of spreading the estimation errors all over the bitstream by keeping the localized errors close to each other in the generated symbol stream. The results from both works show that system performance is further enhanced. These works enhance the rate control performance marginally as they use the same correlation noise model to characterize the noise over all sub-images. More enhancements would, however, be possible if more accurate noise model is used.

III. PROPOSED SOLUTION

In the light of the traditional variable-size block video coding, where the frame to be encoded is partitioned into many coding units, each unit has different DVC parameters. This partitioning process coincides with the spatial variation in the local characteristics of the video sequence, and performed at the decoder during the side information generation to preserve the encoder low complexity. The partitioning map compressed and communicated to the encoder via the feedback channel. So the proposed solution does not add new process and it only utilizes the existing components. Although there is rate penalty to communicate the compressed partitions map but the overall system performance is expected to be improved. The region based improves the dependency channel model accuracy by incorporating location-specific noise model for each region. Also with region format it is possible to have tight control on the rate control process; since by applying region format even the encoder "blindly" sends the parity bits; the locations of decoding errors are limited into one region rather than a whole bit-plane. The region based also allows applying appropriate source coding for each region targeting a certain operational RD point and adaptively allocate rate based on the local video characteristics. Typically the frame partitioning is a process of classifying the side information into several reliability classes. Since the estimation of the DVC parameters at finer level enhances the overall performance, the DVC parameters are estimated per region. The scheme of partitioning frame into regions can be seen as allocating different channel and rate for each region, which conforms to the unequal error protection concept. The simplest method of unequal error protection is to allocate different channels for different types of data [19].

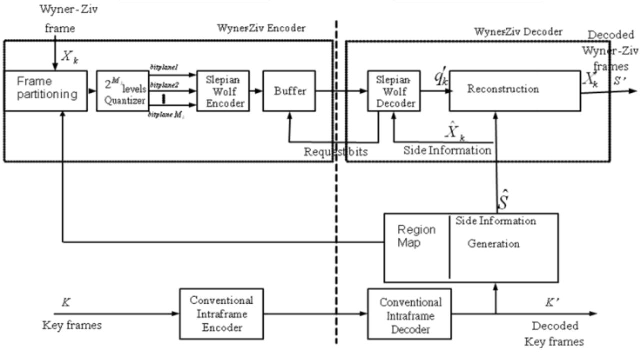

Figure 2: Proposed solution architecture

The frame partitioning is meant to classify the estimated side information into several reliability classes, since the side information is unavailable at the encoder, this task cannot be performed at the encoder if its complexity is to be kept low. The alternative partitioning technique makes use of some estimated reliability information regarding the estimated motion field in the SI generation process which allows quantifying the confidence on the side information estimation. It is common to use the two neighboring motion-compensated key frames to online estimate the dependency model [11]. The proposed frame partitioning is alternatively performed at the decoder by using the two neighboring motion-compensated key frames at the decoder.

$$\hat{X}_B(i,j) = \Gamma_B[X_B(i,j)] \quad (1)$$

$$\hat{X}_F(i,j) = \Gamma_F[X_F(i,j)] \quad (2)$$

where Γ_B and Γ_F represent the backward and the forward motion compensation operators respectively and *(i, j)* corresponds to the pixel location. The regions are constructed in two phases: an initial segmentation phase and iterative region merging phase. In the first phase the two motion-compensated key frames ($\hat{X}_F, \hat{X}_B$) are divided into smaller blocks (block size 8). The residual between each block and the co-located block in the two motion compensated frames ($\hat{X}_F(i,j), \hat{X}_B(i,j)$) is computed as follows

$$z = \sum_{j=1}^{J}\sum_{i=1}^{I}(\hat{X}_F(i,j) - \hat{X}_B(i,j)) \quad (3)$$

where z is the residual value and (i,j) denotes the location of pixel in the considered block.

In the second phase we iteratively merge blocks, this operation can be considered as a simple content segmentation method, with which the part of the frame with residual lower than certain threshold can be grouped into a region. The range of the blocks residual values, starting from the minimum residual value to the maximum residual value, is divided partitions. Each partition contains blocks with similar residual values, those block grouped together to compose a region. Partitioning to the range of the blocks residual values is performed based on MSE (σ_{total}) in such a way that, the sum of all partitions variances is minimized. This is corresponding to the minimum variance of each partition, which can be obtained if the partition elements are close and similar. The thresholds values $\{M_1, M_2, \ldots . M_L\}$ (partitions limits) and the mean values $\{\mu_1, \mu_2, \ldots . \mu_L\}$ of each partition.

$$\sigma_{total} = \int_{-\infty}^{M_1}(x-\mu_1)^2 f(x)dx + \int_{M_1}^{M_2}(x-2M_1+\mu_1)^2 f(x)dx + \cdots \quad (4)$$

$$\sigma_{opt} = \underset{(M_1, M_2, \ldots, M_L)}{\operatorname{argmin}} \sigma_{total}(M_1, M_2, \ldots . M_L) \quad (5)$$

Solving (5) gives the values of $\{M_1, M_2, \ldots . M_L\}$.

After assigning the side information blocks into L regions $\{Y_1, Y_2, \ldots . Y_L\}$ the decoder transmits the regions map to the encoder to partition the original WZ frame into corresponding regions. Straight forward coding of the region map is not very efficient. Furthermore, permutations of the region numbering do not affect the decoding process. We therefore consider a trans-formation of the region map into quadtree. The use of quadtree is ubiquitous in image coding [20-22] because of the simplicity and efficiency in coding the block configuration, requiring only 1 bit per node of the tree.

IV. Region Based-DVC coding parameters estimation

The region format allows allocating reliable DVC parameters that take into account the local characteristics. In this section the DVC coding parameters (dependency channel paramettter, the quantizer and the rate) estimation per region is preseneted.

A. Dependency channel

The residual statistics between correspondent pixels in $\hat{X}_l$ and Y_l is described by a Laplacian distribution, as in [5, 7]. The Laplacian parameter is estimated online by using residual statistics between co-located regions in the two motion-compensated key frames region's as follows.

$$Z_l(i,j) = \frac{\hat{X}_{Fl}(i,j) - \hat{X}_{Bl}(i,j)}{2} \quad (6)$$

$$\sigma_l^2 = E_Z\left[Z_l(i,j)^2\right] - \left(E_Z\left[Z_l(i,j)\right]\right)^2 \quad (7)$$

where σ_Z^2 is a confidence measure of the side information frame creation process which expresses how good the frame interpolation outcome is and $E_Z[Z(i,j)]$ is the expectation operation over the residual region. The Laplacian distribution parameter α_l defined by

$$\alpha_l = \sqrt{\frac{2}{\sigma_l^2}} \quad (8)$$

B. Source coding selection

The rate-distortion characteristics of each region are computed at the decoder based on a statistical model that is used to estimate the correlation between the original source information and the side information. Each region is quantized adaptively, i.e, for each WZ frame a set of scalar quantizer, based on the region rate-distortion characteristics. The quantizer set selection is to reduce the overall distortion to maximize the quality (PSNR) while adhering to rate constraints. To select the best quantizer set for each WZ frame first the distortion reduction and the corresponding rate for each region with respect to each quantizer is computed. The quantizer set that gives the most distortion reduction ΔD_{WZ} and overall rate less than R_{WZ} is:

$$\Delta D = \max_{(Q_1, Q_2, \ldots Q_L)} \Delta D_{WZ}(Q_1, Q_2, \ldots . Q_L) \quad (9)$$

Subject to $R \geq R_{WZ}$

One can also define the over all bitrate and the proposed system to select the best quantizer that gives the larger distortion reduction while maintaining the rate constraints for each region. The decoder after assigning appropriate quantizer to each region transmits the set of quantizers to the encoder through the feedback channel.

The bitrate for each quantized pixel value is the numbers of bits that are needed to Turbo decode the quantized pixel. Since each bit plane is separately decoded, the bitrate estimation is performed at the bitplane level. Let us assume

Wyner-Ziv region to be encoded is denoted by X_l , and the generated side information is denoted by Y_l. Also assume that the most significant $k-1$ bits of the pixel value $x \in X_l$ have already been error free decoded. Note that the encoder and the decoder know from already decoded k-1 bits, $\{b_1, b_2, \ldots. b_n\}$, that x is in the interval $[X_L \;\; X_U]$, where X_L and X_U are lower and upper limits to the x. For example, when $k = 1$, $X_L = 0$ and $X_U = 255$. At the encoder, the bit value reduces the interval $[X_c \;\; X_U]$ in such way that $x \in [X_L \;\; X_c]$ if $b_k = 0$ and $x \in [X_c + 1 \;\; X_U]$ if $b_k = 1$ where $X_c = \frac{X_L + X_U}{2}$ An error in b_k occurs, when pixel, x, lies within X_L and X_c i.e., $x \in [X_L \;\; X_c]$ and the corresponding estimated pixel value , y, lies within X_c and X_U , i.e., $x \in [X_c + 1 \;\; X_U]$. Or, similarly, an error occurs when $x \in [X_c + 1 \;\; X_U]$ and the corresponding estimated pixel $y \in [X_L \;\; X_c]$. Assuming that the difference between the original source information and the side information is Laplacian distributed, the conditional pdf of y given x and $X_L \leq y \leq X_U$ is

The parameters α can is estimated as in (7), the original source information is modeled as random variable (X'),

$$p(y|x, X_L \leq y \leq X_U) = \begin{cases} \dfrac{\frac{\alpha e^{-\alpha|x-y|}}{2}}{p(x, X_L \leq x \leq X_U)} & if \quad X_L \leq x \leq X_U \\ 0 & otherwise \end{cases} \tag{10}$$

with Laplacian distribution similar to the one for Z since Y is non-random

$$X' = Y + Z \tag{11}$$

From (9), the error probability of bit value b_k of pixel x is estimated using

$$P_e(x_k) = \begin{cases} \int_{x_c+0.5}^{x_U} p(y \mid x_L \leq y \leq x_U) dy \\ \\ \int_{x_L}^{x_c+0.5_U} p(y \mid x_L \leq y \leq x_U) dy \end{cases} \tag{12}$$

The integration intervals are expanded by 0.5 in order to cover the whole interval $[x_L \;\; x_U]$. Finally the average error probability, P_k for the bitplane k is estimated using

$$P_k = \sum_{x=0}^{x=255} H(x) P_e(b_k) \tag{13}$$

Where $H(x)$ is the histogram of the region, which provides the relative frequency of occurrence for each pixel value x. The rate corresponding rate is obtained from

$$R_k = -P_k.\log(P_k) - (1-P_k).\log(1-P_k) \tag{14}$$

The overall bitrate for the entire region R_l is obtained from

$$R_l = \sum_{k=1}^{K} R_k \tag{15}$$

where K is number of bitplanes. In other words, K represents the bit-depth of the scalar quantizer that is used to quantize of the region pixels. The overall bitrate for the entire frame R_{WZ} is sum of rates for each region, i.e,

$$R_{WZ} = \sum_{l=1}^{L} R_l \tag{16}$$

The reduction in the distortion is the difference between the initial distortion and the reconstruction distortion. The initial distortion is distortion between the estimated pixel and the original source pixel. This distortion corresponds to bitrate zero. For region with $(I \times J)$ pixel this distortion is obtained from

$$D_l = \sum_{i=1}^{I} \sum_{j=1}^{J} (X_l'(i,j) - Y_l(i,j)) \tag{17}$$

To compute the reduction in the distortion we must first compute the distortion between the reconstructed pixel when a R_l bitrate is transmitted and the original source pixel. The value of R_l is estimated as shown in (14). The reconstructed region's pixel value $\hat{X}_l$ when bitrate R_l is used can be estimated by using the reconstruction method as described by (19). Where $\hat{X}$ is computed by

$$\hat{X} = \begin{cases} \left\lfloor \frac{Y}{Q} \right\rfloor .Q + Q & if \left\lfloor \frac{X'}{Q} \right\rfloor > \left\lfloor \frac{Y}{Q} \right\rfloor \\ Y & if \left\lfloor \frac{X'}{Q} \right\rfloor = \left\lfloor \frac{Y}{Q} \right\rfloor \\ \left\lfloor \frac{Y}{Q} \right\rfloor .Q & if \left\lfloor \frac{X'}{Q} \right\rfloor < \left\lfloor \frac{Y}{Q} \right\rfloor \end{cases} \tag{18}$$

Here Q is quantization step, and is obtained using $Q = \dfrac{255}{2^{K-1}}$ where K is quantizer bit-depth.

The estimated distortion D_l' can then be computed as in (19)

$$D_l' = \sum_{i=1}^{I} \sum_{j=1}^{J} X_l'(i,j) - \hat{X}_l(i,j) \tag{19}$$

The distortion reduction now is computed by

$$\Delta D_l = D_l - D_l' \tag{20}$$

The over distortion reduction is obtained by

$$\Delta D_{WZ} = \sum_{l=1}^{L} \Delta D_l \tag{21}$$

V. Region-Based Performance Evaluation Methodology

In this study a MATLAB simulation for the proposed DVC scheme is developed. Video sequences in the uncompressed YUV/QCIF format (144X176) are used as simulation materials. These video sequences contain different numbers of frames but for the simulation only the first 101 frames are considered. The temporal resolution (number of frames per second) is fixed at 30 frame/second during this simulation for all materials. The performance of the proposed solution is tested only by using the luminance data for these video sequences. The group of picture (GOP) is chosen to be 2 during the simulation in this group and the odd frame is conventionally encoded and the even frame is encoded by the proposed solutions. The number of regions for each WZ frame (even frame) is predetermined to vary between 4 to 5 regions based on each video sequence characteristics. Turbo codes simulation is borrowed from the Code Modulation Library (CML) with two parallel recursive systematic encoders (RSC). The numbers of iterations run by Turbo decoder is fixed at 20 iterations. The rate-distortion curves contain the rate and the PSNR values for the even frames of the given sequence are used for RD performance evaluation. For comparison the proposed region-based DVC solution is impelemented with a fixed quantizer per frame (for all regions). Frame based DVC slotuion is also implemented to compare the proposed sloution performance. MPEG 2 (I-P-I-P) as conventional video coding is used for comparison to evaluate the proposed system.

VI. Simulation Results

The frame partitioning is performed for the two video sequences and the figure 3 shows the result of partitioning the frames into regions. As expected the partitioning groups the static (Background area) blocks into one region and the foreground blocks are partitioned into a few regions since the amount of motion is different and hence the reliability of the side information for these blocks is expected to vary from one block to another. The number of regions is predetermined to be 5 for "News" and "Mother-Daughter" video sequences based on observation that there are three objects that move along the sequence. Different part of the human body can be seen as different object if their motion does not depend on each other. Table 1 shows the feedback channel bitrates of the region-based. The regions map are compressed and sent through the feedback channel to the encoder, with bitrate less than 3% of the overall bitrates. High motion (carphone) sequence has larger feedback channel bitrate because of the video contents.

TABLE 2 Feedback Channel Bitrate For Different Video Sequence

Video sequence	Feedback rate per 15 FPS ($144X176X15$) (kbps)
News	1.9
Mother-Daughter	1.85
Miss America	1.54

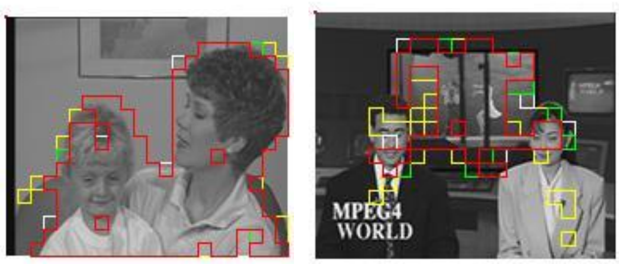

Figure 3 Partitioning result into 4 regions for Mother-Daughter and 5 regions for News sequence

Figure 4 shows the RD performance of the proposed technique, comparing with the state-of-the-art frame-based for "Mother-Daughter" sequence. It is observed that system performance is further enhanced by Region-Based adaptive selection to the DVC parameters. The PSNR gain can outperform up to 0.9 dB. The system improvement achieved by Region-Based with adaptive DVC parameters selection largely attributed to process of allocating higher rates to the poorly estimated regions and to the appropriate dependency channel model. Table 2 shows the average selected quantizer set corresponding to each rate point. It can be observed that there is still a large performance gap when compared to the MPEG+ inter-frame coding performance.

TABLE 3 Average Quantizer Set For Different Rate Points

Frame based [6]	Region Adaptive Quantizers			
	Region1	Region2	Region3	Region4
1	1	1	1	1
2	2	2	2	2
3	2	2	4	4
4	3	4	4	5

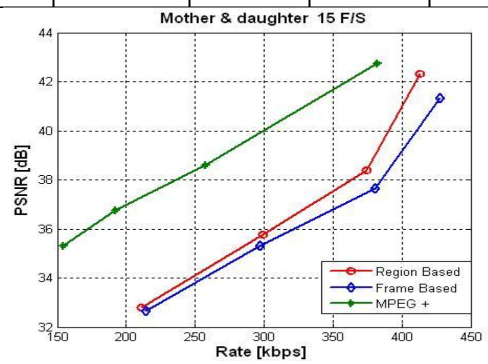

Figure 4 RD performance for the Mother-daughter test sequence Pixel domain

Figure 5 shows the RD performance of the proposed technique, comparing with the state-of-the-art frame-based for "News" sequence. It is observed that system performance is further enhanced by Region-Based adaptive DVC parameters selection. The PSNR gain can outperform up to 1.5 dB. The system improvement achieved by Region-Based adaptive DVC parameters selection is largely attributed to process of allocating higher rates to the poorly estimated regions. Table 3 shows the average selected quantizer set corresponding to each rate point. It can be observed that there is a performance gap when compared to the MPEG+ inter-frame coding performance.

TABLE 4 Average Quantizer Set for Different Rate Point

Frame-Based[6]	Region Adaptive Quantizers			
	Region 1	Region 2	Region 3	Region 4
1	1	1	1	1
2	1	2	2	3
3	2	3	3	4
4	3	4	4	5

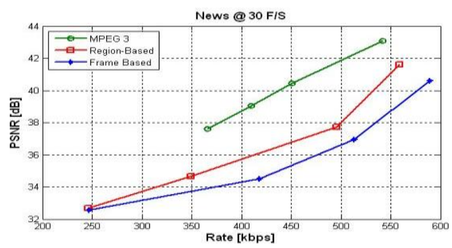

Figure 5 RD performance for the News test sequence Pixel domain

The results presented in this section show that the RD performance of region based adaptive DVC codec is considerably above the state-of-the-art frame-based DVC codec for all bitrates and all the test sequences. However, comparing RD performance of the region based adaptive DVC codec to the MPEG+ interframe coding with I-P-I-P structure, it can be seen that there is still a compression gap. This gap is more for video sequences characterized by high amount of motion. Since the partitioning process in the proposed scheme makes use of a-priori chosen 4 or 5 regions only, it cannot match the performance of MPEG+ schemes which adaptively use more number of partitions. The proposed scheme deliberately does not use more regions as it causes more bits from the region map to be transmitted back, resulting in rate penalty. It is also desirable that the number of bits from any given bit plane is large enough so the block codes (Turbo codes) used for generating the parity bits work efficiently. This requirement restricts the number of regions to be no more than 4 or 5.

The work reported herein suggests that a trade-off exists between the number of partitions employed and the size of codes used.

VII. Conclusion

This paper presented region-based adaptive DVC solution that takes into account the local characteristics of the video. The proposed DVC solution adheres to the encoder complexity constraints by performing most of the new proposed tasks at the decoder part. To avoid adding extra complexity to the encoder; the solution reutilizes the existing state-of-art components, such as side information generation and the feedback channel. It overcomes the problem of an accurate dependency channel model; by incorporating location-specific dependency channel model for each region. Region based enhanced the rate control process thus the parity bits are sent more accurately and adaptively allocates rate based on the local video characteristics. This allows applying appropriate source coding for each region targeting a certain operational RD point. As a result a finer quantizer is allocated to poorly estimated region. Experimental results show that this adaptive DVC parameters selection performed by using region based enhances the overall PSNR when compared with frame based approach.

Acknowledgment

The authors are grateful to Universiti Technology PETRONAS for funding this research. We also thank Ibraheem M. Dooba for the editing the paper.

References

[1] A. Aaron and B. Girod, "Wyner-Ziv video coding with low-encoder complexity "Invited Paper"," in Proc. Picture Coding Symposium San Francisco, CA, 2004.
[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions. Information.Theory, vol. 19, pp. 471-482, july 1973.
[3] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, vol. 22, pp. 1-10, 1976.
[4] R. Puri and K. Ramchandran, "PRISM: an uplink-friendly multimedia coding paradigm," in Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, 2003, pp. IV-856-9 vol.4.
[5] C. Brites, "Advances on distributed video coding," in Instituto Superior Técnico, . vol. MSc Lisbon: Technical University of Lisbon, 2005, p. 143.
[6] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video "Invited paper"," in Proc. Asilomar Conference on Signals and Systems Pacific Grove, CA: IEEE, 2002.
[7] A. Aaron and B. Girod, "Compression with side information using turbo codes," in Proc. IEEE Data Compression Conference, Snowbird, UT, 2002.
[8] R. Puri and K. Ranichandran, "Prism: A "Reversed" Multimedia Coding Paradigm," in Image Processing International Conference, Sardinia, Italy, 2003, pp. 617-20
[9] J. E. Fowler, "An Implementation of PRISM Using QccPack," Mississippi State University 2005.
[10] R. P. Westerlakena, R. K. Gunnewiekb, and R. Lagendijka, "The Role of the virtual channel in distributed source coding of video," in ICIP IEEE International Conference image processing: 2005.
[11] C. Brites and F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner-Ziv video coding," IEEE transactions on Circuit and systems for video technology, vol. 18, September 2008
[12] P. F. A. Meyer, R. P. Westerlaken, R. K. Gunnewiek, and a. R. L. Lagendijk, "Distributed source coding of video with non-stationary side-information," in Proc. Visual Communications and Image Processing 2005.
[13] P. F. A. Meyer, "Slepian-Wolf turbo coding of video with non-stationary side-information," in Faculty of engineering. vol. Master: Delf university of Technologoy, 2004, p. 28.
[14] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. I. L. Lagendijk, "Finding a near optimal dependency channel model for a ldpc-based wyner-ziv video compression scheme," in 12th annual conference of the Advanced School for Computing and Imaging, Lommel, Belgium, 2006.
[15] J. Ascenso, C. Brites, and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding," in IEEE International Conference on Advanced Video and Signal Based Surveillance Como,Italy:, 2005.
[16] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation," in Picture Coding Symposium Lisbon, Portugal, 2007.
[17] [Z. Xue, "Research and Developments of Distributed Video Coding," in School of Engineering and Design. vol. PhD Brunel: Brunel University 2009, p. 163.
[18] W.-J. Chien, L. J. Karam, and G. P. Abousleman, "Block-Adaptive wyner-Ziv coding for Transform-domain Distributed video coding," in 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.
[19] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal Error Protection: Some fundamental limits and optimal strategies," in Information theory , ISIT IEEE International Symposium, 2008, pp. 2222-2226
[20] J. Lee, "Optimal quadtree for variable block size motion estimation," in Proc. IEEE Int. Conf. Image Processing Washington, DC: 1995, pp. 480–483.
[21] Z. Jiajun, M. Omair Ahmad, and M. N. S. Swamy, "Quadtree structured region-wise motion compensation for video compression," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 9, pp. 808-822, 1999.
[22] P. Strobach, "Quadtree-structured recursive plane decomposition coding of images," IEEE Trans. Signal Processing, vol. 39, pp. 1380–1397, June 1991.

A. Elamin Phd Student. Sudan (July 1973). Bs.c Electrical engineering U of K (Sudan), MSc multimedia technology ITU of Copenhagen (Denmark), recent research in Distributed video coding.

AP.Dr Varun Jeoti, Telecomm cluster head here at UTP Malaysia. BS.c, MSc and PhD IIT (Indea)

Dr. Samir Belhaouri, .