TRANSMISSION OF SCALABLE H.264 CODED VIDEO OVER WIRELESS MIMO SYSTEMS WITH OPTIMAL BANDWIDTH ALLOCATION

Mohammad K. Jubran, Manu Bansal, Lisimachos P. Kondi

Department of Electrical Engineering,
State University of New York at Buffalo,
Buffalo, NY 14260.
Email: {mkjubran, mbansal, lkondi}@eng.buffalo.edu

ABSTRACT

We propose an optimal strategy for the transmission of scalable video over multiple-input multiple-output (MIMO) wireless systems. In this paper, we use the latest scalable H.264 codec which provides combined temporal, quality and spatial scalability. At the transmitter, we have developed a method for the estimation of the video distortion at the receiver for given channel conditions. The accuracy of this method is validated using experimental results. In our proposed system, we use a MIMO system with orthogonal space-time block codes (O-STBC) that provides spatial diversity and guarantees independent transmission of different symbols within the block code. Rate-compatible punctured convolutional (RCPC) codes are used for unequal error protection of the scalable layers. In the constrained bandwidth allocation framework considered here, we use the estimated decoder distortion to optimally select the application layer parameter, i.e. quantization parameter (QP), and physical layer parameters, i.e. RCPC and modulation.

Index Terms— Scalable H.264 codec, Distortion estimation, Wireless video, MIMO systems, Optimal bandwidth allocation.

1. INTRODUCTION

The H.264/AVC standard [1], [2] and its recently proposed scalable extension [3], [4] have error-resilient network adaptation layer (NAL) structure and provide superior compression efficiency. The scalable H.264 provides a combined scalability in the form of temporal scalability, fine granular quality scalability (FGS) and spatial scalability. We consider the efficient transmission of temporal and quality scalable layers over packet-based wireless networks, with optimization of source coding, channel coding and physical layer parameters on a per-GOP basis. For that to be possible, a good knowledge of the total end-to-end decoder distortion should be available at the encoder. In [5], [6] and [7], a recursive per-pixel based decoder distortion estimation algorithm, ROPE was proposed for non-scalable and scalable H.263+ codec. In this paper, we develop a method for the accurate estimation of the distortion of scalable H.264/AVC coded video at the receiver for given channel conditions. Our scalable decoder distortion estimation (SDDE) algorithm takes into account loss of both temporal and SNR scalable layers as well as error concealment at the decoder.

Diversity techniques, such as space-time coding (STC) have been proven to help overcome the degradations due to wireless channels by providing the receiver with multiple replicas of the transmitted signal over different channels. Orthogonal space-time block codes (O-STBC), which were first proposed by Alamouti [8] and later gen-eralized by Tarokh *et. al.* [9] are used here for video transmission over the MIMO system. They guarantee independent transmission and low-complexity maximum-likelihood (ML) decoding for each symbol in the codeword. This enables us to independently choose the elements of the codeword from different constellations.

In only a few publications such as [10], [11], [12], wireless video transmission using STC has been studied. In [10], a joint source-channel matching framework for image transmission using the SPIHT encoder over an OFDM system with space-time block codes is proposed. Zhao *et. al.* in [11], proposed progressive video transmission over a space-time differentially coded OFDM system with optimal rate and power allocation among multiple layers. However, in all the above-mentioned work, the orthogonal structure of STBC codes has not been exploited by independent transmission of the layered video over different symbols of the STBC code modulated with different constellations. In [13], an approach for using the scalable H.264 with unequal erasure protection (UXP) over wireless IP networks has been proposed.

In this paper, we propose a system that integrates video coding with combined scalability, forward error correction (FEC) through unequal channel coding, modulation scheme selection and spatial diversity for wireless video transmission. Temporal and quality scalable layers are obtained using a scalable H.264 codec and are given unequal protection using rate-compatible punctured convolutional (RCPC) [14] codes with cyclic redundancy check (CRC) [15] error detection. The channel coded layers are then modulated and encoded using O-STBC for transmission over multiple antennas. The bandwidth constrained problem is addressed by minimizing the expected end-to-end distortion (estimated using the SDDE algorithm) and optimally selecting QP, RCPC rate and the symbol constellation for the MIMO transmission.

2. DECODER DISTORTION ESTIMATION FOR SCALABLE H.264 CODEC

The latest scalable extension of H.264/AVC is based on a hierarchical prediction structure as shown in Figure 1. A GOP consists of a key picture and all other pictures temporally located between the key picture and the previously encoded key picture. These key pictures are considered as the lowest temporal resolution of the video sequence and are called temporal level zero (TL0) and the other pictures encoded in each GOP define different temporal levels (TL1, TL2, so on). Each of these pictures is represented by a non-scalable base layer (FGS0) that includes the corresponding motion and an approximation of the intra and residual data, and zero or more quality scalable enhancement (FGS) layers. Also, the priority of the base
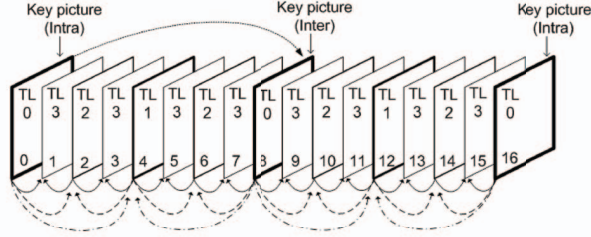
Fig. 1. Hierarchical prediction structure for scalable H.264 for GOP size = 8, two GOPs are shown.

layer (FGS0) of each temporal level decreases from the lowest to the highest temporal level, and each FGS layer for all the frames is considered as a single layer. Further, each layer of each frame is packetized into constant size packets ($\gamma = 100$ bytes) for transmission. At the receiver, any unrecoverable errors in each packet would result in dropping of that packet and hence would mean loss of the layer to which the packet belongs. We assume that the base layers of all the key pictures are received error free. Using the fact that the scalable H.264 encoding and decoding are done on a GOP basis, it is possible to use the frames within a GOP for error concealment purposes. In the event of losing a frame, temporal error concealment at the decoder is applied such that the lost frame is replaced by the nearest available frame in the decreasing as well as increasing sequential order but from only lower or same temporal levels. We start towards the frames that have a temporal level closer to the temporal level of the lost frame, e.g. in a GOP of eight frames, if frame f_6 is lost, the order in which the frames are used for concealment is f_4 and then f_8. For the frame in the center of the GOP (like f_4), the key picture at the start of the GOP is used for concealment. The SDDE algorithm can be modified to work for any error concealment technique.

Without loss of generality, in the following derivation of the proposed distortion estimation algorithm, we consider a base layer and two FGS layers. We assume the frames are lexicographically ordered and the distortion of each macroblock (and hence, each frame) is the summation of the distortion estimated for all the pixels in the macroblock of that frame. Let f_n^i denote the original value of pixel i in frame n and $\hat{f}_n^i$ denote its encoder reconstruction. The reconstructed pixel value at the decoder is denoted by $\tilde{f}_n^i$. The mean square error for this pixel is

$$d_n^i = \mathrm{E}\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} = \left(f_n^i\right)^2 - 2f_n^i \mathrm{E}\left\{\tilde{f}_n^i\right\} + \mathrm{E}\left\{\left(\tilde{f}_n^i\right)^2\right\} \quad (1)$$

where d_n^i is the distortion per pixel. As mentioned earlier, the base layer of all the key pictures are guaranteed to be received error free, the sth moment of the ith pixel of the key pictures n is calculated as follows:

$$\mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\} = P_{nE1}\left(\hat{f}_{nB}^i\right)^s + (1 - P_{nE1}) P_{nE2}\left(\hat{f}_{n(B+E1)}^i\right)^s + (1 - P_{nE1})(1 - P_{nE2})\left(\hat{f}_{n(B+E1+E2)}^i\right)^s \quad (2)$$

where $\hat{f}_{nB}^i$, $\hat{f}_{n(B+E1)}^i$, $\hat{f}_{n(B+E1+E2)}^i$ are the reconstructed pixel values at the encoder using only the base layer, the base along with the first FGS layer and the base layer with both of the FGS layers of frame n, respectively. P_{nE1} and P_{nE2} are the probabilities of losing the first and the second FGS layer of frame n, respectively.

For all the frames except the key pictures of a GOP, let us denote $\hat{f}_{nB\_uv}^i$ as the ith pixel value of the base layer of frame n reconstructed at the encoder. Frames $u(<n)$ and $v(>n)$ are the reference

pictures used in the hierarchical prediction structure for the reconstruction of frame n. In the decoding process of scalable H.264, the frames of each GOP are decoded in the order starting from the lowest to the highest temporal level. At the decoder,

- If frame u is not available as the reference picture for frame n (where frame n does not belong to the highest temporal level), then frame u' is selected as the new reference picture such that $u' < n$ and $TL(u') \leq TL(n)$ where $TL(.)$ is the temporal level to which the corresponding frame belongs. For the frames in the highest temporal level, $u' < n$ and $TL(u')$ is strictly less than $TL(n)$. Let us define $\mathbf{L}_n$ as the set consisting of frame u and all the possible choices of u' for frame n.

- If frame v is not available as the reference picture for frame n, then frame v' is selected as the new reference picture such that $v' > n$ and $TL(v') < TL(n)$. In this case, we define $\mathbf{R}_n$ as the set consisting of frame v and all the possible choices of v' for frame n.

The sth moment of the ith pixel of frame n when at least the base layer is received correctly is defined as:

$$\mathrm{E}\left\{\left(\tilde{f}_n^i(\mathbf{L}_n, \mathbf{R}_n)\right)^s\right\} = \sum_{j=1}^{|\mathbf{L}_n|} \sum_{k=1}^{|\mathbf{R}_n|} \left(1 - P_{\mathbf{L}_n(j)}\right)\left(1 - P_{\mathbf{R}_n(k)}\right)$$
$$\prod_{c=1}^{j-1} P_{\mathbf{L}_n(c)} \prod_{d=1}^{k-1} P_{\mathbf{R}_n(d)} \mathrm{E}\left\{\left(\tilde{f}_{n\_\mathbf{L}_n(j)\mathbf{R}_n(k)}^i\right)^s\right\} \quad (3)$$

where,

$$\mathrm{E}\left\{\left(\tilde{f}_{n\_\mathbf{L}_n(j)\mathbf{R}_n(k)}^i\right)^s\right\} = P_{nE1}\left(\hat{f}_{nB\_\mathbf{L}_n(j)\mathbf{R}_n(k)}^i\right)^s + P_{nE2}(1 - P_{nE1})\left(\hat{f}_{n(B+E1)\_\mathbf{L}_n(j)\mathbf{R}_n(k)}^i\right)^s + (1 - P_{nE2})(1 - P_{nE1})\left(\hat{f}_{n(B+E1+E2)\_\mathbf{L}_n(j)\mathbf{R}_n(k)}^i\right)^s \quad (4)$$

where, $P_{\mathbf{L}_n(j)}$ and $P_{\mathbf{R}_n(k)}$ are the probabilities of losing the base layer of the reference frames j and k from the sets $\mathbf{L}_n$ and $\mathbf{R}_n$, respectively.

Now to get the distortion per-pixel after error concealment, we will define a set $\mathbf{Q} = \{f_n, f_{q1}, f_{q2}, f_{q3}, ..., f_{GOPend}\}$, where f_n is the frame to be concealed, f_{q1} is the first frame, f_{q2} is the second frame to be used for concealment of f_n, and so on till one of the GOP ends is reached. The sth moment of the ith pixel using the set $\mathbf{Q}$ is defined as $\mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\}$,

$$\mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\} = (1 - P_n) \mathrm{E}\left\{\left(\tilde{f}_n^i(\mathbf{L}_n, \mathbf{R}_n)\right)^s\right\}$$
$$+ P_n (1 - P_{q1}) \mathrm{E}\left\{\left(\tilde{f}_{q1}^i(\mathbf{L}_{q1} \setminus \{f_n\}, \mathbf{R}_{q1} \setminus \{f_n\})\right)^s\right\}$$
$$+ P_n P_{q1}(1 - P_{q2}) \mathrm{E}\left\{\left(\tilde{f}_{q2}^i(\mathbf{L}_{q2} \setminus \{f_n, f_{q1}\}, \mathbf{R}_{q2} \setminus \{f_n, f_{q1}\})\right)^s\right\}$$
$$+ ... + P_n \prod_{z=1}^{|\mathbf{Q}|-2} P_{qz} \mathrm{E}\left\{\left(\tilde{f}_{GOPend}^i\right)^s\right\} \quad (5)$$

where P_n and P_{qz} are the probabilities of losing the base layer of frame n and qz, respectively; $\mathbf{L}_n \setminus \{f_w\}$ is the set of all the reference frames $\mathbf{L}_n$ excluding frame f_w, and $\mathbf{R}_n \setminus \{f_w\}$ is the set of all the reference frames $\mathbf{R}_n$ excluding frame f_w. The mean square error in (1) is obtained by calculating the the 1st and 2nd moments of pixel i of frame n using (2), (3), (4) and (5).

The performance of the SDDE algorithm is evaluated by comparing it with the actual decoder distortion averaged over 200 channel realizations. Figure 2(a) show the results for the "Foreman" sequence for GOP = 8. Each of these layers is considered to be affected with different loss rates as follows: $P_{TL0} = 0\%$, $P_{TL1} = 10\%$, $P_{TL2} = 20\%$, $P_{TL3} = 30\%$, $P_{E1} = 50\%$ and $P_{E2} = 60\%$
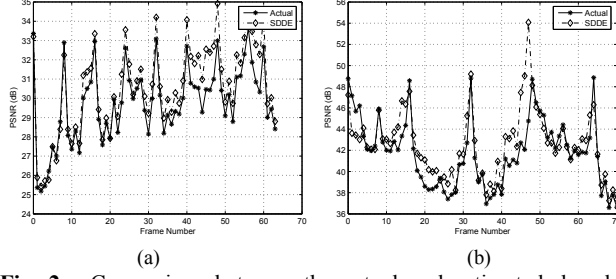
Fig. 2. Comparison between the actual and estimated decoder PSNR, (a) "Foreman" sequence and (b) "Akiyo" sequence.

where P_{TLx} is the probability of losing the base layer of a frame that belongs to TLx, P_{E1} and P_{E2} are the probabilities of losing FGS1 and FGS2 of each frame, respectively. It is evident that the proposed SDDE algorithm provides an accurate estimate of the scalable H.264 decoder distortion at the encoder. Similar results are also shown in Figure 2(b) using the "Akiyo" sequence with GOP =16 and packet error rates: $P_{TL0} = 0\%$, $P_{TL1} = 10\%$, $P_{TL2} = 15\%$, $P_{TL3} = 20\%$, $P_{TL4} = 25\%$, $P_{E1} = 30\%$ and $P_{E2} = 40\%$.

3. SYSTEM DESCRIPTION

After video encoding, the base and FGS layers of each frame are divided into packets of constant size ($= \gamma$) which then channel encoded using 16-bit CRC for error detection and rate-compatible punctured convolutional (RCPC) codes for UEP. These channel encoded packets are further encoded using O-STBC for transmission over MIMO wireless system. A Rayleigh flat-fading channel with AWGN is considered and ML decoding is used to detect the transmitted symbols which are then demodulated and channel decoded for error correction and detection. All the error-free packets for each frame are buffered and then fed to the source decoder with error concealment for video reconstruction.

For the MIMO system, we consider $M_t = 4$ transmit and $M_r = 1$ receive antennas. We used the O-STBC design, $\mathbf{G}_4(x_1, x_2, x_3)$ of rate $3/4$ (proposed by Tarokh *et. al* [9]), where x_1, x_2 and x_3 are the symbols that can be chosen from either same or different constellations, transmitted in $T = 4$ time slots.

$$\mathbf{G}_4(x_1, x_2, x_3) = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ -x_2^* & x_1^* & 0 & x_3 \\ -x_3^* & 0 & x_1^* & -x_2 \\ 0 & -x_3^* & x_2^* & x_1 \end{bmatrix} \quad (6)$$

The signal model is given as $\mathbf{Y} = \sqrt{\frac{\rho}{M_t}}\mathbf{C}\mathbf{H} + \mathbf{N}$, where $\mathbf{C}_{T \times M_t}$ is the energy-normalized transmitted signal matrix and is given as $\mathbf{C} = \sqrt{\frac{T}{K}}\mathbf{G}_{M_t}(x_1, x_2, \ldots, x_K)$; K is the number of different symbols in a codeword. $\mathbf{H}_{M_t \times M_r}$ is the channel coefficient matrix; $\mathbf{Y}_{T \times M_r}$ is the received signal matrix and $\mathbf{N}_{T \times M_r}$ is the noise matrix. The noise samples and the elements of $\mathbf{H}$ are independent samples of a zero-mean complex Gaussian random variable with variance 1. The fading channel is assumed to be quasi-static. The factor $\sqrt{\frac{\rho}{M_t}}$ is to ensure that ρ is the SNR at each receiver antenna and is independent of M_t. We assume perfect channel state information is known at the receiver, and the ML decoding is used to detect the transmitted symbols, i.e. $x_1, x_2, \ldots, x_K$ independently.

4. OPTIMAL BANDWIDTH ALLOCATION

We consider the minimization of the expected end-to-end distortion by optimally selecting the QP, the RCPC coding rate and the symbol constellation choice for the MIMO transmission on a GOP-by-

Table 1. Layer allocation on O-STBC symbols.

	x_1	x_2	x_3
A_1	$\mu_1, \mu_2, \mu_3, \mu_4$	μ_5	μ_6
A_2	μ_1, μ_2, μ_3	μ_4, μ_5	μ_6
A_3	μ_1, μ_2	μ_3, μ_4, μ_5	μ_6
A_4	μ_1	$\mu_2, \mu_3, \mu_4, \mu_5$	μ_6

GOP basis. We consider the combined temporal and FGS scalability and define a total of L layers for a GOP. The first $L - 2$ layers ($\mu_1, \ldots, \mu_{L-2}$) are the base layers (FGS0) of the frames associated with the lowest to the highest temporal level in decreasing order of importance for video reconstruction. The other two FGS layers (FGS1 and FGS2) of all the frames in a GOP are defined as individual layers (μ_{L-1}, μ_L) of even lesser importance.

$$\{\mathbf{QP}^*, \mathbf{R}_c^*, \mathbf{M}^*\} = \underset{\{\mathbf{QP}, \mathbf{R}_c, \mathbf{M}\}}{\arg \min} \ E\{D_{s+c}\} \ s.t. \ B_{s+c} \leq B_{budget} \quad (7)$$

where B_{s+c} is the transmitted symbol rate, B_{budget} is the total available symbol rate and $E\{D_{s+c}\}$ is the total expected end-to-end distortion which is accurately estimated using the SDDE algorithm. $\mathbf{QP}$, $\mathbf{R}_c$ and $\mathbf{M}$ are the admissible set of values for RCPC coding rates, symbol constellations and QP values, respectively. For each of the layers, $\mathbf{R}_c^* = \{R_{c,\mu_1}, \ldots, R_{c,\mu_L}\}$; $\mathbf{M}^* = \{M_{\mu_1}, \ldots, M_{\mu_L}\}$ and $\mathbf{QP}^* = \{QP_{\mu_1}, \ldots, QP_{\mu_L}\}$ define the RCPC coding rates, the symbol constellations and QP values, respectively after optimization. In solving the problem defined in (7), we take advantage of the independent transmission of each symbol in the O-STBC in (6) by allocating L layers to three different groups corresponding to x_1, x_2 and x_3. Table 1 shows the allocations (A_1, A_2, A_3, A_4) considered here for GOPsize = 8 (four temporal and two FGS layers). It is necessary to emphasize that the optimal allocation is done on a GOP-by-GOP basis and is decided upon the expected distortion (PSNR) for each O-STBC symbol (under the bandwidth constraint) which is obtained as follows:

- Based on (8), the optimal parameter set $\mathbf{X}_1^* = \{\mathbf{QP}_{x_1}^*, \mathbf{R}_{c,x_1}^*, \mathbf{M}_{x_1}^*\}$ for all the layers transmitted over the O-STBC symbol x_1 is obtained by using the admissible set of values of each of the parameter. $PSNR_{x_1}$ and B_{x_1} are the estimated PSNR and the symbol rate allocated for x_1, respectively.

$$\mathbf{X}_1^* = \underset{\{\mathbf{QP}, \mathbf{R}_c, \mathbf{M}\}}{\arg \max} PSNR_{x_1} \ s.t. \ T \times B_{x_1} \leq B_{budget} \quad (8)$$

- Given $\mathbf{X}_1^*$, the optimal parameter set $\mathbf{X}_2^*$ is obtained using (9).

$$\mathbf{X}_2^* = \underset{\{\mathbf{QP}, \mathbf{R}_c, \mathbf{M}\}/\mathbf{X}_1^*}{\arg \max} PSNR_{x_2} \ s.t. \ T \times B_{x_2} \leq B_{budget} \quad (9)$$

- Finally, having obtained $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ the optimal set $\mathbf{X}_3^*$ is obtained using (10).

$$\mathbf{X}_3^* = \underset{\{\mathbf{QP}, \mathbf{R}_c, \mathbf{M}\}/\{\mathbf{X}_1^*, \mathbf{X}_2^*\}}{\arg \max} PSNR_{x_3} \ s.t. \ T \times B_{x_3} \leq B_{budget} \quad (10)$$

PSNR values in (8), (9) and (10) is calculated using the SDDE algorithm as in (2) and (5). Let us define the packet error rate for the constant size packets as $PER(R_{c,\mu_l}, M_{\mu_l})$, which depends on the channel parameters. Now, the probabilities P_n, P_{nE1} ($P_n(l = L - 1)$) and P_{nE2} ($P_n(l = L)$) are obtained as:

$$P_n = 1 - (1 - PER(R_{c,\mu_l}, M_{\mu_l}))^{\left\lceil \frac{N_{n,\mu_l}}{\gamma} \right\rceil}, \quad l \in \{1, 2, \ldots, L - 2\} \quad (11)$$

where N_{n,μ_l} is the size of FGS0 of the frame n which belongs to the layer μ_l; $N_{n,\mu_{L-1}}$ and N_{n,μ_L} are the size of the layers FGS1
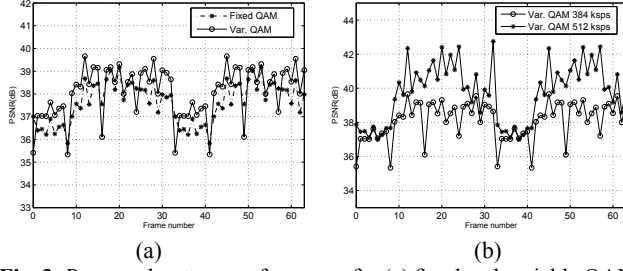
Fig. 3. Proposed system performance for (a) fixed and variable QAM for bandwidth of 384 ksps, (b) variable QAM for target bandwidth 512 and 384 ksps.

and FGS2 of frame n, respectively. Next, the bandwidth allocated to each O-STBC symbol is obtained as

$$B_{x_a} = \sum_{\mu_l \in x_a} \frac{R_{s,\mu_l}}{R_{c,\mu_l} \times \log_2(M_{\mu_l})} \quad (12)$$

where R_{s,μ_l} is the source rate for layer μ_l, it is in bits/sec and depends on the quantization parameter value used for that layer. Each of the problems in (8),(9) and (10) is a constrained optimization problem and is solved as an unconstrained one by using the Lagrangian method as in [16].

5. EXPERIMENTAL RESULTS

For all the experimental results, "Foreman" sequence is encoded at 30 fps, GOP=8 and constant Intra-update (I) at every 32 frames. For optimization, we consider QP value in the range of 20 to 50 and RCPC coding rates of $\mathbf{R}_c = 8/N : N = 32, 30, \ldots, 10$, which are obtained by puncturing a mother code of rate $8/32$ with constraint length of 3 and a code generator $[23;35;27;33]_o$. Quadrature amplitude modulation (QAM) is used with the possible constellations size $\mathbf{M}=\{4, 8, 16\}$. In Figure 3(a), we show the performance of the proposed system for optimal selection of parameters and the comparison of optimal constellation versus fixed constellation across all the layers transmitted over the MIMO system. It is seen that choosing optimal modulation across the layers (variable QAM) for every GOP (avg. PSNR = 38.61 dB) outperforms the case when the modulation for all the layers is optimally selected but kept fixed (fixed QAM) for the whole sequence (avg. PSNR = 38.32 dB). Figure 3(b) shows the the system performance for two target bandwidth values 384 ksps and 512 ksps after optimal parameter selection.

6. CONCLUSIONS

We proposed a wireless video transmission system that integrated the latest scalable H.264 coding providing a combined scalability and spatial diversity technique using O-STBC over broadband MIMO systems. We developed an accurate decoder distortion estimation algorithm and validated its accuracy using wireless video transmission simulations. Using the decoder distortion estimation algorithm, the bandwidth-constrained optimization problem has been solved. We exploited the orthogonal structure of the O-STBC codes by allocating layers over different codeword symbols modulated using different constellations. The optimization was carried on a GOP-by-GOP basis and the results for different target bandwidth values were presented. The results indicate the advantage of updating the optimal modulation selection every GOP as compared to selecting it optimally but keeping it fixed for the whole sequence.

7. REFERENCES

[1] T. Stockhammer, M. H. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments," *IEEE Trans. CSVT*, vol. 13, no. 7, pp. 657 – 673, Jul. 2003.

[2] S Wenger, "H.264/AVC over IP," *IEEE Trans. CSVT*, vol. 13, no. 7, pp. 645 – 656, Jul. 2003.

[3] H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "Combined scalability support for the scalable extension of H. 264/AVC," in *proc. ICME 2005*, Jul. 2005, pp. 446– 449.

[4] J. Reichel, H. Schwarz, and M. Wien, "Scalable working draft - working draft 1," Joint Video Team (JVT), Doc. JVT-N020, Hong Kong, CN, January 2005.

[5] R. Zhang, S.L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE JSAC*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[6] Y. Shen, P. C. Cosman, and L. B. Milstein, "Video coding with fixed-length packetization for a tandem channel," *IEEE Trans. IP*, vol. 15, no. 2, pp. 273– 288, Feb. 2006.

[7] S. Regunathan, R. Zhang, and K. Rose, "Scalable video coding with robust mode selection," *Signal Proc. Image Comm.*, vol. 16, pp. 725–732, 2001.

[8] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE JSAC*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[9] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Info. Th.*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[10] J. Song and K. J. R. Liu, "Robust progressive image transmission over OFDM systems using space-time block code," *IEEE Trans. Multimedia*, vol. 4, no. 3, pp. 394–406, Sept. 2002.

[11] S. Zhao, Z. Xiong, X. Wang, and J. Hua, "Progressive video delivery over wideband wireless channels using space-time differentially coded OFDM systems," *IEEE Trans. Mobile Computing*, vol. 5, no. 4, pp. 303– 316, Apr. 2006.

[12] C. Kuo, C. Kim, and C.-C. J. Kuo, "Robust video transmission over wideband wireless channel using space-time coded OFDM systems," in *proc. WCNC2002*, Mar. 2002, vol. 2, pp. 931– 936.

[13] T. Schierl, H. Schwarz, D. Marpe, and T. Wiegand, "Wireless broadcasting using the scalable extension of H. 264/AVC," in *proc. ICME 2005*, Jul. 2005, pp. 884– 887.

[14] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Comm.*, vol. 36, no. 4, pp. 389–400, April 1988.

[15] T. V. Ramabadran and S. S. Gaitonde, "A tutorial on CRC computations," *IEEE Micro*, vol. 8, Aug 1998.

[16] L. P. Kondi, D. Srinivasan, D. A. Pados, and S.N. Batalama, "Layered video transmission over wireless multirate DS-CDMA links," *IEEE Trans. CSVT*, vol. 15, no. 12, pp. 1629–1637, Dec. 2005.