CERIAS Tech Report 2001-73

A Hybrid Embedded Video Codec Using Base Layer Information For Enhancement Layer Coding

by Eugene T. Lin and Christine I. Podilchuk and Arnaud Jacquin and Edward J. Delp

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

A HYBRID EMBEDDED VIDEO CODEC USING BASE LAYER INFORMATION FOR ENHANCEMENT LAYER CODING

Eugene Lin, Christine Podilchuk*

Bell Labs, Lucent Technologies
Multimedia Research Laboratory
600 Mountain Avenue
Murray Hill, NJ
USA

Arnaud Jacquin, Edward Delp

Video and Image Processing Laboratory (VIPER)
School of ECE
Purdue University
West Lafayette, Indiana
USA

ABSTRACT

Scalability has become an important feature for video coding algorithm design for heterogeneous networks due to variable bandwidth, variable wired and wireless network conditions and variable terminal capabilities. However, the best video compression algorithms are based on temporal prediction through motion compensation which do not lend themselves naturally to a scalable framework. More recently, hybrid coding schemes have been introduced that combine a base-layer motion-compensation coder with an enhancement layer which offers fine-grain scalability through an embedded or progressive coder. Such a framework usually results in some compression efficiency loss over the best single layer motion compensated scheme. Here, we introduce a hybrid coding scheme which combines a base-layer MC coder and a finely scalable enhancement layer where the information from the base-layer is used to determine the location of the high energy signal in the enhancement layer. This coder provides better compression results than embedded approaches which do not rely on base layer information.

1. INTRODUCTION

Scalability has become an important feature for video coding algorithms for transmission over heterogeneous networks with variable bandwidth, variable wired and wireless network conditions and variable terminal capabilities. Ideally, scalability should provide tha ability to scale back the bitrate at the compressed bitstream level. Both the ITU H.26x family of coders and the MPEG coders provide some scalability using a layered approach [1]. The layered scalable options include temporal, spatial and PSNR (quality) scalability. A layered coder provides coarse scalability where each enhancement layer can be added to the base layer for improved quality, temporal or spatial resolution. Layered coders can also be used as the framework for transmission over networks which support different levels of priority for QoS or for unequal error protection.

Each enhancement layer builds on the information in the previous base and enhancement layers adding more details to the reconstructed source at the decoder and increasing the total overall bitrate. Introducing scalability in a coder whose framework is based on temporal prediction results in compression loss. Each additional layer which is introduced into such a framework results in a greater gap in performance between the layered results and the optimum single layer result.

Fully embedded or progressive coders have also been proposed and codecs such as Shapiro's Embedded Zero Tree Wavelet (EZW) and Said and Pearlman's Set Partitioning in Hierarchical Trees (SPIHT) have been extremely successful for still image coding; yielding both superior compression and the ability to progressively decode the image. In this case, the wavelet decomposition and bit-plane encoding provides a natural framework for a scalable bitstream both in terms of quality (PSNR) and spatial resolution. Such a scheme produces a fully embedded bitstream which can be decoded to any given intermediate bitrate resulting in a reconstructed image of lower quality and/or resolution. Such fine granular scalability is desirable for varying network conditions and varying terminal capabilities in a streaming environment. For video applications, 3D embedded wavelet schemes provide good results as well but compression improvement comes at the expense of delay and additional memory requirements (longer temporal filters). Also, scaling back significantly in bitrate does not always produce satisfactory results due to the temporal smoothing that occurs systematically. However, a 3D-embedded wavelet scheme provides an elegant solution in terms of scalability and elim-

*This work was performed during a summer 2000 internship at Bell Labs.

0-7803-6725-1/01/$10.00 ©2001 IEEE

inates the need for explicit rate control. It also lends itself naturally to unequal error protection in a lossy environment where bitstream prioritization can be done on a bit exact level. It is useful to see how we can combine the compression efficiency of a motion compensated approach with the added network flexibility of an embedded approach [2].

Recently, MPEG-4 has adopted Fine Granular Scalability (FGS) for streaming applications [3]. The FGS framwork includes a base layer which is a traditional motion compensation/discrete cosine transform (MC/DCT) coder and an enhancement layer which is a finely scalable (embedded) layer. Several embedded coders were tested for the coding of the enhancement layer. A DCT-based bitplane coder was shown to yield similar results to a wavelet-based embedded coder such as the Embedded Zerotree Wavelet (EZW) algorithm and was chosen since the DCT is already an integral component of the standard. This hybrid approach as well as the previously proposed layered scalable options such as Annex O in H.263, introduce scalability at the expense of compression efficiency due to a framework that takes advantage of temporal correlation through motion compensation only at a coarse level at the base layer. However, the added flexibility of scalability is viewed to be an important component for network applications. The applications for FGS include multicasting where the original content is precompressed and stored on a server and can adapt to varying network conditions and end-user terminal capabilities on a bitstream level.

2. ALGORITHM DESCRIPTION

We investigate a general framework for a FGS type coder which combines a base-layer MC/DCT coder and a progressive enhancement layer where the information from the base-layer is used to determine the location of the high energy residual in the enhancement layer. Once the high energy residual is located, it is encoded using a wavelet-based scheme on a bitplane level for PSNR scalability.

The work here is motivated by a video coding algorithm based on dense motion field compensation [4]. In the previous work, the accurate motion field information and predicted frame information is used to determine the location of the high energy residual in the displaced frame difference (DFD). This information is available at the decoder so that the location information does not have to be explicitly coded. The original work was motivated by the observation that the encoding of the DFD frame is done using the same techniques which have been designed for coding still image data. However, the characteristics of the DFD frame, sparse data composed of high frequency components, is very different from the lowpass type of data typically found in still images and it may be beneficial to encode the residual data differently. Also, because dense motion field information is more accurate than traditional block-based motion vector information, fewer bits are available to encode the residual. The motion data and predicted frame data are used to predict where the significant DFD energy is located, namely motion boundaries and uncovered regions. A predicted frame is generated using the encoded motion field and the encoded previous frame. This process can be duplicated at the decoder. A Sobel edge detector is used to find the edges in the predicted frame which determines the location of the high energy DFD residual. The motion field discontinuities are also identified using a Canny edge detector. The intensity-based edges and motion-based discontinuities are used as a mask to identify the location of the high energy residual. The magnitude of the motion vectors are also used to expand the areas predicted to have high DFD energy due to uncovered regions which cannot be predicted accurately from the previous frame.

In the hybrid MC/progressive scheme introduced here, the energy distribution in the enhancement layer is similar to the energy distribution of the DFD so that the same approach can be used to locate the high energy data in the enhancement layer. The decoder has access to the base layer and can reproduce the steps needed to locate the high energy locations without any additional overhead. One drawback of this approach is that the coding performance saturates due to the prediction step which does not allow for better coding results once all the locations in the prediction mask are encoded. This is not a problem for many coding rates of interest that do not approach lossless quality. The algorithm could also be simply modified to encode the undetected locations once the detected areas are fully encoded.

The high energy signal is located using the reconstructed image and motion information available from the base layer. The identified regions are encoded using an embedded technique. Here we choose to encode the enhancement layer using the progressive wavelet coder (PWC) introduced by Malvar [5]. Other successful progressive coders such as EZW and SPIHT depend on reordering the data through data structures called zero trees in order to cluster the insignificant data (zeros) for efficient compression. Malvar introduces a less complex technique with similar coding performance that encodes the data in a data-independent way scanning the subbands from lowest to highest subband coefficient in a predetermined way, resulting in a similar clustering of zero values. This technique, like the previous wavelet schemes, does bitplane encoding resulting in an embedded scheme which scales in PSNR as well as spatial resolution.

3. RESULTS

Some preliminary results are shown in Figures [1-5]. Here we encode the "Mother Daughter" sequence and the "Hall" sequence at CIF resolution (360x240), 10 frames per sec-

ond. The base layer is encoded using a MC/DCT codec at 128 kbps. The enhancement layer is encoded using the PWC at 256 kbps for a total bitrate of 384 kbps. These rates are of interest for Third Generation (3G) Mobile applications. A Sobel edge detector is used to locate discontinuities in the reconstructed frame. These locations are used to predict the locations of the high energy signal in the enhancement layer to be encoded. The results of using the edge information to predict the location of the high energy residual is compared to a straightforward encoding of the entire enhancement layer. The PSNR curves are illustrated in Figure 1 for the "Mother Daughter" sequence. Here the lowest curve is the base-layer only result at 128 kbps using the ITU H.263 coder and 16x16 motion estimation. The two upper curves illustrate the base layer + enhancement layer at a total bitrate of 384 kbps. The solid curve shows the PSNR for the case where the base-layer edge information is used with PWC and the dotted curve is the result of using PWC without additional base-layer information. Figure 4 illustrates how well the edge information does at predicting the high energy locations in the enhancement layer. The gold areas correspond to correct detection, the red areas correspond to false positives (the edge information identifies areas in the enhancement layer that do not have significant energy), and the green areas correspond to false negatives, (significant signal energy that was not predicted by the base layer edge data). Figure 5 shows the corresponding coded frames from the "Mother Daughter" sequence at 384 kbps with the base layer + enhancement layer using base-layer side information. Figures 4 and 5 correspond to encoding the "Hall" sequence using base-layer edge information and the motion vectors generated from the base-layer encoding. The motion vector information is used to prune the edge-based locations so that we can effectively differentiate between boundaries of moving objects where occlusions and newly exposed areas can occur and static areas. The base layer is again encoded at 128 kbps with an enhancement layer of 256 kbps. Figure 4 illustrates the effectiveness in using base layer information to predict the location of the high energy information in the enhancement layer. The gold areas correspond to the high energy enhancement signal that was correctly identified by the base layer information. Blue corresponds to areas of the image that are pruned due to the motion vector information from the encoded base layer and red corresponds to correctly identified enhancement signal data that we were unable to encode due to insufficient bit budget. Figure 5 shows the reconstructed base and enhancement layer using the combined edge and MV approach.

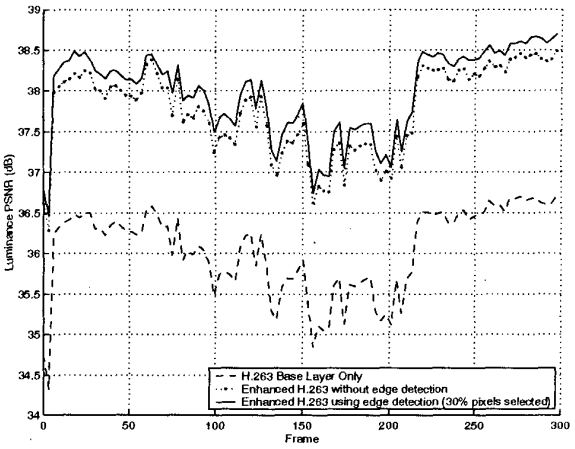

Fig. 1. PSNR curves for "Mother Daughter" Sequence; base layer results, enhancement layer using PWC, enhancement layer using high energy prediction and PWC

4. REFERENCES

[1] Barry G. Haskell, Atul Puri, and Arun N. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman & Hall, New York, New York, 1997.

[2] K. Shen and E.J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 109–122, Feb. 1999.

[3] ISO/IEC JTC1/SC29/WG11, *MPEG4 Video Verification Model 5.0*, Nov. 1996.

[4] S Han and C Podilchuk, "Modeling and coding of dfd using dense motion fields in video compression," in *Proc. IEEE Int. Conf. Image Processing*. 2000, Vancouver, Canada.

[5] Henrique S. Malvar, "Fast progressive wavelet coding," in *Proc. DCC'99, IEEE Data Compression Conference*. Mar. 1999, pp. 336–343, Snowbird, UT.

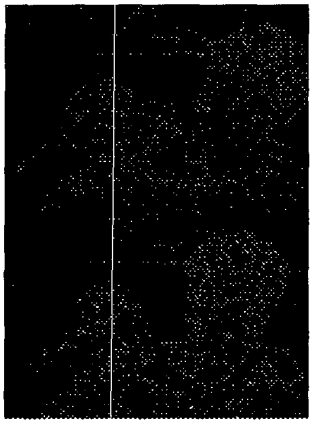

Fig. 2. Prediction of high energy data in enhancement layer; correct detection (yellow), false positives (red), false negatives (green)

Fig. 4. Prediction of high energy data in enhancement layer using edge data and MV data; correct detection (yellow), pruned areas due to MVs (blue)

Fig. 3. Encoded base layer + enhancement layer for the "Mother Daughter" sequence using edge-based high energy prediction

Fig. 5. Encoded base layer + enhancement layer for the "Hall" sequence using edge and mv data for prediction