MODE SELECTION AND OPTIMAL RATE CONTROL FOR VIDEO CODING USING AN AND-OR TREE REPRESENTATION

Tsung-Hang John Lee and Wen-Liang Hwang

Institute of Information Science, Academia Sinica, Taiwan

ABSTRACT

We propose an AND-OR tree-based approach for structuring a variety of mode selections in a hybrid video coding system. The proposed approach can systematically analyze the input sequence and transform it into an AND-OR tree representation to allocate bits to each node of the tree. The motion vector of our system is estimated in the wavelet domain, where wavelet coefficients in different scales can be obtained by using different wavelets. We demonstrate the performance of the new features, and compare our codec's performance with that of an H.264/AVC-like implementation.

Index Terms— Wavelet, Video coding, Rate-distortion optimization, Mode selection

1. INTRODUCTION

The recently proposed H.264/AVC video coding standard is one of the most important developments in video coding. The bitstream produced by H.264/AVC achieves a significant improvement in compression efficiency compared to previous coding standards, such as MPEG-2, H.263, and MEPG-4. Many features contribute to the success of H.264/AVC, for example, the adoption of various prediction modes [8], and the improved rate-constrained control achieved by using a Lagrangian bit-allocation technique [7, 4].

Since using more prediction modes increases the number of coding options, the prediction accuracy is increased at the cost of increasing the complexity of obtaining optimal coding parameters. Nevertheless, the side information required to transfer the chosen modes is also increased. The coding options in H.264/AVC can be summarized as two relations: an OR relation and an AND relation. These AND and OR relations can be structured to form an AND-OR tree [1]. The AND-OR tree perspective allows us to employ a systematic approach to manage the variety of coding option selections and the sophisticated interactions between the temporal and spatial dependencies among blocks.

In this paper, we propose a number of new approaches. We use an AND-OR tree to structure various modes of a P-frame, incorporate residual coding into the motion vector selection process, and present an optimal bit-allocation algorithm for an AND-OR tree. The new approaches are demonstrated by a hybrid video coding system in which variable block size motion estimation is performed in the wavelet domain. Our system has two exclusive features compared to a DCT-based system: (1) our codec can explore the effects of using more than one wavelet filter in motion estimation and residual coding; and (2) the wavelet coefficients of a block of any size can be obtained efficiently by re-arranging the coefficients of the dyadic wavelet transform (un-decimated wavelet transform) of a frame [3]. We perform experiments on various QCIF sequences in our codec and demonstrate that the performance improvement is 0.5-1 dB.

The remainder of the paper is organized as follows. Section 2 presents variable block size motion estimation methods in the wavelet domain. Section 3 describes the methods used to obtain the optimum R-D curve in any node of an AND-OR tree. Section 4 provides an AND-OR tree-based rate control algorithm that can obtain the optimal graph solution. Section 5 discusses the implementation parameters and compares the performance of the proposed approach with that of other codecs. Finally, in Section 6, we present our conclusions.

2. WAVELET-BASED VARIABLE BLOCK SIZE MOTION ESTIMATION

Motion estimation in the wavelet domain is a new method that uses different wavelet filters in different scales to obtain wavelet coefficients. For example, the wavelet coefficients in one scale can be obtained by using the Harr wavelet, while the bi-orthogonal 5-3 wavelet can be used in another scale. The proposed system raises the following issue: How can we relate motion estimation error in the wavelet domain to that in the image?

Because motion estimation is performed in the wavelet domain, the estimation errors between the wavelet domain and the image must be related. This issue has been studied in [6, 5]. Let us assume that a wavelet signal is divided into N blocks, and let $e_{ij,k}$ denote the residual error of the k-th block in subband ij. The residual error in subband ij is the sum of the residual errors in each block, i.e., $\sigma_{i,j}^2 = \sum_{k=1}^{N} e_{ij,k}^2$. The minimum square error (MSE) can now be written as

$$\sigma_e^2 = \sum_{k=1}^{N} (w_{10}\sigma_{10,k}^2 + w_{11}\sigma_{11,k}^2 + w_{01}\sigma_{01,k}^2). \qquad (1)$$

where w_{ij} is derived from the the reconstruction filters that transforms the coefficients in subband j and level i (called subband ij for short) to the image domain. Let MV_k denote the motion vector of the k-th block, and **MV** be the motion vectors of all the blocks. We make the dependence of motion vector selection in Eq. 1 explicit as follows:

$$\begin{aligned}\sigma_e^2(\mathbf{MV}) &= \sum_{k=1}^{N}(w_{10}e_{10,k}^2(\mathbf{MV}) + w_{11}e_{11,k}^2(\mathbf{MV}) \\ &\quad + w_{01}e_{01,k}^2(\mathbf{MV})). \end{aligned} \tag{2}$$

According to the above equation, to minimize the overall distortion in the image domain, each subband must be multiplied by a weight before the spatial dependency between all the motion vectors in the blocks can be explored.

Estimating spatially dependent motion vectors is a complex procedure [9, 2]. In practice, to reduce complexity, the motion prediction and residual coding are performed independently for each block. This corresponds to finding a suboptimal solution for Eq. 2 by the following simplifications:

$$\begin{aligned}\min_{\mathbf{MV}} \sigma_e^2(\mathbf{MV}) &\leq \min_{MV_k} \sum_{k=1}^{N}(w_{10}e_{10,k}^2(MV_k) \\ &\quad + w_{11}e_{11,k}^2(MV_k) + w_{01}e_{01,k}^2(MV_k)) \\ &= \sum_{k=1}^{N} \min_{MV_k}(w_{10}e_{10,k}^2(MV_k) \\ &\quad + w_{11}e_{11,k}^2(MV_k) + w_{01}e_{01,k}^2(MV_k)). \end{aligned} \tag{3}$$

The above derivations can be extended to M-level decomposition as follows:

$$\begin{aligned}\sigma_e^2(\mathbf{MV}) &= \sum_{j=1}^{N}(\sum_{i=0}^{M-1} w_{i1}e_{i1,j}^2(\mathbf{MV}) \\ &\quad + w_{(M-1)0}e_{(M-1)0,j}^2(\mathbf{MV})), \end{aligned}$$

and also to 2-D images without much effort.

3. R-D CURVE OPTIMIZATION

Because residual coding is not used in the Lagrangian cost function, the method does not produce the optimal coding performance for all bit-rates. In the following, we present a new method that estimates the residual rate-distortion (R-D) in motion vector selection.

3.1. The Labeled R-D Curve of "AND" and "OR" Nodes

A labeled R-D curve is an R-D curve labeled by parameters to indicate that the curve is the encoding result of using the parameters. An R-D curve may perform better than other curves at one bit-rate, but worse at another bit-rate. Therefore, the R-D curves obtained by encoding with different parameters should be processed to obtain the optimum R-D curve. The curve can be optimal at any bit rate. Also, the curve must be labeled so that the parameter setting that achieves the optimal distortion at a specific bit-rate can be distinguished from the labels on the curve.

The proposed R-D curve combination (R-D combination for short) combines many R-D curves into an optimal curve for all bit-rates. The basic idea is to calculate all R-D points of all the modes in R-D plane, and then discard the R-D points located in the upper-right quadrant of some R-D point in the rate-distortion plane. For example, an R-D point (R, D) in the upper-right quadrant of another point (R', D') is removed, since encoding with the parameters to obtain (R, D) uses more bits ($R \geq R'$) and achieves a worse performance ($D \geq D'$) than the encoding used to obtain (R', D'). After discarding those points, we connect the remaining points to obtain the optimal R-D curve for the combination process. The curve is the lower envelope curve of all R-D points. Note that it is not necessarily a convex curve. Figure 1 shows a simple example where two curves are combined. We also label a point on the curve to indicate the parameters used to obtained it.

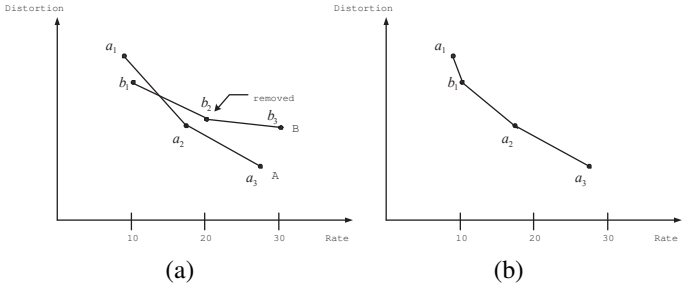

Fig. 1. Two steps of R-D combination: (a) The points b_2 and b_3 in the upper-right quadrant of a_2 are removed; (b) The remaining points are connected. The labels of points a_1, a_2, and a_3 indicate that the points originated from motion vector A, while b_1 originated from motion vector B.

In an AND-OR tree, an R-D combination is classified as either an AND R-D combination or an OR R-D combination. For a node in the OR state, only one of its child nodes is selected at bit rate b:

$$D_{OR}(b) = \min_{k \in Child} D_k(b), \tag{4}$$

where $Child$ is the set of child nodes. For this reason, the R-D combination method is applied to the R-D points of the child nodes to obtain the optimal R-D curve $D_{OR}(b)$. On the other hand, for a node in the AND state, we need to solve the following minimization problem at bit rate b:

$$D_{AND}(b) = \min_{\{b_k\}} \sum_{k \in Child} D_k(b_k), \tag{5}$$

with the constraint that $\sum_{k \in child} b_k \leq b$. Because of the rate constraint, before applying the R-D combination, we first explore every possible combination of the R-D points that are in agreement with the constraint. The R-D combination is then applied to obtain the optimal R-D curve $D_{AND}(b)$.

4. AND-OR TREE STRUCTURE FOR VIDEO CODING

The basic approach of using an AND-OR tree for video compression determines whether each operational mode is an AND relation or an OR relation, and organizes them as an AND-OR tree. Although many parameters are needed to encode a frame, we only list those related to INTER prediction decisions, namely: wavelet filters (called the wavelet filter selection mode), variable block size motion prediction (called the INTER variable block size mode), motion estimation, and residual coding.

- Wavelet filter selection mode: We decompose a macroblock of n levels using m wavelets. In the setting, the mode has n^m values, but only one of the values is chosen for a macroblock. Because only one method is chosen, the operation mode is an OR relation.
- INTER variable block size mode: INTER-16x16, INTER-16x8, INTER-8x16, INTER-8x8. A macroblock can be partitioned into any of those modes and each one can be further sub-divided. Therefore, this mode begins with an OR relation, and is followed by an AND relation. Note that INTER-16x16 is not partitioned. To ensure that an AND relation follows the OR relation, a singleton AND node is created as a child of INTER-16x16.
- Motion estimation and residual coding: For a macroblock with a given wavelet filter selection mode and INTER variable block size mode, a motion vector and residual coding parameters are selected for a given bit rate. This mode is a selection mode; hence, it is an OR relation.

4.1. AND-OR Tree DFS Rate-control Algorithm

The AND-OR tree structure for encoding a P-frame using our codec is shown in Figure 2. It is explored with a depth-first-search (DFS) algorithm. Let $D = 0$ and $P = \phi$ denote the distortion and the sequence of the coding parameters from the root to the current node, respectively. The algorithm explores the tree from the root node v. The child nodes of an OR node are explored by calling *SearchAND*, which is as follows:

$SearchOR(v, b, D, P)$: v is an OR node, and b is the rate.

1. For each child node w of v do
   - 1.1 If w is not visited, then
     - 1.1.1 Let s be a sequence of sampling bit rates: $[b_1, b_2, \cdots] \leq b$.
     - 1.1.2 For each sampling bit rate $b_s \in s$ do
       - 1.1.2.1 Call $SearchAND(w, b_s, D, P)$. Obtain an R-D point of w at bit rate b_s.
     - 1.1.3 End for-loop
     - 1.1.4 Obtain the optimal labeled R-D curve of w.

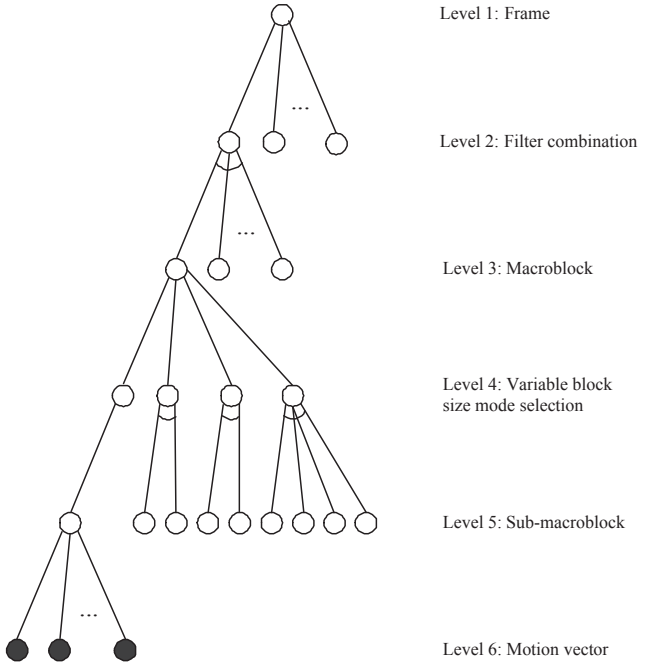

Fig. 2. The frame-level AND-OR tree structure of our codec. The root is an OR node; and the leaf nodes are highlighted with solid circles. The parameter from the root to a leaf is the label of the leaf.

2. End for-loop

We can apply the OR R-D combination on the combined R-D curve afterwards, and find out the child node that gives the optimal distortion D^* at rate b. The *SearchAND* algorithm explores the child nodes of an AND node by recursively calling the *SearchOR* algorithm. The steps of the *SearchAND* algorithm are as follows:

$SearchAND(v, b, D, P)$: v is an AND node, and b is the rate.

1. If v is a leaf, then perform residual coding. Return the optimal distortion at bit rate b and the residual coding parameters to obtain the optimum.
2. For each child node w of v do
   - 2.1 If w is not visited, then
     - 2.1.1 Let s be a sequence of sampling bit rates: $[b_1, b_2, \cdots] \leq b$.
     - 2.1.2 For each sampling bit rate $b_s \in s$ do
       - 2.1.2.1 Call $SearchOR(w, b_s, D, P)$. Obtain an R-D point of w at bit rate b_s.
     - 2.1.3 End for-loop
     - 2.1.4 Obtain the optimal labeled R-D curve of w.
3. End for-loop

After the $SearchAND$ ends, we apply the AND R-D combination to obtain the optimal labeled R-D curve $D_{AND}(b)$.

5. PERFORMANCE EVALUATIONS AND COMPARISONS

The video sequences used in the experiments are Y-luminance signals in QCIF format, ten frames per second. The first frame of each sequence is DCT encoded, but the others are raw data. Except for the first frame, motion prediction is always based on the previous decoded frame. We also implement a H.264-like codec for the experiments. As with the 4x4 integer transform of H.264/AVC, we use an integer wavelet transform to remove potential drifting effects when decoding residual values. We use two wavelets: the Haar filter and the 5-3 filter for performing three wavelet decompositions on each QCIF P-frame. A dyadic wavelet transform is applied to a whole frame and the wavelet coefficients are weighted. The discrete wavelet coefficients of a block are obtained by re-arranging the weighted dyadic wavelet coefficients of the frame. Motion vectors are in integer precision and obtained within a search range ± 16 on both the x-axis and the y-axis. The number of bits allocated to frames is the same in H.264-like codec and the proposed system. The motion vector of a block is derived independently and encoded in the same way as in H.264/AVC. The EZBC encoder is used to generate the encoded bitstream and context-based adaptive binary arithmetic entropy coding is used to code the parameters of our codec.

Figure 3 compares the performance of the proposed method with that of the H.264-like codec. Although H.264 outperforms our method sometimes, the overall improvement of our method is not in doubt, especially at low bit-rates. The figure illustrates the effectiveness of our algorithm. For high bit-rates, our codec's PSNR performance is close to that of H.264-like codec, but for low bit-rates, the performance difference for most sequences is substantial.

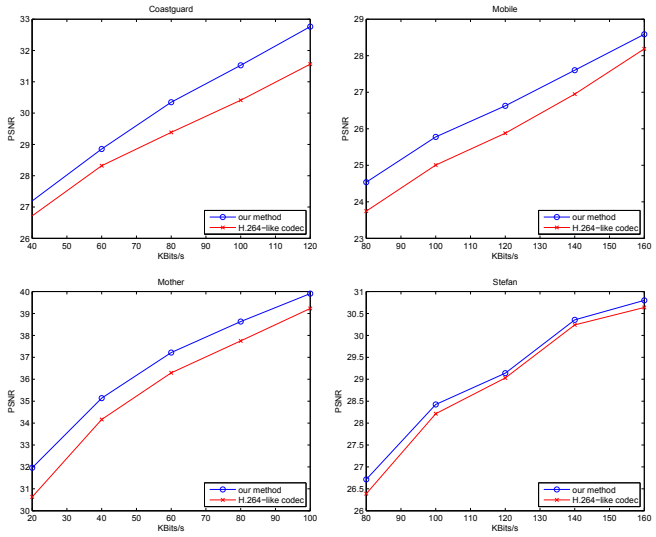

Fig. 3. The PSNR performance of our algorithm compared to H.264

6. CONCLUSION

We propose an AND-OR tree-based approach for organizing the mode selections in a hybrid video codec. The rate-distortion cost of the tree is analyzed and the optimal number of bits is allocated to each node. We estimate motion vectors in the wavelet domain and show that a weight is needed for each subband in order to derive the ratio of the wavelet errors to image errors. Our codec has a exclusive feature that uses wavelets in different scales. The experiments show that using the proposed feature, the proposed motion estimation algorithm, and applying the optimal bit allocation algorithm to a tree improves the coding performance. However, the performance of our codec improves at a cost of increasing the encoding complexity. Our future research direction is to reduce the complexity in order to efficiently obtain the optimal graph solution of an AND-OR tree.

Acknowledgement: We would like to express our gratitude to Karsten Suehring for his assistance in explaining many issues regarding to the execution of JM 10.2.

7. REFERENCES

[1] N. J.Nilsson. *Principles of Artificial Intelligence*. Springer Verlag, 1983.

[2] K. K. Lin and R. M. Gray. Wavelet video coding with dependent optimization. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(4), April 2004.

[3] H.-W. Park and H.-S. Kim. Motion estimation using low-band-shift method for wavelet-based moving-picture coding. *IEEE Trans. on Image Processing*, 9(4), April 2000.

[4] G. J. Sullivan and T. Wiegand. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, pages 74–90, November 1998.

[5] B. E. Usevitch. Optimal bit allocation for biorthogonal wavelet coding. *Proc. Data Compression Conference*, 1996.

[6] B. E. Usevitch. A tutorial on modern lossy wavelet image compression: Foundations of jpeg 2000. *IEEE Signal Processing Magazine*, September 2001.

[7] T. Wiegrand, H. Schwarz, A. Joch, F.Kossentini, and G. J. Sullivan. Rate-constrained coder control and comparison of video coding standards. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(7), July 2003.

[8] M. Wien. Variable block-size transforms for h.264/avc. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(7):604–613, July 2003.

[9] Y. Yang and S. Hermami. Generalized rate-distortion optimization for motion-compensated video coders. *IEEE Trans. on Circuits and Systems for Video Technology*, 10, September 2000.