OPTIMAL INTRA-CODED VIDEO SUMMARIZATION FOR TWO BANDWIDTH LIMITED CHANNELS

[1]*Reto Ansorge,* [1]*Erwin Kirchmeier,* [2,3]*Zhu Li,* [1]*Guido M. Schuster, and* [2]*Aggelos K. Katsaggelos*

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

ABSTRACT

For video communication over a wireless ad hoc network, multiple channels with different rates are common. This paper considers the transmission of intra-coded video sequences through two bandwidth limited channels. More specifically, with the proposed framework a rate-distortion optimal summary of the video sequence is generated first, utilizing the cumulative bit rate of the two channels. In a second step the frames of the summary are distributed between the two channels according to the individual channel rates. Toward this task, a greedy suboptimal distribution algorithm and an optimal distribution algorithm, based on the knapsack theory, are shown. Experimental results are presented.

1. INTRODUCTION

For video communication over a wireless ad hoc network, multiple channels with different rates are common. Two or more channels can be put together to form a logical channel with higher bandwidth [1], but there are two problems that have to be considered. First, the bandwidth of the logical channel might be smaller than the bandwidth of a given video sequence and second, the channels can drop out.

A solution for the first problem is to further compress the video sequence. The second problem can be solved by using, for example, multiple description coding [2].

An alternative approach in considering the above described scenario is followed in this work. We assume that two channels are available and the video to be transmitted is intra-coded. We then generate a rate-distortion optimal summary[3, 4] of the video sequence utilizing the total rate of the video sequence of the logical channel. The distortion introduced this way is temporal (omission of frames in the summary), while the spatial quality of the frames utilized in the summary is preserved. The resulting frames belonging to the summaries are then distributed into the two available channels, so that the rate constraints imposed by each of them are satisfied.

In the following the problem to be solved is formulated in Sec. 2. In Sec. 3 we relax the problem and propose a framework to solve it. In Sec. 4 a greedy and an exact algorithm are presented to solve the relaxed problem. Experimental results are described in Sec. 5 and conclusions in Sec. 6.

2. PROBLEM FORMULATION

Let a video sequence of n frames be denoted by $V = \{f_0, f_1, \ldots, f_{n-1}\}$, and its summary of m frames $S = \{f_{l_0}, f_{l_1}, \ldots, f_{l_{m-1}}\}$, in which l_k denotes the k-th summary frame's location in the original sequence V. The reconstructed sequence $V'_S = \{f'_0, f'_1, \ldots, f'_{n-1}\}$ from the summary S is obtained by substituting missing frames with the most recent frame that belongs to the summary S, that is,

$$f'_j = f_{i=\max(l):s.t.\, l\in\{l_0,l_1,\ldots,l_{m-1}\}, i\le j} \;\forall f'_j \in V'_S. \tag{1}$$

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 defined in this paper as the maximum frame distortion, that is,

$$D(S) = \max_{k\in[0,n-1]} d(f_k, f'_k). \tag{2}$$

We use the maximum distortion, since minimizing the maximum distortion results in a best worst case, which in turn results in similar frame distortions [5]. We assume the existence of two communication channels with rates R_1 and R_2. We also assume that each frame f_k in V is intra coded and has a rate r_k. Then the problem becomes the generation of two summaries S_1 and S_2, so that:

$$\min_{S_1,S_2} D\left(S_1 \cup S_2\right), s.t. R(S_1) \le R_1, R(S_2) \le R_2$$
$$S_1 \cap S_2 = \{f_0\}, \tag{3}$$

where $R(S_i)$ is the total number of bits of the frames in summary S_i. Note that frame f_0 is included in both summaries S_1 and S_2.

3. RELAXED PROBLEM

The problem, as formulated above, is very hard to solve. For long video sequences an exhaustive search would require prohibitive time. A frame can either be transmitted in the first channel, in the second channel or not transmitted at all, except for the first frame which is always transmitted in both channels. Therefore there are 3^{n-1} possibilities to distribute the frames.

In order to relax the problem, we build an optimal summary for the logical channel first, with total rate $R_1 + R_2 - r_0$. The following problem is solved:

$$\min_{S} D(S), s.t. R(S) \leq R_1 + R_2 - r_0, \tag{4}$$

where r_0 is the rate for the first frame. In solving (4) we can also enforce a skip constraint, which determines the maximum number of successive frames that be skipped in the summary. In a second step the resulting summary will be split into the two channels if possible, so that

$$R(S_1) \leq R_1, R(S_2) \leq R_2, s.t. S_1 \cup S_2 = S$$
$$S_1 \cap S_2 = f_0. \tag{5}$$

If the resulting summary S cannot be split into the two channels so that (5) holds then the first step has to be recalculated with a lower rate constraint. This leads to

$$\min_{S} D(S), s.t. R(S) \leq R_1 + R_2 - r_0 - \varepsilon, \tag{6}$$

where ε is varied (increased), until a solution that can be split is found. The summarization algorithm (4) which generates the optimal summary is described in [3]. The two splitting algorithms that we propose, will be presented in Sec. 4.

3.1. Limitations

Even if the summarization and splitting are optimal, it cannot be guaranteed that the resulting solution is an optimal solution to the unrelaxed problem. The problem is that the summarization algorithm will only find solutions, that are on the operational rate-distortion (ORD) staircase [3] of the corresponding one channel problem. The optimal solution for the unrelaxed problem however, does not have to be on this ORD-staircase. Assume the following scenario, depicted in Fig. 1

- $R_1 + R_2 - r_0$ is the initial upper bound of the rate $R(S)$
- The operating points indicated by a triangle (∇) cannot be split so that (5) holds
- The operating points indicated by a circle ($\bigcirc$) can be split so that (5) holds

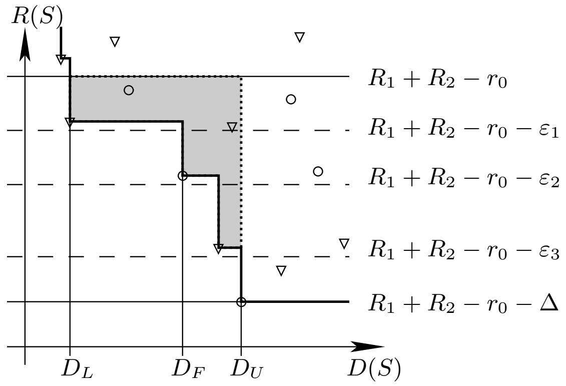

Fig. 1. Problem with splitting the optimal one channel summary.

The summarization algorithm finds the first operating point (OP) below $R_1 + R_2 - r_0$ on the ORD-staircase which has distortion D_L. If this OP cannot be split by the chosen splitting algorithm, the upper bound has to be lowered by ε, as shown in Fig. 1. The summarization has to be recalculated until ε_3, where an OP on the ORD-staircase is found, which can be split.

The first OP on the ORD-staircase, which can be split has distortion D_F. The first OP on the ORD-staircase, which can be split by the chosen splitting algorithm, has distortion D_U. In the shaded region between D_L and D_U could be other OPs that can be split. The OP in the shaded region, which has the smallest $D(S)$ is the optimal OP for the unrelaxed problem. If D_L and D_U are close to each other, the probability that there is an OP in the shaded region, that can be split, is small. If $D_U = D_L$ then the optimal solution for the relaxed and unrelaxed problems has been found. If $D_U = D_F$ then the optimal solution for the relaxed problem has been found.

4. FRAME SPLITTING ALGORITHMS

In this section we will present two splitting algorithms. First, a greedy suboptimal algorithm and then an exact splitting algorithm are shown. Using the exact splitting algorithm D_U is always equal to D_F whereas this can not be guaranteed for the greedy algorithm.

4.1. Greedy Algorithm

Given a summary S this algorithm distributes the frames between the two channels in a straightforward manner. A summary is well distributed when

$$\frac{R(S_1)}{R(S_2)} = \frac{R_1}{R_2}, \tag{7}$$

which is the same as

$$R_2 R(S_1) - R_1 R(S_2) = 0. \quad (8)$$

The algorithm first puts frame f_0 into both channels. Every other frame f_{l_k} in S is put either in S_1 or S_2, such that

$$|R_2 R(S_1^k) - R_1 R(S_2^k)| \quad (9)$$

is minimized, where S_1^k and S_2^k are the summaries for each channel up to frame f_{l_k}. This achieves a distribution of the frames between the two channels, which will have low $D(S_1)$ and $D(S_2)$ for most natural video sequences.

4.2. Exact Algorithm

The splitting problem has an interesting property that can be exploited. Because all frames have to be assigned to one of the two channels, it is sufficient to find an optimal splitting for one of the two channels. For a logical channel $R_1+R_2-r_0$ and given summary S we can calculate the remaining free space Δ (See Fig. 1), which is given by

$$\Delta = (R_1 + R_2 - r_0) - R(S). \quad (10)$$

A feasible splitting solution has been found when

$$R_1 - \Delta \le R(S_1) \le R_1. \quad (11)$$

If (11) holds, then $R(S_2)$ will also be smaller or equal to R_2. Replacing the left part of (11) with a maximum function and using the right part as its constraint we get

$$\max\{R(S_1)\},\ s.t.\ R(S_1) \le R_1, \quad (12)$$

which can be regarded as the following subset-sum knapsack problem,

$$max\left(\sum_{k=1}^{m} r_{l_k} x_k\right),\ s.t.\ \sum_{k=1}^{m} r_{l_k} x_k \le R_1,$$
$$x_0 = 1,$$
$$x_j = 1\ or\ 0 \quad \forall j > 0. \quad (13)$$

If $x_j = 1$ then frame f_{l_k} is in S_1. There exist many fast algorithms to solve this problem [6][7] but it is still computational intensive. We are using Pisingers 'decomp' algorithm [6], for which the C code is available on it website. The code was adapted so that it stops when it has found a solution that can be split as denoted in (11).

The strategy of the 'decomp' algorithm is to put first as many consecutive frames as possible into summary S_1. If the rate constraint (11) holds, the algorithm terminated. If the rate constraint does not hold, the algorithm starts to exchange frames at the end of summary S_1. If the algorithm finds a solution, the exchanging stops. All frames that are not assigned to S_1 will be assigned to S_2. A possible resulting splitting is shown in Fig. 2 (a). The size of the exchanged region depends on how fast a solution has been found.

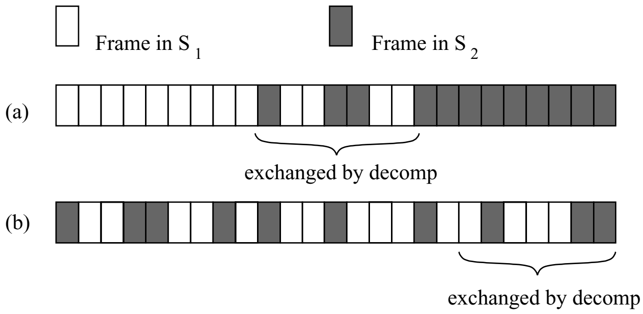

Fig. 2. Splitting of S without (a) and with reordering (b)

As one can see in Fig. 1 (a) the strategy of 'decomp' distributes the frames of summary S locally in an unequal fashion between S_1 and S_2. This can introduce a rather high distortion $D(S_1)$ and $D(S_2)$. With an appropriate reordering this problem can be avoided somewhat. For the reordering we use the greedy split algorithm without checking the rate constraints. With the output S_1' and S_2' of the greedy algorithm a reordered summary $\hat{S}$ is formed, according to

$$\hat{S} = S_1' \cup (\tilde{S_2} \backslash \{f_0\}), \quad (14)$$

where $\tilde{S_2}$ is S_2' in reverse order. The union $\cup$ is used so that the order of the elements is preserved. If 'decomp' is used on $\hat{S}$ to get S_1 and S_2, the exchanging is now performed at the end of both summaries. A resulting splitting is shown in Fig. 2 (b).

5. RESULTS

As distortion metric we use the Euclidian distance in the Principal Component space of the frames [8] as described in [3]. As video coder we used the TMN5 implementation of H.263. As test sequence we used frames 150–299 of the 400 frame long foreman sequence.

Calculations for different splitting ratios (x-axis) and different logical channels (y-axis) have been made. All together there have been considered 250×10000 channel combinations . In all tables and figures 'Greedy' and 'Exact' denotes the greedy splitting algorithm and the exact splitting algorithm, respectively. $R_1 + R_2 - r_0$ is the rate of the logical channel, while $R(V)$ is the rate of the video sequence that should be transmitted.

Fig. 3 shows the difference between D_L and D_U. If $D_L = D_U$ the optimal solution for the unrelaxed problem (3) has been found (white area). If 'Exact' is used the optimal solution is found in the overwhelming cases.

$R_1 : R_2$	Split	$D(S)$	$D(S_1)$	$D(S_2)$	D_U–D_L
1 : 1	Greedy	2.45	**8.01**	**8.14**	0.12
	Exact	**2.33**	29.8	20.6	0
1 : 10	Greedy	2.38	51.3	8.00	0.05
	Exact	**2.33**	68.4	5.27	0

Table 1. Half-rate summarization

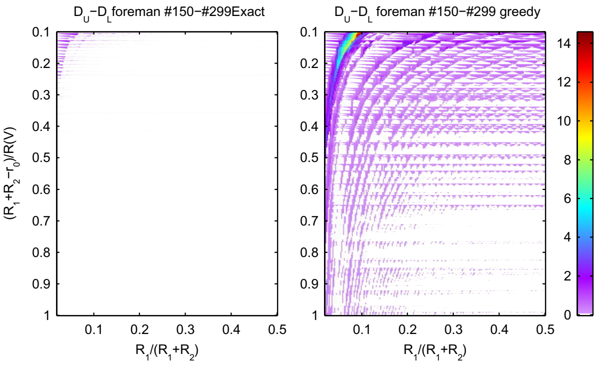

Fig. 3. Check if the optimal solution is maybe missed

If 'Greedy' is used the optimal solution will not be found so often. That means that ε has to be increased several times. Every time ε is increased a new summary S has to be calculated which needs some time.

Fig. 4 shows the distortion of summary S_1. If 'Exact' is used, the distortion can be rather high. In contrast the results are much better if 'Greedy' is used. This is because 'Greedy' achieves a more natural distribution of the frames between the channels.

Table 1 shows the results for two channel ratios, and a logical channel size of $\frac{1}{2}R(V)$. As it can be seen in it the individual channel distortion when 'Greedy' is used is considerably smaller when compared to 'Exact', as expected.

6. CONCLUSION

The proposed framework first computes a rate-distortion optimal summary for the logical channel. In a second step this summary is split into the two channels. So it is possible to extend the optimal single channel summarization theory to two channels. The well implemented Subset-sum algorithm from Pisinger [6] was used for the 'Exact' splitting algorithm. This splitting, however, causes rather high and unbalanced $D(S_1)$ and $D(S_2)$.

By using 'Greedy' for splitting, $D(S_1)$ and $D(S_2)$ are much lower compared to the case when 'Exact' is used. On the other hand one has to pay for the lower distortions in specific channels with a higher $D(S)$. But the price is not too high and can be accepted for most practical applications.

A next step is to extend the proposed framework to more

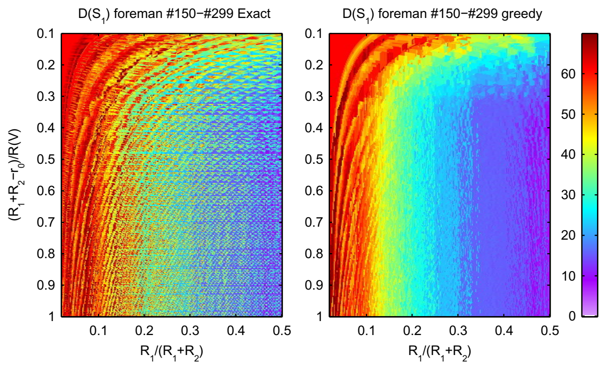

Fig. 4. Distortion in channel 1

than two channels. This means that the summary S for the logical channel has to be split into multiple channels.

7. REFERENCES

[1] P. H. Fredette "The past, present, and future of inverse multiplexing", in *IEEE Comm, Magazine*, vol. 32, no. 4, pp.42–46, April 1994.

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

[3] Z. Li, A. K. Katsaggelos and G. M. Schuster "Rate-Distortion based optimal video summary", in *Intelligent Multimedia Processing with Soft Computing*, Springer-Verlag, Heidelberg, 2004.

[4] Z. Li, G. M. Schuster, A. K. Katsaggelos and B. Gandhi "MINMAX Optimal Video Summarization", in *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services*, 2004.

[5] G. M. Schuster, G. Melnikov and A. K. Katsaggelos. A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers In *IEEE Transactions on Multimedia*, Vol. 1, No.1, pp. 3 - 17, March 1999.

[6] D. Pisinger, "An exact algorithm for large multiple knapsack problems", in *European Journal of Operational Research*, vol. 114, pp. 528–541, 1999.

[7] N.Y. Soma and P. Toth, "An exact algorithm for the Subset–sum problem", in *European Journal of Operational Research*, vol. 136, pp. 57–66, 2002.

[8] H. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Chapter 14, Springer Series in Statistics, New York, 2001.