# H.264 Scalable Video over Finite-State Markov Chain Wireless Channels*

Guogang Hua[1], Daewon Song[1], David Hench[2], and Chang Wen Chen[1]
[1]Dept. of ECE, Florida Institute of Technology, Melbourne, FL 32901
[2]Air Force Research Lab, Information Directorate/IFGC, Rome, NY 13441

## ABSTRACT

In this paper, a wireless channel is viewed as a heterogeneous network in the time domain, and an adaptive video transmission scheme for H.264 scalable video over wireless channels modeled as a finite-state Markov chain processes is presented. In order to investigate the robustness of adaptive video transmission for H.264 scalable video over wireless channels, statistical channel models can be employed to characterize the error and loss behavior of the video transmission. Among various statistical channel models, a finite-state Markov model has been considered as suitable for both wireless links as Rayleigh fading channels and wireless local area networks as a combination of bit errors and packet losses. The H.264 scalable video coding enables the rate adaptive source coding and the feedback of channel parameters facilitates the adaptive channel coding based on the dynamics of the channel behavior. As a result, we are able to develop a true adaptive joint source and channel based on instantaneous channel estimation feedback. Preliminary experimental results demonstrate that the estimation of the finite-state Markov channel can be quite accurate and the adaptive video transmission based on channel estimation is able to perform significantly better than the simple channel model in which only average bit error rate is used for joint source and channel coding design.

**Keywords:** H.264, Scalable video coding, finite-state Markov chain, channel state estimation, unequal error protection

## 1. INTRODUCTION

The demand for high-quality mobile wireless communication services (multimedia broadcasting, video streaming, video telephony, etc) beyond conventional voice communication is increasing at an explosive rate. However, the time-varying characteristics of wireless channel still remain a significant obstacle for reliable wireless communication. From the link layer point of view, the time-varying characteristic of the wireless channel is its dynamic bit error and packet over time. From application layer point of view, its bandwidth is dynamic with time, so wireless channel can be viewed as a heterogeneous network in time domain. Scalable video, designed for heterogeneous networks, can be used with a wireless channel to adapt the time variance in bit error and packet loss.

The scalable extension of H.264/MPEG4-AVC is a current standardization project of the Joint Video Team (JVT) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). This amendment to the H.264 standard is expected to be finalized at the 24th JVT meeting 29 June 2007. Most components of H.264/MPEG4-AVC are used as specified in the standard. The base layer of an SVC bit-stream is generally coded in compliance with H.264/MPEG4-AVC, and each standard conforming H.264/MPEG-4 AVC decoder is capable of decoding this base layer representation when it is provided with an SVC bit-stream. The experimental results show that coding efficiency of H.264 scalable video coding is much better than previous scalable video coding (such as MPEG-4 FGS). For SNR scalability, the performance of H.264 scalable video coding is very close to that of single layer coding [1].

* Acknowledgement: This research work has been supported by Florida Institute of Technology Allen S. Henry Endowment Fund and Air Force Grant FA8750-07-2-0025 by AFRL/IFGC.

Multimedia Systems and Applications X, edited by Susanto Rahardja, JongWon Kim, Jiebo Luo,
Proc. of SPIE Vol. 6777, 677702, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.748662

In this paper, we investigate the new H.264 scalable video coding on an error prone wireless channel. As we all know, due to multi-path, mobility, terrain and a lot of other complicated reasons, the wireless channel are unreliable and dynamic. In order to investigate the performance of adaptive video transmission for H.264 scalable video over the wireless channels, we model the channel as a finite-state Markov model. Finite-state Markov modeling of a communication channel is a simple and effective approach for communication channel description [2]. To transmit scalable video over heterogeneous networks, the encoder must know the capacity of each link in order to deliver a suitable video stream (suitable base layer and enhancement layer descriptions) to each link. This video stream can make sure the decoder at each link can get the best quality of video under the given link capacity. When applied to the wireless channel, i.e., the time domain heterogeneous network, the encoder should know the channel state at the each transmission time. If the channel is good, the encoder can generate a better description and use lighter FEC protection; if the channel is not so good, the encoder can generate a suitable description and use stronger FEC protection. So channel estimation is a very important task in delivering H.264 scalable video coding over the wireless channel. We propose a channel estimation method based on the decoding process in the decoder, and feed back the channel state to the encoder to help the encoder to deliver best quality video based on the estimated channel condition.

The rest of this paper is organized as follows: In Section 2, an overall system description of the proposed scheme is given. We will also discuss both scalable H.264/AVC codec and channel state estimation based on finite-state Markov Chain. In Section 3, experimental results are reported to demonstrate that the proposed scheme indeed is able to perform significantly better than the non-rate adaptive system. Section 4 concludes this paper with summary and discussion.

## 2. SYSTEM DESCRIPTION

The proposed scalable video transmission over a time-varying wireless system is shown in Figure 1. At the encoder side, base and enhancement layers are generated by scalable extension of H.264/AVC [3]. With the feedback of estimated channel state information, channel codes are adaptively assigned, check sum bits are added, and the bit stream of enhancement layer is controlled in order to meet the available bandwidth. At the receiver side, the transmitted corrupted signals might be corrected by channel decoding and residual errors are detected by check sum bits. Then, channel state will be estimated using finite-state Markov chain. For video reconstruction, corrupted packets are dropped and finally transmitted video sequences are reconstructed using error concealment.

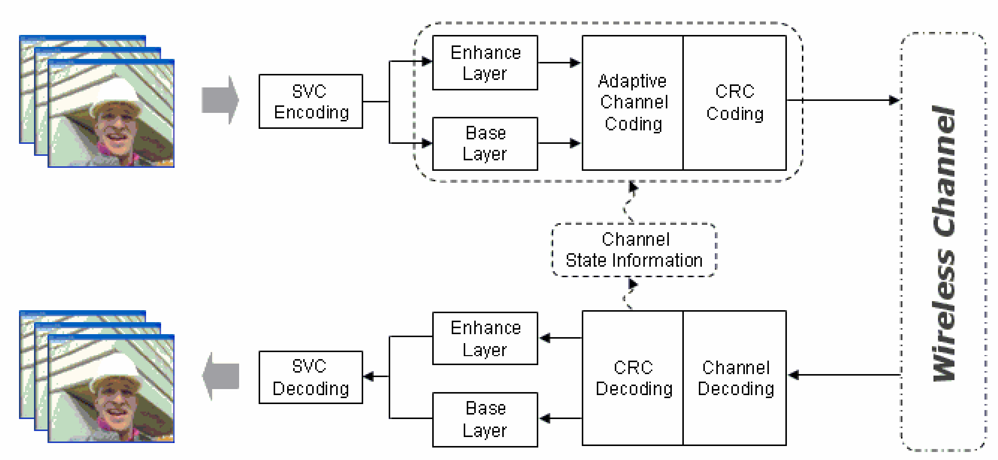


Figure 1: The proposed MIMO system

### 2.1 Scalable Extension of H.264/AVC

With the proposed video transmission scheme over the wireless heterogeneous networks, multi-layered video bit streams are essential and thus created by SVC [3] in this research. In this sub-section, we briefly review this video codec which is an extension of the H.264/AVC [5][6] video coding standard. Traditionally, there are two different ways for scalable video codec: either by using a technique that is intrinsically scalable (such as bit-plane arithmetic coding) or by using a layered approach. SVC supports a combination of the two approaches so that a full spatio-temporal and quality scalable codec is achieved. A coded SVC video sequence consists of a series of Network Abstraction Layer (NAL) units, each containing the layer information. A 4-byte SVC NAL unit header, an extension of H.264/AVC NAL, indicates the decoding dependency relationship of spatial, temporal, and quality scalability.

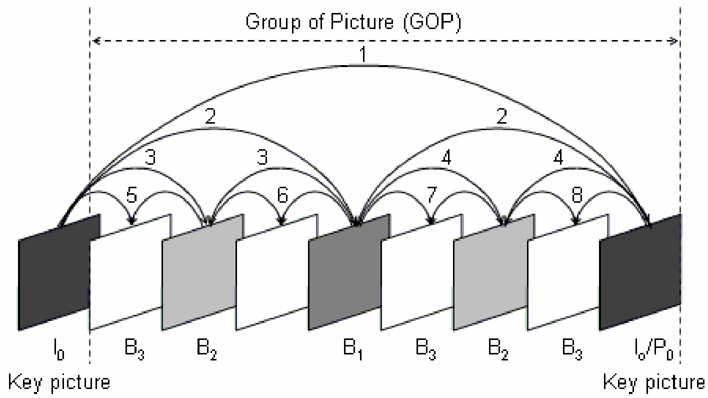


Figure 2: an example of temporal decomposition

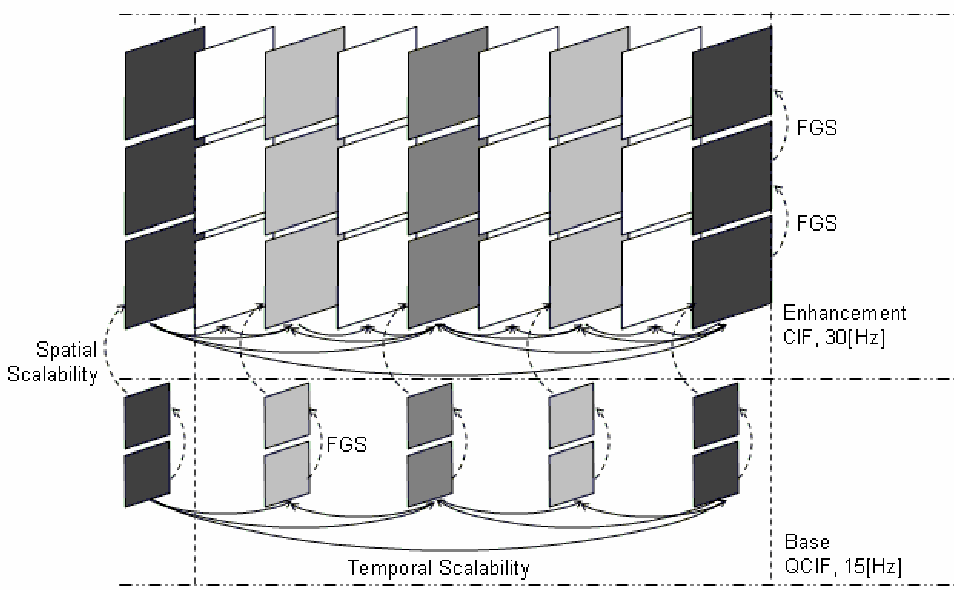


Figure 3: an example of combined scalability

The temporal scalability of SVC is typically given based on the principle of hierarchical B pictures. A hierarchical prediction structure with 4 dyadic hierarchy stages (or 4 temporal scalability levels) is described in Figure 2. In a video sequence, the first picture (key picture) is intra-coded as IDR picture and next key picture will be either intra-coded or inter-coded using previous key picture. All pictures between two key pictures are hierarchically predicted and encoded. Therefore, a group of picture (GOP) is created by a key picture and all pictures that are temporally located between a key picture and the previous key picture. Spatial scalability is also supported based on existing multi-layered coding approach. As depicted in Figure 3, the difference of spatial resolution between base layer and enhancement makes it achieve spatial scalability.

For quality (or SNR) scalability, SVC supports two types: coarse grain scalability (CGS) using various inter-layer prediction techniques and fine grain scalability (FGS) known as progressive refinement. In this work, we adopt FGS for SNR scalability in order to both satisfy each layer's target bit-rate and increase the error robustness. Within each spatial resolution FGS is achieved by encoding successive refinements of the transform coefficients. Therefore, a picture is represented by base representation and FGS refinement representations by repeatedly decreasing the quantization step size. The NAL units of FGS refinement layers can be truncated at any arbitrary point at the encoder and thus error robustness is increased by the decoder capable of arbitrarily discarding corrupted NAL unit streams. An example of combined scalability with group of picture (GOP) 8 is applied in experiments and described in Figure 3. For a more detailed explanation on SVC, see [3][4].

### 2.2 Finite State Markov Chain

A finite-state Markov model is defined as a finite set of states $\boldsymbol{S} = \{\boldsymbol{s_0}, \boldsymbol{s_1}, \cdots, \boldsymbol{s_{K-1}}\}$, where $\boldsymbol{K}$ is the number of states (Figure 4 shows an example of 2-state Markov model). A binary symmetric channel (BSC) with a given crossover probability is associated with each state $\boldsymbol{s_k}, \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots\}$, so the channel quality for each state can be identified. Let $\{\boldsymbol{S_n} \mid \boldsymbol{S_n} \in \boldsymbol{S}, \boldsymbol{n} = \mathbf{0,1,2,}\cdots\}$ be a constant Markov process. Since the constant Markov process has the property of stationary transitions, the transition probability is independent of time index $\boldsymbol{n}$, so

$$\boldsymbol{t_{j,k}} = \mathbf{Pr}\{\boldsymbol{S_{n+1}} = \boldsymbol{s_k} \mid \boldsymbol{s_n} = \boldsymbol{s_j}\} \quad \boldsymbol{k}, \boldsymbol{j} \in \mathbf{0,1,2,}\cdots \boldsymbol{K} - \mathbf{1} \tag{1}$$

is a constant $\boldsymbol{K} \times \boldsymbol{K}$ matrix, which is called the state transition probability matrix $\boldsymbol{T}$. The state transition probability matrix is an important parameter for a finite-state Markov model. Note that the sum of the elements on each row of $\boldsymbol{T}$ is equal to 1.

$$\sum_{l=0}^{K-1} \boldsymbol{t_{kl}} = \mathbf{1}, \forall \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots\} \tag{2}$$

Another set of parameters for a finite-state Markov model are the crossover probabilities in each state, $\boldsymbol{e_k}, \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots\}$ .The last parameter for a finite-state Markov model is the steady state probability of each state, $\boldsymbol{p_k}, \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots\}$. But the steady state probabilities are not independent. The steady state probability $\boldsymbol{p_k}$ for state $\boldsymbol{s_k}$ can be viewed as the probability of all states transitioning to state $\boldsymbol{s_k}$, so

$$\sum_{j=0}^{K-1} \boldsymbol{p_j t_{j,k}} = \boldsymbol{p_k}, \forall \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots, \boldsymbol{K} - \mathbf{1}\} \tag{3}$$

The functions only have zero solution. But $\sum_{k=0}^{K-1} \boldsymbol{p_k} = \mathbf{1}$ (4)

With any $\boldsymbol{K} - \mathbf{1}$ functions from (Eq. 3) and the function (Eq. 4), the $\boldsymbol{p_k}, \boldsymbol{k} \in \{\mathbf{0,1,2,}\cdots\}$ can be determined with any given state transition probability matrix $\boldsymbol{T}$.

So the K-state Markov model is decided by its parameters: state transition probability matrix $T = [t_{jk}], j,k \in [0,1,...K-1]$ and crossover probabilities $e_k, k \in [0,1,...,K-1]$. These states and its crossover probabilities determine the channel quality for any wireless links modeled by such a Markov model. When an estimation of the channel model can be obtained, an adaptive video transmission scheme can be designed to take full advantage of the known channel parameter and the dynamics of the channel behavior. The parameters of the finite-state Markov channel model can be obtained by collecting traces under various network conditions.

### 2.3 Channel State Estimation

We propose the following channel state estimation at the decoder. The decoder estimates the channel state for the next one or several transmissions (according to application latency) based on the current received packet and the packet after channel decoding. Each video packet is protected by different strong level of FEC, depending on the important of video coding layer and the current channel state. After receiving one packet, the decoder first does channel decoding.

Based on the received original packet and the packet after channel decoding, the channel estimation method is as following:

If the channel decoding is successful, we can compare the received original packet and the packet we get after decoding. The different number of bits should be the number of transmission error. The number of transmission error divided by the length of the packet, is the error probability $p_e$.

If current state is good, then the next state is estimated as good;

Else if current state is bad:

If $p_e < e_0$ (the crossover probabilities of good state), then the next state is estimated as good;

Else, the next state is estimated as bad.

If channel decoding is unsuccessful, then the next state is estimated as bad.

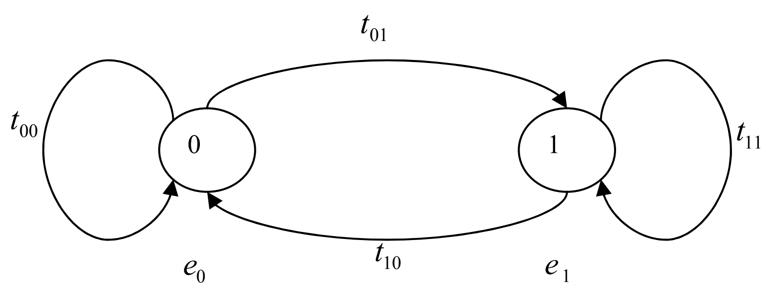


Figure 4: A two-state Markov chain model

### 2.4 Video Reconstruction

If the bit errors are above the capacity of the assigned channel coding, the transmitted packets might be corrupted and are dropped before video reconstruction. Then, error concealment is processed at the decoder. The detection of one or more missing pictures is done by calculating the frame number gap, picture order count (POC) gap, and the GOP size. If a base layer packet is lost, the corresponding enhancement layer packet is invalid and will be regarded as a lost packet also. We adopt picture copy (PC) algorithm in this work. With this error concealment, each sample of value of the concealed picture is copied from the corresponding sample of the first picture in the reference list 0. The base layer PC will only be done when decoding a low resolution sequence. A lost enhancement layer picture will be concealed using the first picture in the reference picture list 0. The list 0 is generated from the decoding process of the enhancement layer and it contains only high resolution pictures.

## 3. EXPERIMENT RESULTS

We provide numerical examples to show how the proposed video transmission scheme with adaptive channel coding is able to achieve stable reconstructed video over time-varying wireless channel.

### 3.1 Performance of Channel State Estimation

Using the proposed method above, we conduct the prediction experiments based on different Markov models. We tested different state transition matrix. In all cases, the crossover probabilities of the good state is 0.1%; and the crossover probabilities of the bad state is 10%. Table 1 gives the prediction error with different Markov models. From the table we can see that when the transition probabilities between the states become larger, the prediction error also becomes larger. This is because when the transition probabilities between the states are large, the channel becomes "unstable", it is not easy to predict the next state based on previous states.

Table 1. Prediction errors with different Markov models

| State transition matrix | Prediction Error (%) |
|---|---|
| $\begin{bmatrix} 0.99875 & 0.00125 \\ 0.005 & 0.995 \end{bmatrix}$ | 0.202 |
| $\begin{bmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{bmatrix}$ | 1.35 |
| $\begin{bmatrix} 0.98 & 0.025 \\ 0.025 & 0.97 \end{bmatrix}$ | 2.48 |

### 3.2 Performance of Reconstructed Video PSNR

In this sub-section, we conduct experiments to show the performance of the proposed system with reconstructed PSNR of the decoded video sequences. As shown in sub-section 3.1, the prediction of channel state (CS) estimation would be quite accurate if the transition probabilities between the states become large. Thus, in this sub-section, we conduct an experiment with ideal case that the channel state estimation is perfect and no delay. Three video sequences, 'Mobile', 'Carphone', and 'Foreman', are tested. All test sequences are 64 frames with CIF 15[Hz] and encoded by scalable H.264/AVC reference model JSVM 7.0 [4] to generate 2-layer scalable video bit-streams with GOP size 8. Reed-Solomon (RS) Codes [7] are adopted to protect the transmitted bit-streams since it maintains maximum erasure protection while produces a minimum of redundancy. We assume that 640 [Kbits/sec] is available as data rate and then 256[Kbits/sec] is assigned for base layer and 384[Kbits/sec] is utilized for enhancement layer. Then, in order to achieve unequal error protection (UEP), different channel coding redundancies are assigned for base and enhancement layers. Fixed UEP RS codes are assigned in no CS feedback system while assigning adaptive UEP RS codes in the proposed system that has the feedback of CS. Note that the enhancement layer bit streams are adaptively dropped to adjust the RS channel coding according to the channel state so as to achieve the adaptive transmission scheme with the fixed available data rate

The average reconstructed PSNR (luminance component) for 'Mobile', 'Foreman', and 'Carphone' are shown in Figure 5, Figure 6, and Figure 7, respectively. In these figures, we clearly demonstrate that the proposed adaptive transmission scheme (with CS Feedback) based on channel state estimation achieves steady reconstructed video sequences over time-varying wireless networks as compared with non-rate adaptive source coding system (w/o CS Feedback). In table 2, we illustrated the average reconstructed PSNR of three test sequences.

Table 2. Average reconstructed PSNR of test sequences

| | Mobile | Carphone | Foreman |
|---|---|---|---|
| w/o CS Feedback | 26.13 [dB] | 33.76 [dB] | 28.35 [dB] |
| with CS feedback | 32.53 [dB] | 37.42 [dB] | 35.82 [dB] |

## 4. CONCLUSION

We propose in this research an adaptive video transmission based on H.264 scalable video coding and the adaptive channel estimation based on finite-state Markov model (See Figure 4). The H.264 scalable video coding enables the rate adaptive source coding and the feedback of channel parameters facilitates the adaptive channel coding based on the dynamics of the channel behavior. As a result, we are able to develop a true adaptive joint source and channel based on instantaneous channel estimation feedback. Preliminary experimental results demonstrate that the estimation of the finite-state Markov channel can be quite accurate and the adaptive video transmission based on channel estimation is able to perform significantly better than the simple channel model in which only average bit error rate is used for joint source and channel coding design. The proposed methods can be easily extended to multi-state Markov model based on different application scenarios.

## REFERENCES


[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/MPEG4-AVC extension", Proc. IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, USA, October 8-11, 2006

[2] F. **b**abich and G. Lombardi, "A Markov model for the mobile propagation channel", IEEE Trans. On Vehicular Technology, vol. 49, No.1 Jan. 2000.

[3] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, Joint Draft of SVC Amendment, Joint Video Team (JVT), Doc. JVT-T201, Klagenfurt, Austria, July 2006

[4] J. Reichel, H. Schwarz, and M. Wien, JSVM-7 Software, Joint Video Team (JVT), Doc. JVT-T202, Klagenfurt, Austria, July 2006

[5] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for generic audiovisual services," 2003

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. on Circuits and Systems for Video Tech, vol. 13, no. 7, pp. 560-576, July 2003

[7] S. Lin and D. J. Costello, Error control coding-fundamentals and applications, New Jersey: Prentice-Hall, 1993


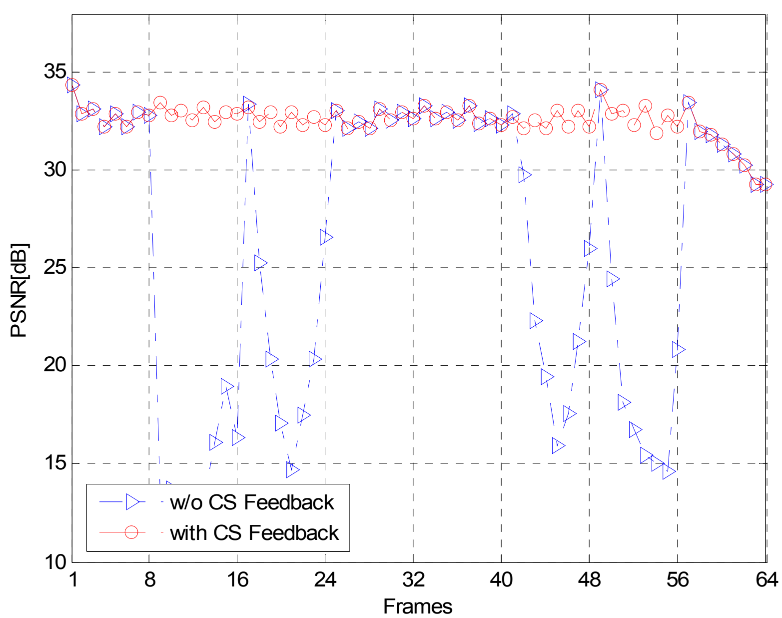


Figure 5: Reconstructed PSNR of Mobile

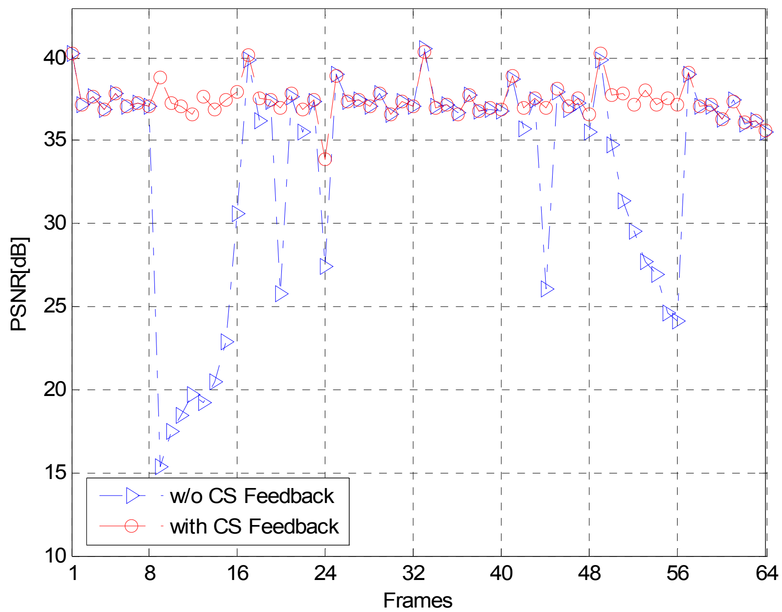


Figure 6: Reconstructed PSNR of Carphone

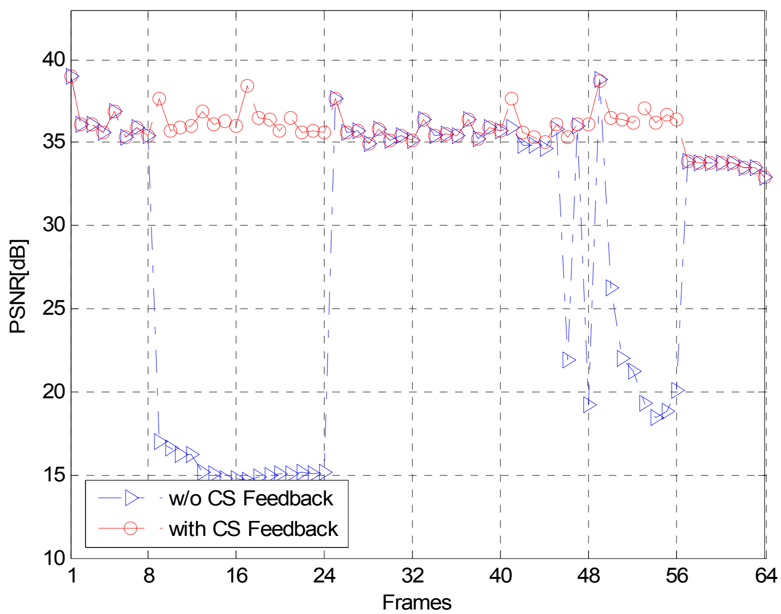


Figure 7: Reconstructed PSNR of Foreman