RELIABLE VIDEO COMMUNICATION WITH MULTI-PATH STREAMING USING MDC

Ivan Lee
School of Elec. and Info. Engineering
University of Sydney, NSW, Australia, 2006
Email: ivanlee@ieee.org

Ling Guan
Department of Elec. and Comp. Engineering
Ryerson University, Ontario, Canada, M5B-2K3
Email: lguan@ee.ryerson.ca

ABSTRACT

Video streaming demands high data rates and hard delay constraints, and it raises several challenges on today's packet-based and best-effort internet. In this paper, we propose an efficient multiple-description coding (MDC) technique based on video frame sub-sampling and cubic-spline interpolation to provide spatial diversity, such that no additional buffering delay or storage is required. The frame dropping rate due to packet loss and drifting error under the multi-path streaming environment is analyzed in this paper.

I. INTRODUCTION

Tremendous amount of research has been taken place to ensure the quality of service for video streaming applications. Techniques such as Integrated Service (IntServ), Differential Service (DiffServ), and Resource Reservation Protocols (RSVP) have been standardized. These techniques require the support on the network backbone, which is not readily available on today's Internet. Providing these services require higher computational resources on the network equipments for additional packet processing, which is a costly practice. In addition, service providers need to sustain a non-interrupted service during the upgrade process, which is another difficult task to deploy such techniques.

To get around with the network backbone upgrade while maintaining a low packet loss rate, numerous distributed streaming techniques have been proposed. Layered coding techniques such as Fine Granularity Scalable (FGS) coding [1] and Progressive Fine Granularity Scalable (PFGS) coding [2] divides the video bitstream with different priority levels, which is best applied with DiffServ technique mentioned above. If DiffServ is unavailable on the backbone, it is also possible to appoint statistical feedback from the receiver to the sender to adjust the amount of video layers to be transmitted. The major drawbacks of this approach are the round-trip feedback delay and the capability of adapting bandwidth varying channels. To overcome these drawbacks, combining source coding and distributed streaming infrastructure have been proposed in [3][4], with a layered coding technique applied to the video bitstream and a collaborative streaming infrastructure.

Multiple-Description Coding (MDC) [5] is another approach proposed to address some of the challenges described above. MDC is a source coding technique in which the source is encoded into multiple descriptions (MD), which can be transmitted via independent paths to the receiver. Vaishampayan proposed a Multiple Description Scalar Quantization (MDSQ) technique in [6], to generate two sub-streams by producing two indices for each quantization level [7].

This paper introduces an efficient MDC technique based on spatial domain up-scaling or super-resolution. The coding efficiencies and the reconstructed video qualities of the proposed MDC technique are further explored. The multi-path video streaming framework is simulated using the two-state Markovian Gilbert-Elliot Model, and comparisons between the frame loss rates by single-path streaming and multiple-path streaming are analyzed in this paper. Other factors affecting the frame loss rate, such as the number of MDC streams and the transition parameters of the Gilbert-Elliot Model, are further analyzed in the experiment.

II. MULTI-DESCRIPTION VIDEO CODING WITH SPATIAL DIVERSITY

The Internet consists of arbitrary interconnections between sources and destinations with heterogeneous propagation delays. Most of the routing protocols today select the best path for data transmission, and a path handover will take place upon any failure or congestions appearing at the network. Unfortunately, the delay for the routing update during the path handover can produce severe impact to real-time applications such as video streaming. Therefore, source coding technical with embedded error resilience has become a popular approach to resolve errors caused by the network connections.

MDC technique is one attempt to resolve the drawback caused by single path transmission. Given a sequence of video frames $\{F_1,F_2,\ldots\}$, where F_k represents the k-th frame of the video. Using CIF video format for example: $F_k = \{y,u,v\}$, $\forall y \in [0,255]^{352x288}$

0-7803-9332-5/05/$20.00 ©2005 IEEE

and $\forall u, v \in [0,255]^{176x144}$. MDC consists of an encoder E_{MDC} and a decoder D_{MDC}. The MDC encoder E_{MDC} maps a given frame into m codes:

$$E_{MDC} : F_k \rightarrow \{C^1_{MDC,k}, \ldots, C^m_{MDC,k}\} \quad (1)$$

The MDC decoder D_{MDC} maps a subset of $\{x_1, \ldots, x_n\} \subset \{1, \ldots, m\}$ codes into a reconstructed frame, $\hat{F}^n_k$:

$$D_{MDC} : \{C^{x_1}_{MDC,k}, \ldots, C^{x_n}_{MDC,k}\} \rightarrow \hat{F}^n_k \quad (2)$$

with the property that

$$d(F_k, \hat{F}^m_k) \leq d(F_k, \hat{F}^n_k) \quad (3)$$

where distortion metric $d(F_1,F_2)$ represents the similarities between the two video frames F_1 and F_2, for example, mean squared error (MSE) or peak signal to noise ratio (PSNR), and a lower $d(F_1,F_2)$ value represents a higher similarity between F_1 and F_2.

In this paper, an efficient super-resolution based video MDC codec is constructed based on the strong inter-pixel similarity between neighboring pixels within a video frame. In the encoding stage, E_{MDC} performs spatial sub-sampling for each individual frame in the video sequence.

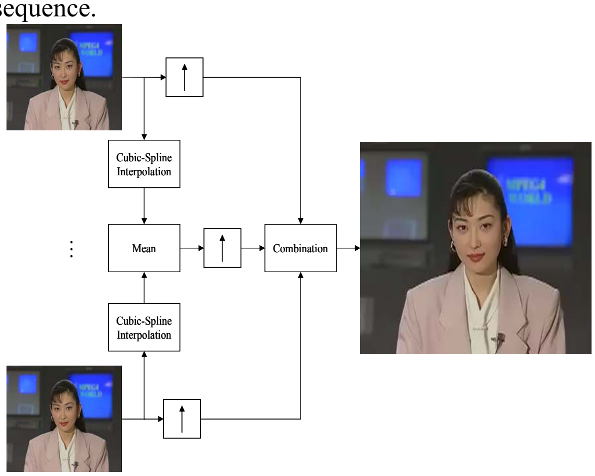


Figure 1: MDC video reconstruction using Cubic Spline Interpolation

	Raw (kbps)	MDC (kbps)	ζ	PSNR (dB)
Akiyo	126.24	68.91	1.8320	33.10
		137.70	0.9168	36.36
		206.18	0.6123	36.88
		274.60	0.4597	37.82
Foreman	464.51	293.15	1.5845	28.19
		589.70	0.7877	28.87
		887.22	0.5236	32.96
		1188.02	0.3910	34.70
News	271.64	141.89	1.9144	28.69
		284.11	0.9561	32.37
		425.22	0.6388	34.13
		566.92	0.4792	36.40

Table 1: Coding Efficiencies for MDC with Temporal Diversity

The down-sampled video sequences are encoded using conventional video codec. In our experiment, advanced video codec standard using H.264 is chosen. The decoded sequences by H.264 are used to reconstruct the video. The missing MDC streams are predicted from the received MDC streams using cubic-spline interpolation. If there are more than one received MDC streams, the mean of the predicted frame will be used. Figure 1 shows the block diagram of the proposed D_{MDC} process.

MDC provides higher error resilience to the raw codec scheme, with the trade-off in coding efficiency. The coding efficiency ζ is defined in Eqn (4).

$$\varsigma = r \Big/ \sum_{m=1}^{M} r_m \quad (4)$$

Where r denotes the bitrate for the raw video codec, r_m denotes the bitrate for stream in the MDC. Coding efficiencies of the proposed MDC technique for the standard Akiyo, Foreman, and News sequences are summarized in Table 1.

III. MULTI-PATH TRANSMISSION AND ITS IMPACT ON RECONSTRUCTED VIDEO QUALITY

The two-state Markovian Gilbert-Elliot Model is used as the channel model [8]. This model has been shown to be able to effectively capture the bursty packet loss behavior. The two states of this model are denoted as S_0 (good) and S_1 (bad). S_0 (good) represents the state where packets are received correctly and satisfies real-time constraint. S_1 (bad) represents the state where packets are missing for reconstructing the video. The state transition probabilities for S_0-to-S_1 and S_1-to-S_0 are denoted as P_{01} and P_{10}, respectively. The probability of remaining in the same state are denoted as P_{00} and P_{01} where P_{00}=1-P_{01} and P_{11}=1-P_{00}. $P_x \in \Re$ and $0 < P_x < 1$, $\forall x \in \{00, 01, 10, 11\}$. Steady state analysis of being in state S_0 and S_1 satisfies the criteria in Eqn (5).

$$\begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} P_0 \\ P_1 \\ 1 \end{bmatrix} \quad (5)$$

where P_0 and P_1 are the steady state probabilities of being in S_0 and S_1 states, respectively. Solving Eqn (5), the probabilities of successful and unsuccessful packet transmissions can be represented in Eqn (6) and Eqn (7).

$$P_0 = \frac{P_{11} - 1}{P_{00} + P_{11} - 2} \quad (6)$$

$$P_1 = \frac{P_{00} - 1}{P_{00} + P_{11} - 2} \quad (7)$$

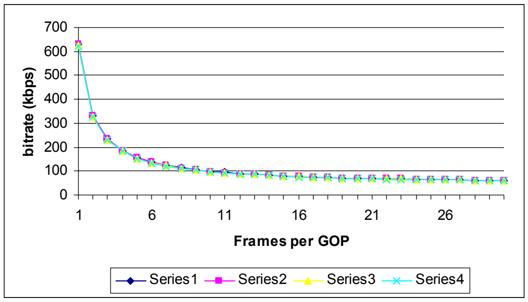


Figure 2: Video bitrate for the down-sampled Akiyo sequence at different I-intervals

H.264 baseline profile is used in our analysis, and only inter- (I) and predictive (P) frames are used in the bitstream. I-frames are encoded as discrete frames, independent of adjacent frames. P-frames are encoded with respect to a past I-frame or P-frame with motion vector prediction and compensation. A group of frames that starts with an I-frame and ends at the frame before the next I-frame is called a Group of Pictures (GOP), denoted as *T*. Losing an I-frame or P-frame will corrupt the remaining frames within the same GOP in the reconstructed video sequence, and this is referred as the drifting error. Figure 2 shows the bitrate for each MDC stream with different lengths of GOP.

Considering the drifting error, each packet loss will propagate until the end of the GOP. Take the extreme case for example, if the I-frame is dropped, the remaining P-frames cannot be reconstructed the output video without a reference frame for motion prediction, and the whole GOP will be missing. In general, the probability of receiving n frames followed by a lost frame is shown in Eqn (8).

$$S_0 \cdot P_{00}^{n-1} \cdot P_{01} \quad (8)$$

The frame dropout rate for receiving n frames followed by a lost frame is given by

$$\frac{(T-n)}{T} \quad (9)$$

It is easy to show that the mean frame dropout rate is given by Eqn (10), where the first term represents a lost I-frame and the second term shows that several I- and P-frames are received before a lost P-frame.

$$P_1 + P_0 \sum_{n=1}^{T-1} \frac{(T-n)P_{00}^{n-1} P_{01}}{T}$$
$$= P_1 + P_0 \sum_{n=1}^{T-1} \frac{(T-n)P_{00}^{n-1}(1-P_{00})}{T}$$
$$= P_1 + \frac{P_0(1-P_{00})}{T}\left(T\frac{P_{00}^{T-1}-1}{P_{00}-1} - \sum_{n=1}^{T-1} nP_{00}^{n-1}\right) \quad (10)$$

Let $H = \sum_{n=1}^{T-1} nP_{00}^{n-1}$ (11)

and H can be expanded as

$$H = P_{00}^0 + 2P_{00}^1 + 3P_{00}^2 + \ldots + (T-1)P_{00}^{T-2}$$
$$= \sum_{i=1}^{T-1}\left(\sum_{j=i}^{T-1} P_{00}^{j-1}\right)$$
$$= \frac{1}{(P_{00}-1)^2}\left(T \cdot P_{00}^T - T \cdot P_{00}^{T-1} - P_{00}^T + 1\right) \quad (12)$$

Substituting H back to Eqn (11) yields

$$P_1 + \frac{P_0(1-P_{00})}{T}\left(T\frac{P_{00}^{T-1}-1}{P_{00}-1} - \frac{T \cdot P_{00}^T - T \cdot P_{00}^{T-1} - P_{00}{}^T + 1}{(P_{00}-1)^2}\right)$$
$$= P_1 + P_0 + P_0\frac{(P_{00}^T - 1)}{T(1-P_{00})} \quad (13)$$

Next, let us consider the case the video is divided into M streams using MDC, and transmitted over M-independent paths. Assume all the paths have identical loss paths, and therefore share the same state transition parameters for the Gilbert Model.

For MDC with M independent streams, output video can be reconstructed from any stream. The video frames cannot be reconstructed only if all streams are lost. Denote the GOP period of the MDC as *T*. Then the probability that a stream loses a frame at or before t_1 is:

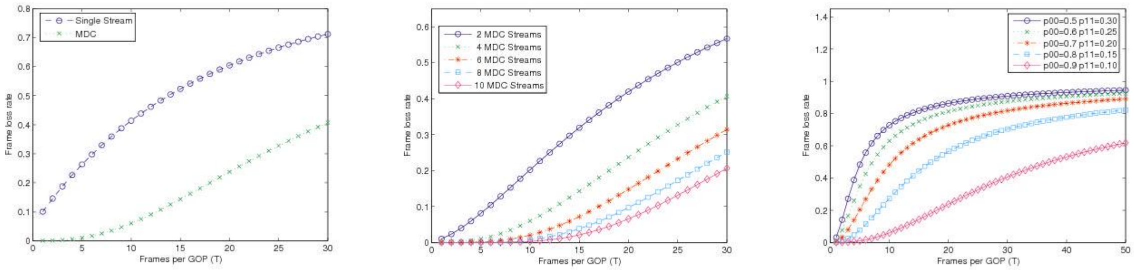
$$P_1 + P_0 P_{00}^0 P_{01} + P_0 P_{00}^1 P_{01} \ldots + P_0 P_{00}^{t_1-1} P_{01}$$
$$= P_1 + P_0 P_{01}\frac{P_{00}^{t_1}-1}{P_{00}-1} \quad (14)$$

Hence, the probability of losing all M streams at but not before t_1 is:

$$\left(P_1 + P_0 P_{01}\frac{P_{00}^{t_1}-1}{P_{00}-1}\right)^M - \left(P_1 + P_0 P_{01}\frac{P_{00}^{t_1-1}-1}{P_{00}-1}\right)^M \quad (15)$$

(a) 1 MDC stream (PSNR Y = 28.80 dB) | (b) 2 MDC streams (PSNR Y = 32.88 dB) | (c) 3 MDC streams (PSNR Y = 34.63 dB) | (d) 4 MDC streams (PSNR Y = 37.16 dB)

Figure 3: Reconstructed News video frame, frame 1



(a) (b) (c)

Figure 4: (a) Frame Loss Rate for single stream and 4 MDC streams (b) Comparisons of the Frame Loss Rate for different level of multiple description coding (c) Effect of state transition probability to the frame loss rate

It can be shown that the mean frame dropout rate for MDC is given by Eqn (16)

$$P_1^M + \sum_{n=1}^{T_M-1}\left[\left(P_1 + P_0P_{01}\frac{P_{00}^n-1}{P_{00}-1}\right)^M - \left(P_1 + P_0P_{01}\frac{P_{00}^{n-1}-1}{P_{00}-1}\right)^M\right]\frac{(T_M-n)}{T_M}$$

$$= \frac{1}{T}\sum_{n=1}^{T}\left(P_1 + P_0P_{01}\frac{P_{00}^{n-1}-1}{P_{00}-1}\right)^M \qquad (16)$$

IV. EXPERIMENTS

In this paper, we divide CIF video sequences into four sub-sampled QCIF sequences. Each QCIF sequence is coded using H.264. The decoded sequences are assembled using cubic-spline interpolation, as described in section II. Examples of reconstructed MDC video frames are shown in Figure 3 (News sequence, frame 1).

The frame loss rate of the single video stream transmission and the MDC transmission using the Gilbert model are shown in Eqn (14) and Eqn (17), respectively. Both equations show the frame loss rate is a function of length of GOP (*T*) and the transition probabilities (P_{01} and P_{10}). As shown in Figure 4(a), the frame loss rate increases with *T*, due to a higher drifting error associated with a larger *T* value. We observe that for all *T* values, MDC outperforms single stream transmission in terms of frame dropping rate (the lower the frame loss rate, the higher the perceived video quality). Figure 4(b) compares different number of MDC streams. We observe that the more MDC streams are used - a higher value for *m* in Eqn (1), the lower the frame loss rate. Figure 4(c) compares different transition probabilities P_{00} and P_{11} and their impacts to the frame loss rate. We observe that lowering increasing P_{00} (higher probability of staying in good state) and decreasing P_{11} (lower probability of staying in bad state) yield lower frame loss rate.

V. CONCLUSIONS

In this paper, we propose an efficient MDC video coding technique with spatial diversity. Sub-sampled video frames are H.264 encoded, transmitted over multiple paths, H.264 decoded, and reconstructed using cubic-spline interpolation. The proposed MDC trades-off the coding efficiencies with the improvement in the frame loss rate, yielding a better perceived video quality.

REFERENCES

[1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard", IEEE Tran. on Circuits and Systems for Video Technology, vol. 11, pp. 301-317, Mar. 2001.

[2] F. Wu, S. Li, and Y-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," IEEE Tran. on Circuits and Systems for Video Technology, vol. 11, pp. 332-344, Mar. 2001.

[3] I. Lee, and L. Guan, "Centralized peer-to-peer streaming with layered video", IEEE Int. Conf. on Multimedia and Expo, Baltimore, USA, vol. 1, pp. 513-516, Jul 2003.

[4] I. Lee, and L. Guan, "Centralized Peer-to-Peer Streaming for PFGS Video", Pacific-Rim Conference on Multimedia, vol. I, pp. 131-138, Tokyo Waterfront City, Japan, Dec 2004.

[5] V. K. Goyal, "Multiple description coding: compression meets the network", IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[6] Vinay Anant Vaishampayan, "Design of Multiple Description Scalar Quantizers", IEEE Tran. on Information Theory, vol. 39, pp.821-834, May 1993.

[7] Vivek K Goyal, J. Kovacevic, and M. Vetterli, "Multiple description transform coding: Robustness to erasures using tight frame expansions", Proc. IEEE, Aug. 1998.

[8] J. R. Yee and E. J. Weldon, "Evaluation of the performance of Error-Correcting Codes on a Gilbert Channel", IEEE Trans. Comm., vol. 43, pp. 2316-2323, Aug. 1995.