An HVS-based Adaptive Computational Complexity Reduction Scheme for H.264/AVC Video Encoder using Prognostic Early Mode Exclusion

Muhammad Shafique, Bastian Molkenthin, and Jörg Henkel
Karlsruhe Institute of Technology, Chair for Embedded Systems, Karlsruhe, Germany
{shafique, molkenthin, henkel}@informatik.uni-karlsruhe.de

Abstract—The H.264/AVC video encoder standard significantly improves the compression efficiency by using variable block-sized Inter (P) and Intra (I) Macroblock (MB) coding modes. In this paper, we propose a novel Human Visual System based Adaptive Computational Complexity Reduction Scheme (ACCoReS). It performs Prognostic Early Mode Exclusion and a Hierarchical Fast Mode Prediction to exclude as many I-MB and P-MB coding modes as possible (up to 73%) even before the actual Rate Distortion Optimized Mode Decision (RDO-MD) and Motion Estimation while keeping a good quality. In the best case, ACCoReS processes exactly one MB Type and one corresponding near-optimal coding mode, such that the complete RDO-MD process is skipped. Experimental results show that compared to state-of-the-art approaches ([10], [22]-[26]), ACCoReS achieves a speedup of up to 9.14x (average 3x) with an average PSNR loss of 0.66 dB. Compared to exhaustive RDO-MD, ACCoReS provides a performance improvement of up to 19x (average 10x) for an average 3% PSNR loss.

I. INTRODUCTION
The advanced video coding standard H.264/AVC [1] was developed by the Joint Video Team (JVT) to provide a bit rate reduction of 50% as compared to MPEG-2 with similar subjective visual quality [7]. However, this improvement comes at the cost of significantly increased computational complexity [8], thus posing serious challenges on high-throughput (real-time) encoder implementations. High computational complexity of H.264 is mainly due to its complex Prediction, Motion Estimation (ME) and Rate Distortion Optimized Mode Decision (RDO-MD) processes that operate on multiple (variable) block sizes (as shown in Fig. 1). Large effort has been made in developing fast algorithms in ME for H.264 to reduce its complexity [15], [21]. However, RDO-MD is the most critical functional block in H.264, as it determines the number of ME iterations. Therefore, it becomes the primary research focus for complexity reduction.

A Macroblock (MB, i.e. 16x16 pixels) in a video frame can be divided into 16x16, 16x8, 8x16, or 8x8 blocks. Each 8x8 block can be further divided into 8x4, 4x8, or 4x4 sub-blocks. Altogether, there are 7 different block combinations per MB are evaluated in RDO-MD [11]. An example scenario is presented in Fig. 1.

![Fig. 1: Variable Block Sizes for Inter Prediction used in H.264/AVC](image)

Each MB Type can be predicted using one of the following coding modes' with variable block sizes:

1 In this paper, we only target Baseline and Main profiles, therefore, we do not consider 18x8. However, the decisions and steps for 14x4 in our proposed scheme are scalable for 8x8.
RDO Mode Elimination – as a fast RDO-MD – is used that eradicates the unlikely coding modes depending upon the output of previously processed coding mode. Note: ACCoReS facilitates the integration of previously predicted fast RDO-MD schemes (see [10]-[14], [18]-[26]) in the stage of Sequential RDO Mode Elimination.

The principal distinctions of our proposed ACCoReS compared to the state-of-the-art approaches are the Prognostic Early Mode Exclusion and the Hierarchical Fast Mode Prediction that exclude more than 70% of the possible coding modes even before starting the fast RDO-MD and ME while keeping the bit rate and distortion loss imperceptible (see Section VI). The ultimate goal of our ACCoReS is to predict exactly one MB Type and one corresponding near-optimal mode in the best case, thus skipping the complete RDO-MD and ME. Therefore, the computational load is significantly decreased. Up to the best of our knowledge, none of the state-of-the-art has done it before.

Our results show that the proposed ACCoReS results in a significant performance improvement (up to 19x for QCIF and 14.5x for CIF videos). This benefit comes at the cost of an average 3% PSNR loss and insignificant (compared to the benefit) processing overhead of computing video frame statistics (as we will see in Section D).

Paper Organization: Section II presents the related work. Section III presents a case study for video analysis considering the HVS properties. Section IV presents our ACCoReS in detail followed by the QP-based thresholding in Section V. Section VI presents the results and evaluation followed by conclusion in Section VII.

II. RELATED WORK

The fast RDO-MD algorithms either simplify the used cost function or reduce the set of candidate modes iteratively depending upon the output of the previous mode computation. [10] uses Mean Absolute Difference (MAD) of MB to reduce the number of candidate block types in ME. On average, it processes 5 out of 7 block types. [23] uses the RD-cost of neighboring MBs to predict the possible coding mode for the current MB. Similar approach is targeted by [24] and [26] that use the residue texture or residue of current and previously reconstructed MB for fast P-MB RDO-MD. [25] uses the mode information from previous frame to predict the modes of MBs in the current frame. [22] provides a fast SKIP and P16x16 prediction as an early predicted mode option. In [11], smoothness and SAD of the current MB are exploited to extend the Skip prediction and exclusion of smaller block mode types. Even if all conditions are satisfied, still 152 out of 168 RD costs are evaluated (Luminance component only), else all RD costs are evaluated as the exhaustive RDO-MD.

[12] exploits the local edge information by creating an edge map and an edge histogram for fast I-MB RDO-MD. Using this information, only a part of available I-MB modes are chosen for RDO, more precisely, 416x4 and 2 out of 4 116x16 out of 36 116x16 are chosen. The fast I-MB scheme in [13] uses partial computation of the cost function and selective computation of highly probable modes. 416x4 blocks are down-sampled and the predicted cost is compared to variable thresholds to choose the most probable mode.

A limited work has been done that jointly performs fast MD for both I-MB and P-MB. In [14], a scalable mode search algorithm is developed where the complexity is adapted jointly by parameters that determine the aggressiveness of an early stop criteria, the number of re-ordered modes searched, and the accuracy of ME steps for the P-MB modes. At the highest complexity point, all P-MB and I-MB modes are processed with highest ME accuracy. [17] proposes a scalable fast RDO-MD for H.264 that uses the probability distribution of the coded modes. It prioritizes the MB coding modes such that the highly probable modes are tried first, followed by less probable ones.

Unlike the related work, our scheme performs an extensive mode-exclusion before fast RDO-MD and ME thus providing a significant reduction in the computational complexity. Our proposed scheme in most of the cases (up to 70%) skips the complete RDO process and predicts the near-optimal coding mode and MB Type (see Section VII) that up to the best of our knowledge, related work have not achieved yet. We will now present the analytical case study for finding the important spatial and temporal video statistics as used by ACCoReS.

III. ANALYTICAL CASE STUDY OF VIDEO SEQUENCES

Although the digital image and video processing fields are built on a foundation of mathematical and probabilistic formulations, human intuition and analysis play the central role in the choice of one technique vs. another [3]. Therefore, important properties of the Human Visual System (HVS) are considered in the scope of work to account for subjective quality. Some important HVS properties are as follows (see [3], [4], and [5] for details):

a) The perceived brightness is a function of contrast and light intensity. Visual system tends to overshoot and undershoot at the boundary of regions of different intensities.

b) The total range of distinct intensity levels that an eye can discriminate simultaneously is rather small when compared with the total adaptation range. Below that level, all stimuli are perceived as featureless or indistinguishable between the extremes.

c) The Human eye is more sensitive to brightness compared to color.

d) At low levels of illumination, vision is carried out by activity of the Rods (part of the human eye; they are not involved in color vision), therefore, under low ambient light human eye can only extract the luminance information.

e) Moving objects capture more attention of the eye compared to the stationary objects.

We have carried out an extensive investigation of several video sequences [9] to subjectively learn the HVS response to different statistics of video frames and their corresponding coding modes. Fig. 2 shows the coding mode distribution (P-MBs in green and I-MBs in purple) for the 7th frame of American Football sequence encoded with the fast RDO-MD using JM13.2 software [2]. This analysis revealed that MBs with high texture and fast motion (e.g. fast moving players) are more probable to be encoded as 16x4 or P8x8 and blow. On the contrary, homogeneous or low-textured MBs with slow motion (e.g. grassy area) are more probable to be encoded as P16x16, P16x8, or P8x16 because the Motion Estimation (ME) has high probability to find a good match. Similar behavior was found in various other video sequences leading to the conclusion that majority of coding modes of a video frame can truly be predicted using spatial and temporal statistics of the current and previous video frames.

From our analytical study, we have learnt that five primitive characteristics of a video frame are sufficient to categorize an MB, thus to predict a probably correct coding mode. The decision of which video frame property to choose can be made considering the tradeoff between computational overhead and the provided precision in the early mode-prediction:

a) Average Brightness is used to categorize an MB as dark or bright. It is the average of luminance values \( I(i,j) \) of an MB (Eq. 1).

\[
\mu_{I} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} I(i,j) \geq 8 \quad (Eq. 1)
\]

b) Contrast is the difference in visual properties that makes an object distinguishable from the background and other objects. In our scheme – due to its simplicity – we have used a modified version of Michelson Contrast [6] as shown in Eq. 2.

\[
C_{MB} = \max_{0 \leq (i,j) \leq 16} \{ I(i,j) - \min_{0 \leq (i,j) \leq 16} \} \geq 8 \quad (Eq. 2)
\]

c) Variance is a measurement for statistical dispersion (Eq. 3), thus it is used as a descriptor of smoothness or measurement of texture. If all samples have the same brightness, then it is a flat/smooth area and the corresponding Variance is zero.

\[
\sigma_{MB}^2 = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (I(i,j) - \mu_{I})^2
\]

d) Gradient: Gradient is defined as the rate of change of luminance. In our case, it measures the average rate of change of luminance over a whole 16x16 MB, vertically (Gx) and horizontally (Gy). Therefore, it is regarded as an approximation of texture. The first order Gradient (G) along a particular direction is approximated by using the difference between two pixel along that direction (Eq. 4).
### Text and Edges
In addition to Gradient, a more precise edge detection — operating on a finer granularity — is required to predict the smaller coding modes more precisely. A Sobel Edge Filter is applied to obtain the magnitude and the direction of edges for every 4x4 sub-block. The Sobel Edge Filter has the advantage of providing both differentiating and smoothing effect. The total edge values for a 4x4 sub-block, 8x8 block, and 16x16 MB are computed using Eq. 5.

\[
G_i = \left( \sum_{j=1}^{4} \sum_{k=1}^{4} \sqrt{G_{xx}^2 + G_{yy}^2} \right) \cdot \frac{1}{2^{i-1}}
\]  
(Eq. 4)

The direction angle (in degrees) with respect to the x-axis is calculated as \( \alpha = \tan^{-1}(G_y / G_x) \). It is used to classify an edge into one of the following four directional groups (Fig. 4).

### Step 1: HVS-Based Macroblock Categorization
Video Frame Statistics based Categorization:

- **Group-A**: High-textured MB containing medium to fast motion
- **Group-B**: Flat, homogeneous regions with slow motion

**A. Step 1:** HVS-based categorization of MBs is performed using the spatial and temporal video statistics. The QP-based thresholds (as discussed in Section V) are used for this categorization.

**B. Step 2:** Prognostic Early Mode Exclusion for I-MB and P-MB coding modes is incorporated that excludes the highly unwanted modes. In many cases the curtailed set of modes is left with either I-MB or P-MB modes, especially for slow-motion sequences.

**C. Step 3:** Hierarchical Fast Mode Prediction further analyzes the curtailed set of modes and provides a set of candidate coding modes.

**D. Step 4:** In the last step, Sequential RDO Mode Elimination is done. It processes the candidate coding modes one-by-one starting from the bigger partitions. After a mode is processed, it is evaluated for the termination condition or to exclude further irrelevant modes. Now we will explain each processing stage of ACCoReS in detail.

### Average Brightness \((\mu_{SB})\): very dark \((\mu_{1})\), dark \((\mu_{2})\), bright \((\mu_{3})\), very bright \((\mu_{4})\)

### Contrast \((C_{SB})\): low \((C_{1})\), high \((C_{2})\) contrast

### Variance \((\sigma_{SB})\): very low \((\sigma_{1})\), low \((\sigma_{2})\), high \((\sigma_{3})\) variance

### Gradient \((G_{SB})\): very low \((G_{1})\), low \((G_{2})\), high \((G_{3})\) gradient

**Edge Map:** low \((S_{1})\), highly \((S_{2})\) edged

Combinations of the above-defined categories are used to predict the MB content characteristics as follows:

\[
H_{high_{MB}}^{Textured} = \left( S_{H} \land V_{H} \right) \lor \left( S_{H} \land G_{H} \right) \lor \left( G_{H} \land V_{H} \right)
\]

\[
S_{StrongThick}^{MB} = \beta G_{H} \land \beta S_{H} \land \gamma G_{H} \land \gamma S_{H} \land \mu \mu \mu \mu
\]

\[
S_{ManyThin}^{MB} = \beta S_{H} \land \gamma G_{H} \land \gamma V_{H} \land \mu \mu \mu \mu
\]

### Directional Statistics:
An edge direction is called dominant if the edge sum belonging to an edge direction group \( \gamma \) (see Fig. 3) significantly contributes to the total edge sum of this MB.

\[
E_{Dir}^{Dominant} = \begin{cases} 1, & \alpha > \epsilon \cdot \mu_{S_{MB}}; \ i \in \{0, 1, 2, 3\} \\ 0, & \text{Otherwise} \end{cases}
\]  
(Eq. 7)

**Motion-Field Statistics** are obtained using the motion characteristics of the neighboring MBs as follows:

\[
E_{Dir}^{H_{MB}} = \begin{cases} S_{i} > 0.5 \cdot S_{MB}, & \text{EDir}^{Dominant} \equiv 1 \\ 0, & \text{Otherwise} \end{cases}
\]  
(Eq. 8)

**Coding-Mode-Field Statistics** are obtained considering the coding modes of the spatial (in the current frame F_i) and temporal (in the previous frame F_{i-1}) neighboring MBs encoded as an I-MB.

\[
\begin{align*}
INB_{Spatial}^{MB} &= isI\left(MB_{Fi} \cdot MB_{Fi}^{1} \cdot MB_{Fi}^{2} \cdot MB_{Fi}^{3}\right) \\
INB_{Temporal}^{MB} &= isI\left(MB_{Fi}^{1} \cdot MB_{Fi-1} \cdot MB_{Di} \cdot MB_{Di-1} \cdot MB_{Di-1}^{1} \cdot MB_{Di-1}^{2}\right) \\
INB_{TemporalTotal}^{MB} &= isI\left(MB_{Fi}^{1} \cdot MB_{Fi}^{1} \cdot MB_{Fi-1} \cdot MB_{Di} \cdot MB_{Di-1} \cdot MB_{Di-1}^{1} \cdot MB_{Di-1}^{2}\right) \\
\end{align*}
\]  
(Eq. 9)

**B. Step 2:** Prognostic Early Mode Exclusion

The Prognostic Early Mode Exclusion scheme starts with a classification of MBs into two distinct groups using Eq. 10:

\[
\begin{align*}
A_{(High_{Trans} \land SAD_{Collected} \land Th_{SA})} &= \left( \mu_{High_{Trans}} \land \mu_{SAD_{Collected}} \land \mu_{Th_{SA}} \right) \\
B_{(Low_{Trans} \land SAD_{Collected} \land Th_{SA})} &= \left( \mu_{Low_{Trans}} \land \mu_{SAD_{Collected}} \land \mu_{Th_{SA}} \right)
\end{align*}
\]  
(Eq. 10)

Fig. 5 and Fig. 7 present the pseudo-codes of Prognostic Mode Exclusion for both Group-A and Group-B, respectively. In case of Group-A, P16x16 is excluded (line 3) due to high texture and the best choice would most probably be P8x8 or 14x4. However, exclusion of P16x16 at this point is critical as a wrong exclusion may result in a significantly increased bit rate. Therefore, the exclusion decision of P16x16 is performed in the Hierarchical Fast Mode Prediction step. Lines 4-7 and 8-11 check for slow motion using the motion statistics of the spatial neighboring MBs and exclude the smaller block partitions and 14x4 (line 5, 9). Lines 12-15 detect a high texture and hectic motion region. In this case, 14x4 coding mode is selected and all other modes are excluded.

In case of Group-B, a more sophisticated scheme systematically excludes the most unlikely modes. Lines 3-5, 6-12, 13-27 check for slow motion, flat and homogeneous region, respectively. In these cases, 14x4, P8x8 and smaller partition modes are excluded. If a homogenous MB is stationary, P16x16 is predicted to be the most probable coding mode; otherwise, 11x16 is additionally processed (line 8). Lines 15-18, 19-25, 28-31 detect low motion and low-to-moderate textural to exclude 14x4 mode; otherwise, 14x4 mode is re-enabled to avoid significant visual quality loss. Lines 33-39 assure that modes with smaller block partitions are only excluded if low motion and/or low texture are detected.
If all modes except P16x16 are already processed, only the P16x16 is processed. Hence, if the main edge direction is determined to be horizontal or vertical, P16x8 or P8x16 block type is chosen, respectively.

**Sub-P8x8 Mode Prediction:** Based on the assumption "the pixels along the direction of local edge exhibit high correlation, and a good prediction could be achieved using those neighboring pixels that are in the same direction of the edge", the main edge direction is investigated to split the MB accordingly.

**Skip Mode Prediction:** If SAD of an MB in P16x16 mode is significantly low, a perfect match could be very well predicted by ME. Such MBs are highly probable to be SKIP, thus saving complete ME computational load. Similarly, if the collocated MB is highly correlated with the current MB, then the probability of SKIP is very high e.g., the complete region is homogeneous.

---

**GROUP-A:** High-textured MB containing medium-to-fast motion

1. M = {P16x16, P16x8, P8x16, P8x8, P4x4, P4x4, 116x16, 14x4}

2. \(M \leftarrow \{116x16\};\) // Include 116x16

3. If \((SAD_{Spatial}^\mu_b < \delta_b^\mu_b \text{ Th})\) Then

4. \(M \leftarrow \{P8x8, P8x4, P4x8, P4x4\};\) // Exclude 14x4, P8x8 and below

5. return;

6. Go to Step-3 (Section C)

7. End If

8. If \((SAD_{Spatial}^\mu_b < \delta_b^\mu_b \text{ Th})\) Then

9. \(M \leftarrow \{14x4\};\) // Exclude 14x4

10. return;

11. Go to Step-3

12. If \((\mu_b \in \{P8x8, P8x4, P4x8, P4x4\};\) // Exclude P8x8 and below

13. return;

14. Go to Step-3

15. End If

16. End return;

---

**GROUP-B:** Flat, homogenous regions with slow-to-medium motion

1. M = {P16x16, P16x8, P8x16, P8x8, P4x4, 116x16, 14x4}

2. \(M \leftarrow \{116x16\};\) // Include 116x16

3. If \((SAD_{Spatial}^b < \delta_b^\mu_b \text{ Th})\) Then

4. \(M \leftarrow \{P8x8, P8x4, P4x8, P4x4, 14x4\};\) // Exclude 14x4, P8x8 and below

5. return;

6. Go to Step-3

---

**Step-3:** Hierarchical Fast Mode Prediction

Our Hierarchical Fast Mode Prediction (Fig. 6) performs a refined second-level mode exclusion to obtain a set of candidate coding modes, which is later evaluated by the RDO-MD process with an integrated Sequential RDO Mode Elimination mechanism.

**P16x16 Mode Prediction:** If all modes except P16x16 are already excluded, then P16x16 is processed unless SKIP mode is detected in the last step of Fig. 6. On the contrary, P16x16 is excluded if the MB has fast motion and high texture.

**P16x16, P16x8, P8x16 and P8x8 Mode Prediction:** Based on the assumption "the pixels along the direction of local edge exhibit high correlation, and a good prediction could be achieved using those neighboring pixels that are in the same direction of the edge", the main edge direction is investigated to split the MB accordingly. Hence, if the main edge direction is determined to be horizontal or vertical, P16x8 or P8x16 block type is chosen, respectively. A very small edge sum points out the presence of a homogenous region, so only the P16x16 is processed.

**Sub-P8x8 Mode Prediction:** In case the SAD of the neighboring MBs is too high, P4x4 mode is predicted. In case the dominating horizontal or vertical edge direction is detected, P8x4 or P4x8 partition is selected, respectively.

**Skip Mode Prediction:** If SAD of an MB in P16x16 mode is significantly low, a perfect match could be very well predicted by ME. Such MBs are highly probable to be SKIP, thus saving complete ME computational load. Similarly, if the collocated MB is highly correlated with the current MB, then the probability of SKIP is very high e.g., the complete region is homogeneous.
Moreover, if the MB lies in a dark region, the human eye cannot perceive small brightness variations. Thus, the insignificant distortion introduced by a forceful SKIP is tolerable here.

D. Step-4: Sequential RDO Mode Elimination

An integrated Sequential RDO Mode Elimination mechanism re-evaluates the candidate coding modes for sequential elimination, i.e. after P16x16 is processed, P16x8, P8x16, P8x8, and below are re-evaluated for elimination as specified in Fig. 6. However, for Sequential RDO Mode Elimination, the spatial SDV and MV values are replaced by the actual SAD and MV of the previously evaluated mode.

V. QP-BASED THRESHOLDING

As discussed in Section A, QP-based thresholds are used to categorize different features of video frames. For higher QP values, the effect of texture and motion becomes bluer due to the increased number of zero coefficients. It follows the fact that finding a good prediction is easier for ME, thus the number of injected I-MBs decreases. Therefore, with changing QP values, the thresholds (related to the decisions operating on the referenced frames) need to be adapted. We have performed extensive experimentation using different QPs (12 to 40) and several video sequences (only a small subset all of sequences used for validation in Section IV) to evaluate these thresholds. Afterwards, we have performed polynomial curve fitting using MATLAB to obtain threshold equations as a function of QP, see Eqs. 11-12. Our empirical analysis revealed that only the thresholds for SADs are similar to MV (thus the major characteristics for motion and texture detection) react to the changing QPs. Table 1 presents the remaining thresholds (which are not affected by changing QPs).

\[
TH_{SAD} = \begin{cases} 
2500, & QP_{prev} < 20 \\
9000, & QP_{prev} \geq 20 
\end{cases}
\]

\[
TH_{MV} = \begin{cases} 
-0.3QP_{prev} + 38, & 3QP_{prev} - 115, \quad QP_{prev} \geq 497, \quad \text{Otherwise} \\
10000, & QP_{prev} < 20 \\
13000, & QP_{prev} \geq 28 
\end{cases}
\]

\[
TH_{MV}^{(M/MV/4/MV)} = \begin{cases} 
(20, 45, 30), & QP_{prev} \leq 28 \\
(30, 55, 40), & QP_{prev} > 28 
\end{cases}
\] (Eq. 12)

VI. RESULTS AND EVALUATION

For evaluation and validation of our ACCoReS, we compare it with several state-of-the-art fast RDO-MD schemes and exhaustive RDO-MD. Common test conditions are: JM 13.2, IPPP, 1 reference frame, search range = 16. We have encoded various QCIF and CIF video sequences (low to fast motion) with different QPs (12, 16, 20, 24, 28, 32, 36, and 40) using an Intel 6600 (2.4 GHz, 2GB RAM, Windows XP) PC. Note: All speedup results include the overhead of ACCoReS and computation of video statistics in software.

A. Comparison with State-of-the-Art RDO-MD Schemes

We have compared our ACCoReS with several state-of-the-art fast RDO-MD schemes for quality (a positive ΔPSNR shows PSNR loss) and performance using the similar coding conditions as specified by the corresponding scheme. Table 2 shows that, compared to state-of-the-art approaches ([10], [22]-[26]), ACCoReS achieves a speedup of up to 9.14x (average 3.05x) at the cost of an average PSNR loss of 0.66 dB. The significant speedup comes from the Prognostic Early Mode Exclusion and Hierarchical Fast Mode Prediction that curtails the set of candidate coding modes for further evaluation.

B. Comparison with Exhaustive RDO-MD

Table 3 provides the comparison (average and maximum) of ACCoReS with the exhaustive RDO-MD for distortion, bit rate (a positive ΔBit Rate shows the bit rate saving) and speedup. Each result for a sequence is the summary of 8 encodings using different QP values. The average PSNR loss is approximately 3%, which is visually imperceptible. However, our ACCoReS provides a significant reduction in the computational complexity i.e. performance improvement of up to 19x (average 10x) compared to the exhaustive RDO-MD. The major speedup comes from slow motion sequences (Susie, Hall, Akiyo, Container, etc.) as smaller block partitions and I-MB coding modes are excluded in the Prognostic Early Mode Exclusion stage.
requirements are (#statistics)*#MBs*16bits, where (#\text{statistics} + \#\text{temporal statistics} = 5 + 2). Note: Similar to fast RDO-MD schemes, ACCoReS (Fig. 5, Fig. 7, Fig. 8) is implemented in software for flexibility (as they are not fixed by the standard).

VII. CONCLUSION

We have presented a novel HVS-based Adaptive Computational Complexity Reduction Scheme (ACCoReS) that performs Prognostic Early Mode Exclusion and Hierarchical Fast Mode Prediction to curtail the set of possible coding modes. Compared to exhaustive RDO, our ACCoReS provides a performance improvement of up to 19x (average 10x) with an average 3% PSNR loss. ACCoReS excludes more than 70% of the possible coding modes even before starting the RDO-MD and ME. Our scheme is especially beneficial for low-cost performance and/or power-critical embedded systems where the available computational resources are limited. Our proposed scheme is quick and easy to be deployed in the real-world video encoding applications and exhibit a great industrial potential.

VIII. REFERENCES