Two-Dimensional Map Based Fast Mode Decision for H.264/AVC*

Tiesong Zhao, Hanli Wang, Xiangyang Ji, and Sam Kwong

Department of Computer Science, City University of Hong Kong, Hong Kong, China
ztiesong2@student.cityu.edu.hk, wanghl@cs.cityu.edu.hk,
{xiangyji,cssamk}@cityu.edu.hk

Abstract. The state-of-the-art video coding standard H.264/AVC achieves significant coding performance improvement by choosing the rate-distortion (RD) optimized encoding mode from a list of candidate modes at the cost of intensive computations, which limits the practical application of H.264/AVC. A number of fast mode decision algorithms have been recently presented in the literature in order to reduce H.264/AVC encoding computations. In this paper, a novel two-dimensional (2D) map based fast mode decision scheme is proposed, which produces a flexible mode checking order for fast mode decision. Experimental results demonstrate that with the proposed algorithm, the computational complexity of H.264/AVC coding is significantly reduced while the video quality and compression efficiency are preserved.

Keywords: H.264/AVC, mode decision, temporal-spatial correlation, 2D map.

1 Introduction

As more and more multimedia products have been invented and taken into application, video coding has played a more and more important role in storage and transmission of broadcast and entertainment media. As the state-of-the-art video coding standard, H.264/AVC [1] adopts several advanced coding techniques to achieve higher compression ratio, while the computational complexity is synchronously increased too. Therefore, it is important to decrease the encoding time of H.264/AVC for real-time transmission. In H.264/AVC, each macroblock (MB) has seven different block sizes for INTER mode coding and three block sizes for INTRA mode coding. Moreover, the SKIP and DIRECT modes are also used in H.264/AVC. All of these modes could be checked for every MB to search the best in terms of the least RD cost. This is one of the most important modules for H.264/AVC to achieve higher compression ratio, but brings intensive computations.

A number of fast mode decision algorithms have been proposed to reduce H.264/AVC computations [2]-[6]. Yin *et al.* [2] proposed a monotonic error surface based prediction (MESBP), which used the RD costs of INTER16x16, INTER8x8 and INTER4x4 modes to determine whether or not to check the other modes, but with a

* This work is supported by Hong Kong RGC Competitive Earmarked Research Grant Project 9041236 (CityU 114707).

Y.-M.R. Huang et al. (Eds.): PCM 2008, LNCS 5353, pp. 148–157, 2008.

fixed mode checking order. In [3] the characteristics of video objects including spatial homogeneity and temporal stationarity are employed for fast mode decision with about 30% of encoding time reduced. Wang *et al.* [4] [5] proposed a novel algorithm to detect all-zero blocks in H.264 and thus mode decision and ME could be early terminated with adaptive thresholds. Especially, in [5] it is proposed that the mode information of temporal-spatial neighboring MBs could be used for fast mode decision, resulting in the temporal-spatial checking (TSC) method. However, TSC is a conservative method that could not decrease the encoding time very much. In [6] the spatial neighboring MBs' optimal mode pattern is used to build a probable mode list from which the optimal mode for current MB could be predicted. It is reported that the encoding time is largely decreased for mobile applications, however at the noticeable cost of coding performance degradation with maximum 0.3 dB PSNR loss and 9% bitrate increment.

In this paper, a novel scheme for fast mode decision is proposed by virtue of the pre-coded mode information of temporal-spatial neighboring MBs. The modes of temporal-spatial neighboring MBs are firstly mapped into a 2D space with each mode represented as a 2D point. Then a distance based adaptive scheme is used to discard low-probability modes for fast mode decision. Experimental results show that the proposed algorithm decreases the encoding time by about 30% - 40% while keeping the coding performance almost intact. The rest of the paper is organized as follows. In Section 2, a 2D space with mode representation is defined. The details of the proposed 2D distance based fast mode decision algorithm are presented in Section 3. Experimental results are shown in Section 4. Section 5 concludes this paper and gives future research directions.

2 Mode Projection in 2D Space

As mentioned before, there are seven block sizes (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4), as shown in Fig. 1, for INTER mode decision, three block sizes (16x16, 8x8, 4x4) for INTRA mode decision, and SKIP mode and DERICT mode are also considered. For video objects with slow motion or many homogenous regions, the SKIP/DIRECT mode and larger block size modes such as INTER16x16 mode are probable to be chosen; whereas, for high motion or complex texture video objects, smaller block size modes such as INTER4x4 or INTRA mode are usually employed.

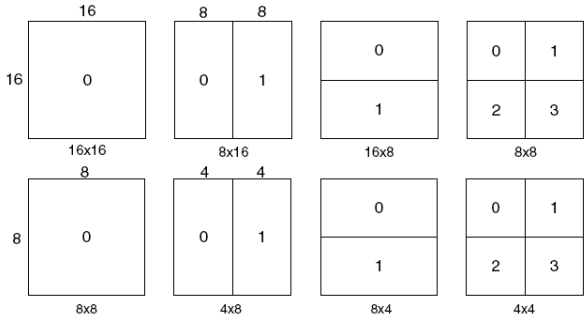

Fig. 1. MB and sub-MB partitions

In general, the mode decision aims to minimize the RD cost function for each mode defined as:

$$J(s,c,MODE \mid Q_p,\lambda_{MODE}) = SSD(s,c,MODE \mid Q_p) + \lambda_{MODE} * R(s,c,MODE \mid Q_p) \quad (1)$$

In the above equation, J denotes the cost function, which is dependent on the original signal s, the reconstructed signal c, the quantization parameter Q_p, the Lagrange multiplier λ_{MODE} and $MODE$ that indicates a candidate encoding mode. SSD denotes the sum of the square difference between s and c. R stands for the number of bits for coding the corresponding MB.

In order to model the temporal-spatial correlation of MBs, a 2D coherence value is defined for each encoding mode as given in Table 1, where the coherence value is calculated as the (horizontal block size, vertical block size) of the corresponding mode divided by 4. The larger the block size, the higher the coherence value. Therefore, the best mode information of temporal-spatial neighboring MBs can be clearly described in a 2D map, as shown in Fig. 2, in terms of the corresponding coherence values. In the next section, we will describe how to utilize the temporal-spatial correlated 2D map for the derivation of the proposed fast mode decision algorithm.

Table 1. Coherence values of different modes

MODE	Coherence value
SKIP/DIRECT*, INTER16x16, INTRA16x16	(4,4)
INTER16x8	(4,2)
INTER8x16	(2,4)
DIRECT*,INTER8x8, INTRA8x8	(2,2)
INTER8x4	(2,1)
INTER4x8	(1,2)
INTER4x4, INTRA4x4	(1,1)

* DIRECT mode can be applied to 16x16 and 8x8 blocks.

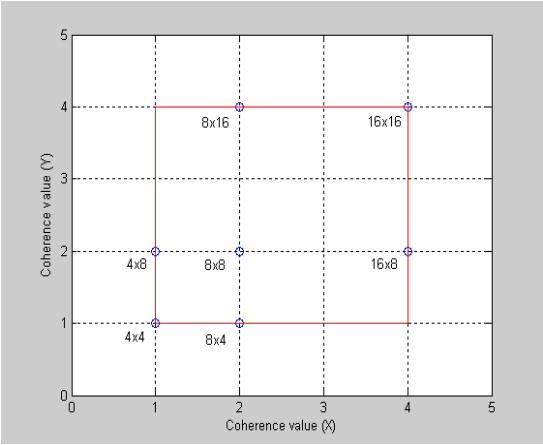

Fig. 2. Modes projected into 2-D space

3 Proposed 2D Map Based Mode Decision Algorithm

3.1 Distance-Based Adaptive Mode Checking

Natural video sequences usually exhibit strong temporal-spatial correlations. In [5] the temporal-spatial sets of a MB $M(n, x, y)$, located in the nth frame with the upper-left pixel (x, y) is defined as:

$$S_T = \{M(n-1, x+\Delta x, y+\Delta y \mid \Delta x, \Delta y = 0, \pm 16\} \tag{2}$$

$$\begin{aligned} S_S = \{&M(n, x-16, y-16), M(n, x, y-16), \\ &M(n, x+16, y), M(n, x-16, y)\} \end{aligned} \tag{3}$$

where S_T is the temporal set and S_S is the spatial set. If the MBs in both the temporal set and spatial set have been coded as SKIP/DIRECT or INTER16x16 modes, then $M(n, x, y)$ is determined to be homogeneous and only the INTER16x16 and SKIP/DIRECT modes are examined, which is the TSC method. However, TSC method is so conservative that it can not achieve significant computational reduction.

In this paper, we propose a novel 2D map based mode decision algorithm by fully making use of the temporal-spatial characteristics of MBs, which can significantly reduce the computations for H.264/AVC encoding. To achieve this, the coherence value based 2D map as described in the preceding section is utilized. Firstly, the correlation between the best encoding mode of a MB and those of its neighboring MBs in S_T and S_S are studied. Statistical analysis shows that the best encoding modes of S_T except $M(n\text{-}1, x, y)$ has a lower relativity than S_S. Hence, we only use the mode information of $M(n\text{-}1, x, y)$ and S_S with a new utility set S defined as:

$$\begin{aligned} S = \{&M(n, x-16, y-16), M(n, x, y-16), \\ &M(n, x+16, y), M(n, x-16, y), M(n-1, x, y)\} \end{aligned} \tag{4}$$

The best mode information of MBs in S is used to construct a coherence based 2D map. Secondly, a predicted point in the 2D map is generated to approximate the current MB's best encoding mode. Considering the correlation of mode information in the 2D map, the mean of the coherence values of set S is used for computing the predicted point. It is also possible to use other generation methods such as median and activity based methods to further improve the performance, which will be the future research direction for the current work. Thirdly, after getting the predicted point, a candidate mode list is constructed with an adaptive checking priority according to the distance between the 2D point of the corresponding mode and the predicted point, i.e., the nearest mode point will be checked first and then the second nearest mode point and so on, due to the strong temporal-spatial correlations of video objects in natural video sequences.

From exhaustive statistical analysis, we find that it is only necessary to check a subset of modes instead of the entire candidate mode list given a distance threshold. To verify this, Table 2 gives the percentage of the distance between the best mode and the predicted point for three benchmark video sequences. It is noted that the best mode has

Table 2. Distribution of the distance between the best mode and the predicted point

News, QCIF, IBBP				
distance	dist<=1.0	1.0<dist<=2.0	2.0<dist<=3.0	dist>3.0
Q_p=26	85.13%	14.65%	0.22%	0.00%
Q_p =30	86.42%	13.40%	0.18%	0.00%
Q_p =36	90.02%	9.94%	0.04%	0.00%
Forman, QCIF, IBBP				
distance	dist<=1.0	1.0<dist<=2.0	2.0<dist<=3.0	dist>3.0
Q_p =26	62.97%	36.14%	0.89%	0.00%
Q_p =30	70.12%	29.32%	0.56%	0.00%
Q_p =36	80.78%	18.85%	0.37%	0.00%
Mobile Calendar, QCIF, IBBP				
distance	dist<=1.0	1.0<dist<=2.0	2.0<dist<=3.0	dist>3.0
Q_p =26	60.68%	39.14%	0.18%	0.00%
Q_p =30	59.98%	39.58%	0.44%	0.00%
Q_p =36	68.95%	30.53%	0.52%	0.00%

a very high probability within a small area in the 2D map with the center at the predicted point. Therefore, we can discard the modes which are far away from the predicted point for fast mode decision.

3.2 Overall Fast Mode Decision Algorithm

Based on the above analysis, a novel fast mode decision algorithm is proposed with the step-by-step description as follows.

Step 1: Predict the best point for the current MB with temporal-spatial neighboring MBs in set *S*, go to Step 2.

Step 2: Compute all the distances between the predicted point and each INTER mode point, construct the candidate mode list by sorting all the modes according to their distances from the nearest to the furthest, go to Step 3.

Step 3: INTER mode decision: if dist $\geq T_a$, then go to Step 4, i.e., only check the modes with dist $< T_a$ in the candidate mode list, which is composed of the following sub-steps. T_a is a predetermined threshold for distance evaluation. In the proposed 2D map ($1 \leq x \leq 4$, $1 \leq y \leq 4$), a valid T_a falls in the range of $(0, 3\sqrt{2}]$. Based on our statistical analysis (see Table 2), we set $T_a = 3$ in this work.

Step 3.1: If the candidate mode is INTER16x16, and DIRECT/SKIP mode has not been checked before, check DIRECT/SKIP mode first, then check INTER16x16 mode, go to Step 3.4.

Step 3.2: If the mode is INTER16x8 or INTER8x16, check it and go to Step 3.4.

Step 3.3: If the mode is INTER8x8, INTER8x4, INTER4x8 or INTER4x4 and DIRECT mode (for B frame only) has not been checked before, check the DIRECT mode first and then the corresponding sub-MB modes, go to Step 3.4.

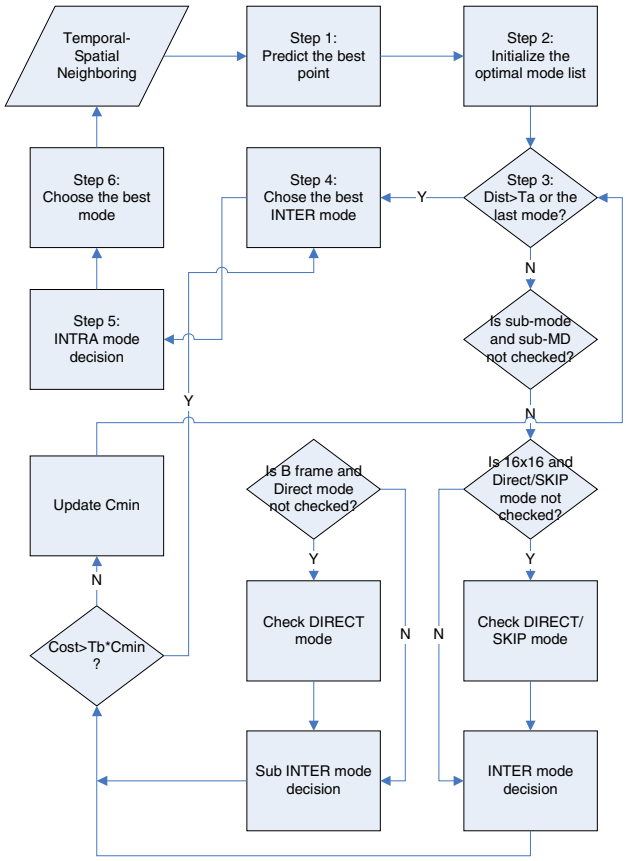

Fig. 3. Flow chart of the overall fast mode decision algorithm

Step 3.4: If the RD cost of the current mode $C < C_{min}$, set $C_{min} = C$, where C_{min} is the minimum RD cost found so far. Otherwise, if $C > T_b * C_{min}$, i.e., a larger RD cost is searched, then the INTER mode decision for the current MB is early stopped and go to Step 4. Otherwise, the INTER mode decision is continued by going back to Step 3.1 for the next mode.

The reason for proposing the early termination scheme triggered by $C > T_b * C_{min}$ lies in the fact that if a relatively larger RD cost is found, it is most likely that all the subsequent modes in the candidate mode list have larger RD costs than C_{min}, and thus are definitely not the best mode. This is because the modes in the candidate mode list are arranged in the ascending order based on their distances from the predicted point, and it is highly probable that the longer the distance, the larger the corresponding RD cost. T_b is a predefined threshold used for RD cost evaluation and $T_b \geq 1$. From exhaustive experimental results, it can achieve a good trade-off between encoding complexity reduction and video quality degradation by setting $T_b = 1$.

Step 4: Choose the best INTER mode, go to Step 5.

Step 5: INTRA mode decision: if the optimal mode point, which is the nearest mode point from the predicted point on 2D map, represents the INTER16x16 or SKIP/DIRECT mode, then skip checking the INTRA4x4 mode. Otherwise, if the optimal mode point represents a sub-MB mode and all of the four sub-MB modes are INTER4x4, then skip checking INTRA16x16. Go to Step 6.

Step 6: Choose the best coding mode for the current MB among all the tested modes.

The overall fast mode decision algorithm is also depicted in Fig. 3.

4 Experimental Results

To examine the performance of the proposed algorithm, the H.264/AVC reference software JM13.2 [7] is applied. The UMHexagonS motion estimation (ME) algorithm is used with quarter pixel resolution enabled. The motion vector search range is 16 and only 1 reference frame is used. The RD optimization and 8x8 transform are disabled, and CABAC is enabled.

Firstly, some statistical results are presented to verify the rationality of the proposed algorithm. The difference of RD cost of the best mode found by the proposed algorithm as compared to the original encoder is defined as:

$$E_c = (C - C_0) / C_0 \times 100\% \tag{5}$$

where C is the RD cost resulted from our algorithm and C_0 is the RD cost resulted from the original encoder. Due to the space limit, only the distribution of E_c results for *Foreman* (QCIF) sequence with an IntraPeriod of 10 and various Q_p values are shown in Table 3. For other video sequences, similar statistics can be obtained. The results demonstrate that the proposed algorithm can achieve a good result of RD costs and thus it could be verified in a sense.

Table 3. Distribution of RD cost difference

	IPPP				IBBP			
E_c(%)	Q_p=28	Q_p=32	Q_p=36	Q_p=40	Q_p=28	Q_p=32	Q_p=36	Q_p=40
<=0.0	60.47	63.60	59.97	60.37	59.85	61.56	59.63	59.72
0.0~1.0	7.24	6.46	6.84	5.89	6.58	5.68	5.54	5.62
1.0~2.0	6.16	4.81	4.28	4.41	5.68	4.83	4.22	4.10
2.0~3.0	5.93	4.31	4.91	4.31	4.61	4.50	3.88	3.52
3.0~4.0	4.41	3.77	3.40	3.23	4.08	3.70	3.47	3.27
4.0~5.0	3.03	2.82	3.47	2.90	3.24	3.06	2.94	3.02
5.0~6.0	2.42	2.39	2.49	1.95	3.02	2.98	2.89	2.15
6.0~7.0	1.95	2.19	2.32	2.12	2.44	2.12	2.43	2.45
7.0~8.0	1.78	2.15	1.99	1.99	2.13	2.23	2.08	2.15
8.0~9.0	1.31	1.41	1.58	1.55	1.73	1.54	1.52	1.88
>9.0	5.28	6.06	8.75	11.28	6.62	7.78	11.39	12.08
$E_{c(mean)}$	**-0.49**	**-0.92**	**-0.44**	**-0.90**	**-0.86**	**-1.23**	**-0.02**	**-0.75**

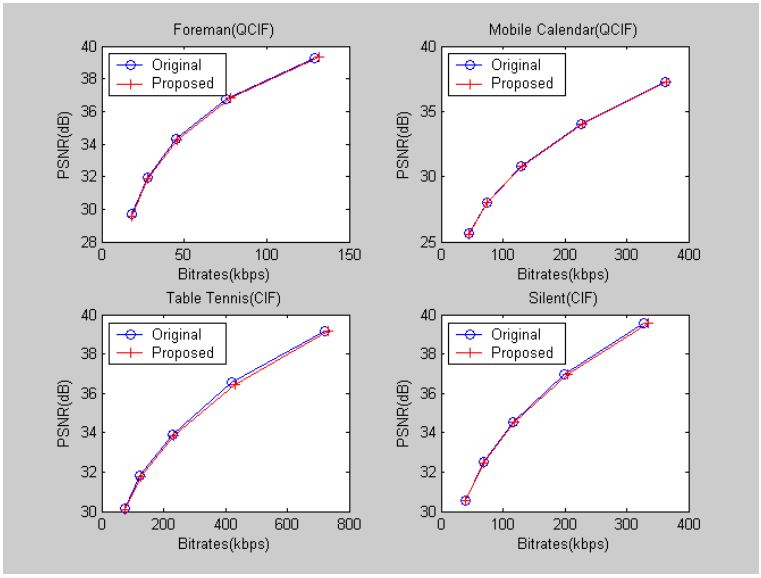

Fig. 4. RD curves of IPPP sequences

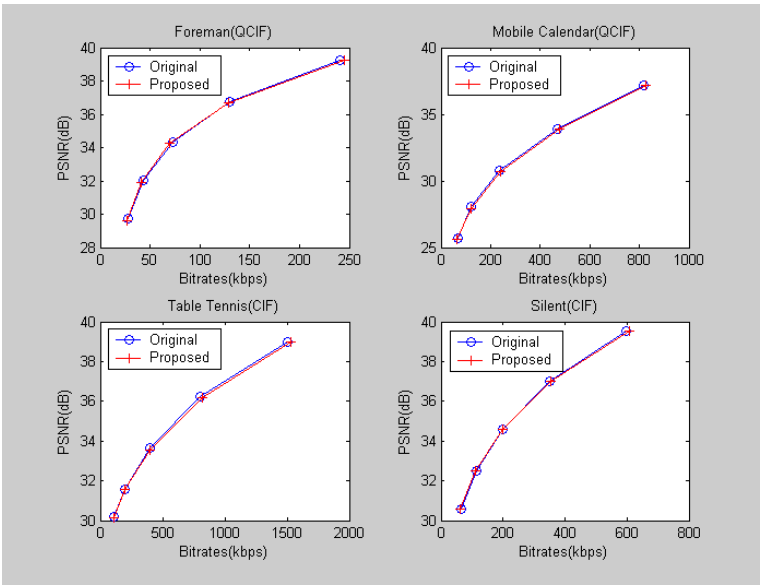

Fig. 5. RD curves of IBBP sequences

The RD performances of the proposed algorithm are shown in Figs. 4 and 5 for four benchmark video sequences with IPPP and IBBP coding structures, respectively, as compared to the original encoder. Two of the video sequences are in QCIF format including *Foreman* and *Mobile Calendar*; the other two are in CIF format including

Table Tennis and *Silent*. From Figs. 4 and 5, it can be seen that the proposed algorithm can achieve almost the same RD performance as the original encoder.

Finally, the encoding computational complexity is studied. Several video sequences, including *Foreman* (QCIF), *Mobile Calendar* (QCIF) and *News* (QCIF) with IPPP and IBBP structures are tested. In addition, different Q_p values, including 28, 32, 36 and 40, are used. The performance of the proposed algorithm as compared to the original encoder is evaluated in terms of the criteria as follows: the entire encoding time reduction (%) ΔT_E, ME time reduction (%) ΔT_M, PSNR degradation (dB) ΔP, and bitrates increment (%) ΔR. Due to the space limit, only the average results are given in Table 4, where the average results of TSC [5] is also presented for comparison.

Table 4. Average results

		ΔT_E (%)	ΔT_M (%)	ΔP(dB)	ΔR (%)
Proposed	IPPP	-31.81	-53.13	-0.054	1.10
	IBBP	-37.11	-46.75	-0.082	-1.10
TSC		-9.0	-10.0	-0.005	0.15

From Table 4, it is observed that the proposed algorithm can keep the video quality almost intact, while reducing the encoder complexity efficiently, about 30% for IPPP sequences and about 40% for IBBP sequences, as compared with the TSC method.

5 Conclusions and Future Works

In this paper, a novel 2D map based fast mode decision algorithm is proposed, which utilizes the temporal-spatial correlated characteristics of video objects. Experimental results demonstrate that the proposed algorithm can reduce the H.264/AVC coding complexity remarkably with the video quality and compression efficiency almost intact.

The proposed algorithm could be further improved by utilizing more coding information of temporal-spatial neighboring MBs, e.g., the activity information such as the number of non-zero quantized DCT coefficients and the RD cost information of temporal-spatial neighboring MBs could be used for generating the predicted point and designing more efficient scheme to skip redundant modes. In addition, the proposed algorithm can be further optimized by incorporating more efficient early termination conditions such as the early mode decision termination (EMDT) scheme in [4] [5].

References

1. Advanced Video Coding for Generic Audiovisual Services, ISO/IEC 14496-10:2005(E) ITU-T Rec. H.264 (E) (March 2005)
2. Yin, P., Tourapis, H.C., Tourapis, A.M., Boyce, J.: Fast mode decision and motion estimation for JVT/H.264. In: Proc. IEEE ICIP 2003, vol. 3, pp. 853–856 (September 2003)
3. Wu, D., Pan, F., Lim, K.P., Wu, S., et al.: Fast intermode decision in H.264/AVC video coding. IEEE Trans. Circuits Syst. Video Technol. 15(7), 953–958 (2005)

4. Wang, H., Kwong, S.: Hybrid Model to Detect Zero Quantized DCT Coefficients in H.264. IEEE Tran. Multimedia 9(4), 728–735 (2007)
5. Wang, H., Kwong, S., Kok, C.-W.: An efficient mode decision algorithm for H.264/AVC encoding optimization. IEEE Tran. Multimedia 9(4), 882–888 (2007)
6. Jo, Y., Kim, Y.-G., Choi, Y.: Fast mode decision algorithm using optimal mode predictions for H.264-based mobile devices. In: Proc. ICCE 2007, pp. 1–2 (January 2007)
7. H.264/AVC reference software JM13.2, `http://iphome.hhi.de/suehring/tml/`