Evaluation of Interleaved Source Coding (ISC) over Channel with Memory

Jin Young Lee, *Member, IEEE* and Hayder Radha, *Senior Member, IEEE*

***Abstract* - Network impairments such as delay and packet losses have severe impact on the presentation quality of many predictive video sources. *Interleaved Source Coding* (ISC) [5][6][7] is one of the error resilient coding methods for predictive video coded frames transmitted over a single erasure channel. ISC has shown to significantly improve the overall quality of predictive coded video stream over a lossy channel [5][6][7]. In this paper, ISC is evaluated over channels with memory. In particular, we analyze the impact of packet correlation [11][13] of the popular Gilbert model on ISC-based packet video over a wide range of packet loss probabilities. Simulations have shown that ISC advances the traditional method as either the loss rate increases or the packet correlation decreases.**

Keywords-Interleaving; packet correlation; markov decision process; dynamic programming

I. INTRODUCTION

With the expansion of the underlying infrastructure of the IP network, streaming video has become one of the most popular on-line realtime Internet applications. However, despite of the growth of internet infrastructure, due to the unreliable nature of IP networks, such realtime streaming services often lack Quality-of-Service (QoS) guarantees. To improve playback quality of realtime streaming video under such network condition, coding techniques that are resilient to packet losses are in need. Coding techniques (e.g., [1][3]-[8]) are few examples of methods to be resilient to packet losses.

In this paper, we evaluate interleaved source coding (ISC) [5][6], a recently proposed packet loss resilient coding technique, over erasure channels with memory. ISC codes a single video sequence into two sub-sequences and transmits them over a single erasure channel to reduce the frequency and impact of the cascaded effect of packet losses and related propagation of errors of predictive coded video. In addition, to avoid complex modifications to standard video coders, ISC uses frozen frame technique for the decoder failed frames from packet losses. The objective of ISC is to find an optimum interleaving sub-sequence set for a given erasure channel model such that the impact of packet losses is minimal, in other words, the number of loss impacted frames are minimum. ISC employs Markov Decision Process (MDP) and a Dynamic Programming algorithm with a realistic packet loss model and some coarse measure of the temporal correlation among pictures within a given video sequence, to find interleaving sets that are unique to each video sequence.

In this paper we evaluate the performance of ISC with various packet loss rates in conjunction with a *packet correlation* model [12][13]. The *packet correlation* model uses the average loss rate p and the packet correlation ρ to represent the state transition probabilities of the two state Markov model (a.k.a. *Gilbert model*).

The remainder of this paper is organized as follows: In Section II, we briefly describe the ISC coding method following by the description of the *packet correlation* model in Section III. In the last section, ISC is evaluated over *packet correlation* based lossy channel models with various packet loss rates and streams coded using the MPEG-4 video standard.

II. INTERLEAVED SOURCE CODING

A. General Interleaving

In general, a predictive video coding partitions a single lengthy sequence into a number of shorter length Group Of Video object planes (GOVs) to limit the impact of possible errors or losses into individual GOVs.

The *interleaved source coding* (ISC) [5][6][7] is a pre- and post-process of predictive source coders to reduce the impact of losses within a given GOV, hence improves the playback quality of predictive video over lossy packet networks. As introduced in [5][6][7], ISC separates a single video sequence into two sub-sequences and the resulting sub-sequences are encoded using separate video encoders. Once encoded, the sub-sequences are merged into a single stream in the original-sequence frame order. Prior to the transmission of the merged ISC stream, information on the interleaving pattern employed by the ISC must be transmitted to the decoder. Once the merged stream is transmitted, with the interleaving pattern information, the stream is separated for the decoder and the separated streams are decoded independent to each other. For playback, the separated sub-sequences' frames are merged into the proper order.

J. Y. Lee is with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, on leave from Michigan State University, East Lansing, MI 48824 USA (Phone: +82-42-860-5383, Fax: +82-42-860-5440 Email: jinlee@etri.re.kr)

H. Radha is with Michigan State University, East Lansing MI 48824 USA (radha@egr.msu.edu)

$$
\begin{gathered}
s = \left\{0 \quad 1 \quad \cdots \quad 2N-1\right\} = \bigcup_{j=1}^{2} s^{(j)} \\
\bigcap_{j=1}^{2} s^{(j)} = \varnothing, \qquad \forall j, size\left(s^{(j)}\right) = N \qquad (1) \\
\sum \left\{s^{(j)}(2) - s^{(j)}(1), \cdots, s^{(j)}(N) - s^{(j)}(N-1)\right\} > N-1
\end{gathered}
$$

For an interleaving sub-sequence set $\mathbb{S} = \left\{s^{(1)}, s^{(2)}\right\}$, the numbers in $s^{(j)}$ represent the frame locations in the non-interleaved sequence and the coded stream's frame transmission order and this interleaving information must be transmitted (e.g., as meta data) prior to the coded stream transmission.

When separating a single sequence into two sub-sequences, $s^{(j)}$, represented by an index set, $j = \{1,2\}$, ISC uses the interleaving constraints shown in (1). Here, $2N$ represents the number of frames in the original non-interleaved sequence s where it could be the number of frames in a GOV. Hence, the same interleaving is applied to all GOVs in the sequence or a scene.

For a given GOV size, the size of the possible number of the interleaving set $\mathbb{K}$ could be quite large for any reasonable GOV size $2N$, over a lossy network channel. Therefore, identifying an optimum interleaving set that produces the best quality decoded video could be very computationally expensive task, hence ISC employs a decision-based search algorithm to elect an optimal interleaving set for a given erasure-channel model and a video sequence.

B. Decision Based Interleaving

Prior researches on the analysis and modeling of packet losses over the Internet (e.g., [11]) and wireless networks (e.g., [4]) have proven that these losses exhibit Markovian properties. A *Markov Reward Process* (MRP) (e.g., [2][9]) can estimate the system's performance after n packet transmissions using the lossy channel's Markov transition probabilities and some model for the *rewards* associated with each system state. For a MRP of an erasure-channel model, the state space of two state Markov model is mapped to good (0) and bad (1) packet transmission state. Whenever the process reaches state i, the instant rewards r_i assigned for each state are awarded to the process. Hence the aggregated reward $v(n-1)$ ([2][9]) is a function of the number of transmitted packets which represents the performance of predictive sequence transmission after n packet transmissions over a lossy channel with a channel's state transition matrix p. For a Gilbert channel model, $\{r_0, r_1\} = \{1,0\}$ are the instant reward constants and the process is awarded with *1* for a successful packet and *0* for a lost packet transmission, hence $v_i(n-1)$, represent the expected number of good packet transmissions with the initial process state i.

Associating Markov reward process with a series of actions and decision criteria [2][9] gives a base model for a Markov Decision Process (MDP). For ISC, MDP is employed to find an interleaving set that is most suitable for a given decision criteria, maximize the number of frames (or associated packets) that can be decoded *correctly*. An aggregated MDP equation is obtained by incorporating (2) with an *interleaving set indicator* k, $k \in \mathbb{K}$ along with a set of *policies,* mappings from states to actions, and a set of *discount factors*, γ_a [2][9].

$$
\begin{aligned}
v(0) = r = \left\{r_0, r_1\right\}^T, &\quad v(1) = r + p \bullet v(0) \\
v(n-1) = \left[v_0(n-1) \quad v_1(n-1)\right]^T & \qquad (2) \\
= r + p \bullet v(n-2) &= \left(I + \sum_{m=1}^{n-1} p^m\right) \bullet r
\end{aligned}
$$

$$
v^{(k)}(n-1) = r_{a^{(k)}(n-1)} + \gamma_{a^{(k)}(n-1)} \times p \bullet v^{(k)}(n-2) \qquad (3)
$$

For predictive video coding, one of the two actions, Coding (C), or Skip (S) is taken for each state iteration. $\mathbb{S}^{(k)} = \left\{s^{(k,1)}, s^{(k,2)}\right\}$ indicates an ISC sub-sequence set with respect to k, and to start the sub-sequence's reward computation from time instance 0, the frame numbers are rewritten as following.

$$
s^{*(k,j)}(n) = s^{(k,j)}(n) - s^{(k,j)}(0), \quad \text{for } 0 \le n \le N-1 \qquad (4)
$$

In ISC, frames are coded (action C is performed) at frame locations specified in $s^{*(k,j)}$. When the difference between two adjacent numbers in $s^{*(k,j)}$ exceeds 1, which indicates the presence of skipped frames, action S is performed.

ISC requires modification of the transition matrix p in association with actions. For the policy $\{C,1\}$, the state *1* for action C is considered as a trapping state since the decoder of predictive coding is forced to stop when a lost packet is detected. Once the decoder is stopped, ISC uses the last successfully decoded frame to replace the missing and effected frames and waits for a new GOV with successfully decoded *I* frame to restart decoding process. Hence the transition probability to the next good state is set to *0* for the policy. Simultaneously, the discount factor for the policy $\{C,1\}$ is set to *0*, since no further decoding is possible, the process is in the trapping state, hence, propagation of the aggregated reward is prohibited for the GOV. For all other policies, the aggregated rewards are propagated to the next state and the discount factors are set to *1*.

For the initial state, the instant reward is multiplied by a stationary probability π since it is assumed that the first packet in *I*-frame arrives to the process with the stationary probability since *I*-frames do not have any temporal dependencies to the previously decoded frames. When frames are packetized, each frame is coded into different number of packets depending on the bitrate, the frame rate of the encoder, the packet size, the frame coding type, and the motion of the sequence. Hence, ISC incorporates this unpredictability of the variation of the number of packets per each coded frame with an average number of packets per frame and the sum of aggregated rewards estimates the expected number of successfully decoded frames for each interleaving set k .

$$v^{(k)} = \sum_{j=1}^{2} \sum_{n=0}^{N-1} v^{(k,j)} \left(s^{*(k,j)}(n) \right) \tag{5}$$

At last, ISC elects an optimal interleaving set k that satisfies our decision criterion, a set with the highest $v^{(k)}$.

$$\arg\max_{k} \left[v^{(k)} \right] \tag{6}$$

When predictive video coding uses frozen frames for the decoder failed frames, the distances between the replacement and the replaced frames have effects on the smoothness of the sequence flow and the overall quality of the playback sequence since the shorter distance between the replacing frames indicates highly correlated frame replacement in place of decoder failed frames. Hence, ISC is expected to produce smoother and higher quality video over erasure channels when frames are replaced, since the distance from a replacement frame to the replaced ones are expected to be shorter.

To incorporate frame replacement action, ISC adopts a Dynamic Programming to find an ISC set with the highest MDP sum of the aggregated reward with correlation gain $g^{(k)}$. To compute the sequence specific correlation gain, following steps are used: First, temporal correlation of the transmitting sequence is computed with average PSNR between original sequence and temporally shifted sequences (7), where d is temporal distance between the original and shifted sequence. Second, Minimum Mean Square Estimator (MMSE) is employed to obtain a function that represents temporal correlation of a given sequence (8). Third, a distance matrix $D^{(k)}$ is generated for each ISC set k for single-packet-loss per GOV cases since the main purpose of ISC is to isolate decode failure to one sub-sequence. $D^{(k)}$ is a $2N$ by $2N$ upper triangular matrix and its diagonal indices indicate the first frame location in a GOV impacted by a single packet loss. Hence, the non-zeros entries represent the distances from replacement frames to the replaced ones. At last, the correlation weight matrix $W^{(k)}$ (9) with respect to the distances from replacement frames to the replaced ones, and the correlation computed aggregated reward gain matrix $G^{(k)}$ (11) is computed to obtain $g^{(k)}$ (12). Hence, the optimal interleaving set with correlation model is found with (13).

$$\rho = \bigcup_{d=1} \frac{avg.PSNR(s, s+d)}{avg.PSNR(s, s)} \tag{7}$$

$$\arg\min_{\{a,b,c\}} \left[MSE \left\{ \rho, a \times \exp\left(-d^{b} \right) + c, \forall d \right\} \right] \tag{8}$$

$$W_{x,y}^{(k)} = \begin{Bmatrix} a \times \exp\left(-\left(D_{x,y}^{(k)} \right)^{b} \right) + c, & \forall D_{x,y}^{(k)} \neq 0 \\ 0, \quad \text{otherwise} \end{Bmatrix} \tag{9}$$

$$V^{(k)^*} = \begin{bmatrix} V^{(k)}(2N-1) & V^{(k)}(0) & \cdots & V^{(k)}(2N-2) \end{bmatrix} \tag{10}$$

$$G_{x,y}^{(k)} = \begin{Bmatrix} W_{x,y}^{(k)} \times V^{(k)^*} \left(y - D_{x,y}^{(k)} \right), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), \quad \text{otherwise} \end{Bmatrix} \tag{11}$$

$$g^{(k)} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)}, \qquad \forall x, y \leq \text{GOV SIZE} \tag{12}$$

$$\arg\max_{k} \left[v^{(k)} + g^{(k)} \right] \tag{13}$$

However, due to delay, complexity, and memory constraints, measuring the temporal correlation among video frames within a complete GOV may not be always feasible for realtime applications. Therefore, a generic correlation model may be required if the actual correlation cannot be computed.

$$\begin{gathered} VI^{(k)} = \bigcup_{j} VI^{(k)}(s^{(k,j)}) \\ \text{where } VI^{(k)}(s^{(k,j)}(0)) = v^{(k,j)}\left(s^{*(k,j)}(0) \right), \\ VI^{(k)}(s^{(k,j)}(n)) = v^{(k,j)}\left(s^{*(k,j)}(n) \right) - v^{(k,j)}\left(s^{*(k,j)}(n-1) \right), \\ \text{for } 1 \leq n \leq N-1 \end{gathered} \tag{14}$$

$$\begin{gathered} W^{(k)^*} = \left(\left(D^{(k)^*} \times VI^{(k)^T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)^*}, \forall x \right)^{\wedge} D^{(k)}, \\ D^{(k)^*} = \begin{Bmatrix} 0, & \forall D_{x,y}^{(k)} \neq 0 \\ 1, & \text{otherwise} \end{Bmatrix} \end{gathered} \tag{15}$$

$$G_{x,y}^{(k)^*} = \begin{Bmatrix} W_{x,y}^{(k)^*} \times V^{(k)^*} \left(y - D_{x,y}^{(k)} \right), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), \quad \text{otherwise} \end{Bmatrix} \tag{16}$$

$$g^{(k)^*} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)^*}, \qquad \forall x, y \leq \text{GOV SIZE} \tag{17}$$

Here, $VI^{(k)}$ is the set of the reward increments at each sub-sequences' reward calculation iteration. In (15), the term $\left(\left(D^{(k)^*} \times VI^{(k)^T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)^*}, \forall x \right)$ is the average reward increment of the successfully decoded frames in case of a single error in a GOV and the decrement of the weight matrix is assumed to be exponential with respect to temporal distances from the replacement frames to the replaced ones. Hence, the optimal interleaving set using the above generic correlation model can be found using the following equation.

$$\arg\max_{k} \left[v^{(k)} + g^{(k)^*} \right] \tag{18}$$

III. GILBERT MODEL PARAMETER PAIRS USING PACKET CORRELATION

In the Gilbert channel model, the steady state probabilities in good state and bad state is represented as following.

$$\pi(0) = \frac{p_{10}}{p_{01} + p_{10}} \text{ and } \pi(1) = \frac{p_{01}}{p_{01} + p_{10}} \tag{19}$$

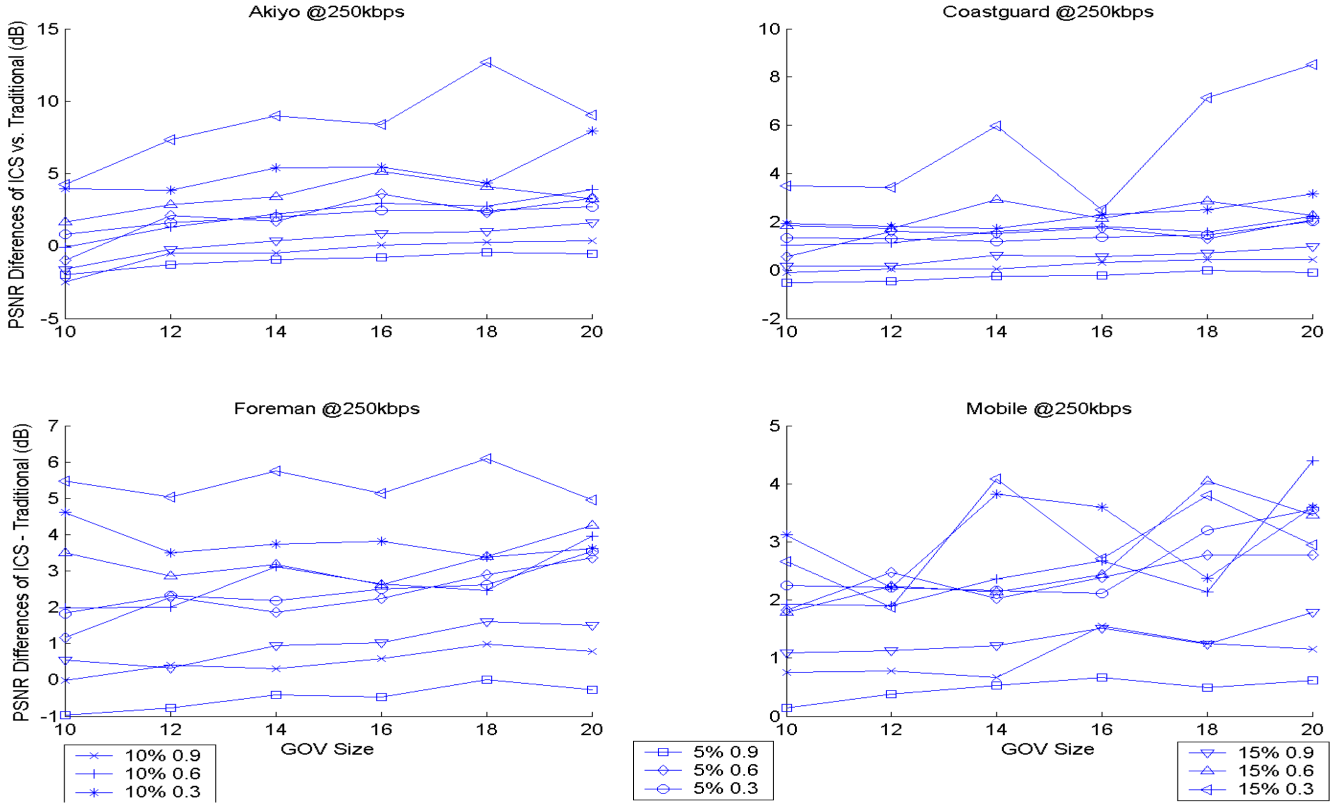

Figure 1. PSNR Differences between ISC and Non-ISC : Sequence Specific ISC vs. Non-ISC *@250kbp*

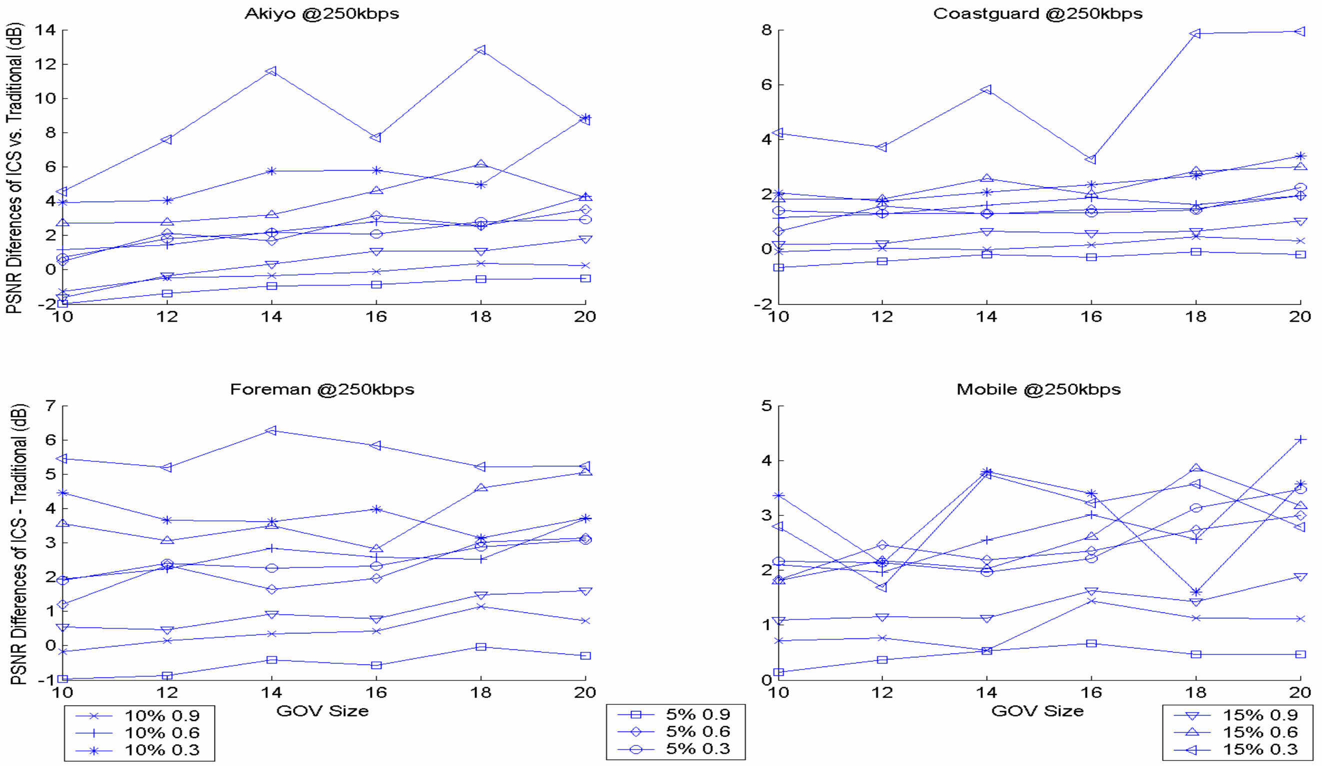

Figure 2. PSNR Differences between ISC and Non-ISC : Generic ISC vs. Non-ISC *@250kbps.*

These values give coarse measure of a given channel's packet transmission behavior. However, for statistical channel modeling, instead of the above probabilities, the transition probabilities p_{01} and p_{10} (or $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$) could be used to characterize Gilbert channel model. Since it is difficult to properly model a Gilbert channel with arbitrary

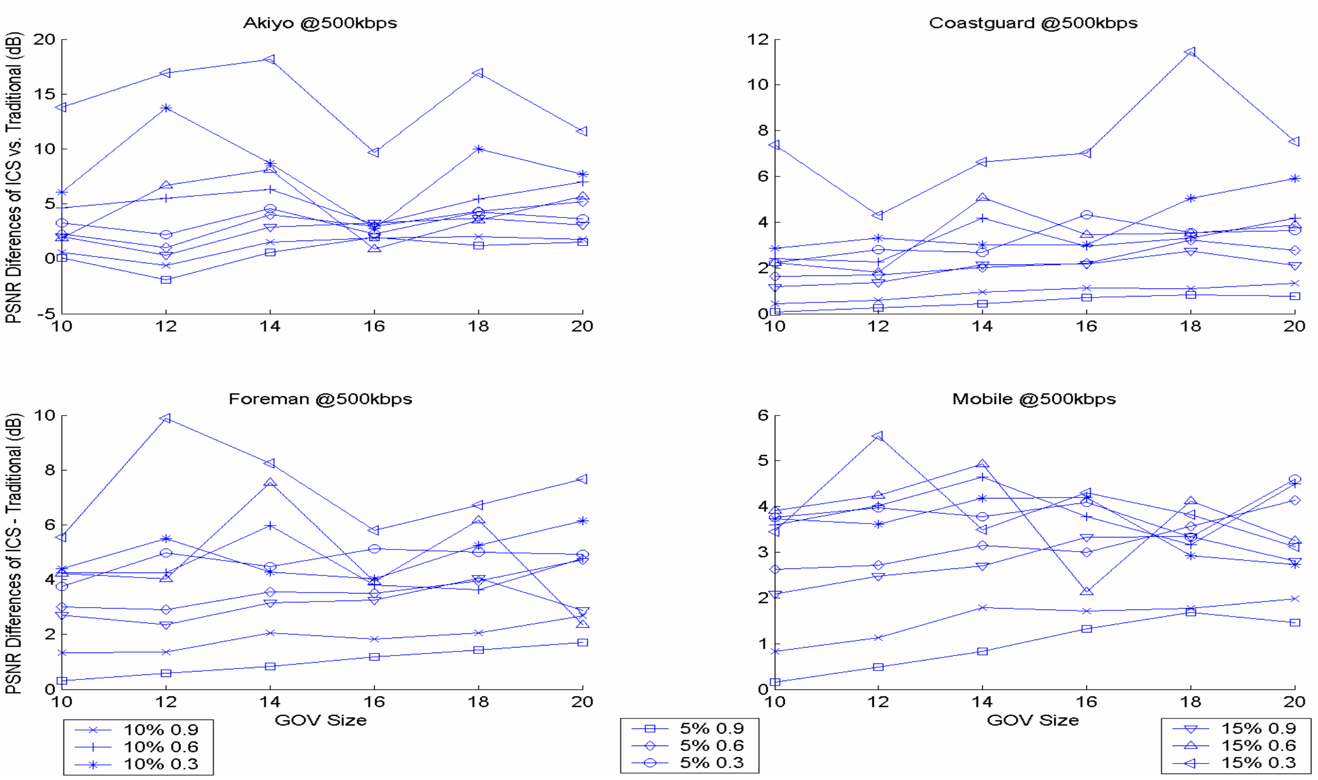

Figure 3. PSNR Differences between ISC and Non-ISC : Sequence Specific ISC vs. Non-ISC *@500kbps*.

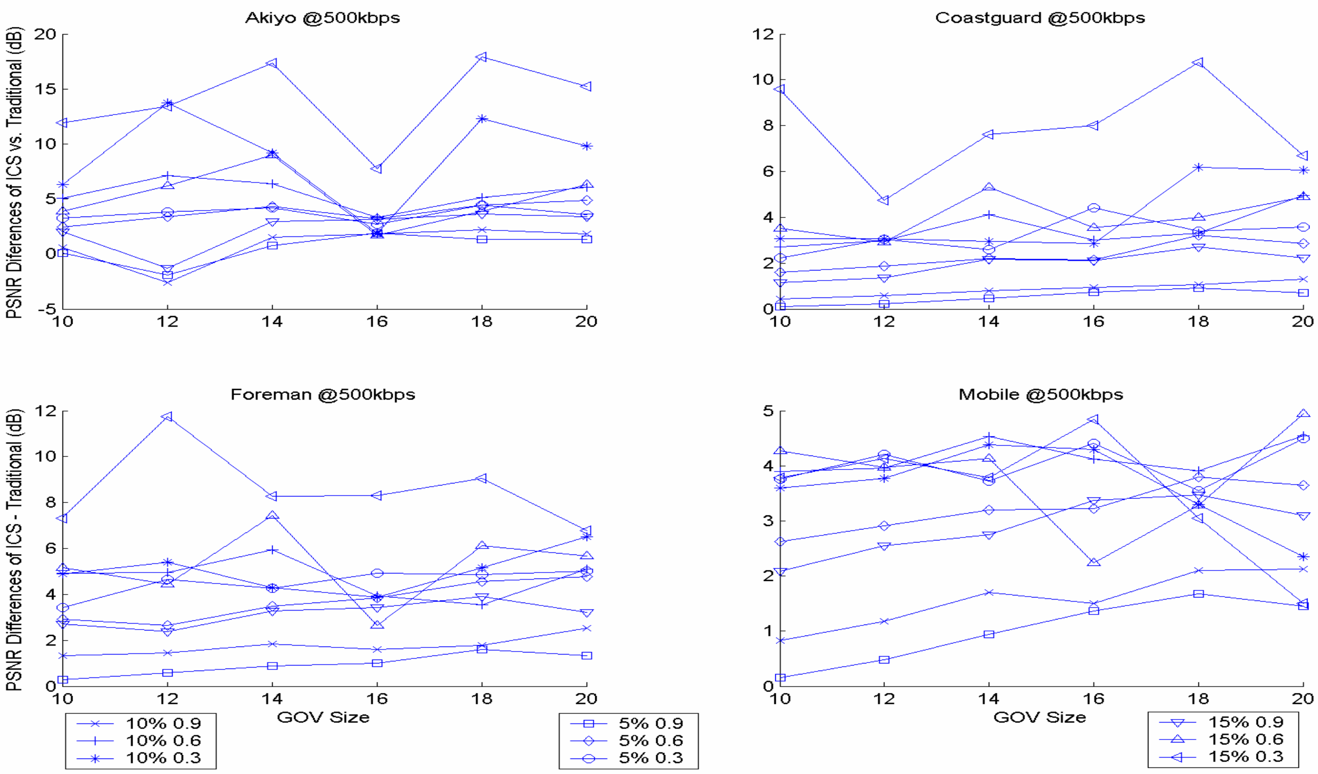

Figure 4. PSNR Differences between ISC and Non-ISC : Generic ISC vs. Non-ISC *@500kbps*

transition probabilities, p_{01} and p_{10}, a more meaningful pair of parameters are the average loss rate, p_1, and the *packet correlation*, ρ; this pair can provide a practical and useful insight of the channel while representing the state transition probabilities.

The average loss rate and the correlation between two consecutive packets can be defined as follows [11][13]:

$$p_1 = \frac{p_{01}}{p_{01} + p_{10}}, \qquad \rho = p_{01} + p_{10} - 1 \qquad (20)$$

Hence, the transition probabilities represented by p_1 and ρ are:

$$\begin{aligned} p_{00} &= 1 - p_1(1-\rho), & p_{01} &= p_1(1-\rho) \\ p_{10} &= (1-p_1)(1-\rho), & p_{11} &= 1-(1-p_1)(1-\rho) \end{aligned} \qquad (21)$$

In addition, the steady state probabilities are directly related to the loss rate p_1; $\pi(0) = 1 - p_1$ and $\pi(1) = p_1$. Furthermore, the packet erasure correlation ρ provides an average measure of the correlation of two consecutive packets. In particular, when $\rho = 0$, $p_{01} + p_{10} = 1$, the loss process is memory-less, and the above probability measures reduce to the special case of a memory-less Binary Erasure Channel (BEC). In the sequel, we analyze the impact of the level of correlation among consecutive packets, as represented by ρ, on ISC-based packet video over a wide range of loss rate p_1 values.

IV. SIMULATIONS AND RESULTS

For Simulation, CIF sequences, *Akiyo*, *Foreman*, *Coastguard*, and *Mobile*, are encoded using MPEG-4 encoder. Only Intra- (*I*) and Inter- (*P*) coded frames are used to form GOV. Un-interleaved GOV sizes of 10, 12, 14, 16, 18, and 20 were used and frame rate of *15fps*, bitrate of *250kbps* and *500kbps*, and packet size of *512Byte* were used when coding the sequences. 5%, 10%, and 15% packet loss rates were used and packet correlation value of 0.3, 0.6, and 0.9 were used to represent low, medium, and high correlation between the transmitted packets.

To limit the impact of a single packet loss to a single frame, no packets are shared among two consecutive coded frames. Furthermore, partial decoding is not allowed for the frames with errors and they are replaced with the last successfully decoded frames (frozen frames) for both ISC and traditional (non-ISC) cases. To simulate statistically viable experiments and to capture a realistic network loss patterns, ten error traces were generated for each $p_1 - \rho$ pair. Each ISC pattern is fitted into these error traces and the PSNR values are averaged to provide statistically satisfying results for analysis.

As shown in Fig 1-4., ISC shows improvements on most of the evaluation cases. It is clearly seen that ISC advances the traditional method as the channel loss rate increases or the packet correlation rate decreases. This is due to the fact that ISC reduces impact of packet losses to the GOV by isolating errors to one of the two sub-sequences and decreases frame replacement distances for decoder failed frames. However, with the increment of the packet correlation constants, the frequency of the long error bursts increases, hence increases the chance that both sub-sequences are impacted by the long error bursts. In addition, when comparing the two different correlation model sets, the generic correlation model shows competitive results, and it is reasonable to use the generic model in cases when the actual temporal correlation for a given sequence is not feasible to compute.

V. CONCLUSION

In this paper, we have evaluated the performance of an *interleaved source coding* (ISC) method for packet video over channels with memory. ISC is resilient to packet losses since it limits the errors due to packet losses to one of two sub-sequences (generated by ISC) and minimizes cascaded effects of packet losses over a *single* erasure-channel model and increases the number of successfully decoded frames and overall playback quality of the decoded video sequence. It is clearly shown that ISC advances traditional method for most of the cases; however, the performance highly depends on the channel's loss rate and the packet correlation increases. Since most of the packet losses over the best-effort network channel are caused by the buffer overflow of the routers on the path, hence the high frequency, short burst errors are more likely to be observed, therefore, adopting ISC to the realtime streaming services over the best-effort network would be beneficial.

VI. REFERENCES

[1] Begen, A., Altunbasak, and Y., Ergun, O., "Multi-path selection for multiple description encoded video streaming," *IEEE Proc. ICC*, May 2003.

[2] Gallager, R., *Discrete Stochastic Processes*, Kluwer Academic Publishers, 1996.

[3] Girod, B., "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Trans. Image Processing*, vol. 9, no. 2, pp. 173-183, February 2000.

[4] Khayam, S. and Radha, H., "Markov-based Modeling of Wireless Local Area Networks," *ACM Mobicom Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, September 2003

[5] Lee, J. and Radha, H., "Interleaved Source Coding (ISC) for Predictive Video over ERASURE Channel," *Packet Video*, December 2004.

[6] Lee, J. and Radha, H., "Interleaved Source Coding (ISC) for Predictive Video over the Internet," *IEEE Proc. ICC,* May, 2005.

[7] Lee, J. and Radha, H., "Evaluation of the Interleaved Source Coding (ISC) under Packet Correlation," *IEEE Proc. ICME,* July, 2005.

[8] Lin, S. and Wang, Y., "Error resilience property of multihypothesis motion-compensated prediction," *IEEE Proc. ICIP*, September, 2002.

[9] Puterman, M., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc., New York, NY, 1994.

[10] Radha, H., van der Scharr, M., and Chen, Y., "The MPEG-4 FGS video coding method for multmedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, issue 1, pp. 53-68, March 2001.

[11] Yajnik, M., Moon, S., Kurose, J., and Towsley, D., "Measurement and modeling of the temporal dependence in packet loss", *IEEE Proc. INFOCOM*, 1999.

[12] Yee, J., Weldon, E., "Evaluation of the performance of error-correcting codes on a Gilbert Channel," *IEEE Trans. Communications*, vol. 43, pp. 2316-2323, August, 1995.

[13] Wu, M. and Radha, H. "Network Embedded FEC over Multi-hop Wireless Channels with Memory," (To Appear), *IEEE Proc. ICC,* May, 2005.