# SCALABLE H.264/AVC VIDEO TRANSMISSION IN MIMO SYSTEM WITH CONTENT AWARE PACKET IMPORTANCE ORDERING


*Tuo Huang*

Department of EECS
Northwestern University



## ABSTRACT

In this paper, we propose a content-aware scalable video transmission scheme over time-varying multiple-input multi-output (MIMO) wireless channels. At transmitter, distortion is calculated based on both EL truncation and drift. We then order the packets according to their importance scores. Each packet has an importance score calculated from the expected distortion if the packet is lost during transmission. The MIMO channel model used in the simulation is frequency non-selective and Rayleigh-fading. The channels are decomposed into parallel single-input single-output (SISO) channels using estimated channel state information (CSI). Ordered packets in buffer are allocated to sub-channels to achieve minimum delay. Also, channel protection (redundancy) codes are added into the packets to guarantee target packet loss rate. Under constrained bandwidth and time-varying channel states, this scheme outperforms existing MIMO video transmission schemes and guarantees high-quality video transmission.



***Index Terms—*** Spatial Multiplexing, MIMO, Content Aware, Scalable H.264/AVC, Robust Transmission


## 1. INTRODUCTION

Delivering video data to mobile user is challenging due to limited channel bandwidth and unpredictability of wireless propagation channel. In this paper, we combine the efficient scalable H.264 AVC codec and MIMO system to ensure robust video streaming in a mobile environment. With H.264/SVC, video sub streams can be efficiently extracted to fit the needs of the channel, when fading is severe. Also, in MIMO wireless systems two approaches may be utilized to increase channel capacity : spatial multiplexing and spatial diversity. With spatial multiplexing [1], orthogonal data streams are transmitted across different channels to increase data rate and channel state information (CSI) is required to calculate precoder and decoder matrices. With spatial diversity [2], bit error rate can be reduced with space time block code. In this paper we consider spatial multiplexing MIMO, where CSI is available. With this architecture, channel capacity can be maximized with water-filling (WF) solution [3]. By WF algorithm, optimal power is allocated to each sub-channel and short-term SNR of sub-channel is available to transmitter. With sub-channel SNR,, which adds channel coding protection to transmission blocks can be performed toin order to achieve a target packet loss rate.

Several approaches to transmit video data in MIMO have been developed. In [4],spatial time block codes are used to reduce bit

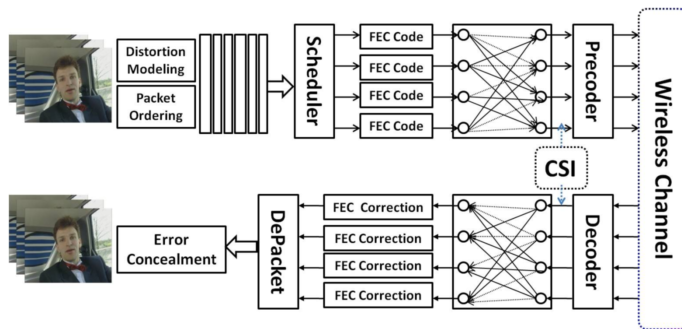


**Fig. 1**. Proposed MIMO video streaming framework.

error rate to provide robust video communication. In [5], an adaptive channel selection system is proposed to periodically launch the bitstream for each layer to the proper sub-channel according to the importance of layers. In [6], adaptive modulation is performed in each session to maximize throughput.

In the proposed approach, we use optimized bit extraction [7] [8] to obtain importance ordering based on both EL truncation and drift distortion. Meanwhile, with CSI, MIMO channels are decomposed into parallel SISO channels. Each Video packet is scheduled to be transmitted over different sub-channels according to its distortion-based importance score.

## 2. SYSTEM ARCHITECTURE

The proposed framework is described in Fig. 1 for a 4 $\times$ 4 MIMO System. At the transmitter, raw video frames are encoded with scalable H.264/AVC into 5 layers. The scalable bit stream is analyzed with distortion model, which assigns an importance score to each packet based on the expected distortion in the reconstructed video stream at the receiver when that packet is lost. That is, if the loss of a packet is expected to result in a larger distortion at the receiver, its importance score will be higher. The video packets are ordered in a manner to be described in Section 3.

The MIMO channels are decomposed into four parallel sub-channels based on CSI feedback. Then, the scheduler allocates the ordered packets into different channels based on the optimization algorithm described in Section 5. Then, Forward Error Correction (FEC) codes are added into those selected packets to guarantee a target loss rate. The target loss rate for each packet is again determined by its relative importance score: the more important (resulting in larger expected distortion) a packet is, the lower its target loss rate will be set. At the receiver side, the received packets are corrected with FEC error detection and the bit stream is decoded.

## 3. PACKET IMPORTANCE ORDERING

For accurate distortion estimates, we use the method developed by Ehsan Maani and Aggelos K. Katsaggelos in [7]. This bit extraction approach considers distortion caused by EL truncation and error drift.

### 3.1. Decoder Distortion Estimation for H.264 SVC

The total distortion of frame $f_n$ due to drift and EL truncation is obtained according to

$$\begin{aligned} D_n^t(q) &= \parallel f_n - f_n^d + e_n(q) \parallel^2 & (1)\\ &= D_n^d + D_n^e(q) + 2(f_n - f_n^d)^T e_n(q) & (2)\\ &\approx D_n^d + D_n^e(q) + 2k\sqrt{D_n^d}\sqrt{D_n^e(q)} & (3)\\ &\leq D_n^d + D_n^e(q) + 2\sqrt{D_n^d}\sqrt{D_n^e(q)} & (4)\end{aligned}$$

where $D_n^d$ and $D_n^e(q)$ represent respectively the distortion, i.e., sum of squared errors (SSE), due to drift and EL truncation and $k$ is a constant in the range $0 \leq k \leq 1$.

The drift distortion of frame $f_n$ is computed with parent child relationships.

$$D_n^d \approx \gamma + \sum_{i \in \Lambda_n} \alpha_i D_i^t + \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n} \beta_{ij} D_i^t D_j^t \qquad (5)$$

where $\Lambda_n$ represents the set of parent frames associated with frame $f_n$

### 3.2. Packet Importance Ordering

With this formulation the total distortion for each frame in the Group of Pictures (GOP) and thus the packet importance ordering are calculated.

Packet importance data are represented with a triplet $\pi(i, n_i, d_i)$, where $i$ is the order of importance, $n_i$ is the packet number in the GOP and $d_i$ is the distortion decrease when packet $n_i$ is included. Packet importance graph is accumulate graph of function $\pi(i, n_i, d_i)$.

The selection process is based on Hill algorithm. The base layers of the key pictures are given the highest priority and, therefore, are the first packets to be included in the importance ordering. Then, packets are added to the transmission queue one at a time based on their global distortion gradient. In other words, initially, the selection function $\phi(n) = 1$ if $s_n$ is a key frame, otherwise $\phi(n) = 0$ . Then, at each time step $i$, a packet is added $\pi(i, n_i^*, d_i^*)$ and $\phi(n_i^*)$ is incremented by one where is obtained by

$$n_i^* = \arg\max_n \mid \frac{\frac{\partial D(\phi)}{\partial \phi(n)}}{\frac{\partial R_s(\phi)}{\partial \phi(n)}} \mid \qquad (6)$$

Here, $R_s(\phi)$ represents the source rate associated with the current selection function $\phi$. This process continues until all available packets within the optimization window (i.e., GOPs) are added to the transmission queue. The follwing GOPs are then processed similarly.

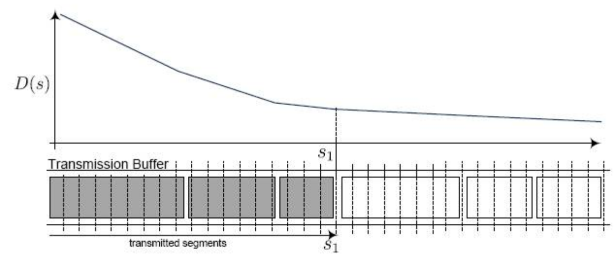


**Fig. 2**. Packet Importance Ordering

## 4. MIMO WIRELESS SYSTEMS

In this architecture, we mainly consider frequency non-selective and Rayleigh-fading MIMO channel. For a MIMO system with $M_T$ transmitter antennas and $M_R$ receiver antennas, the system equation is

$$y = Hx + n \qquad (7)$$

where $y$ is $M_R \times 1$ vector of the received signal, $x$ is the $M_T \times 1$ vector of the transmitted signal, and $n$ is the $M_R \times 1$ vector of additive white Gaussian noise with variance $\sigma^2$. $H$ is the channel matrix obtained from CSI feedback.

### 4.1. MIMO Wireless channel decomposition

With singular value decomposition, MIMO channel matrix H can be decomposed as

$$H = WDV^* \qquad (8)$$

where $W$ and $V^*$ are unitary, orthogonal matrix and $D$ is diagonal matrix, $D = Diag(\sqrt{\lambda_1}, \sqrt{\lambda_2} ... \sqrt{\lambda_m})$ The singular value is always positive and is regarded as the gain for the sub-channel.

A sample decomposition of a $4 \times 4$ channel is $\lambda_1$ = 2.6881, $\lambda_2$ = 1.3597, $\lambda_3$ = 0.7912, $\lambda_4$ = 0.3840. For the four sub-channel, channel is enhanced for the first two sub-channel and is faded for the last two sub-channel. SNR of i-th sub-channel can be written as

$$SNR_i = \rho_i \lambda_i \qquad (9)$$

where $\rho_i$ is the power allocated to the $i$th subchannel using Water Filling (WF) algorithm. The WF algorithm also returns the total number of usable subchannels.

Let $\tilde{y} = W^* y$, $\tilde{x} = V^* x$, and $\tilde{n} = W^* n$, (1) could be written as

$$\tilde{y} = D\tilde{x} + \tilde{n} \qquad (10)$$

where D is a diagonal matrix. Thus, the MIMO channels are decomposed into independent SISO channels, which can be considered separately.

### 4.2. Bit Error Rate and Packet Loss Rate

In rectangular $M$-QAM modulation systems, $k = \log_2 M$ bits are transmitted by each symbol.

Thus, the SNR of a sub-channel can also be written as

$$\Gamma_i = \frac{E_b k}{N_0} \qquad (11)$$

where $E_b$ is the energy per bit and $N_0$ is the AWGN power density.

Using SNR1 and SNR2, the bit error rate of $M$-QAM is given approximately by

$$P_b \simeq \frac{2}{\log_2 \sqrt{M}}\left(1-\frac{1}{\sqrt{M}}\right)Q\left[\sqrt{S\frac{E_b}{N_0}}\right] \quad (12)$$

$$= \frac{2}{\log_2 \sqrt{M}}\left(1-\frac{1}{\sqrt{M}}\right)Q\left[\sqrt{S\frac{\rho_i \lambda_i}{N_0 k}}\right] \quad (13)$$

where $S = 3\log_2 \sqrt{M}/M - 1$

If an $(n, l)$ inter-package channel coding is used for a packet of length $l$ and is able to correct up to $t = Floor(\frac{n-l}{2})$ errors, the packet loss rate is given by

$$P_e = 1 - \sum_{i=0}^{t} C_i^n P_b^i (1-P_b)^{n-i} \quad (14)$$

## 5. SOLUTION ALGORITHM

In a GOP, the packet are ordered according to content importance as described in Sec. 3. One example of importance ordering is shown in Fig. 2.

It is obvious that, given importance scores calculated from expected distortion (reduction in PSNR) at reconstruction, it is less preferable to lose a high-importance packet than to losing a low-importance packet. Thus, if the packet is more important, we set a lower target loss rate for the packet, while higher target loss rates can be tolerated for less important packets. This mechanism is formulated with:

$$P_{loss} * Q = const \quad (15)$$

where $P_{loss}$ is the target loss rate for a packet and $Q$ is the importance score of the packet. Considering the effect of time delay and varying channel states, it is better to send higher-importance packets first because the channel state may change over time, resulting in inaccurate calculation precoder/decoder matrices, and also because the limited time slot/bandwidth may only allow some packets to be transmitted, not the others. With these two main objectives in mind, we propose an optimization framework that minimizes the total time slots in the usable sub-channels to send packets that satisfy target error rates.

The size of the ordered packages are represented as $l_1, l_2, ..., l_n$, where $n$ is the number of packets in the GOP. Using CSI feedback we calculate the bit error rate $P_{bj}$ for each subchannel using the method described in the previous section, where $j$ represents the $j$th sub-channel. We need to find the minimum bits of channel coding (redundancy) to guarantee the target loss rate $P_loss$.

$$\begin{aligned} \underset{i\in N_p, j\in N_c}{\text{minimize}} \quad & n_{ij} \\ \text{subject to} \quad & P_{ej} * Q_i <= const \\ & P_{ej} = 1 - \sum_{i=0}^{n_{ij}-l_i} C_i^{n_i} P_{bj}^i (1-P_{bj})^{n_{ij}-i} \end{aligned} \quad (16)$$

where $N_C$ and $N_p$ represent the number of channels and packets available, respectively.

At each time slot, buffer size of each channel is $buf_j$, which represents the bits that has not been transmitted over the channel. Optimal allocation of packet $i$ is accomplished with the solution below.

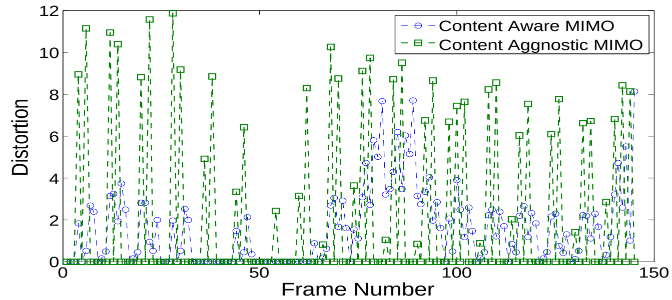


**Fig. 3**. Performance of content agnostic and content aware transmission scheme on Foreman CIF sequence at 250 kbps and $SNR = 18$.

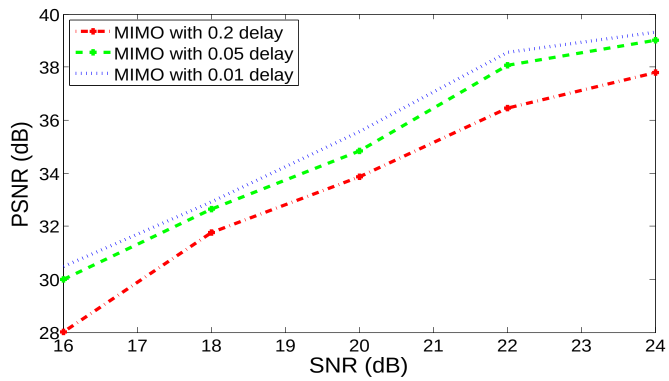


**Fig. 4**. Performance of content aware transmission scheme on Mobile CIF sequence at different feedback delay.

$$j^* = \arg\min_j (buf_j + n_{ij}) \quad (17)$$

Summary of proposed optimal allocation solution is described as:

1. At each time slot, get the bit error rate for each sub-channel $P_{bj}$ according to Equ. 14.
2. Calculate the number of parity bits to guarantee loss probability with Equ. 17.
3. Find optimal channel $j^*$ for packet $i$ with Equ. 18.

## 6. EXPERIMENT

Various CIF sequences at 30fps. The total power across all the channels is $P_t$. Using the MIMO model of Rayleigh-fading and frequency non-selection channels, the real and imaginary parts of each of the elements of the MIMO channel matrix H is obtained from an i.i.d. complex Gaussian distribution with zero mean, and variance 1/2.

Video sequences are encoded using JSVM 9.10 codec into two layers, a base layer and a quality enhancement layer, with basis quantization parameters $QP = 36$ and $QP = 24$, respectively. Furthermore, the enhancement layer is divided into 5 MGS layers.

First experiment is conducted to compare content-aware and content-agnostic transmission schemes. The experiment is done on

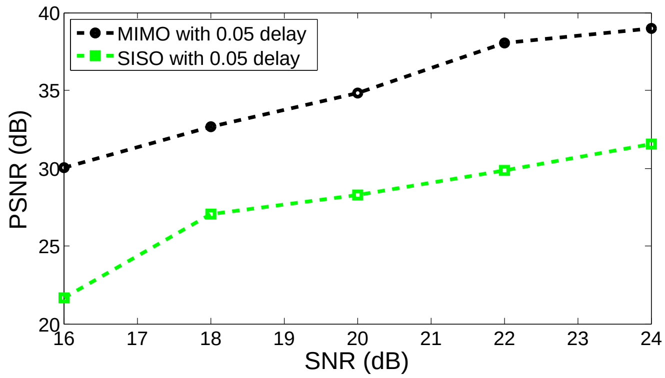


**Fig. 5**. Performance of content aware transmission scheme on MIMO and SISO at same feedback delay

Foreman CIF sequence at 250kbps data rate for each sub-channel and $SNR = 18$ for each subchannel. For agnostic transmission, we order the sequence by distortion calculated with only truncation error. As shown in Fig. 3, we can see content aware scheme outperform content agnostic scheme at almost every frame. The result can be explained that the distortion model in [7] is more accurate than using only truncation error.

The second experiment is conducted to investigate the influence of feedback delay on the transmission scheme. The experiment is done on Mobile CIF sequence at 250kbps data rate for each sub-channel at different SNRs. As shown in Fig. 4, we can see that the MIMO scheme is robust to feedback delay [9, 10, 11, 12, 13].

The third experiment is conducted to investigate the influence of feedback delay on MIMO and SISO[14, 15, 16, 17] systems. The experiment is done on Mobile CIF sequence at 250kbps data rate for each sub-channel at different SNR. As shown in Fig. 5, we can see that the MIMO [18, 19, 20, 21, 22, 23, **?**]scheme is more robust to feedback delay than SISO.

## 7. CONCLUSION

In this framework, we provide robust video streaming with spatial multiplexing MIMO combined with packet ordering. With priority based transmission, we can guarantee transmission for more important packets to provide decent video quality even when bandwidth is severely constrained. Although packet ordering can be time-consuming, this framework can be easily extended to take advantage of empirical packet ordering. Future work will involve examining empirical packet ordering algorithms and priority based transmission in spatial diversity MIMO scenarios.

## 8. REFERENCES


[1] G. D. Golden P. W.Wolniansky, G. J. Foschini and R. A. Valenzuela, "V-blast; an architecture for realizing very high data rates over the rich-scattering wireless channel," *Proc. IEEE ISSSE*, p. 295300, 1998.

[2] S. M. Alamout, "A simple transmit diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun*, p. 14511458, 1998.

[3] D. Tse and P. Viswanath, "Fundamentals of wireless communication," *Cambridge University Press*, 2005.

[4] M. Farshchian and W. A. Pearlman, "Real-time video transmission over mimo ofdm channels using space-time block codes," *Proc. 40th Annual Conf. Inf. Sci. Syst*, 2006.

[5] Daewon Song and Chang Wen Chen, "Scalable h.264/avc video transmission over mimo wireless systems with adaptive channel selection based on partial channel information," *IEEE Trans. CSVT*, 2007.

[6] Daewon Song and Chang Wen Chen, "Maximum-throughput delivery of svc-based video over mimo systems with time-varying channel capacity," *J. Vis. Commun. Image R*, 2008.

[7] Ehsan. Maani and Aggelos. K. Katsaggelos, "Optimized bit extraction using distortion modeling in the scalable extension of h.264 avc," *IEEE Trans. on Image Processing*, 2009.

[8] Ehsan Maani and Aggelos K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 407–416, 2010.

[9] Z. Li, Xin Xin, and A.K.Katsaggelos, "Image topological coding for visual search," in *US Patent*, 2012.

[10] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "System and method for compact descriptor for visual search," in *US Patent*, 2012.

[11] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "System and method for compact descriptor for visual search," in *US Patent*, 2012.

[12] Z. Li, Xin Xin, and A.K.Katsaggelos, "Image topological coding for visual search," in *US Patent*, 2012.

[13] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "Scalable query for visual search," in *US Patent*, 2013.

[14] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "Query compression for mobile visual search," in *Proc. of IEEE ICIP, Submitted*, 2012.

[15] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "A novel image retrieval framework exploring inter cluster distance," in *Proc. of IEEE ICIP*, 2010.

[16] TJ Liu, HY Han, Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "A robust and lightweight feature system for video fingerprinting," in *Proc. of EUSIPCO*, 2010.

[17] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "Robust feature selection with self-matching score," in *Proc. of IEEE ICIP*, 2013.

[18] Xin Xin, Zhu Li, and Aggelos K. Katsaggelos, "Laplacian sift in visual search," in *Proc. of IEEE ICASSP*, 2012.

[19] M. Iliadis, Seunghwan Yoo, Xin Xin, and Aggelos K. Katsaggelos, "Virtual touring: A content based image retrieval application," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1–4.

[20] Xin Xin, A. Nagar, G. Srivastava, Zhu Li, F. Fernandes, and A. K. Katsaggelos., "Large visual repository search with hash collision design optimization," *MultiMedia, IEEE*, vol. 20, no. 2, pp. 62–71, 2013.

[21] Xin Xin, Zhu Li, Zhan Ma, and A. K. Katsaggelos, "Spectral approximation to point set similarity metric," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1–4.

[22] J. Springer, Xin Xin, Zhu Li, J. Watt, and A. K. Katsaggelos, "Forest hashing: Expediting large scale image retrieval," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 1681–1684.

[23] Xin Xin and A. K. Katsaggelos, "A novel image retrieval framework exploring inter cluster distance," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 3213–3216.