ELSEVIER

Signal Processing: *Image Communication* 16 (2001) 725–732

SIGNAL PROCESSING:
IMAGE
COMMUNICATION

www.elsevier.nl/locate/image

Scalable video coding with robust mode selection

Shankar Regunathan, Rui Zhang, Kenneth Rose*

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

Abstract

We propose to improve the packet loss resilience of scalable video coding. An algorithm for optimal coding mode selection for the base and enhancement layers is developed, which limits error propagation due to packet loss, while retaining compression efficiency. We first derive a method to estimate the overall decoder distortion, which includes the effects of quantization, packet loss and error concealment employed at the decoder. The estimate accounts for temporal and spatial error propagation due to motion compensated prediction, and computes the expected distortion precisely *per pixel*. The distortion estimate is incorporated within a rate-distortion framework to optimally select the coding mode as well as quantization step size for the macroblocks in each layer. Simulation results show substantial performance gains for both base and enhancement layers. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Video coding; Error resilience; Packet loss; Scalable coding; Source-channel coding; Error propagation; Error concealment

1. Introduction

Scalable coding is an important tool for efficient transmission of video over packet switched networks. The scalable coder transmits essential information on the video source in the base layer which can be decoded independently to obtain a coarse quality of reconstruction. Additional information is transmitted in higher enhancement layers, which complements the base layer information, to improve the video reconstruction at the decoder. The syntax for scalable coding is provided in the H.263+ and MPEG standards.

Scalable video coding offers means for robustness, as base-layer reconstruction may be used as a fall-back option in case of severe packet loss [1,5]. For example, ATM networks can assign higher priority to the base-layer cells in case of congestion. In wireless networks, base-layer packets may be protected by stronger error correction codes than enhancement-layer packets. However, in practice, some packet loss is inevitable even in the base-layer. Moreover, error propagation will amplify the effect of packet losses in both base and enhancement layers, and will further degrade the performance. In this paper, we propose an optimal strategy for coding mode selection per macroblock (MB) in both the base and enhancement layers, which substantially improves the robustness of scalable video coding systems. While there is a considerable volume of published work on mode selection for packet loss resilience in the single-layer (non-scalable) video coding (e.g. [2,3,6,7]), very little work has been reported on the corresponding problem in scalable video coding.

* Corresponding author. Tel.: + 1-805-893-7024; fax: + 1-805-893-3262.

E-mail address: rose@ece.ucsb.edu (K. Rose).

PII: S0923-5965(01)00003-0

We focus on an SNR scalable system, which provides layers with the same spatial-temporal resolution but different reconstruction quality. The key step in our derivation is the estimation of the overall decoder distortion that takes into account the effects of quantization, packet loss and error concealment. To calculate this estimate, we extend the recursive optimal per-pixel estimate (ROPE) which we had proposed for non-scalable video coding [6,7]. The extended ROPE is shown to accurately account for both temporal and spatial error propagation, and to compute the total distortion in each layer at pixel-level precision. For each MB, the prediction mode and quantization step size are jointly selected to minimize the rate-distortion (RD) cost. Simulation results show substantial gains in reconstructed video PSNR at the base and enhancement layers. While simulation results are presented in the context of H.263+ scalable coding, it is important to note that ROPE-based mode selection can improve the robustness of MPEG-4 and other block-based scalable video coders.

The paper is organized as follows. In Section 2, we derive the extended ROPE model that computes the optimal estimate of the overall distortion in decoder reconstruction for each layer. We incorporate the estimate within an RD framework for optimal selection of mode and quantizer parameters in Section 3. Section 4 presents simulation results to demonstrate the performance of the method.

2. Recursive optimal per-pixel estimate of decoder distortion in scalable coding

2.1. Preliminaries

The standard video coder segments the video frame into MBs. In the base layer, the MBs may be encoded in either inter-mode or intra-mode. In inter-mode, the MB is "predicted" from the previously decoded frame via motion compensation, and the prediction error is encoded. In intra-mode, the original MB data is encoded directly. The enhancement layer typically offers three possible prediction modes [8]: MBs can be predicted from the current base layer (upward), from the previous enhancement layer (forward), or via combined prediction using both (bi-directional). The prediction residue is then transform coded.

Mode selection is a powerful standard-compatible tool to trade compression efficiency for packet loss resilience. The use of intra-mode in the base layer, or upward prediction in the enhancement layer, can limit error propagation and is more effective during scene changes. However, in general, they are more costly in quantization bits. An optimal mode selection strategy at the encoder should minimize the overall distortion in decoder reconstruction, including both quantization and packet loss effects, for the given bit-rate. Thus, a key task at the encoder is the estimation of overall decoder distortion.

However, this task is complicated by two factors. Spatial error propagation beyond MB boundaries (due to motion compensation) can only be accurately accounted for by computing the distortion per pixel. Further, distortion due to quantization and packet loss are not additive, but are instead combined in a highly complex fashion to produce the overall distortion. In this section, we derive an algorithm to accurately estimate the total distortion in decoder reconstruction at the various layers of a scalable coder.

We assume that groups of blocks (GOB) are packetized such that odd and even rows are carried in different packets. Further, by prohibiting inter-GOB motion vector prediction, we ensure that the packets are independently decodable. Such packetization allows missing blocks to be concealed using information recovered from neighbouring GOBs at the cost of slight degradation in compression performance. In this case, the pixel loss rate equals the packet loss rate. The number of packets per frame can vary with the target rate so as to minimize the overhead due to packet headers.

We model the channel as a Bernoulli process with packet loss rate p_b for the base layer, and packet loss rate p_e for the enhancement layer. Note that this model is only assumed for presentation simplicity, and more complex models may be incorporated. We also assume that the error concealment scheme is known to the encoder.

Let f_n^i denote the original value of pixel i in frame n, let $\hat{f}_n^i(b)$ and $\hat{f}_n^i(e)$ denote its encoder reconstruction at the base and enhancement layer, respectively. The reconstructed values at the decoder, possibly after error concealment, are denoted by $\tilde{f}_n^i(b)$ and $\tilde{f}_n^i(e)$. For the encoder, $\tilde{f}_n^i(b)$ and $\tilde{f}_n^i(e)$ are random variables. Assuming mean square error distortion, the overall expected distortion for this pixel, at the base and enhancement layers, is given by

$$d_n^i(b) = E\{(f_n^i - \tilde{f}_n^i(b))^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(b)\} + E\{(\tilde{f}_n^i(b))^2\}. \quad (1)$$

$$d_n^i(e) = E\{(f_n^i - \tilde{f}_n^i(e))^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(e)\} + E\{(\tilde{f}_n^i(e))^2\}. \quad (2)$$

We observe that the computation of $d_n^i(b)$ and $d_n^i(e)$ requires the first and second moments of the corresponding random variables, and develop recursion formulae to sequentially compute these two moments.

2.2. ROPE for the base layer

It is easy to see that the problem of base layer mode selection is identical to that of non-scalable coding. Thus, the ROPE algorithm derived in [6,7] may be directly applied for calculating the total decoder distortion. We briefly summarize the algorithm in this subsection.

We assume, for presentation simplicity, that the temporal error concealment technique is in use at the decoder. If the MB containing pixel i is lost, temporal replacement is used for error concealment, i.e., the motion vector of this MB is estimated as the median of the motion vectors of the nearest three MBs in the previous GOB (above). Let the estimated motion vector associate pixel i with pixel k in the previous frame. We thus have $\tilde{f}_n^i(b) = \tilde{f}_{n-1}^k(b)$. In terms of our packet loss model, the probability of this event is $p_b(1 - p_b)$. If the previous GOB is also lost, the estimated motion vector is set to zero, and we have $\tilde{f}_n^i(b) = \tilde{f}_{n-1}^i(b)$, with probability p_b^2. If the MB is correctly received and has been intra-coded, we have $\tilde{f}_n^i(b) = \hat{f}_n^i(b)$ with probability $(1 - p_b)$. Thus, for a pixel in an intra-coded MB,

$$E\{\tilde{f}_n^i(b)\} = (1 - p_b)(\hat{f}_n^i(b)) + p_b(1 - p_b)E\{\tilde{f}_{n-1}^k(b)\} + p_b^2 E\{\tilde{f}_{n-1}^i(b)\},$$

$$E\{(\tilde{f}_n^i(b))^2\} = (1 - p_b)(\hat{f}_n^i(b))^2 + p_b(1 - p_b)E\{(\tilde{f}_{n-1}^k(b))^2\} + p_b^2 E\{(\tilde{f}_{n-1}^i(b))^2\}. \quad (3)$$

If an inter-coded MB is correctly received, the decoder has access to the quantized residue, $\hat{e}_n^i(b)$, and the motion vector. Let the motion vector be such that pixel i is predicted from pixel j in the previous frame. The encoder's prediction is given by $\hat{g}_n^i(b) = \hat{f}_{n-1}^j(b)$, and its reconstruction is given by $\hat{f}_n^i(b) = \hat{e}_n^i(b) + \hat{g}_n^i(b)$. The decoder must use its prediction, $\tilde{g}_n^i(b) = \tilde{f}_{n-1}^j(b)$. The corresponding decoder reconstruction is given by $\tilde{f}_n^i(b) = \hat{e}_n^i(b) + \tilde{g}_n^i(b)$, with probability $(1 - p_b)$. As the decoder's prediction is not identical to encoder's prediction, error propagation occurs even if the residue is received correctly. Thus, for a pixel in an inter-coded MB,

$$E\{\tilde{f}_n^i(b)\} = (1 - p_b)(\hat{e}_n^i(b) + E\{\tilde{g}_n^i(b)\}) + p_b(1 - p_b)E\{\tilde{f}_{n-1}^k(b)\} + p_b^2 E\{\tilde{f}_{n-1}^i(b)\},$$

$$\begin{aligned} E\{(\tilde{f}_n^i(b))^2\} &= (1 - p_b)E\{(\hat{e}_n^i(b) + \tilde{g}_n^i(b))^2\} + p_b(1 - p_b)E\{(\tilde{f}_{n-1}^k(b))^2\} + p_b^2 E\{(\tilde{f}_{n-1}^i(b))^2\} \\ &= (1 - p_b)((\hat{e}_n^i(b))^2 + 2\hat{e}_n^i(b)E\{\tilde{g}_n^i(b)\} + E\{(\tilde{g}_n^i(b))^2\}) \\ &\quad + p_b(1 - p_b)E\{(\tilde{f}_{n-1}^k(b))^2\} + p_b^2 E\{(\tilde{f}_{n-1}^i(b))^2\}. \end{aligned} \quad (4)$$

2.3. ROPE for the enhancement layer

We now extend the ROPE algorithm to estimate the decoder distortion at the enhancement layers. If an MB in the enhancement layer is lost, the decoder uses the corresponding base-layer block for error concealment.

Let us denote the prediction value at the encoder side as $\hat{g}_n^i(e)$, and that of the decoder side as $\tilde{g}_n^i(e)$. Let the transmitted residue be denoted by $\hat{e}_n^i(e)$. Note that $\hat{g}_n^i(e)$ and $\tilde{g}_n^i(e)$ are not identical. Thus, even if the packet containing the current pixel is received correctly (with probability $(1 - p_e)$), the reconstruction at the encoder, $\hat{f}_n^i(e) = \hat{e}_n^i(e) + \hat{g}_n^i(e)$, is different from the reconstruction at the decoder, $\tilde{f}_n^i(e) = \hat{e}_n^i(e) + \tilde{g}_n^i(e)$. Note that $\tilde{f}_n^i(e)$ and $\tilde{g}_n^i(e)$ must be treated as random variables by the encoder.

Thus, we have the following recursion functions for the expected moments of $\tilde{f}_n^i(e)$:

$$E\{\tilde{f}_n^i(e)\} = (1 - p_e)(\hat{e}_n^i(e) + E\{\tilde{g}_n^i(e)\}) + p_e E\{\tilde{f}_n^i(b)\},$$

$$\begin{aligned} E\{(\tilde{f}_n^i(e))^2\} &= (1 - p_e)E\{(\hat{e}_n^i(e) + \tilde{g}_n^i(e))^2\} \\ &\quad + p_e E\{(\tilde{f}_n^i(b))^2\} \\ &= (1 - p_e)((\hat{e}_n^i(e))^2 + 2\hat{e}_n^i(e)E\{\tilde{g}_n^i(e)\} \\ &\quad + E\{(\tilde{g}_n^i(e))^2\}) + p_e E\{(\tilde{f}_n^i(b))^2\}, \end{aligned} \qquad (5)$$

where the base layer moments are calculated as described in the previous section.

Let the MB motion vector associate pixel i with pixel j in the previous frame. The encoder and decoder predictors are specified for each prediction mode as:
upward prediction mode:

$$\begin{aligned} \hat{g}_n^i(e) &= \hat{f}_n^i(b), \\ \tilde{g}_n^i(e) &= \tilde{f}_n^i(b); \end{aligned} \qquad (6)$$

forward prediction mode:

$$\begin{aligned} \hat{g}_n^i(e) &= \hat{f}_{n-1}^j(e), \\ \tilde{g}_n^i(e) &= \tilde{f}_{n-1}^j(e); \end{aligned} \qquad (7)$$

bi-directional prediction mode:

$$\begin{aligned} \hat{g}_n^i(e) &= (\hat{f}_{n-1}^j(e) + \hat{f}_n^i(b))/2, \\ \tilde{g}_n^i(e) &= (\tilde{f}_{n-1}^j(e) + \tilde{f}_n^i(b))/2. \end{aligned} \qquad (8)$$

We reemphasize that these recursions are performed per-pixel at the encoder in order to estimate the expected total distortion at the decoder as accurately as possible. While for simplicity the recursions have been derived within a two-layer scalable coding setup, they can be extended in a straightforward manner to compute the total decoder distortion at each layer of a multi-layer video coder.

Note that the estimate is precise for integer-pixel motion estimation. In the half-pixel case, the bilinear interpolation makes the exact computation of the second moment highly complex. The estimate may be approximated by the simpler recursion of integer-pixel motion compensation. An alternate approach to approximate computation of the second moment in the half-pixel is presented in [4]. Further, for bi-directional prediction, we make the approximation:

$$E\{\tilde{f}_n^i(b)\tilde{f}_{n-1}^j(e)\} = E\{\tilde{f}_n^i(b)\}E\{\tilde{f}_{n-1}^j(e)\}. \qquad (9)$$

Although above approximations are sub-optimal, substantial gains are achieved.

The computational complexity of implementing ROPE algorithm at each layer is comparable to that of performing DCT [7]. It is important to note that the additional complexity is incurred only at the encoder.

2.4. Simplified ROPE for the special case of guaranteed base layer

An important practical scenario in scalable video coding is when the base-layer packets are transmitted with guaranteed reception or with negligible packet loss rate. In this case, the decoder reconstruction at the base-layer can be well approximated by the encoder reconstruction, i.e., $\tilde{f}_n^i(b) = \hat{f}_n^i(b)$. In this special case, we can use a simplified ROPE to calculate the enhancement-layer distortion. The recursions for the enhancement layer may be rewritten as

$$E\{\tilde{f}_n^i(e)\} = (1 - p_e)(\hat{e}_n^i(e) + E\{\tilde{g}_n^i(e)\}) + p_e\hat{f}_n^i(b),$$

$$\begin{aligned} E\{(\tilde{f}_n^i(e))^2\} &= (1 - p_e)E\{(\hat{e}_n^i(e) + \tilde{g}_n^i(e))^2\} + p_e(\hat{f}_n^i(b))^2 \\ &= (1 - p_e)((\hat{e}_n^i(e))^2 \\ &\quad + 2\hat{e}_n^i(e)E\{\tilde{g}_n^i(e)\} + E\{(\tilde{g}_n^i(e))^2\}) + p_e(\hat{f}_n^i(b))^2, \end{aligned} \qquad (10)$$

where the base-layer prediction is given for the three prediction modes by:
upward prediction mode:

$$\hat{g}_n^i(e) = \tilde{g}_n^i(e) = \hat{f}_n^i(b); \qquad (11)$$

forward prediction mode:

$$\begin{aligned} \hat{g}_n^i(e) &= \hat{f}_{n-1}^j(e), \\ \tilde{g}_n^i(e) &= \tilde{f}_{n-1}^j(e); \end{aligned} \qquad (12)$$

bi-directional prediction mode:

$$\begin{aligned} \hat{g}_n^i(e) &= (\hat{f}_{n-1}^j(e) + \hat{f}_n^i(b))/2, \\ \tilde{g}_n^i(e) &= (\tilde{f}_{n-1}^j(e) + \hat{f}_n^i(b))/2. \end{aligned} \qquad (13)$$

3. RD optimized mode selection algorithm for scalable coding

We next incorporate the distortion estimate, produced by the ROPE algorithm within an RD framework. The overall method selects the coding mode and quantization step size of each MB so as to minimize the decoder distortion at the given bit-rate.

The fundamental rate-distortion problem at hand is that of jointly selecting the coding modes for all the MBs to minimize the total distortion, D, subject to a given rate constraint, R. Equivalently, we may recast the problem as an unconstrained Lagrangian minimization, $J = D + \lambda R$, where λ is the Lagrange multiplier. Note that individual MB contributions to this cost are additive and, hence, the cost may be independently minimized for each MB. The coding modes are optimized for the base and enhancement layers sequentially.

For the base layer, the optimal mode and quantization step size for each MB are chosen by the simple minimization:

$$\min_{\text{mode}} (J_{\text{MB}}(b)) = \min_{\text{mode}} (D_{\text{MB}}(b) + \lambda_b R_{\text{MB}}(b)), \quad (14)$$

where the distortion of the MB is the sum of the distortion contributions of the individual pixels,

$$D_{\text{MB}}(b) = \sum_{i \in \text{MB}} d_n^i(b). \quad (15)$$

For the enhancement layer, the prediction mode and quantization step size are chosen to minimize

$$\min_{\text{mode}} (J_{\text{MB}}(e)) = \min_{\text{mode}} (D_{\text{MB}}(e) + \lambda_e R_{\text{MB}}(e)), \quad (16)$$

where the distortion of the MB is given by

$$D_{\text{MB}}(e) = \sum_{i \in \text{MB}} d_n^i(e). \quad (17)$$

Note that we use ROPE to calculate the distortion per pixel, while the coding mode and quantization step size are selected per MB via (14) and (16). The rate is controlled by using the "buffer status" to update λ_b and λ_e as in [7].

4. Simulation results

For the simulations, we implemented the ROPE-RD mode selection strategy by appropriately modifying the UBC H.263+ codec with two-layer scalability [9]. The RTP payload format [10] is assumed for packetization, and each packet contains one GOB. A random packet loss generator is used to drop packets at a specified loss rate. In the proposed system, the ROPE-RD algorithm is used to select mode and quantizer parameters in both layers. The comparison group consists of methods that use random intra-update (RIU) [2] in the base layer, where MBs are randomly intra-coded at the rate of $1/p_b$. In the enhancement layer, we compare the proposed scheme with two standard approaches for prediction mode selection. One method employs the quantization distortion estimate (QDE) within an RD framework to make the selection among the three prediction modes. The second approach only uses the upward prediction (UP) mode. UP ensures that there is no error propagation when the base-layer is loss free. 250 frames from QCIF video sequence "carphone" and CIF video sequence "LTS" are compressed. The PSNR of luminance reconstruction is computed for the sequence and averaged over 30 different channel simulations (with different packet loss patterns).

Fig. 1 shows the results for packet loss rates in the base and enhancement layer of 5% and 15%, respectively. In the base layer, the proposed ROPE based mode selection outperforms RIU by 0.4–1.0 dB on "carphone" and 0.6–1.2 dB on "LTS". In the enhancement-layer, ROPE based robust mode selection achieves PSNR gains of 0.9–1.8 dB on the "carphone" sequence and 1.2–2 dB on the "LTS" sequence, over the competing methods. This corresponds to additional improvement of 0.5–0.8 dB.

Figs. 2 and 3 and Table 1 present the results when reception of base layer packets is guaranteed. In this case, base-layer performance is identical for all competing methods. Enhancement layer PSNR is shown versus packet loss rate in Fig. 2, and versus enhancement layer bit-rate (as a fraction of total rate) in Fig. 3. The performance on different video sequences is shown in Table 1. Note that the relative performance of QDE and UP depends on

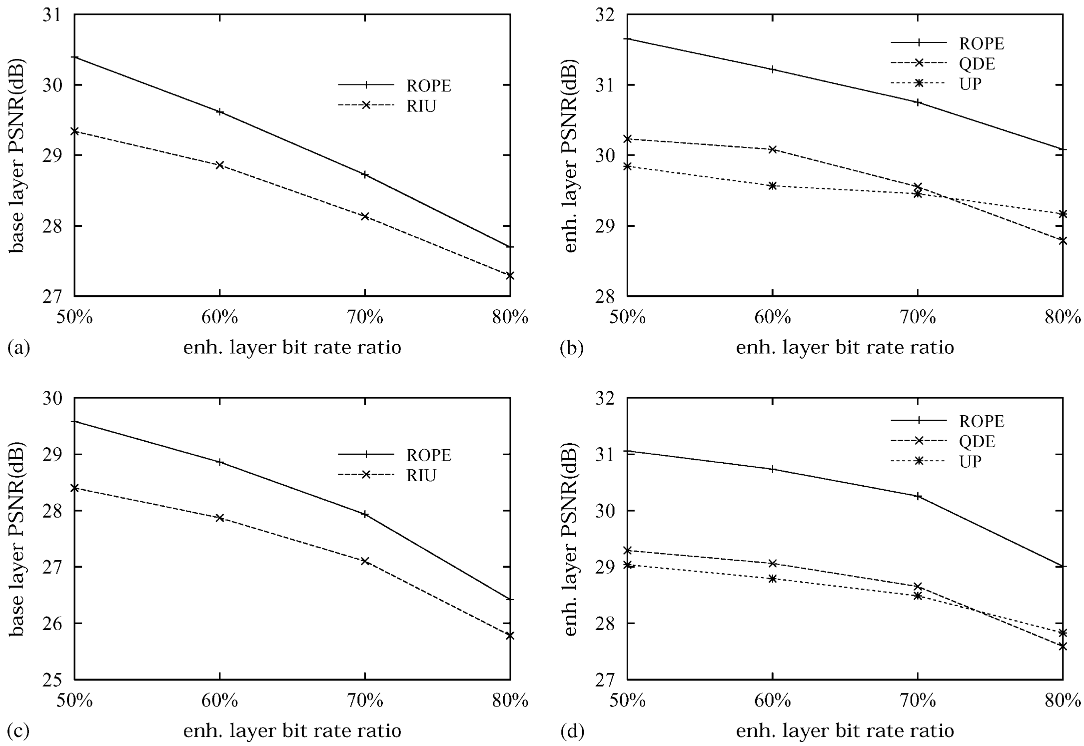

Fig. 1. PSNR versus enhancement layer bit-rate (as a fraction of total rate). The base layer is lossy. Base layer methods: ROPE (proposed), RIU [2]; enhancement layer methods: ROPE (proposed), QDE, UP. Base layer packet loss rate = 5%, enhancement layer packet loss rate = 15%. QCIF sequence "carphone" (frame rate = 10 fps, total bit-rate = 100 kbps): (a) base layer PSNR, (b) enhancement layer PSNR. CIF sequence "LTS" (frame rate = 15 fps, total bit-rate = 600 kbps): (c) base layer PSNR, (d) enhancement layer PSNR.

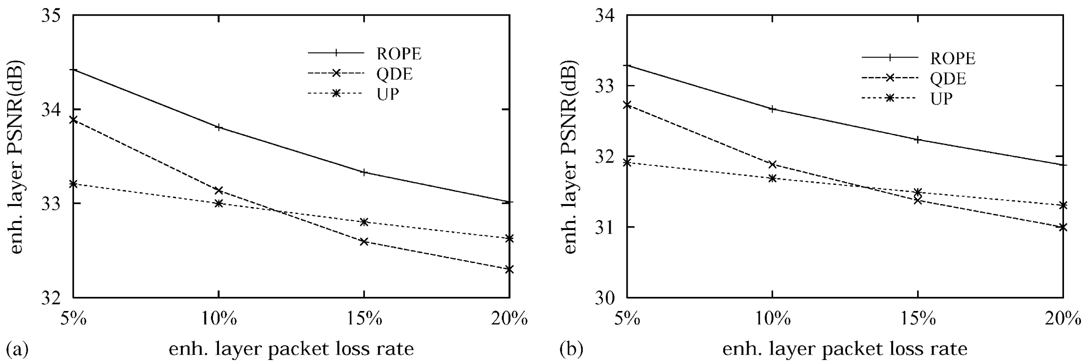

Fig. 2. PSNR versus enhancement layer packet loss rate. The base layer is loss free. Methods: ROPE (proposed), QDE, UP. Enhancement layer bit-rate ratio = 75%. (a) QCIF sequence "carphone" (frame rate = 10 fps, total bit-rate = 100 kbps), (b) CIF sequence "LTS" (frame rate = 15 fps, total bit-rate = 600 kbps).

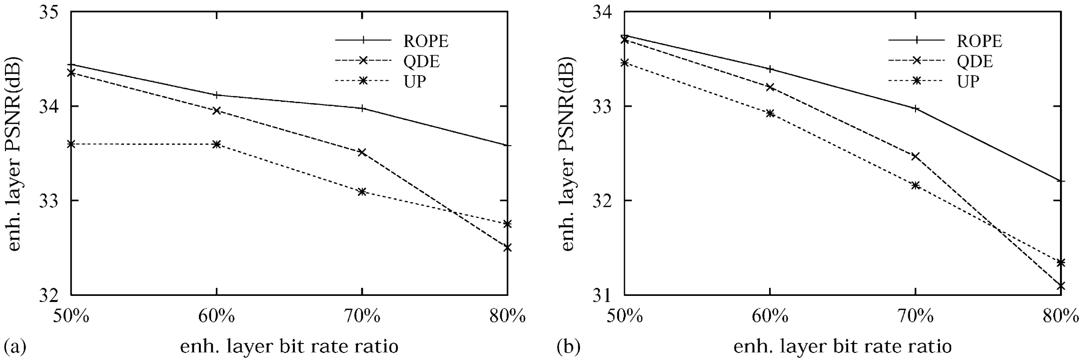

Fig. 3. PSNR versus enhancement layer bit-rate (as a fraction of total bit-rate). The base layer is loss free. Methods: ROPE (proposed), QDE, UP. Enhancement layer packet loss rate = 10%. (a) QCIF sequence "carphone" (frame rate = 10 fps, total bit-rate = 100 kbps), (b) CIF sequence "LTS" (frame rate = 15 fps, total bit-rate = 600 kbps).

Table 1
Performance comparison of prediction mode selection methods (Enhancement layer rate = 75% of total rate, enhancement layer packet loss rate = 10%. Frame rate: 10 fps for QCIF sequences, 15 fps for CIF sequence. Total bit-rate: 100 kbps for QCIF sequences, 600 kbps for CIF sequence)

Sequence	ROPE	UP	QDE
Carphone	33.81 dB	33.00 dB	33.14 dB
Foreman	31.74 dB	31.08 dB	31.13 dB
Mother&daughter	35.77 dB	35.06 dB	35.07 dB
LTS	32.67 dB	31.69 dB	31.89 dB

the packet loss rate and the enhancement layer bit-rate. The proposed ROPE, however, consistently outperforms both the competing methods.

The generality of the approach should be re-emphasized, and, in particular, similar performance gains are expected when ROPE-RD is incorporated into other scalable video coding schemes such as MPEG.

5. Conclusion

We propose a method for optimal mode selection in scalable video coding, which enhances robustness to packet loss. The method accurately estimates the overall decoder distortion for each layer at pixel-level precision by accounting for quantization, error propagation due to packet loss, and error concealment scheme employed at the decoder. The estimate is then incorporated within an RD framework for optimal macroblock mode selection in each layer. Simulation results show that the proposed method consistently outperforms conventional mode selection methods, and achieves significant PSNR gains in both base and enhancement layers. The algorithm requires no modification of the coding syntax or the decoder. Thus, it is compatible with standards such as H.263+ and MPEG.

Acknowledgements

This work was supported in part by the National Science Foundation under grants MIP-9707764 and EIA-9986057, the University of California MICRO program, Conexant Systems, Inc., Fujitsu Laboratories of America, Inc., Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., MedioStream Inc., Qualcomm, Inc., and Texas Instruments, Inc.

References

[1] R. Aravind, M.R. Civanlar, A.R. Reibman, Packet loss resilience of MPEG-2 scalable video coding algorithms, IEEE Trans. Circuits Systems Video Technol. 6 (5) (October 1996) 426–435.

[2] G. Cote, F. Kossentini, Optimal intra coding of blocks for robust video communication over the Internet, Signal Processing: Image Communication 15 (1–2) (September 1999) 25–34.
[3] E. Steinbach, N. Farber, B. Girod, Standard compatible extension of H.263 for robust video transmission in mobile environments, IEEE Trans. Circuits Systems Video Technol. 7 (6) (December 1997) 872–881.
[4] K. Stuhlmuller, Modeling and optimization of video transmission systems, Ph.D. Dissertation, University of Erlangen-Nuremberg, 2000.
[5] J. Villasenor, Y.-Q. Zhang, J.-T. Wen, Robust video coding algorithms and systems, Proc. IEEE 87 (10) (October 1999) 1724–1733.
[6] R. Zhang, S.L. Regunathan, K. Rose, Optimal intra/inter mode switching for robust video communication over the Internet, 33rd Asilomar Conference on Signals, Systems, and Computers, 24–29 October 1999.
[7] R. Zhang, S.L. Regunathan, K. Rose, Video coding with optimal intra/inter mode switching for packet loss resilience, IEEE J. Selected Areas Commun. (special issue on Error-Resilient Image and Video Transmission) 18 (6) (June 2000) 966–976.
[8] ITU-T Recommendation H.263, Video coding for low bit rate communication, 1998.
[9] H.263+ codec, http://spmg.ece.ubc.ca/
[10] RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+), Internet Draft, RFC2429,ftp://ftp.isi.edu/in-notes/rfc2429.txt.