Rate-Distortion Optimized Video Summary Generation and Transmission Over Packet Lossy Networks

Peshala V. Pahalawatta[a], Zhu Li[b], Fan Zhai[c], and Aggelos K. Katsaggelos[a]

[a]Department of Electrical & Computer Engineering
Northwestern University, Evanston, IL 60208;
[b]Multimedia Research Lab (MRL) Motorola Labs, Schaumburg, IL 60196;
[c]Texas Instruments, Dallas, TX 75243;

ABSTRACT

The goal of video summarization is to select key frames from a video sequence in order to generate an optimal summary that can accommodate constraints on viewing time, storage, or bandwidth. While video summary generation without transmission considerations has been studied extensively, the problem of rate-distortion optimized summary generation and transmission in a packet-lossy network has gained little attention. We consider the transmission of summarized video over a packet-lossy network such as the Internet. We depart from traditional rate control methods by not sacrificing the image quality of each transmitted frame but instead focusing on the frames that can be dropped without seriously affecting the quality of the video sequence. We take into account the packet loss probability, and use the end-to-end distortion to optimize the video quality given constraints on the temporal rate of the summary. Different network scenarios such as when a feedback channel is not available, and when a feedback channel is available with the possibility of retransmission, are considered. In each case, we assume a strict end-to-end delay constraint such that the summarized video can be viewed in real-time. We show simulation results for each case, and also discuss the case when the feedback delay may not be constant.

Keywords: video summarization, video streaming, Internet

1. INTRODUCTION

Video summarization is the process of generating a shorter version of an original video sequence. The summarized sequence contains key frames extracted from the original video sequence based on some criterion specified by the user. Video summarization may be required when a system is operating under limited bandwidth conditions, or under tight constraints on viewing time or storage capacity. A typical application that could require the generation and transmission of summarized video would be a remote surveillance application in which video must be recorded over long lengths of time, and important video segments must be transmitted to a base station in real-time in order to be viewed by a human operator.

Video summarization is a widely studied topic. Comprehensive reviews of past work on video summarization can be found in Ref. 1 and Ref. 2. Most work on video summarization techniques consist of segmenting a video sequence into a series of video shots. Then, one or more key frames are selected to represent each video shot. In Refs. 1 and 3, clustering methods are used, in which each cluster of a video sequence is represented by a key frame. Although these methods tend to provide a compact summary of the visual content in each image of the video sequence, they do not capture the temporal information that is also representative of the sequence. Other efforts at video summarization can be found in Refs. 4, 5, 6. However, these methods tend to be computationally complex and are difficult to implement for real-time video transmission.

In order to combat some of the weaknesses of the above methods, Refs. 7 and 8 developed a video summarization scheme that formulates the problem as a rate-distortion optimization. The key observation in this scheme is that the distortion of the summarized video sequence is related to the number of frames that are used

Further author information: (Send correspondence to P.V.P.)
P.V.P.: E-mail: pesh@ece.northwestern.edu, Telephone: 1 847 912 8479

Image and Video Communications and Processing 2005, edited by Amir Said,
John G. Apostolopoulos, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 5685 © 2005 SPIE and IS&T · 0277-786X/05/$15

in the summary, as well as the choice of frames. Therefore, given a constraint on the rate allocated for the video sequence, the optimal set of summary frames could be found using dynamic programming methods.

In Ref. 9, real-time video streaming over a wireless channel is considered. The authors use a two step process where the channel conditions are first estimated based on available feedback statistics, and then, the optimal summary is found that satisfies a rate constraint for lossless transmission. However, delay constraints on the individual video frames in real-time Internet video streaming applications make it difficult to provide lossless transmission of every frame in the sequence. Therefore, in this paper, we extend our previous work in Ref. 7 to the case when the generated video summary must be transmitted over a packet lossy network.

In this paper, our approach is to take into account packet loss and use the expected end-to-end distortion to evaluate the video delivery quality at the sender side. Our goal is to minimize the total end-to-end distortion for the whole sequence by optimally dropping frames in order to meet the constraints on bandwidth and viewing time. This problem can also be regarded as temporal rate control with frame dropping. Note that this approach differs from the traditional rate control methods such as that in Ref. 10 in that video summary applications do not sacrifice the quality of each video frame but, instead, they drop frames to meet the bandwidth constraint. We also consider the case when the sender receives feedback and can retransmit frames/packets based on whether or not they were received after the previous transmission. Similar work has been shown in Ref. 11 for non-video summary applications but unlike in Ref. 11, we take into account the effects of error concealment on the end-to-end distortion of the sequence.

This paper is organized as follows. In Sec. 2 we discuss some preliminary details and assumptions about our system. This section details how the video summary is defined, and also discusses the metric used to find the expected distortion of the video sequence. In Sec. 3, the problem of video summary transmission under stringent end-to-end delay constraints and limited feedback is considered, where the sender cannot retransmit frames. In Sec. 4, we discuss the problem of video summary generation and transmission when feedback is available with the possibility of retransmission. Simulation results are shown in Sec. 5, and we end with some ideas for future work in Sec. 6.

2. PRELIMINARIES

2.1. Internet Video Summary

Let a video sequence of n frames be denoted by, $V = \{f_0, f_1, ..., f_{n-1}\}$. Let its video summary of m frames be $S = \{f_{l_0}, f_{l_1}, ..., f_{l_{m-1}}\}$, in which l_k denotes the k^{th} frame selected for the summary S. The summary S is completely determined by the frame selection process,

$$L^m = \{l_0, l_1, ..., l_{m-1}\}, \tag{1}$$

which has an implicit constraint that $l_0 < l_1 < ... < l_{m-1}$.

The reconstructed sequence $V_S^{'} = \{f_0^{'}, f_1^{'}, ..., f_{n-1}^{'}\}$ from the summary S is obtained using a zero-order hold technique by substituting missing frames with the most recent frame that belongs to the summary S. Therefore,

$$f_k^{'} = f_{i=max(l):s.t.l\in\{l_0,l_1,...,l_{m-1}\},i\leq k} \tag{2}$$

The temporal rate of the summarized sequence is defined as the ratio of the number of frames in the summarized sequence to the total number of frames in the original sequence. Therefore,

$$R(S) = \frac{m}{n} \tag{3}$$

2.2. Packetized Real-Time Video Transmission

In this paper, we assume that the summary frames are first packetized into video packets at the encoder such that each packet consists of one frame from the summary. We also assume that the frames are inter coded using previous summary frames as reference pictures. Therefore, the packets are dependent on each other.

We also assume that there is an initial setup time T_{max}, which is the time between the transmission of the packet at the encoder, and the playback of that packet at the decoder. Once playback begins, all video frames must arrive at the decoder buffer on time to be played back in sequence.

If a packet does not arrive at the decoder on time, it is considered lost. We assume that the packet losses can be modeled by a Bernoulli process. Therefore, each packet may be lost with some probability, ρ,as a result of the current network state, and the losses can be assumed to be independent of each other.

2.3. Expected Distortion

The temporal expected distortion of the video summary, S, can be calculated as,

$$E[D(S)] = \sum_{k=0}^{n-1} E[d(f_k, f_k^{'})], \tag{4}$$

where $d(f_k, f_k^{'})$ is the distortion between frame k and its corresponding summary frame from $V_S^{'}$. For the purposes of this paper, we use the mean squared error (MSE) between the two frames as the metric for calculating the distortion.

The expectation occurs due to the probability of loss that can be incurred by any given summary frame. Since we assume that the summary frames are inter coded, even a received summary frame will incur some random distortion at the decoder. The expected distortion between a pixel x_k in frame f_k of the original sequence, and pixel $x_k^{'}$ in frame $f_k^{'}$ of the reconstructed sequence can be found as:

$$E[d(x_k, x_k^{'})] = E[(x_k - x_k^{'})^2] \tag{5}$$

which, when expanded into its individual terms, becomes:

$$E[d(x_k, x_k^{'})] = x_k^2 - 2x_k E[x_k^{'}] + E[x_k^{'\,2}] \tag{6}$$

noting that x_k is known at the encoder. Therefore, when using the MSE criterion for distortion, only the first and second order moments of the probability distribution of each reconstructed pixel is required in order to calculate the distortion. Given that the first frame in the sequence is guaranteed to have been received at the decoder, and using the simple zero-order hold error concealment scheme for frames that are not in the summary, the first and second order moments, $E[x_k^{'}]$ and $E[x_k^{'\,2}]$ can be recursively calculated at the encoder using an algorithm called ROPE (Recursive Optimal Per-pixel Estimate) Ref. 10.

3. CASE 1: STRINGENT END-TO-END DELAY CONSTRAINTS WITH NO FEEDBACK

3.1. Problem Formulation

As we mentioned in Section 2.2, a packet must arrive by its scheduled playback time in order to be played as part of the video sequence. Even when some buffering can be afforded at the decoder, if the end-to-end delay constraint is stringent, the encoder cannot afford to retransmit past frames, or wait for feedback. However, even in this case, the encoder can be provided with some general feedback, which it can use to determine the network state, and thereby, the probability of loss per packet.

In order for the k^{th} frame to be useful, it must arrive at the receiver by its playback deadline. If we assume that the time to decode one frame of the video summary is T_F, and the initial setup time is T_{max}, then the delay constraint for the k^{th} frame of the original video sequence can be written as:

$$T_k \leq T_{max} + kT_F \tag{7}$$

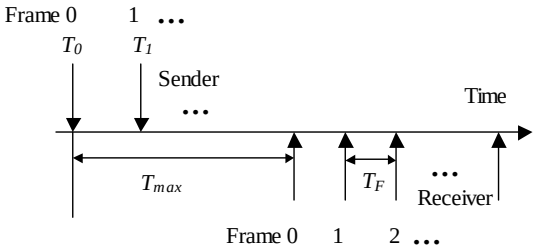

Figure 1. Illustration of the end-to-end delay constraint with initial setup time of T_{max} and image decoding time of T_F.

where T_k is the instant that the k^{th} frame starts to be transmitted from the sender. This delay constraint is illustrated in Fig. 1. Thus, we want to minimize the total distortions for the n frames of the original sequence subject to the delay constraints for each of them. Note that not every frame has to meet its corresponding delay constraint. Only the transmitted (summary) frames do. Those packets that are not transmitted do not meet the delay constraints by default. Then, the problem can be formulated as:

$$\begin{aligned} \min \quad & \textstyle\sum_{k=0}^{n-1} E[D_k] \\ s.t. \quad & T_k \leq T_{max} + kT_F, \quad k = 0, 1, ..., n-1 \end{aligned} \tag{8}$$

where $E[D_k]$ is the expected distortion for frame k.

While the original problem can be formulated as above, in a practical system, all the frames of the original sequence may not be available at the sender prior to the beginning of transmission. Therefore, a common tactic employed in such systems is to perform the optimization in stages, as the video frames become available at the encoder Ref. 12. In this case, the optimization can be performed over a window of size A frames while keeping track of the delay constraints for the frames in that window.

In our case, we can assume that a higher level rate controller can determine the size of the decoder buffer, and that it allocates the temporal rate for the video summary sequence based on T_{max} and the decoder buffer size, i.e., for a small window of size A in the original video sequence, the rate controller will determine the maximum allowable temporal rate, $R_{max} = \frac{m}{A}$, such that each summary frame will arrive prior to its playback deadline. Therefore, the optimization problem for each frame window can be written as:

$$\begin{aligned} \min_{\Pi} \quad & \textstyle\sum_{k=0}^{A-1} E[D_k] \\ s.t. \quad & \textstyle\sum_{k=0}^{A-1} \pi_k \tau_k \leq T_{max,w} \quad \pi_k \in \{0, 1\} \end{aligned} \tag{9}$$

where $\Pi = \{\pi_0, \pi_1, ..., \pi_{A-1}\}$ such that, $\pi_k = 1$ implies that frame k in the window w is in the transmitted summary, and $\pi_k = 0$ implies otherwise. τ_k denotes the time taken to transmit frame k, and $T_{max,w}$ denotes the transmission deadline for window w, which will be a function of T_{max} as determined by the rate controller. We assume that the variance in bit rate across frames is limited, and thus, τ_k is relatively constant across all the frames in the window. This allows the rate controller to specify the temporal rate constraint, R_{max}, for the optimization window.

3.2. Dynamic Programming Solution

In general, there are $\binom{A}{m} = \frac{A!}{m!(A-m)!}$ feasible solutions to the above summarization problem. Therefore, an exhaustive search solution would be prohibitive in a real-time situation. Another possibility would be a greedy

algorithm, in which the algorithm selects the first frame that minimizes the expected distortion of the sequence, given that no other frames are in the summary. Once this frame is picked, the algorithm would pick the next best frame given that the previous frame is already in the summary. However, it can be shown that the above method does not perform well even in the lossless case Ref. 7. Therefore, in our previous work, we have shown a dynamic programming (DP) solution to the lossless summarization problem. In this paper, we extend that solution while taking into account the packet loss probabilities, so that the summarization scheme will pick the best set of summary frames given that they, and their neighboring summary frames, may never arrive at the receiver. Given reasonably well-behaved sequences, this approach will arrive at the optimal solution in the packet lossy case.

Let the distortion state D_t^k be the minimum expected distortion incurred by a summary that has t frames and ends with frame f_k, i.e., $l_{t-1} = k$. Therefore,

$$D_t^k = \min_{l_0,l_1,\ldots,l_{t-2}} \sum_{j=0}^{A-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,k\}}] \tag{10}$$

where the distortion is calculated as in Eq. (6). Note that $l_{t-1} = k$ is removed from the minimization since we assume that frame k is being picked. Also, l_0 can be less than 0, in which case it would represent the last summary frame to be picked in the previous optimization window. $d(f_j, f_i)$ denotes the distortion between the original video frame and the summary frame which represents it. However, unlike in Ref. 7, in this case the pixel values in f_i are random, and their distribution depends on the probability of packet loss, and the error concealment technique. Observing that the choice of frame k to the summary does not affect the expected distortions for any of the previous frames, we can rewrite Eq. (10) as:

$$D_t^k = \min_{l_0,l_1,\ldots,l_{t-2}} \left\{ \sum_{j=0}^{k-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\}}] + \sum_{j=k}^{A-1} E[d(f_j, f_k)] \right\} \tag{11}$$

Now, by adding and subtracting the same term to the above equation, we get,

$$\begin{aligned} D_t^k = \min_{l_0,l_1,\ldots,l_{t-2}} \quad & \{\textstyle\sum_{j=0}^{k-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\}})] + \sum_{j=k}^{A-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\}})] \\ & - \textstyle\sum_{j=k}^{A-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\}})] + \sum_{j=k}^{A-1} E[d(f_j, f_k)]\} \end{aligned} \tag{12}$$

Therefore, the distortion state can be broken down to two parts as:

$$D_t^k = \min_{l_0,l_1,\ldots,l_{t-2}} \left\{ \sum_{j=0}^{A-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\},i\le j)})] - \sum_{j=k}^{A-1} E[d(f_j, f_{l_{t-2}}) - d(f_j, f_k)] \right\} \tag{13}$$

where the first part represents the problem of finding the minimum distortion summary that contains $t-1$ frames, while the second part represents the reduction in expected distortion given that frame k is added to the summary. Therefore, we can recursively find D_t^k as:

$$D_t^{k=l_{t-1}} = \min_{l_{t-2}} \left\{ D_{t-1}^{l_{t-2}} - \sum_{j=k}^{A-1} E[d(f_j, f_{i=max(l):s.t.l\in\{l_0,l_1,\ldots,l_{t-2}\}})] + \sum_{j=k}^{A-1} E[d(f_j, f_k)] \right\} \tag{14}$$

Note that the above frame selection assumes that the change in distortion due to adding a new summary frame after the k^{th} frame will only be coupled with the choice of the k^{th} frame but not with any previous frames in the summary. This is a reasonable assumption for most video sequences.

In Fig. 2, we show the ensuing distortion state trellis for a summarization scheme where $A = 5$ and $m = 3$. Each node in the trellis represents a distortion state, D_t^k, where t denotes the stage at which frame k is transmitted. The optimal frame selection can be obtained by backtracking from the node that has the minimum distortion at stage $t = 3$.

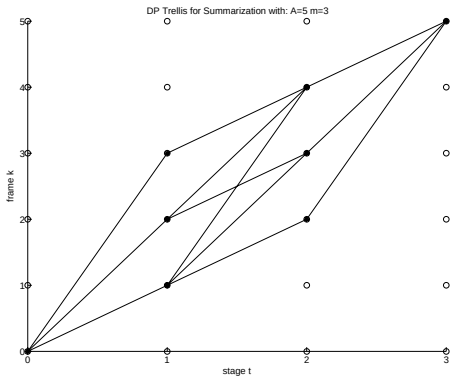

Figure 2. Distortion state trellis for optimal frame selection with temporal rate $\frac{m}{A}$.

4. CASE 2: SENDER DRIVEN RETRANSMISSION BASED ON FEEDBACK WITH DELAY CONSTRAINTS

4.1. Problem Formulation

In this case, we assume that feedback is available at the sender, and that each packet/frame can have multiple transmission opportunities. We simplify the problem by assuming a constant network delay, which ensures that the round trip time (RTT) is constant over an optimization period. We also assume that the feedback channel is error-free, which is a reasonable assumption, since the acknowledgments sent by the receiver will require significantly less bits than the actual source packets. Therefore, if an acknowledgment for a transmitted packet is not received by the sender within the packet's RTT, the sender assumes that the packet has not been received at all.

Note that in the video summarization scenario, unlike in a usual video transmission scenario, a frame can be skipped, so that the time saved from skipping one frame can be used for other frames that are more representative of the video sequence. Frame skipping will lead to higher expected distortion for the frame being skipped, but it will lower the expected distortion for transmitted frames. The transmission policy for the sequence can be written as, $\mathbf{\Pi} = \{\pi(0), \pi(1), ..., \pi(n-1)\}$. Here, $\pi(k)$ represents the transmission policy for the k^{th} frame where $\pi(k) = \{\pi_0(k), \pi_1(k), ..., \pi_{N_{k-1}}(k)\}$, $\pi_j(k) \in \{0, 1\}$, $j = 0, 1, ..., N_{k-1}$. $\pi_j(k) = 1$ implies that the k^{th} frame will be transmitted at the j^{th} transmission opportunity while $\pi_j(k) = 0$ implies otherwise. The maximum number of transmission opportunities for frame k, given as N_{k-1} can be found based on the end-to-end delay constraint T_{max}. We assume that only one frame can be transmitted at one transmission opportunity.

Our goal is to find the best transmission policy $\mathbf{\Pi} = \{\pi(0), \pi(1), ..., \pi(n-1)\}$ for the video sequence of n frames such that the total end-to-end distortion is minimized. Again, as in the previous case, performing the optimization for the entire sequence at once is both intractable and impractical. We can, however, assume that a higher level rate controller can take into account the decoder buffer, as well as other time constraints such as the initial setup time at the decoder, and specify a temporal rate at which to transmit the frames in a given window of the video sequence. Based on this assumption, we can write the optimization problem as:

$$\begin{aligned} \min_{\mathbf{\Pi}} \quad & E[D_k] \\ s.t. \quad & \textstyle\sum_{k=0}^{A-1} |\pi(k)| = m \end{aligned} \tag{15}$$

where $|\pi(k)| = \sum_{j=0}^{m-1} \pi_j(k)$ is the total number of times that frame k is transmitted, and the temporal rate specified by the higher level rate controller for that frame window is $\frac{m}{A}$. Note that the transmission policy also includes the summary policy, in that, if $\pi(\mathbf{k}) = \{0, 0, ..., 0\}$, that would be equivalent to skipping frame k of the original sequence when selecting the summary frames.

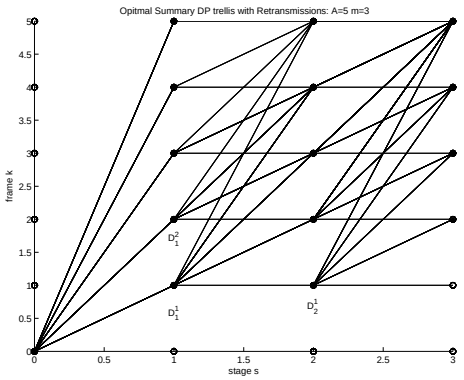

Figure 3. Distortion state trellis for optimal frame selection with retransmissions.

The probability of loss for a particular frame now depends on the number of times it is transmitted, after the last acknowledgment (ACK) is received. In this work, we assume a Bernoulli model for the probability of loss. Thus, the probability of loss for a packet that has been transmitted k times after the last ACK will be ρ^k, where ρ is the probability of loss for each transmission. It must also be noted that the optimization above can take into account the feedback received from previously transmitted frames in order to determine $E[D_k]$ for future frames. This is due to the fact that the frames are inter coded, and if acknowledgments, ACKs, or NACKs (negative acknowledgments) are available at the sender, then the sender can perfectly reconstruct the state at the decoder at the times that the ACKs/NACKs were transmitted. Then, future frames can be encoded using this information about the reconstructed frames at the decoder.

4.2. Algorithmic Solution

The truly optimal solution to the above problem using exhaustive search is intractable since it grows exponentially with the number of frames, and the number of transmission opportunities. However, assuming that the RTT is a small multiple of the frame transmission time, we extend our solution to the previous problem, and find a greedy solution that optimizes the frame selection given the current known state at the decoder.

Given a sequence of frames that need to be summarized and transmitted within a time constraint, we use the dynamic programming method discussed in Section 3.2 to develop a three-step approach to solve the problem. In the first step, we generate a video summary with retransmissions of summary frames using the DP approach. Then, we develop a transmission policy for the generated video summary. Finally, we dynamically update the optimization as feedback is received in the form of an ACK.

For the first step of generating a video summary with retransmissions, the major difference between this case and the previous one is in the topology of the DP trellis. Fig. 3 shows the new topology for the case when retransmissions are allowed. This is equivalent to relaxing the implicit constraint on the summary frames such that $l_0 \leq l_1 \leq ... \leq l_{m-1}$.

Although an optimal summary policy with retransmissions can be obtained using the above method, it does not specify the transmission policy for the frames. For a scenario where the RTT is a small multiple of one packet's transmission time, we consider an efficient packet transmission policy that will ensure that the transmission opportunities available to the sender will be optimally utilized. The algorithm essentially ensures that at any given transmission opportunity, first preference for transmission will be given to the earliest frame that has not been transmitted within the last RTT.

The above transmission policy is continued until an ACK is received by the sender. Then, it will redo the optimization using the remaining frames in the encoder buffer, while taking into account the acknowledgment, and also taking into account the frames that have already been transmitted but not acknowledged. Essentially, this process can be considered a greedy solution to the above problem, where the solution is dynamically adjusted based on feedback from the receiver.

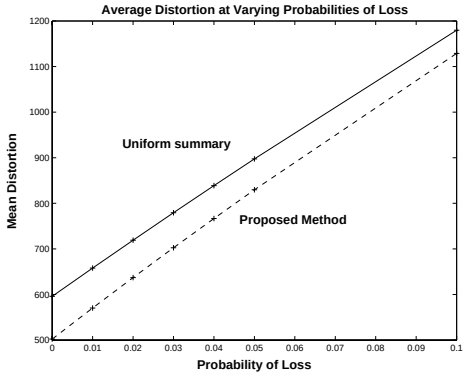

Figure 4. Comparison of average end-to-end distortion between uniform summary and optimal summary.

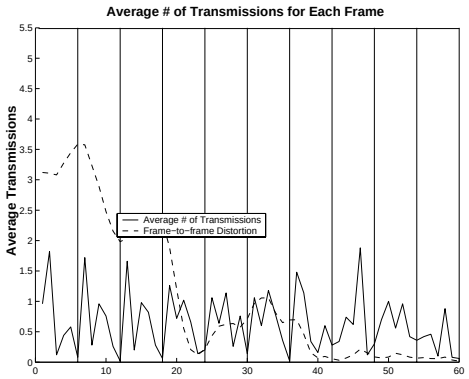

Figure 5. Average number of transmissions plotted along with a scaled plot of frame-by frame distortion. The vertical lines denote the optimization window boundaries. Probability of packet loss was set at 0.2.

5. SIMULATIONS

5.1. Case 1

For our simulations, we used an H.263+ codec Ref. 13 to perform the source coding of the video sequence, and we considered frames 210-270 of the QCIF (176x144) foreman sequence. The window length, A, was set at 15 frames, while the temporal rate for the summary, $\frac{m}{A}$ was set at $\frac{3}{15}$. The average distortions were calculated for the optimal summary and compared with a uniform summarization method that uniformly picks one of every 5 frames for the summary. Fig. 4 shows the average distortion of the optimal summary compared to that of a uniform summarization method.

5.2. Case 2

Simulations for this case were also done using the same video sequence. The window length for this case was set at 6 frames, while the temporal rate was set at $\frac{4}{6}$. RTT was set at twice the transmission time of a packet. Fig. 5 shows the average number of transmissions of each frame in the original video sequence, which is averaged over multiple loss realizations. A scaled plot of the frame-by-frame distortion, which is defined as the distortion between two consecutive frames in the original sequence is shown alongside the average number of transmissions. It can be seen that within a transmission window, the algorithm generally chooses to retransmit the key frames more often than others.

6. CONCLUSIONS AND FUTURE WORK

We have shown two schemes for rate-distortion optimized video summarization and transmission in packet-lossy networks. We have considered the scenarios when no feedback is available at the sender, and when feedback is available with the possibility of retransmitting summary frames. In the second scenario, we have assumed that the network delay is constant. However, a third important case is when the network delay is not constant. In this case, the RTT can be a random variable with some distribution, and if an ACK is not received, then the probability of loss of a packet must be calculated based on the time elapsed since it was transmitted.

We have also limited our discussion to the MSE based distortion metric. However, the perceptual validity of this metric for video summarization applications can be questioned. In Ref. 7 , a distortion metric based on PCA transforms of the image is shown to perform well under video summarization conditions. Therefore, exploring the extension of that metric for the lossy case would also be a useful future research direction.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Yiftach Eisenberg for his useful advice in formulating this work.

REFERENCES

1. A. Hanjalic, and H. Zhang, "An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis," *IEEE Trans. on. Circuits and Systems for Video Technology*, vol. 9, December 1999.
2. Y. Wang, Z. Liu, and J-C. Huang, "Multimedia Content Analysis" *IEEE Signal Processing Magazine*, vol. 17, November 2000.
3. Y. Zhuang, Y. Rui, T.S. Huan, and S. Mehrotra, "Adaptive Key Frame Extraction Using Unsupervised Clustering," *Proc. Int'l Conf. on Image Processing (ICIP)*, Chicago, Illinois, 1998.
4. D. DeMenthon, V. Kobla and D. Doerman, "Video Summarization by Curve Simplification," *Proc. of ACM Multimedia Conference*, Bristol, UK, 1998.
5. N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias, "Video Content Representation Using Optimal Extraction of Frames and Scenes," *Proc. Int'l Conf. on Image Processing (ICIP)*, Thessaloniki, Greece, 2001.
6. H. Sundaram and S-F. Chang, "Constrained Utility Maximization for Generating Visual Skims," *IEEE Workshop on Content-Based Access of Image & Video Library*, 2001.
7. Z. Li, A.K. Katsaggelos, G. Schuster and B. Gandhi, "Rate-Distortion Optimal Video Summary Generation," *IEEE Trans. on Image Processing*, to appear.
8. Z. Li, G. Schuster, A.K. Katsaggelos and B. Gandhi, "Bit Constrained Optimal Video Summarization," *Proc. Int'l Conf. on Image Processing (ICIP)*, Singapore, 2004.
9. Y-H. Ho, W-R. Chen and C-W. Lin, "A Rate-Constrained Key Frame Extraction Scheme for Channel-Aware Video Streaming," *Proc. Int'l Conf. on Image Processing (ICIP)*, Singapore, 2004.
10. R. Zhang, S.L. Regunathan and K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience," *IEEE J. Selected Areas in Communications*, vol. 18, pp. 966-976, June 2000.
11. P.A. Chou and Z. Miao, "Rate-Distortion Optimized Streaming of Packetized Media," *IEEE Trans. on Multimedia*, submitted, 2001.
12. F. Zhai, Y. Eisenberg, T.N. Pappas, R. Berry and A.K. Katsaggelos, "Rate-Distortion Optimized Hybrid Error Control for Real-Time Packetized Video Transmission," *IEEE Trans. on Image Processing*, accepted, 2004.
13. ITU-T, *Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263, Jan. 1998, Version 2.