Runtime-Optimised Intra-4×4 Mode-Decision for H.264/AVC Video Encoding

Harald Jordan[1], Florian H. Seitner[2], Michael Bleyer[2] and Margrit Gelautz[2]
Vienna University of Technology
[1]Institute of Automation and Control
Gusshausstrasse 26, A-1040 Vienna, Austria
jordan@acin.tuwien.ac.at
[2]Institute for Software Technology and Interactive Systems
Favoritenstrasse 9-11, A-1040 Vienna, Austria
{seitner;bleyer;gelautz}@ims.tuwien.ac.at

Abstract

We describe a method that considerably improves the computational behaviour of H.264 Intra-only encoders. Such Intra-only encoders come to use in video-cutting and low-latency video coding where temporal prediction via using Inter-frames is no feasible option. We identify the spatial prediction step as the computational bottleneck in Intra-only encoders. In this step, the encoder tests various modes that represent predictions of the current macroblock's or sub-macroblock's texture from spatial neighbouring pixels in order to find the mode of lowest residuum. Unfortunately, testing the complete set of allowed modes is computational expensive. However, as is demonstrated by an analysis provided in this paper, it is reasonable to assume that a large percentage of blocks preserve their prediction modes over time. Based on this assumption we develop two algorithms that improve the computation time in the prediction step. These algorithms differ by their criteria used to decide whether a block's coding mode can be propagated from a temporal preceding frame. Computational speed is enhanced, since we test the full set of modes only for blocks that fail these criteria. Experimental results show that our methods considerably improve the execution time of an Intra-only encoder and only show small impact on data-rate and image quality.

1. Introduction

H.264 represents a powerful, but also challenging video compression standard. It achieves its excellent compression performance by using various advanced encoding techniques. Compared to previously defined standards, H.264 accomplishes lower bitrates while simultaneously increasing the image quality. This is achieved by introducing improved coding tools for reducing the temporal and spatial redundancies. The respective coding modes are called Intra (spatial) and Inter (temporal) modes. However, the price for the improved coding quality is an increased computational demand that often makes H.264 real-time encoding intractable on many hardware configurations.

In this paper, we focus on Intra-only coding. Typical scenarios for Intra-only encoders are represented by video-cutting and low-latency applications. Here, the requirement for low latencies prohibits the use of temporal prediction (Inter-frames). The computational bottleneck in Intra-only encoders is formed by the spatial prediction step. It is therefore reasonable to focus on this prediction step in order to enhance the encoder's run-time behaviour, which also represents the strategy followed in this paper. An overview of the H.264 spatial prediction method is given in the following two paragraphs. The reader is referred to [7] for further details.

The H.264 video standard partitions the frame into so-called macroblocks (MB). A MB thereby comprises 16×16 pixels and can further be partitioned into 16 sub-macroblocks of 4×4 pixels (4×4-block). In the spatial prediction step, a prediction block is formed by using already encoded pixels in the spatial proximity of the current block. H.264 defines various possibilities for building a prediction, which are denoted as Intra-modes. In the H.264 standard, there exist four coding modes for MBs and nine for the 4×4-blocks. These coding modes differ in that they use different directions along with the texture pattern which is extrapolated from already encoded pixels. The directional modes for 4×4-blocks are depicted in Figure 1.

To judge the quality of a prediction (or coding mode), the encoder computes the Sum of Absolute Differences (SAD) in intensity values between the predicted block and the original uncompressed block. We refer to the resulting value as SAD-costs or prediction error. The coding mode that is finally selected is the one that shows minimum SAD-costs. To find this "optimal" mode, it is required to calculate the prediction error for each existing Intra-mode. This procedure is commonly denoted as mode-decision. It is important to notice that this mode-decision procedure represents the bottleneck in common Intra-only encoder implementations. For our encoder, we have measured that approximately 50 percent of the encoding time is consumed by this operation. When considering the mode-decision procedure separately from the other encoder functions, we have found that over 80 percent of the processing time is consumed by testing the nine modes for the 4×4-blocks and less than 20 percent is needed to test the four MB modes. We have therefore decided to solely focus on optimising the run-time of

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.

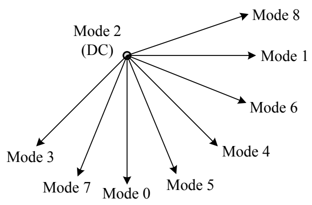


Figure 1. The nine Intra-4×4 coding modes supported by H.264. Eight directional modes and one DC-mode are defined.

4×4-block predictions, as this yields the highest potential for speeding up the encoder.

In the context of prior work, Fritts et al. [2] describe a method for reducing the number of Intra-4×4 prediction modes. Instead of checking all nine modes, only a sub-set of the most probable modes are tested. A similar approach is described in [1]. The authors of [5] present a fast Intra-mode selection strategy based on a prediction method that uses various MB properties. These properties are required for selecting a sub-set of candidate modes. The work of [3] and [8] simplifies the mode decision by using the transformation coefficients. Exploitation of spatial properties of stationary MBs within P-slices is described in [6] and [4]. These methods reuse Intra-modes at co-located positions in a given reference-frame. The approach presented in [9] compares the pixels of temporally adjacent MBs to judge the reuse of coding modes.

Our method extends the idea of [9] by exploiting temporal as well as spatial relationships between co-located MBs of consecutive frames. We further attempt to reuse already computed intermediate results (e.g. SADs) for optimising the runtime-behaviour. Our major assumption is that a large number of MBs keep their coding modes in the next frame. We justify this assumption via empirical analysis (see Section 2). Based on this analysis, we develop two new methods that exploit temporal and spatial characteristics for improving the run-time behaviour of the Intra-4×4 prediction while maintaining similar data-rates and image qualities (see Section 3). This is demonstrated by our experiments in Section 4.

2. Analysis

This section analyses the encoding behaviour of the Intra-4×4 coding modes for various sequences. The aims are to evaluate the validity of three assumptions (A1-A3):

A1 A 4×4-block that retains its coding mode over time will also have similar SAD-costs.

A2 Static 4×4-blocks tend to retain their coding mode between consecutive frames (i.e. the most efficient coding mode typically outperforms the other modes significantly). A 4×4-block is static if no motion occurs at this frame-position.

A3 Spatially adjacent 4×4-blocks (e.g. the left and upper 4×4-blocks) tend to have similar coding modes.

These assumptions form the starting point of the algorithms described in Section 3. We derive them from the following theoretic considerations. Except for sensor-related effects, a static 4×4-block should not change significantly over time. Therefore, the SAD-costs of the 4×4-block should remain nearly constant (*A1*). Furthermore, the changes in this 4×4-block are typically too small for causing alternating coding modes (*A2*). For the Intra-4×4 prediction we exploit the fact that spatially adjacent pixels in a picture are typically correlated. This implies that neighbouring 4×4-blocks often contain similar texture patterns and can be efficiently described by similar Intra coding modes (*A3*).

2.1. Test Sequences

We have selected a set of eight CIF sequences containing different amounts of motion and texture. The sequences have been captured in progressive format at a capturing rate of 25 frames per second. We encoded the sequences with a H.264/AVC baseline-profile encoder using three different quantisation values (QP). Figure 2 shows the resulting PSNR quality and the bitrates of the encoded sequences.

2.2. Evaluation

In this section, we evaluate the three assumptions *A1*-*A3*. For examining assumption *A1*, we analyse the SAD-cost changes of co-located 4×4-blocks with constant coding mode. Figure 3 shows the results for the eight test sequences.

We see that for around 50 percent of the 4×4-blocks the SAD-costs tend to change by less than 25 percent. This result indicates that assumption *A1* is valid for more than half of the static 4×4-blocks in the test sequences. For 65 percent of the 4×4-blocks, the changes are below 50 percent.

Figure 4 shows the percentage of co-located 4×4-blocks with equal coding modes. It can be seen that on average approximately 50 percent of the coding-modes remain unchanged between two consecutive frames. In sequences with little motion and a high amount of static regions, 4×4-blocks tend to keep their coding mode. For sequences with a high amount of motion, typically a lower percentage of 4×4-blocks with constant coding can be observed. Motion at sub-pixel level causes frequent textural changes in the 4×4-blocks and consequently alternating coding modes. However, moderate motion for homogeneous regions does not necessarily lead to coding mode changes. This can be seen in the high percentage of constant 4×4-block coding for strongly dynamic sequences such as *f1car*. According to our observations, assumption *A2* is valid for a major part of the static 4×4-blocks. Homogeneous regions in dynamic sequences show a similar coding behaviour. Exploiting this knowledge allows the optimisation heuristics in Section 3 to cope with dynamic sequences in an efficient way. For examining assumption *A3*, we compared the coding mode

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.

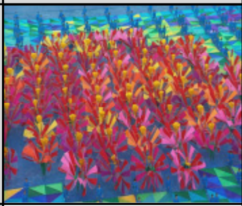
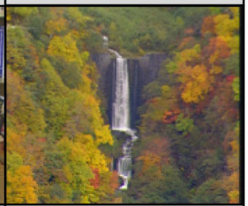
Sequence	Barcelona	Bus	Canoe	F1Car
				
PSNR [dB]	33 / 29 / 26	34 / 30 / 27	35 / 31 / 28	34 / 31 / 28
Bitrate [bpp]	2.1 / 1.3 / 0.8	1.1 / 0.7 / 0.4	1.0 / 0.6 / 0.4	1.1 / 0.7 / 0.3
Motion	low	medium	high	high
Texture	high	medium	medium	medium
Sequence	Mobile	Mother	Paris	Waterfall
				
PSNR [dB]	33 / 29 / 25	38 / 34 / 31	35 / 31 / 27	34 / 30 / 28
Bitrate [bpp]	2.0 / 1.3 / 0.8	0.5 / 0.4 / 0.2	1.2 / 0.8 / 0.5	1.1 / 0.6 / 0.3
Motion	medium	low	low	low
Texture	high	low	medium	low

Figure 2. Seven test-sequences at CIF resolution using Intra-4×4 encoding and QP=30/35/40. Rows 2 and 3 show the Peak Signal-to-Noise Ratio (PSNR) and the resulting data-rate. For each 4×4-block of a sequence, the motion has been measured. In row 4, the average of these motion vectors is given. The last row describes the texture contained in each sequence.

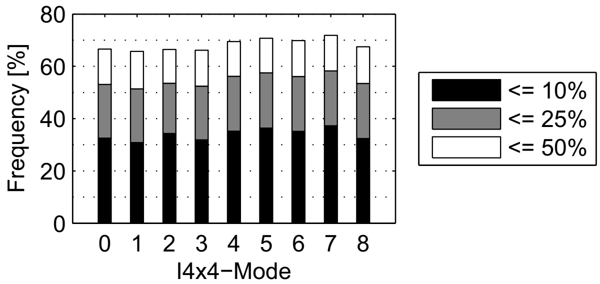


Figure 3. Changes in the SAD-costs of co-located 4×4-blocks with constant coding mode. We have divided the 4×4-blocks of each sequence into three groups: 4×4-blocks, with SAD changes between 0 and 15 percent, between 15 and 25 percent, and 25 and 50 percent. The frequency of the three groups is shown for the eight sequences.

of each constant 4×4-block to the coding mode of its left and upper neighbour 4×4-blocks (Figure 5). For calculating the correlation between the coding mode of the current 4×4-block and the upper/left 4×4-blocks, four alternative sequences have been used.

This ensures independent results when analysing the heuristics in Section 3 with the eight test sequences. These sequences are *carphone* (high motion, medium texture), *claire* (low motion, low texture), *coastguard* (high motion, high texture) and *news* (medium motion, medium texture). Figure 5 shows the statistical correlation of the three 4×4-blocks. We can see that the coding modes of neighbouring

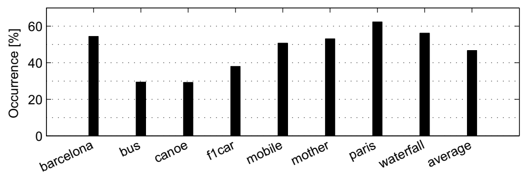


Figure 4. Percentage of co-located 4×4-blocks with equal coding: The percentage of 4×4-blocks with equal intra coding mode between consecutive frames is shown for each test-sequence for QP=25.

4×4-blocks are only weakly correlated. For most cases the best coding mode for the current 4×4-block can be predicted with a certainty of 25 to 30 percent. However, if the left/upper neighbour 4×4-blocks are both using the same coding modes in the range of 0 to 2, the current 4×4-block uses the same coding mode with a probability of about 65 percent. This implies that assumption *A3* is only partially valid for certain conditions. However, we can exploit these conditions in our optimisation heuristics.

3. Algorithms for Lowering the Complexity of the Intra-Mode Decision

For lowering the complexity of the Intra-mode decision, the proposed algorithms attempt to reuse the encoding information of previously encoded 4×4-blocks. It is therefore required to develop heuristics for deciding whether to keep the coding mode or not.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.

Mode of left 4x4-block \ Mode of upper 4x4-block	0	1	2	3	4	5	6	7	8
0	0 (65)	0 (35)	0 (35)	0 (30)	0 (35)	0 (40)	0 (35)	0 (40)	0 (35)
1	1 (40)	1 (65)	1 (40)	1 (35)	1 (35)	1 (35)	1 (40)	1 (40)	1 (45)
2	2 (30)	1 (35)	2 (65)	2 (30)	2 (25)	2 (30)	1 (30)	2 (35)	1 (30)
3	0 (25)	1 (30)	2 (25)	2 (25)	2 (25)	2 (30)	1 (25)	2 (25)	8 (25)
4	0 (25)	1 (30)	2 (25)	1 (25)	6 (25)	0 (25)	1 (25)	0 (25)	1 (25)
5	0 (30)	1 (25)	0 (25)	0 (25)	0 (25)	5 (30)	1 (25)	0 (25)	8 (25)
6	0 (25)	1 (35)	1 (25)	1 (25)	1 (25)	1 (25)	6 (35)	1(25)	1 (30)
7	0 (30)	1 (25)	2 (25)	7 (25)	0 (25)	0 (25)	1 (25)	7 (25)	8 (25)
8	1 (25)	1 (35)	1 (30)	1 (25)	1 (30)	1 (25)	1 (30)	1 (30)	1 (30)

Figure 5. Correlation between the coding mode of the current 4×4-block and the coding modes of its left/upper neighbours: The table shows the most probable mode of the current 4×4-block and the probability for this mode. Example: When using Mode 0 for coding the left and upper 4×4-blocks, the most probable mode for the current 4×4-block is also mode 0 with a probability of 65 percent.

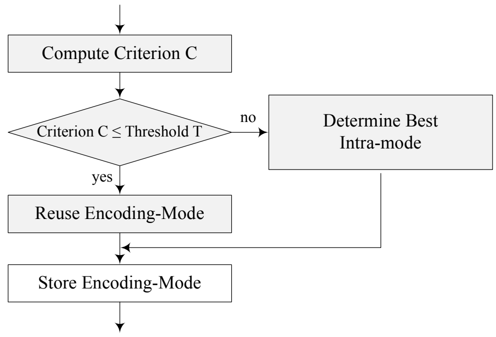


Figure 6. The work-flow of the proposed algorithm: Using a criterion C and a threshold T we decide if the previously used coding mode of a 4×4-block is reused or a full intra mode-decision search is started.

We present simple methods for reusing Intra-4×4 coding modes based on SAD-cost changes in Section 3.1. We extend these heuristics into a more sophisticated algorithm in Section 3.2.

3.1. Single Criterion Mode Reusage

Based on the conclusions of Section 2, we can assume that the SAD-costs provide a good measure for the 'stability' of a 4×4-block's coding mode. The decision whether to reuse the previous coding mode is based on the SAD-cost differences of two co-located 4×4-blocks according to the following equation:

$$SAD_{ratio} = \left| 1 - \frac{SAD_t}{SAD_{t-1}} \right| \tag{1}$$

In this equation SAD_t and SAD_{t-1} are the 4×4-block's SAD values at frame t and $t-1$, respectively. For the case that the SAD ratio is below a predefined threshold (e.g. 10 percent), the previously determined coding mode is reused:

$$reuse\_mode = \begin{cases} 1 & \text{if } SAD_{ratio} \leq T \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Otherwise, we have to determine the correct Intra-4×4 coding mode by using a full mode-decision search. Figure 6 visualizes this decision strategy.

3.2. Multi Criterion Mode Reusage

We have experimented using the single-criterion method and found that the coding mode reusage indeed lowers the computational complexity at the cost of a negligible algorithmic overhead. In this section, an extended method is presented which incorporates spatial as well as temporal features. Instead of the single-criterion shown in Figure 6, a weighted combination of multiple criteria decides on the reusage of a 4×4-block's coding mode. This results in a more stable and coding efficient reusage of Intra-4×4 coding modes.

The proposed decision stage uses four different decision criteria (c_0-c_4). These criteria are explained in detail in the Sections 3.2.1 to 3.2.4. Each criterion c_i is weighted by a weighting factor w_i and their combination is compared to a threshold T:

$$reuse\_mode = \begin{cases} 1 & \text{if } \sum_{i=0}^{4} c_i w_i \leq T \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The weighting factors w_i have been determined by observations from the four sequences used for analysing assumption $A3$. It is important to know, that these sequences are different from the test sequences. This allows us to use the test sequences for evaluating our multi-criterion optimisation technique.

3.2.1 Criterion c_0

This criterion exploits the SAD-cost difference between two co-located 4×4-blocks introduced in Section 3.1.

3.2.2 Criterion c_1

Criterion c_1 observes the 4×4-block's coding mode for multiple frames and incorporates long-term knowledge into our algorithm.

$$c_1 = \begin{cases} 1 & \text{if mode constant for N frames} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.

If a coding mode is constant for more than N frames, the criterion indicates a highly static region and votes for the reusage of the previous coding mode. For our evaluation, $N = 3$ has been used.

3.2.3 Criteria c_2 and c_3

These criteria examine the coding mode reusage of neighbouring 4×4-blocks. For a 4×4-block which is surrounded by 4×4-blocks with fixed coding modes, we assume that the coding mode of this 4×4-block is also constant. In larger static regions, these criteria prevent mode changes of single 4×4-blocks which are surrounded by 4×4-blocks with reused coding modes. Criterion c_2 is true if the left 4×4-block reused its previous coding mode:

$$c_2 = \begin{cases} 1 & \text{if left } 4 \times 4\text{-block reused mode} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Criterion c_3 uses the upper 4×4-block instead of the left one. Note that instead of using only the directly neighbouring 4×4-blocks, these two criteria can propagate knowledge about reused 4×4-block codings to more distant 4×4-blocks.

3.2.4 Criterion c_4

Criterion c_4 uses the statistical characteristics observed in Figure 5. We first evaluate if the left and upper 4×4-blocks use the same coding mode and if it is a mode between 0 and 2. If this case occurs, we predict the coding mode of this current 4×4-block using the table from Figure 5 and compare this result with the mode used in the previous frame. If both modes are equal (i.e. reused mode equals predicted mode), criterion c_4 votes in favour of the mode reusage.

4. Experimental Results

We extended a H.264 baseline-profile encoder with the proposed algorithms and tested it on an Intel P-IV 3.0GHz PC with 2GB RAM. Firstly, the test results of the single-criterion approach (see Section 3.1) are presented. In Figure 7, the average values of the achieved data-rates and the mode-reusages are given for various QPs and a threshold T of 10 percent. This figure shows that the average data-rate increase is below five percent. We can observe a maximum increase of about 10 percent. The maximum PSNR-changes observed for all sequences are less than *-0.2dB*. For the human eye this is nearly imperceptible. The reusage of Intra-coding modes was thereby in a range between 10 to 40 percent.

Secondly, Figure 8 provides the test results of the multi-criterion approach. The average data-rate increases are below three percent. However, outliers are also having higher values, as their absolute data-rate is significant smaller compared to the other sequences. The sequence *mother*, for instance, takes less than half of the data-rate of sequence

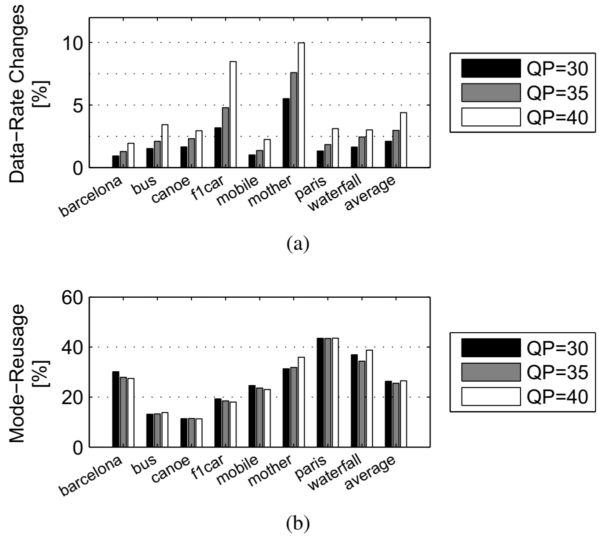


Figure 7. Average increase in the data-rate and mode-reusage using the single-criterion algorithm ($T = 10\%$). (a) The increases of the data-rate compared to the exhaustive Intra-4×4 mode-decision. (b) Percentage of 4×4-blocks with reused coding modes.

canoe. This approach also deteriorates the average PSNR-changes by not more than *-0.02dB*.

The average mode reusage is approximately 30 percent. The reusage for a frame is almost independent of the used QP values. This indicates that the multi-criterion method works well for different bitrates and texture content.

A direct comparison of the implemented algorithms is given in Figure 9. It is shown that both methods achieve similar results, whereas the second approach tends to have better (=lower) data-rates and higher PSNR-values. The second algorithm incorporates knowledge about the coding of neighbouring 4×4-blocks. It applies the mode reusage less aggressively, but achieves lower data-rates than the single-criterion algorithm.

Figure 10 presents the average performance increases when encoding the eight test sequences. Both approaches achieve approximately 20 percent lower encoding times on average and a maximum performance increase of about 40 percent. The multi-criterion approach achieves better data-rates but has a lower mode reusage (i.e. lower run-time performance).

5. Summary

In this paper, we have presented two methods that considerably speed up the run-time of Intra-only H.264 encoders. We have focused on the spatial prediction operation that represents the bottleneck in Intra-only encoders. Our methods for enhancing the run-time of the spatial prediction step are based on the assumption that for a large number of blocks their spatial prediction modes remain constant over time. The validity of this assumption has been

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.

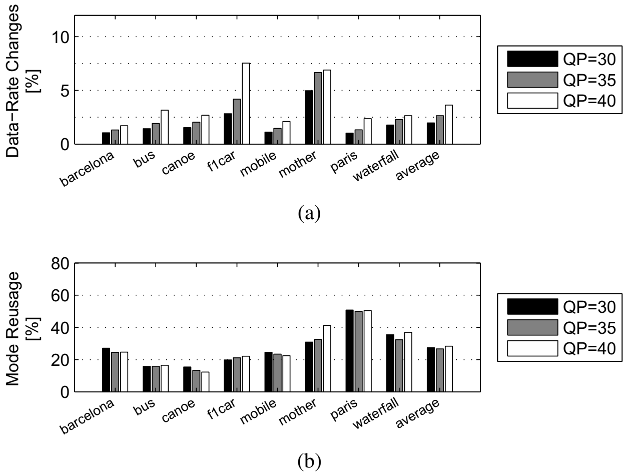


Figure 8. Average data-rate changes and mode-reusage using the multi-criterion algorithm. (a) Data-rates related to the result of an exhaustive Intra-4×4 mode-decision. (b) Reusage of Intra-4×4 coding modes.

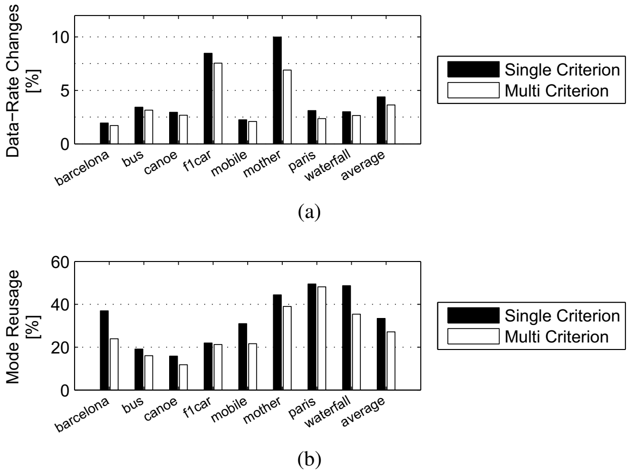


Figure 9. Direct comparison of the two developed algorithms. (a) Increases in the data-rate for both algorithms compared to the exhaustive Intra-mode decision search. (b) Reusage of Intra-4×4 coding modes.

demonstrated on various sequences. Using this assumption we have presented two algorithms that perform the computational expensive Intra-4×4 mode decision only for a subset of the available 4×4-blocks. For the other 4×4-blocks, their modes are propagated from the previous video frame. We have reported in our experiments that this strategy decreases the computational complexity in the range of 20 to 40 percent, while maintaining similar bitrates and image qualities (PSNR).

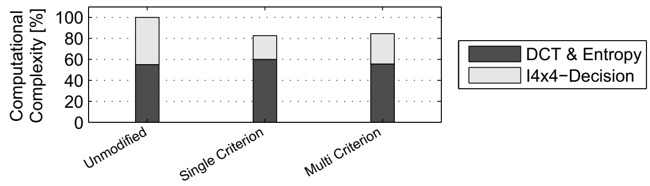


Figure 10. Comparison of the run-time complexity: The figure compares the run-time of the investigated methods and the unmodified encoder. It shows that both methods lower the computational complexity by approximately 20 percent.

6. Acknowledgment

This work has been supported by the Austrian Federal Ministry of Transport, Innovation, and Technology under the FIT-IT project VENDOR (Project No. 812429). Michael Bleyer is financed by the Austrian Science Fund (FWF) under project P19797.

References

[1] A. Elyousfi, A. Tamtaoui, and E. Bouyakhf, ”A new fast intra prediction mode decision algorithm for H.264/AVC encoders”, *Int. Journal of Computer Systems Science and Engineering*, 4 (No. 1), December 2006, pp. 1 – 7.

[2] J. Fritts, H. Zhang, and F. Steiling, ”Fast intra-prediction mode selection for H.264”, *SPIE05*, 2005, pp. 23.

[3] G. Hwang, J. Park, B. Jung, K. Choi, Y. Joo, Y. Oh, and B. Jeon, ”Efficient fast intra mode decision using transform coefficients”, *Advanced Communication Technology*, 1, February 2007, pp. 399 – 402.

[4] M.-C. Hwang, J.-K. Cho, J.-S. Kim, J.-H. Kim, and S.-J. Ko, ”Fast intra prediction mode selection scheme using temporal correlation in H.264”, *IEEE Conf. on Innovative Technologies for Societal Transformation*, November 2005, pp. 1 – 5.

[5] M. Jafari and S. Kasaei, ”Fast intra- and inter-prediction mode decision in H.264 Advanced Video Coding”, *Int. Journal of Computer Science and Network Security*, 8 (No. 5), May 2008, pp. 1 – 6.

[6] H. Ko, K. Yoo, J. Seo, and K. Sohn, ”Fast intra-mode decision using inter-frame correlation from H.264/AVC”, *IEEE Int. Symp. on Consumer Electronics*, April 2008, pp. 1 – 4.

[7] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, John Wiley & Sons Ltd, 2003.

[8] T. Tsukuba, I. Nagayoshi, T. Hanamura, and H. Tominaga, ”H.264 fast intra-prediction mode decision based on frequency characteristic”, *Proc. of EUSIPCO2005*, September 2005, pp. 61 – 66.

[9] J. Xin and A. Vetro, ”Fast mode decision for intra-only H.264/AVC coding”, *Proc. of the 25th Picture Coding Symposium*, April 2006, pp. 7.

Authorized licensed use limited to: Universitatsbibliothek der TU Wien. Downloaded on December 11, 2009 at 04:51 from IEEE Xplore. Restrictions apply.