Video Motion Activity Transmission with Redundant Antennas at the Transmitter Side

Martin B. Obando, Walter C. Freitas Jr and Francisco R. P. Cavalcanti
Federal University of Ceará, Wireless Telecommunications Research Group (GTEL)
CP 6005, Campus do Pici, 60455-760, Fortaleza-CE, Brazil
Email:{martin, walter, rodrigo}@gtel.ufc.br

Abstract—**In this paper, we propose a real-time video motion activity transmission over Multiple-Input Multiple-Output (MIMO) system. An interchange of information between the application and physical layers is realized. In the application layer, a Sum of Absolute Differences (SAD) measurement is used to estimate the variance between consecutive frames in the video sequence, which defines a threshold for Unequal Error Protection (UEP). In the physical layer, an Orthogonal Space-Time Block Code (OSTBC) MIMO system with redundant antennas is used to exploit diversity gain using Antenna Selection (AS) at the transmitter side. The simulation results presented in this paper show that the transmission of video based in the motion activity detection per frame using the diversity scheme with AS is capable of improving the objective and subjective video quality at the end user.**

I. INTRODUCTION

In recent years, the transmission of video over wireless communication systems and their integration with the Internet is a subject of major interest. This is employed in the IP Multimedia Subsystem (IMS) that is standardized by the 3rd. Generation Partnership Project (3GPP) and consists in a system of services for next generation networks [1].

The development of the video applications and services is an opportunity where multimedia applications would bring in large volume of data transmissions onto the wireless networks. This can become a potential challenge for Fourth Generation (4G) systems [2].

Errors which occur in data transmission, through wireless channels, are mainly caused by propagation mechanisms, such as long-term and short-term fading which degrade the quality of the received signal, leading to packet loss.

The cooperation between the different layers that compose the protocol stack of wireless communication networks is an alternative that can be used in order to improve the transmission of multimedia services in wireless systems [2]. In the literature, there are several works that use this mechanism to improve the quality of multimedia services, such as video.

In [3] the authors propose a Hybrid MIMO Structure (HMS) to implement UEP for video delivery in MIMO system. The goal of that work was to exploit the diversity gain to provide better protection to the high priority data, while transmitting the low priority data with spatial multiplexing, to achieve high data rate.

978-1-4244-7173-7/10/$26.00 2010 IEEE

In [4], a layered video transmission scheme over Multiple-Input Multiple-Output (MIMO) Adaptive Channel Selection (ACS) was proposed. In the application layer, Scalable Video Coding (SVC) generates layered bit streams that need prioritized delivery. In the physical layer, the ordering of each subchannel's Signal-to-Noise Ratio (SNR) is done as partial Channel Information (CI) at the receiver, which is acquired via the Channel State Information (CSI), based on training sequences. The strategy used in that paper consists in switching the bitstream automatically to match the ordering of SNR strength for the subchannels achieving Unequal Error Protection (UEP).

A power allocation and Antenna Selection (AS) scheme for scalable video delivery transmission had been proposed in [5], where the power allocation and AS among the multiple layers is subject to constraints on the total transmission rate and the total power level. As the wireless channel condition changes, the authors propose to scale the video streams and to transport the scaled video streams to receivers with a smooth change of perceptual quality.

In [6], the authors propose a cross-layer method for SVC and MIMO transmission using CSI. In this work, the channels with unequal Bit Error Rate (BER) are created for transferring the parts of the scalable video stream that are of unequal importance. The quality of the channel is then matched to the importance of video sub-stream being transferred.

In this work, we propose an algorithm that uses the variance of the Sum of Absolute Differences (SAD) between consecutive frames to switch between schemes of diversity gain, whereas for frames with strong motion activity, a UEP scheme using AS at the transmitter side is configured. The difference between the works mentioned above and our proposed algorithm is that in those, a UEP scheme is configured based in the SVC and transmission rate. On the other hand, in our work, a measurement of variance based in the SAD of consecutive frames is used to configure a UEP scheme. The main contribution of our proposal is to exploit the motion activity of consecutive frames and AS scheme to get UEP in the transmission. Furthermore, the algorithm requires a simple implementation and small amount of feedback.

The remainder of the paper is organized as follows: in section II, we present the system and channel model assumed. In section III, a video coding and video sequence classification background is made. Section IV presents a particular Orthogonal Space-Time Block Code (OSTBC) MIMO system and antenna selection at the transmitter side overview. In section V, we present the proposed algorithm. Finally, section VI presents the performance results and section VII concludes this paper.

II. CHANNEL MODEL

We consider wireless transceivers equipped with M_{tx} transmit antennas and M_{rx} receive antennas. The wireless channel is assumed to have rich scattering and flat fading that could be achieved with Orthogonal Frequency Division Multiplexing (OFDM), even for a high demanding symbol rate, as well as for video transmission. The channel is quasi-static, i.e., the channel does not change significantly in a single block, but can vary from block to block. In each block we can represent the sampled received signal as in [7]:

$$\mathbf{Y} = \sqrt{\frac{\rho}{M_{\mathrm{tx}}}}\mathbf{HS} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{C}^{M_{\mathrm{rx}} \times T}$ is the received signal matrix, T is the number of signaling intervals, $\mathbf{S} \in \mathbb{C}^{M_{\mathrm{tx}} \times T}$ is the transmitted signal matrix, ρ is the average signal to noise ratio at each receiver antenna, $\mathbf{H} \in \mathbb{C}^{M_{\mathrm{rx}} \times M_{\mathrm{tx}}}$ is the random channel matrix and $\mathbf{N} \in \mathbb{C}^{M_{\mathrm{rx}} \times T}$ is the additive noise matrix, whose entries are i.i.d. Zero Mean Circularly Symmetric Complex Gaussian (ZMCSCG).

III. VIDEO CODING AND VIDEO SEQUENCE CLASSIFICATION BACKGROUND

In video coding, the most important standards such as the H.264/AVC and MPEG-4 [2] define three main frames types for the compressive video streams; including the I frame (intra-coded), P frame (predictive-coded) and B frame (bidirectional predictive-coded). The I frames are encoded independently and decoded by themselves. The P frames are encoded using predictions from the preceding I or P frames in the video sequence. The B frames are encoded using predictions from the preceding and the subsequent I or P frames.

The Video Motion Activity Detection (VMAD) is a way of defining the activity in a scene by analyzing data frame (i.e., in this case, a frame represents a picture of the video sequence) and differences in a series of frames. The SAD can reflect the degree of the motion activity very well, where SAD value of the current macroblock (i.e., the basic unit that composes each frame in the video coding process) in video sequence is bigger than the other ones. Thus, the video sequences can be classified as slow motion activity, median motion activity and strong motion activity, following the thresholds Th_1=1300 and Th_2=4000 given in [8]. The Motion Activity (MA) function can be represented by

$$MA = \begin{cases} \text{slow motion}, & \text{if} \quad \text{meanSAD} < \mathrm{Th}_1 \\ \text{median motion}, & \text{if} \quad \mathrm{Th}_1 \leq \text{meanSAD} \leq \mathrm{Th}_2 \\ \text{strong motion}, & \text{if} \quad \text{meanSAD} > \mathrm{Th}_2 \end{cases} \tag{2}$$

We estimated the mean SAD value of nine video sequences, which are shown in Table I.

TABLE I
TEST VIDEO SEQUENCES MEAN SAD VALUES.

Video sequence	mean SAD
Akiyo	153
Bus	5964
Carphone	1406
Claire	258
America Football	5809
Foreman	2063
Grandma	329
Ice	2921
Soccer	5035

Motion activity may describe several attributes that contribute towards the efficient use of these motion descriptors in a number of applications. In this paper, we consider the motion descriptors by frame, in spatial and temporal distribution of activity.

In Figure 1, it is shown an example of the motion activity detection by frame used in this work, where the frame 66 in the video sequence "American Football" presents a strong motion in relation to its anterior and posterior frame.

(a) Frame 65 (b) Frame 66 (c) Frame 67

Fig. 1. Motion activity detection example frames 65, 66 and 67 in the video sequence "American Football".

IV. MIMO SYSTEMS AND ANTENNA SELECTION OVERVIEW

The use of MIMO systems can generally be exploited to achieve two main objectives. The first one is link reliability (i.e., the diversity gain) using for example OSTBC schemes, in which the same symbol stream is transmitted from different transmit antennas in an appropriate manner, to obtain transmit diversity [9]. The second option is to obtain link spectrum efficiency (i.e., spatial multiplexing gain) using for example Vertical Bell Laboratories Layered Space-Time Architecture (VBLAST) schemes, in which different symbol streams are simultaneously transmitted from all transmit antennas [10].

Another way to exploit the MIMO systems is using the Hybrid MIMO Structure (HMS), where the transmission process of a HMS can be divided in layers, somewhat like VBLAST. However, in contrast to VBLAST, in the HMS case, a layer may consist of the stream of symbols at the output of an OSTBC, which is sent to a group of antennas, or of an uncoded stream, which is transmitted from a single antenna.

Based on this concept of layers, the HMS combines pure diversity schemes (e.g. OSTBC) with pure spatial multiplexing schemes (e.g. VBLAST) [11]. In the following, we explain a well known MIMO structure to obtain link reliability, denoted as Alamouti OSTBC.

A. Alamouti OSTBC Structure

This structure consists of two antennas, whose configuration corresponds to an OSTBC known as (G2) [9] and provides diversity gain. The transmission matrix $\mathbf{S}$ of this structure is given by

$$\mathbf{S}[T_1, T_2] = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}, \tag{3}$$

in which $K = 2$ symbols (i.e. s_1 and s_2) are transmitted in $T = 2$ (i.e. T_1 and T_2) consecutive signalling intervals. Therefore, its effective spectral efficiency is given by:

$$\eta = (K/T) \cdot \log_2(\mu) = \log_2(\mu), \tag{4}$$

where μ is considered the cardinality of the modulation scheme.

B. Antenna Selection at the Transmitter Side

One of the drawbacks of using multiple antennas is the complexity and cost that arises from using a separate Radio Frequency (RF) chain for every employed antenna. Thus, AS is a means to alleviate this complexity, while exploiting the advantages provided by multiples transmit and receive antennas [12].

The AS scheme consists in the use of a subset of available antennas in a MIMO system, where the number of RF chains required is reduced to as few as the number of selected antennas, becoming less expensive and more feasible. The AS scheme can be performed at the transmitter side, the receiver side, or both simultaneously.

In this work, we used the AS at the transmitter side considering perfect CSI available at the receiver side. As a performance analysis at the receiver the CSI is given by

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{\mathrm{m_{tx}}} \dots, \mathbf{h}_{M_{\mathrm{tx}}}]_{M_{\mathrm{rx}} \times M_{\mathrm{tx}}}, \tag{5}$$

where $\mathbf{h}_{\mathrm{m_{tx}}}$ is one of the channel matrix columns, representing the gain coefficients vector from the mth transmit antenna to all the $\mathrm{M_{rx}}$ antennas.

The receiver seeks the best $\mathrm{L_{tx}}$ gains out of all the $\mathrm{M_{tx}}$ candidates by calculating the Frobenius norm for the columns of the channel matrix, each given as

$$||\mathbf{h}_{\mathrm{m_{tx}}}||_F^2 = \sum_{m_{\mathrm{rx}}=1}^{\mathrm{M_{rx}}} |\mathbf{h}_{m_{\mathrm{rx}} m_{\mathrm{tx}}}|^2, (1 \leq m_{\mathrm{tx}} \leq \mathrm{M_{tx}}). \tag{6}$$

The selected antennas are those with the $\mathrm{L_{tx}}$ large Frobenius norm. The new MIMO wireless link is formed by the selected transmit antennas, the channel matrix $\mathbf{H}$ for the new MIMO link can be expressed as

$$\mathbf{H}' = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{L_{\mathrm{tx}}}]_{\mathrm{M_{rx}} \times L_{\mathrm{tx}}} \tag{7}$$

The ($\mathrm{M_{tx}}$–$\mathrm{L_{tx}}$) antennas with smaller Frobenius norm are regarded as spatially redundant and thus are deactivated during transmitting the bitstream as in [13].

Algorithm 1 Proposed Algorithm

```
iniScheme(TxG2Scheme);
Threshold ← A;
k ← 0;
varianceSAD ← 0;
for all VideoSequence do
    Frame ← videoSequence(k);
    bitstream ← videoEncoder(Frame);
    if k > 1 then
        vectorSAD(k-1) ← calSAD(Mframe,Frame);
        meanSAD ← calMeanSAD(vectorSAD);
        varianceSAD ← calVarSAD(meanSAD,vectorSAD);
    else
        if k = 1 then
            vectorSAD(k-1) ← calSAD(Mframe,Frame);
        end if
    end if
    if varianceSAD < Threshold then
        TxG2Scheme(bitstream);
    else
        TxASG2Scheme ← AS(TxG2Scheme);
        TxASG2Scheme(bitstream);
    end if
    MFrame ← Frame;
    k ← k+1;
end for
```

V. PROPOSED ALGORITHM FOR VIDEO TRANSMISSION

In the following, we present the algorithm for video transmission in wireless systems, and its explanation step by step.

At the beginning, the G2 MIMO scheme without AS by default is configured and a threshold to detect the strong motion activity by frames in the video sequence is initialized. In this work, the threshold (i.e. Threshold) has a value of 11 which is assumed following the 30% of frequency in video sequences of median motion activity and the 50% of frequency in video sequences of strong motion activity. A counter of each frame in the video sequence (i.e., k) and a variance of SAD between forward consecutive frames (i.e., varianceSAD) are initialized with 0.

Each frame of the input video sequence (i.e., Frame) is encoded giving as result a bitstream of video. To estimate the SAD, two frames are necessary, thus, if k=1 the first SAD is estimated and the result is assigned to the vector that saves the SAD parameter (i.e., vectorSAD). Then, for k>1 the SAD, mean of SAD (i.e., meanSAD), and variance of the SAD can be estimated for the following frames.

If the variance of SAD given in logarithmic scale is lower than the threshold defined for switching between the G2 MIMO structure and G2 MIMO structure with AS, then, the G2 MIMO structure (i.e., TxG2Scheme()) is used for bitstream transmission. On the other hand, G2 with AS (i.e., TxASG2Scheme()) is used for bit-stream transmission. The last memory frame (i.e., MFrame) is actualized with the current frame, as well as the counter frame. This way, the remaining frames in the video sequences are processed for their transmission.

A summary of the proposed algorithm is presented in Algorithm 1.

VI. RESULTS

In this section, the performance of the proposed algorithm is evaluated by using Monte Carlo simulation.

In order to assess the performance of our proposed algorithm we simulated the G2 MIMO structure without AS and the cross-layer algorithm proposed in [6], which we identified in this work as G2 and SVC with AS, respectively.

We considered a turbo code with data rate (R_c=1/3), M_{tx}=3 antennas at the transmitter side and M_{rx}=2 antennas at the receiver side. We considered 16 Quadrature Amplitude Modulation (16QAM) signal constellation for our proposed algorithm and G2 MIMO scheme, while for SVC with AS scheme we considered a VBLAST MIMO scheme with $T = 2$ and Quadri-Phase Shift Keying (QPSK) signal constellation. This is done with the purpose of keeping an equal spectral efficiency in the schemes.

The video coder used was the H.264/AVC baseline profile from [14]. The video encoded bit rate is 64 kbps with Group of Pictures (GOP) structure of eighteen frames in the form (IPPPPPPPPPPPPPPPPP). The video sequences used in the simulation were: “Foreman” and “Carphone” classified as median motion activity and “American Football” and “Soccer” classified as strong motion activity from [15].

The video sequences of median motion activity consist of 240 frames each one and the video sequences of strong motion activity consist of 120 frames each one. The format of the video sequences is Quarter Common Intermediate Format (QCIF) of size 176×144 pixels per frame with 4:2:0 sampling. A packet size of 250 bytes was assumed which is used in video telephone service.

The objective and subjective assessment is realized, following some mechanisms studied in [16].

Fig. 2 and 3 show the occurrence frequency of the variance over different video sequences. They are classified into median motion activity and strong motion activity. The threshold 11 was assumed, which represents 30% of frames protected for median motion activity video sequences, approximately, and 50% of frames protected for strong motion activity video sequences, approximately.

The threshold 11 is assumed due to that part of the frames that compose the video sequences, which will be protected to improve the objective performance in the video transmission. We can observe that a higher percentage of frames in the strong motion activity video sequences is protected, in comparison with the median motion activity video sequences. This is due to the fact that strong motion video sequences are more vulnerable to the packet loss, in our case.

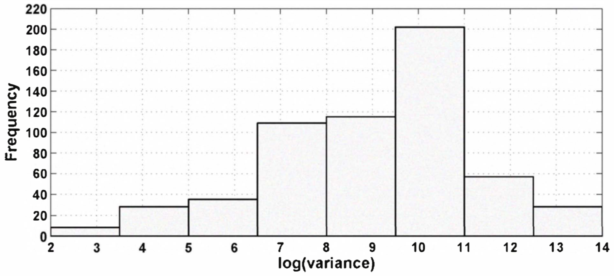

Fig. 2. Frequency vs Variance in logarithmic scale of median motion activity video sequences.

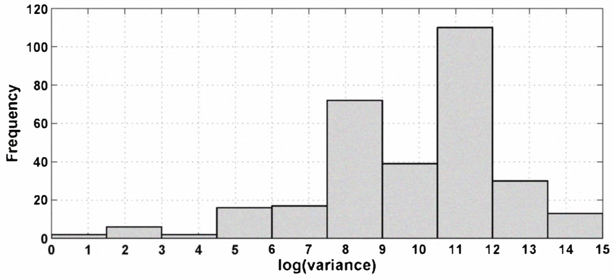

Fig. 3. Frequency vs Variance in logarithmic scale of strong motion activity video sequences.

Fig. 4 shows the performance of the BER for both the G2 MIMO structure with AS and the G2 MIMO structure without AS, where for G2 MIMO with AS we have M_{tx}=3 antennas (i.e., redundant antennas) which used L_{tx}=2 only. We can observe that the G2 MIMO with AS scheme presents better performance in comparison with the G2 MIMO structure without AS. That means that a higher spatial redundancy brings a gain in diversity order, where the diversity gain comes from both the optimal transmit antennas selection and the spatial redundant cancellation.

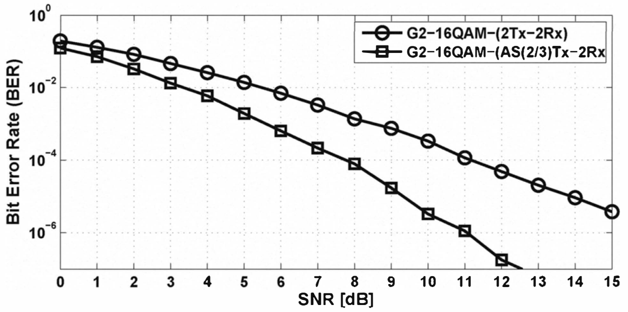

Fig. 4. BER performance of the G2 MIMO structure with and without AS at the transmitter side.

Fig. 5 shows that the proposed algorithm outperforms the G2 MIMO without AS scheme and the SVC with AS scheme in Peak Signal-to-Noise Ratio (PSNR) performance. The proposed algorithm outperforms the PSNR of the G2 structure without AS in 1.5 dB from a SNR=9 dB to SNR=12 dB approximately, for SNR>12 dB , as the SNR increases, the gain performance is reduced. It is due that to same SNR the proposed algorithm gets better diversity gain by using redundant antennas with AS at the transmitter side.

The result obtained with the SVC with AS is due to the UEP scheme where the priority bitstreams are transmitted using better subchannels, while on the other hand, the remaining subchannels present a poor power performance for bitstreams transmission. Thus, the video quality of the received video is degraded. We can observe that with the increase of SNR, it is possible to get a better video quality for the SVC with AS.

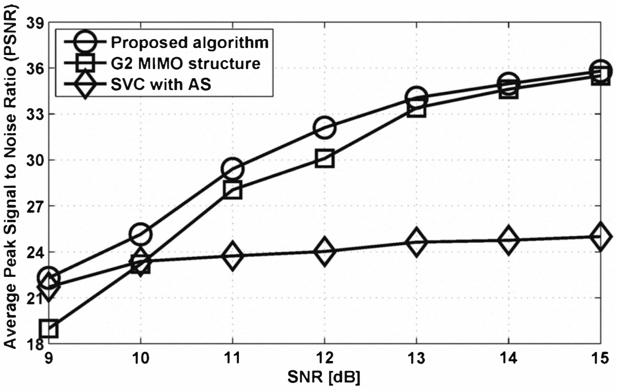

Fig. 5. Average PSNR performance of median motion activity video sequences.

Fig. 6 shows that the proposed algorithm outperforms the PSNR performance of both the G2 MIMO structure and SVC with AS. We can compare the results obtained in Figures 5 and 6, where we can observe that the proposed algorithm gets a better result when transmitting video sequences, classified as strong motion activity which is an interesting result for real time video services transmission.

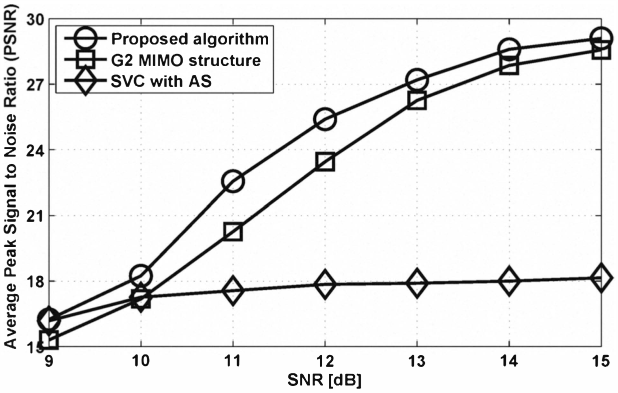

Fig. 6. Average PSNR performance of strong motion activity video sequences.

Table II presents the average of 15 Human Visual System (HVS) quality assessment for each transmitted video sequences, which consists in presenting four different video sequences of the same video, using a web page. The presented videos consist in the original video sequence, the video sequences transmitted by our proposed algorithm, G2 without AS and SVC with AS. The web page used for this subjective evaluation can be found in [17]. Therefore, the subjective assessment can be considered as a mechanism for verifying the effectiveness of the proposed algorithm.

TABLE II
SUBJECTIVE PERFORMANCE

Video Sequence	Proposed algorithm	G2 Structure	SVC with AS
Median motion activity SNR=12 dB			
Carphone	3.2	2.8	2.3
Foreman	3.1	2.0	1.7
Strong motion activity SNR=14 dB			
American Football	2.6	1.5	1.1
Soccer	2.4	1.9	1.2

The results in Table II show that the proposed algorithm outperforms the subjective quality in comparison with the use of the other schemes. For video sequences with strong motion activity, it was selected a SNR=14 dB to get better quality performance for assessment, due to the high degradation to lower SNR. The results obtained in this evaluation show that when video sequences of strong motion activity are transmitted using the proposed algorithm, they get better performance than when video sequences of median motion activity are transmitted.

In Fig. 7 we present an example, where we can observe that the proposed algorithm outperforms the G2 MIMO structure without AS and the SVC with AS scheme. In this example, the original and degrade frames are present when they are transmitted using the different schemes implemented in this work. We can observe that the proposed algorithm gets to control the distortion inserted during the transmission better than the other schemes. This is due to the frame protection, by using the proposed scheme. It is important to say that the quality of the video sequences used is lower, due to the lower bit rate of codification and different conversion formats used for subjective evaluation assessment.

VII. CONCLUSION

In this paper, we proposed a real-time video motion activity transmission MIMO system, with redundant antennas. The strategy used consists in identifying the frames with strong motion activity, using the variance SAD between consecutive frames, to achieve UEP during their transmission. The UEP is got with the use of AS at the transmitter side. Simulation results show that the objective and subjective video quality is improved using this scheme. Therefore, the proposed algorithm is a feasible scheme that can be adopted for video transmission over wireless systems.

ACKNOWLEDGMENT

Martin B. Obando is supported by The National Council for Scientific and Technological Development (CNPq).

(a) Original (b) Proposed algorithm (c) G2 Structure (d) SVC with AS

(e) Original (f) Proposed algorithm (g) G2 Structure (h) SVC with AS

Fig. 7. Examples of subjective quality of the frame 111 and 101 from the video sequences “Foreman” and “American Football” transmitted with a SNR = 12dB, respectively.

REFERENCES

[1] S. Chakraborty., et al., *IMS multimedia Telephony over Cellular Systems*, John Wiley & Sons, Ltd, 2007.

[2] H. Wang, L. Condi, A. Luthra and S. Ci, *4G Wireless Video Communication*, Wiley series on Wireless Communication and Mobile Computing, 2009.

[3] Y. Qian et al., “Unequal Error Protection Scheme for Multiple-Input and Multiple-Output Wireless Systems,” *IEEE Vehicular Technology Conference, 2006. VTC-2006 Fall*, pp. 1–5, 2006.

[4] D. Song et al., “Scalable H.264/AVC video Transmission Over Wireless Systems with Adaptive Channel Selection Based on Partial Channel Information,” *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 17, pp. 1218–1226, 2007.

[5] S. Zhao et al., “Scalable Video Delivery over MIMO OFDM Wireless Systems Using Joint Power Allocation and Antenna Selection,” *Multimedia Signal Processing. IEEE International Workshop on*, pp. 1–6, 2009.

[6] D. Radakodiv et al., “Priority-aware transfer of SVC encoded video over MIMO communications system ,” *IEEE Picture Coding Symposium, 2009. PCS 2009*, pp. 1–4, 2009.

[7] A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communication*, Cambridge University Press, 2003.

[8] H. Weiwei, et al., “H.264 motion estimation algorithm based on video sequences activity,” *Journal of Electronics (China)*, vol. 25, pp.125–128, January 16, 2008.

[9] S.M. Alamouti, “A Simple Transmit Diversity Technique for Wireless Communications,” *IEEE Journal, Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.

[10] M. Haideh, et al., “Spatial Multiplexing and Diversity Gain in OFDM-Based MIMO Systems”, IEEE Topical Conference on Wireless Communication Technology, pp.299-301, 2003.

[11] W. C. Freitas, et al., “Hybrid Transceiver Schemes for Spatial Multiplexing and Diversity in MIMO Systems,” *Journal of Communication and Information Systems*, vol.20, pp. 63–76, 2005.

[12] Byung Joon Oh et al., “A Survey on Antenna Selection for MIMO Communication Systems,” *IEEE* , vol. 16, pp. 1451–1458, October 2008.

[13] F. Wang et al., “Space Time Block Coded MIMO System with Redundant Antennas ,” *IEEE Conferences*, vol. 2, pp. 204–207, 2009.

[14] H.264/AVC, *JM Reference Software*, Download:“http://iphome.hhi.de/suehring/tml/download,” 2010.

[15] YUV video sequences:“http://trace.eas.asu.edu/yuv/index.html”, 2010.

[16] S. Winkler, et al., “The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics,” *IEEE Transactions on Broadcasting*, vol.54, pp. 660–668, September 2008.

[17] Subjective evaluation:“http://www.gtel.ufc.br/∼martin”, 2010.