

14-61-CP
027418

A SUBBAND CODING METHOD FOR HDTV

Wilson Chung, Faouzi Kossentini, and Mark J. T. Smith
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332

This paper introduces a new HDTV coder based on motion compensation, subband coding, and high order conditional entropy coding. The proposed coder exploits the temporal and spatial statistical dependencies inherent in the HDTV signal by using intra- and inter-subband conditioning for coding both the motion coordinates and the residual signal. The new framework provides an easy way to control the system complexity and performance, and inherently supports multiresolution transmission. Experimental results show that the coder outperforms MPEG-2, while still maintaining relatively low complexity.

I. INTRODUCTION

Several methods have been proposed recently for transmission of HDTV [1, 2, 3, 4, 5, 6, 7]. Most employ motion compensation at one stage or another, after which the residual between the original and predicted frames is computed and encoded spatially. DCT-based spatial coders are widely used, most notably in the MPEG standards. However, subband coders are also becoming popular.

There are many important issues that are associated with HDTV coding, such as control over the bit rate and picture quality, error correction and concealment, and multiresolution capability for multisource decoding and progressive transmission applications. In this paper, we introduce a new subband video coder which achieves good performance with low relative complexity, but also provides a framework where most of these issues can be easily addressed. The proposed coder employs motion estimation and compensation independently for each subband, but encodes the motion vectors using a high order conditional entropy coding scheme that exploits statistical dependencies between motion vectors of the same frame and successive frames as well as between the coordinates of the motion vectors, simultaneously. The coder also identifies non-compensatable blocks through the use of statistically optimized thresholding, which are then intra-frame coded. The video coder is described next. This is followed by a discussion of practical design issues. Section 4 presents

experimental results which compare the performance and complexity of the coder with that of MPEG-2.

II. THE VIDEO CODER

First, consider a conventional subband video coder. In the parlance of MPEG, the frames that are coded spatially are called I frames. Those that are forward-predicted are called P frames, and those that are forward- and backward-predicted are called B frames. The sequence of video frames is first grouped into blocks of N frames, where the first frame (or I frame) is coded using an intra-frame subband coder, and the other $N - 1$ frames (or P frames) are predicted using motion estimation and compensation, and the residual frames are coded using another subband coder. In this work, no B frames are used. At the receiver, each video frame is constructed from motion information (if applicable) and the coded residual frame.

There are two important problems associated with the above coder. First, motion compensation using the block matching algorithm with a typical block size of 16×16 and search range of -16 to $+16$ in each dimension is usually computationally intensive. This problem becomes even worse in HDTV coding because both block sizes and search areas have to be somewhat larger to achieve good performance. Second, due to the block matching algorithm, blockiness frequently appears in the residual frame, which introduces *artificial* high frequencies. To solve these two problems, we apply the block matching algorithm to each of the subbands. Figure 1 shows a block diagram of the proposed subband coder and Figure 2 shows the structure of the RVQ coder. Each frame is first decomposed into subbands using a tree-structured IIR analysis filter bank. The filter bank is based on two-band decompositions, which employ allpass polyphase separable IIR filters [8]. A full-search block matching algorithm (BMA) using the mean absolute distance (MAD) is used to estimate the motion vectors. Since the BMA does not necessarily produce the true motion vectors, we employ a thresholding technique for improving the rate-distortion performance. Let d_{min} be the minimum MAD associated with a block to be coded. Also, let T be a threshold, which is a large positive number empirically determined from the statis-

This work was supported by the National Science Foundation under contract MIP-9116113 and the National Aeronautics and Space Administration.

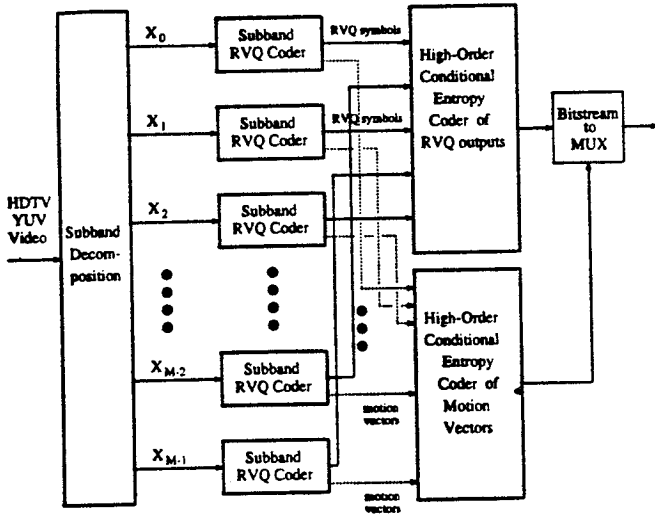


Figure 1: Proposed subband video coder.

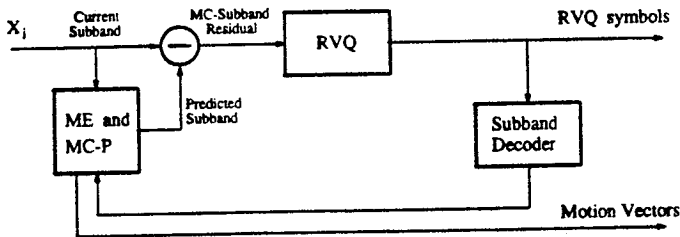


Figure 2: Basic structure of the RVQ coder.

tics of the subband being coded. If $d_{min} > T$, then the block is likely not compensatable. Thus, both the original block and the residual block, obtained by subtracting the motion compensated predicted block from the original one, are coded using the intra-band and residual coders, respectively, and the one leading to better rate-distortion performance is chosen (as will be described shortly). A special symbol, which can be coded as part of the motion information, is sent to the decoder indicating the type of coding used.

In many conventional HDTV subband coders as well as in MPEG, differential entropy coding of motion vectors is employed. Since motion vectors are usually slowly varying, the motion bit rate can be further reduced by exploiting dependencies not only between previous motion vectors within and across the subbands but also between the vector coordinates. For this purpose, we employ a high order conditional entropy coder that is based on finite state machine (FSM) modeling. More specifically, let $(X_{n,m}, Y_{n,m})$ be the pair of random variables representing the current horizontal and vertical motion displacements in the current subband (n, m) in frame n . Also, let $(U_{n,m}, V_{n,m})$ be the pair of state random variables with realizations $u_{n,m} = \{0, 1, \dots, S_{n,m}^u\}$ and $v_{n,m} = \{0, 1, \dots, S_{n,m}^v\}$, which we associate with $X_{n,m}$ and $Y_{n,m}$,

respectively. Each state $u_{n,m}$ is given by

$$u_{n,m} = F_{n,m}(s_{n,m}^0, s_{n,m}^1, \dots, s_{n,m}^{M_{n,m}}),$$

and each state $v_{n,m}$ is given by

$$v_{n,m} = G_{n,m}(t_{n,m}^0, t_{n,m}^1, \dots, t_{n,m}^{N_{n,m}}),$$

where $s_{n,m}^0, s_{n,m}^1, \dots, s_{n,m}^{M_{n,m}}$ and $t_{n,m}^0, t_{n,m}^1, \dots, t_{n,m}^{N_{n,m}}$ are previously coded conditioning symbols. The mappings $F_{n,m}$ and $G_{n,m}$ are generally many-to-one mappings that convert combinations of realizations of the conditioning symbols to a particular state. Assuming that $X_{n,m}$ is entropy coded first, the conditioning symbols for the FSM model associated with $X_{n,m}$ are selected from a region composed of symbols located in all previously coded subbands (i.e., where motion vectors were already coded) in both frames n and $n-1$. When $Y_{n,m}$ is being coded, the horizontal displacements in the same subband can also be included in the conditioning region.

Statistical modeling for entropy coding the motion vectors consists of first selecting, for each subband (n, m) , $M_{n,m}$ ($N_{n,m}$) conditioning symbols for $X_{n,m}$ ($Y_{n,m}$), and then finding mappings $F_{n,m}$ and $G_{n,m}$ such that the conditional entropies $H(X_{n,m}|U_{n,m})$ and $H(Y_{n,m}|V_{n,m})$ are minimized subject to a limit on complexity. The total number of probabilities that must be computed and stored is used here as a measure of complexity. The tree-based algorithms described in [9] are used to find the best values of $M_{n,m}$ and $N_{n,m}$ subject to a limit C_1 on the total number of probabilities. The PNN algorithm [10], in conjunction with the generalized BFOS algorithm [11], is then used to construct mapping tables that represent $F_{n,m}$ and $G_{n,m}$ subject to another limit C_2 ($C_2 \ll C_1$) on the number of probabilities.

The intra-band (I-subband) and residual (P-subband) coders are multistage residual vectors quantizers (RVQs) followed with high order conditional statistical models, which are optimized to the intra-band and residual band statistics, respectively. Multistage RVQs provide an easy way to control the complexity-performance tradeoffs, and allow efficient high order statistical modeling. We restrict the number of code vectors per stage to be 2, which simplifies both statistical modeling and entropy coding used in this work. This also provides the highest resolution in a progressive transmission environment.

The same statistical modeling algorithm used for entropy coding the motion vectors is also used for entropy coding of the output of the RVQs. Both the motion vectors and the output of the RVQs are eventually coded using adaptive binary arithmetic coders (BACs) [12, 13]. These coders are very easy to adapt and require small complexity.

III. PRACTICAL DESIGN ISSUES

To achieve the lowest bit rate, the statistical models used to entropy code the motion vectors should be generated on-line. However, this requires a two-pass process where statistics are generated in the first pass, and the statistical modeling algorithm described above is used to generate the conditional probabilities. These probabilities must then be sent to the BAC decoders so that they can track the corresponding encoders. In most cases, this requires a large complexity. Moreover, even by restricting the number of states to be relatively small (such as 8), the side information can be excessive. Therefore, we choose to initialize the encoder with a generic statistical model, which we generate using a training HDTV sequence, and then employ dynamic adaptation [12] to track the local statistics of the motion flow.

For both the I-subbands and P-subbands, the multi-stage RVQs and associated statistical models are designed jointly using an entropy and complexity-constrained algorithm, which is described in [9, 14]. The design algorithm iteratively minimizes the expected distortion $E\{d(\mathbf{X}, \hat{\mathbf{X}})\}$ subject to a constraint on the overall entropy of the statistical models. The algorithm is based on a Lagrangian minimization and employs a Lagrangian parameter λ to control the rate-distortion tradeoffs. To substantially reduce the complexity of the design algorithm, only separate subband encoders and decoders are used. However, the RVQ stage encoders in each subband are jointly optimized through dynamic M -search, the decoders are jointly optimized using the Gauss-Seidel algorithm.

The most important part of the design algorithm is the encoding procedure, where either an intra-frame or inter-frame subband coder must be chosen for a particular block. Suppose we want to encode a block $B_{n,m}^i$ of size $L_{n,m}$ using the proposed I-subband and P-subband coders with Lagrangian parameters (or quality factors) λ_I and λ_P , respectively. The BMA algorithm is first applied, and the minimum MAD d_{min} is computed. If $d_{min} < T$, then the corresponding motion vector is encoded using the BAC specified by the current state, and the residual block is quantized using the P-subband (residual) RVQ. The output of each RVQ stage is encoded with a separate entropy coder composed of a FSM statistical model and a set of BACs, each specified uniquely by a state. If $d_{min} \geq T$, then the block is both I-subband and P-subband coded. Let $R_x = -\log_2 p(x^i|u^i)$ and $R_y = -\log_2 p(y^i|v^i)$ be estimates of the number of bits required to code the horizontal and vertical coordinates of the motion vector, respectively. Also, let d_P be the distortion and R_P be the rate that compose the minimum Lagrangian $J_P = d_P + \lambda_P R_P$ associated with coding the residual block. Assuming that $J_I = d_I + \lambda_I R_I$ is the minimum Lagrangian associated with coding the original block, then the I-subband coding



Figure 3: The 114th frame of the sequence BRITS

method is selected if

$$J_I < d_P + \lambda_P (R_x + R_y + R_P).$$

The proposed coder has many practical advantages, due to both the subband structure and the multistage structure of RVQ. For example, multiresolution transmission can be easily implemented in such a framework. Another example is error correction, where the more probable of the two stage code vectors is selected if an uncorrectable error is detected. Since each stage code vector represents only a small part of the coded vector, this should not significantly affect the reconstruction or the FSM statistical models.

IV. EXPERIMENTAL RESULTS

The image shown in Figure 3 is frame number 114 of the test sequence BRITS, which we encode using both the proposed coder and MPEG-2. The frame size is 720×1280 . The original RGB color sequence with 8 bits/pixel requires approx 1.3 Gbs. The MPEG-2 software we used resides on <ftp.netcom.com/pub/cfogg/mpeg2> [15].

In our experiments, each frame is decomposed into 64 uniform subbands, but more than half of the subbands are not coded. This is determined based on initial rate-distortion tradeoffs [9]. The BMA algorithm used in our experiments employs a block size of 2×2 and a search area of -2 to $+2$ in each dimension. Motion estimation is performed, and is done only for the Y luminance component and the estimated motion vector field is subsequently used for the motion compensation of U and V chrominance signals. A high order conditional entropy coder is designed for the motion vector coordinates, and one I-subband coder and one P-subband coder with vector size of 2×2 are designed for the each of the YUV components. We set the maximum allowed numbers of conditional probabilities for the motion entropy coder and the I-subband and P-subband entropy coders to $C_1 = 4094$ and $C_2 = 512$. The BACs used employ a skew factor between 1 and 256.

REFERENCES

- [1] J. G. Apostolopoulos and J. S. Lim, "Video compression for digital advanced television systems," in *Motion Analysis and Image Sequence Processing*, Boston: Kluwer Academic Publishers, 1992.
- [2] S. M. Lei, T. C. Chen, and K. H. Tzou, "Subband HDTV coding using high-order conditional statistics," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 65-76, Jan. 1993.
- [3] J. Bellisio and K. Tzou, "HDTV and the emerging broadband isdn network," *SPIE Proc. Visual Communications and Image Processing*, vol. 1001, pp. 772-786, 1988.
- [4] F. Bosveld, R. Lagendijk, and J. Biemond, "Hierarchical coding of HDTV," *Signal Processing: Image Communication*, vol. 4, pp. 195-225, July 1992.
- [5] F. Bosveld, R. Lagendijk, and J. Biemond, "Compatible HDTV transmission using conditional entropy coding," in *Proc. ICASSP*, vol. V, (Minneapolis, MN, USA), Apr. 1993.
- [6] I. Furukawa, M. Nomura, N. Otha, and S. Ono, "Hierarchical coding of super high definition images with adaptive block-size multi-stage VQ," *Signal Processing of HDTV*, vol. III, 1992.
- [7] S. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV," *IEEE Trans. on CVST* vol. 1, pp. 100-112, March 1991.
- [8] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on ASSP.*, vol. 38, pp. 1446-1456, Aug. 1991.
- [9] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on IP.*, July 1994.
- [10] W. H. Equitz, "A new vector quantization clustering algorithm," *IEEE Trans. ASSP.*, pp. 1568-1575, October 1989.
- [11] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on IT*, vol. 37, pp. 400-422, Mar. 1991.
- [12] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Trans. on COM*, vol. 29, no. 6, pp. 858-867, 1981.
- [13] W. B. Pennebaker, J. L. Mitchell, G. G. Langdon, and R. B. Arps, "An overview of the basic principles of the Q-coder adaptive binary arithmetic coder," *IBM J. Res. Dev.*, vol. 32, pp. 717-726, Nov. 1988.
- [14] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," *Abstract Submitted to ICASSP*, (Detroit, MI, USA), Apr. 1995.
- [15] MPEG Software Simulation Group, "MPEG-2 encoder / decoder, version 1.0," *ISO/IEC DIS 13818-2 codec*, May 1995.

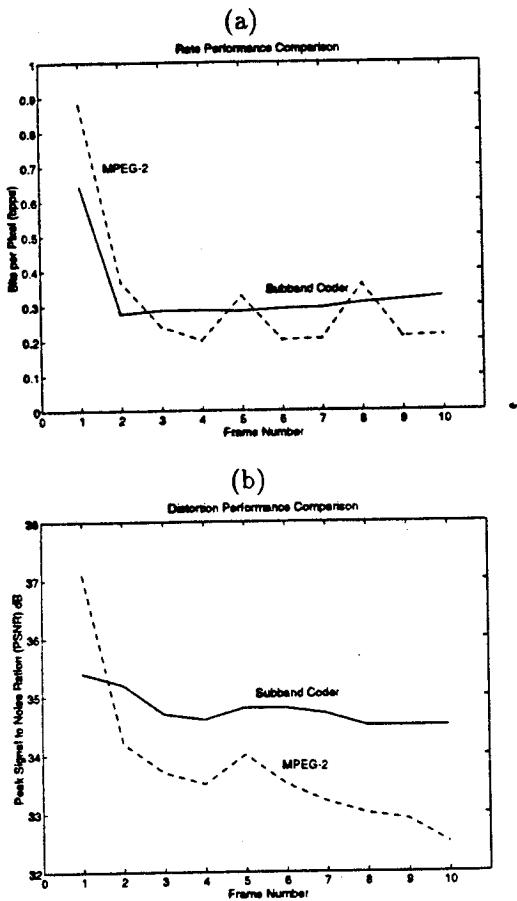


Figure 4: (a) Overall rate usage and (b) PSNR performance for the proposed coder.

For each rate-distortion point, the total memory required to store both the I-subband and P-subband RVQ codebooks and associated tables of conditional entropy codes is approximately 4.6 kilobytes. Moreover, only 512 bytes are required by the motion entropy coder. For analysis, quantization using dynamic M -search, and BAC encoding, approximately 27 multiplies and 32 adds per pixel are required. Only 3 multiplies and 14 adds are required for BAC decoding, inverse quantization, and synthesis. Not only are the encoding complexity and memory relatively small, but the performance is also good. Figure 4 (a) shows the average bit per pixel and Figure 4 (b) shows the PSNR result of our coder in comparison with the MPEG-2 standard for 10 frames of the luminance component of the color test video sequence BRITS. The average bit rate is approximately 18.0 Mbits/sec and the average PSNR is 34.75 dB for the proposed subband coder and 33.70 dB for MPEG-2. As is shown in the figure, the proposed coder clearly outperforms MPEG-2. Moreover, although MPEG-2 requires less encoding complexity and memory, the complexity of our subband coder are still reasonable.