EUROGRAPHICS Workshop on Multimedia (2004)
N. Correia, J. Jorge, T. Chambel, Z. Pan (Editors)

Analysis of inter-frame coding without intra modes in H.264/AVC*

Yun Cheng[1,2] , Zhiying Wang[1] , Kui Dai[1] and Jianjun Guo[1]

1 College of Computer, National University of Defense Technology, Changsha, China
2 Department of Computer, College of Hunan Humanities, Science and Technology, Loudi, China

Abstract

The ABSTRACT is to be in fully-justified italicized text, between two horizontal lines, in one-column format, below the author and affiliation information. Use the word "Abstract" as the title, in 9-point Times, boldface type, left aligned to the text, initially capitalized. The abstract is to be in 9-point, single-spaced type. The abstract may be up to 3 inches (7.62 cm) long.
Leave one blank line after the abstract, then add the subject categories according to the ACM Classification Index (see http://www.acm.org/class/1998/) H.264/AVC is a new international standard for video coding which has great advantage of coding efficiency com-pared with other standards. It can save about 50% bit-rate compared with that of the successful prior coding stan-dards under the same reconstructed picture quality. But the high coding efficiency is acquired by heavily computa-tion. In this paper, the coding mode and algorithm for mode decision are introduced firstly, then transform and quantization are analyzed and experiments on inter-frame coding with or without intra modes are performed. The experiment results illustrate that the encoding method without intra modes in inter-frame coding will decrease the encoding time from 76.03% to 50.09% compared with that of the standard encoding method, while the PSNR-Y will change from -0.45dB to +0.20dB (most cases are ±0.10dB) at the same bit-rates..

Categories and Subject Descriptors (according to ACM CCS): H.4.3 [Information Systems Applications]: Computer conferencing, teleconferencing and videoconferencing.

1. Introduction

In 1998, the ITU-T VCEG (Video Coding Experts Group) started a project called H.26L with the target to double the coding efficiency when compared with any other existing video coding standard. In December 2001, the ITU-T VCEG and the ISO/IEC MPEG (Moving Pictures Expert Group) formed the Joint Video Team (JVT) with the charter to finalize the new video coding standard [WS03] which is known as Recommendation H.264 and also MPEG-4 Part 10: AVC(Advanced Video Coding).

H.264/AVC has great advantage of coding efficiency compared with the successful prior coding standards. It can save 64.46%, 48.80%, and 38.62% bit-rate compared with that of MPEG-2, H.263++ (HLP: High Latency Profile), and MPEG-4 (ASP: Advanced Simple Profile) respectively under the same reconstructed picture quality [SWS03]. But the high coding efficiency is acquired by heavily computation. It is estimated that the complexity of H.264/AVC encoder is about 5~10 times as that of MPEG-4, and the complexity of H.264/AVC decoder is about 2~4 times as that of MPEG-4. Chinese researchers' test results are as followed [ZYJ*03]. The actual time needed for

* This research was supported by the National Science Foundation of China (NSFC) 60173040

H.264/AVC encoder is about 14 times as that of H.263, and the actual time needed for H.264/AVC decoder is about 11 times as that of H.263. The speed of coding limits the application of H.264/AVC technologies in the domain of real time communication.

This paper is organized as follows. Section 2 introduces coding mode and algorithm for mode decision in H.264/AVC. Transform and quantization for mode decision in H.264/AVC are analyzed in section 3. In section 4, experiment results and discussions are presented. Finally, a conclusion is given in the last section.

2. Coding mode and algorithm for mode decision in H.264/AVC

2.1 Coding mode in H.264/AVC

Generally speaking there are two coding modes in H.264/AVC: one is intra-frame coding, and the other is inter-frame coding. In all slice-coding types, two classes of intra coding types are supported, which are denoted as Intra_4×4 and Intra_16×16 in the following. In contrast to previous video coding standards where prediction is conducted in the transform domain, prediction in H.264/AVC is always conducted in the spatial domain by referring to neighbouring samples of already coded blocks.

When using the Intra_4×4 mode, each 4×4 block of the luma component utilizes one of the nine prediction modes. When utilizing the Intra_16×16 mode, which is well suited for smooth image area, a uniform prediction is performed for the whole luma component of a macroblock. Four prediction modes are supported. The chroma samples of a macroblock are always predicted using a similar prediction technique as for the luma component in Intra_16×16 macroblock. Intra prediction across slice boundaries in not allowed in order to keep all slices independent of each other.

For the inter-frame coding mode, the block size can be 16×16, 16×8, 8×16, 8×4, 4×8, or 4×4, which are called corresponding to mode 1~mode 7. Because all of the seven partitions pay critical roles for improving the H.264/AVC's coding efficiency [JK01], we use all of the block-sizes mentioned above in the following analysis.

2.2 Algorithm for mode decision in H.264/AVC [SW03]

2.2.1. Algorithm for mode decision in inter-frame coding

In order to adapt the prediction more precisely than previous standards to movement within one video picture, macroblocks can be partitioned as Fig.1 [WS03].

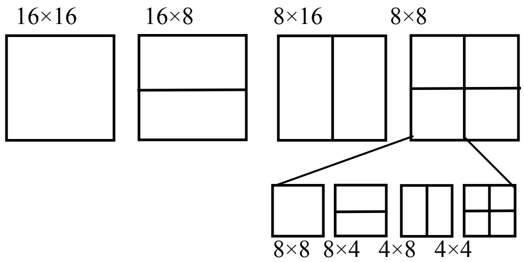


Figure 1:*Macroblock and sub-macroblock partitions*

The macroblock mode decision is done by minimizing the Lagrangian functional

$$J(s,c,MODE\,|\,QP,\lambda_{MODE}) = SSD(s,c,MODE\,|\,QP) + \lambda_{MODE} R(s,c,MODE\,|\,QP) \quad (1)$$

Where QP (Quantization Parameter) is the macroblock quantizer, λ_{MODE} is the Lagrange multiplier for mode decision, and $MODE$ indicates a mode chosen from the set of potential prediction modes:

$$MODE_I \in \{Intra\_4\times4, Intra\_16\times16\} \quad (2)$$

$$MODE_P \in \left\{\begin{matrix} Intra\_4\times4, Intra\_16\times16, SKIP, \\ 16\times16, 16\times8,\ 8\times16,\ 8\times8 \end{matrix}\right\} \quad (3)$$

$$MODE_B \in \left\{\begin{matrix} Intra\_4\times4, Intra\_16\times16, DIRECT, \\ 16\times16, 16\times8,\ 8\times16,\ 8\times8 \end{matrix}\right\} \quad (4)$$

Note that SKIP mode refers to the 16×16 mode where no motion and residual information is encoded. SSD is the sum of the squared differences between the original block s and its reconstruction c given as

$$\begin{aligned} SSD(s,c,MODE|QP) = &\sum_{x=1,y=1}^{16,16}(s_Y[x,y]-c_Y[x,y,MODE|QP])^2 \\ &+\sum_{x=1,y=1}^{8,8}(s_U[x,y]-c_U[x,y,MODE|QP])^2 \\ &+\sum_{x=1,y=1}^{8,8}(s_V[x,y]-c_V[x,y.MODE|QP])^2 \end{aligned} \quad (5)$$

and $R(s,c,MODE \mid QP)$ is the number of bits associated with choosing MODE and QP, including the bits for the macroblock header, the motion, and all DCT blocks. $c_Y[x,y,MODE \mid QP]$ and $s_Y[x,y]$ represent the reconstructed and original luminance values; c_U, c_V, and s_U, s_V the corresponding chrominance values. The Lagrangian multiplier λ_{MODE} is given by

$$\lambda_{MODE,P} = 0.85 \times 2^{QP/3} \quad (6)$$

for I and P frames and

$$\lambda_{MODE,B} = \max\left(2, \min(4, QP/6)\right) \times \lambda_{MODE,P} \quad (7)$$

The mode decision for the 8×8 sub-partitions is done similar to the macroblock mode decision by minimizing Eq.(1), and the $MODE$ indicates a mode chosen from the set of potential prediction modes:

$$MODE_P \in \{Intra\_4\times4, 8\times8, 8\times4, 4\times8, 4\times4\} \quad (8)$$

$$MODE_B \in \begin{Bmatrix} Intra\_4\times4, & DIRECT, & & \\ 8\times8, & 8\times4, & 4\times8, & 4\times4 \end{Bmatrix} \quad (9)$$

2.2.2. Algorithm for mode decision in intra-frame coding

The Intra_16×16 mode decision is performed by choosing the Intra_16×16 mode which results in the minimum *SATD* (Sum of Absolute Transform Differences) value.

For the Intra_4×4 prediction, the mode decision for each 4×4 block is performed similar to the macroblock mode decision by minimizing Eq.(1), and the MODE indicates an intra prediction mode:

$$MODE_I \in \begin{Bmatrix} DC, HOR, VERT, DIAG\_DL, DIAG\_DR, \\ VERT\_R, VERT\_L, HOR\_D, HOR\_U \end{Bmatrix} \quad (10)$$

SSD is the sum of the squared differences between the original 4×4 block luminance signal s and its reconstruction c, and $R(s,c,MODE \mid QP)$ represents the number of bits associated with choosing intra prediction mode. It includes the bits for the intra prediction mode and the DCT-coefficients for the 4×4 luminance block.

3. Analysis of transform and quantization for mode decision in H.264/AVC

In the algorithm for mode decision in H.264/AVC, the transform and quantization are basic operations. By Eq.(5), we can see that the encoding of images are necessary if the distortions need to be computed, and the transform and quantization are necessary if the transform and quantization are necessary if the images need to be encoded. The transform and quantization in H.264/AVC are analyzed under the following condition. Picture format: QCIF (Quarter Common Intermediate Format); types of frame: I-, P-, and B-frame (where I, P, and B stands for Intra, Predictive, and Bi-predictive respectively); reference frames: 1.

3.1 Analysis of transform and quantization for I-frame

In I-frame coding mode, the Intra_4×4 luma blocks have nine intra prediction modes, the 8×8 chroma blocks have four prediction modes, and the Intra_16×16 luma blocks also have four intra prediction modes. The estimated count of 4×4 luma blocks' transform and quantization for the 99 macroblocks in an I-frame is 57024□and the actual count is 50567□the estimated count of 16×16 luma blocks' transform and quantization in an I-frame is 396, and the actual count is 357; the estimated count of 8×8 chroma blocks' transform and quantization is 792(one transform and quantization will process the u and v chroma signal), and the actual count is 714. The only causation for the actual counts are all less than the corresponding estimated values is that some marginal blocks can't use the whole prediction modes.

3.2 Analysis of transform and quantization for P-frame

The transform and quantization for P-frame are quite different from that of I-frame. When the 8×8 sub-macroblock is being searched for the best motion vector, the 4×4 luma blocks' transform and quantization are performed after the motion search for one of the 8×8 sub-macroblock partitions having been finished. One 8×8 sub-macroblock consists of four 4×4 blocks, so the count of 4×4 luma blocks' transform and quantization in one 8×8 sub-macroblock is four. From Figure 1 we can see that one 8×8 sub-macroblock has four sub-block partitions, and one 8×8 sub-macroblock needs sixteen 4×4 luma blocks' transform and quantization after all of the nine best motion vectors having been founded. One 16×16 macroblock consists of four 8×8 sub-macroblock, we can draw the conclusion that the count of 4×4 luma blocks' transform and quantization after all of the 8×8 sub-macroblock's motion vectors having been founded is sixty-four; moreover, one macroblock can be partitioned as 16×16, 16×8, or 8×16 sub-block, after all of the three sub-block's best motion vectors having been founded, the count of 4×4 luma blocks' transform and quantization is forty-eight. For the P-frame in H.264/AVC, after the block-sizes are decided, the Intra_4×4 and Intra_16×16 modes are also checked, the count of 4×4 luma blocks' transform and quantization for the intra modes in a macroblock is at least 103 (the estimated count is 144, but some marginal 4×4 blocks can't use the full prediction modes), and the count of 16×16 luma blocks' transform and quantization is at least 99. From above analysis the count of 4×4 luma blocks' transform and quantization for one macroblock in P-frame is at least 215, and for a P-frame is at least 21285. The count of 8×8 chroma blocks' transform and quantization for a P-frame is at least 594.

3.3 Analysis of transform and quantization for B-frame

For B-frame, the count of motion search is double of P-frame's. The count of 4×4 luma blocks' transform and quantization for a B-frame is at least 27621, the count of 16×16 luma blocks' transform and quantization is similar to that of P-frame, 99, and the count of 8×8 chroma blocks' transform and quantization for a B-frame is at least 891.

3.4 Comparison of transform and quantization in inter-frame coding with or without intra modes in H.264/AVC

Table 1 is the comparison of transform and quantization for inter-frame coding with or without intra modes in H.264/AVC.

From Table 1 we can see that for the encoding of P-frame, if the intra-macroblock coding modes is omitted, the time for 4×4 luma blocks' transform and quantization will decrease about 50%, the time for 16×16 luma blocks' transform and quantization will derease 100□□and the time for 8×8 chroma blocks' transform and quantization will derease 1/3, so the encoding speed will improve greatly if the inter-frame coding without intra-modes is adopted in H.264/AVC.

4. Experiment results and discussions

On the base of H.264/AVC reference software JM8.1 [JVT04], we compare the performance in two aspects which are operational complexity and coding efficiency for the inter-frame coding without intra modes. Full search(the searching range is 16×16 rectangular window and the accuracy is 1/4 pixel), all block-size(mode 1~mode 7), RDO , and CABAC coding are adopted in the H.264/AVC encoder with 2 reference frames. The simulation environment of test is: P4 1.8GHz, 256MB SDRAM, VC++ 6.0 compiler.

Many images sequences are tested in the experiments, and two typical image sequences with QCIF format are selected out to be compared here, which are Foreman and Silent. Silent is a simple sequence of moving pictures whose background is stable, while Foreman is a complicated sequence of moving pictures whose background and objects are all moving. Each of them has 100 frames, with the first frame being coded as I-frame and the residual 99 frames being coded as P-frame. Some more experiment results are included in the Appendix.

By the analysis above, we know that it will decrease greatly the time for the transform and quantization if the intra modes are not used in the encoding of P-frames. Encoding time represents the operational complexity. Tab.2 shows the comparison of encoding time under the typical QPs.

From Table 2 we can find that the encoding time without intra modes in P-frame is distinctly less than that of with intra modes under the same test conditions, and the cut down percentage of encoding time is increasing with the QP decreasing. For example, the encoding time for the sequences of Foreman and Silent decreased 70.53% and 76.03% respectively when QP is 0, while the percentage of decreased time for Foreman and Silent are 52.64% and 52.67% respectively when QP is 51.

The coding efficiency can be shown as figure of PSNR-Y and bit-rate. Fig.2 shows the comparison of coding efficiency for the sequences of Foreman and Silent.

From Figure 2 we can find that the encoding efficiency of inter-frame coding without intra modes is almost the same as that of with intra modes when the image sequence is complex (foreman.qcif); for the simple image sequence (silent.qcif), the coding efficiency of inter-frame coding without intra modes is nearly the same as that of with intra modes when the bit-rate is high, while the PSNR will decrease about 0.15dB compared with that of with intra modes when the bit-rate is low.

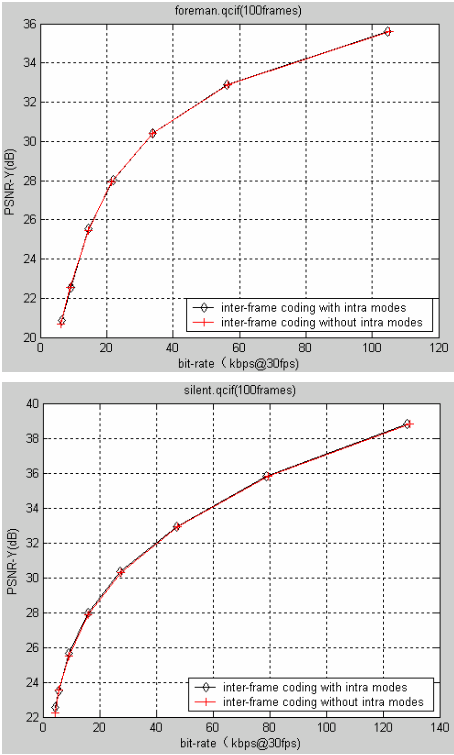


Figure 2:*Comparison of PSNR-Y and bit-rate*

5. Conclusions

In this paper, theoretical analysis and experiment research are performed for the inter-frame coding with or without intra modes in H.264/AVC. The experiment results illustrate that the encoding method without intra modes in inter-frame coding will decrease the encoding time from 76.03% to 50.09% compared with that of the standard encoding method, while the PSNR-Y will change from -0.45dB to +0.20dB (most cases are ±0.10dB) at the same bit-rates.

References

[JK01] A. Joch, F. Kossentini: Performance Analysis of H.26L Coding Features, *ITU-T/SG 16/VCEG, 15th Meeting, Pattaya, Thailand* ,Doc. VCEGO42, Dec.2001.

[JVT04] Joint Video Team (JVT) Test Model JM8.1, April 2004 at *http://bs.hhi.de/~suehring/tml/download/*

[SW03] G. Sullivan, T. Wiegand, K.P. Lim: Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Joint Model Reference Encoding Methods and Decoding Concealment Methods, *document JVT-I049d0.doc, San Diego, USA*, September, 2003

[SWS03] R. Schäfer, T. Wiegand, H. Schwarz: The emerging H.264/AVC standard EBU Technical Review, January 2003 at *http://www.ebu.ch/trev_293-contents.html*

[WS03] T. Wiegand, G. Sullivan: Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification(ITU-T Rec. H.264|ISO/IEC 14496-10 AVC),*document JVT-G050d35.doc,7th Meeting: Pattaya, Thailand*, March,2003

[ZYJ*03] J.L. Zhou, Y. Jin, S.S. Yu, J.H. Zheng: Research on video coding technologies based on H.264.*Journal of Huazhong University of Sci. and Tech.(Nature Science Edition)*, Vol.31(8) Aug. 2003□32□34

Abbreviations

ASP (MPEG-4) Advanced Simple Profile

AVC Advanced Video Coding

CABAC Context-Adaptive Binary Arithmetic Coding

CAVLC Context-Adaptive Variable Length Coding

CIF Common Intermediate Format

DCT Discrete Cosine Transform

HLP (H.263++) High Latency Profile

IEC International Electrotechnical Commission

ISO International Organization for Standardization

ITU International Telecommunication Union

ITU-T ITU - Telecommunication Standardization Sector

JVT (MPEG/VCEG) Joint Video Team

MPEG (ISO/IEC) Moving Picture Experts Group

PSNR Peak Signal-to-Noise Ratio

QCIF Quarter Common Intermediate Format

QP Quantization Parameter

RDO Rate Distortion Optimization

SAD Sum of Absolute Differences

SATD Sum of Absolute Transform Differences

SDRAM Synchronous Dynamic Random Access Memory

SSD Sum of Squared Differences

VCEG (ITU-T) Video Coding Experts Group

Table 1: Comparison of transform and quantization for inter-frame coding with or without intra modes in H.264/AVC

Types of transform	P-frame With Intra modes	P-frame Without - Intra modes	P-frame Decreasing Percentage	B-frame With Intra-modes	B-frame Without-Intra modes	B-frame Decreasing Percentage
4×4 luminance	21285	11088	47.91%	27621	17424	36.92%
16×16 luminance	99	0	100.00%	99	0	100.00%
8×8 chrominance	594	396	33.33%	891	693	22.22%

Table 2 Comparison of Encoding Time

QP	Foreman With Intra-modes	Foreman Without Intra modes	Foreman Decreasing Percentage	Silent With Intra-modes	Silent Without Intra modes	Silent Decreasing Percentage
0	349.855	103.107	70.53%	332.274	79.644	76.03%
4	318.083	97.427	69.37%	299.405	72.573	75.76%
8	286.640	90.298	68.50%	265.950	66.237	75.09%
12	255.979	84.079	67.15%	237.089	63.088	73.39%
16	229.770	79.485	65.41%	215.157	62.813	70.81%
20	205.443	75.111	63.44%	195.087	63.148	67.63%
24	186.075	72.842	60.85%	180.145	64.956	63.94%
28	173.162	72.720	58.01%	168.630	66.335	60.66%
32	163.377	71.941	55.97%	159.485	68.048	57.33%
36	157.105	72.846	53.63%	152.608	68.229	55.29%
40	152.136	72.656	52.24%	146.809	68.024	53.67%
44	147.870	71.122	51.90%	142.152	67.713	52.37%
48	143.256	69.380	51.57%	140.531	65.509	53.39%
51	139.420	66.037	52.64%	135.105	63.946	52.67%

Appendix

Some more results of the experiments

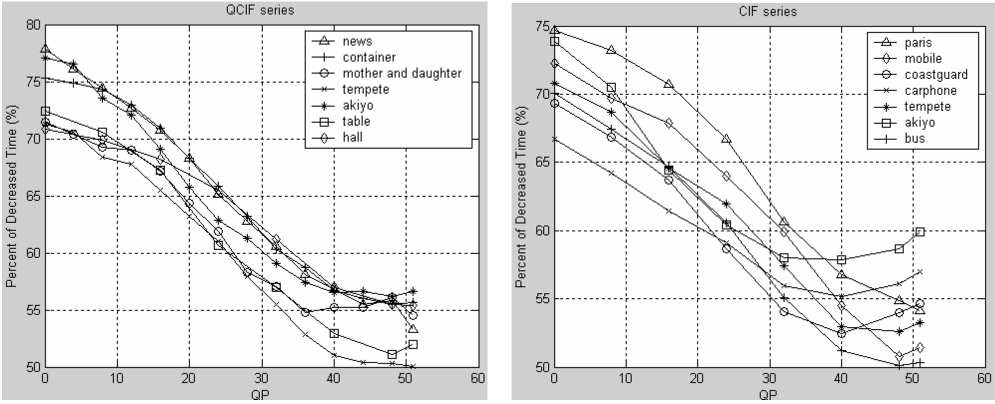
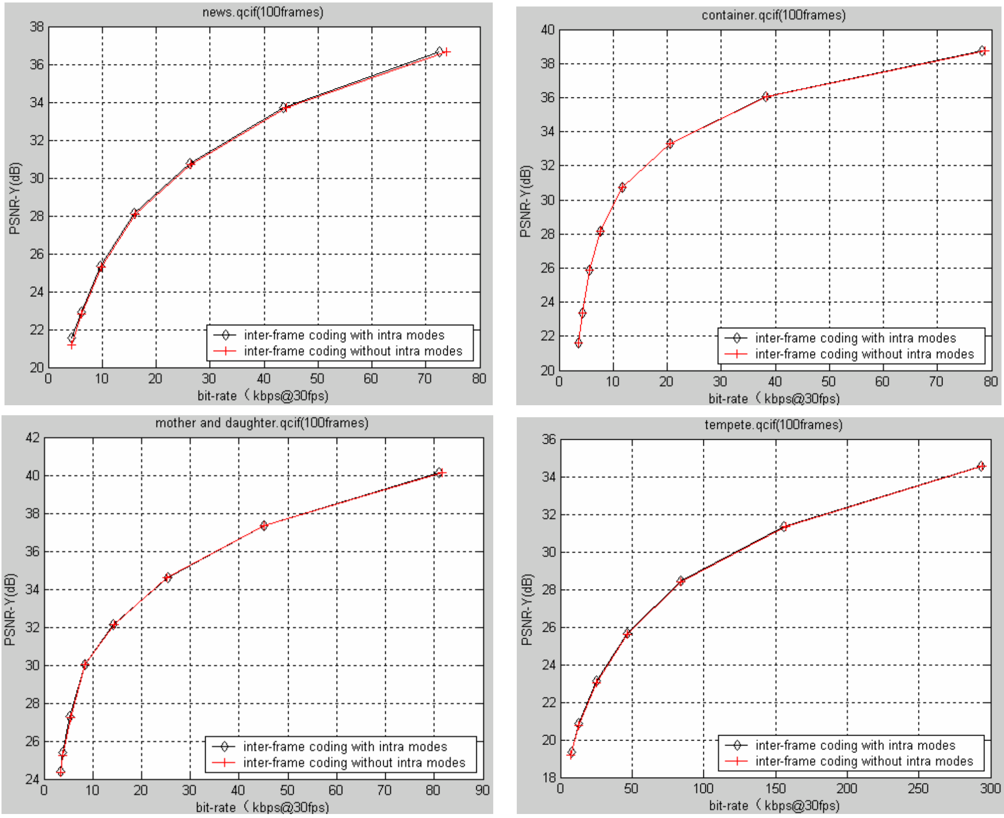
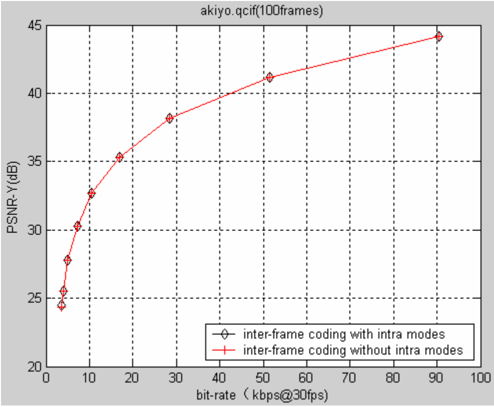


Figure 3:*Percentage of decreased time in different QPs for some of the QCIF and CIF series*




akiyo.qcif(100frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes

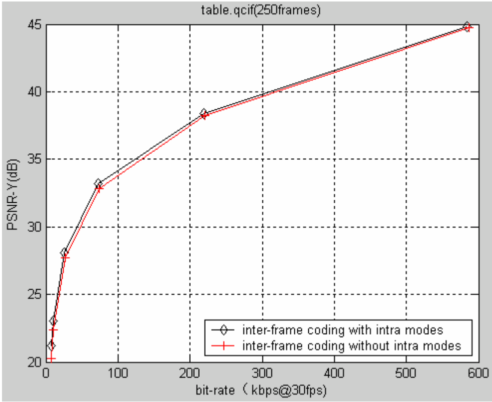
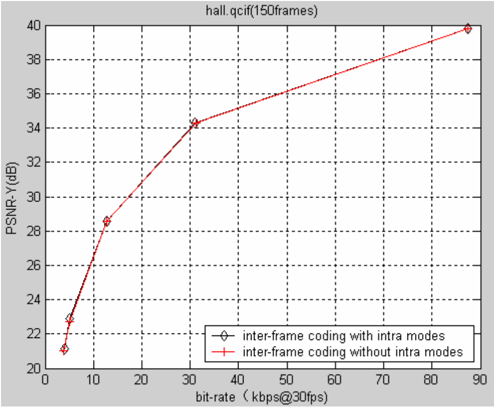
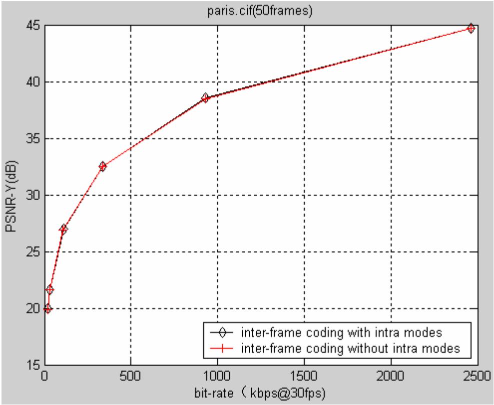
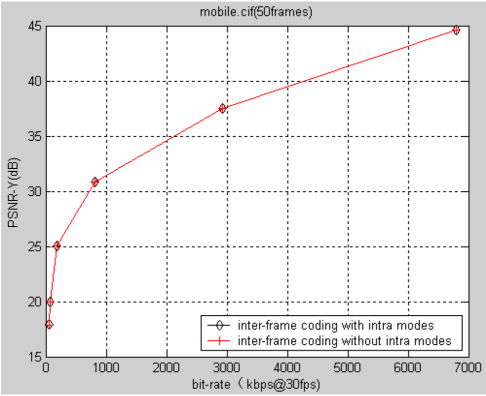
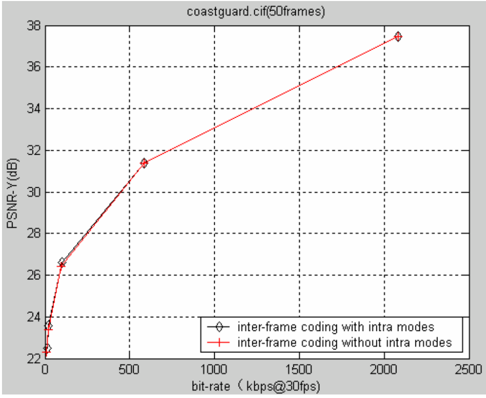


table.qcif(250frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes


hall.qcif(150frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes


paris.cif(50frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes


mobile.cif(50frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes


coastguard.cif(50frames)
PSNR-Y(dB)
bit-rate (kbps@30fps)
inter-frame coding with intra modes
inter-frame coding without intra modes

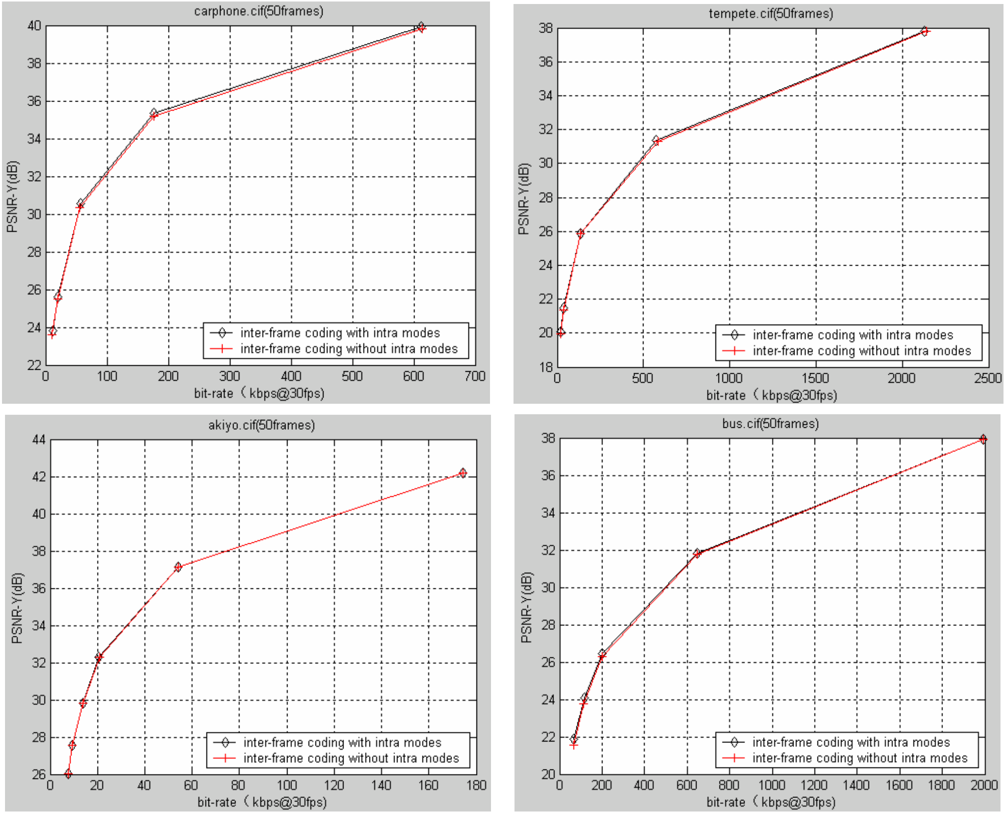


Figure 4: *Comparison of PSNR-Y and bit-rate for some of the QCIF and CIF series*

.