A temporal mode selection in the MPEG-2 encoder scheme

Laurent Piron
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2605; fax +41 21 693 7600
e-mail: piron@ltssg2.epfl.ch

ABSTRACT

This paper deals with the mode decision in an MPEG–2 framework. An algorithm for mode decision is introduced. This algorithm is based on a rate-distortion criterion and takes into account the temporal dependency of the frames. This approach can allow a quality gain of more than one dB compared to the Test Model 5 (TM5) mode decision algorithm.

1 INTRODUCTION

As for all the standards (H-263, MPEG-1), the MPEG-2 (Motion Pictures Expert Group, phase 2) standard [2] defines only the syntax of the bit-stream. There is no constraint on the encoder. This allows complete freedom in the design of the encoder. However, the encoder has to be block-based, that is the input image is divided in blocks and the encoding process operates on individual block at a time. There are several *modes* for coding each block. For example, a block may be coded with its DCT coefficients or it may be predicted from a block in the previous image. Thus, *mode decision* (choosing the optimal mode for a block) is an important problem.

In the Test Model 5, each frame is encoded independently of the others. Each block is coded using one of several modes. The mode decision is based on the variance of the difference between the original block and the decoded block. The mode that produces the smallest variance is selected. This approach does not take into account the target bit-rate of a particular application. In this paper, we propose a mode decision algorithm based on a rate distortion criterion. This algorithm also takes into account the temporal dependency between frames.

Section 2 presents the relevant aspects of MPEG–2, while section 5 introduces the proposed algorithm. Simulations results are described in the section 6. Some conclusions are presented in section 7.

2 MPEG–2 CODING MODES

The MPEG–2 standard divides a video sequence into Groups Of Pictures (GOP). Each group is considered independently in the coding process. In each GOP, the first frame is always coded without any temporal reference. This type of frame is called Intra (I). The other frames in the GOP are Predicted (P) or Bi-directionally interpolated (B) frames. A P-frame is predicted from the previous I- or P-frame. A B-frame is predicted from the previous and next I- or P-frames. A common GOP structure is shown in Fig. 1. In this figure, a GOP is composed of 12 frames. Each I- or P-frame is followed by two B-frames.

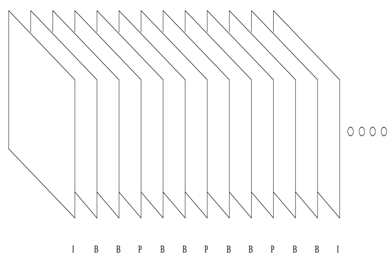

Figure 1: GOP structure. The GOP consists of 12 frames. Each I- or P-frame is followed by two B-frames. The last I-frame marks the beginning of the next GOP.

Each frame is divided into smaller blocks, called *macro-blocks*, of size 16×16 pixels. A macro-block may be Intra-coded (without any temporal reference), Forward-predicted (with reference to the previous I- or P- frame), Backward-predicted (with reference to the future I- or P-frame) or Interpolated (based on the previous and the future I- or P-frame).

The type of the frame (I, P or B) defines the modes which are allowed for a macro-block in that frame. For instance, in a P-frame, only the Intra and Forward modes are allowed. The coding modes allowed in MPEG–2 are:

- I-frame
  1. Intra
- P-frame
  1. Intra

2. Forward prediction with two motion vectors
3. Forward prediction with one motion vector
4. Forward prediction with no motion compensation
5. Dual prime prediction (only when B-frames are not used)

- B-frame

1. Intra
2. Forward prediction with two motion vectors
3. Forward prediction with one motion vector
4. Backward prediction with two motion vectors
5. Backward prediction with one motion vector
6. Interpolated prediction with two motion vectors
7. Interpolated prediction with one motion vectors

3 RATE DISTORTION OPTIMIZATION

In this work, the Rate Distortion, R(D), criterion is used for the mode selection. More specifically, all possible coding modes for a macro-block are investigated. The coding mode that minimizes R(D) is selected for coding this macro-block.The problem of mode decision can be stated as the following optimization procedure:

$$\min_{GOP} \mathcal{D}_{\mathcal{MB}} \quad \text{such that } \mathcal{R}_{\mathcal{MB}} \leq \mathcal{R}_{target}$$

where R_{target} is the rate target for a macro-block, $\mathcal{D}_{\mathcal{MB}}$ is the distortion for the macro-block and $\mathcal{R}_{\mathcal{MB}}$ is the rate for the macro-block.

This constrained problem is first converted into an unconstrained problem by using Lagrangian multipliers [1]. We define the value $J(\lambda) = \mathcal{D}_{\mathcal{MB}} + \lambda \mathcal{R}_{\mathcal{MB}}$ where λ is the constraint parameter. This yields the following optimization problem:

$$\min_{GOP} J(\lambda) \quad \text{for a given macro-block}$$

with $J(\lambda) = \mathcal{D}_{\mathcal{MB}} + \lambda \times \mathcal{R}_{\mathcal{MB}}$, where $\lambda > 0$ is the constraint parameter. The mode that minimizes $J(\lambda)$ is selected.

4 TEMPORAL DEPENDENCY

Frames in a video sequence are not independent. This dependency is exploited by motion estimation algorithms. In this work, temporal dependency is also taken into account in the optimization problem presented in section 3. The mode selection for a macro-block depends on the modes selected for macro-blocks at the same position in previous coded frames [4]. For P-frames, the optimal $J(\lambda)$ for a macro-block is the one that also minimizes the distortion of the relevant macro-blocks in B-frames that are interpolated using this particular P-frame. In B-frames, the optimization does not consider macro-blocks in other frames.

5 PROPOSED MODE DECISION METHOD

The coding possibilities for a macro-block are represented in a graph structure. The nodes of the graph represent information about macro-blocks in P- and I-frames. The edges represent informations about macro-blocks in B-frames. For convenience, we use the term I-node, P-node and B-edge to refer to the nodes and edges of this graph. Fig. 2 shows the graph structure for the GOP shown in Fig. 1.

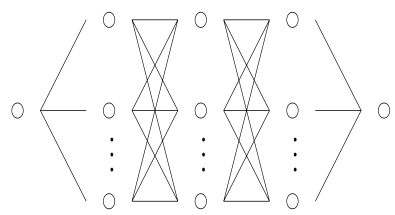

Figure 2: Graph representation for a macro-block in a GOP. The edges represent informations about macro-blocks in B-frames. The nodes represent information about macro-blocks in P- and I-frames.

A possible coding mode for a macro-block is associated to each node. The information assigned to each node is $J_I(\lambda)$ if the node represents a macro-block from a I-frame and $J_P(\lambda)$ if the node represents a macro-block from a P-frame. For the GOP shown in Fig. 1, each edge is assigned the value $\min(J_{B_1}(\lambda)) + \min(J_{B_2}(\lambda))$ where B_1 and B_2 are two successive B frames.

This graph offers several paths between macro-blocks of consecutive I-frames. We use the Viterbi algorithm [3] to select the minimum-cost path. The optimal coding mode for a macro-block lies on this path.

The optimization algorithm is as follows:

1. Generate a graph representation of the GOP structure for each macro-block, as explained above.
2. Assign the cost $J_I(\lambda)$ to the I-node.
3. Assign the cost $J_P(\lambda)$ to the P-nodes.
4. Assign $\min(J_{B_1}(\lambda)) + \min(J_{B_1}(\lambda))$ to each B-edge.
5. Use Viterbi algorithm to find the minimum-cost path between the two I-nodes in the graph.

The coding modes on this path are used for coding the macro-blocks in this GOP.

6 SIMULATIONS RESULTS

Simulations results are shown for three different sequences: “Mobile-calendar”, “Flower-garden” and “Basket”. These sequences have been coded at two different bitrates: 3 Mbits/s. and 5 Mbits/s. The sequences are in

CCIR-601 format (720×576), and are five seconds long each.

Figures 3, 4 and 5 show comparisons between the proposed algorithm and the TM5 algorithm. The comparison is based on peak signal-to-noise ratio (PSNR) of the decoded sequences. In each figure, the solid line represents the PSNR obtained using the proposed algorithm while the dashed line corresponds to the PSNR resulting from the TM5 algorithm.

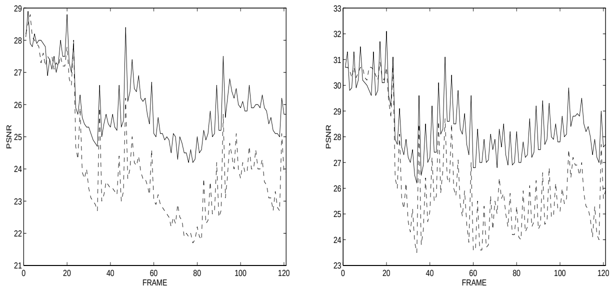

Figure 3: PSNR results for "Basket" at 3 Mbits/s. (left) and 5 Mbits/s. (right). The solid line represents the PSNR obtained using the proposed algorithm while the dashed line corresponds to the PSNR resulting from the TM5 algorithm.

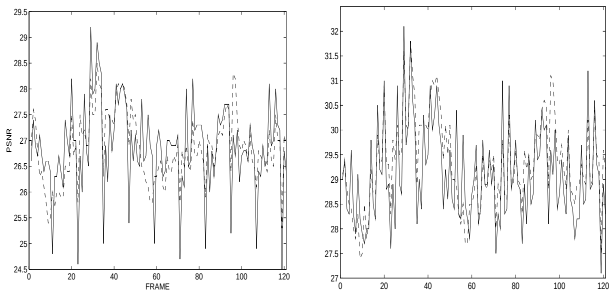

Figure 4: PSNR results for "Mobile-calendar" at 3 Mbits/s. (left) and 5 Mbits/s. (right). The solid line represents the PSNR obtained using the proposed algorithm while the dashed line corresponds to the PSNR resulting from the TM5 algorithm.

These figures show that the proposed algorithm performs as well as the TM5 algorithm or better. The performances of both techniques are comparable for the "Flower-garden" and "Mobile-calendar" sequences. Tables 1 and 2 show the average PSNR (over five seconds) for the proposed algorithm and for the TM5 algorithm. For the "Basket" sequence, the proposed method provides an improvement of around 1.5 dB for a target bit-rate of 3 Mbits/s. and around 2 dB at 5 Mbits/s. The "Basket" sequence shows a basket-ball game. The motion information in this sequence is quite

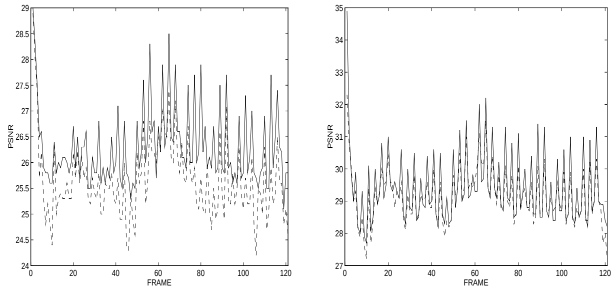

Figure 5: PSNR results for "Flower-Garden" at 3 Mbits/s. (left) and 5 Mbits/s. (right). The solid line represents the PSNR obtained using the proposed algorithm while the dashed line corresponds to the PSNR resulting from the TM5 algorithm.

	Proposed algorithm	Test Model 5
Mobile-calendar	26.87	26.84
Flower-garden	26.29	25.63
Basket	25.98	24.23

Table 1: Average PSNR for the test sequences at 3 Mbits/s.

erratic. This is probably the reason why the proposed algorithm performs better than the TM5 algorithm on this sequence. The proposed method makes better use of the motion information than the TM5 algorithm. In the proposed algorithm, the rate-distortion optimization takes into account the (differentially coded) motion information.

7 CONCLUSION

In this paper, a mode selection algorithm for MPEG–2 has been presented. This algorithm is based on a rate distortion criterion which takes temporal dependency of the macro-blocks into account. Simulation results show that a gain around one dB in the PSNR is possible, using the proposed algorithm. The main drawback of this approach is the high computational complexity. In order to select the optimal mode for the macro-blocks, all possible coding modes have to be investigated. It

	Proposed algorithm	Test Model 5
Mobile-calendar	29.09	29.31
Flower-garden	29.40	29.12
Basket	28.44	26.42

Table 2: Average PSNR for the test sequences at 5 Mbits/s.

is possible to reduce the computational complexity by introducing pruning conditions in the graph-generation procedure.

8 ACKNOWLEDGEMENTS

The author would like to thank Sushil Bhattacharjee for his help in this work.

References

[1] H. Everett. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operation research*, 11:399–417, 1963.

[2] International Organization for Standardization. *MPEG-2 DRAFT INTERNATIONAL STANDARD ISO/IEC DIS 13818-2*, 1994.

[3] G. D. Forney. The viterbi algorithm. *Proc. IEEE*, 61:268–278, March 1973.

[4] Kannan Ramchandran, Antonio Ortega, and Martin Vetterli. Bit allocation for dependent quantization with application to multi-resolution and MPEG video coders. *IEEE transactions on image processing*, 3(5), September 1994.