PREDICTION TYPE DECISION FOR ADVANCED VIDEO CODING: AMACHINE LEARNINGAPPROACH

Muhammad Asif[1],Maaz Bin Ahmed[2], Rafi Ullah[2] and Syed Ali Hassan[1]
[1]Department of Computer Science and Information Technology, Lahore Leads University, Lahore, Pakistan
[2]COCIS Department, PAF, Karachi Institute of Economics and Technology, Karachi, Pakistan
Corresponding Author: astz786@yahoo.com

ABSTRACT: The modern video coding standards such as High Efficiency Video Coding (HEVC) and H.264/MPEG-4 Advanced Video Coding (AVC) used variable block-size intra and inter-prediction modes to improve the coding efficiency. These coding standards employ an exhaustive search algorithm rate-distortion optimization (RDO) to select the coding parameters for each macroblock (MB) such as prediction type, modes and block sizes. The use of the RDO process drastically increases the computational complexity of the encoder. This paper presents a computationally efficient methodology to decide the prediction type for an MB based on machine learning technique that exploit the spatial and temporal statistics of video sequence, modes ofpreviously encoded spatial and temporal neighboring MBs and motion-field statistics. The experimental results for H.264/AVC shows that the proposed technique is about 25.61% faster than the full search method RDO with negligible coding loss in terms of BDPSNR and BDBR i.e.by the amount of 0.006 dB and 0.169%, respectively.

KEYWORDS: H.264/AVC, HEVC, Prediction type, RDO, Inter Prediction, Intra Prediction

INTRODUCTION

Video coding standards mainly evolved through two well-known organizations Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG) [1]. These organizations jointly developed the latest video coding standards High Efficiency Video Coding (HEVC) [2]-[4] and H.264/AVC [5]. These coding standards outperform the previous coding standards in terms of better compression, visual quality and enhanced peak signal-to-noise ratio (PSNR) [6]. In order to obtain the better performance, these coding standards incorporated many new techniques that increase the coding efficiency at the cost of increase in computational complexity of the encoder [7]-[8]. In H.264/AVC coding standard, a macroblock (MB, i.e. 16×16 pixels) is basic processing unit in a video frame. It can be encoded as intra-Predicted (I-MB) or Inter-Predicted (P-MB). In case of Intra-Predicted (I-MB); an MB is predicted using the reconstructed pixels of the neighboring MBs in the current frame. On the other hand, in case of Inter-Predicted (P-MB); the prediction of an MB is performed using the reconstructed pixels of the MBs in the previous frame. For better representation of spatial and temporal details of an MB, H.264/AVC provides various coding modes with variable block sizes to perform intra and inter-prediction. For intra-prediction of luma component, two block sizes 4×4 and 16×16 are supported [9]. The nine prediction modes are offered for a luma 4×4 and four modes for a luma 16×16 and chroma 8×8 blocks [10]. For inter-prediction, seven different prediction block sizes are supported [11[-[12]. In H.264/AVC, RDO technique [13] is employed to select the coding parameters for an MB. The RDO calculates the rate distortion cost (RDcost) for all possible parameters and selects those that give minimum RDcost. Therefore, for each MB, the numbers of possible intra-prediction mode combinations are 592, and for inter-prediction there are 20 different possible block size combinations. Thus, to select the prediction type for one MB, the H.264/AVC encoder performs 592+20= 612 RDO calculations. As a result of this brute-force searching, the computational complexity of the encoder increases tremendously because it involves both encoding and decoding processes. To achieve real-time encoding, this computational complexity becomes a bottleneck. So, it is highly desirable to decrease the encoder complexity without any significant coding loss for the wide range of video encoding applications. In literature several efforts have been made in the area of prediction type decision (I-MB or P-MB). Chen et al. [14] proposed a fast prediction type selection technique based on simple features. They used variance and the sum of absolute differences (SAD) of an MB to model the costs of intra- and inter-coding methods, respectively. Some times, they also used motion vector and quantization parameters. Turaga and Chen [15] presented a classification-based prediction type decision algorithm that exploited the maximum likelihood (ML) criterion in order to facilitate the video transmission over networks. Kim [16] proposed a fast intra/inter mode decision scheme based on a risk-minimization criterion to reduce the encoder complexity of the H.264 encoder. It consists of three steps. At first step, 3D feature vector is formed by extracting the three features from the current MB. Secondly, the feature space is partitioned into three regions, i.e. risk-free, risk-tolerable, and risk-intolerable regions. Finally, mechanisms of different complexities are applied for the final mode decision depending on the location of the feature vector in the feature space. However, these schemes are not suitable for the prediction type decision for H.264/AVC because the selected features are too simple to provide the most suitable prediction type.

This paper presents an efficient technique to decide an MB prediction type (I-MB or P-MB) to reduce the computational complexity and overheads related RDO process in video encoding. The proposed technique exploit the spatial and temporal statistics of video sequence, modes of previously encoded spatial and temporal neighboring MBs and motion-field statistics to select an appropriate prediction type before starting the RDO process. The experimental results show that the presented methodology has resulted in significant decrease of computational complexity without much degradation in rate-distortion performance.

The rest of the paper is organized as follows. Section II presents the observation and analysis. The proposed technique is presented in Section III. Section IV presents the

experimental analysis. Finally, the conclusion is drawn in Section V.

I. OBSERVATIONS AND ANALYSIS

Extensive experiments are performed on variety of video sequences using exhaustive parameters selection technique RDO of the H.264/AVC reference software to acquire the data for statistical analysis of prediction type decision. The test conditions are set as follows: MV search range is 32 pels, entropy coding is set to CABAC, RDO and fast motion estimation are enabled in encoder main profile, MV resolution is 1/4-pel, number of reference frame is set to 1, and 300 frames are encoded. To compute the probability of coding parameters, encoding results at five different QPs including 24, 28, 32, 36 and 40 are used. Table 1 lists the averaged probability of selecting each prediction type when each test video sequence is encoded with IPPPPP structure.

Table 1: Probability (%) of Prediction types

Sequence	Format	I-MB	P-MB
Coastguard	QCIF (176x144)	99.75	0.25
Claire		99.92	0.08
Container		99.83	0.17
Foreman		99.47	0.53
Highway		99.87	0.13
Akiyo	CIF (352x288)	100	0
Mobile		99.9	0.1
MaD		99.77	0.23
Silent		98.79	1.21
Tempete		98.01	1.99
Flower	NTSC (720x480)	99.36	0.64
Football		90.07	9.93
Intros		96.96	3.04
Mobile		99.58	0.42
Vtc1nw		99.99	0.01
Parkrun	720p (1280x720)	99.75	0.25
Shield		99.01	0.99
Stockholm		99.27	0.73
Average		98.85	1.15

Table 1 illustrates that for 98.85% MBs inter-prediction type is selected and intra-prediction type is selected forthe remaining 1.15% of MBs. It indicates that the P-MB prediction type indeed dominates in inter frame encoding particularly for those sequences encompassing slow motion, homogeneous motion or motionless content. On the other hand, the probability of selecting I-MB prediction type is high incase of random motionand for MBs belonging to low motion regions with low texture.

Similarly, an extensive investigation is performed on several video sequences to observe the relationship between different statistics of video frames and their corresponding prediction type. It is observed that an MB prediction type is highly correlated with the prediction type of itsspatial and temporal neighboring MBs. This statistical analysis shows that the optimum selection of prediction type is also highly correlated with spatial and temporal statistics of the video content. Therefore, it can be concluded that spatial and temporal features of the video sequences are adequate to differentiate an MB thus to foretell a probable prediction type.

II. PROPOSED METHODOLOGY

The block diagram of the proposed methodology is shown in Fig. 1. It consists of two major steps. Initially, spatial and temporal features for current MB are extractedto form 8-D feature vector. Finally, based on feature vector, prediction type decision is made using machine learning algorithm.

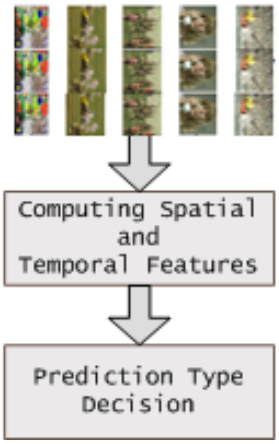


Fig. 1 Block Diagram of the Proposed Technique

A. Computing Spatial and Temporal Features

In this work, following spatial and temporal features are selected to predict the prediction type of an MB.

i. Average Brightness

It is the mean of luminance component values X (i,j) of an MB. The average brightness of each MB is calculated as

$$\mu_{m,n} = 1/MN \sum_{i=1}^{N} \sum_{j=1}^{M} X(i,j) \quad (1)$$

ii. Variance

Variance of an MB gives the information about its statistical dispersion and is used as measurement of texture. It can be roughly estimated as

$$\sigma_{m,n} = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(|X(i,j) - \mu_{m,n}| \right) \quad (2)$$

iii. Zero SAD (Z_SAD)

The sum of absolute difference between current MB and its collocated MB in the previous frame in display order is known as Zero SAD (Z_SAD). It gives the information about degree of variation between two MBs i.e. motion or stillness. The Zero SAD is calculated as

$$Zero\ SAD = \sum_{i=1}^{N} \sum_{j=1}^{M} |X(i,j) - Y(i,j)| \quad (3)$$

Where X indicates the current MB and Y is its collocated MB in previous frame.

iv. MB residual complexity (MB_RC)

In order to calculate the MB_RC, motion estimation on 8×8 block size is performed. For performing motion estimation, 3-D Recursive Search (3-D RS) [17] motion estimator is used because it is light weight and tends towards actual or true motion of objects. The MB_RC for each MB is the average SAD of its corresponding 8×8 blocks. If an MB is located at the m^{th} row and n^{th} column, it is denoted as MBm,n . The

SAD ofits corresponding 8×8 blocks are represented as SADi, j, i ∈ [8m, 8m + 1], j ∈ [8n, 8n + 1]. The residual complexity of an MB is calculated as

$$MB\_RC_{m,n} = 1/4 \sum SAD_{i,j} \quad (4)$$

v. Coding-Mode-Field Statistics

Coding-Mode-Field statistics are attained by the use of coding mode information of the temporal (in the previous frame Ft-1) and spatial (in the current frame Ft) neighboring MBs encoded as an intra MB i.e. I-MB.

$I\text{-}MB_{Spatial}=isI(MB_{TL}) + isI(MB_T) + isI(MB_{TR}) + isI(MB_L)$ (5)

$I\text{-}MB_{Temporal}= isI(MB_{TL})+ isI(MB_T)+ isI(MB_{TR})+ isI(MB_L)+ isI(MB_{Collocated})+ isI(MB_R)+ isI(MB_{DL})+ isI(MB_D)+ isI(MB_{DR})$ (6)

vi. Motion-Field Statistics

These statistics are acquired through motion characteristics (SAD) of the temporal and spatial neighboring MBs as follows:

$SAD_{Spatial}= (SAD_{TL}+ SAD_T + SAD_{TR} + SAD_L)/4$ (7)

$SAD_{Temporal}= (SAD_{TL}+ SAD_T + SAD_{TR} + SAD_L + SAD_{Collocated}+ SAD_R + SAD_{DL} + SAD_D + SAD_{DR})/9$ (8)

The temporal and spatial neighboring MBs of current MB are shown in Fig.2. In order to calculate the average brightness, variance and Zero SAD of an MB each frame is down sized by factor four. In down sized frame, 4×4 block size of an actual frame is represented by one pixel and 16×16 block size is mapped to 4×4 block size. This down sizing helps to reduce the computational complexity related to feature extraction.

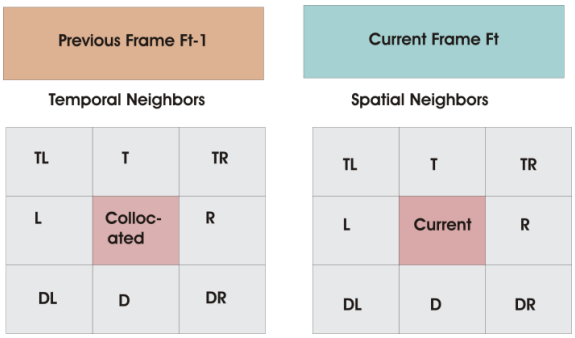


Fig.2 Spatial and Temporal Neighboring Macroblocks of current MB

B. Prediction Type Decision

In this work, macroblock prediction type decision is taken as classification problem. To solve this problem a machine learning based solution is presented in which each MB is classified into one of the three following classes:

• Class 1: MB is predicted as intra (I-MB)
• Class 2: MB is predicted as inter P-MB
• Class 3: MB can be predicted as intra I-MB or inter P-MB.

The decision of MB prediction type class 3 is made through RDO. The reason behind the consideration of Class 3 is to minimize the coding performance degradation. To select an appropriate prediction type, Adaboost classifier is trained using spatial and temporal features mentioned in section III.A. The training data is obtained from variety of video sequences that contains 10000 samples. Then, the aforementioned features and the appropriate prediction type determined by the RDO are taken as the training set for Adaboost classifier. The numbers of trained week classifier are 75.

The class decision of each MBm,n with 8-D feature vector F Vm,n = {μ, σ, Zero SAD, MB_RC, $SAD_{Spatial}$, $SAD_{Temporal}$, $I\text{-}MB_{Spatial}$, $I\text{-}MB_{Temporal}$} is made based on class conditional probabilities as follows

• If P1 <P2 and P1 >τ, then MBm,n belongs to Class 1
• If P2 >P1 and P2 >τ, then MBm,n belongs to Class 2
• Otherwise, MBm,n belongs to Class 3

Where P1 and P2 are the probabilities of each macroblock MBm,n which belongs to Class 1 and Class 2, respectively. The value of τ is set to 0.6 after performing large number of experiments on variety of test sequences. The outcome of this classification is probable prediction type for an MB this helps to reduce the computational complexity of the RDO process.

III. EXPERIMENTAL ANALYSIS

The proposed methodology is integrated into the H.264/AVC JVT Reference Software (Version JM 12.2) [18].To evaluate the performance of the proposed technique, experiments are conducted on a PC with Intel core i3-2100 CPU @ 3.1 GHz x and 2 GB RAM by using the wide range of test video sequences. The test conditions are set as follows: encoder main profile is used, fast motion estimation and RDO is enabled, MV search range is -32 to +32 pels, MV resolution is ¼ pel, and number of reference frames is set to 5. Four different quantization parameters including 28, 32, 36 and 40 are used. Test sequences with three different frame formats, QCIF (144×176), CIF (352×288) and NTSC(720×480) are used. For each test sequence, 100 frames are encoded in IPPPP structure. All the frames are encoded as P-frames except the first one which is encoded as I. Each test sequence is encoded three times independently for each quantizer to compute the mean results. In order to evaluate the performance of the proposed schemes three metrics are used including Bjontegaard delta bit-rate (BDBR) [19], Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and time saving (TS). TS can be defined as follow

$$Ts = \frac{Tp-Tr}{Tr} \times 100 \quad (9)$$

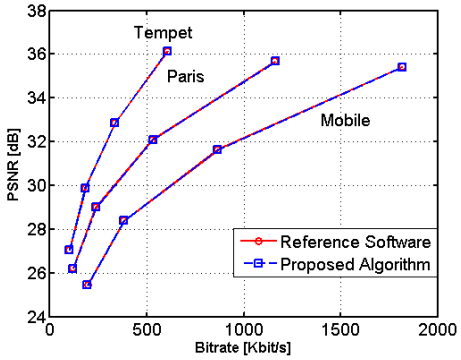
Where Tp and Tr are the encoding time taken by the proposed methodology and reference software, respectively. The negative values of the performance measuring metrics BDPSNR, BDBR and TS indicate decrease whereas positive values represent an increase. All training processes of classifier are accomplished offline. During the encoding, the trained models are loaded at the beginning. The proposed prediction type decision algorithms used these models to decide the prediction type for an MB based on the run time features. Table 2 shows the experimental results for wide range of test sequences. It demonstrate that the presented scheme is about 25.61% faster than the exhaustive full search technique RDO with negligible coding loss in terms of BDBR and BDPSNR i.e. by the amount of 0.169% and 0.006 dB, respectively. The proposed scheme shows a consistent gain in encoding time savings for all sequences ranging from 19.77% in Akiyo to 32.51% in Washdc. This encoding gain is

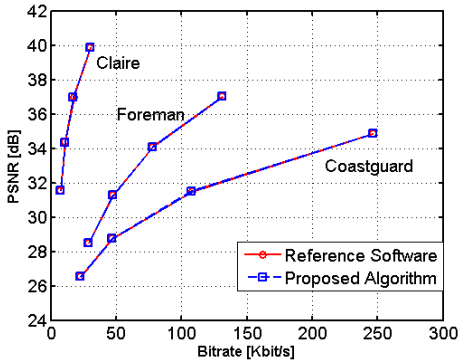
Table 2: Experimental Results

		Performance Analysis			Class Frequency (%)		
Sequences	Format	Ts	BDBR	BDPSNR	Class 1	Class 2	Class 3
Foreman	QCIF (176x144)	-31.53	0.167	-0.009	0.02	92.59	7.39
Clarie		-20.56	-0.576	0.037	0.01	66.83	33.16
Coastguard		-30.9	-0.249	0.008	0.03	90.56	9.41
Container		-22.36	-0.021	0	0	72.21	27.79
Hall		-23.73	-0.011	0	0	79.89	20.11
Highway		-28.34	-0.508	0.019	0.03	81.04	18.93
Akaiyo	CIF (352x288)	-19.77	0.001	0	0.06	68	31.94
Mobile		-29.02	0.069	-0.003	0.2	89.7	10.1
MaD		-19.93	0.548	-0.025	0.18	66.04	33.78
Paris		-30.78	0.072	-0.004	0.01	92.21	7.78
Silent		-27.11	0.638	-0.027	0.46	87.09	12.45
Tempet		-30.15	0.429	-0.018	0.27	89.57	10.16
Flower	NTSC (720x480)	-22.19	0.023	-0.002	0.05	65.58	34.37
Football		-25.76	1.072	-0.045	1.35	78.71	19.94
Mobile		-21.87	-0.062	0.003	0.04	69.28	30.68
vtc1nw		-21.48	0.34	-0.011	0	67.24	32.76
washdc		-32.51	0.419	-0.018	0.01	93.93	6.06
Galleon		-22.93	0.699	-0.022	0.64	71.44	27.92
Average		-25.61	0.169	-0.006	0.19	78.99	20.82

achieved with a maximum increase in BDBR of 1.072% or a maximum BDPSNR loss of 0.045 dB, and is thus negligible. Table 2 also shows the frequency of each class that is averaged using the results under four different Qps including 28, 32, 36 and 40. It can be inferred from the percentage of MBs belonging to particular class that the reduction in encoding time is maximum for the sequences for which most of the MBs are classified into Class 1 and Class 2. For instance, in case of Washdc sequence, only 6.06% of MBs are classified to Class 3 and therefore, for the remaining 93.94% MBs, RDcost calculation is performed for either intra or inters prediction type only. This results in reduction of around 32.51% in encoding time. On the other hand, in case of Akiyo sequence, 31.94% of MBs are classified to Class 3 and for these MBs time consuming RDcost calculation is performed for both intra and inter-prediction types resulting in comparatively lower time saving i.e. 19.77%.

The Rate Distortion (RD) curves of the JM reference exhaustive prediction type decision algorithm and suggested prediction type decision scheme are demonstrated in Fig.3. It demonstrates that the presented prediction type decision technique attains asimilar RD performance as that of the JM reference software full search algorithm.





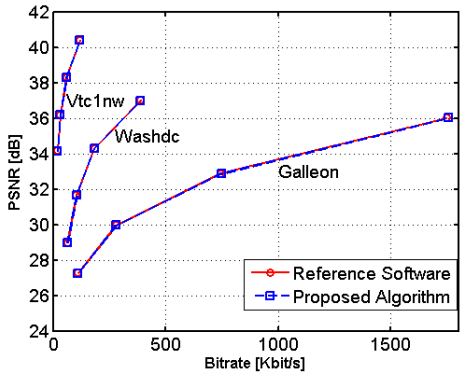


Fig.3 RD curves for Prediction type Decision Scheme

IV. CONCLUSION

This paper presented a machine learning-based algorithm for fast prediction type decision in H.264 encoding. The proposed technique exploited the spatial and temporal statistics of video sequence, modes of previously encoded spatial and temporal neighboring MBs and motion-field statistics to decide an appropriate prediction type before starting the RDO process. The experimental results showed that the presented

methodology 25.61% speedup the encoding process without significant degradation in rate-distortion performance.

REFERENCES

[1] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transaction on circuits and systems for video technology*, **22**(12), 1649-1668 (2012).

[2] P. K. Podder, M. Paul and M. Murshed, "Fast Mode Decision in the HEVC Video Coding Standard by Exploiting Region with Dominated Motion and Saliency Features,"*PLoS ONE*, **11**(3), 1-22 (2016).

[3] S. Radicke, J. Hahn, Q. Wang and C. Grecos, "A Parallel HEVC Intra Prediction Algorithm for Heterogeneous CPU+GPU Platforms,"*IEEE Transactions on Broadcasting*, **62**(1), 103-119 (2016).

[4] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm and T. Wiegand,"High Efficiency Video Coding (HEVC) Text Specification Draft 9, document JCTVCK1003,"*ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCTVC)* (2012).

[5] T. Wiegand, GJ. Sullivan, G. Bjontegard and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, **13**(7), 560-576 (2003)

[6] G. J. Sullivan J.-R. Ohm B. Bross, W.-J. Han and T. Wiegand,"High efficiency video coding (hevc) text specification draft 9, document jctvc- k1003,"*ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT- VC)* (2012)

[7] M. Asif, S.Majeed, Imtiaz A. Taj, S. M. Ziauddin, and Maaz Bin Ahmad,"Exploiting MB level parallelism in H.264/AVC encoder for multi-core platform," *The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)* ,125–130 (2014)

[8] S. Momcilovic, N. Roma, and L Sousa,"Exploiting task and data parallelism for advanced video coding on hybrid cpu + gpu platforms"*J, Real Time Image Proc.*, (2013)

[9] M. Asif, Imtiaz A. Taj, S. M. Ziauddin, Maaz Bin Ahmad, and M. Tahir,"An efficient scheme for intra prediction block size and mode selection in advanced video coding" *Frontier of Information Technology (FIT)*(2015)

[10] W.-J. Han J. Min J. Lainema, F. Bossen and K. Ugur, "Intra coding of the hevc standard" *IEEE Transactions on Circuits Syst. Video Technol.*, **22**(12):1792–1801 (2012)

[11] J. Lee, S. Kim, K. Lim, H. J. Kim, and S. Lee, "Fast intermode decision algorithm based ongeneral and local residual complexity in h.264/avc,"*EURASIP J. Image and Video Process.*,(2013)

[12] M. Asif, Imtiaz A. Taj, S. M. Ziauddin, Maaz Bin Ahmad, and Atif Raza. An efficient inter prediction mode selection scheme for advanced video coding based on motion homogeneity and residual complexity.*"IEEJ Transactions on Electrical and Electronic Engineering,"***11**(6), 760-767, (2016)

[13] G. Sullivan and Wiegand T., "Rate-distortion optimization for video compression". *In IEEE Signal Processing Magazine*, (1998)

[14] Y.-K. Chen, A. Vetro, H. Sun, and S. Y. Kung, "Optimizing intra/inter coding mode decisions," *in the Proceedings of International Symposium on Multimedia Information Processing*, Taipei, (1997)

[15] D. S. Turaga and T. Chen, "Classification based mode decisions for video over networks," *IEEE Trans. Multimedia, Special issue on Multimedia over IP*, **3**(1), 41-52 (2001)

[16] C. Kim and C. –C. Jay Kuo, "Fast intra/inter mode decision for H.264 encoding using a risk-minimization criterion", *Proc. SPIE*, 55-58 (2004)

[17] G. de Haan, P.W.A.C. Biezen, H. Huijgen and O. A. Ojo, "True motion estimation with 3D recursive search block matching", *IEEE Trans. Circuits and Systems for Video Technology*, 3, 368-379 (1993)

[18] Joint Video Term (JVT), "H.264/AVC reference software",

[19] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", *In VCEG-M33, 13th Meeting*, (2001)