

An Adaptive Slice Group Multiple Description Coding Technique for Real-time Video Transmission over Wireless Networks[†]

Viswesh Parameswaran, Sudheendra Murthy, Arunabha Sen and Baoxin Li
 Department of Computer Science and Engineering
 Arizona State University, Tempe, Arizona 85281
 Email: {vparames, sudhi, asen, baoxin.li}@asu.edu

Abstract—This paper addresses the problem of transmitting real-time video over wireless networks. We propose improvements to multiple description coding technique using the slice group coding tools provided in H.264/AVC. The Macroblocks in each frame are mapped into different slice groups based on their motion vectors and distortion parameters. The slice group containing significant motion and distortion is fine quantized, while the other slice group is coarse quantized. The encoded quality scaled video is transmitted normally over the first path, but with a quantization offset on the second path. The proposed system provides a better quality video for the regions containing interesting targets. Simulation results in Network Simulator NS-2 confirm robust performance of the scheme under different packet loss conditions.

Index Terms—Wireless networks, Multiple Description Coding (MDC), Single Description Coding (SDC), Flexible Macroblock Ordering (FMO)

I. INTRODUCTION

There are many applications, especially in the military domain, that can greatly benefit from the availability of real-time video captured by mobile imaging sensors. For example, one can envision equipping a group of field soldiers with networked and wearable computing devices including cameras and displays so that the soldiers can share live video with each other. The video source may also originate from other compact Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs). Since potentially the soldiers and the UAVs/UGAs have different field of view, sharing video among them can greatly enhance the tactical options available to the soldiers and thus help accomplishing

a mission with reduced casualty. In such and similar applications, we cannot assume the availability of an infrastructure network and thus, a key technical problem is the transmission of real-time video over the mobile ad hoc wireless network.

Recent developments in the fields of wireless communication and video compression have made it feasible to transmit video over wireless networks. However, significant challenges still exist since the wireless channel is highly error-prone and has limited bandwidth capacity. In parallel with the development on the network side, extensive research has been carried out to develop intelligent visual processing techniques, which can analyze video sequence and determine the threat targets automatically. The object of interest usually appears in the field of view only for a very short interval of time. An intelligent video coding system can potentially conserve bandwidth by selectively transmitting only the sequences containing the target object, resulting in appreciable bandwidth savings. This also reduces the workload on the operator since the number of sequences that he needs to go through, is dramatically reduced.

In this paper, we focus on developing a system for transmitting quality scaled video over mobile ad hoc wireless networks. The application scenario is illustrated in Figure 1.

The system contains multiple source nodes transmitting their individual field of view to a destination node over a wireless channel. Multiple intermediate nodes may be available between the source and destination at any point of time and thus, multiple paths could be formed from any source to the destination. Each of these individual paths is unreliable because of the following reasons (a) changes in physical characteristics of the channel due to fading and noise (b) mobility of the nodes resulting in destruction of existing paths and

[†] This material is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contract/grant number W911NF-06-1-0354.

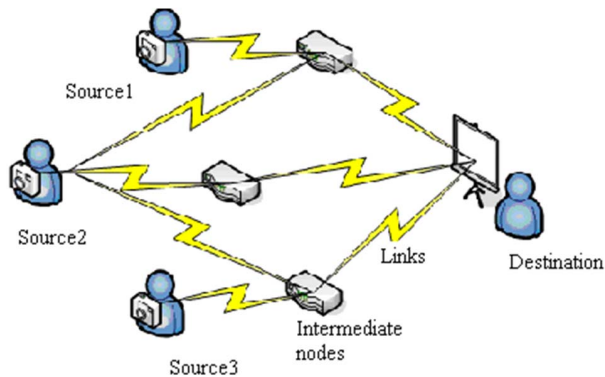


Fig. 1. A typical example of video over wireless network

creation of newer paths. We propose a scheme based on Multiple Description Coding (MDC) to address these issues. Specifically, the main contributions of this paper are as follows.

- Region of interest (ROI) classification based on the motion vector and block distortions
- Smoothed update of slice group-macroblock mapping vectors
- Utilization of quantization offset to achieve video of better quality in case both descriptions are received at the decoder

The rest of the paper is structured as follows. In Section II we provide a brief literature survey of the popular methods. The proposed system design is described in section III. The simulation results are presented in section IV with concluding remarks in section V.

II. RELATED WORK

In order to transmit video over a lossy wireless network, we need to use error resilient techniques at the encoder and error concealment at the decoder. The real time video is highly sensitive to delay, so the conventional retransmission techniques cannot be used for video transmission. One popular error resilient technique is to employ Forward Error Correcting (FEC) codes for channel coding. This approach was employed in [1] where Reed-Solomon codes were used for error recovery in wireless channel. The use of FEC is a practical solution in cases where all the nodes are stationary. But in case of mobile nodes, there could be outages because of nodes moving out of the source transmission range. In these cases we could encode the video using several independent descriptions and transmit them over multiple channels. This method termed as MDC achieves error resilience through path diversity. The theoretical analysis for MDC was carried out in [2].

The different descriptions in MDC are correlated so that they can be independently decoded at the receiver.

This means that there is a trade-off between the redundancy and the error resilience associated with the MDC scheme. One popular method for MDC is temporal splitting where the even frames are sent on one channel and the odd frames on the other. If only one description is received, then the lost frames can be reconstructed from the other description. The problem with this approach is the poor coding efficiency because of the increased temporal distance that degrades motion-based prediction coding. Another method is the spatial splitting technique in which even lines are sent on one path and odd lines on the other path. Some of these methods have been evaluated in [3]. The system proposed in [3] adaptively chooses the best mode depending upon end-to-end distortion. Another MDC method based on overlapping quantization [4] was proposed in [5] for MPEG-4 video. In this method the spatial and temporal smoothness properties of video are used for frame reconstruction in case of a lost description.

The video compression standards have been evolving continuously, reducing the required bit rates for good quality video transmissions. H.264/AVC is the latest video coding standard proposed by ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. In [6], salient features of this standard are covered. The Flexible Macroblock Ordering (FMO) tool and its applications have been described in [7]. In [8], the quantization levels are offset by a fixed value to achieve better performance. In [9], the macroblocks in the frame are classified on their distortion measures and an unequal error protection scheme based on Reed-Solomon codes is applied for single description coding. In [10], the frame is divided into blocks of high and low importance depending on the motion vector information and the slice groups are differentially quantized in the primary and the secondary paths. This method maintains satisfactory quality for the important regions in the video in case only one description is received at the decoder. But in cases where both descriptions are received at the decoder, the information from the secondary path is discarded. In our paper, we refine this approach for MDC based video transmission.

III. PROPOSED SYSTEM

The flow diagram of the proposed MDC encoder scheme is shown in Figure 2. The system consists of two symmetric H.264/AVC encoders processing the same source stream. The motion estimation output from the first encoder is used to determine the regions of interest in the frame. This information is used for subsequent coding by both encoders. The second encoder also contains a quantization offset block to offset the quantization

levels. The reference frame used for encoders on both ends is the same as the reference frames at the respective decoders. This is essential to avoid drift effects between the encoder and decoder systems.

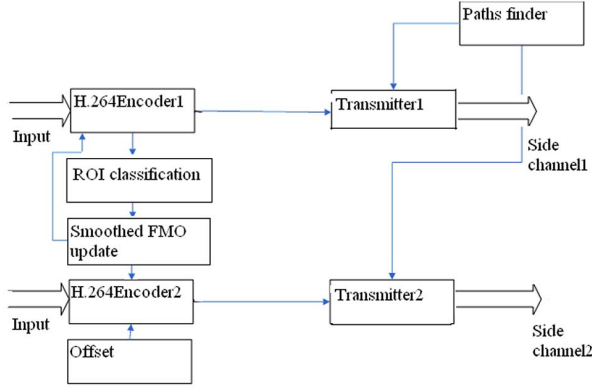


Fig. 2. Flow diagram of the proposed MDC encoder

We discuss these components in detail in the subsequent subsections.

A. ROI Classification

In H.264/AVC encoder, the macroblocks in each slice are predicted using either of the two methods (a) Intra prediction, where the macroblocks are predicted from previously encoded macroblocks in the current slice or (b) Motion compensated inter prediction, where the current macroblock is predicted from previously encoded slice in one or more reference frames. The difference between the current macroblock and the predicted macroblock value, termed as the residual error is transformed, quantized and transmitted over the channel. In this paper, we use the following measures to classify regions of interest (i) Motion vector information for inter-predicted macroblocks and (ii) Distortion measure for both intra-predicted and inter-predicted macroblocks.

In this context, we should note that the process of region classification and video encoding are inter-dependent and we cannot separate the two processes from each other. In case of temporally smooth video, the regions of interest do not change appreciably between consecutive frames. Under this assumption, it is possible to classify the regions of interest for the current frame based on the previous encoded frame measures. We follow this approach for our classification. Also, we use only the luminance component of the macroblock for region of interest partitioning. This is sufficient since the luminance information constitutes the more significant part of the video.

1) *Motion vector grouping*: H.264/AVC performs tree structured motion compensation for inter-prediction. The 16x16 macroblock is sub-partitioned into two 16x8 partitions, two 8x16 partitions or four 8x8 partitions depending on the residual error. The 8x8 partition is further sub-divided into two 8x4 sub-partitions, two 4x8 sub-partitions or four 4x4 sub-partitions. This means that for a given macroblock we can have multiple motion vectors. The first step in the classification is to determine the maximum values of the motion vectors for each macroblock.

$$MV_{max}(i, x) = \max_{l \text{ is a MB sub-partition index}} MV(i, l, x)$$

$$MV_{max}(i, y) = \max_{l \text{ is a MB sub-partition index}} MV(i, l, y) \quad (1)$$

In equation (1), $MV_{max}(i, x)$ and $MV_{max}(i, y)$ are the maximum motion vectors of the i th macroblock in x and y directions. The mean motion vector for all the inter-predicted macroblocks is determined using equation (2).

$$E[MV_x] = \frac{\sum_{i=1}^{N_{inter}} MV_{max}(i, x)}{N_{inter}}$$

$$E[MV_y] = \frac{\sum_{i=1}^{N_{inter}} MV_{max}(i, y)}{N_{inter}} \quad (2)$$

In the equation 2, N_{inter} is the total number of inter-predicted macroblocks in the frame. Once the mean motion vector is estimated, we group the macroblocks on the basis of motion significance. This is done by comparing the individual motion vector components against the mean motion vector as shown in equation (3).

$$MV_{ind}(i) = \begin{cases} 1, & \text{if } MV_{max}(i, x) > E[MV_x] \\ & \text{or } MV_{max}(i, y) > E[MV_y] \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

2) *Distortion measure grouping*: The quantizer distortion in H.264/AVC is dependent upon the selected block partitioning mode used for prediction. The mode selection algorithm selects the best possible mode so as to minimize the rate distortion cost in equation (4).

$$D(i, m) = \frac{\sum_{i=1}^n \sum_{j=1}^n |x(i, j) - q(i, j, m)|}{N \times N}$$

$$RDcost_{min}(i) = \min_{all\ modes} (D(i, m) + \lambda R) \quad (4)$$

In the above equations, $D(i, m)$ is the distortion of the i^{th} macroblock when encoded with a particular mode m , $N \times N$ is the size of the macroblock, R is the rate constraint if rate control is enabled and λ is the lagrangian penalty function. We use the estimated rate distortion costs for our classification. The mean distortion of all macroblocks is estimated by using equation (5).

$$E[D] = \frac{\sum_{i=1}^N RDcost_{min}(i)}{N} \quad (5)$$

Here, N is the total number of intra-predicted and inter-predicted macroblocks in the frame. In the next step all the macroblocks with rate distortion cost greater than mean distortion are classified as significant distortion regions.

$$D_{ind}(i) = \begin{cases} 1, & \text{if } RDcost_{min}(i) > E[D] \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

B. Smoothed FMO update

The motion vectors and quantizer distortion are estimated every frame and this could result in a fast changing region of interest. This is not a desired situation in many cases since we require a stable ROI between frames. This can be achieved by smoothing out the variations using a first order IIR filter before updating the slice group-macroblock mapping vectors. The input and the output equations for the filter are shown in equation (7).

$$x(i) = \begin{cases} 1, & \text{if } MV_{ind}(i) = 1 \text{ or } D_{ind}(i) = 1 \\ 0, & \text{otherwise} \end{cases}$$

$$y(i) = \begin{cases} \alpha_1 x(i) + (1 - \alpha_1)y(i - 1), & \text{if } x(i) = 1 \\ \alpha_2 y(i - 1), & \text{otherwise} \end{cases} \quad (7)$$

The input is set to 1 if the particular macroblock has significant motion or distortion. The output of the filter is a function of both input as well as the previous output sample. The weight coefficient α_1 determines how fast the system adapts to a change in ROI and α_2 determines how long it retains an already determined ROI. The final step in classification is to group the macroblocks into appropriate slice group by comparing the output against a threshold value. That is,

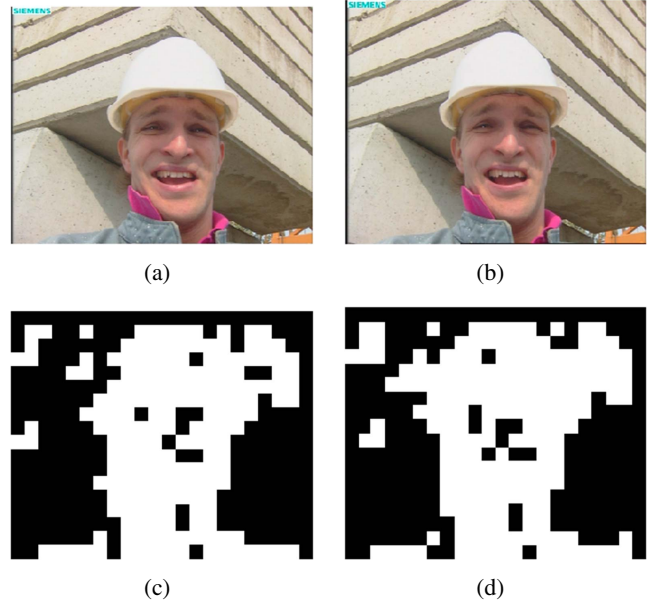


Fig. 3. Classification results for Foreman sequence (a) frame 20 (b) frame 22 (c) slice grouping result for frame 20 (d) slice grouping result for frame 22 (white - important slice group, black - rest of the scene)

$$i \in \begin{cases} 1, & \text{slice group 1 if } y(i) > y_{thresh} \\ 0, & \text{slice group 1 otherwise} \end{cases} \quad (8)$$

The classification results obtained for frames 20 and 22 of the Foreman sequence is shown in Figure 3.

C. Quantization offsetting

Once the macroblocks are classified as described earlier, we can relatively assign the quantization parameters for each slice group. The slice group containing important regions is fine quantized with a parameter Q_1 and the less important slice group is coarse quantized with a parameter Q_2 in both paths. The quantization levels in the second path are offset by a quantization offset value. This is done so that if we receive descriptions from both paths, we can reconstruct the video at a higher quality than in cases where only one description is received.

The quantization and the inverse quantization processes in H.264/AVC are described by equation (9).

$$Z_{i,j} = \lfloor \frac{|W_{i,j}|MF_{i,j} + f_{i,j}}{2^{15+\lfloor QP/6 \rfloor}} \rfloor \cdot \text{sgn}(W_{i,j})$$

$$W'_{i,j} = Z_{i,j} \cdot V_{i,j} \cdot 2^{\lfloor QP/6 \rfloor} \quad (9)$$

In equation (9), W is the original transform coefficient, Z is the quantized level and W' is the reconstructed transform coefficient. QP is the quantization parameter of the macroblock. A higher value for the quantization parameter implies a higher quantization step

size and reduced picture quality and viceversa. The factor $MF_{i,j}$ and $V_{i,j}$ are position dependent scaling factors and $f_{i,j}$ is the fixed shift factor taking into account the Laplacian distribution of the coefficients. In the proposed method, we add a varying offset at the time of quantization. The modified quantization process is given in equation (10).

$$Z_{i,j} = \lfloor \frac{|W_{i,j}|MF_{i,j} + \text{deltaqp}_{i,j} + f_{i,j}}{2^{15+\lfloor QP/6 \rfloor}} \rfloor \cdot \text{sgn}(W_{i,j})$$

$$\text{deltaqp}_{i,j} = \lfloor \frac{2^{\lfloor QP/6 \rfloor} - f_{i,j}}{2} \rfloor \quad (10)$$

The offset selection transforms the coefficients in the range $(Q^{-1}(Z_{i,j}) + \text{deltaqp}, Q^{-1}(Z_{i,j} + 1) - \text{deltaqp})$ to $Z_{i,j}$ in the first path and $(Z_{i,j} + 1)$ in the second path. This additional quantization level improves the performance of the system when both descriptions are received at the decoder. The normal decoding process can be carried out if we receive the description only from one path. In case we receive descriptions from both paths the mean of the corresponding pixel values is taken to reconstruct the final image. The decoder side processing is shown in Figure 4.

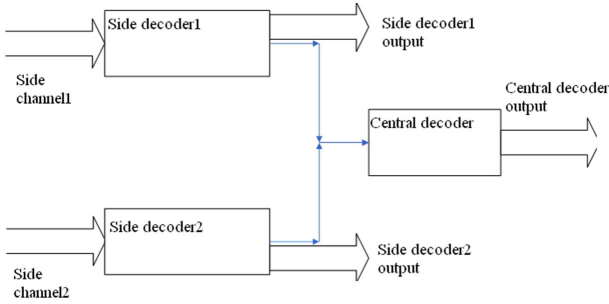


Fig. 4. Proposed decoder configuration

IV. EXPERIMENTAL RESULTS

The JM 12.2 version of H.264/AVC encoder was used for simulations. The encoder profile was set to baseline mode. The B-frame count was set to 0 so that no B-frames were inserted in the Group of pictures (GOP) sequence. The Context Adaptive Binary Arithmetic Coding (CABAC) and transform 8x8 modes were disabled. The rate distortion optimized mode selection was enabled. The frame skip counter was set to 0 to avoid skipping of frames. The frame rate of the video sequence was set to 25fps. The intra frame was forced every 300 frames. The encoders on both paths were symmetrically set except for the quantization offset on the second path. The slice mode was set in customizable mode for dynamic slice grouping. The picture parameter set

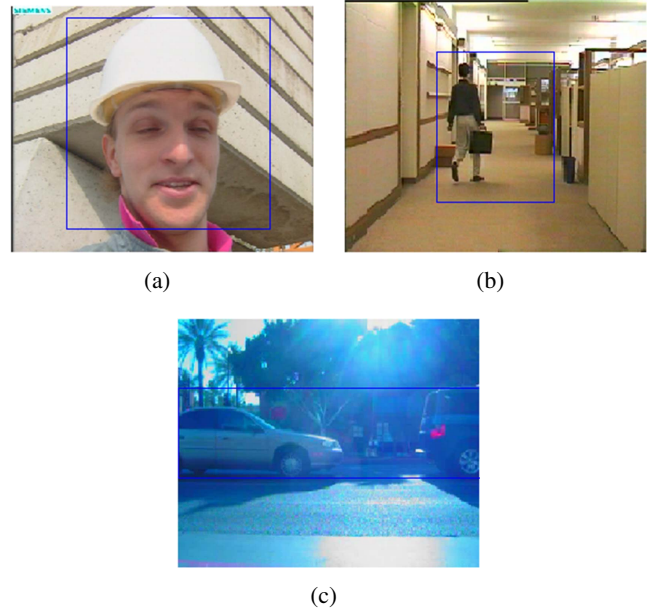


Fig. 5. Reference region of interest in (a) Foreman sequence (b) Hall video sequence (c) Real traffic video sequence

was transmitted every 15 frames from the encoder side. The quantization parameters for I-slices and P-slices in the default configuration were set to the same value. The important slice group was fine quantized with a quantization parameter $(QP - m)$ and the less important slice group was coarse quantized with a quantization parameter $(QP + n)$ in both paths.

The following video sequences were used for testing (a) Foreman sequence (b) Hall video sequence (c) Real traffic video sequence. The performance of the proposed system in all cases was compared against standard H.264 single slice group encoding. In addition to normal peak signal to noise ratio (PSNR), an objective performance measure, termed ROI PSNR, was defined to compare the performance around the region of interest. The ROI PSNR metric is defined as follows.

$$ROI\text{PSNR} = 10 \log \frac{255 \times 255}{\frac{1}{A_{ROI}} \sum_{(i,j) \in ROI} (c(i,j) - I(i,j))^2} \quad (11)$$

In the above equation, A_{ROI} is the area of region of interest, c is the reconstructed frame and I is the original frame. The region of interest for comparison was manually segmented for three sequences as shown in Figure 5. These regions undergo maximum motion in the respective video sequences.

The maximum size of the slice in both cases was fixed to 800 bytes. In case of single slice group H.264 encoding, each frame contains multiple slices, limited by the maximum size criteria. Each slice corresponds to one

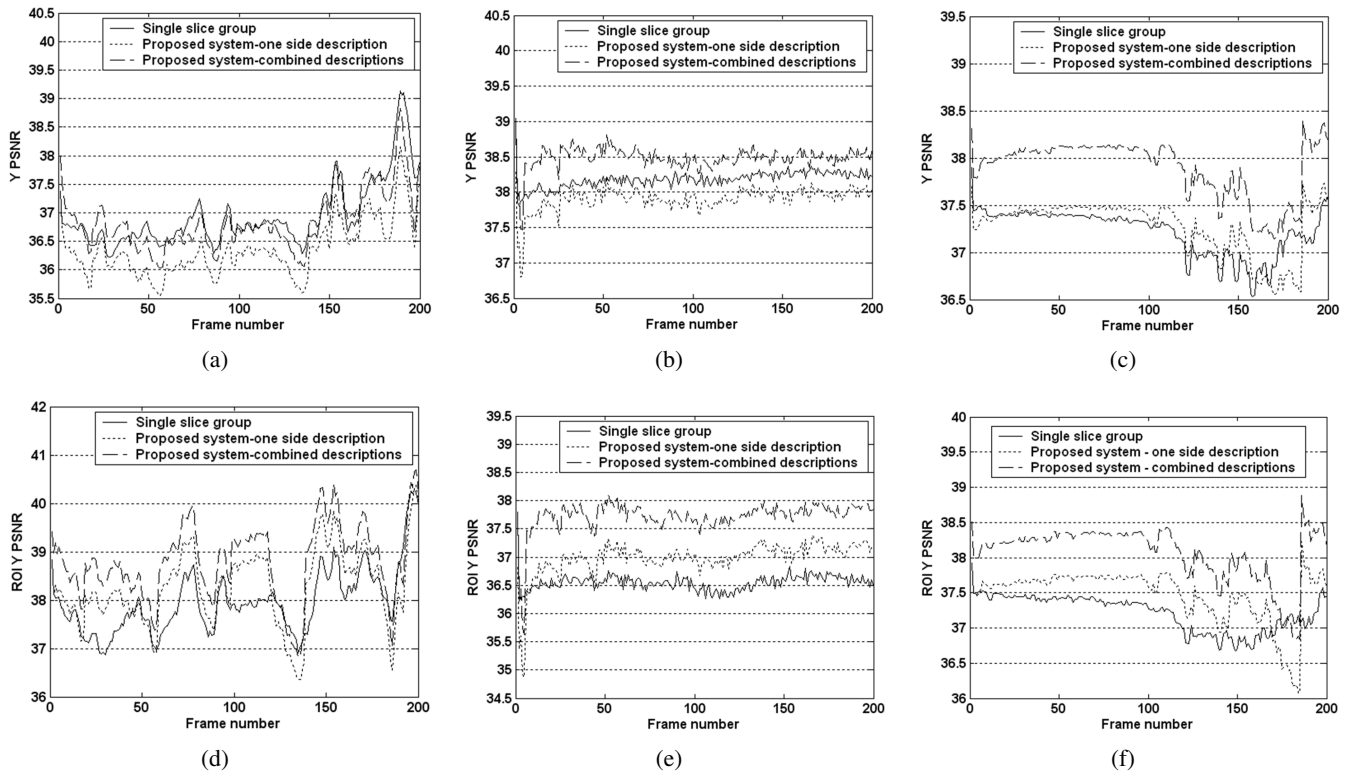


Fig. 6. Obtained graphs under lossless conditions (a) PSNR for foreman sequence (b) PSNR for hall sequence (c) PSNR for real video sequence (d) ROI PSNR for foreman sequence (e) ROI PSNR for hall sequence (f) ROI PSNR for real traffic video sequence

network transmission unit. In the proposed system, each frame consists of two slice groups; each slice group in turn could contain multiple slices. The PSNR and ROI PSNR measures were computed for both schemes.

Figure 6 shows the obtained measures for test video sequences under no loss condition. The plots (a)-(c) show the variation of PSNR for different frames of the video sequences. In all cases we see that the PSNR values for the proposed system with one description is slightly less than that of the single slice group case. This is expected because we are coarse quantizing the non-significant region so as to allocate more bits near the regions of interest. Also the use of FMO slightly reduces the coding efficiency and creates an additional overhead. The video quality is significantly improved in cases where we receive both descriptions. The plots (d)-(f) show the variation of ROI PSNR for the respective video sequences. In all cases we see that the ROI PSNR for the proposed system with one description is much higher than the single slice group case. The effect is very evident in case of hall and real traffic video sequences. In these sequences the moving objects occupy a much smaller region compared with rest of the scene. This results in better accuracy for the motion based segmentation algorithm. This is typically the scenario in surveillance video. The camera is generally located

far away from the target object. Hence the size of the object in the image will be much smaller than rest of the scene. Also the target object appears in the vicinity of the camera only for a very short duration of time. In the case of foreman sequence there are lots of moving regions, which reduces the performance benefit of using FMO. Also we see that the ROI PSNR quality is significantly improved by combining the descriptions from the two paths in all cases. The observed results conform to the expected notion that quality trade off is possible between the regions of interest and rest of the scene in surveillance systems.

The network was configured in NS-2 simulator as a grid of 25 nodes arranged in 5 rows and 5 columns. The distance between every pair of neighbors was set to 200 m. The transmission range was the default transmission range in NS-2, which is 250 m. In the grid, two node-disjoint paths from the source to the destination each having 5 hops were manually selected. Each 1-hop cross-traffic flow was generated by choosing a random source, a random destination. The rate of the cross-traffic flow was randomly generated from a uniform distribution between 0 and 100 kB. A frame copy error concealment scheme was employed at the decoder. Table I shows the results obtained averaged over 100 frames of the respective video sequences.

TABLE I
AVERAGE PSNR VALUES FOR DIFFERENT CROSS TRAFFIC FLOWS

Sequence	# Cross-traffic flows	Mean PSNR for system with single path	Mean PSNR for proposed system over two paths
Foreman	0	36.8	37.2
	10	33.2	35.6
	15	30.8	34.8
	20	29.6	33.5
Hall	0	37.8	38.5
	10	37.2	38.2
	15	36.5	37.0
	20	35.8	36.4

V. CONCLUSION

The proposed method provides significant benefits for multiple description video transmission over wireless channels. The characterization of frames into slice groups based on the motion and the distortion characteristics helps in relative quality adjustment for regions of interest. The quantization offset between the paths help in combining the information in case descriptors are received from both paths. The simulation results show improved performance of the system compared to the single slice group H.264/AVC encoding scheme.

REFERENCES

- [1] E. Ayanoglu, P. Pancha, A. Reibman, and S. Talwar, "Forward error control for mpeg-2 video transport in a wireless atm lan," in *International Conference on Image Processing 1996*, Sep. 1996.
- [2] A. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, Nov. 1982.
- [3] B. Heng, J. Apostolopoulos, and J. Lim, "End-to-end rate-distortion optimized md mode selection for multiple description video coding," *EURASIP Journal on Applied Signal Processing*, vol. vol.2006, 2006.
- [4] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, May. 1993.
- [5] Y.-C. Lee, Y. Altunbasak, and R. Mersereau, "Coordinated application of multiple description scalar quantization and error concealment for error-resilient mpeg video streaming," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 4, 2005.
- [6] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuit and System for Video Technology*, Jul. 2003.
- [7] Y. Dhondt and P. Lambert, "Flexible macroblock ordering: an error resilience tool in h.264/avc," in *Fifth FTW PhD Symposium, Faculty of Engineering, Ghent University*, no. 106, Dec. 2004.
- [8] T. Wedi and S. Wittmann, "Quantization offsets for video coding," in *ISCAS 2005. IEEE International Symposium*, May. 2005.
- [9] N. Thomos, S. Argyropoulos, N. Boulgouris, and M. Strintzis, "Robust transmission of h.264/avc video using adaptive slice grouping and unequal error protection," in *IEEE International Conference on Multimedia and Expo, 2006*, July. 2006.
- [10] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding using motion vector estimation," in *IEEE Int. Conf. Image Proc.*, Oct. 2004.