PERCEPTUAL VIDEO QUALITY OPTIMIZATION IN AWGN CHANNEL USING LOW COMPLEXITY CHANNEL CODE RATE ALLOCATION

Ting-Lan Lin and Pamela C. Cosman

University of California, San Diego – ECE

Review Topics : H-7 = Speech, Image and Video Processing (Image and Video Coding)

ABSTRACT

In error-prone channels, forward error correction is necessary for protecting important data. In this paper, we use a packet loss visibility model to evaluate the visual importance of video packets to be transmitted. With the loss visibility of each packet, we use the Branch and Bound method to optimally allocate rates of Rate-Compatible Punctured Convolutional codes. The complexity of our prior algorithm can be reduced by k-means clustering before using the Branch and Bound method. Experimental results show that the proposed unequal error protection algorithm can improve upon the received video quality compared to our prior work with much lower complexity.

Index Terms— Unequal error protection, packet loss visibility model, perceptual quality.

1. INTRODUCTION

Packets of compressed video induce different levels of quality degradation when lost. Therefore they should be protected by channel codes based on their importance. FMO (Flexible Macroblock Ordering) in H.264 is employed in [1] to group macroblocks of similar estimated distortion in a frame into a slice, with different levels of channel protection over slices. This method is extended in [2] with Converged Motion Estimation, which performs motion estimation for the current frame using mostly the highly-protected MBs in the previous frame as reference. In [3], frames closer to the end of a GOP have less error protection. Traditionally and in the methods mentioned above, video quality degradation is measured with MSE (mean-squared error). However, MSE is not correlated well with human perception [4]. Therefore, a metric developed/verified by subjective experiments is necessary for evaluating the video quality.

We evaluate the visual importance of each packet by the packet loss visibility model developed by extensive subjective experiments from [5, 6]; the model estimates for each packet the probability that the end user will observe the packet loss artifact if the packet is lost. Based on this metric, we aim to optimally allocate Rate-Compatible Punctured Convolutional (RCPC) codes to minimize the visual quality degradation when transmitting video over an AWGN channel, given a total rate budget. We solve the optimization problem by the Branch and Bound method (BnB) in [7]. However, BnB has worst case exponential complexity [8]. In this paper, we reduce the complexity of our algorithm presented in [7] by preprocessing the packet information using k-means clustering. We evaluate the decoded lossy video using Video Quality Metric (VQM) [9, 10], a full reference (FR) metric developed by the National Telecommunication and Information Administration, shown to be well correlated with human perception compared to other FR video quality metrics [11]. The proposed algorithm has much lower complexity, and performs slightly better than the prior algorithm.

The organization of this paper is as follows. In Section 2, we explain the packet loss visibility model and its predictors. Section 3 formulates the RCPC rate allocation problem as an integer programming problem. In Section 4, the branch and bound method is introduced to solve the optimization problem. Section 5 describes how we apply K-means algorithm to reduce the complexity of our prior algorithm. Experimental results and conclusions are shown in Section 6.

2. PACKET LOSS VISIBILITY MODEL

In this paper, we evaluate the visual importance of each packet by our packet loss visibility model [5, 6]. Packet loss visibility is the probability that the end user will observe the packet loss artifact if the packet is lost. Therefore, higher loss visibility means a packet is more visually important. This model was built from three subjective experiments with various codecs (H.264 and MPEG-2), encoding rates, GOP structures and error concealment methods. The packet loss visibility is modeled using a Generalized Linear Model, whose link function is a logit function [6]. In the following, we describe the factors used in the model.

This work was supported in part by the National Science Foundation, the Center for Wireless Communications at UCSD, and the UC Discovery Grant program.

IMSE, **ISSIM** (the average MSE and SSIM [12] among all MBs in an initial packet loss) and **MaxIMSE** (the maximum per-MB MSE over all MBs in the initial packet loss) are significant to the packet loss visibility. A forward motion estimation using 16x16 motion blocks from the *uncompressed* signal $f(t)$ was used to measure the motion information (x,y) per MB that is independent of any codec. (MOTX,MOTY) is the average motion vector, and **ResidEng** is the average residual energy after motion compensation, over MBs in a packet. The boolean variable **HighMOT** is TRUE if $\sqrt{MOTX^2 + MOTY^2} > \sqrt{2}$.

We also consider Reference-Scene-related factors. A method for detection of quick scene cuts was presented in [13]. Each packet loss was labeled by the distance in time between the first frame affected by the packet loss and the nearest scene cut, either before or after. This is DistFromSceneCut, and is positive if the packet loss happens after the closest scene cut in display order, and negative otherwise. DistToRef per MB describes the distance between the current frame (with the packet loss) and the reference frame used for concealment. This variable is positive if the frame with the packet loss uses a previous (in display order) frame as reference, and negative otherwise. **FarConceal** is TRUE if MaxDistToRef (maximum of |DistToRef| in a slice) ≥ 3. In this inequality, MaxDistToRef has units of frames. The Boolean variable, **OtherSceneConceal**, is TRUE if |DistFromSceneCut| $<$ |MaxDistToRef|, where the compared variables must be of the same sign (same direction). In this inequality, the compared variables have units of seconds. If the compared variables have different signs, OtherSceneConceal is FALSE. OtherSceneConceal describes whether the packet loss will be concealed by an out-of-scene reference frame which will increase the visibility of packet loss. To account for the depressed visibility immediately *before* a scene cut, **BeforeSceneCut** is TRUE if $-0.4sec < \text{DistFromSceneCut} < 0sec$. Scenes are classified based on four camera-motion types: still, panning, zooming, or complex camera motions. Since significantly fewer viewers see packet loss in still scenes than in panning or zooming scenes [5], the Boolean variable **NotStill** is TRUE if motion type is not still. If a packet loss induce two consecutive slices lost, **SXTNT2** is TRUE. **SXTNTFrame** is TRUE when all slices in the frame are lost. **Error1Frame** is TRUE if the packet loss lasts only one frame. The factors and coefficients of our final model are summarized in Table 1, taken from [6].

Factors	Coeff. for Final Model
Intercept	4.18061
$\log(1 - \text{ISSIM} + 10^{-7})$	0.22871
SXTNT2	-0.41208
SXTNTFrame	-1.47672
Error1Frame	-0.33009
$\log(\text{MaxIMSE} + 10^{-7})$	0.27578
$\log(\text{ResidEng} + 10^{-7})$	-0.61219
HighMOT	0.18290
NotStill	0.73364
BeforeSceneCut	-1.14434
OtherSceneConceal	2.08966
$\log(\text{IMSE} + 10^{-7})$	0.30492
$\log(\text{IMSE} + 10^{-7})$: FarConceal	0.25720

Table 1. Factors of the final model. Note that the colon (:) means "interact with"

SNR	-2dB	-1dB	0dB	1dB	2dB
$a(SNR)$	-1.59	-2.15	-2.59	-3.11	-3.43
$b(SNR)$	1.82	2.35	2.46	2.5	2.01

Table 2. The coefficients of the linear model for BER in Equation (3) for different AWGN SNR.

3. RCPC RATE ALLOCATION

Packets transmitted through the wireless network are subjected to bit errors. Therefore, to minimize the video quality degradation from the transmission over the error-prone channel, we aim to optimally allocate RCPC codes, for a given total output bandwidth, to video packets. The basic idea is to assign lower RCPC rates (more protective) to packets with higher packet loss visibility (more visually important). We formulate the problem as follows: assume we have N packets in an optimization group (e.g., $N = 450$ is the number of packets in a GOP), with size S_i and packet loss visibility V_i, $i = 1, 2, ..., N$. We seek the optimal RCPC rate selection r_i for the ith packet from the RCPC candidate set $\{R_1, R_2, ..., R_K\}$, so as to minimize the end-to-end expected packet loss visibility, while the outgoing total rate budget is constrained to be B. The problem can be formulated as:

$$\begin{aligned} \min_{\mathbf{r}} \ & \tfrac{1}{N} \textstyle\sum_{i=1}^{N} V_i \times PacketErrorRate(SNR, S_i, r_i) \\ & s.t. \ \textstyle\sum_{i=1}^{N} S_i \times \tfrac{1}{r_i} \leq B \\ & r_i \in \{R_1, R_2, ..., R_K\} \\ & i = 1, 2, ..., N \end{aligned} \tag{1}$$

where $\mathbf{r} = [r_1, r_2, ..., r_N]$. Here we define a packet to be in error (undecodable by the source decoder) when any of the bits in the packet is incorrect. Therefore,

$$\begin{aligned} & PacketErrorRate(SNR, S_i, r_i) \\ & = 1 - (1 - BER(SNR, r_i))^{S_i} \end{aligned} \tag{2}$$

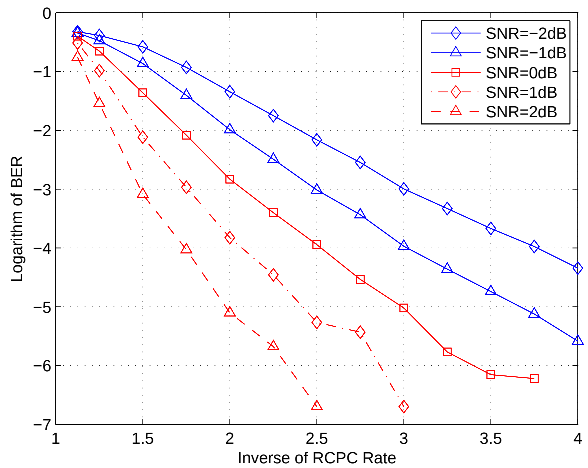

Fig. 1. Logarithm of BER versus the inverse of RCPC rates for different SNRs of an AWGN channel.

Also, from [14] and Figure 1, the logarithm of BER (bit error rate after channel decoding) is approximately linearly related to the inverse of the RCPC rate for a given SNR in an AWGN channel. Thus we approximate

$$\begin{aligned} BER &= 10^{a(SNR)\frac{1}{r}+b(SNR)} \\ r &\in \{R_1, R_2, ..., R_K\} \end{aligned} \tag{3}$$

where the regression coefficients a and b are functions of SNR, as listed in Table 2. From here we fix SNR for discussion simplicity. Substituting the expression of packet error rate and BER into problem (1), we have the following:

$$\begin{aligned} \min_{\mathbf{r}} \quad & \tfrac{1}{N}\textstyle\sum_{i=1}^{N} V_i \times \{1-(1-10^{a\frac{1}{r_i}+b})^{S_i}\} \\ s.t. \quad & \textstyle\sum_{i=1}^{N} S_i \times \tfrac{1}{r_i} \leq B \\ & r_i \in \{R_1, R_2, ..., R_K\} \\ & i = 1, 2, ..., N \end{aligned} \tag{4}$$

This problem can be recognized as a nonlinear integer programming problem, and can be solved by a well-developed method called branch-and-bound (BnB). We transform the original discrete optimization variables into binary boolean variables [15] to use the BnB algorithm. The optimization variables are transformed into $x_{ij} \in \{0, 1\}$ defined as

$$x_{ij} = \begin{cases} 1 & \text{if packet } i \text{ uses rate } j \\ 0 & \text{otherwise} \end{cases}$$

Since one packet can only use one rate, we have the following linear equality constraint:

$$\sum_{j=1}^{K} x_{ij} = 1, \forall i = 1, 2, ..., N$$

Therefore, our problem in (4) can be written as:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{K} x_{ij} \times V_i \times \{1-(1-10^{a\frac{1}{R_j}+b})^{S_i}\} \\ s.t. \quad & \textstyle\sum_{i=1}^{N}\sum_{j=1}^{K} x_{ij} \times S_i \times \tfrac{1}{R_j} \leq B \\ & \textstyle\sum_{j=1}^{K} x_{ij} = 1, \forall i = 1, 2, ..., N \\ & x_{ij} \in \{0, 1\} \\ & i = 1, 2, ..., N, j = 1, 2, ..., K \end{aligned} \tag{5}$$

where $\mathbf{X} = \{x_{ij}\}_{j=1,...,K}^{i=1,...,N}$. The problem can now be solved by BnB as described next.

4. BRANCH AND BOUND

BnB partitions the original problem into smaller subsets by tree-growing, and eliminates further consideration of the feasible solutions that can not be better than the current one [15]. The algorithm solves the binary-variable problem as follows :

1. The original problem is solved with the integer constraint relaxed to allow $0 \leq x_{ij} \leq 1$. A lower bound (LOWER) of the optimized value to the original problem is the optimized value of the relaxed problem since it has a larger feasible set. An upper bound (UPPER) of the optimized value to the original problem can be obtained by substituting the rounded solutions (to zero or one so that they are feasible) into the problem. This is an upper bound of the optimal value to the original problem since it is the best feasible value we can find at this stage. The optimal value to the original problem must be less than or equal to UPPER, and greater than or equal to LOWER.

2. For non-zero-or-one entries (which are infeasible) in the solution from the previous stage, the algorithm grows binary subtrees that fix one entry to zero or one, then solves the problem while relaxing other entries. The optimized value is a lower bound to this subproblem. If this lower bound is greater than the UPPER, we prune this branch, since the solutions to any feasible combination of the trees growing from this point are not going to be better (less) than the UPPER we currently have. Otherwise, we keep the node for further growing. The upper bound of this subproblem can again be yielded by substituting the rounded solution to the problem. If the upper bound to this subproblem is less than UPPER, we update UPPER, and mark this rounded solution (feasible) as the best solution so far.

3. Step 2 will be repeated until there is no node to be grown.

Details of the BnB algorithm can be found in [15].

Ideally, we should perform BnB on all packets ($N = 450 = 15 packets/frame \times 30 frames/GOP$ for our experiment) in a GOP (Group of picture) using all ($K = 13$) RCPC codes for the algorithm to select from. However, the complexity of BnB has worst-case exponential time [8]. In our prior work [7], we select from only 4 RCPC codes, and perform BnB on every N=15 packets, grouped by selecting the 2 most visible and 13 most invisible packets from the packets in a GOP, so that we can better reallocate channel rates among visible and invisible packets. For example, after sorting the packets by the estimated loss visibility in ascending order, the first group of packets includes packets 1-13, 449, 450 from the sorted packets, and the second 14-26, 447, 448, etc. This way, for each GOP, we need to perform a (N, K)=(15,4) BnB 30 times. We denote this the **SORTED** algorithm.

5. K-MEANS CLUSTERING

An algorithm involving (N, K)=(15,4) BnB 30 times for a GOP may still be complicated for real-time processing, therefore, we aim to reduce the complexity of the algorithm. From Equation 4 we see that the optimization problem depends on not only the visibility of the packet (V_i), but also the size of the packet (S_i). The idea is to group packets of similar visibility and packet size together and consider this group as one packet. Thus we propose to use k-means clustering to group packets before using BnB. Considering N packets in a GOP, each has a 2-dimensional vector (V_i,S_i), k-means partitions the N vectors into P clusters that minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances. Each of the P clusters has a quantized vector ($\bar{V}_z$,$\bar{S}_z$) for the cluster, $z = 1, 2, ..., P$. We then use these P vectors to perform the BnB. Equation 4 can now be rewritten as:

$$\begin{aligned} \min_{\bar{\mathbf{r}}} \ & \frac{1}{N}\sum_{z=1}^{P} \bar{V}_z \times Num_z \times \{1-(1-10^{a\frac{1}{\bar{r}_z}+b})^{\bar{S}_z}\} \\ s.t. \ & \sum_{z=1}^{P} \bar{S}_z \times Num_z \times \frac{1}{\bar{r}_z} \leq B \\ & \bar{r}_z \in \{R_1, R_2, ..., R_K\} \\ & z = 1, 2, ..., P \end{aligned} \quad (6)$$

where $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, ..., \bar{r}_P]$, and Num_z is the number of vectors in cluster z. After solving this problem, the optimal rate allocation for each packet $i, i = 1, 2, ..., N$, is the optimal $\bar{r}_z$ where packet i is in cluster z. In our experiment, we use $P = 15$. Therefore, for each GOP, we only need to do one $(N = P, K)$=(15,4) BnB, which is 30 times less complicated than SORTED is. The proposed algorithm is denoted **KMEANS**.

6. EXPERIMENTAL RESULTS

In this section we compare the performance of SORTED and KMEANS, in terms of perceptual video quality and algorithm complexity.

The video sequence used in our experiment is encoded by H.264/AVC JM Version 12.1 in SIF resolution (352×240)

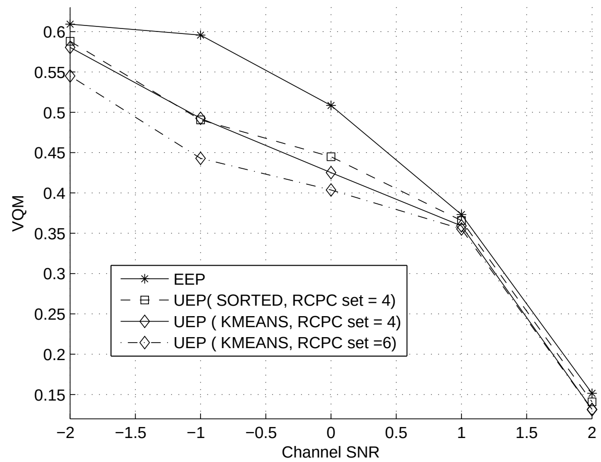

Fig. 2. The average VQM comparison among EEP, SORTED and KMEANS algorithms over 100 realizations of each AWGN channel.

with GOP structure IPPP, frame rate 30 fps, and encoding rate 600 $kbps$. We define a packet (a NAL unit) as a horizontal row of macroblocks. Therefore, there are 15 packets in a frame. Here we consider all packets in a GOP ($N = 450 = 15 \times 30$). The convolutional coder to produce the mother code of the RCPC code has rate $\frac{1}{L}$ where $L = 4$, with memory $M = 4$. The puncturing period of the RCPC code is $P = 8$. The channel we simulate for the wireless communication is AWGN. As mentioned previously, the logarithm of the bit error rate at a given SNR can be linearly related to the inverse of the RCPC rates, as shown in Fig 1. The coefficients of models for different SNRs are shown in Table 2. In this simulation, the RCPC rate used by EEP is $\frac{8}{14}$, and the RCPC rates from which our method can select are $\{\frac{8}{12}, \frac{8}{14}, \frac{8}{16}, \frac{8}{18}\}$. The budget for the optimization problem will be the number of bits used by the EEP in the optimization group. The simulated AWGN channel SNR ranges from -2 to 2 dB, corresponding to bit error rates after decoding from about 10^{-1} to 10^{-4} when the RCPC rate of EEP ($\frac{8}{14}$) is considered. The resulting source-decoded videos are evaluated by a full-reference metric VQM (Video Quality Metric) [9] which has been shown to be much closer than other perceptual video quality metrics to human perception. VQM ranges from 0 (excellent quality) to 1 (poor quality).

As discussed earlier, for a GOP, KMEANS is 30 times less complicated than SORTED. However, with much less computational complexity, we can see from Figure 2 that the VQM result of KMEANS is competitive to and even slightly better than the result by SORTED when the number of RCPC codes is the same ($K = 4$). This is because for SORTED, each optimization is performed for the heuristically grouped packets in a GOP, so the bit budget allocation is restricted in each small group of N=15 packets. However, KMEANS can assign the resources to representative packets from the total budget of N=450 packets, and therefore achieve better performance.

Since the complexity is lower for KMEANS, we try to improve the performance by offering the algorithm six RCPC code rates $\{\frac{8}{10}, \frac{8}{12}, \frac{8}{14}, \frac{8}{16}, \frac{8}{18}, \frac{8}{20}\}$ to select from. Figure 2 shows that at 0 dB, by using 6 RCPC codes, the improvement from EEP increases to 0.1 in VQM score, while at -1 dB, the VQM difference to EEP increases from 0.1 to 0.15. Note that a 0.1 improvement in VQM is considered to be a significant difference in VQM score (see, e.g., [16]). This means that the performance of our visibility-based algorithm could make more difference in the comparison if more RCPC codes are used in the optimization scheme, with consideration of reasonable algorithm complexity. In future work, we aim to lower complexity further to solve our problem with all 13 RCPC code rates.

In conclusion, we proposed a much less complicated algorithm than the one in our prior work to solve the channel rate allocation problem for end-to-end video perceptual quality. The proposed algorithm with low complexity performs better since no heuristic grouping is used. And because of the lower complexity, the code rate set can be enlarged and better visual performance is achieved by our visibility-based unequal error protection algorithm.

7. REFERENCES

[1] N. Thomos et al., "Robust Transmission of H.264/AVC Video using Adaptive Slice Grouping and Unequal Error Protection," *IEEE ICME*, pp. 593–596, July 2006.

[2] J.Y. Shih and W.J. Tsai, "A new unequal error protection scheme based on FMO," *ICIP*, pp. 3068–3071, 2008.

[3] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Processing: Image Communication*, vol. 19, no. 4, pp. 313 – 323, 2004.

[4] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[5] A. R. Reibman et al., "Predicting packet-loss visibility using scene characteristics," *Packet Video*, pp. 308–317, Sept 2007.

[6] T.-L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, and A. Reibman, "A Versatile Model for Packet Loss Visibility and its Application to Packet Prioritization," *IEEE Transactions on Image Processing*, accepted for publication.

[7] T.-L. Lin and P. Cosman, "Optimal RCPC Channel Rate Allocation in AWGN channel for perceptual video quality using Integer programming," *First International Workshop on Quality of Multimedia Experience (Qomex), IEEE*, 2009.

[8] B. Krishnamoorthy, "Bounds on the size of branch-and-bound proofs for integer knapsacks," *Operations Research Letters*, vol. 36, no. 1, pp. 19 – 25, 2008.

[9] VQM, "http://www.its.bldrdoc.gov/n3/video/," .

[10] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, September 2004.

[11] M. H. Loke, E. P. Ong, W. Lin, Z. Lu, and S. Yao, "Comparison of video quality metrics on multimedia videos," *IEEE ICIP*, October 2006.

[12] Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, Apr 2004.

[13] A. Hanjalic, "Content-based analysis of digital video," *Kluwer Academic Publishers, Boston*, 2004.

[14] B. Wimmer and T. Stockhammer, "Unequal error protection for H.263 video using RCPC codes," *ITU - Telecommunications Standardization Sector*, vol. Q11-G-12, July 1998.

[15] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2nd edition, 2004.

[16] S. L. P. Yasakethu, W. A. C. Fernando, S. Adedoyin, and A. Kondoz, "A rate control technique for off line H.264/AVC video coding using subjective quality of video," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1465–1472, Aug. 2008.