Scalable H.264 Wireless Video Transmission over MIMO-OFDM Channels

Manu Bansal[1], Mohammad Jubran[2], and Lisimachos P. Kondi[3,*]

[1] Pillsbury Winthrop Shaw Pittman LLP, Intellectual Property Dept., McLean, VA, USA
[2] Birzeit University, Dept of Electrical Engineering, Birzeit, Palestine
[3] University of Ioannina, Dept. of Computer Science, Ioannina, GR-45110, Greece

Abstract. A cross-layer optimization scheme is proposed for scalable video transmission over wireless Multiple Input Multiple Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) systems. The scalable video coding (SVC) extension of H.264/AVC is used for video source coding. The proposed cross-layer optimization scheme jointly optimizes application layer parameters and physical layer parameters. The objective is to minimize the expected video distortion at the receiver. Two methods have been developed for the estimation of video distortion at the receiver, which is essential for the cross-layer optimization. In addition, two different priority mappings of the SVC scalable layers are considered. Experimental results are provided and conclusions are drawn.

1 Introduction

Recent advances in computer technology, data compression, high-bandwidth storage devices, high-speed networks, and the third and the fourth generation (3G and 4G) wireless technology have made it feasible to provide the delivery of video over multicarrier wireless channels at high data rates [1]. Transmission over Multiple Input Multiple Output (MIMO) channels using Orthogonal Frequency Division Multiplexing (OFDM) provides such high data rates for multimedia delivery and therefore is of great interest in the area of wireless video applications.

Diversity techniques, such as space-time coding (STC) for multiple antenna systems (i.e., MIMO systems) have been proven to help overcome the degradations due to the wireless channels by providing the receiver with multiple replicas of the transmitted signal over different channels. MIMO systems employ orthogonal space-time block codes (O-STBC) [2], [3], which 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.

[*] This research was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme.

J. Blanc-Talon et al. (Eds.): ACIVS 2010, Part II, LNCS 6475, pp. 34–45, 2010.

On the other hand, OFDM mitigates the undesirable effects of a frequency-selective channel by converting it into a parallel collection of frequency-flat subchannels. OFDM is basically a block modulation scheme where a block of N information symbols is transmitted in parallel on N subcarriers. The subcarriers have the minimum frequency separation required to maintain orthogonality of their corresponding time domain waveforms, yet the signal spectra corresponding to the different subcarriers overlap in frequency. Hence, the available bandwidth is used very efficiently. An OFDM modulator can be implemented as an inverse discrete Fourier transform (IDFT) on a block of N information symbols. To mitigate the effects of intersymbol interference (ISI) caused by channel time spread, each block of N IDFT coefficients is typically preceded by a cyclic prefix (CP) or a guard interval consisting of G samples, such that the length of the CP is at least equal to the channel length. As a result, the effects of the ISI are easily and completely eliminated. Recent developments in MIMO techniques promise a significant boost in performance for OFDM systems. A parallel transmission framework for multimedia data over spectrally shaped channels using multicarrier modulation was studied in [4]. A space-time coded OFDM system to transmit layered video signals over wireless channels was presented in [5]. Video transmission with OFDM and the integration of STC with OFDM have been studied recently [6,7,8]. In [9] an optimal resource allocation method was proposed for multilayer wireless video transmission by using the large-system performance analysis results for various multiuser receivers in multipath fading channels. However, the above approaches have not exploited wireless video transmission over MIMO-OFDM systems with bandwidth optimization.

In this paper, we consider the bandwidth constrained transmission of temporal and quality scalable layers of coded video over MIMO-OFDM wireless networks, with optimization of source coding, channel coding and physical layer parameters on a per group of pictures (per-GOP) basis. The bandwidth allocation problem is addressed by minimizing the expected end-to-end distortion (for one GOP at a time) and optimally selecting the quantization parameter (QP), channel coding rate and the constellation for the STBC symbols used in this MIMO-OFDM system. At the source coding side, we use the scalable video coding (SVC) extension of the H.264/AVC standard which has an error-resilient network abstraction layer (NAL) structure and provides superior compression efficiency [10]. The combined scalability provided by the codec is exploited to improve the video transmission over error-prone wireless networks by protecting the different layers with unequal error protection (UEP). In [11,12], we proposed bandwidth optimization algorithms for SVC video transmission over MIMO (non-OFDM) channels using O-STBC.

A good knowledge of the total end-to-end decoder distortion at the encoder is necessary for such optimal allocation. Accordingly, we use the low-delay, low-complexity method for accurate distortion estimation for SVC video as discussed in [11] and also propose a new modified version of this distortion estimation method. The two distortion estimation methods differ in the priority order in which different types of scalability inherent in the SVC codec, namely temporal and Signal to Noise Ratio (SNR), are considered for estimation purposes. We also propose two different priority mappings of the scalable layers produced by SVC. Comparison results for the two priority mappings are presented for bandwidth constrained and distortion optimized video transmission over a MIMO-OFDM system. The results exemplify the advantages of the use of each priority mapping for different video sequences.

The rest of the paper is organized as follows. In section 2, the proposed system is introduced. In section 3, the scalable extension of H.264 is described. In section 4, the cross-layer optimization problem is formulated and solved. In section 5, the two video distortion estimation methods are discussed. In section 6, the priority mapping of the temporal and FGS layers of SVC is discussed. In section 7, experimental results are presented. Finally, in section 8, conclusions are drawn.

2 System Description

In our packet-based video transmission system, we utilize channel coding followed by orthogonal space-time block coding for MIMO-OFDM systems. After video encoding, the scalable layers of each frame are divided into packets of constant size γ, which are then channel encoded using a 16-bit cyclic redundancy check (CRC) for error detection and rate-compatible punctured convolutional (RCPC) codes for UEP. These channel-encoded packets are modulated with a particular constellation size and further encoded using O-STBC for each subcarrier for transmission over the MIMO wireless system. A 6-ray typical urban (TU) channel model with AWGN is considered (details shown in Table 1) and ML decoding is used to detect the transmitted symbols at each subcarrier, 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-OFDM system used here, we consider $M_t = 2$ transmit and $M_r = 2$ receive antennas. We used the O-STBC design for MIMO-OFDM systems in which two codewords (corresponding to two time instances) are transmitted. The channel is assumed to be quasi-static for these two codeword time periods. The codeword structure is as follows:

Table 1. Six-ray typical urban (TU) channel model

Delay (μs)	0.0	0.2	0.5	1.6	2.3	5.0
Power (mean)	0.189	0.379	0.239	0.095	0.061	0.037

$$\mathbf{C}_{OFDM1} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \\ | & | \\ | & | \\ | & | \\ x_{2N-1} & x_{2N} \end{bmatrix}, \tag{1}$$

and

$$\mathbf{C}_{OFDM2} = \begin{bmatrix} -x_2^* & x_1^* \\ -x_4^* & x_3^* \\ | & | \\ | & | \\ | & | \\ -x_{2N}^* & x_{2N-1}^* \end{bmatrix} \tag{2}$$

where N is the number of subcarriers and $(.)^*$ denotes the complex conjugate. The two codewords take two time instances and each row represents transmission over one subcarrier. Hence, the two codewords together form a 2 × 2 O-STBC for each subcarrier. In such a design, we gain spatial diversity but no frequency diversity.

The signal model at the j-th receive antenna for the n-th subcarrier at time t $(t = 1, 2)$ is given as

$$y_t^j\left(n\right) = \sqrt{\frac{\rho}{M_t}} \sum_{i=1}^{M_t} c_t^i\left(n\right) h_{ij}\left(n\right) + \eta_t^j\left(n\right), \tag{3}$$

where ρ is the channel SNR, $c_t^i(n)$ is the energy-normalized transmitted symbol from the i-th transmit antenna at the n-th tone, and $\eta_t^j(n)$ are independent Gaussian random variables with zero mean and variance 1. $h_{ij}(n)$ is the channel frequency response from the i-th transmit antenna to the j-th receive antenna at the n-th tone. t takes values 1 and 2 since there are two codewords that take two time instances, as mentioned earlier. 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 the transmitted symbols independently.

3 Scalable H.264 Codec

In this work, the scalable extension of H.264/AVC is used for video coding. We will use the acronym "SVC" to specifically refer to the scalable extension of H.264/AVC and not to scalable video coding in general. 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, and so on). Each of these pictures is represented by a non-scalable base layer (FGS0) and zero or more quality scalable enhancement fine granularity scalability (FGS) layers. The hierarchical coding structure of SVC is shown in Figure 1.

4 Optimal Bandwidth Allocation

The bandwidth allocation problem is defined as the minimization of the expected end-to-end distortion by optimally selecting the application layer parameter, QP

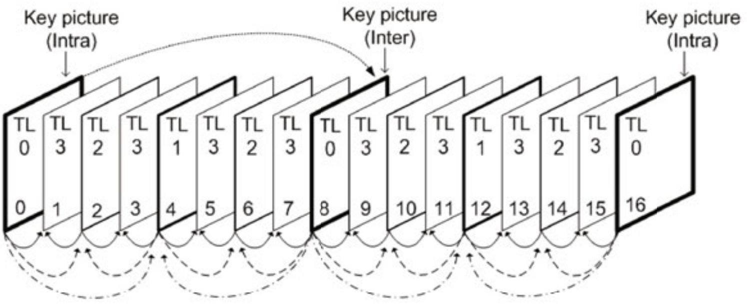

Fig. 1. Hierarchical prediction structure for SVC for a GOP size of 8

value for video encoding, and the physical layer parameters, RCPC coding rate and the symbol constellation choice for the STBC block code. The optimization is considered on a GOP-by-GOP basis and is constrained by the total available bandwidth (symbol rate) B_{budget}. We assume that the SVC codec produces L layers $\mu_1, \mu_2, \ldots, \mu_L$ via a combination of temporal and FGS scalability. Then, the bandwidth allocation problem can be described as:

$$\{\mathbf{QP}^*,\ \mathbf{R}_c^*,\ \mathbf{M}^*\} = \underset{\{\mathbf{QP},\ \mathbf{R}_c,\ \mathbf{M}\}}{\arg\min}\ E\left\{D_{s+c}\right\}\ s.t.\, B_{s+c} \leq B_{budget} \tag{4}$$

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 due to source and channel coding, which needs to be estimated as discussed in Section 5. $\mathbf{QP}$, $\mathbf{R}_c$ 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 rates and the symbol constellations for each scalable layer, respectively, obtained after optimization. The transmitted symbol rate B_{s+c} can be obtained as

$$B_{s+c} = \sum_{l=1}^{L} \frac{R_{s,\mu_l}}{R_{c,\mu_l} \times \log_2(M_{\mu_l})} \tag{5}$$

where R_{s,μ_l} is the source coding rate for layer μ_l in bits/s and depends on the quantization parameter value used for that layer; R_{c,μ_l} is the channel coding rate for layer μ_l and is dimensionless; M_{μ_l} is the constellation used by layer μ_l and $\log_2(M_{\mu_l})$ is the number of bits per symbol.

The problem of Eq. (4) is a constrained optimization problem and is solved using the Lagrangian method.

5 Decoder Distortion Estimation

In order to perform the optimization of Eq. (4), it is necessary to estimate the expected video distortion at the receiver $E\{D_{s+c}\}$. In this paper, we use the distortion estimation technique of [11] and we also propose a new technique.

As mentioned previously, SVC produces video frames which are partitioned into FGS layers. We assume that each layer of each frame is packetized into constant size packets of size γ 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 channel coding rate and constellation used for the transmission of the base layers of all key pictures is such that they 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.

As discussed in [11], 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 of even lesser priority. We will refer to this method as ***Temporal-SNR*** scalable decoder distortion estimation (SDDE) method. Alternatively, we can consider both the base and the FGS layers of the reference frames to be used for the encoding and the reconstruction of the frames of higher temporal levels (non-key pictures). In such a case, both the base and the FGS layers of the reference frames (from the lower temporal levels) are considered of the same importance, and of higher importance than the frame(s) (from a higher temporal level) to be motion-compensated and reconstructed. We will refer to this case as the ***SNR-Temporal*** SDDE method. Next we will present the derivations of the two above-mentioned SDDE methods.

5.1 Temporal-SNR SDDE

In the following derivation of the Temporal-SNR SDDE method, we consider a base layer and one FGS layer. We assume that the frames are converted into vectors via lexicographic ordering 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 [13]:

$$d_n^i = \mathrm{E}\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} = \left(f_n^i\right)^2 - 2f_n^i\mathrm{E}\left\{\tilde{f}_n^i\right\} + \mathrm{E}\left\{\left(\tilde{f}_n^i\right)^2\right\} \tag{6}$$

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

$$\mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\} = P_{nE1}\left(\hat{f}_{nB}^i\right)^s + \left(1 - P_{nE1}\right)\left(\hat{f}_{n(B,E1)}^i\right)^s \tag{7}$$

where $\hat{f}_{nB}^{i}$, $\hat{f}_{n(B,E1)}^{i}$ are the reconstructed pixel values at the encoder using only the base layer, and the base along with the FGS layer of frame n, respectively. P_{nE1} is the probability of losing the FGS layer of frame n.

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 Temporal-SNR SDDE method, the s^{th} moment of the i^{th} pixel of frame n when at least the base layer is received correctly is defined as

$$\begin{aligned} E\left\{\left(\tilde{f}_n^i\left(u_n, v_n\right)\right)^s\right\} &= \left(1-P_{u_n}\right)\left(1-P_{v_n}\right) P_{nE1}\left(\hat{f}_{nB\_u_n v_n}^i\right)^s \\ &+ \left(1-P_{u_n}\right)\left(1-P_{v_n}\right)\left(1-P_{nE1}\right)\left(\hat{f}_{n(B,E1)\_u_n v_n}^i\right)^s \end{aligned} \tag{8}$$

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

$$\begin{aligned} \mathrm{E}\left\{\left(\tilde{f}_n^i\right)^s\right\} &= \left(1-P_n\right) \mathrm{E}\left\{\left(\tilde{f}_n^i\left(u_n, v_n\right)\right)^s\right\} \\ &+ \left(1-\bar{P}_n\right)\left(1-P_{q1}\right) \mathrm{E}\left\{\left(\tilde{f}_{q1}^i\left(u_{q1}, v_{q1}\right)\right)^s\right\} \\ &+ \left(1-\bar{P}_n \bar{P}_{q1}\right)\left(1-P_{q2}\right) \mathrm{E}\left\{\left(\tilde{f}_{q2}^i\left(u_{q2}, v_{q2}\right)\right)^s\right\} \\ &+ ... + \left(1-\bar{P}_n \prod_{z=1}^{|\mathbf{Q}|-2} \bar{P}_{qz}\right) \mathrm{E}\left\{\left(\tilde{f}_{GOPend}^i\right)^s\right\} \end{aligned} \tag{9}$$

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.

5.2 SNR-Temporal SDDE

Similar to the Temporal-SNR SDDE case, in this method the base layer of all the key pictures are assumed to be received error-free and the s^{th} moment of the i^{th} pixel of the key pictures n is again calculated using Eq. (7). For all the frames except the key pictures of a GOP, let us denote $\hat{f}_{nB\_u_{(B,E1)n} v_{(B,E1)n}}^{i}$ as the i^{th} pixel value of the base layer of frame n reconstructed at the encoder. Frames $u_{(B,E1)n} (< n)$ and $v_{(B,E1)n} (> n)$ are the reference pictures (including both base and FGS layers) used in the hierarchical prediction structure for the reconstruction of frame n. In case of losing the FGS layers of the reference pictures, only the base layers of the frames u_n and v_n are used as the reference for frame n. As discussed above, in SVC the decoding of all the frames in a GOP is done from the lowest to the highest temporal level. Similar to the Temporal-SNR method, we will use the "true" reference frames for distortion estimation and hence, the loss of base layer of either or both the reference frames will result in the concealment of the frame n. The s^{th} moment of the i^{th} pixel of frame n when at least the base layer is received correctly is calculated as:

$$\begin{aligned} &E\left\{\left(\tilde{f}_n^i\left(u_n, v_n\right)\right)^s\right\} = P_{uvB} P_{nE1}\left(\hat{f}^i_{nB\_u_{Bn}v_{Bn}}\right)^s \\ &+P_{uvB}\left(1-P_{nE1}\right)\left(\hat{f}^i_{n(B,E1)\_u_{Bn}v_{Bn}}\right)^s \\ &+P_{uvB,E1} P_{nE1}\left(\hat{f}^i_{nB\_u_{(B,E1)n}v_{(B,E1)n}}\right)^s \\ &+P_{uvB,E1}\left(1-P_{nE1}\right)\left(\hat{f}^i_{n(B,E1)\_u_{(B,E1)n}v_{(B,E1)n}}\right)^s \end{aligned} \tag{10}$$

where, $P_{uvB} = (1 - P_{u_nB})(1 - P_{v_nB})P_{u_nE1}P_{v_nE1}$ is the probability of correctly receiving the base layers and not receiving the FGS layers of the frames u_n and v_n. Similarly, $P_{uvB,E1} = (1 - P_{u_nB})(1 - P_{v_nB})(1 - P_{u_nE1})(1 - P_{u_nE1})$ is the probability of correctly receiving the base layers and the FGS layers of the frames u_n and v_n. In case the base layer of frame n is lost, the complete frame has to be concealed. To get the distortion per-pixel after error concealment, we use Eq. (9).

The performance of the two SDDE methods 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 $P = \{P_{TL0}, P_{TL1}, P_{TL2}, P_{TL3}, P_{E1}\}$, where P_{TLx} is the probability of losing the base layer of a frame that belongs to TLx and P_{E1} is the probability of losing FGS1 of a frame. For performance evaluation, packet loss rates considered are $P1 = \{0\%, 0\%, 5\%, 5\%, 10\%\}$ and $P2 = \{0\%, 10\%, 20\%, 30\%, 50\%\}$. In Table 1, the average Peak Signal to Noise Ratio (PSNR) performance is presented for the "Foreman", "Akiyo" and "Carphone" sequences. As can be observed, both the Temporal-SNR and the SNR-Temporal methods result in good average PSNR estimates and hence they are used to solve the optimization problem of section 4.

Table 2. Average PSNR comparison for the proposed distortion estimation algorithms

	Foreman 363 kbps	Akiyo 268 kbps	Carphone 612 kbps
Actual P1 (dB)	36.40	45.91	40.85
Temporal-SNR SDDE (dB)	35.48	45.84	40.12
SNR-Temporal SDDE (dB)	36.00	45.43	40.35
Actual P2 (dB)	30.82	41.46	35.32
Temporal-SNR SDDE (dB)	29.80	41.20	35.10
SNR-Temporal SDDE (dB)	30.22	40.86	35.28

6 Priority Mapping of Scalable Layers

We considered two different mappings of the temporal and FGS layers of SVC into the scalable layers μ_i. Let us assume that the number of temporal layers is T. For a GOP size of 8, as used here, we have $T = 4$. For the first mapping, which we call the Temporal-SNR mapping, the first $L - 1$ layers $(\mu_1, \ldots, \mu_{L-1})$ 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. So, $L-1 = T$ and the number of scalable layers is $L = T+1$. The FGS layer of all the frames in a GOP are defined as a single layer μ_L of even lesser importance. The Temporal-SNR distortion estimation method uses exactly the same priorities as the Temporal-SNR mapping, so we used it for our experimental results for this mapping. The second mapping is the SNR-temporal mapping. For this mapping, there are two layers, base and enhancement for each of the T temporal layers. In this case, the FGS layer of the lower temporal layer has more importance than the base layer of the higher temporal levels. Thus, there are a total of $L = 2T$ scalable layers. For the SNR-Temporal mapping, we used the SNR-Temporal distortion estimation method, as it uses exactly the same priorities.

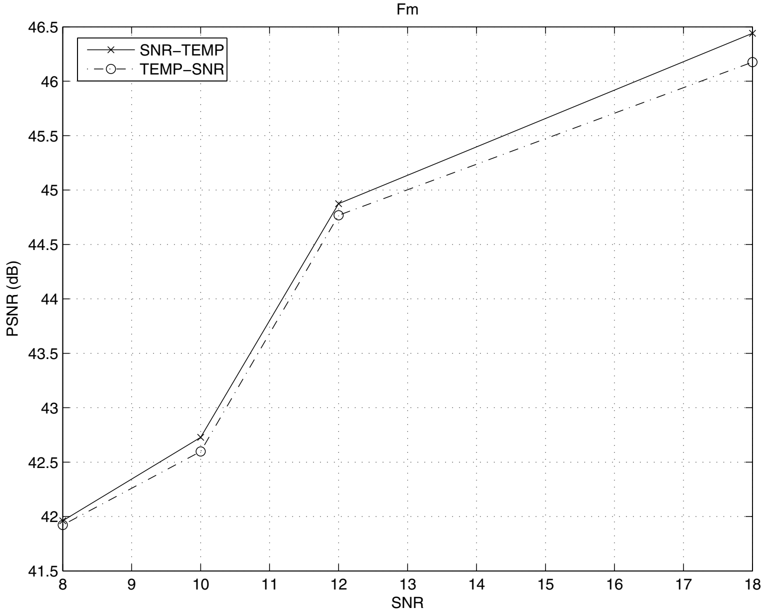

Fig. 2. Performance of the cross-layer optimization using the Temporal-SNR and the SNR-Temporal mappings of scalable layers ("Foreman" sequence)

7 Experimental Results

For experimental results, the "Foreman" and 'Akiyo" video sequences are encoded at 30 fps, GOP=8 and constant Intra-update (I) at every 32 frames. We

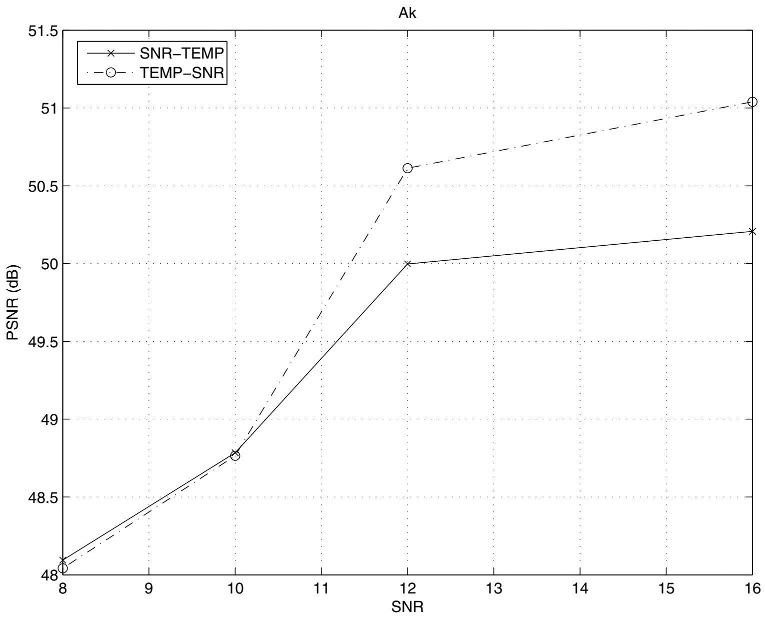

Fig. 3. Performance of the cross-layer optimization using the Temporal-SNR and the SNR-Temporal mappings of scalable layers ("Akiyo" sequence)

consider the video encoding QP values in the range of 16 to 50 and RCPC coding rates of $\mathbf{R}_\mathrm{C} = 8/K : K \in \{32, 28, 24, 20, 16, 12\}$, 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 $\mathbf{M}$=$\{4, 8, 16\}$. The total number of subcarriers N for the OFDM system is fixed at 64. This includes a cyclic prefix (CP) of 1/8 and guard interval (GI) of 1/8 of the total number of subcarriers. The packet size is chosen as $\gamma = 100$ bytes. Both the Temporal-SNR and SNR-Temporal priority mappings are considered.

Average PSNR results obtained for transmission of the "Foreman" sequence over the MIMO-OFDM system after the optimal selection of the application layer and physical layer parameters (on a GOP-by-GOP basis) for a channel SNR of 8dB, 10dB, 12 dB and 18dB are shown in Figure 2. Overall, we can see that the SNR-Temporal mapping performs better (in the PSNR sense) than the Temporal-SNR mapping.

Similarly, in Figure 3, we show the average PSNR value comparison of the SNR-Temporal and Temporal-SNR mappings for the "Akiyo" sequence. The PSNR results are obtained after the optimal parameter selection for a channel SNR of 8dB, 10db, 12dB and 16dB. However, we can clearly see that the behavior (in the PSNR sense) for a low motion sequence is opposite compared to the previous case, i.e., the Temporal-SNR mapping performs better than the SNR-Temporal mapping.

8 Conclusions

We have proposed a novel cross-layer optimization scheme for wireless video transmission over MIMO-OFDM channels. The scalable extension of H.264 is used for source coding and the compressed video is divided into scalable layers. For each of these scalable layers, the cross-layer optimization scheme determines the quantization parameter, channel coding rate, and symbol constellation. In order to carry out the optimization, an accurate estimation of the expected video distortion at the receiver is required. We have developed two expected distortion estimation methods, the Temporal-SNR SDDE method and the SNR-Temporal SDDE method. These methods differ in the priority order in which temporal and SNR scalability are considered. We have also proposed two different priority mappings of the scalable layers, the Temporal-SNR mapping and the SNR-Temporal mapping. We have presented experimental results that show the outcome of the cross-layer optimization using both distortion estimation methods. The SNR-Temporal mapping performs better for high-motion video sequences, while the Temporal-SNR mapping performs better for low-motion video sequences.

References

1. Wang, H., Kondi, L.P., Luthra, A., Ci, S.: 4G Wireless Video Communications. John Wiley and Sons, Ltd., Chichester (2009)
2. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE Journal on Selected Areas in Communications 16(8), 1451–1458 (1998)
3. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space-time block codes from orthogonal designs. IEEE Transactions on Information Theory 45(5), 1456–1467 (1999)
4. Zheng, H., Liu, K.J.R.: Robust image and video tansmission over spectrally shaped channels using multicanier modulation. IEEE Transactions on Multimedia (March 1999)
5. Kuo, C., Kim, C., Kuo, C.C.J.: Robust video tansmission over wideband wireless channel using space-time coded OFDM systems. In: Proc. WCNC, vol. 2 (March 2002)
6. Zhang, H., Xia, X.-G., Zhang, Q., Zhu, W.: Precoded OFDM with adaptive vector channel allocation for scalable video transmission over frequency-selective fading channels. IEEE Trans. Mobile Computing 1(2) (June 2002)
7. Kuo, C., Kim, C., Kuo, C.-C.J.: Robust video transmission over wideband wireless channel using space-time coded OFDM systems. In: Proc. IEEE Wireless Comm. and Networking Conf., WCNC 2002 (March 2002)
8. Dardari, D., Martini, M.G., Milantoni, M., Chiani, M.: MPEG-4 video transmission in the 5Ghz band through an adaptive ofdm wireless scheme. In: Proc. 13th IEEE Intl Symp. Personal Indoor, and Mobile Radio Comm., vol. 4 (2002)
9. Zhao, S., Xiong, Z., Wang, X.: Optimal resource allocation for wireless video over CDMA networks. IEEE Trans. Mobile Computing 4(1) (January 2005)
10. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Transactions on Circuits and Systems for Video Technology 17(9), 1103–1120 (2007)

11. Jubran, M.K., Bansal, M., Kondi, L.P.: Low-delay low-complexity bandwidth-constrained wireless video transmission using SVC over MIMO systems. IEEE Transactions on Multimedia 10(8), 1698–1707 (2008)
12. Jubran, M.K., Bansal, M., Kondi, L.P., Grover, R.: Accurate distortion estimation and optimal bandwidth allocation for scalable H.264 video transmission over MIMO systems. IEEE Transactions on Image Processing 18(1), 106–116 (2009)
13. Zhang, R., Regunathan, S.L., Rose, K.: Video coding with optimal inter/intra-mode switching for packet loss resilience. IEEE Journal on Selected Areas in Communications 18(6), 966–976 (2000)