MPEG-2 NONLINEAR TEMPORALLY SCALABLE CODING AND HYBRID QUANTIZATION

Sadik Bayrakeri   Russell M. Mersereau

Center for Signal and Image Processing
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
e-mail: sadik@eedsp.gatech.edu

ABSTRACT

In this paper, we investigate the MPEG-2 temporal scalability syntax and introduce a new approach to temporally scalable coding. Temporal scalability is provided by employing various nonlinear prediction and demultiplexing schemes. A nonlinear deinterlacing algorithm is presented and the related issues on interlaced, progressive and mixed mode video processing are addressed. In addition to the considered scalability techniques, a lookahead quantization scheme is presented for P- and B-type picture coding, which improves the coding performance by selective combination of the DCT domain scalar quantization and entropy-constrained vector quantization. Remarkable performance improvement over the simulcast coding is achieved.

1. INTRODUCTION

As a generic international standard, MPEG-2 video is intended to serve a wide range of digital video applications in the range of 2 to 15 Mbits/sec [1]. The standard is defined in terms of profiles and levels where each profile-level combination supports the features needed by an application of concern while limiting the decoder complexity.

Interlaced processing is one of the main additional features of MPEG-2. Each frame of interlaced video consists of two fields and field based temporal prediction modes are employed to find the best motion compensated prediction. Scalable coding is another new feature of the standard, the key property of which is to permit the division of a coded bit stream into two or more coded bitstreams representing the video at different resolutions. The scalable modes of MPEG-2, which are SNR, spatial, and temporal, correspond to multiple quality, multiple spatial resolution, and multiple temporal resolution video coding, respectively. Data partitioning is another type of scalability described in the standard for single resolution video coding, in which critical information is transmitted in a channel with better error performance.

In this paper, we analyze the principles of the MPEG-2 temporal scalability and develop a new approach to temporally scalable video coding. A nonlinear deinterlacing algorithm is presented to be employed as an additional prediction unit in temporally scalable coding. In addition to the considered scalability techniques, we also introduce a lookahead quantization scheme based on a selective combination of the DCT domain scalar quantization and the spatial domain entropy-constrained vector quantization (ECVQ). The next section describes the proposed nonlinear deinterlacing algorithm. A new approach to temporally scalable coding is developed in Section 3. In Section 4, the hybrid lookahead quantization scheme is presented. Performance evaluation of the proposed algorithms is given in Section 5. Conclusions follow as Section 6.

2. A NONLINEAR DEINTERLACING ALGORITHM

Interlaced scanning is a well known bandwidth reduction technique that has been extensively used in common television broadcasting systems. In each scan, only half the scan lines, as either the odd or the even field, are displayed instead of the complete frame. *Deinterlacing* is the process of converting an interlaced video signal into a progressive format by predicting the missing opposite parity fields.

An effective solution to the deinterlacing problem is to choose a prediction for each region of the missing field either from the temporal or the spatial information that yields the minimum error. Assume that we deinterlace an *odd field* and try to recover the the *missing even field.* The spatial information, in this particular problem, is the odd field and the temporal information is the temporal prediction of the missing even field. Depending on the application, one can search for the best possible method to form the temporal prediction of the missing field. One solution is to approximate the motion vectors of the missing field by projecting the motion vectors of the available neighboring fields onto the missing field. However, in temporally scalable video coding, we have the freedom of evaluating the temporal prediction of the missing field based on itself, as it is available in the enhancement layer.

Assume that a *temporal prediction* of the missing even field is available. The question is how to combine it with the odd field to find the best possible prediction of the missing even field. For that purpose, the proposed nonlinear deinterlacing algorithm is depicted in Figure 1.

This work was supported in part by the Joint Services Electronics Program under Contract DAAH-04-93-G-0027.

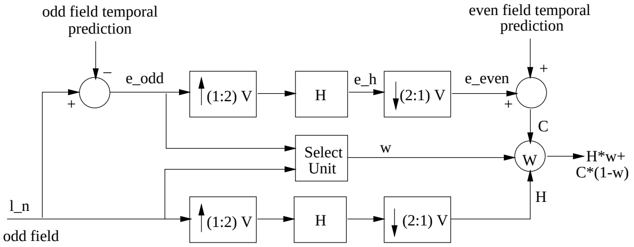

Figure 1: A spatio-temporal deinterlacing scheme

The algorithm can be described as follows: as the first step, a temporal prediction of the odd field is formed from the neighboring available fields using motion compensated prediction. The temporal prediction error of the odd field, which is shown by *e_odd*, is evaluated by subtracting the odd field temporal prediction from the odd field. This step measures the accuracy of temporal prediction. The error image *e_odd* is interpolated and subsampled vertically by keeping the even lines, and then added back to the even field temporal prediction. The corrected even field temporal prediction is labeled *C*. By this step, the algorithm detects the inaccurately motion estimation regions in the odd field temporal prediction, then uses the same criteria for the even field temporal prediction to correct the even field temporal prediction. The lower resolution image, *l_n*, is also vertically interpolated and subsampled by keeping the even field.

A further step in the algorithm is taken by the weighted combination of *H* and *C*. The reason for this final step is to evaluate the variation of *e_odd* and *l_n* for the same regions and to assign to each a relative interpolation accuracy weight. In this way, we can eliminate the signal that has higher chance of resulting in wrong estimates. The box labeled as *selection unit* is designed to fulfill this requirement. It is a very simple unit that evaluates the variances of *e_odd* and *l_n* for a given region. For each higher resolution pixel to be interpolated, the variances of *e_odd* and *l_n* are measured over a 4x3 window. The spatial (*w*) and temporal (*1-w*) weights are evaluated as the normalized inverse variances. The adaptive algorithm, without any side information transmitted, achieves extremely accurate prediction and minimizes common prediction artifacts such as aliasing, blocking effects, and occlusion. A related video interpolation algorithm for progressive-to-progressive spatial scalability can be found in [2].

3. A NEW APPROACH TO TEMPORALLY SCALABLE CODING

Temporal scalability is a tool intended for use in a range of diverse applications from telecommunications to HDTV for which migration to higher temporal resolution is necessary. The video input sequence at full temporal rate is *temporally demultiplexed* to form two video sequences as base and enhancement layers. The base layer is coded independently. To encode the enhancement layer, the MPEG-2 temporal scalability syntax employs predictions of the enhancement layer from either base or enhancement layer reproduced pictures. In the enhancement layer, a P-type picture forms its prediction from either the decoded base layer, or from a previously reproduced enhancement layer picture. For B-type pictures, prediction is formed in one of two ways: One way is to form two predictions from the base layer. Another alternative is to form one prediction from each of the base and enhancement layers. Within the enhancement layer, it is prohibited to use backward prediction. This restriction is imposed to avoid the need for frame reordering and to reduce complexity. More details can be found in [1, 3].

This structure is applicable to a number of base and enhancement layer picture processing formats. In particular, progressive input: progressive-to-progressive, progressive input: interlace-to-interlace, and interlaced input: interlace-to-interlace temporal scalability forms are supported in the MPEG-2 standard. The *progressive: progressive-to-progressive* temporal scalability refers to the coding scheme in which a progressive input sequence is temporally demultiplexed into two progressive sequences. The base layer is formed by the odd frames and the enhancement layer is formed by the even frames, or vice versa. In *interlace: interlace-to-interlace* temporal scalability, the interlaced input frames are temporally demultiplexed into two interlaced sequences. Demultiplexing is performed similar to the progressive-to-progressive case where odd interlaced frames form the base layer and even interlaced frames form the enhancement layer. In both temporal scalability cases, the enhancement layers can be predicted as described in the above paragraph.

Progressive: interlace-to-interlace temporal scalability is one of the possible future applications of temporal scalability. A representative example is the progressive HDTV broadcast along with the conventional interlaced TV. The process of temporal demultiplexing involves progressive input to two interlace format sequence conversion. The interlaced TV sequence and a complementary interlaced sequence are extracted from the progressive HDTV input. The demultiplexing operation is illustrated in Figure 2.

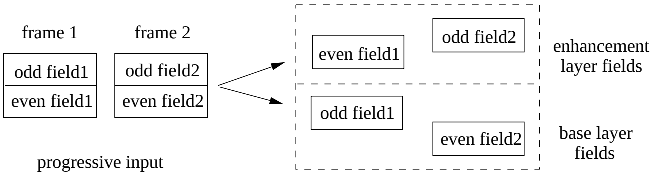

Figure 2: Progressive: Interlace-to-Interlace Demultiplexer

The base layer interlaced fields are formed from the odd field of frame 1 and the even field of frame 2. The complementary fields, even field1 and odd field2, are used to form the enhancement layer. The base layer is coded independently by an MPEG-2 encoder. The enhancement layer complementary fields are spatio-temporally predicted using the reproduced base layer opposite parity fields and the temporal prediction formed within the enhancement layer. The enhancement layer prediction unit is shown in Figure 3.

For each complementary layer frame to be predicted, the base layer odd and even interlaced fields are deinterlaced into two progressive frames and the opposite parity field is

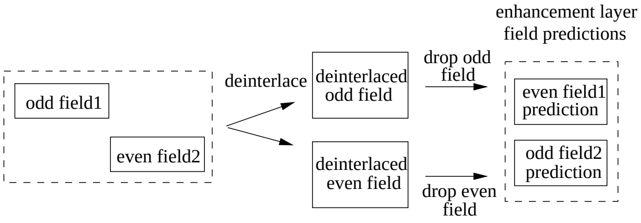

Figure 3: Progressive: Interlace-to-Interlace Prediction

extracted to form the enhancement layer field predictions. Deinterlacing of the base layer fields are performed by the proposed deinterlacing algorithm. The proposed deinterlacing algorithm employs the base layer temporal prediction to measure the local accuracy of the enhancement layer temporal prediction. While it is possible to use the actual base layer field temporal predictions, which are available at the decoder site, instead we approximate them by interpolating, and subsampling by keeping the opposite parity field, of the enhancement layer field temporal predictions. This process is preferred to reduce the decoder storage complexity.

The proposed nonlinear deinterlacing algorithm function depends partially on the degree of distortion introduced into the decoded base layer fields. The problem is solved by evaluating the prediction error of the proposed deinterlacing algorithm along with the spatial and temporal predictions on a macroblock-by-macroblock basis. More precisely, for each macroblock, we consider the deinterlacing prediction (V), spatial prediction (H), temporal prediction (P), (V+P)/2, (V+H)/2, (P+H)/2, and (V+P+H)/3, as seven candidate prediction modes. For each macroblock, four luminance block prediction errors are evaluated for each mode and one of the seven methods, which gives the least error, is chosen. For each macroblock, three bits are transmitted to the decoder as the additional information.

4. A LOOKAHEAD HYBRID QUANTIZER DESIGN

The standard MPEG-2 quantization scheme, in principle, is a DCT-based transform coder with scalar quantization, which is close to optimum only if the signal is highly correlated. While scalar quantization is chosen for its simplicity, it is possible to achieve more compression for prediction error signals by using block coding. Among the various block-coding methods, ECVQ is a common choice that jointly minimizes the codebook rate and distortion. One approach is to employ ECVQ for P- and B-type pictures. While it is possible to apply the same method in DCT domain, spatial domain implementation of ECVQ is chosen to reduce the number of codebooks. Direct application of ECVQ requires large codebooks to achieve performance improvement over the DCT-Scalar quantization scheme. On the other hand, a *lookahead* design method that employs both quantization schemes can achieve the best performance with the minimum codebook sizes. The proposed lookahead method can be explained as follows: for each 8x8 block, scalar quantization is applied to the DCT coefficients with a step size *mquant* defined by the MPEG-2 coding scheme. The resultant scalar quantization error is calculated. Then, the same block is divided into four 4x4 vectors in spatial domain and the best match of each vector is found from an operating codebook. The quantization error corresponding to each vector is calculated and added to find the total ECVQ error for the considered 8x8 block. The 8x8 block DCT-Scalar quantization error and the total ECVQ error is compared and the method that gives the least error is chosen for that block. To send the codeword indices, four arithmetic coders [4], corresponding to each codebook, are run during the coding interval. The additional cost of the hybrid lookahead quantization method is one bit for an 8x8 block of pixels. Considering YUV color format video coding, for P-type pictures, one codebook is designed for Y component and another codebook is designed for both U and V components. The same process is repeated for B-type pictures. The rate allocation of each codebook is performed as follows: after an overall coding bit-rate (excluding the headers, motion vectors, ...) is defined, it is distributed between I-. P-, and B-type pictures, based on the MPEG-2 global complexity measures. The relative rates of I-, P-, and B-type pictures are defined as $\{4, 1.5, 1\}$, respectively. A fixed rate-ratio is assigned between Y and UV color components. Each of the four codebooks are then generated with the ECVQ algorithm [5]. Performance improvement over the scalar quantization is guaranteed due to the lookahead mode selection scheme.

5. PERFORMANCE EVALUATION

In this section, the performance of the proposed scalable scheme is evaluated using two SIF size (352x240 at 60 Hz) video sequences, FOOTBALL, and CYCLE GIRL. Simulations are performed by using a software implementation [6] of the MPEG-2 standard.

Before we proceed with temporally scalable coding results, the performance of the proposed deinterlacing technique is depicted in Figure 4 for a P-type FOOTBALL field. The energy of the scalable prediction error image is considerably reduced compared to the simulcast technique. The common temporal prediction artifacts such as blocking and uncovered background are also minimized to a great extend.

Temporally scalable coding performance is evaluated based on the average PSNR values of 96 fields. The average PSNR values of the scalable, simulcast, and single layer coding schemes are presented in Table 1 for the two video sequences. For each sequence, the base layer (352x120 at 60 Hz) is coded at 1.6 Mbits/sec and the enhancement layer (352x120 at 60 Hz) is coded at 1 Mbits/sec. The progressive input sequence (352x240 at 60 Hz) is coded as single layer at 2.6 Mbits/sec, which is the total rate of the base and enhancement layers. For the enhancement layer hybrid coding, four ECVQ codebooks are generated for P- and B-type picture coding. Y space codebook sizes are chosen as 512 and UV spaces codebook sizes are chosen as 256. During the coding process, each codebook is run at the rate it is designed. However, better performance improvement is expected by employing a hybrid bit-allocation algorithm that incorporates the ECVQ system into the MPEG-2 bit-allocation scheme. Based on the results depicted in Table 1,

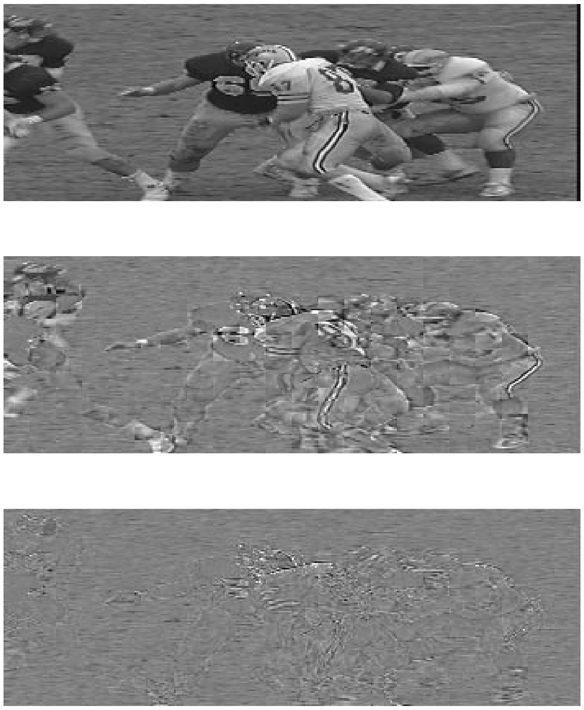

Figure 4: Prediction error images of a P-type FOOTBALL picture: (top) the original luminance field, (middle) simulcast prediction error, (bottom) scalable prediction error.

the scalable scheme achieves 2-3 db higher PSNR than the simulcast technique. It is observed that, for each video sequence, nearly more than half of the blocks are predicted by using the proposed deinterlacing algorithm and its related modes, i.e the sum of the percentages of the selected prediction schemes V, (V+P)/2, (V+H)/2, (V+P+H)/3. The scalable coding performance is also improved by the hybrid quantization scheme. Our second observation is the performance decrease in interlaced sequence coding. The interlaced base layer coding performance is lower than the single layer progressive coding performance, although the progressive single layer is coded at a lower relative rate. The single layer (352x240 at 60 Hz) coding at 2.6 Mbits/sec bit-rate corresponds to 1.3 Mbits/sec at the base layer dimensions (352x120 at 60 Hz). The main reason for this difference is the temporal prediction. At the same temporal rate, motion estimation is harder in interlaced coding than progressive coding because of the missing fields. Therefore, even with 2-3 db PSNR improvement in the enhancement layer, scalable coding performance is lower than that of single layer progressive coding at the same bit-rate. The performance decrease in interlace coding temporal prediction effects both the base and the enhancement layers.

Cycle Girl	**Y**	**U**	**V**	**Rate**
Base	29.14	33.74	36.58	1.6 Mbits/sec
Simulcast	26.13	32.76	35.96	1 Mbits/sec
Scalable	28.68	33.81	36.68	1 Mbits/sec
Scal-Hybrid	29.32	34.12	36.89	1 Mbits/sec
Single Layer	29.76	35.05	37.97	2.6 Mbits/sec

Football	**Y**	**U**	**V**	**Rate**
Base	32.63	34.96	37.06	1.6 Mbits/sec
Simulcast	29.21	32.78	35.58	1 Mbits/sec
Scalable	32.22	34.84	37.07	1 Mbits/sec
Scal-Hybrid	32.68	35.16	37.24	1 Mbits/sec
Single Layer	34.24	37.14	38.89	2.6 Mbits/sec

Table 1: PSNR (db) comparison of simulcast, scalable, and single layer coding

6. CONCLUSIONS AND COMMENTS

The principles of temporally scalable video coding is analyzed. A novel spatio-temporal deinterlacing algorithm is presented. The deinterlacing algorithm can also be used for other purposes such as interlace-to-progressive scan conversion. Based on the deinterlacing algorithm, a new approach to temporally scalable video coding is developed. The performance improvement of the new temporally scalable scheme over the simulcast technique is shown both visually and experimentally. While the scalable coding performance is improved by the hybrid lookahead quantization scheme, better results can be achieved by employing a hybrid bit-allocation algorithm that incorporates the ECVQ system into the MPEG-2 bit-allocation scheme. It is also observed that interlaced video coding shows lower performance than progressive coding at the same temporal rate. This is reflected on the base layer and the enhancement layer coding performances, including the scalable coding case. Further research on field motion estimation algorithms is required to improve the interlaced coding performance.

7. REFERENCES

[1] ISO/IEC 13818-2 Recommendation H.262, ISO/IEC JTC1/SC29 WG11/602, Seoul, *MPEG-2 committee draft*, Nov. 1993.

[2] S. Bayrakeri and R. M. Mersereau, "MPEG-2/ECVQ lookahead quantization and spatially scalable coding." *to be presented in VCIP'97.*

[3] A. Puri, L. Yan, and B. G. Haskell, "Temporal resolution scalable video coding," in *IEEE Intl. Conf. Image Processing*, vol. 2, pp. 947–951, 1994.

[4] I. H. Witten, R. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, pp. 520–540, June 1987.

[5] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 31–42, Jan. 1989.

[6] MPEG Software Simulation Group, *MPEG-2 Video Codec 1.1*, June 1994.