PRESCIENT MODE SELECTION FOR ROBUST VIDEO CODING

Rui Zhang, Shankar L. Regunathan and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Abstract

In standard predictive video coders, intra-mode coding of macroblocks (MBs) provides packet loss resilience, at the cost of reduced compression efficiency. Conventional mode selection algorithms are "greedy" as they focus on minimization of distortion of current MB, while ignoring the effect of this selection on subsequent frames. This paper proposes an algorithm for prescient mode selection: the coding modes of MBs are chosen while taking into account the distortion of subsequent frames. The problem is formulated as one of joint selection of MB coding modes, for a group of pictures, so as to minimize the rate-distortion cost. The straightforward solution based on dynamic programming requires enormous computational complexity. We propose an iterative algorithm which obtains a locally optimal solution at feasible complexity. The total decoder distortion is computed using recursive optimal per-pixel estimate (ROPE) which accurately accounts for the effects of quantization, packet loss, error propagation, and error concealment. Simulation results show consistent improvement over our previous "greedy" ROPE-RD mode selection, and substantial gains over other (non-ROPE) mode selection schemes.

1. INTRODUCTION

In packet-switched networks, packets may be discarded at intermediate nodes, or be considered lost due to long queuing delays. In the case of predictive video coding, the prediction loop propagates errors, and causes additional deterioration of the performance. Mode selection is a "standard-compatible" tool for mitigating the effects of packet loss. Intra-mode coding of macroblocks (MBs) stops error propagation while consuming more bits than inter-coding.

The problem of MB coding mode selection to balance the tradeoff between compression efficiency and robustness has received much attention [1] [2] [3]. While state-of-the-art mode selection algorithms improve the robustness of video coders, they share a common limitation. They perform the optimization for each frame independently, and ignore the effect of mode selection on subsequent frames.

In this work, we consider the problem of optimal MB mode selection by a "prescient" encoder, which has access to the current frame as well as to a limited number of subsequent frames. This problem is related to that of "dependent quantization" [4], and can be formulated as one of joint MB coding modes selection for a group of pictures. Solutions based on dynamic programming are globally optimal, but require enormous computational complexity. Instead, we propose an iterative algorithm that guarantees a local optimal solution at feasible complexity. Moreover, we show that the impact of mode selection on the distortion of future frames can be approximated to achieve further reduction in complexity. The scheme explicitly accounts for the impact of quantization, packet loss as well as error propagation on the total decoder distortion. This is achieved by an extension of the recursive optimal per-pixel estimate (ROPE) [2].

In section 2, we analyze the effect of the choice of MB mode on subsequent frames. In section 3, we formulate the joint optimization problem and derive an iterative solution. ROPE is used in section 4 to achieve further simplification of the algorithm. Simulation results in section 5 demonstrate the performance gains.

2. DEPENDENT MODE SELECTION

The standard video coder employs inter-frame prediction to remove temporal redundancies. Although inter-mode coding generally achieves higher compression efficiency, it is more sensitive to channel errors as it promotes error propagation.

It is widely recognized that adaptive intra-update is an important tool for mitigating the effects of packet loss. However, state-of-art approaches for coding mode selection perform optimization of the MB coding modes of each frame independently. Consequently, they neglect the impact of the choice of current coding mode on the rate-distortion performance in subsequent frames. We formulize this inter-frame dependency by extending our error propagation model [2].

Let f_n^i denote the value of pixel i in frame n, and let $\hat{f}_n^i$

This work is supported in part by the NSF under grant MIP-9707764, the University of California MICRO Program, Conexant Systems, Inc., Dolby Laboratories, Inc., Lucent Technologies, Inc., MedioStream, Inc. and Qualcomm, Inc.

represent its reconstruction at the encoder. The reconstructed value at the decoder, possibly after error concealment, is denoted by $\tilde{f}_n^i$. For the encoder, $\tilde{f}_n^i$ is a random variable. Let f_{n+1}^m, $\hat{f}_{n+1}^m$ and $\tilde{f}_{n+1}^m$ denote the original value, encoder reconstruction and decoder reconstruction of pixel m in frame $n+1$, which is motion compensated from pixel i in frame n. Thus, the predictor used by the encoder for pixel m is $\hat{f}_n^i$. The corresponding predictor at the decoder is $\tilde{f}_n^i$. If $\hat{r}_{n+1}^m$ represents the corresponding quantized residual, the encoder's reconstruction is given by $\hat{f}_{n+1}^m = \hat{r}_{n+1}^m + \hat{f}_n^i$. We assume that temporal error concealment is used to reconstruct this pixel in case of packet loss. Let this replacement be $\tilde{f}_n^k$. If packet loss rate is p, we have

$$\tilde{f}_{n+1}^m = (1-p)\{\hat{r}_{n+1}^m + \tilde{f}_n^i\} + p\tilde{f}_n^k. \qquad (1)$$

Our goal, here, is to formulize how the coding mode of frame n affects the distortion at frame $n+1$.

The overall expected decoder distortion, of pixel m in frame $n+1$, is

$$\begin{aligned} d_{n+1}^m &= E\{(f_{n+1}^m - \tilde{f}_{n+1}^m)^2\} \\ &= (1-p)E\{(f_{n+1}^m - \hat{r}_{n+1}^m - \tilde{f}_n^i)^2\} \\ &\quad +pE\{(f_{n+1}^m - \tilde{f}_n^k)^2\} \\ &\approx (1-p)E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\} + (1-p)(f_{n+1}^m - \hat{f}_{n+1}^m)^2 \\ &\quad +pE\{(f_{n+1}^m - \tilde{f}_n^k)^2\} \\ &= \bar{d}_{n\to n+1}^{i\to m} + \hat{d}_{n+1}^m + \tilde{d}_{n+1}^m. \qquad (2) \end{aligned}$$

Here, $\hat{d}_{n+1}^m$ and $\tilde{d}_{n+1}^m$ denote the quantization distortion and the error concealment distortion for pixel m in frame $n+1$. More importantly, $\bar{d}_{n\to n+1}^{i\to m}$ represents the effect of distortion propagation from frame n. Note that $\hat{f}_n^i$ and $\tilde{f}_n^i$ are the predictors used by the encoder and decoder respectively, and the propagation of distortion is due to the mismatch between the two. If the MB containing pixel i in frame n was intra-coded, the mismatch term would be strictly smaller than if it was inter-coded. Thus, the mode decision in frame n affects the distortion of frame $n+1$ and other frames in the future. Further, the final approximation assumes that the effect of quantization and error propagation are additive. This approximation will be useful in section 4 to reduce the complexity of the optimization scheme.

3. OPTIMIZATION VIA ITERATIVE DESCENT SEARCH

We now propose to optimize MB coding parameters while accounting for the temporal dependency analyzed in the previous section. The coding mode and the quantization step size are the parameters that can be optimized for each MB.

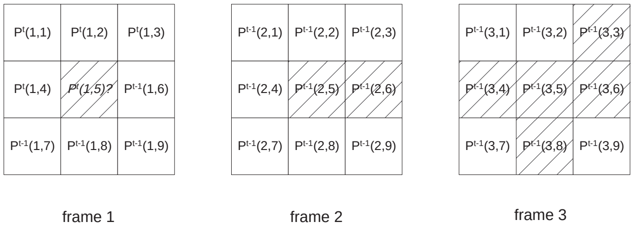

Fig. 1. Temporal-spatial dependency. As parameter for current MB changes, RD performance of some corresponding MBs in the following frames changes too.

Let P_n^m denote the coding parameters of MB m in frame n. Let the set of the parameters of all the MBs in frame n be denoted by $\mathcal{P}_n$. For a group of pictures with L frames, the problem is to select parameters $\{\mathcal{P}_n, \mathcal{P}_{n+1}, ..., \mathcal{P}_{n+L-1}\}$ jointly so that the total decoder distortion

$$\begin{aligned} D_n(\mathcal{P}_n) + D_{n+1}(\mathcal{P}_n, \mathcal{P}_{n+1}) + ... \\ +D_{n+L-1}(\mathcal{P}_n, \mathcal{P}_{n+1}, ..., \mathcal{P}_{n+L-1}) \end{aligned} \qquad (3)$$

is minimized, while satisfying the constraint on the rate,

$$\begin{aligned} R_n(\mathcal{P}_n) + R_{n+1}(\mathcal{P}_n, \mathcal{P}_{n+1}) + ... \\ +R_{n+L-1}(\mathcal{P}_n, \mathcal{P}_{n+1}, ..., \mathcal{P}_{n+L-1}) \leq R_{budget}. \end{aligned} \qquad (4)$$

Note here that the rate-distortion performance of each frame is dependent not only on the parameters of current frame, but also on parameters of previous frames.

This constrained minimization problem can be recast to an unconstrained minimization of

$$\begin{aligned} J_n + J_{n+1} + ... + J_{n+L-1} = \\ D_n + D_{n+1} + ... + D_{n+L-1} + \\ \lambda(R_n + +\lambda R_{n+1} + ... + \lambda R_{n+L-1}), \end{aligned} \qquad (5)$$

where λ is the Lagrangian multiplier. As this formulation is similar to the problem of dependent quantization [4], dynamic programming can be used to search for the best combination of coding mode and quantizer step size. However, an extremely large number of states are required to account for all possible combination of the parameters for all the MBs. Due to motion compensation, the RD curve of each MB may depend on the parameter choice of several MBs in the previous frame. If there are M MBs in a frame and N choices for the quantizer step size, the number of states for each frame is then $(2N)^M$. Optimization over L-frames introduces $(2N)^{ML}$ states. For example, if there are 5 possible quantization step sizes for each MB, even QCIF sequences would require 10^{99L} states. The enormous computational complexity of dynamic programming makes it impractical even if joint optimization is restricted to two frames.

We next derive a low-complexity iterative algorithm to jointly optimize MB parameters for a group of frames. Each

MB is associated, due to the spatio-temporal dependency introduced by motion compensation, with some MBs in subsequent frames. If the parameter of this MB is varied, while fixing all other parameters of this group of pictures, the rate-distortion performance of only the associated MBs are affected. Figure 1 depicts this inter-frame dependency of MBs. Thus, the optimal choice of coding mode and quantization step size for current MB, depends only on the RD cost of encoding the current MB as well as those of the associated MBs in subsequent frames. Based on this observation, we derive an iterative descent search algorithm:

- Step 1: Initialize parameters $\{\mathcal{P}_n^0, \mathcal{P}_{n+1}^0, ..., \mathcal{P}_{n+L-1}^0\}$. In our simulations, we used ROPE-RD [2] to obtain the initial parameter values.
- Step 2: Update parameter for each MB to minimizes *RD cost of encoding all associated MBs*, for fixed choice of parameters of *all other MBs*. Let the parameter set at the end of tth iteration be denoted by $\{\mathcal{P}_n^t, \mathcal{P}_{n+1}^t, ..., \mathcal{P}_{n+L-1}^t\}$.
- Step 3: Check for convergence: If parameters are identical with those of previous iteration, stop. Otherwise, go back to step 2.

Note that the total RD cost of encoding the frame never increases. Therefore, the algorithm produces a locally optimal set of parameters. The complexity of this algorithm is linear in the number of choices per parameter ($2N$), the number of MBs per frame (M), and the number of iterations (T). If A denotes the number of MBs in the next frame associated with each MB in current frame (A is typically around 2), The complexity of the algorithm is $2NTMA^{L-1}$. In comparison, the the complexity of dynamic programming is in the level of $(2N)^{ML}$. In simulations, the algorithm usually converged after $4 \sim 6$ iterations.

Note that this iterative algorithm is not specific to coding mode selection for packet loss resilience. In fact, it can be applicable for parameter optimization for the general dependent quantization [4] scenario, for both error free and lossy channels.

4. SIMPLIFIED DESCENT SEARCH

In this section, we further reduce the complexity of iterative search algorithm for mode selection. Note that the computationally intensive steps are the extra DCT and quantization operations required to re-encode and compute the rate-distortion performance of all associated MBs. To simplify the algorithm, we approximate the total decoder distortion as a combination of quantization distortion and error propagation, as in equation (2). This allows explicit computation of the impact of mode selection on the distortion of future frames, without the need for re-encoding those MBs.

We present the derivation for the special case when only one future frame is available for optimization. The extension to multiple frame dependcy is straightforward. Recall that the inter-frame dependency in distortion can be captured explicitly by the predictor mismatch term. From (2), we have

$$\begin{aligned} D_n + D_{n+1} &= \sum_{i \in frame(n)} d_n^i + \sum_{m \in frame(n+1)} d_{n+1}^m \\ &\approx \sum_{i \in frame(n)} (d_n^i + \omega_i \bar{d}_{n \rightarrow n+1}^{i \rightarrow m}) \\ &+ \sum_{m \in frame(n+1)} (\hat{d}_{n+1}^m + \tilde{d}_{n+1}^m) \\ &= D_n + \bar{D}_{n \rightarrow n+1} + \hat{D}_{n+1} + \tilde{D}_{n+1}. \quad (6) \end{aligned}$$

where ω_i is the number of pixels in frame $n+1$ that is motion compensated by the pixel i in frame n. Note that ω_i is associated with the mode selection (and motion vector) of frame $n+1$, while $\bar{d}_{n \rightarrow n+1}^{i \rightarrow m}$ is the result of coding mode decision in frame n. Therefore, the mismatch term is determined by the mode selection in both frames. The distortion due to error concealment, $\tilde{D}_{n+1}$, is independent of the coding mode in frame $n+1$. The dependency of the quantization distortion, $\hat{D}_{n+1}$, on the coding parameter of frame n is small relative to the other terms, and can be neglected. Based on these observations, we derive the simplified iterative algorithm to solve the unconstrained problem in equation (5):

- Step 1: Initialize the parameter sets $\mathcal{P}_n$ and $\mathcal{P}_{n+1}$.
- Step 2: For the given $\mathcal{P}_{n+1}$, compute ω_i for each pixel i in frame n.
- Step 3: Select the coding parameters, $\mathcal{P}_n$ for frame n, to minimize $D_n + \bar{D}_{n \rightarrow n+1} + \tilde{D}_{n+1} + \lambda R_n$. This is equivalent to minimizing $J_n + J_{n+1}$ for fixed $\mathcal{P}_{n+1}$.
- Step 4: Select the coding parameters, $\mathcal{P}_{n+1}$ for frame $n+1$, to minimize $D_{n+1} + \lambda R_{n+1}$. This is equivalent to minimizing $J_n + J_{n+1}$ for fixed $\mathcal{P}_n$.
- Step 5: If the mode selection convergence criteria is not satisfied, go to step 2; Otherwise, stop.

Standard ROPE-RD based mode selection is used to execute step 3 and step 4 of the algorithm. As the impact of mode selection on the distortion of future frames is explicitly computed, the DCT and quantization for frame $n+1$ are unnecessary when optimizing the parameters for frame n. Thus, complexity is further reduced.

5. SIMULATION RESULTS

We call the proposed mode selection algorithm Prescient ROPE-RD (PROPE-RD). We simulated both the iterative

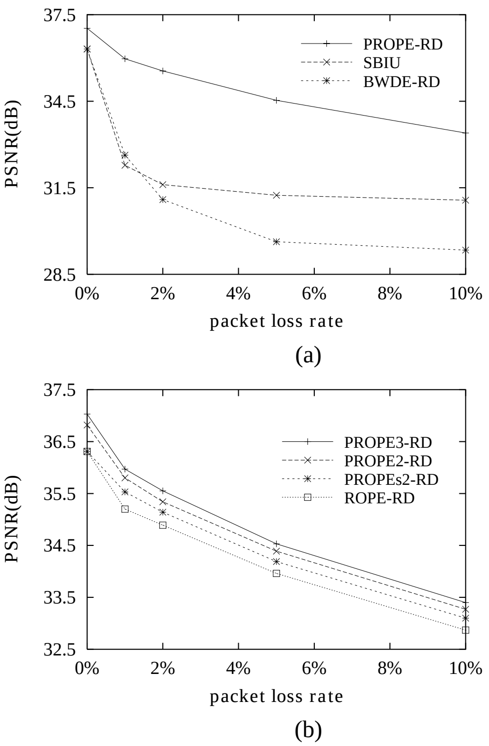

Fig. 2. PSNR vs. Packet loss rate for sequence "carphone".

algorithm and its simplified approximation (denoted by suffix 's'). We used a sliding window to form the group of pictures, and hence maintained a fixed delay. We considered simple two-frame and three-frame dependency models. The algorithms were implemented by modifying the UBC H.263+ coder [5]. The RTP payload format [6] is assumed for packetization. A random packet loss generator is used to drop packets at a specified loss rate. 250 frames of the QCIF sequence "carphone" are compressed. The PSNR of luminance reconstruction is computed and averaged over 30 different channel realizations (with different packet loss patterns).

The PROPE-RD scheme is compared with another two non-ROPE mode selection approaches: (i) "Scattered block intra-update" (SBIU) which arbitarily assigns MBs to $1/p$ groups, and cyclically intra-updates one group per frame, (ii) "block weighted distortion estimate" (BWDE) [7] which performs an approximate computation of decoder distortion. Figure 2 (a) presents the PSNR of decoder reconstruction at various packet loss rates. The PROPE-RD schemes yield significant gains over the other two methods.

We further compared with our previous ROPE-RD approach, where the coding decision is made greedily for each frame. Results in Figure 2 (b) demonstrates the additional gain of around 0.2∼0.7dB over ROPE-RD, depending on the number of frames in the group of pictures and the accuracy of the calculation. These improvements are achieved consistently for varying packet loss rate. Note that there is gain even when the packet loss rate is 0%. This shows that iterative search algorithm can be used for parameter optimization in dependent quantization even when the channel is loss free.

6. CONCLUSION

We proposed a "non-greedy" mode selection algorithm to improve the robustness of video coding to packet loss. The problem is formulated as one of joint optimization of the coding parameters of a group of pictures. An iterative algorithm is used to obtain a locally optimal set of parameters at feasible complexity. The complexity of the algorithm can be further reduced by approximating the total decoder distortion as sum of quantization and error propagation. ROPE is used to precisely calculate the decoder distortion, and for initializing the algorithm. Simulation results demonstrate consistent gains over greedy ROPE-based mode selection, and substantial gains over non-ROPE based mode selection methods.

7. REFERENCES

[1] E. Steinbach, N. Farber and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. on Circuits and Systems for Video Technology*, pp. 872-881, Vol. 7, No. 6, Dec. 1997.

[2] R. Zhang, S. L. Regunathan and K. Rose, "Video coding with optimal intra/inter mode switching for packet loss resilience", *IEEE Journal of Selected Areas in Communications*, pp. 966-76. vol. 18, June 2000.

[3] G. Cote, S. Shirani and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks". *IEEE Journal on Selected Areas in Communications*, pp. 952-65, vol. 18, June 2000.

[4] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantization with applications to MPEG video coders", *Proceedings of ICASSP'93*, vol. 5, pp. 381-384.

[5] H.263+ codec, http://spmg.ece.ubc.ca/

[6] "RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+)", Internet Draft, RFC2429, http://www.faqs.org/rfcs/rfc2429.html

[7] G. Cote and F. Kossentini, "Optimal Intra Coding of Blocks for Robust Video Communication over the Internet," *Image Communication*, pp. 25-34, Sept. 1999.