SCALABLE H.264 WIRELESS VIDEO TRANSMISSION USING QUASI-ORTHOGONAL SPACE-TIME BLOCK CODES

Mohammad K. Jubran
Department of Electrical Engineering
Birzeit University
Birzeit, Palestine
mjubran@birzeit.edu

Manu Bansal
Intellectual Property Group
Goodwin Procter LLP
Boston, MA 02109 USA
mbansal@goodwinprocter.com

Lisimachos P. Kondi
Department of Computer Science
University of Ioannina
Ioannina 45110 Greece
lkon@ieee.org

ABSTRACT

We propose a low-delay low-complexity end-to-end video transmission system that integrates the latest scalable H.264 codec and the full-rate full-diversity quasi-orthogonal space-time block codes. At the encoder of the scalable H.264 codec, we have developed a low-delay low-complexity method for the estimation of the video distortion at the receiver for given channel conditions. The distortion estimation algorithm is validated using experimental results. In this work, we show clearly the advantage of using the quasi-orthogonal space-time block codes (STBC), which provide a higher rate compared to the orthogonal space-time block codes in a wireless video transmission system. This is in spite of the use of a non-optimal signal constellation structure. The performance results are obtained after addressing the bandwidth allocation problem that employs the scalable decoder distortion estimation algorithm for the optimal selection of the application layer and physical layer parameters.

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

1. INTRODUCTION

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) [1], [2], exploit the orthogonality property of the code matrix to achieve the full diversity gain and have the advantage of low complexity maximum-likelihood (ML) decoding. Later in [3], Su and Xia proposed a full-rate full-diversity quasi-orthogonal STBC by appropriately choosing the signal constellations of the symbols in the block codes. The full-rate quasi-orthogonal codes guaranteed fast maximum likelihood (ML) decoding for each pair of transmitted symbols instead of single symbols as in the O-STBC. The resultant code outperformed the O-STBC considering equal spectral efficiency and optimal signal constellations selection [3]. The end-to-end video transmission scheme proposed here employs STBC codes over a multiple-input multiple-output (MIMO) system. The independent transmission and decoding of symbols (individual or pair-wise) in a given STBC enables us to independently choose the elements of the codeword from different constellations and an additional unequal error protection can be provided for the elements of the codeword.

The scalable extension of H.264/AVC (SVC) has an error-resilient network adaptation layer (NAL) structure and provides superior compression efficiency [4], [5]. The combined scalability provided by the codec is exploited here to improve the video transmission over error-prone wireless networks by protecting the different layers with unequal error protection (UEP). In [6], progressive video transmission is proposed over a space-time differentially coded OFDM system with optimal rate and power allocation. In [7], an integrated system of data-partitioned video coding, layered space-time block coding and OFDM modulation is proposed but no optimization for resource allocation is addressed. However, in all the above-mentioned work, the orthogonal or quasi-orthogonal structure of STBC codes has not been exploited by independent transmission of the layered video over different symbols of the STBC code world. In [8], we presented a bandwidth optimization algorithm for SVC video transmission using O-STBC.

In this paper, we show the advantage of using full-rate full-diversity quasi-orthogonal STBC over O-STBC for high rate video transmission systems, even for non-optimal signal constellation choices for block code symbols. We consider the bandwidth constrained optimization problem for both types of codes and propose specific allocations of the temporal and quality scalable layers to different STBC symbols. The bandwidth allocation problem is addressed by minimizing the expected end-to-end distortion (for one group of pictures (GOP) at a time) and optimally selecting the quantization parameter (QP), rate-compatible punctured convolutional (RCPC) code rate and the constellation(s) for STBC symbols. A good knowledge of the total end-to-end decoder distortion at the encoder is necessary for such optimal allocation. In [9], 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 low-delay low-complexity method for the accurate estimation of the distortion of scalable SVC coded video at the receiver. The algorithm has a lower complexity than the algorithm in [8] and takes into account loss of both temporal and SNR scalable layers as well as error concealment at the decoder. The performance of the algorithm is validated by comparing it with simulated decoder distortion estimation results using different packet loss rate values.

2. SCALABLE H.264 CODEC AND DECODER DISTORTION ESTIMATION

SVC is based on a hierarchical prediction structure in which 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) and zero or more quality scalable enhancement (FGS) layers. Also, the priority of the base 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 (γ = 100 bytes for this work) for transmission. At the receiver, any unrecoverable errors in each packet would result in dropping the 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 SVC encoding and decoding is 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. For the frame in the center of the GOP, the key picture at the start of the GOP is used for concealment.

In the following derivation of the proposed low-complexity low-delay scalable decoder distortion estimation algorithm (*SDDE TrueRef*), we consider a base layer and two FGS layers. We assume that 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 defined as

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

where d_n^i is the distortion per pixel. As mentioned earlier, the base layers of all the key pictures are guaranteed to be received error free. The s^{th} moment of the i^{th} pixel of the key pictures n is calculated as

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

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 $\hat{f}_{nB\_u_n v_n}^i$ as the i^{th} pixel value of the base layer of frame n reconstructed at the encoder. Frames $u_n(< n)$ and $v_n(> n)$ are the reference pictures used in the hierarchical prediction structure for the reconstruction of frame n. We will refer to these frames (u_n and v_n) as the "true" reference pictures for frame n. In the decoding process of SVC, the frames of each GOP are decoded in the order starting from the lowest to the highest temporal level. At the decoder, if either or both of the true reference frames are not received correctly, the non-key picture(s) will be considered erased and will be concealed.

For the *SDDE TrueRef* algorithm, the s^{th} moment of the i^{th} pixel of frame n when at least the base layer is received correctly is defined as

$$\begin{aligned}&E\left\{\left(\tilde{f}_n^i\left(u_n, v_n\right)\right)^s\right\} = \left(1 - P_{u_n}\right)\left(1 - P_{v_n}\right) P_{nE1}\left(\hat{f}_{nB\_u_n v_n}^i\right)^s \\ &+ \left(1 - P_{u_n}\right)\left(1 - P_{v_n}\right) P_{nE2}\left(1 - P_{nE1}\right)\left(\hat{f}_{n(B,E1)\_u_n v_n}^i\right)^s \\ &+ \left(1 - P_{u_n}\right)\left(1 - P_{v_n}\right)\left(1 - P_{nE2}\right)\left(1 - P_{nE1}\right)\left(\hat{f}_{n(B,E1,E2)\_u_n v_n}^i\right)^s\end{aligned} \tag{3}$$

where, P_{u_n} and P_{v_n} are the probabilities of losing the base layer of the reference frames u_n and v_n, respectively. Now to get the distortion per-pixel after error concealment, we 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 s^{th} moment of the i^{th} pixel using the set $\mathbf{Q}$ is defined as

$$\begin{aligned}\mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\} &= \left(1 - P_n\right)\mathrm{E}\left\{\left(\tilde{f}_n^i\left(u_n, v_n\right)\right)^s\right\} \\ &+ \left(1 - \bar{P}_n\right)\left(1 - P_{q1}\right)\mathrm{E}\left\{\left(\tilde{f}_{q1}^i\left(u_{q1}, v_{q1}\right)\right)^s\right\} \\ &+ ... + \left(1 - \bar{P}_n \prod_{z=1}^{|\mathbf{Q}|-2} \bar{P}_{qz}\right)\mathrm{E}\left\{\left(\tilde{f}_{GOPend}^i\right)^s\right\}\end{aligned} \tag{4}$$

where $\bar{P}_n = (1 - P_n)(1 - P_{u_n})(1 - P_{v_n})$ is the probability of correctly receiving the base layers of frame n and the base layers of its reference pictures.

The performance of the *SDDE TrueRef* algorithm is evaluated by comparing it with the actual decoder distortion estimation averaged over 200 channel realizations. Different video sequences encoded at 30 fps, GOP size of eight frames and six layers are used in packet-based video transmission simulations. Each of these layers is considered to be affected with different loss rates $P = \{P_{TL0}, P_{TL1}, P_{TL2}, P_{TL3}, P_{E1}, P_{E2}\}$, where P_{TLx} is the probability of losing the base layer of a frame that belongs to TLx and P_{E1}, P_{E2} are the probabilities of losing FGS1 and FGS2 of each frame, respectively. The performance of the *SDDE TrueRef* algorithm is evaluated for packet loss rates of $P1 = \{0\%, 0\%, 5\%, 5\%, 10\%, 20\%\}$, $P2 = \{0\%, 10\%, 20\%, 30\%, 50\%, 60\%\}$ and $P3 = \{0\%, 0\%, 10\%, 20\%, 30\%, 40\%\}$. In Table 1, the average PSNR performance is presented for the "Foreman", "Akiyo" and "Carphone" sequences. As can be observed, the *SDDE TrueRef* algorithm results in good average PSNR estimates and hence is used to solve the optimization problem as discussed in the following sections.

Table 1. Average PSNR comparison for the proposed *SDDE TrueRef* algorithm.

	Foreman 363 kbps	Akiyo 268 kbps	Carphone 612 kbps
Actual P1 (dB)	36.20	45.11	40.85
SDDE TrueRef P1 (dB)	36.46	45.81	41.12
Actual P2 (dB)	30.61	40.74	35.32
SDDE TrueRef P2 (dB)	30.52	41.47	35.27
Actual P3 (dB)	33.46	42.84	38.03
SDDE TrueRef P3 (dB)	34.10	43.67	38.78

3. SYSTEM DESCRIPTION

After video encoding, the base and FGS layers of each frame are divided into packets of constant size γ, which are then channel coded using combined 16-bit CRC for error detection and RCPC codes for UEP. These packets are then transmitted using space-time block codes (STBC) over the MIMO system. A Rayleigh flat-fading channel with AWGN is considered between each transmit and receive antenna pair. 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. For the purpose of this work we consider two STBCs:

- O-STBC: we consider the code matrix $\mathbf{G}_4(x_1, x_2, x_3)$ of rate $3/4$ (as proposed by Tarokh *et. al* [2]), where x_1, x_2 and x_3

are codeword symbols transmitted in $T = 4$ time slots.

$$\mathbf{G}_4\left(x_1, x_2, x_3\right) = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ -x_2^* & x_1^* & 0 & x_3 \\ -x_3^* & 0 & x_1^* & -x_2 \\ 0 & -x_3^* & x_2^* & x_1 \end{bmatrix} \qquad (5)$$

- Quasi-orthogonal STBC: we consider the full-rate full-diversity quasi-orthogonal STBC as proposed by [3]. This quasi-orthogonal STBC improves the symbol transmission rate by employing the code matrix $\mathbf{G}_4(x_1, x_2, x_3, x_4)$ of rate 1:

$$\mathbf{G}_4\left(x_1, x_2, x_3, x_4\right) = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2^* & x_1^* & -x_4^* & x_3^* \\ x_3 & x_4 & x_1 & x_2 \\ -x_4^* & x_3^* & -x_2^* & x_1^* \end{bmatrix} \qquad (6)$$

  where $x_1 \in \mathcal{A}$, $x_2 \in \mathcal{B}$, $x_3 \in e^{j\theta}\mathcal{A}$ and $x_4 \in e^{j\phi}\mathcal{B}$ for some signal constellations $\mathcal{A}$ and $\mathcal{B}$ and are transmitted in $T = 4$ time slots. It is necessary to emphasize that $\mathcal{A}$ and $\mathcal{B}$ could be the same or different constellations. The constellation rotation angles θ and ϕ are determined according to the corresponding signal constellation as proposed in [3].

For both the STBCs described above, the signal model is given as $\mathbf{Y} = \sqrt{\frac{\rho}{M_t}}\mathbf{CH}+\mathbf{N}$, where $\mathbf{C}_{T\times M_t} = \sqrt{\frac{T}{K}}\mathbf{G}_{M_t}(x_1, x_2, \ldots, x_K)$ is the energy-normalized transmitted signal matrix; 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. We assume that perfect channel state information is known at the receiver and the ML decoding is used to detect each of the symbols, i.e. x_1, x_2, x_3 independently for O-STBC. However, for the quasi-orthogonal STBC, the symbol pair x_1, x_3 is detected jointly using the ML decoding, and so is the symbol pair x_2, x_4. The decoding of these two symbol pairs is carried out independently.

4. OPTIMAL BANDWIDTH ALLOCATION

The bandwidth allocation problem for the two systems is defined as the minimization of the expected end-to-end distortion by optimally selecting the application layer parameter (QP) and the physical layer parameters (RCPC coding rate and symbol constellation choice). In both the cases the optimization is considered on a GOP-by-GOP basis and is constrained on the total available bandwidth (symbol rate) B_{budget}. We consider the combined temporal and FGS scalability and define a total of L layers, which are unequally protected by optimally selecting the physical layer parameters. 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. The bandwidth allocation problem is described as:

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

where B_{s+c} is the transmitted symbol rate, B_{budget} is the total available symbol rate and $E\left\{D_{s+c}\right\}$ is the total expected end-to-end distortion which is accurately estimated using the *SDDE TrueRef* algorithm. $\mathbf{QP}$, $\mathbf{R}_c$ and $\mathbf{M}$ are the admissible set of values for QP, RCPC coding rates and symbol constellations, respectively. For all the layers of each GOP, $\mathbf{QP}^* = \{QP_{\mu_1}, \ldots, QP_{\mu_L}\}$, $\mathbf{R}_c^* = \{R_{c,\mu_1}, \ldots, R_{c,\mu_L}\}$ and $\mathbf{M}^* = \{M_{\mu_1}, \ldots, M_{\mu_L}\}$ define the QP values, the RCPC coding rate values and the symbol constellations, respectively obtained after optimization. The transmitted symbol rate B_{s+c} can be obtained as

$$B_{s+c} = \sum_{l=1}^{L} \frac{R_{s,\mu_l}}{R_{c,\mu_l} \times \log_2(M_{\mu_l})} \times \frac{T}{K} \qquad (8)$$

where R_{s,μ_l}, R_{c,μ_l} and M_{μ_l} are the source coding rate, channel coding rate and signal constellation, respectively used for layer μ_l. T is the number of time slots required to transmit K symbols in each codeword over the MIMO system.

Table 2. Layer allocation for 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

Table 3. Layer allocation for quasi-orthogonal STBC symbols.

	$x_{1,3}$	$x_{2,4}$
A_1	$\mu_1, \mu_2, \mu_3, \mu_4$	μ_5
A_2	μ_1, μ_2, μ_3	μ_4, μ_5
A_3	μ_1, μ_2	μ_3, μ_4, μ_5
A_4	μ_1	$\mu_2, \mu_3, \mu_4, \mu_5$

4.1. Optimal bandwidth allocation for O-STBC

For the O-STBC system, the problem defined in (7) is solved by taking advantage of the independent transmission of each symbol in the O-STBC. This is done by allocating L layers to three different groups corresponding to O-STBC symbols, x_1, x_2 and x_3 as shown in Table 2 and solving for the bandwidth allocation problem one O-STBC symbol at a time. Table 2 shows the possible allocations (A_1, A_2, A_3, A_4) considered here for GOPsize = 8 where each of the six layers is associated with one of the O-STBC symbols. It is necessary to emphasize that here the optimal selection is done on a GOP-by-GOP basis for each allocation structure and the best allocation is selected after considering the expected distortion (PSNR) criteria for all the O-STBC symbols. This is defined in more detail as follows:

- Based on (9), 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.

$$\mathbf{X}_1^* = \underset{\{\mathbf{QP},\mathbf{R}_c,\mathbf{M}\}}{\arg\max}\ PSNR_{x_1}\ \ s.t.\ \ B_{x_1} \leq B_{budget} \qquad (9)$$

- Given $\mathbf{X}_1^*$, the optimal parameter set $\mathbf{X}_2^*$ is obtained using (10).

$$\mathbf{X}_2^* = \underset{\{\mathbf{QP},\mathbf{R}_c,\mathbf{M}\}/\mathbf{X}_1^*}{\arg\max}\ PSNR_{x_2}\ \ s.t.\ \ B_{x_2} \leq B_{budget} \qquad (10)$$

- Finally, having obtained $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ the optimal set $\mathbf{X}_3^*$ is obtained using (11).

$$\mathbf{X}_3^* = \underset{\{\mathbf{QP},\mathbf{R}_c,\mathbf{M}\}/\{\mathbf{X}_1^*,\mathbf{X}_2^*\}}{\arg\max}\ PSNR_{x_3}\ \ s.t.\ \ B_{x_3} \leq B_{budget} \qquad (11)$$

where B_{x_a}, $a \in \{1, 2, 3\}$ is the symbol rate allocated to each O-STBC symbol and is obtained as

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

4.2. Optimal bandwidth allocation for quasi-orthogonal STBC

To solve the optimization problem for the quasi-orthogonal STBC, the independence between the symbol pairs x_1, x_3 and x_2, x_4 is exploited by considering independent transmission and decoding of each symbol pair. In this case, we combine both FGS layers (μ_{L-1} and μ_L) into a single layer μ_{L-1} and allocate the $L-1$ layers to the two independent pairs. Table 3 shows the possible allocations (A_1, A_2, A_3, A_4) considered here for GOPsize = 8. As in the previous case, the optimal parameter selection is done on a GOP-by-GOP basis for each allocation structure and the best allocation is selected as per the PSNR criteria as explained below:

- Based on (13), the optimal parameter set $\mathbf{X}^*_{1,3} = \{\mathbf{QP}^*_{x_{1,3}}, \mathbf{R}^*_{c,x_{1,3}}, \mathbf{M}^*_{x_{1,3}}\}$ for all the layers transmitted over the quasi-orthogonal STBC symbols x_1 and x_3 is obtained.

$$\mathbf{X}^*_{1,3} = \underset{\{\mathbf{QP},\mathbf{R}_c,\mathbf{M}\}}{\arg\max}\ PSNR_{x_{1,3}}\ \ s.t.\ B_{x_{1,3}} \leq B_{budget} \quad (13)$$

- Given $\mathbf{X}^*_{1,3}$, the optimal parameter set $\mathbf{X}^*_{2,4}$ is obtained using (14).

$$\mathbf{X}^*_{2,4} = \underset{\{\mathbf{QP},\mathbf{R}_c,\mathbf{M}\}/\mathbf{X}^*_{1,3}}{\arg\max}\ PSNR_{x_{2,4}}\ \ s.t.\ B_{x_{2,4}} \leq B_{budget} \quad (14)$$

where B_{x_b}, $b \in \{(1,3),(2,4)\}$ is the bandwidth (symbol rate) allocated to each pair of the quasi-orthogonal STBC symbols and is obtained as

$$B_{x_b} = \sum_{\mu_l \in x_b} \frac{R_{s,\mu_l}}{R_{c,\mu_l} \times \log_2(M_{\mu_l})} \times \frac{T}{2} \quad (15)$$

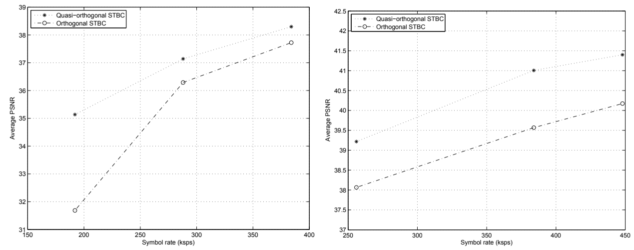
PSNR values in (9), (10), (11), (13) and (14) is calculated using the SDDE TrueRef algorithm. It is clear from section 2 that the accurate estimation of decoder distortion is dependent upon the probabilities of losing each layer (P_n, P_{nE1} and P_{nE2}). Let us define the packet error rate for the constant size packets (= γ bytes) 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:

$$P_n = 1-(1-PER(R_{c,\mu_l}, M_{\mu_l}))^{\left\lceil \frac{N_{n,\mu_l}}{\gamma} \right\rceil}, \quad l \in \{1,\ldots,L-2\} \quad (16)$$

where N_{n,μ_l} is the size of the layer 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 and FGS2 of frame n, respectively. The problems defined in (9), (10), (11), (13) and (14) are constrained optimization problems and are solved using the Lagrangian method.

5. EXPERIMENTAL RESULTS

For experimental results, various video sequences are encoded at 30 fps, GOP=8 and constant Intra-update (I) at every 32 frames. We consider the video encoding QP values in the range of 16 to 50 and RCPC coding rates of $\mathbf{R_C} = 8/N : N \in \{32, 28, 24, 20, 16, 12\}$. Rectangular quadrature amplitude modulation (QAM) is used with the possible constellation sizes **M**={4, 8, 16}. In Figure 1 (a), we compare the performance of the quasi-orthogonal STBC and the O-STBC systems for the transmission of "Foreman" sequence after the optimal selection of the application layer and physical layer parameters. The comparison in this figure is shown for multiple target symbol rate values and it is clearly shown that the full-rate full-diversity quasi-orthogonal STBC system outperforms the full-diversity O-STBC system in spite of the use of non-optimal signal constellations. Similar results are shown in Figure 1 (b) for the "Carphone" video sequence.



(a) (b)

Fig. 1. Performance comparison for O-STBC and quasi-orthogonal STBC Systems for (a) "Foreman", (b) "Carphone" video sequences.

6. CONCLUSIONS

We developed an accurate low-delay low-complexity decoder distortion estimation algorithm and validated its accuracy experimentally. Using the distortion estimation algorithm, we employed the full-rate full-diversity quasi-orthogonal STBC in an optimal bandwidth allocation framework for wireless video transmission over a 4×1 MIMO system. We have shown that using this code, which achieves a higher rate than the O-STBC, results in a better end-to-end video system performance in the PSNR sense. This is in spite of the use of non-optimal signal constellations.

7. REFERENCES

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[2] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[3] W. Su and X. Xia, "Signal constellations for quasi-orthogonal space-time block codes with full diversity," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2331– 2347, Oct 2004.

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

[5] 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.

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

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

[8] M. K. Jubran, M. Bansal, and L. P. Kondi, "Transmission of scalable H.264 coded video over wireless MIMO systems with optimal bandwidth allocation," in *Proc. ICASSP*, Honolulu, HI, April 2007.

[9] R. Zhang, S.L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, June 2000.