RATE-DISTORTION OPTIMIZED VIDEO CODING WITH STOPPING RULES: QUALITY AND COMPLEXITY

Marcos Moecke and Rui Seara

LINSE – Circuits and Signal Processing Laboratory
Department of Electrical Engineering
Federal University of Santa Catarina
88040-900 – Florianópolis – SC – Brazil

ABSTRACT

This paper presents a new motion estimation (ME) strategy for video coding. Such a strategy makes use of stopping rules in the ME process, which considers both the motion vector coding cost and desired minimum quality for each macroblock used. In a rate-distortion sense, the obtained results by the proposed algorithm are superior to the ones achieved with the full search (FS) algorithm. Moreover, the strategy reduces the mean number of searching points. Through simulations, by using an H.263 encoder, we verify that the proposed algorithm is more effective for video coding than the one used in the TMN11-LC (low complexity mode decision) FS implementation.

1. INTRODUCTION

Block-based hybrid video coders generally consist of two stages: block-matching motion compensation (MC) and displaced frame difference (DFD) coding. In these coders, blocks are coded by using multiple modes (*Skip*, *Inter*, and *Intra*). Block sizes of 16 × 16 and 8 × 8 pixels are used in the video coding standards H.263 [1], MPEG-2 [2] and MPEG-4 [3]. Smaller block sizes as far as 4 × 4 pixels are supported in the H.264/AVC standard [3].

Several research works have allowed for choosing the appropriate coding mode, the motion parameters and the block size to be used. The computational complexity required and the type of application in question are primary factors in the choice of the ME technique to be used. For high coding quality, a complex coding process should be used. However, in this case, the application should tolerate an off-line coding procedure. On the other hand, for on-line applications, such as video conference and video telephony, real-time low complexity coders are needed [4].

The motion vector (MV) for each block can be obtained by a full search (FS) algorithm, minimizing a chosen distortion function. Fast algorithms, such as NTSS [5], BBGDS [6], DS [7], among others, reduce the ME computational complexity by restricting the number of search points (NSP). However, such approaches lead to an MC quality reduction [7]. In contrast, by using partial distortion techniques, the search complexity is reduced without loss of quality [8]. Nevertheless, for the most cases, the isolated use of these techniques does not lead to a satisfactory complexity reduction. In addition, all mentioned techniques use only distortion measurements as a merit figure for assessing the MVs in the ME process. For this reason, a criterion that associates bit rate with the distortion is more adequate, leading to a better rate-distortion (RD) performance.

Now, taking into account the coding mode selection, the choice of the block sizes for MC must be considered. Since most coders work with several block sizes, the ME process must be carried out for all the possible block sizes, leading to different results of distortion and MV coding cost. An example is the solution adopted in the TMN11-LC [7]. Such a strategy simplifies both the ME process and mode selection. However, the RD results obtained are very poor in performance. This is due to the fixed penalties used to compare the distortions which do not take into account the overall bit rate. The way to obtain the best trade-off between distortion and bit rate is to minimize the distortion considering a defined bit rate. To this end, the use of Lagrangian optimization techniques [9] can be an interesting option, irrespective of the high computational complexity involved when the mode selection and the ME are jointly optimized. A widely accepted intermediate solution is to optimize sequentially the ME and the mode selection. This approach leads to a result near the optimum.

This paper presents a new ME and mode selection strategy based on a Lagrangian cost criterion, which lead to a significant improvement as compared with the ones used in the TMN11-LC. In our approach ME computational complexity depends directly on the required bit rate, resulting in a considerable computational load reduction for low bit rate applications. The proposed algorithm assesses the MVs in an increasing coding cost (ICC) order. The ME process is ended when an equivalent distortion, corresponding to the coding cost, is larger than the current minimum distortion. As a second ME stopping criterion a distortion threshold (DT) is used, which ends the process when a desired minimum quality is achieved. By using such criteria, a significant improvement in RD performance along with an NSP reduction is obtained.

2. RATE-DISTORTION OPTIMIZED MOTION ESTIMATION

To implement the cost-based stopping rule, initially all the possible MVs [$p = (h, v)$] must be organized in an increasing coding cost order. The set of MVs, here defined as $\mathbf{P} = (p_1, p_2, \cdots, p_{\text{last}})$, is partitioned in subgroups

---

Marcos Moecke is also with the Telecom. Dept., Federal Technological Education Center of Santa Catarina (CEFETSC), São José-SC, 88103-310, Brazil. This work is partially supported by the Brazilian National Research Council (CNPq).

0-7803-8554-3/04/$20.00 ©2004 IEEE.

$\mathbf{P}=(\mathbf{S}_{k1},\cdots,\mathbf{S}_{kn})$, such that each subgroup $\mathbf{S}_k$ only contains MVs with an equal coding cost, $k=R_{\mathrm{MV}}(p)$, where $R(\cdot)$ represents the number of bits used in the coding. When the MV components h and v are coded independently, the cost can be determined by the sum of the cost of each component, $R_{\mathrm{MV}}(p)=R(h)+R(v)$. Furthermore, the coding cost of each MV can be obtained *a priori* if a variable length code (VLC) is used for coding the MVs. In that case, the subgroups S_k are fixed.

To optimize the ME process in RD sense, the Lagrangian cost function, $J_d(p)$, is used. Thus, the optimum MV for each block is obtained by

$$p_{\mathrm{opt}}=\min_{p\in\mathbf{P}}\left\{J_d(p)=D_d(p)+\lambda^d R_{\mathrm{MV}}(p)\right\}, \tag{1}$$

where λ is the Lagrange multiplier; d characterizes the distortion type used, either sum of absolute differences (SAD) ($d=1$) or sum of squared differences (SSD) ($d=2$); and $D_d(p)$ denotes the distortion function [3]. For the sake of simplicity, from now on the index d is only used as needed to distinguish the type of distortion measurement used.

We have used $\lambda=f(Q)$ [9], where Q represents the quantizer parameter which controls the overall bit rate. Note that $J(p)$ can be interpreted as an equivalent distortion measure and λ as a coding cost to distortion penalty converter $\lambda^d R_{\mathrm{MV}}(p)$. In this way, the distortion penalty is now directly dependent on the bit rate.

Since ME is carried out in an ICC order, it is possible to determine for each new S_k whether the current minimum Lagrangian cost, $J_{\min}(p_x)$, is or not the optimum one [$J_{\mathrm{opt}}(p_x)$]. Thus, if $J_{\min}(p_x)\le\lambda k$, then $J_{\mathrm{opt}}(p_x)=J_{\min}(p_x)$ and the ME process is ended. Then, the solution of (1) is $p_{\mathrm{opt}}=p_x$. With such a cost-based stopping rule, the search space is reduced, maintaining the ME process optimized in RD.

The inclusion of a long-term memory [3] in the ME/MC process can be considered, but to simplify our analysis, we have used only the previous frame as a reference.

Depending on the application and the required quality, a certain distortion threshold $D_{\mathrm{thr}}\ge 0$ can be used. So, the ME process can be stopped, since a minimum quality level in the corresponding block has been attained. Considering that the MV search is carried out as previously described, and the ME process is stopped when $J_{\min}(p_x)\le D_{\mathrm{thr}}$, the found MV p_x is the one of least cost for the considered D_{thr}. The value D_{thr} must be determined taking into account the distortion function and the block size used.

3. MODE SELECTION APPROACH

The right mode selection for each block is a primary factor for an efficient coding. When the Lagrangian cost function is used for optimizing the mode selection, the coder normally needs to compute the bit rate and distortion for all the upheld modes [9], which leads to a high computational burden if various candidate modes are considered [7]. To overcome such a drawback, we propose an alternative approach that ignores *a priori* some of those modes. In the following we describe in details the procedure adopted.

According to the bitstream syntax, an initial coding cost $R_{\mathrm{init}}(m)$ is determined for each mode m. That cost is defined as the number of mandatory bits required for the block coding. In this way, an initial distortion penalty is computed as

$$D_{d,\mathrm{init}}(m)=\lambda^d R_{\mathrm{init}}(m)\,. \tag{2}$$

Then, by using the $D_{\mathrm{init}}(m)$ values determined, a mode selection test order can be established. Thus,

$$\mathbf{M}=\left\{m_1\prec m_2\prec m_3\cdots m_N\right\}\,|\,m_i\prec m_j \Leftrightarrow D_{\mathrm{init}}(m_i)\le D_{\mathrm{init}}(m_j) \tag{3}$$

where $\prec$ denotes a precedence relation between the elements of the arranged set $\mathbf{M}$. Moreover, as the *Skip* mode has the lowest initial distortion penalty, it is generally the first mode, m_1, in the test order.

3.1. *Skip* mode pre-selection

The next step is to test if the *Skip* mode can be pre-selected. To this end, its coding cost is computed. Thus,

$$J_d(\mathrm{Skip})=D_d(p_{\mathrm{null}})+\lambda^d R_{\mathrm{H}}(\mathrm{Skip})\,, \tag{4}$$

where p_{null} represents the null MV; $D(p_{\mathrm{null}})$ denotes the current frame block distortion with respect to the reference frame; and $R_{\mathrm{H}}(m)$ is the coding cost of the block header, which defines the used mode. Now, if

$$J_d(\mathrm{Skip})\le D_{d,\mathrm{init}}(\mathrm{Inter})+D_{\mathrm{thr}}\,, \tag{5}$$

then the *Skip* mode is pre-selected and ME is not carried out, else the ME process is started.

3.2. Block size choice

Coders that comply with the H.26X and MPEG standards use a macroblock (MB) structure that merges different block sizes, defined as $\mathrm{MB}=\left\{B_1,B_2,\cdots,B_M\right\}$. To attain the best RD performance for each MB, usually the coder has to carry out the ME process with all the possible block sizes. Once known the optimum MV for each block [computed by (1)] the Lagrangian cost of the MB $J_{d,\mathrm{MB}}(m)$, for all kinds of *Inter* mode, can be computed as

$$J_{d,\mathrm{MB}}(m)=\sum_{\forall B_i\in\mathrm{MB}} D_d(p_{B_i})+\lambda^d\left[\sum_{\forall B_i\in\mathrm{MB}} R_{\mathrm{MV}}(p_{B_i})+R_{\mathrm{H}}(m)\right]. \tag{6}$$

So, the best *Inter* mode can be obtained by minimizing (6). Now, aiming to reduce the computational load of this minimization process, the ME is accomplished following the order of the test defined by $\mathbf{M}$ [see (3)]. The distortion penalty $D_{\mathrm{init}}(m)$, for each *Inter* mode, is determined by

$$D_{d,\mathrm{init}}(m)=\lambda^d\left[\sum_{\forall B_i\in\mathrm{MB}} R_{\mathrm{MV}}(p_{B_i,\mathrm{PMV}})+R_{\mathrm{Hinit}}(m)\right]. \tag{7}$$

where $R_{\mathrm{Hinit}}(m)$ denotes the mandatory bits of the MB header; and $p_{B_i,\mathrm{PMV}}$ is the predicted MV for the block B_i [1]-[3], corresponding to the one of lower cost. Since, this distortion penalty is proportional to the number of blocks forming the MB, the ME process must start with the block of larger size (for instance, 16×16). As a consequence, the ME for smaller blocks can be avoided when

$$\min_{m_i \in \mathbf{M} \mid i<j} \left\{ J_{\mathrm{MB,min}}(m_i) \right\} \le D_{\mathrm{init}}(m_j) + D_{\mathrm{thr}} , \qquad (8)$$

where $J_{\mathrm{MB,min}}(m)$ is the minimum Lagrangian cost of the MB for the mode m .

3.3. Rate-distortion optimized mode selection

If the *Skip* mode is not selected by (5), then the optimum mode is selected by

$$m_{\mathrm{opt}} = \min_{m_i \in \mathbf{M}} \left\{ D_{2,\mathrm{REC}}(m_i \mid Q) + \lambda^2 R_{\mathrm{REC}}(m_i \mid Q) \right\} , \qquad (9)$$

where $D_{2,\mathrm{REC}}$ is the distortion measured between the original and reconstructed pixels of the MB; R_{REC} is the overall bit rate of the MB, considering a coding mode m_i and a quantization parameter Q .

Therefore, by using the proposed approach the measurements of distortion and bit rate are only computed for the remaining modes determined by (5) and (8). For this reason, our approach leads to a significant reduction in computational complexity as compared with the other established procedures as TMN11-HC (high complexity model) [7] and [9].

4. PROPOSED ALGORITHM APPLIED TO H.263-BASED VIDEO CODING

To illustrate the use of the proposed algorithm, we integrate it into the University of British Columbia's (UBC) H.263+ encoder [10]. We have considered the case in which the VLC for unrestricted MV mode (Annex D of H.263) has been used for obtaining the coding cost of the MV differences (MVDs). The mode selection and ME (using $\lambda = 0.92Q$ [7]) have been phased as shown in the next sections.

4.1. *Skip* mode pre-selection

The first phase is to pre-select the *Skip* mode by using (5) and (7). For the VLC considered, $R_{\mathrm{H}}(\mathrm{Skip}) = 1$, $R_{\mathrm{MV}}(p_{\mathrm{PMV}}) = 2$ and $R_{\mathrm{Hinit}}(\mathrm{Inter}_{16\times16}) = 4$ bits are used. Thereby, the *Skip* mode is selected as

$$J(\mathrm{Skip}) \le 5.1Q^2 + D_{\mathrm{thr}} . \qquad (10)$$

In this way, the Skip mode selection is controlled by the Q^2 factor, leading to favor this mode in low bit rate as well as reduce the NSP average. Such advantages are not achieved simultaneously either in TMN11-LC or in other mode selection strategies [7], [9].

After all, if the MV for *Inter*$_{16\times16}$ mode and the DFD coefficients (after the discrete cosine transform and their quantization) are null, the *Skip* mode can be selected in the same way as in [7] and [9].

4.2. Block size choice and mode selection

Secondly, the ME process with 16 × 16 blocks is accomplished by minimizing (6). The search center, for full-pixel precision, is defined as the predicted MV. Following, a half-pixel refinement is performed. Moreover, if the condition (11) below is fulfilled [considering (7) and (8) with $R_{\mathrm{MV}}(p_{\mathrm{PMV}}) = 2$ and $R_{\mathrm{Hinit}}(\mathrm{Inter}_{8\times8}) = 6$], the J_{opt} previously obtained cannot be improved by any combination of MV. Otherwise, the 8×8 block size ME process should be carried out.

$$\min \left\{ J(\mathrm{Skip}), J_{\mathrm{opt}}(\mathrm{Inter}_{16\times16}) \right\} > 15.3\, Q^2 + D_{\mathrm{thr}} . \qquad (11)$$

In contrast with the TMN11 approach, for *Inter*$_{8\times8}$ mode selection, the predicted MV now corresponds to the search center. Note that the order of the B_1, B_2, B_3 and B_4 blocks (defined in the H.263 standard) must be followed to preserve neighboring relationships between the blocks for the MV predictions.

Finally, once completed the ME process for each MB, the coding mode selection is accomplished as previously discussed (see Section 3.3).

5. EXPERIMENTAL RESULTS

In this section the proposed ME and mode selection approaches are compared with the ones used in the H263+ reference implementation (TMN11-LC) built by UBC and Telenor [10] for performance. For such, we consider two cases: first the Baseline encoder, and next the Baseline-DFIT encoder (Baseline including Annexes D, F, I and T). For the comparison tests, the ten initial seconds of the QCIF sequences *Foreman*, *Mother & Daughter* (*M&D*), *News, Silent, Carphone* and *Akiyo* have been used. The sequences have been coded with different *Q* levels in IPPP structure at 10 and 15 Hz using a search range of ±15.5 pixels. The distortion threshold as defined is computed by $D_{d,\mathrm{thr}}(r) = L\,H\,r\,Q^d$, where $L \times H$ is the block size; and r is a coefficient ranging from 0 to 0.25. For half-pixel refinement, bilinear interpolation [1] and Wiener interpolation filter [3] have been used. We have selected some representative results to highlight the advantages of our approach. The peak signal-to-noise ratio (PSNR) of the luminance component (Y) is used for quality assessment.

Fig. 1 shows the rate-distortion plots for the *Akiyo* sequence with 15 Hz. We can note a significant improvement in RD as compared with the reference implementation considered. By using a Wiener interpolation filter along with our approach, a PSNR performance gain of 1-2 dB is achieved. This gain is due to three main differences between our and the TMN11-LC approaches: ME process and mode selection optimized in RD, and the improved subpixel interpolation using Wiener filter.

To measure the computational complexity, we have used the NSP (which corresponds to the number of MB distortion computed), since in the ME process the largest load is due to the distortion computation. Fig. 2 shows the NSP percentage relative to the FS algorithm *versus* PSNR performance. In this figure, we can verify a significant NSP reduction for the entire PSNR range, mainly for the low bit-rate case. Such a gain is because of both the *Skip* mode pre-selection (see Fig. 3) and stopping criteria used (cost-based and distortion-based). In addition, for 8 × 8 blocks, a higher reduction is achieved by bypassing the ME in a lot of MB [see condition (11)].

Fig. 4 shows a significant influence of the DT in the computational burden reduction as compared with the FS algorithm. This improvement is accomplished preserving the RD performance practically unchanged (less than 0.1 dB) for an *r* range between 0 and 0.125. For an *r* coefficient above 0.125, the RD performance is reduced mainly at a high bit-rate.

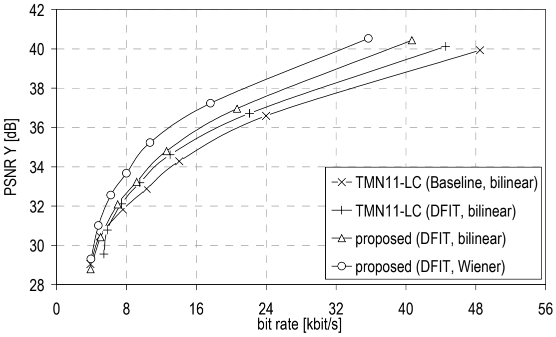

Fig. 1. Rate-distortion curves for the *Akiyo* sequence with 15 Hz.

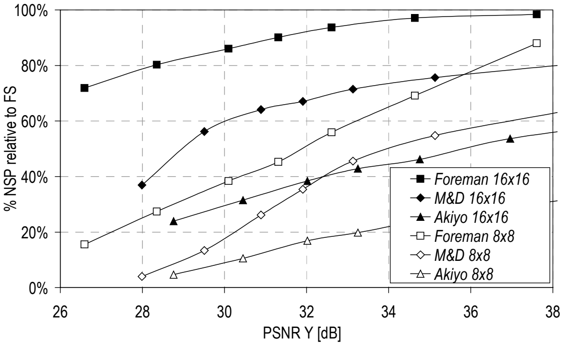

Fig. 2. NSP relative to TMN11-LC with FS-ME algorithm.

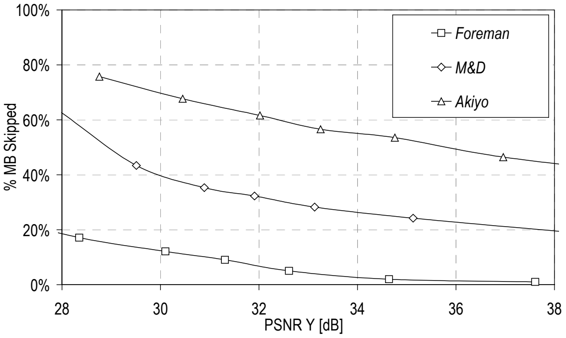

Fig. 3. Number of pre-selected MB as *Skip* mode for $r = 0.125$.

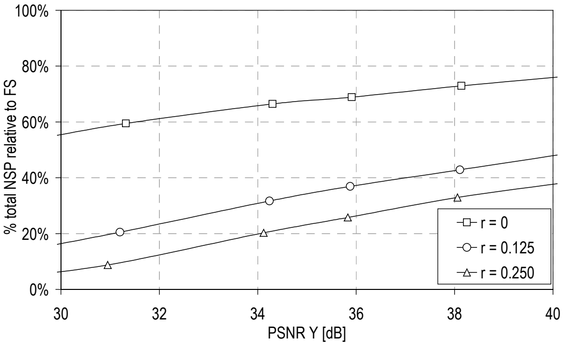

Fig. 4. NSP of the proposed algorithm relative to FS, with DT varying from $r = 0$ to 0.25, for the *Akiyo* sequence with 15 Hz.

Furthermore, comparing our approach with the one presented in [9] for the *Akiyo* sequence, we verify that it is similar in RD performance and requires a lower computational load. Note that such results are achieved though the long-term memory prediction mode (H.263++ Annex U) is not used.

6. CONCLUSIONS AND REMARKS

This paper presents an improved ME strategy for video coding, based on the RD optimization. Such approach outperforms the reference implementation (TMN11-LC) in both the computational complexity and RD. The introduced cost-based stopping rule does not affect the ME optimization process. The distortion-based stopping rule along with the former rule reduces considerably the ME search space, while maintaining the RD performance approximately unchanged. In addition, the computational load becomes dependent on the quantizer, leading to an automatic adjustment of complexity with the bit rate used. The mode selection is also simplified by bypassing the *Inter* and *Intra* mode in many MBs.

We expect that the proposed approach will result in similar outcomes when applied to MPEG video encoders and even higher reduction in the search space when more coding modes are supported as in H.264/AVC.

7. REFERENCES

[1] *Video Coding for Low Bit Rate Communication*, ITU-T Rec. H.263, Feb. 1998.

[2] *Generic Coding of Moving Picture and Associated Audio Information – Part 2: Video*, ITU-T Rec. H.262 - ISO/IEC 13 818-2, 1995.

[3] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next_generation Multimedia*. West Sussex: John Wiley & Sons, 2003.

[4] S. M. Akramullah, I. Ahmad, and M. L. Liou, "Optimization of H.263 Video Encoding Using a Single Processor Computer: Performance Tradeoffs and Benchmarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 901-915, Ago. 2001.

[5] R. Li, B. Zeng, and M. L. Liou, "A New Three Step Search Algorithm for Block Motion Estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 438-442, Ago. 1994.

[6] L. K. Liu and E. Feig, "A Block Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 419-422, Ago. 1996.

[7] *TMN11: Test Model 11*, ITU, Monterey, Feb. 1999.

[8] W. Li and E. Salari, "Successive Elimination Algorithm for Motion Estimation," *IEEE Trans. Image Processing*, vol. 4, pp. 105-107, Jan. 1995.

[9] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained Coder Control and Comparison of Video Coding Standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688-703, Jul. 2003.

[10] *TMN (H.263) coder version 3.0*, University of British Columbia, Canada.