VIDEO SUMMARIZATION FOR MULTIPLE PATH COMMUNICATION

[1,2]Zhu Li, [3]Guido M. Schuster, and [2]Aggelos K. Katsaggelos

[1]Multimedia Research Lab (MRL), Motorola Labs, Schaumburg, Illinois, USA
[2]Department of Electrical & Computer Engineering, Northwestern University, Evanston, Illinois, USA
[3]Hochschule für Technik Rapperswil, Switzerland

ABSTRACT

For video communications over wireless ad hoc networks, multiple paths with limited bandwidth are common. It therefore presents new challenges to the video encoding. In this paper, we formulate the problem as a multiple path video summarization problem under bit rate constraints, where video summaries are generated to satisfy each channel's rate constraint, while the combined summary at the receiving end achieves the minimum summarization distortion. The optimal solution (within a convex hull approximation) is found by Lagrangian relaxation and dynamic programming. Simulation results demonstrate the effectiveness of the approach.

1. INTRODUCTION

For communicating video over a channel with limited bit rate, instead of sending all frames with severe frame SNR distortion, a better option is to transmit a subset of the frames with higher SNR quality. A video summary generator that can "optimally" select frames based on an optimality criterion is essential for these applications. Past work in this area can be found in [1]-[6], [11]-[12]. In [8] we provided a solution for the bit rate constrained summarization problem with a single channel. The resulting video summaries show that reasonably good visual qualities can be achieved at low bit rates.

The problem arises naturally when there are multiple paths with different rates for communication. One solution is to summarize the sequence and split the encoded bit stream to match individual channel rates. However due to the dependency among bit streams, when a single channel fails, the data carried by the other channels also become useless. Instead, a multiple description coding (MDC) [6], [7] like scheme is more applicable, where each channel carries an independently decodable summary that meets the channel rate constraint, while the combined summary at the receiving end achieves the minimum distortion possible.

In this paper we formulate the multipath video summarization problem as a summarization distortion minimization problem under multiple rates constraints, and find solution through Lagrangian relaxation and dynamic programming. The paper is organized into the following sections: in section 2, we give the definitions and formulation of the multi path summarization problem, in section 3, we develop the solution to the problem, in section 4, we present some simulation results, and in section 5 we draw conclusions and outline our future work.

2. DEFINITIONS AND PROBLEM FORMULATION

A video summary is a shorter version of the original video sequence. Video summary frames form a subset of the frames selected from the original video sequence. Let an n-frame video sequence be denoted by $V= \{f_0, f_1, \ldots, f_{n-1}\}$, and its video summary of m frames $S= \{f_{l_0}, f_{l_1}, \cdots f_{l_{m-1}}\}$, in which l_k denotes the k-th summary frame's location in the original sequence V.

Let the reconstructed video sequence, $V_S' = \{f_0', f_1', \cdots f_{n-1}'\}$, be generated from the video summary by substituting the missing frames with the previous frames in the summary (zero-order hold), that is,

$$f_j' = f_{i=\max(l):s.t.l\ \{l_0,l_1,\cdots l_{m-1}\}, i\le j}, \quad \forall f_j'\ \ V_S'. \tag{1}$$

Apparently the number of frame available to the summary will affect the fidelity of the reconstructed sequence. To state the trade off between the quality of the reconstructed sequences and the number of frames in the summary, we utilize the following definitions. Let the distortion between two frames j and k be denoted by $d(f_j, f_k)$, then the sequence distortion introduced by the summary is given by the average (total) frame distortion

$$D(S) = \sum_{j=0}^{n-1} d(f_j, f_j') \tag{2}$$

Also, let the summarization distortion for a segment of V_S' starting with summary frame $f_{l_{t-1}}$ and ending with frame f_{l_t} be,

$$G_{l_{t-1}, l_t} = \sum_{k=l_{t-1}}^{l_t - 1} d(f_k, f_{l_{t-1}}) \tag{3}$$

Let us assume that two paths with rates R_1 and R_2 exist for video communication, assuming constant PSNR quality coding of the video summary, and a rate profiler

0-7803-9134-9/05/$20.00 ©2005 IEEE

(for example, [14]) that will give the rate estimates, $r(f_k)$, for encoding a summary frame f_k. Then the problem of multi path summarization is given by,

$$\min_{S_1,S_2} D(S_1 \cup S_2), s.t. R(S_1) \le R_1, R(S_2) \le R_2 , \quad (4)$$

where "U" denotes set union, and $R(S)$ is the number of bits needed to code a summary S, that is,

$$R(S) = \sum_{\forall f_{l_k} \ S} r(f_{l_k}) . \quad (5)$$

Assuming that the first frame f_0 is included in both S_1 and S_2, that is $l_0^1=l_0^2=0$, formulation (4) can be also written as,

$$\min_{l_1^1,l_2^1,\cdots,l_{m1}^1,l_1^2,l_2^2,\cdots,l_{m2}^2,} D(S_1 \cup S_2), \quad s.t. R(S_1) \le R_1, R(S_2) \le R_2 . \quad (6)$$

The problem becomes the selection of summary frames for each channel to meet rate constraints, while also minimizing the combined summary distortion at the receiving end.

In addition, for each summary we may also impose an optional frame skip constraint K_{max}, that is the maximum number of frames can be skipped between any two summary frames.

2. SOLUTION TO THE MULTI-PATH SUMMARIZATION PROBLEM

The direct solution of the problem in Eq. (6) is difficult, due to the large searching space and the complicated inter-dependence among operating parameters and constraints. Instead we relax the problem with the use of Lagrangian multipliers. The relaxed problem is solved first, then followed by a search on the Lagrangian multiplier to obtain the optimal solution to the original problem within a convex hull approximation. Such an approach has been successfully applied to a number of video coding and summarization problems (for example, [8], [9] and [10]).

By introducing two Lagrangian multipliers, λ_1 and λ_2, the original constrained problem becomes,

$$L(\lambda_1, \lambda_2) = \arg \min_{l_1^1,\cdots,l_{m1-1}^1,l_1^2,\cdots,l_{m2}^2,} D(S_1 \cup S_2) + \lambda_1 R(S_1) + \lambda_2 R(S_2) \quad (7)$$

where λ_1, λ_2 are positive real numbers that control the trade off between rates and the combined summarization distortion.

We observe that the minimization in Eq. (7) has a interesting recursive strucutre. Let two summaries S_1 and S_2 all start from frame f_0, and stop when the final virtual frame, f_n, is reached. Let us define for a given multiplier pair, λ_1 and λ_2, the state $L_{j,k}$ be given by,

$$L_{j,k} = \min_{l_1^1,\cdots,l_t^1=j,l_1^2,\cdots,l_h^2=k} D(S_1 \cup S_2) + \lambda_1 R(S_1) + \lambda_2 R(S_2) \quad (8)$$

where t and h are the number of frames in the summaries S_1 and S_2 so far. Frames f_j and f_k are the last frames for S_1 and S_2 respectively. There are many feasible transitions from other states into the current state $L_{j,k}$, corresponding to adding frame f_j to S_1 and f_k to S_2, that is,

$$L_{j,k} = \min_{1 \le p \le K_{\max}, 1 \le q \le K_{\max},} L_{j-p,k-q} + A(j-p, k-q, j, k) , \quad (9)$$

where $(j\text{-}p)$ and $(k\text{-}q)$ are the last frames for the feasible states that can transition into $L_{j,k}$. Skip constraint K_{max} is the maximum number of frames that can be skipped between any two summary frames, which enforces certain degree of smoothness within each individual summary. It does not affect the generality of the approach, since it is optional. The edge cost, $A(j\text{-}p,\ k\text{-}q,\ j,\ k)$, for transition from state $L_{j-p,\ k-q}$ into $L_{j,k}$, must account for the rates increase and combined summarization reduction, and is given by,

$$\begin{aligned} A(j-p, k-q, j, k) &= \\ &= (G_{i1,i2} + G_{i2,n}) - (G_{r1,r2} + G_{r2,r3} + G_{r3,r4} + G_{r4,n}) \\ &\quad + \lambda_1 r(f_j) + \lambda_2 r(f_k) \\ &= (G_{i1,i2} + G_{i2,n}) - (G_{r1,r2} + G_{r2,r3} + G_{r3,r4} + G_{r4,n}) \\ &\quad + \begin{Bmatrix} \lambda_1 r_j + \lambda_2 r_k, & \text{if intra - coding} \\ \lambda_1 r_{j,j-p} + \lambda_2 r_{k,k-q}, & \text{if inter - coding} \end{Bmatrix} \end{aligned} \quad (10)$$

in which $[r1\ r2\ r3\ r4]$=sort($[j\text{-}p, k\text{-}q, j, k]$), $[i1\ i2]$=sort($[j$, k]). r_j is the intra-rate estimate given by the rate profiler, and $r_{j,k}$ is the inter-rate estimate of coding frame j with prediction from frame k. The initial state is given by,

$$L_{0.0} = G_{0,n} + \lambda_1 r_0 + \lambda_2 r_0 \quad (11)$$

With the recursion given by Eqs. (10) and (11), we can compute a trellis for the given problem, and once the trellis ends at the virtual final node f_n, backtrack for the optimal solution [15].

An example of a two-path summarization trellis and the optimal solution path are illustrated in Fig. 1. The sequence is a 12-frame segment from the "foreman" sequence. The trellis starts at the lower left corner and ends at the upper right corner. All possible two path summaries can be represented as a trellis path starting at the lower left corner (f_0) and ending at the virtual final frame (f_n) at the upper right corner.

The horizontal and vertical axes indicate the summary frames for S_1 and S_2, respectively. The dotted (green) lines are the minimum cost incoming arcs to each node, and the solid (red) line with circles is the optimal path obtained by back tracking, corresponding to the given Lagrangian multiplier values of λ_1=4.17e-05 and λ_2=8.33e-06.

The solution to the original problem can be found by searching for the Lagrangian pair λ_1^* and λ_2^*, such that the resulting summaries have the tightest bound to the rates constraints, $R(S_1(\lambda_1^*, \lambda_2^*)) \le R_1$, and $R(S_2(\lambda_1^*, \lambda_2^*)) \le R_2$. This can be achieved through a variety of numerical methods, like cutting plane [16] and coordinate descent.

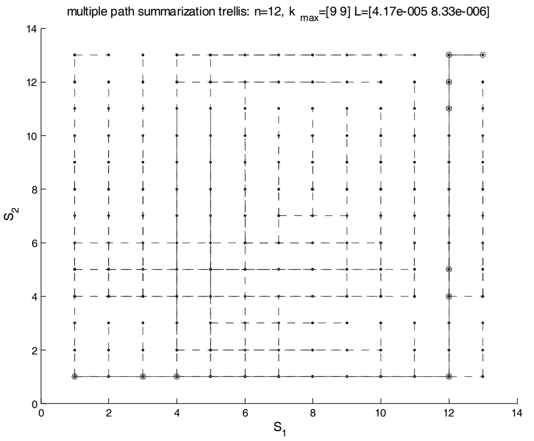


Figure 1. Relaxed Problem DP solution example

3. SIMULATION RESULTS

In our simulation, we use a frame distortion metric for summarization distortion based on the distance between scaled frames in the principal component space [13], that is,

$$d(f_j, f_k) = \| x_j - x_k \|, \quad (12)$$

where x_j and x_k are the feature points for frames f_j and f_k obtained by first scaling the frame luminance field to a resolution of 8x6, and then applying a PCA. This has been found to be an effective and computationally simple metric [13].

In an example shown in Fig. 2, we use a 120-frame segment from the “foreman” sequence (frames 170~289) for the simulation, with the frame skip constraint set at K_{max}=30.

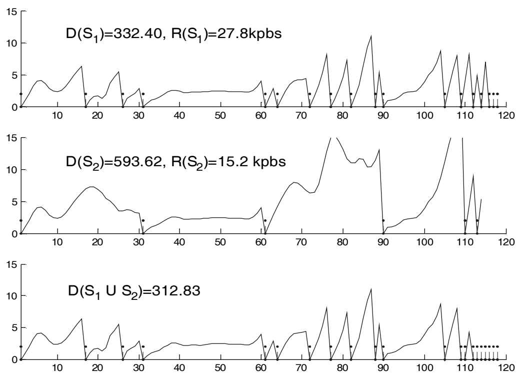
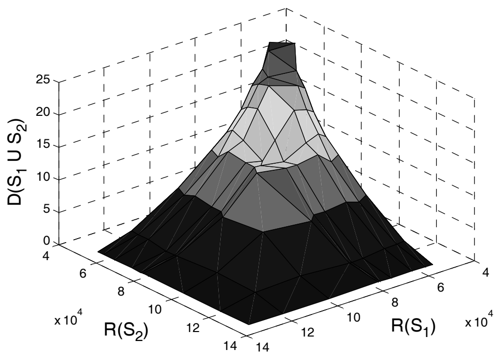


Figure 2. Two-path summarization example

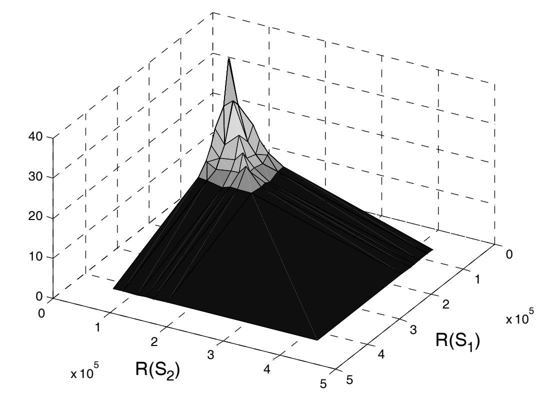
For the relaxed problem with λ_1=0.16e-4 and λ_2=0.5e-4, the solution is illustrated in Fig. 2. The vertical dotted lines indicate frame selections for the summaries, while the curves in solid line represent the frame distortion between the original and the reconstructed sequence from the summary. Notice that the frame distortion is equal to zero at the location where a frame is selected into the summary. The upper and middle plots are for the summaries S_1 and S_2 respectively. The combined summarization distortion is equal to 312.83, while the summary distortions for two individual paths are equal to $D(S_1)$=332.40, and $D(S_2)$=593.62. Notice that a larger Lagrangian value places more emphasis on minimizing the rates, that is why smaller number of frames is selected for S_2 than for S_1, since $\lambda_1 = \lambda_2/3$. The resulting inter-coding rates for S_1 and S_2 are 27.8 *kpbs* and 15.2 *kpbs* respectively.

The video summaries S_1, S_2 and S in Fig. 2 are also available for subjective evaluation at: *http://ivpl.ece.northwestern.edu/~zli/new_home/demo/multipath/multipath.html*.

The overall multiple rates-distortion performance is illustrated in Fig. 3. For a 40-frame segment of the “foreman” sequence, with frame skip constraint K_{max}=16, the distortion function in terms of the Lagrangian multipliers, $D(R(S_1), R(S_2))$, is plotted in Figs. 3a and 3b for the inter-coding and the intra coding cases, respectively.



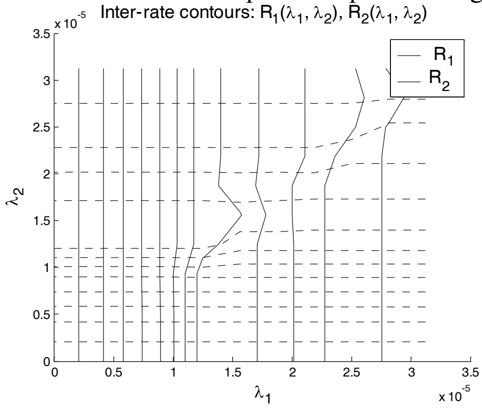
(a) inter-coding case



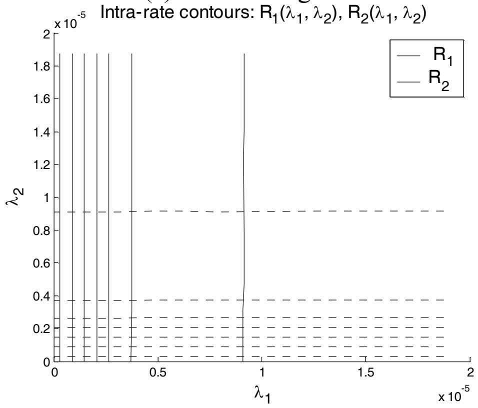
(b) intra-coding case

Figure 3. Distortion as a function of $\lambda_1, \lambda_2 : D(\lambda_1, \lambda_2)$

Notice that the skip constraint enforces some level of smoothness for each individual summary, which is a desirable feature when one channel fails, but on the other hand it reduces the combined summary efficiency. This is an inherent dilemma in multiple description coding.



(a) Inter-coding case



(b) Intra-coding case

Figure 4. Summary rates functions $R(S_1)$ and $R(S_2)$.

The summary rates functions contours for $R(S_1(\lambda_1, \lambda_2))$ and $R(S_2(\lambda_1, \lambda_2))$ for the same sequence are plotted in Figs 4a and 4b for the inter-coding and the intra-coding cases, respectively. For the inter-coding case when $\lambda_1 \quad \lambda_2$, a cutting plane like method needs to be applied, while for the rest, a coordinate descent search will be able to find the summaries that meet the rate constraints.

4. CONCLUSION AND FUTURE WORK

In this paper we developed an optimal solution to the two-path summarization problem. According to the proposed algorithm, Optimal (within a convex hull approximation) combined video summary is delivered through two channels with different rate constraints. The technique can be generalized to handle more paths, and is useful in multi-path video streaming for consumer and security applications.

We are also working on a more efficient searching algorithm for matching the rates constraints, and investigating the scenarios where the maximum frame distortion is used as the summarization distortion. We are also extending the formulation to handle quality constraints on each individual path.

7. REFERENCES

[1] D. DeMenthon, V. Kobla and D. Doermann, "Video Summarization by Curve Simplification", *Proceedings of ACM Multimedia Conference*, Bristol, U.K., 1998

[2] N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias, "Video Content Representation Using Optimal Extraction of Frames and Scenes", *Proc. of Int'l Conference on Image Processing*, Chicago, Illinois, 1998.

[3] A. Girgenshohn and J. Boreczky, "Time-Constrained Key frame Selection Technique", *Proc. of IEEE Multimedia Computing and Systems (ICMCS),* 1999.

[4] Y. Gong and X. Liu, "Video Summarization with Minimal Visual Content Redundancies", *Proc. of Int'l Conference on Image Processing*, 2001.

[5] 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.

[6] V. K. Goyal, "Multiple description coding: compression meets the network", *IEEE Signal Processing Magazine*, September, 2001.

[7] A. R. Reibman, H. Jafarhkani, Y. Wang, M.T. Orchard and R. Puri, "Multiple description coding for video using motion compensated prediction", *Proceedings of Int'l Conference on Image Processing* (ICIP), Kobe, Japan, 1999.

[8] Z. Li, G. Schuster, A. Katsaggelos and B. Gandhi, "Optimal video summarization with a bit budget constraint", *Proceedings of Int'l Conference on Image Processing (ICIP)*, Singapore, 2004.

[9] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video Frame Compression and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.

[10] Y. Shoham and A. Gesho, "Efficient bit allocation for an arbitrary set of quantizers", *IEEE Trans. on Acoustics, Speech, Signal Processing*, vol. 36, pp. 1445-1453, September, 1988.

[11] H. Sundaram and S-F. Chang, "Constrained Utility Maximization for Generating Visual Skims", *IEEE Workshop on Content-Based Access of Image & Video Library*, 2001.

[12] Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, "Adaptive Key Frame Extracting Using Unsupervised Clustering", *Proc. of Int'l Conference on Image Processing*, Chicago, Illinois, 1998.

[13] Z. Li, A. Katsaggelos, G. Schuster, and B. Gandhi, "Rate-Distortion Optimal Video Summary Generation", to appear, *IEEE Trans. on Image Processing*. .

[14] Zhihai He and Sanjit K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Trans. on Circuits and System for Video Technology*, vol. 11, pp. 1221 – 1236, December 2001

[15] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Trans. on Information Theory*, vol. IT-13, pp. 260-269, April 1967.

[16] S. Boyd and L. Vandenberghe, "Localization and Cutting-plane methods", Lecture notes, Stanford University, www.stanford.edu/class/ee392o/localization-methods.pdf