


การควบคุมอัตราสำหรับการส่งวิทยุทัศน์ H.264 โดยใช้แบบจำลองอัตราบิต-ความเพี้ยนแบบโคซี่



นางสาวนงลักษณ์ เขียมจำรัส

ศูนย์วิทยุทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A RATE CONTROL FOR H.264 VIDEO TRANSMISSION USING CAUCHY RATE DISTORTION MODEL



Miss Nongluk Eiamjumrus

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University

นางลักษณ์ เขียมจำรัส : การควบคุมอัตราสำหรับการส่งวิดีโอ H.264 โดยใช้แบบจำลองอัตราบิต-ความเพี้ยนแบบโคซี. (RATE CONTROL FOR H.264 VIDEO TRANSMISSION USING CAUCHY RATE DISTORTION MODEL) อ.ที่ปรึกษา : ผศ. ดร. สุภาวดี อร่ามวิทย์, 182 หน้า.

ในการประยุกต์ใช้งานเทคโนโลยีการสื่อสารมัลติมีเดีย ต้องอาศัยเทคโนโลยีการเข้ารหัสสัญญาณวิดีโอที่มีประสิทธิภาพเพื่อให้ผู้ใช้บริการได้รับบริการที่มีคุณภาพยอมรับได้ และเนื่องจากแบนด์วิดท์ของช่องสัญญาณมีปริมาณจำกัด ดังนั้นสัญญาณวิดีโอจะถูกบีบอัดลงไปที่อัตราบิตต่ำกว่าก่อนส่งออกไปยังช่องสัญญาณ ซึ่งอัตราบิตเฉลี่ยจะเป็นตัวบ่งบอกถึงคุณภาพของสัญญาณวิดีโอที่ถูกถอดรหัส กรรมวิธีที่ใช้ในการจัดสรรจำนวนบิตที่เหมาะสมให้แก่ละภาพวิดีโอ เรียกว่า การควบคุมอัตรา เพื่อให้ได้ภาพวิดีโอที่มีคุณภาพดีที่สุดในภายใต้เงื่อนไขอัตราบิตที่จำกัด กรรมวิธีการควบคุมอัตราจึงต้องถูกออกแบบอย่างมีประสิทธิภาพโดยใช้แบบจำลองอัตรา-ความเพี้ยน เพื่อรองรับหลากหลายสถานการณ์โดยเฉพาะสถานการณ์ที่มีเงื่อนไขอัตราบิตและเวลาประวิงต่ำ ซึ่งเป็นปัจจัยที่สำคัญมากในงานประยุกต์การสื่อสารวิดีโอที่จริง

ในวิทยานิพนธ์นี้ได้ยืนยันจากการวิเคราะห์และการทดลองว่าลักษณะการแจกแจงความน่าจะเป็นแบบโคซีสามารถนำมาประมาณลักษณะของอัตราบิตและความเพี้ยนของสัญญาณวิดีโอได้อย่างแม่นยำกว่าการแจกแจงอื่นที่มีการใช้ เช่น การแจกแจงแบบลาปลาเซียน ดังนั้นจึงได้นำเสนอแบบจำลองที่หาค่าเหมาะที่สุดของอัตราบิตและความเพี้ยนโดยใช้ฟังก์ชันความหนาแน่นความน่าจะเป็นแบบโคซี เพื่อใช้ในการหาค่าระดับชั้นของการควอนไทซ์ที่เหมาะสมที่สุดโดยเทคนิคตัวคูณแบบ Lagrange แบบจำลองที่นำเสนอจะถูกประยุกต์ใช้ในการควบคุมอัตรา โดยวิทยานิพนธ์นี้ได้นำเสนอการควบคุมอัตราแบบใหม่สำหรับการเข้ารหัสอัตราบิตต่ำของสัญญาณวิดีโอ H.264 นอกเหนือจากนี้ ได้พิจารณาสถานการณ์เวลาประวิงต่ำในการส่งสัญญาณวิดีโอที่อาจมีเหตุการณ์การกระโดดของเฟรมวิดีโอเพื่อลดเวลาประวิงของบัฟเฟอร์ เพื่อลดปัญหาดังกล่าว วิทยานิพนธ์ได้นำเสนอการควบคุมอัตราอีกวิธีหนึ่งสำหรับการเข้ารหัสอัตราบิตต่ำของสัญญาณวิดีโอ H.264 ในสถานการณ์เวลาประวิงต่ำ ผลการทดสอบเปรียบเทียบวิธีการควบคุมอัตราทั้ง 2 วิธีที่นำเสนอพบว่าสัญญาณวิดีโอที่มีคุณภาพที่ดีขึ้นรวมถึงมีจำนวนเฟรมวิดีโอที่กระโดดลดลง เมื่อเปรียบเทียบกับมาตรฐาน H.264 JM8.6

ภาควิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่อ.....นางลักษณ์ เขียมจำรัส.....
 สาขาวิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
 ปีการศึกษา.....2550.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4671848221 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: H.264 / RATE CONTROL / RATE DISTORTION MODEL / LAGRANGE MULTIPLIER TECHNIQUE / CAUCHY DISTRIBUTION

NONGLUK EIAMJUMRUS : A RATE CONTROL FOR H.264 VIDEO TRANSMISSION USING CAUCHY RATE DISTORTION MODEL THESIS ADVISOR : ASST.PROF.SUPAVADEE ARAMVITH,Ph.D., 182 pp.

Many multimedia applications require video coding schemes that can provide an acceptable quality of service to the end users. Due to the bandwidth of channel are limited, video source is usually encoded with a low-bit-rate video compression. In the coding process, the mean bit rate of the coded data is proportional to the decoded video quality. The process to allocate suitable number of bits to each video frame is called rate control. To achieve best video quality subject to the bit rate constraints, rate control algorithm has to be efficiently designed with the underlying rate-distortion model to cope with several scenarios especially low bit-rate and low delay scenario which are very important factors for real time video communication.

In this dissertation, we verified through analysis and simulations that Cauchy distribution provides more accurate estimates of rate and distortion characteristics of video sequences than the previously distribution used such as Laplacian distribution. A new rate-distortion optimization model based on Cauchy probability density function is then proposed and optimized to find the optimum choice of quantization step size in video coding by using Lagrange multiplier technique. The proposed rate-distortion optimization model is then applied in an application of rate control. Thus, we propose a new rate control for low bit-rate H.264 video coding. Considering the scenario of low delay in real time video transmission that frames may be skipped to reduce the buffer delay. This results in the motion discontinuity and low visual quality in video communication system. To alleviate the problem, we propose another rate control scheme for low bit-rate H.264 video coding under low delay constraint. Simulation results for both proposed rate control schemes indicate video quality improvement with less number of frame skipped compared to H.264 JM 8.6.

Department.....Electrical Engineering.....Student's signature.....*Nongluk Eiamjums*
 Field of study...Electrical Engineering.....Advisor's signature.....*Supavadee Aramvith*
 Academic year.....2007.....Co-advisor's signature.....

Acknowledgements

The author wishes to express her sincere appreciation to Assist.Prof.Dr.Supavadee Aramvith, the advisor for her criticisms, guidance and suggestions. The author also wishes to acknowledge to all friends in video research group, such as Nat, Wisut, Nattachai, Pichai, Navin, Nontharat, Juntana and Rhands.

Deep appreciation is also expressed to my father, Mr. Paijit Eiamjumrus, who was dead on 8 April 2550. Thank for my mother, both of my brothers and Mr. Chakrit Suvanjumrat.

This dissertation is in part supported by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University and Thailand Research Fund.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

List of contents

	Page
Abstract (Thai).....	IV
Abstract (English).....	V
Acknowledgements.....	VI
Table of contents.....	VII
List of tables.....	X
List of figures.....	XII
List of Abbreviations.....	XXI
Chapter	
I Introduction.....	1
1.1 Background and significance of the research problem.....	1
1.2 Literature review.....	5
1.3 Objective.....	10
1.4 Scope.....	10
1.5 Research procedure.....	10
1.6 Expected prospects.....	11
II. Background.....	12
2.1 Background on video coding standard.....	12
2.1.1 Transform domain coding.....	15
2.1.2 Quantization.....	16
2.1.3 Variable length coding.....	17
2.1.4 Motion estimation and Motion compensated prediction.....	18
2.1.5 Rate control.....	19
2.1.6 Test model for video standards.....	20
2.2 H.264/MPEG-4 Part 10(Advance video coding) standard.....	21
2.2.1 Intra Prediction.....	25
2.2.2 Motion Compensation Prediction.....	28
2.2.3 Transform and quantization.....	31
2.2.4 Entropy coding.....	33
2.2.5 Deblocking filter.....	34
2.2.6 Flexible Macroblock ordering (FMO).....	36
2.2.7 Profiles and Levels of H.264.....	37

Chapter	Page
2.3 Rate - distortion theory.....	39
2.3.1 Background of rate-distortion theory.....	39
2.3.2 Rate-distortion functions.....	41
2.4 H.264 Rate control.....	43
2.5 Objective quality measurement.....	46
III An improved rate-distortion optimization model based on Cauchy density	
function for H.264 video coding.....	48
3.1 Analysis of DCT-coefficients by curve fitting.....	49
3.2 Background of rate and distortion model based on	
Cauchy density function.....	54
3.2.1 Cauchy based rate model	55
3.2.2 Cauchy based distortion model	59
3.3 Proposed generalized Cauchy R-D optimization model	
using Lagrange multiplier technique.....	60
3.4 Proposed a linear prediction rate and distortion model parameter	
by linear regression analysis.....	65
3.4.1 The linear regression model.....	65
3.4.2 Cauchy rate model parameter.....	69
3.4.3 Cauchy distortion model parameter.....	70
IV An improved rate control based on Cauchy rate-distortion	
optimization model for H.264 video coding.....	72
4.1 GOP layer rate control.....	72
4.2 Frame layer rate control.....	73
4.3 Basic unit layer rate control.....	74
4.4 Simulation results.....	76
4.5 Summary.....	100
V An improved rate control based on Cauchy rate-distortion optimization model	
for H.264 video coding under low delay constraint.....	101
5.1 Impact of low delay constraint of H.264 JM 8.6 rate control.....	101
5.2 Proposed rate control scheme using Cauchy R-D optimization model	
under low delay constraint.....	106

List of tables

x

Table	Page
Table 3.1 Comparison of the error curve fitting between Cauchy and Laplacian curve fitting with the histogram of DCT coefficients.....	54
Table 4.1 Parameters of simulation H.264 software version JM 8.6 with main profile.....	78
Table 4.2 Performance of proposed scheme compared with H.264 JM8.6 rate control at 16, 32 and 64 kbps as target bit rate	98
Table 4.3 Performance of proposed scheme compared with H.264 JM8.6 rate control at 128 and 256 kbps as target bit rate	99
Table 5.1 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 16 kbps for delay time 100 ms.....	139
Table 5.2 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 32 kbps for delay time 100 ms.....	139
Table 5.3 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 64 kbps for delay time 100 ms.....	140
Table 5.4 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 128 kbps for delay time 100 ms.....	140
Table 5.5 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 256 kbps for delay time 100 ms.....	141
Table 5.6 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 16 kbps for delay time 150 ms.....	141
Table 5.7 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 16 kbps for delay time 200 ms.....	142
Table 5.8 Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [15] at 16 kbps for delay	

Chapter	Page
5.2.1 GOP layer rate control for low delay constraint.....	106
5.2.2 Frame layer rate control for low delay constraint.....	107
5.2.3 Basic unit layer rate control for low delay constraint.....	108
5.2.4 Simulation results	111
5.2.5 Summary.....	143
VI Conclusion.....	144
References.....	146
Appendix.....	151
Publications.....	152
Curriculum vitae.....	182



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

List of tables

XI

Table

Page

time 400 ms..... 142



ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

List of figures

XII

Figure	Page
figure 1.1 Block diagram of a video encoder	2
figure 2.1 Progression of the ITU-T Recommendations and ISO/IEC standards	12
figure 2.2 A typical quantizer.....	16
figure 2.3 Block matching approach for motion compensation.....	18
figure 2.4 Variable bit rate versus constant bit rate channels.....	20
figure 2.5 Generalized block diagram of a hybrid video encoder with motion compensation: The adaptive deblocking filter and intra-frame prediction are two new tools of H.264.....	22
figure 2.6 Partitioning of an image into several slices.....	23
figure 2.7 4:2:0 sampling patterns (progressive).....	24
figure 2.8 4:2:0 samples to top and bottom fields.....	24
figure 2.9 Labeling of prediction samples 4x4.....	25
figure 2.10 Nine possible intra prediction modes for the intra prediction type INTRA_4x4.....	26
figure 2.11 Four possible intra prediction modes for the intra prediction type INTRA_16x16.....	27
figure 2.12 Intra prediction for chroma macroblock.....	28
figure 2.13 Partitioning of a macroblock and a sub-macroblock for motion compensated prediction.....	29
figure 2.14 Motion-compensated prediction with multiple reference frames. In addition to the motion vector, also an image reference parameter Δ_i is transmitted.....	30
figure 2.15 Matrices H_1 , H_2 and H_3 of three different transforms applied in H.264/AVC.....	31
figure 2.16 Transmission orders of all coefficients of a macroblock.....	32
figure 2.17 Quantization step sizes in H.264/AVC codec.....	33
figure 2.18 Performance of deblocking filter for highly compressed pictures. (a) Without the deblocking filter (b) with deblocking filter.....	34
figure 2.19 Edge filtering order in a macroblock.....	35

List of figures

XIII

Figure	Page
figure 2.20 Samples adjacent to vertical and horizontal boundaries.....	36
figure 2.21 Difference type of FMO.....	37
figure 2.22 H.264/AVC profiles and corresponding tools.....	38
figure 2.23 Transmission system.....	40
figure 2.24 Rate - Distortion curve.....	40
figure 2.25 Block diagram of rate control process for H.264.....	47
figure 3.1 Probability density functions of each 8x8 DCT-coefficients.....	50
figure 3.2 Laplacian probability Density Function at different x and λ	51
figure 3.3 Cauchy probability Density Function at different x_0 and μ	52
figure 3.4 Histogram of DCT coefficients of Akiyo sequence.....	52
figure.3.5 Entropy of Akiyo sequence versus approximated entropy functions at different quantization step size (Q).....	57
figure.3.6 Entropy of Foreman sequence versus approximated entropy functions at different quantization step size (Q).....	57
figure.3.7 Entropy of Carphone sequence versus approximated entropy functions at different quantization step size (Q).....	58
figure.3.8 Entropy of Tempete sequence versus approximated entropy functions at different quantization step size (Q).....	58
figure.3.9 Distortion of Akiyo sequence versus approximated distortion functions at different Quantization step size (Q).....	60
figure 3.10 Plot of regression line.....	66
figure 3.11 Plot of data and regression line	69
figure 4.1 Test video sequence uses in simulation : (a) Carphone sequence, (b) Foreman sequence, (c) Silent sequence, (d) News sequence, (e) Akiyo sequence and (f) Claire sequence (g) Missam_Suzie.....	77
figure 4.2 PSNR versus frame for our proposed and JM8.6 rate control of Akiyo sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps	

List of figures

XIV

Figure	Page
(e) Simulation results at 256 kbps	81
figure 4.3 PSNR versus frame for our proposed and JM8.6 rate control of Carphone sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps	83
figure 4.4 PSNR versus frame for our proposed and JM8.6 rate control of Claire sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps	85
figure 4.5 PSNR versus frame for our proposed and JM8.6 rate control of Foreman sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps.....	87
figure 4.6 PSNR versus frame for our proposed and JM8.6 rate control of News sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps	89
figure 4.7 PSNR versus frame for our proposed and JM8.6 rate control of Silent sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps	91
figure 4.8 PSNR versus frame for our proposed and JM8.6 rate control of Missam_Suzie sequence (a) Simulation results at 16 kbps (b) Simulation results at 32 kbps (c) Simulation results at 64 kbps (d) Simulation results at 128 kbps (e) Simulation results at 256 kbps	93

List of figures

Figure	Page
figure.4.9 PSNR versus bit rate in each frame for our proposed and JM8.6 rate control of seven video sequence.	
(a) Average PSNR versus bit rate in each frame for Akiyo sequence	
(b) Average PSNR versus bit rate in each frame for Carphone sequence	
(c) Average PSNR versus bit rate in each frame for Claire sequence	
(d) Average PSNR versus bit rate in each frame for Foreman sequence	
(e) Average PSNR versus bit rate in each frame for News sequence	
(f) Average PSNR versus bit rate in each frame for Silent sequence	
(g) Average PSNR versus bit rate in each frame for Missam_Suzie sequence.....	95
figure 4.10 Decoded frame of rate control algorithm for Missam_Suzie sequence of our proposed compared with the H.264 JM 8.6 rate control.	
(a1) 62 th frame of the proposed (PSNR= 34.56 dB) of 16 kbps	
(a2) 62 th frame of the JM 8.6 (PSNR= 30.55 dB) of 16 kbps	
(b1) 71 th frame of the proposed (PSNR= 32.90 dB) of 32 kbps	
(b2) 71 th frame of the JM 8.6 (PSNR= 28.86 dB) of 32 kbps.....	96
figure 4.11 Decoded frame of rate control algorithm for News sequence of our proposed compared with the H.264 JM 8.6 rate control at 16 kbps.	
(a1) 50 th frame of the proposed (PSNR= 38.67 dB)	
(a2) 50 th frame of the JM 8.6 (PSNR= 37.67 dB)	
(b1) 63 th frame of the proposed (PSNR= 39.91 dB)	
(b2) 63 th frame of the JM 8.6 (PSNR= 38.92 dB).....	97
figure 5.1 Simulation using Foreman sequence encoded by H.264 JM8.6 rate control at 16 kbps with delay constraint of 100 ms	
(a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR per frame.....	103
figure 5.2 Simulation using Carphone sequence encoded by H.264 JM8.6 rate control at 16 kbps with delay constraint of 100 ms	
(a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR per frame.....	104
figure 5.3 Simulation using Coastguard sequence encoded by H.264 JM8.6	

List of figures

XVI

Figure	Page
rate control at 16 kbps with delay constraint of 100 ms (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR per frame.....	105
figure 5.4 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR of each frame with delay constraint.....	113
figure 5.5 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR of each frame with delay constraint.....	114
figure 5.6 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR of each frame with delay constraint.....	115
figure 5.7 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR of each frame with delay constraint.....	116
figure 5.8 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR of each frame with delay constraint.....	117
figure 5.9 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay 100 ms. (a) Buffer fullness level (bits) per frame with	

List of figures

XVII

Figure	Page
maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	118
figure 5.10 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	119
figure 5.11 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	120
figure 5.12 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	121
figure 5.13 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	122
figure 5.14 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	123
figure 5.15 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay 150 ms. (a) Buffer fullness level (bits) per frame with	

List of figures

XVIII

Figure	Page
maximum buffer size (dashed line).	124
(b) Average PSNR of each frame with delay constraint.....	
figure 5.16 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	125
figure 5.17 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	126
figure 5.18 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	127
figure 5.19 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	128
figure 5.20 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	129
figure 5.21 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay 200 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	

List of figures

XIX

Figure	Page
(b) Average PSNR of each frame with delay constraint.....	130
figure 5.22 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	131
figure 5.23 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	132
figure 5.24 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	133
figure 5.25 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	134
figure 5.26 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	135
(b) Average PSNR of each frame with delay constraint.....	
figure 5.27 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay 400 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).	
(b) Average PSNR of each frame with delay constraint.....	136

List of figures

xx

Figure	Page
figure 5.28 (a1)-(c1) show frames 40,41,42 of Akiyo video sequence coded with our proposed rate control scheme at 32 kbps , (a2)-(c2) are the same frames that are coded with. H.264 JM 8.6 rate control.....	137
figure 5.29 (a1)-(c1) show frames 59,60,61 of Claire video sequence coded with our proposed rate control scheme at 32 kbps , (a2)-(c2) are the same frames that are coded with. H.264 JM 8.6 rate control.....	138



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

List of Abbreviations

AVC	Advanced Video Coding
Cb,Cr	Chrominance
CBR	Constant Bit Rate
DCT	Discrete Cosine Transform
FMO	Flexible Macroblock Ordering
ISO	International Standards
ITUT	International Telecommunication Union
JM	Joint Model
MPEG	Motion Picture Expert Group
MSE	Mean Squared Error
MB	Macroblock
MV	Motion Vector
PDF	Probability Density Function
QP	Quantization Parameter
R-Q	Rate-Quantizer
R-D	Rate-Distortion
SRC	Scalable Rate Control
TMN	Test Model Near Term
TM	Test Model
UTQ	Uniform Threshold Quantizer
VM	Verification Model
VO's	Video Objects
VLC	Variable Length Coding
Y	Luminance

Chapter 1

Introduction

1.1 Background and significance of the research problem

Multimedia communications has experienced rapid growth and commercial success in the last decade. Many multimedia applications such as, digital television broadcasting, video streaming, video conferencing and video-on-demand require video coding schemes that can provide an acceptable quality of service to the end users. However, the bandwidth of channel is limited, video source is usually encoded with low-bit-rate video compression standards. Practical video coding standards are usually “lossy” coding, that is, some of the data in the original sequence are lost during the encoding process. Lossy video coding technique uses transform coding, and motion compensated prediction, quantization, zig-zag scan, variable length coding (VLC), as the building blocks diagram of typical video encoder is shown in Fig 1. The general method of block-based transform coding such as discrete cosine transform (DCT) and integer transform, involves the subdivision of the original image into smaller blocks of size $N \times N$, typically 8×8 pixels. For intra coding, the transform are applied to the image itself, for nonintra coding, only residual images obtained by performing a motion compensated prediction and the transform is applied to these residuals. These coefficients are then quantized and entropy coded.

For the source coding, efficient rate control strategy is employed to obtain the best picture quality subject to the bit rate constraint. In the coding process, the mean bit rate of the coded data is proportional to the decoded video quality. As in a block-based hybrid video coding, quality loss occurs during the quantization process. A large quantization step size leads to poorer visual video quality when decoded sequence. But a large quantization step size produces a more highly compressed coded video data. So that, the quantization step size is the one of the important key parameters that can be adjusted to control the bit rate of the encoded video sequence. For fixed bit rate channel, encoded video data must be transmitted across the channel at a constant bit rate. This creates a problem since video data encoded using any of the

standards has an inherently highly variable bit rate depend on the complexity of each frame.

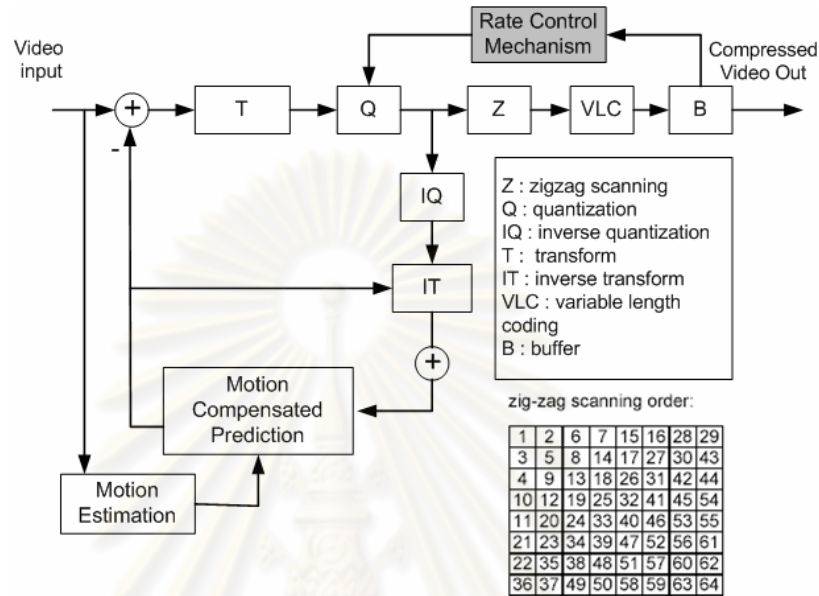


Figure 1.1 Block diagram of a video encoder

To provide effective and reliable video communication, rate control plays a key role to assign suitable number of bits to each video frame and to ensure the generation of constant bit rate stream into channel. Due to its importance, rate control algorithms have always been challenging research issues. Several rate control schemes in various international video coding standards such as Reference Model 8 (RM8) for H.261 [40], Simulation model 3 (SM3) for MPEG-1 [42], Test Model 5 (TM5) [1-3] for MPEG-2, Test Model Near Term 8 (TMN8) [4-6] for H.263, Verification model 8 (VM-8) [7-11] for MPEG-4 and Joint Model (JM) for H.264/MPEG-4 Part 10 [12-15] have been proposed. These video coding standards employ efficient compression techniques to remove the spatial and temporal redundancy within and between frames. Due to the limited storage size or communication bandwidth, quantization step size is introduced to compress the bit rate of the video signal such that the size or bandwidth constraints can be met properly. Generally, the rate control part is an informative part in video coding

standards. Standard body leaves the flexibility for designers to develop the suitable scheme for specific applications.

In real time video transmission, delay is very critical because frames of video must be presented to the viewer at a constant interval. Real-time coded video cannot tolerate large variations in end-to-end delay, because frames of video cannot be displayed at the decoder beyond the time limit. In video coding process, the instantaneous bit-rate after compression may be very different from the channel bit-rate. In order to map the instantaneous video bit-rate with constant bit rate (CBR) channel, data produced by a video encoder is buffered before transmission. This has been done to smooth out the variations in data rate. This means that the quality of the decoded sequence varies depending on the output bit rate of the encoder. This variation in quality can be particularly obvious at low transmission bit rates. While matching the desired bit rate, smooth video quality and low delay constraint are the factors rate control needs to consider.

To develop efficient rate-control algorithm, it requires knowledge of information theory, in particular, Rate-Distortion theory [16]. The basic problem in rate-distortion theory can be stated as follows: given a source distribution and a distortion measure, find the minimum expected distortion achievable at a particular rate. Equivalently, find the minimum rate description required to achieve a particular distortion. In conventional rate distortion theory, the bit rate is usually understood as the number of bits per data sample to be stored or transmitted. In general, the notion of distortion is defined as the variance of the difference between input and output signals, i.e., the mean squared error (MSE).

The functions that relate the rate and distortion are found as the solution of the following minimization problem, as shown in eq.(1),

$$\operatorname{argmin} D(Q), \text{ subject to } H(Q) \leq R_{MAX} \quad (1.1)$$

,where $D(Q)$ and $H(Q)$ are defined the distortion and the entropy functions-, i.e., rate function due to the quantization (Q), respectively. R_{MAX} denotes the maximum bit rate. Thus, if using uniform quantizer, one should be able to find rate and distortion of

any quantization step size and input signal distribution of transform coefficients input signal.

In addition to R-D modeling, the knowledge of the statistical behavior in term of distribution of the transform coefficients is important to the design of the encoder. This implies that we can find rate model and distortion model based on input statistical distribution. From there, the rate model is proposed to find optimal or sub-optimal choice of quantization parameter which is directly related to distribution. Later on, without losing generality, we will use DCT as an example of block-based transform coding, as it is widely used in standard video coding. For DCT, the transform coefficients are categorized into DC and AC coefficients. Since DC coefficient represents the average of energy of the signal, its values are uniformly distributed. However, the characteristics of the rest of AC coefficients are varied according to the nature of each pixel in a particular block. Several studies on the statistical probability distribution then focus on the distribution of AC-coefficients have been studied. Among several, Laplacian distribution [17-22] is widely used in practice with different rate models for each video coding standard, for example, TMN8 rate model in H.263 [6] and quadratic model in H.264 [23]. Cubic spline model [24] in MPEG-2 has shown more accurate estimation rate characteristics of video sequence, but it requires very high computational complexity. Recently the work in [25] presented the observation that Cauchy distribution provides more accurate estimate of the statistical distribution of DCT coefficients in typical video sequence than that of Laplacian distribution and also proposed frame bit allocation using Cauchy rate models for H.264 encoder. However the algorithm in [25] did not applicable for low bit rate and low delay applications. Moreover, the rate and distortion characteristics are considered separately. The parameters of the rate model are updated based on the previously encoded frame and a constant factor with the assumption that the distortion is constant regardless of the frames. In this connection, our work is based on the assumption that AC coefficients follow a Cauchy distribution.

In this dissertation, three main contributions in video coding research have been proposed: Cauchy based rate-distortion model, Cauchy based rate control for low bit-rate H.264 video coding, and Cauchy based rate control for low bit-rate H.264 video coding under low-delay constraint. Firstly, as already mentioned, we investigate

several mathematical models for rate-distortion in video coding. Based on the simulation results, we found that Cauchy distribution is better fit the AC coefficients distribution of video sequences. Thus, we model rate and distortion model based on Cauchy density function. The Lagrange Multiplier technique is used to find the rate and distortion model subject to get the target bit rate constraint resulting in the optimum choice of quantization step sizes. We then propose a new rate control that applies Cauchy rate-distortion optimization model for low bit-rate H.264 video coding. Simulation results indicate our proposed rate control improves video quality with the same computational time compared to H.264 JM 8.6 reference software. In real time video communication, low delay constraint is a considerable factor we need to consider in the rate control otherwise the video quality degradation in terms of frames skipped would happen. In this dissertation, we also propose a new rate control that works in the low-delay constraint scenario. Simulation results indicate our proposed rate control improves video quality especially in terms of the reduction in the number of frames skipped with the same computational time compared to H.264 JM 8.6 reference software.

1.2 Literature review

In this section, we review some recent researches of model based rate distortion functions and some recent advanced rate control techniques in video coding.

To achieve the best video quality, rate-distortion theory [16] is the theoretical foundation of rate control. It is originated from Shannon's paper [27],[28] and forms a basic foundation part of information theory [29] and lossy source coding [30-32]. A lossy source coding scheme, such as video coding, concentrates on the tradeoff between the distortion and bit rate, so called rate –distortion function. The basic principle is that decreasing distortion implies increasing rate, i.e., increasing source quality and vice visa. In rate distortion theory, the rate distortion function is defined as the optimization of a rate–distortion function to find the optimal quantizer in a lower bound of the source rate at a given distortion level. In general, the optimization techniques can be implemented using two approaches, Lagrange optimization [26],[33] and dynamic programming technique [34]-[37] are two popular

techniques to find the optimal or nearly optimal bit-allocation for each video frame. One difficulty is that the optimization scheme requires a huge amount of computations especially for dynamic programming. Then in our work, we proposed to use Lagrange optimization to optimally allocate available bits for each video frame to best represent pictures in rate control because of the application of our work is proposed for real time video coding with low complexity and low delay constraints.

Another type of approach to find optimal quantizer for rate-control is through mathematical modeling. Formula is derived from optimization based on the input statistical distribution such as DCT coefficient probability distribution and the rate and distortion models. In the earliest studies, the AC coefficients of the DCT were conjectured to have Gaussian distributions [38],[39]. Later, several other distribution models were reported, including generalized Gaussian and Laplacian distributions [17]-[22]. Other studies modeled the DCT coefficients using more complex probability density functions. In [19], Muller used a generalized Gaussian function that includes Gaussian and Laplacian probability density function as special cases. *Eude et al.* reported that the DCT coefficients could be modeled by proper linear combination of a number of Laplacian and Gaussian probability density function [20]. Comparing their models with Laplacian, Gaussian and Cauchy probability density function, they claimed that the distribution of DCT coefficients follow neither Cauchy nor Laplacian only but were most accurately modeled as a mixture of Gaussian. Although a generalized Gaussian density can model the statistics of the DCT coefficients more accurately, it is not widely used in practice because it is difficult to analyze mathematically. In [25], *N.Kamaci* and *Y.Altunbasak* collected the actual distributions of DCT coefficients in image and video data and found that those distributions differ significantly from the Laplacian distribution in most cases. In their findings, the actual distributions of DCT coefficients are resemblance to Cauchy distribution, and thus Cauchy distribution should provide more accurate estimate of the rate and distortion than that of Laplacian distribution.

In video coding research, mathematical expressions of rate - distortion characteristics have been developed for various video coding standards. *Ding et al.* [45] proposed a generic rate-quantizer model according to the changes in picture activity and feedback-based bit allocation. In [46], a source model is derived from the

classical rate-distortion (R-D) theory to describe the relationships between the rate, distortion and quantization parameters whilst in [47] an adaptive model - driven bit allocation method based on a parametric rate - distortion model was proposed which incorporated a region classification scheme. *Ribas-Corbera et al.* [6] presented a logarithmic model for bit rate and distortion by using Lagrange optimization to minimize the distortion subject to the bit rate constraint. A spline method was reported in [24] and a quadratic rate-quantization model was proposed in [23], respectively. *Cheng et al.* [48] studied the linear relationship between the activity measure and bit rate and derived an empirical first-order bits model. *He. et al.* [49],[50] proposed a linear ρ -domain R-D model and the corresponding rate distortion analysis framework, where ρ is measured by the percentage of zeros of the DCT coefficients.

Most of rate control algorithms adopted in video coding standards are designed for both CBR and variable bit rate (VBR) applications. In each of video coding standards, rate control schemes were developed for the different simulation models and were implemented in the reference software. The name of the reference software in each standard is different. Examples include RM8 for H.261 [40], SM3 for MPEG-1 [42], TM5 for MPEG-2 [2], TMN8 for H.263 [5], VM8 for MPEG-4 [7] and JM for H.264/MPEG-4 part 10 [13]. H.261 is the first video coding standard intended for ISDN video conferencing application. The H.261 rate control implemented in RM8 simply monitors the buffer status to adjust the quantization step size. The quantization step size, q , is calculated as a linear function of the buffer level as in eq.(1.1),

$$q = 2 \left[\frac{B}{200 \times p} \right] + 2 \quad (1.1)$$

, where p is the multiplier used in specifying the bits rates, B is the buffer level.

The MPEG-1 standard [42] is a multimedia standard focused on storage of multimedia compression contents. The MPEG-1 rate control implemented in SM 3 allocates the total number of bits among the various types of pictures. In MPEG-1, the adaptive quantization algorithm [43] is proposed. Each macroblock (MB) is classified

using class $cl(r,c)$, i.e., the coding difficulty. The quantization step size in each class, $Q(r,c)$, with overall minimum step size, Q_{low} , is assigned according to the eq.(1.2),

$$Q(r,c) = Q_{low} + \Delta Q \times cl(r,c) \quad (1.2)$$

, where r,c is denoted the row and column coordinates of macroblock, respectively and ΔQ is typically 1 or 2.

MPEG-2 [2] is a continuation standard after the adoption of MPEG-1. MPEG-2 was designed to use in the high quality bit rates application such as broadcasted digital television and compression video contents stored in high capacity storage media as namely Digital Video Disc (DVD). The TM5 rate control algorithm [1] for MPEG-2 is described as follows. The quantization parameter for each i th MB (Q_i) is defined as in eq.(1.3),

$$Q_i = \left(\frac{F_i \times 31}{r} \right) \quad (1.3)$$

, where $r = 2 \times \frac{R}{f}$, R and f denote the bit rate and the frame rate respectively, and F_i is the fullness of the appropriate virtual buffer.

H.263 is the first video coding standard the supports lower bit rate than earlier H.261 standard. The TMN8 rate control algorithm [5] is designed for H.263 video coding. The quantization parameter (QP) is adapted to achieve the target bit rate. It is based on the logarithmic R-Q model [6] as in eq.(1.4),

$$R = \begin{cases} \frac{1}{2} \log_2 \left(2e^2 \frac{\sigma^2}{Q^2} \right), & \text{if } \frac{\sigma^2}{Q^2} > \frac{1}{2e} \\ \frac{e}{\ln 2} \frac{\sigma^2}{Q^2}, & \text{if } \frac{\sigma^2}{Q^2} \leq \frac{1}{2e} \end{cases} \quad (1.4)$$

, where Q is defined the quantization step size and σ is defined as the standard deviation of each MB, respectively. Logarithmic R-Q model was derived from the entropy of a Laplacian distributed random variable with variance σ^2 .

The later standard, MPEG-4 [7], is proposed to support object based coding and currently is the standard of streaming video in the internet. In MPEG-4, a scene is viewed as a composition of several video objects (VO's) with intrinsic properties such as shape, motion, and texture. MPEG-4 adopted a scalable rate control (SRC) scheme [9] in the Verification Model (VM10) [10]. To compute the quantization parameter, the quadratic rate-quantizer (R-Q) model [23],[44] as in eq.(1.5),

$$R = \frac{X_1 \times S}{Q} + \frac{X_2 \times S}{Q^2} \quad (1.5)$$

, where X_1, X_2 denotes the model parameter, and S is the coding complexity which is denoted by mean absolute difference (MAD) between input and output signals.

Most recently, ITU-T and ISO/IEC developed the newest video coding standard, H.264/MPEG-4 Part 10 or Advanced Video Coding (AVC) standard [12]. It employs some different features such as 4×4 integer transform, the adaptive deblocking filter and intra-frame prediction are new tools of H.264. The JM rate control algorithm for H.264/AVC is used to compute the quantization step size by the quadratic rate-distortion model [12] as in eq.(1.6),

$$R = \frac{a_1 \times \sigma}{Q} + \frac{X_2 \times \sigma}{Q^2} - h \quad (1.6)$$

, where h is the number of header and motion vectors, a_1 and a_2 are model parameters, and σ is the predicted MAD of the current picture i which is predicted by a linear model according to the actual MAD of the previous picture $i - 1$.

Another important aspect of rate control algorithm is the ability to cope with low delay constraint. In real time video transmission, real time coded video cannot tolerate large variations in end-to-end delay. Therefore, a small buffer is used to reduces the delay and avoid the buffer overflow but in rate control video coding standard, when the encoder buffer higher number of bits than speculated, frame will be skipped. In additional, the visual quality is degraded and more fluctuated. Based on this problem, the earliest research has been developed efficient rate-control algorithm for the application of real time video transmission in each video coding standard, such as, *Pan Feng and Z.G.Li* [51] proposed new algorithm include an adaptive threshold

technique to adaptive target bit allocations. *Song et al.* [52] considered the optimal frame skipping based on spatial and temporal tradeoff in a rate-distortion sense. *Jiang.et.al* [53],[54] proposed a PSNR-based frame complexity estimation to improve H.264 video coding rate control. The aim is to allocate bits more accurately, especially for frames with scene changes and high motion.

1.3 Objectives

To develop new rate control algorithm using Cauchy rate-distortion model for low bit-rate H.264 video coding under normal and low delay constraints for real time video communication applications.

1.4 Scope

- 1) Propose rate-distortion (R-D) model based on Cauchy probability density function and optimize R-D model to find optimum choice of quantization parameter in video coding.
- 2) Propose new rate control scheme that applies Cauchy rate-distortion optimization model for low bit-rate H.264 video coding.
- 3) Propose new rate control scheme that applies Cauchy rate-distortion optimization model for low bit-rate H.264 video coding under low-delay constraint.
- 4) Evaluate and compare the performance of the proposed algorithm with the H.264 JM8.6 rate control in several standard test video sequences at different target bit rates between 16-256 kbps with delay time of 100-400 ms.

1.5 Research procedure

- 1) Study previous research papers relevant to the research works of the dissertation.
- 2) Investigate, develop and design new rate control incorporate Cauchy R-D optimization model.
- 3) Implement proposed methods in to H.264 reference software.
- 4) Test the proposed techniques.

- 5) Collect and analyze computational results obtained from simulation programs.
- 6) Summarize the major findings as we found in step 5, compare and conclude the performance of the proposed techniques in all concerned aspects.
- 7) Check whether the conclusions meet all the objectives of the research work of the dissertation.
- 8) Write the dissertation international conference papers and international journal paper.

1.4 Expected prospects

- 1) Acquire a basic knowledge of video coding.
- 2) H.264 video coding software implementation of proposed Cauchy based rate control.
- 3) Publications in international conference proceedings, and international journal.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Chapter 2

Background

2.1 Background on video coding standards

The continuous development of video coding technology has resulted in series of international standards for image and video coding. Each of these standards supports a particular application of video coding such as video conferencing or digital television. Two standard bodies, the International Standardization Organization and International Electrotechnical Commission (ISO)/IEC and the International Telecommunications Union-Telecommunication Standardization Sector (ITU-T), have developed series of standards that have shaped the development of the visual communications industry. Each standard describes a method of representation for compressed video. The developers of each standard have attempted to incorporate the best developments in video coding technology, i.e., coding efficiency, codec complexity, functionalities, etc. Fig.2.1 show the progress of the video coding standards development.

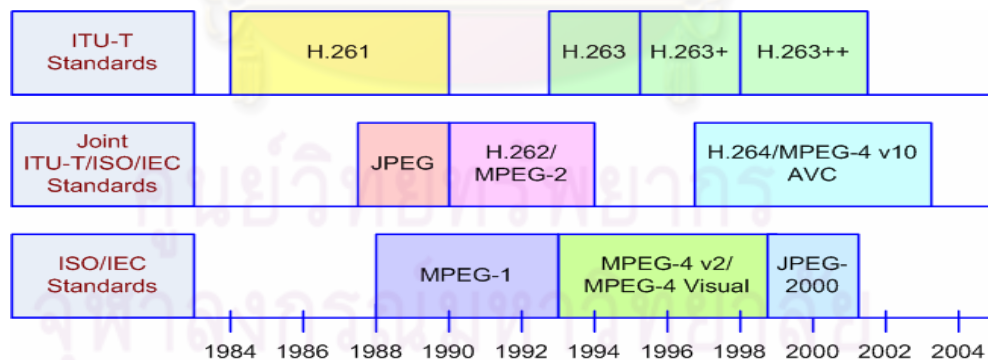


Figure 2.1 Progression of the ITU-T Recommendations and ISO/IEC standards.

As shown in Fig.2.1, we will describe the development of video coding technology of two international standards for video coding. The ITU-T has

concentrated on standards to support real-time, two-way video communications and they are responsible for the following standards;

H.261 (1990) : Video telephony over constant bit rate channels and primary aimed at ISDN channels. It was designed for data rates which are multiples of 64Kbit/s, and is sometimes called $p \times 64\text{Kbit/s}$ (p is in the range 1-30).

H.263 (1995) : Video telephony over circuit and packet switched networks. It was designed for low bit rate communication, early drafts specified data rates less than 64 kbps.

H.263+ (1998), H.263++(2001) : Extensions to H.263 to support a wider range of transmission scenarios and improved compression performance.

In parallel with the ITU's activities, the ISO/IEC has issued standards to support storage and distribution applications. The two relevant groups are JPEG (Joint Photographic Expert Group) and MPEG (Moving Picture Experts Group) and have issued,

MPEG-1 (1993) : Compression of video and audio for storage and real time play back on CD-ROM at a bit rate 1.5 Mbps.

MPEG-4 (1998) : Video and audio compression and transport for multimedia support a wide range of bit rates from 28-1024 kbps.

JPEG-2000 : JPEG 2000 was developed to provide a more efficient successor to the original JPEG. It uses the Discrete Wavelet Transform (DWT) as its basic coding method.

As shown in Fig.2.1, Joint standards committee of ITU-T and ISO/IEC was jointly development in picture and video coding for following standards;

JPEG (1992) : It aims to support a wide variety of applications for compression of still images. It can code full color images, achieving an average compression ratio of 15:1 for subjectively transparent quality.

MPEG-2/H.262 (1995) : Compression and transmission of video and audio for storage and broadcast applications at bit rates 3-10 Mbps.

H.264 (2002) : H.264/MPEG-4 Part 10 or AVC contains a number of new features that allow it to compress video much more effectively than older standards.

Average bit rate reduction of 50% given fixed fidelity compared to any other standard.

The concept of standardized video coding schemes are based on hybrid video coding. Fig.1.1 shows generalized block diagram of a hybrid video encoder. In general, each pixels in a input video signal consists of three components : R (Red), G (Green), B (Blue). To compress video signals, RGB videos are usually converted into another color model called YCbCr where Y is luminance component which represents the brightness information, Cr and Cb denote chrominance red and chrominance blue representing the color information. The reason behind this conversion is because human visual system (HVS) is most sensitive to Y component. So Y component are encoded with full resolution. But HVS is less sensitive to Cb and Cr components, so we subsample Cb and Cr components. This process is called chrominance subsampling. By doing this, we can reduce data without affecting perceived visual quality. In general video coding standard, the chrominance subsampling is 4:2:0, i.e., keeping 4 pixels of Y component and 1 pixel of Cb and Cr. Three different picture types are supported in video coding standard which are I-frame, P-frame and B-frame. For I-frame is being coded by intra-frame coding method, i.e., similar to the image compression method like JPEG. When encoding I-frame, we only reduce the spatial redundancy in the picture without referencing other pictures. For P-frame and B-frame, the method of inter-frame coding is utilized, i.e., the coding requires the reference to the previous and/or future frames. For encoding process, the input image is divided into macroblocks (MB) of Y, Cr and Cb. Therefore, a MB consists of one block of 16x16 pixels for the luminance component and two blocks of 8x8 pixels for the color components. These MBs are coded in intra-frame coding or inter-frame coding. In Inter-frame coding, a MB is predicted using motion compensation.

For motion compensated prediction a displacement vector is estimated and transmitted for each block (motion data) that refers to the corresponding position of its image signal in an already transmitted reference image stored in memory. In intra mode, the standards set the prediction signal to zero such that the image can be coded without reference to previously sent information. The prediction error, which is the difference between the original and the predicted block is transformed, quantized and entropy coded. In order to reconstruct the same image on the decoder side, the

quantized coefficients are inverses transformed and added to the prediction signal. The result is the reconstructed MB that is also available at the decoder side. These MB is stored in a memory. In order to achieve high compression ratio and coding efficiency, video coding standard uses hybrid coding techniques to reduce both spatial redundancy and temporal redundancy. These techniques in hybrid video coding are described as follows.

2.1.1 Transform domain coding

Transform coding has been studied extensively during the last two decades and has become a prominent compression method for still image and video coding. The purpose of Transform coding is to transform the picture content from pixel domain representation to frequency domain representation and compact the energy, that is, most of the energy in the transformed data should be concentrated into small number of coefficients. To this aim the input images are split into disjoint blocks of pixels (i.e., size $N \times N$ pixels). The transform can be represented as a matrix operation using an $N \times N$ Transform matrix A to obtain the $N \times N$ transform coefficients, c , that is called the forward transform ($c = AbA^T$). A^T denotes the transpose of the transformation matrix A . The transformation is reversible, since the original $N \times N$ block of pixels b can reconstructed using a linear and separable inverses transformation, i.e., $b = A^T cA$.

DCT has chosen to be a transform method for still image and video coding standards due to their energy compaction performance and the availability of fast DCT algorithms suitable for real time implementations. For 8×8 block of pixels are forward transformed into frequency domain to generate the 8×8 DCT coefficients. If we define $f(m, n)$, $0 \leq m \leq 7$, $0 \leq n \leq 7$, as pixels-values, the 2-D of 8×8 DCT coefficients $F(u, v)$, $0 \leq u \leq 7$, $0 \leq v \leq 7$, can be computed by eq.(2.1)

$$F(u, v) = \frac{1}{4} C(u)C(v) \sum_{m=0}^7 \sum_{n=0}^7 f(m, n) \cos\left(\frac{(2m+1)u\pi}{16}\right) \cos\left(\frac{(2n+1)v\pi}{16}\right) \quad (2.1)$$

where, $C(u), C(v) = 1/\sqrt{2}$, for $u, v = 0$, and $C(u), C(v) = 1$, otherwise.

In DCT, coefficient in location (0,0) is called DC coefficient and the other values are called AC coefficients. In general, we use large quantization step in quantizing AC coefficients, and use small quantization step to quantize DC coefficient so as to preserve the important frequency components that are vital to human visual perception.

2.1.2 Quantization

The transformation of the pixels does not actually yield any compression. As we discussed above, a block of 64 pixels is transformed into 64 coefficients. The energy in both the pixel and the transform domains are equal, hence no compression is achieved. However transformation causes the significant part of the image energy to be concentrated at the lower frequency components with the majority of the coefficients having little energy. Quantization is the only lossy compression while Variable Length Coding (VLC) is the lossless compression, i.e., statistical method compression of the DCT coefficients lead to bit rate reduction. Moreover as human eye are less sensitive to picture distortions at higher frequencies, one can apply even coarser quantization at these frequencies to give greater compression. A coarser quantization step sizes force more coefficients to zero and as a result more compression is gained with poorer video quality. On the other hand, a finer quantization step size results in a higher bit-rate or lower compression ratio with better video quality.

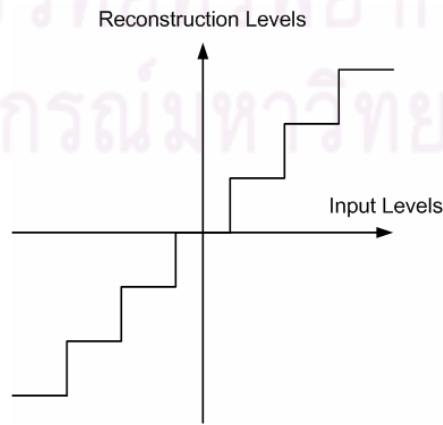


Figure 2.2 A typical quantizer

The typical quantizer used in video coding standards is shown in Fig.2.2. Quantization usually is the only operation that causes information loss in the encoding. The quantization step-size is usually determined by the rate control scheme, which is discussed in the Section 2.1.5. After quantization, the quantized DCT coefficients are zig-zag scanned. After zig-zag scan, variable length coding (VLC) is applied, which is discussed in the section 2.1.3.

2.1.3 Variable length coding

For further bit rate reduction using lossless statistical method, the quantized DCT coefficients and the motion vectors are entropy coded which, DC-coefficient and 63 AC-coefficients are coded separately. We can reduce the average number of bits per coefficients or symbol if the coefficients having lower probability are assigned longer code words, whereas coefficients having higher probability are assigned shorter code words. This method is called variable length coded (VLC) or entropy coding. The length of the codes should vary inversely with the probability of occurrences of the various symbols in VLC. The bit rate required to code these symbols is defined as the logarithm of probability, p . Hence the entropy of symbols which is the minimum average bits required to code the symbol can be calculated as,

$$H(x) = -\sum_{i=1}^n p_i \log_2 p_i \quad (2.2)$$

There are two types of VLC which are employed in the video coding standards, the Huffman coding and Arithmetic coding. The Huffman coding is a practical VLC code, but its compression can never reach as low as the entropy due to the constraint that the assigned symbols must have an integral number of bits. However the Arithmetic coding provides a practical alternative to Huffman coding that can more closely approach theoretical maximum compression ratios. An arithmetic encoder converts a sequence of data symbols into a single fractional number and can approach the optimal fractional number of bits required to represent each symbol.

The concept of entropy coding can be combined with a run-length coding procedure to achieve further data reduction. This method is useful if consecutive pixels along a scan line are highly correlated. With run-length coding one codeword is allocated to a pair of input values (run, length), i.e., the number (run) of consecutive pixels along a scan line with equal values (length) can be encoded by transmitting only one codeword.

2.1.4 Motion estimation and Motion compensated prediction

Motion compensated prediction is a powerful tool to reduce temporal redundancies between frames and is thus used extensively in video coding standards (i.e., H.261, H.263, MPEG-1, MPEG-2 and H.264) as a prediction technique for temporal coding. The concept of motion compensation is based on the estimation of motion between video frames, i.e., if all elements in video scene are approximately spatially displaced, the motion between frames can be described approximately by a limited number of motion parameters (i.e., estimated motion vectors). The best prediction of an actual pixel is given by a motion compensated prediction pixel from a previous frame. Usually both prediction error and motion vectors are transmitted to the receiver. To this aim image are usually separated into disjoint blocks of pixels (i.e., 8x8 pixels, 16x16 pixels) and only one motion vector is estimated and coded for each of these blocks as in Fig.2.3

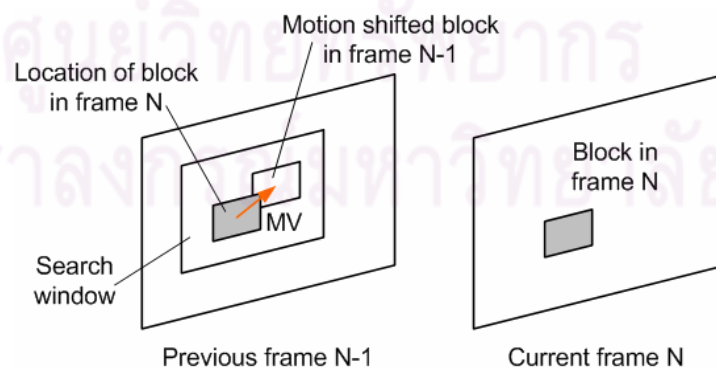


Figure 2.3 Block matching approach for motion compensation

In typical block matching, a frame is divided into blocks of $M \times N$ pixels or, more usually, square blocks of N^2 pixels. Then, for a maximum motion displacement of w pixels per frame, the current block of pixels is matched against a corresponding block at the same coordinates but in the previous frame, within the square window of width, $N + 2w$. That is called search range. Various distortion measures such as, mean squared error (MSE) and mean absolute error (MAE) can be used in the matching criterion by subtracting each pixel in a block with its motion shift counterpart in the reference block of the previous frame which, the two distortion must be minimized. As shown in Fig.2.3, one motion vector (mv) is estimated for each block in the actual frame N to be coded. The motion vector points to a reference block of same size in a previously coded frame $N-1$.

2.1.5 Rate control

The bit rate resulting from the DCT-based coding algorithm fluctuates according to the nature of each video frame. Variations in the speed of moving objects, their size and texture are the main cause for bit rate variation. The objective of rate control algorithm is to assign suitable number of bits to each frame subject to bit rate constraint. In case of constant bit rate channel, rate control will try to adjust to achieve constant bit rate of each frame. A transmission buffer is usually needed to smooth out the bit rate fluctuations and to fit the channel rate. To prevent the buffer from overflow or underflow, and achieve good video quality, a rate control scheme is applied to adjust the quantization step-sizes. Since different quantization step-sizes are used for coding video frames, the PSNR for each frame will vary. On the other hand, if we using a fixed quantization step-sizes results in relatively constant video quality but variable bit rate as shown in Fig.2.4

These examples show that a variable coded bit rate can be adapted to a constant bit rate delivery medium using encoder and decoder buffers. However, this adaptation comes at a cost of buffer storage space and delay and the wider the bit rate variation, the larger the buffer size and decoding delay. It is usually necessary to implement a feedback mechanism to control the encoder output bit rate in order to prevent the buffer from overflow or underflow.

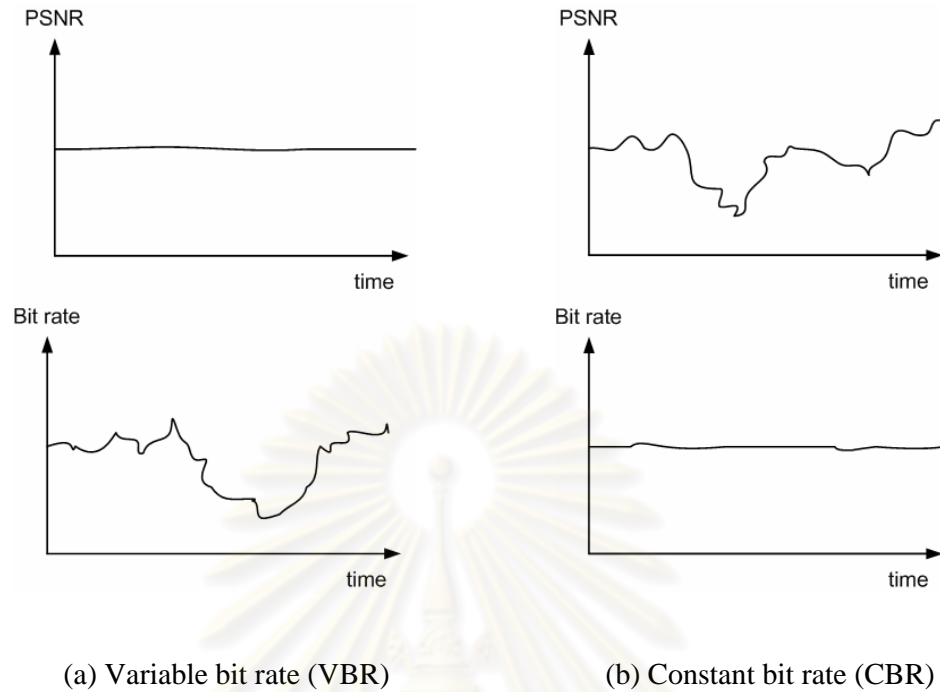


Figure 2.4 Variable bit rate versus constant bit rate channels.

Rate control involves the adjustment of several encoding parameters in order to maintain a target output bit rate. A common approach of rate control in video coding is used to determine the suitable quantization step-sizes for each frame or each macroblock (MB) in a frame to achieve the best visual video quality in order to maintain a target bit rate. How to determine the suitable rate control scheme is not specified in the video coding standards. However, the rate control algorithm has a large impact on the resultant visual video quality. A better rate control algorithm which gives suitable quantization step-sizes for coding MB can produce much better video quality. Currently, rate control is still an achieve research area.

2.1.6 Test model for video standards

Video coding standards only specify those parts that are necessary for interworking between the video encoders and video decoders from different vendors. All the standard video coding algorithms use motion compensated prediction to reduce the temporal redundancy by predicting the macroblock in the current frame from the reference frame. Many parts that do not affect the interworking between the

video encoder and the video decoder (e.g. how to compute the DCT coefficients, how to perform motion estimation, coding mode decision, rate-control, etc.) are left opened to each codec developer.

Since many parts of the standards are left opened, to demonstrate the typical video quality achievable, every video coding standard has developed a reference implementation called reference models or test models. For example, the commonly used reference models are SM3 for MPEG-1, TM5 for MPEG-2, TMN8 for H.263 and JM for H.264, that is the latest in test model of video coding standards. In the test models, specific coding-mode selection, motion estimation, and rate-control algorithm are all specified. These reference models often represent the state-of-the-art algorithms which can produce best video quality with reasonable implementation complexity. The test model is then used as the reference for performance comparison with new proposed coding methods.

2.2 H.264/MPEG-4 Part 10 (Advance video coding) standard

H.264 is a new video coding standard recommendation of ITU-T also known as International Standard or MPEG-4 part 10 Advanced Video Coding of ISO/IEC. That is the latest in a sequence of the video coding standards H.261 (1990) [40], MPEG-1 (1993) [42], MPEG-2 (1994) [3], H.263 (1995,1997) [4] and MPEG-4 (1998) [8]. These previous standards reflect the technological progress in video compression and the adaptation of video coding to different applications and network. The application from video telephony (H.261) to consumer video on CD (MPEG-1) and broadcast of standard definition or high definition TV (MPEG-2). H.264 is a standard for high quality encoding that is resilient to poor network conditions, yet high quality enough to serve as a basis for HDTV and HD-DVD encoding. H.264 providing enhanced video compression performance in view of interactive applications like video telephony requiring a low latency system and non-interactive applications like storage, broadcast and streaming of standard definition TV where focus is on high coding efficiency.

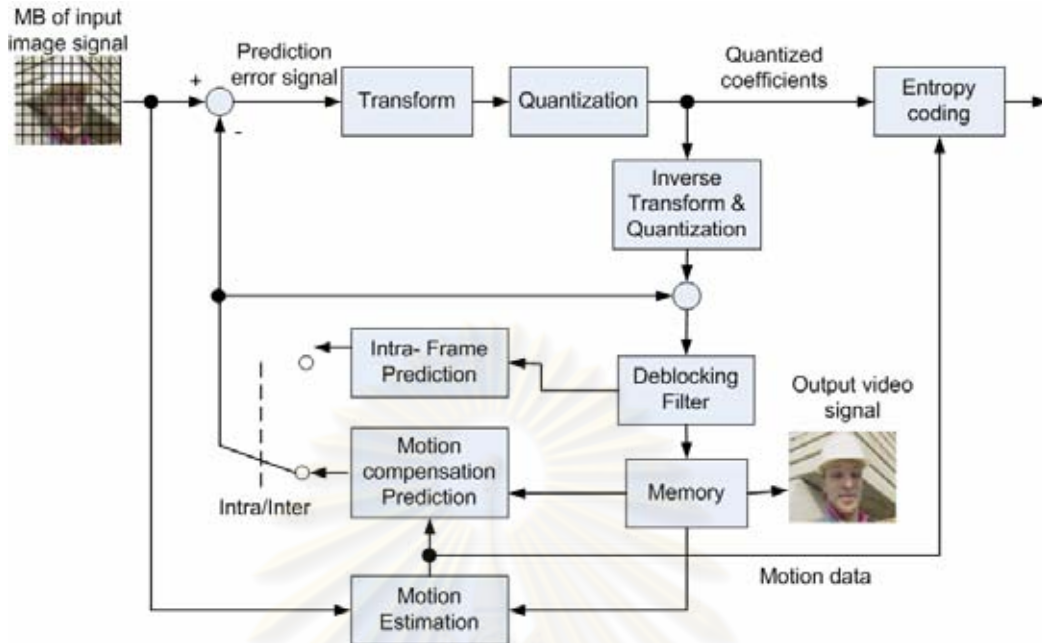


Figure 2.5 Generalized block diagram of a hybrid video encoder with motion compensation: The adaptive deblocking filter and intra-frame prediction are two new tools of H.264.

As shown in Fig.2.5 with respect to the simple block diagram of a hybrid video encoder, H.264/AVC introduces the following changes.

- To reduce the blocking artifacts an adaptive deblocking filter is used in the prediction loop. The deblocked MB is stored in the memory and can be used to predict future MBs.
- In the previous standards, the memory contains only one video frame. But H.264/AVC allows storing multiple video frames in the memory.
- H.264/AVC a prediction schemes is also used in intra mode that uses the image signal of already transmitted MBs of the same image in order to predict the block mode.
- Discrete cosine transform (DCT) used in former standards is replaced by an integer transform with basically the same properties as a 4x4 DCT in H.264.

In H.264/AVC, the macroblocks are processed in so called slices whereas a slice is usually a group of macroblocks processed in raster scan order, as shown in Fig.2.6.

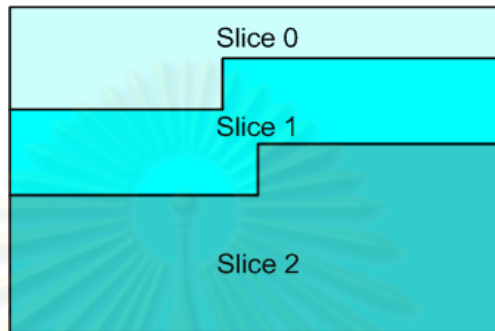


Figure 2.6 Partitioning of an image into several slices.

Five different slice types are supported in H.264 which are I-slices, P-slices, B-slices, SI-slice and SB-slices. In I-slices, all macroblocks are encoded in Intra mode. In P-slices, all macroblocks are predicted using a motion compensated prediction with one reference frame and in a B-slice with two reference frames. SI-slices and SB-slices are specific slices that are used for an efficient switching between two different bitstreams.

For the coding of interlaced video, H.264 supports two different coding modes. The first one is called “frame-mode” and the second one is called “field-mode”. In the field-mode, two fields of a frame are encoded separately. These two different coding modes can be selected for each image or even for each macroblock. The choice of the frame mode is efficient for regions that are not moving. In non-moving regions there are strong statistical dependencies between adjacent lines even though these lines belong to different fields. These dependencies can be exploited in the frame mode. In the case of moving regions the statistical dependencies between adjacent lines are much smaller. It is more efficient to apply the field mode and code the two fields separately.

H.264 supports coding and decoding of 4:2:0 progressive or interlaced video and the default sampling format for 4:2:0 progressive frames is shown in Fig.2.7. In the default sampling format, chrominance (Cb and Cr) samples are aligned horizontally with every 2nd luminance sample and are located vertically between two luminance samples. An interlaced frame consists of two fields separated in time and with the default sampling format shown in Fig.2.8.

In the next sections, we discuss the detail of each coding component of H.264 which are Intra prediction, motion compensated prediction, transform and quantization, entropy coding, the adaptive deblocking filter and the Flexible Macroblock Ordering (FMO).

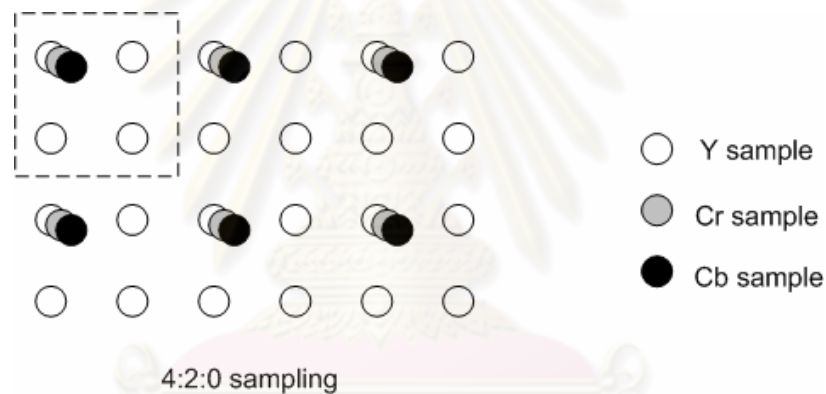


Figure 2.7 4:2:0 sampling patterns (progressive)

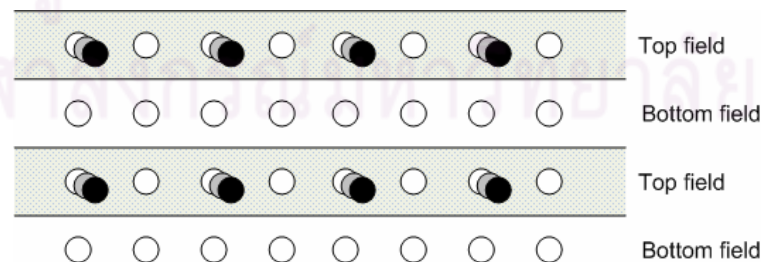


Figure 2.8 4:2:0 samples to top and bottom fields

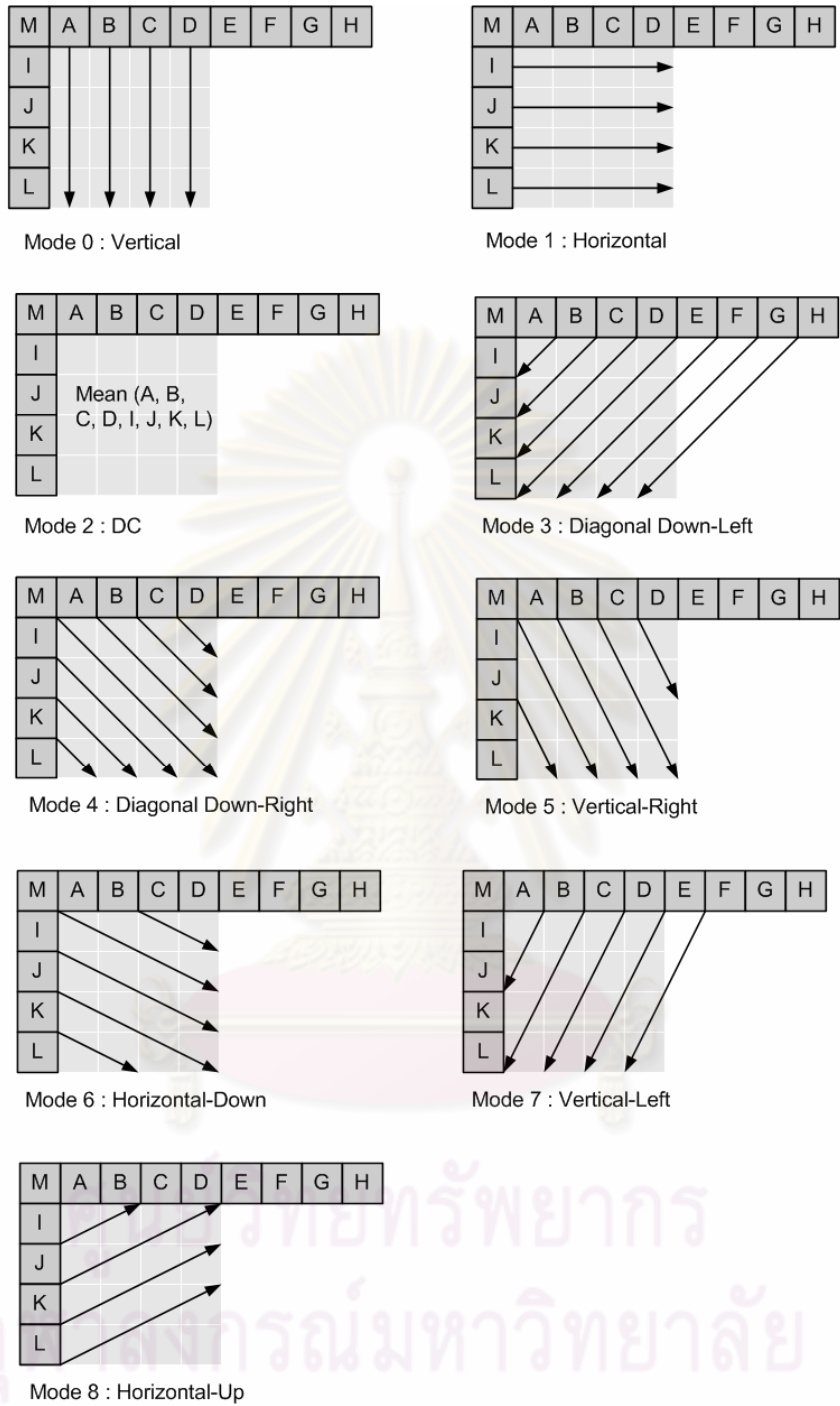
2.2.1 Intra Prediction

The intra prediction means that samples of macroblock are predicted by using only information of already coded macroblocks of the same image. Two different types of intra prediction are used in H.264 for the prediction of the luminance component Y and one intra prediction for chrominance component, Cb and Cr.

The first one for luminance component is called INTRA_4x4 and the second one is called INTRA_16x16. For INTRA_4x4 type, the macroblock which is of the size 16 by 16 pixels (16x16) is divided into sixteen 4x4 subblocks and a prediction for each 4x4 subblock of the luminance signal is applied individually. Nine different prediction modes are supported. One mode is called DC-prediction mode, whereas all samples of current 4x4 subblock are predicted by the mean of all samples neighboring to the left and to the top of current block and which have been already reconstructed at the encoder and at the decoder side, as shown in Fig.2.9 and Fig.2.10

M	A	B	C	D	F	G	H
I	a	b	c	d			
J	e	f	g	h			
K	i	j	k	l			
L	m	n	o	p			

Figure 2.9 Labeling of prediction samples 4x4



A — **M** : Neighboring samples that are already reconstructed at the encoder and at the decoder side
 : Samples to be predicted

Figure 2.10 Nine possible intra prediction modes for the intra prediction type INTRA_4x4

The samples above and to the left (labeled A-M in Fig.2.9) have previously been encoded and reconstructed and are therefore available at the encoder and decoder to form a prediction reference. The samples, a, b, c,...,p of the prediction block P (Fig.2.9) are calculated based on the samples A-M. In addition to DC-prediction mode, eight modes designed for specific prediction of each direction are support. All possible direction are shown in Fig. 2.10

For INTRA_16x16, only one prediction mode is applied for whole macroblock. Four different prediction modes are supported for the type INTRA_16x16 as shown in Fig.2.11: Vertical prediction, horizontal prediction, DC-prediction and plane-prediction. Plane-prediction uses linear function between the neighboring samples. This mode works very well for coding very smooth areas of a picture. The mode of operation of those modes is the same as the one of the 4x4 prediction modes. The difference is that they are applied for the whole macroblock instead of for a 4x4 subblock. The efficiency of this mode is high if the signal is very smooth within the macroblock.

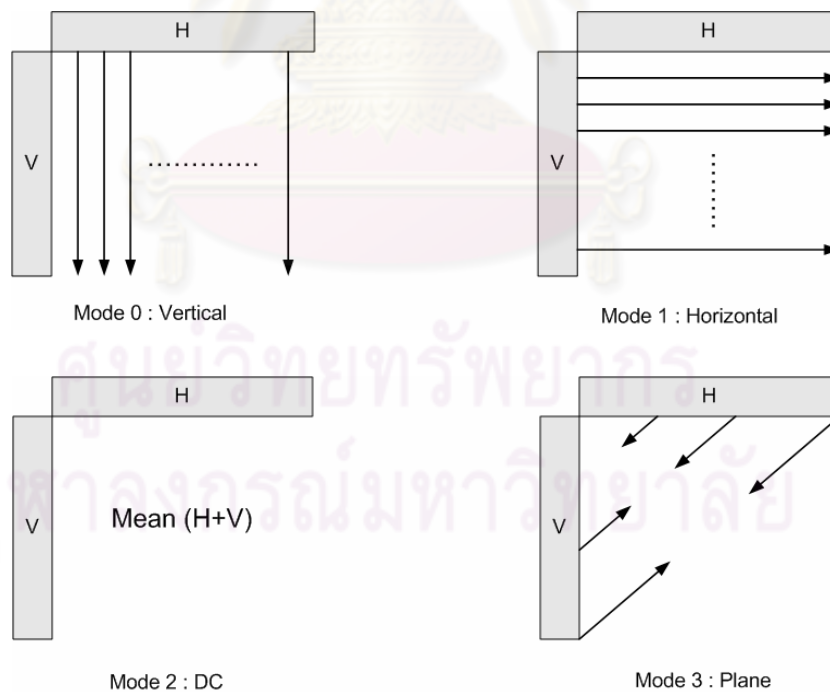


Figure 2.11 Four possible intra prediction modes for the intra prediction type INTRA_16x16

Intra prediction for chrominance blocks support only one mode. A 8x8 chrominance macroblock consists of four 4x4 blocks A, B, C, D as shown in the Fig. 2.12. S0, S1, S2 and S3 are the sum of four neighboring samples.

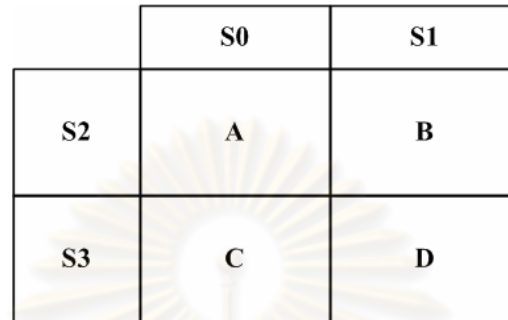


Figure 2.12 Intra prediction for chrominance macroblock

There are four prediction cases depending upon whether S0, S1, S2 or S3. For example,

$$\begin{aligned}
 A &= (S0 + S2 + 4)/8 \\
 B &= (S1 + 2)/4 \\
 C &= (S3 + 2)/4 \\
 D &= (S1 + S3 + 4)/8
 \end{aligned}
 \tag{2.3}$$

2.2.2 Motion Compensation Prediction

In addition to the intra macroblock coding types, various predictive or motion compensated coding types are specific partitioning P-slices macroblocks. For motion compensation prediction, P-types macroblocks are predicted from the image signal of already transmitted reference images. In H.264, each macroblock can be divided into smaller partitions with luminance block sizes of 16x16, 16x8, 8x16, and 8x8 pixels are supported. In case of 8x8 sub-macroblock is further divided into partitions with block sizes of 8x4, 4x8 or 4x4. The partitions of a macroblock and sub-macroblock are shown in Fig.2.13.

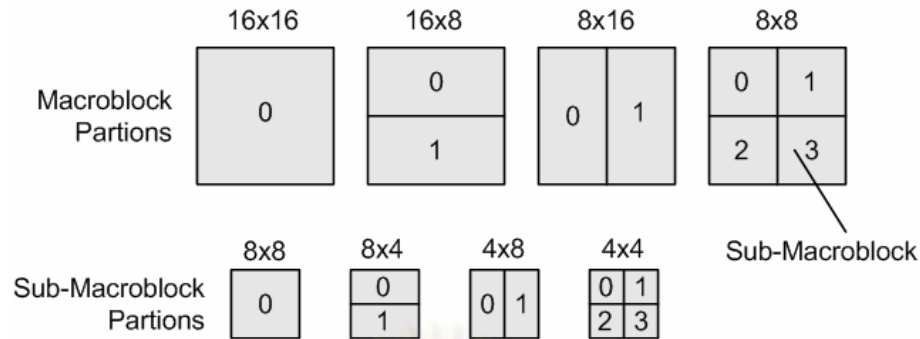


Figure 2.13 Partitioning of a macroblock and a sub-macroblock for motion compensated prediction.

In former standards such as MPEG-4 or H.263, only blocks of the size 16x16 and 8x8 are supported. A displacement vector is estimated and transmitted for each block, refers to the corresponding position of its image signal in an already coded reference image. In former MPEG standards this reference image is the most recent previous image. In H.264 it is possible to refer to several preceding images. For this purpose, an additional picture reference parameter has to be transmitted together with the motion vector. This technique is denoted as motion-compensated prediction with multiple reference frames. The motion vector supports multipicture motion-compensated prediction. That is more than one prior coded picture can be used as reference for motion-compensated prediction. Fig.2.14 illustrates the concept.

Multiframe motion compensated requires prediction both the encoder and decoder have to store the reference pictures used for Inter-picture prediction in a multiple picture buffer. The decoder replicates the multiple reference frames after of the encoder, according to the reference picture buffering type and any memory management control operations that are specified in the bitstream. Unless the size of the multi-picture buffer is set to one picture, the index at which the reference picture buffer has to be signaled. The reference index parameter is transmitted for each motion compensated 16x16, 16x8, 8x16 or 8x8 luma block.

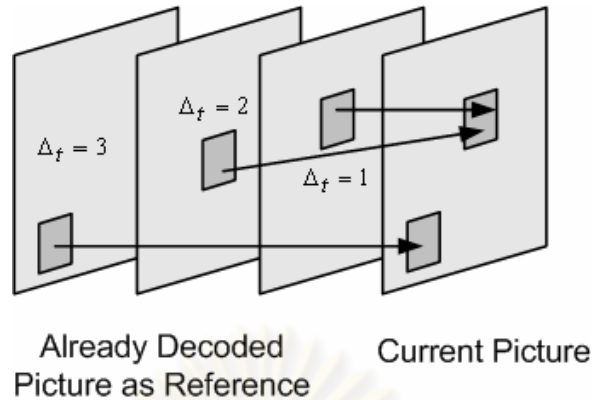


Figure 2.14 Motion-compensated prediction with multiple reference frames. In addition to the motion vector, also an image reference parameter Δ_f is transmitted.

The accuracy of motion compensation is a quarter of a sample distance (quarter-pel or $\frac{1}{4}$ pel). In cases where the motion vector points to an integer-sample position, the prediction signals are the corresponding samples of the reference frame, otherwise they are obtained by using interpolation at the sub-sample positions. The prediction values at half-sample positions are obtained by applying a one-dimensional 6-tap (Finite Impulse Response (FIR) filter). Prediction values at quarter-sample positions are generated by averaging samples at the integer-sample and half-sample positions. The H.264 generally allows unrestricted motion vectors, i.e., motion vectors can point outside the image area. In this case, the reference frame is extended beyond the image boundaries by repeating the edge pixels before interpolation. The motion vector components are differentially coded using either median or directional prediction from neighboring blocks. No motion vector component prediction takes place across slice boundaries.

For B-picture in the classical concept, B-pictures are pictures that are encoded using both past and future pictures as references. The prediction is obtained by linear combination of forward and backward prediction signals. B-picture utilize a similar macroblock partitioning to P-pictures. Besides the P_16x16, P_16x8, P_8x16, P_8x8, and the intra coding types, bi-predictive prediction and another type of prediction

called direct prediction, are provided. The motion vector coding is similar to that of P slices with the appropriate modifications because neighboring blocks may be coded using different prediction modes.

2.2.3 Transform and quantization

Previous standards such as MPEG-1, MPEG-2, and H.263 made use of the 8x8 Discrete transform (DCT) as the basic transform. Instead of the DCT, different integer transforms in mainly 4x4, in special cases 2x2. This smaller block size of 4x4 instead of 8x8 enables the encoder to better adapt the prediction error coding to the boundaries of moving objects, to match the transform block size with smallest block size of the motion compensation, and to generally better the transform to the local prediction error signal. Three different types of transforms are used. The first type is applied to all samples of all prediction error blocks of the luminance component Y and also for all blocks of both chrominance components Cb and Cr regardless of whether motion compensated prediction or intra prediction was used. The size of the transform is 4x4. Its transform matrix H_1 is shown in Fig.2.15.

$$H_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad H_3 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Figure 2.15 Matrices H_1 , H_2 and H_3 of three different transforms applied in H.264

If the macroblock is predicted using the type INTRA_16x16, the second transform, a Hadamard transform with matrix H_2 as in Fig. 2.15, is applied in addition to the first one. It transforms all 16-DC coefficients of the already transformed blocks of the luminance signal. The size of this transform is also 4x4. The transform is also a Hadamard transform but of size 2x2. It is used for the transform of the 4-DC coefficients of each chrominance component. Its matrix H_3 is shown in Fig.2.15.

The transmission order of all coefficients is shown in Fig.2.16. If the macroblock is predicted using the intra prediction type INTRA_16x16 the block with the label “-1” is transmitted first. This block contains the DC coefficients of all blocks of the luminance component. Afterwards all blocks labeled “0”- “25” are transmitted whereas blocks “0”-“15” consist of all AC coefficients of blocks of the luminance component. Finally, blocks “16” and “17” consist of the DC coefficients and blocks “18”-“25” the AC coefficients of the chrominance components.

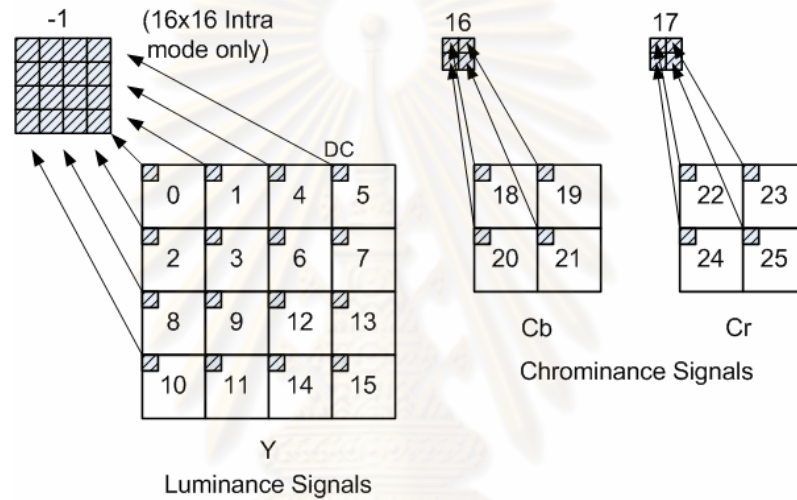


Figure 2.16 Transmission orders of all coefficients of a macroblock

For the quantization of transform coefficients, H.264 uses scalar quantization. The mechanisms of the forward and inverse quantizers are complicated by the needs to avoid division or floating point arithmetic and incorporate the post and pre scaling matrices. The basic forward quantizer operation as follow;

$$Z_{ij} = \text{round}(Y_{ij} / Q_{step}) \quad (2.4)$$

where, Y_{ij} is a coefficient of the transform describe above, Q_{step} is a quantizer step size and Z_{ij} is a quantized coefficient.

A total of 52 value of Q_{step} are supported by the standard and these are indexed by a quantization parameter (QP). The values of Q_{step} corresponding to each QP are shown in Fig.2.17. Note that Q_{step} doubles in size for every increment of 6 in QP, Q_{step} increases by 12.5% for each increment of 1 in QP. The wide range of quantization step sizes makes it possible for an encoder to accurately and flexibly control the trade off between bit rate and quality. After quantization, the quantized transform coefficients of block are generally scanned in zigzag scan and transmitted using entropy coding methods.

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	...
QStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	...
QP	...	18	...	24	...	30	...	36	...	42	...	48	...	51
QStep		5		10		20		40		80		160		224

Figure 2.17 Quantization step sizes in H.264/AVC codec

2.2.4 Entropy Coding

In H.264, two methods of entropy coding are supported. A low-complexity technique based on switched sets of variable length codes (VLC) is called Context Adaptive VLC (CAVLC). In this scheme, VLC tables for various syntax elements are switched depending on already-transmitted syntax elements. Since the VLC tables are well designed to match the corresponding conditioned statistics, the entropy coding performance is improved in comparison to schemes using just a single VLC table. The efficiency of entropy coding can be improved if the computationally more demanding algorithm of context-based adaptive binary arithmetic coding (CABAC) is used. On the other hand, the use of arithmetic coding allows the assignment of non-integer number of bits to each symbol of an alphabet, which extremely beneficial for symbol probabilities much greater than 0.5. On the other hand, the use of adaptive codes permits adaptation to non-stationary symbol statistics. Another important property of CABAC is its context modeling. The statistics of already-coded syntax elements are

used to estimate the conditional probabilities. These conditional probabilities are used for switching several estimated probability models. Both methods (CAVLC and CABAC) represent major improvements in term of coding efficiency compared to the techniques of statistical coding traditionally used in prior video coding standards.

2.2.5 Deblocking Filter

One particular characteristic of block-based video coding is visible block structures. Block edges are typically reconstructed with less accuracy than interior pixels and “blocking” is generally considered to be one of the most visible artifacts with the present compression methods. For this reason H.264 defines an adaptive in-loop deblocking filter, where the strength of filtering is controlled by the values of several syntax elements. The blockiness is reduced without much affecting the sharpness of the content. Consequently, the subjective quality is significantly improved. At the same time the filter reduces bit-rate with typically 5-10 % while producing the same objective quality as the non-filtered video as shown in Fig.2.18.



Figure 2.18 Performance of deblocking filter for highly compressed pictures. (a) Without the deblocking filter (b) with deblocking filter

The filtered image is used for motion compensated prediction of future frames and this can improve compression performance because the filtered image is often a more faithful reproduction of the original frame than a blocky unfiltered image. The

default operation of the filter is as follows; it is possible for the encoder to alter the filter strength or to disable the filter.

Filtering is applied to vertical or horizontal edges of 4×4 blocks in a macroblock (except for edges on slice boundaries), in the following order.

- Filter 4 vertical boundaries of the luma component in order e, f and d in Fig.2.19.

- Filter 4 horizontal boundaries of the luma component in order e, f, g and h in Fig.2.19.

- Filter 2 vertical boundaries of each chrominance component (i, j)

- Filter 2 horizontal boundaries of each chrominance component (k, l)

Each filtering operation affects up to three samples on either side of the boundary. Fig.2.20 shows four samples on either sides of a vertical or horizontal boundary in adjacent blocks p and q (p_0, p_1, p_2, p_3 and q_0, q_1, q_2, q_3).

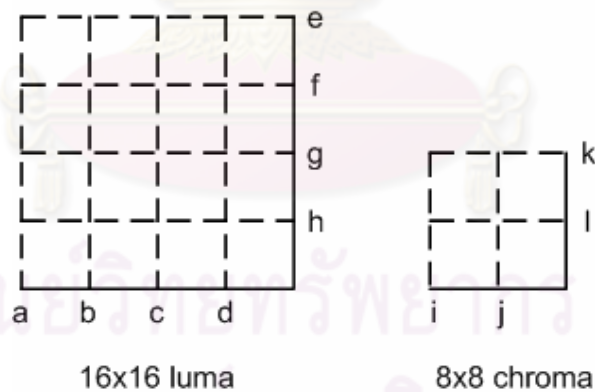


Figure 2.19 Edge filtering order in a macroblock

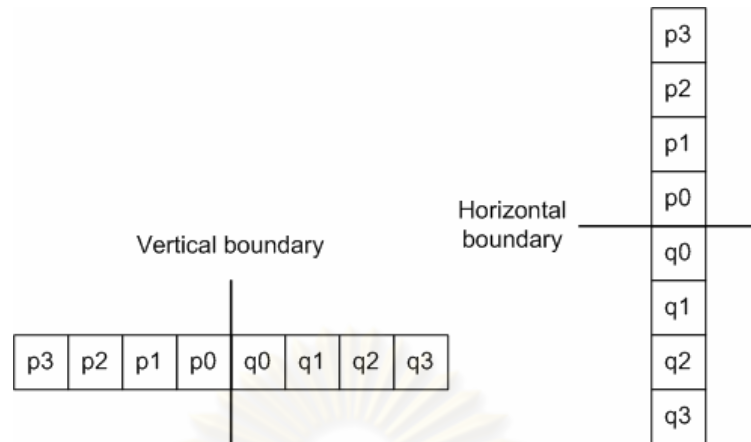


Figure 2.20 Samples adjacent to vertical and horizontal boundaries

2.2.6 Flexible Macroblock ordering (FMO)

In H.264, Flexible Macroblock ordering (FMO) modifies the way how pictures are partitioned into slices and macroblocks by utilizing the concept of slice groups. Each slice group is a set of macroblocks defined by a macroblock to slice group map, which is specified by the content of the picture parameter set and some information from slice headers. The macroblock to slice group map consists of a slice group identification number for each macroblock in the picture, specifying which slice group the associated macroblock belongs to. Each slice group can be partitioned into one or more slices, such that a slice is a sequence of macroblocks within the same slice group that is processed in the order of a raster scan within the set of macroblocks of a particular slice group.

In order to provide efficient methods for concealment in error-prone channel applications, each slice group is transmitted separately. If a slice group is lost, the samples in spatially neighbouring macroblocks that belong to other correctly received slice groups can be used for efficient error concealment. The allowed patterns range from rectangular patterns to regular scattered patterns, such as chess boards, or to completely random scatter patterns. An example is shown in Fig.2.21.

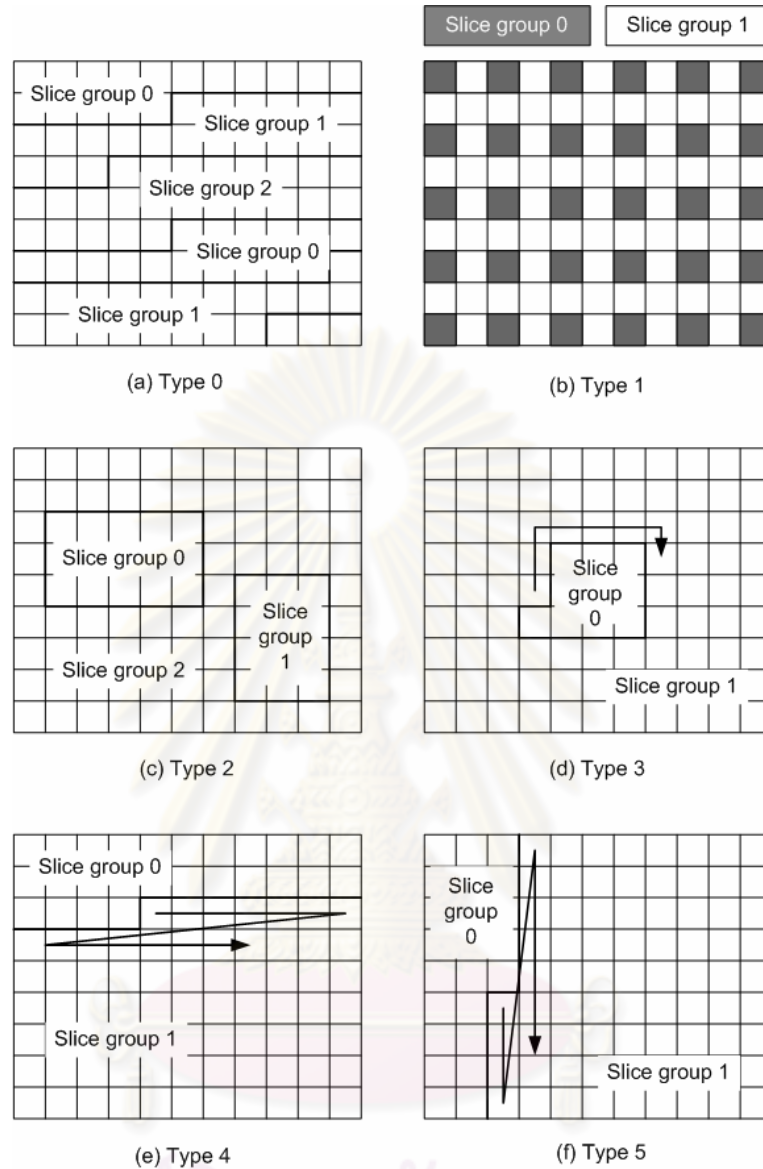


Figure 2.21 Difference type of FMO

2.2.7 Profiles and Levels of H.264

H.264 has been developed to address a large range of applications that support various bit rates, resolutions, qualities, and services. However, different applications typically have different requirements both in term of functionalities, e.g., error resilience, compression efficiency and delay, as well as complexity.

To address the large range of applications considered by H.264, three profiles have been defined as shown in Fig.2.22.

- **Baseline Profile** : This profile supports all features in H.264 except the following tools : B slices, weighted prediction, CABAC, filed (interlaced) coding and macroblock adaptive switching between frame and field coding, SP and SI slices data partitioning. This profile typically target applications with low complexity and low delay requirements.

- **Main Profile** : Supports together with the baseline profile a core set of tools as in Fig.2.22. The main profile is almost a supset of the baseline profile, however, the main profile does not support the FMO. The additional tools provided by main profile are B-slices, weighted prediction, CABAC, filed (interlaced) coding and macroblock adaptive switching between frame and field coding. This profile typically allows the best quality at the cost of higher complexity and delay.

- **Extended Profile** : This profile is a superset of the baseline profile supporting all tools in the specification with the exception of CABAC. The SP/SI slices and slice data partitioning tools are only included in this profile.

Figure 2.22 shows the relationship between the three Profiles and the coding tools supported by the standard. It is clear from this figure that the baseline profile is a subset of the extended profile, but not of the main profile. Performance limits for CODECs are defined by a set of Levels, each placing limits on parameters such as sample processing, rate, picture size, coded bit rate and memory requirements.

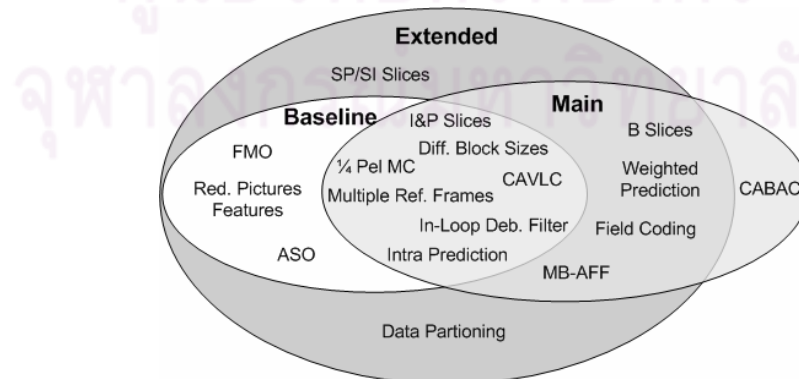


Figure 2.22 H.264/AVC profiles and corresponding tools.

2.3 Rate - distortion theory

2.3.1 Background of rate-distortion theory

Rate-distortion theory is an integral part of information theory. It addresses the problem of determining the minimal amount of entropy or information (R) that should be transmitted over a channel, so that the source (input signal) can be approximately reconstructed at the receiver (output signal) without exceeding a given distortion (D).

Rate-distortion theory gives theoretical bounds for how much compression can be achieved using lossy data compression methods. Many of the existing audio, speech, image, and video compression techniques have transforms, quantization, and bit-rate allocation procedures that design on the general shape of rate-distortion functions. Rate distortion theory was founded by Claude Shannon in his foundational work on information theory [27-28].

In rate-distortion theory, the *rate* is usually understood as the number of bits per data sample to be stored or transmitted. The notion of *distortion* is a subject of ongoing discussion. In the most simple case (which is actually used in most cases), the distortion is defined as the variance of the difference between input and output signal (i.e., the mean squared error of the difference). However, since we know that most lossy compression techniques operate on data that will be perceived by humans (listen to music, watch pictures and video) the distortion measure preferably should include some aspects of human perception.

Consider the block diagram model of a transmission system depicted in Fig.2.23. We have stated that the function of a transmission system is to convey information from the source to the user. The user usually does not require a perfect reproduction of the source output but rather will settle for a sufficiently accurate approximation. In order to determine quantitatively whether or not the performance of a transmission system is satisfactory, it is necessary to assign function relationship between coder, decoder, bit rate and distortion that the system may make.

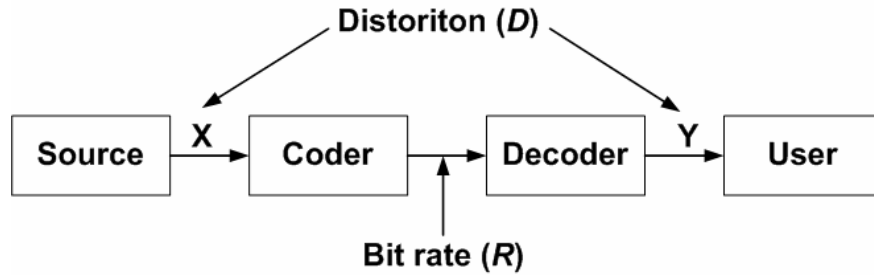


Figure 2.23 Transmission system

In this regard, Shannon [27] has shown that one can associate with most channels of interest a number of C , called the capacity of the channel, which has the significance that information can be sent through the channel with arbitrarily low probability of error at any rate less than C . Function $R(D)$, called the rate-distortion function, which has the following significance. A transmission system can be designed that achieves fidelity D if and only if the capacity of the channel that connects the source to the user exceeds $R(D)$. It should be clear that $R(D)$ is the effective rate at which the source produces information subject to the constraint that the user can tolerate an average distortion of D . The rate at which a source produces information subject to requirement of perfect reproduction is called the entropy of the source. It follows that the rate distortion function is a generalization of the concept of entropy. As D increases, $R(D)$ decreases monotonically and usually reaches to zero at some finite value of distortion. A typical rate distortion function is sketched in Fig. 2.24.

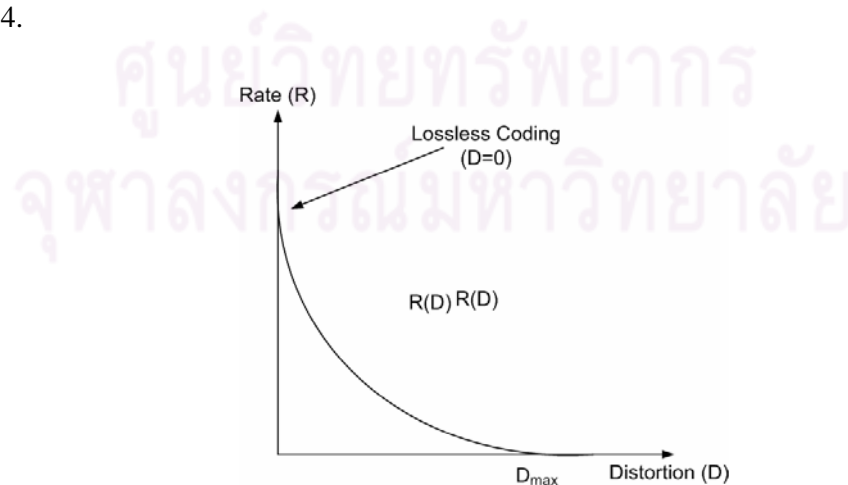


Figure 2.24. Rate - Distortion curve

2.3.2 Rate-distortion functions

The functions that relate the rate and distortion are found as the solution of the following minimization problem.

$$\min_{Q_{Y|X}(y|x)} I_Q(Y; X) \quad \text{subject to} \quad D_Q \leq D^* \quad (2.5)$$

,where $Q_{Y|X}(y|x)$ is the conditional probability density function (PDF) of the transmission channel output (compressed signal : Y) for a given input (original signal : X) and $I_Q(Y; X)$ is the mutual information between Y and X defines as shown in eq. (2.6).

$$I_Q(Y; X) = H(Y) - H(Y|X) \quad (2.6)$$

, where $H(Y)$ and $H(Y|X)$ are entropy if the output signal Y and the conditional entropy of the output signal given the input signal as in eqs.(2.7)-(2.8).

$$H(Y) = \int_{-\infty}^{\infty} P_Y(y) \log_2(P_Y(y)) dy \quad (2.7)$$

$$H(Y|X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{Y|X}(y|x) P_X(x) \log_2(Q_{Y|X}(y|x)) dx dy \quad (2.8)$$

The problem can also be formulated as distortion-rate function, where we find the supremum over achievable distortions for given rate constraint. The relevant expression is shown in eq.(2.9).

$$\min_{Q_{Y|X}(y|x)} E[D_Q(Y; X)] \quad \text{subject to} \quad I_Q(Y; X) \leq R \quad (2.9)$$

The two formulations lead to functions which are the inverse function of each other.

The mutual information can be understood as a measure for prior uncertainty the receiver has about the sender's signal ($H(Y)$), diminished by the uncertainty that is left after receiving information about the sender's signal ($H(Y|X)$). The decrease in uncertainty is due to the communicated amount of information, which is $I_Q(Y; X)$.

In the definition of the rate-distortion function, D_Q and D^* are the distortion between X and Y for a given $Q_{Y|X}(y|x)$ and the prescribed maximum distortion, respectively. When we use the mean squared error as distortion measure, we have the formula as in eq.(2.10).

$$\begin{aligned} D_Q &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{X,Y}(x,y)(x-y)^2 dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{Y|X}(y|x)P_X(x)(x-y)^2 dx dy \end{aligned} \quad (2.10)$$

The above equation show, that calculating a rate distortion function requires the description of the input X in term of the PDF ($P_X(X)$), and then aims at finding the conditional PDF ($Q_{Y|X}(y|x)$) that minimize rate for a given distortion (D^*).

For example, if we assume that $P_X(X)$ is Gaussian with variance σ^2 , and if we assume that successive samples of the signal X are independent, the analytical expression for the rate-distortion function can be stated as shown in eq.(2.11).

$$R(D) = \begin{cases} \frac{1}{2} \log_2(\sigma_x^2 / D), & \text{if } D \leq \sigma_x^2 \\ 0, & \text{if } D > \sigma_x^2 \end{cases} \quad (2.11)$$

From the background given above, it is clear that we can find rate model and distortion model based on input statistical distribution. The functions that relate to the rate and distortion are found as the solution of the minimization problem. From there, the rate-distortion model is proposed to find optimal or sub-optimal choice of quantization parameter. This dissertation has a main objective to propose a new rate–distortion model based on input statistical distribution of test video sequences. We will discuss in the next chapter.

2.4 H.264 Rate control

In H.264 rate control, a quantization parameter is determined by using linear and quadratic rate – distortion models. The rate control in H.264 [15] is composed Group of picture (GOP) layer, frame layer and basic unit layer. The basic unit in H.264 video coding is defined as a group of contiguous macroblock in a frame ,i.e., basic unit can be a macroblock, a slice or a frame.

- **GOP layer rate control** : GOP rate control calculates the total bits for the rest pictures in this GOP and the initial quantization parameter of I-frame and of the first P-frame. When encoded the current frame, rate control will compute the occupancy of encoder buffer by using fluid traffic model as shown in eq. (2.12). The initial buffer fullness is set to zero. The N_{gop} denotes the total number of GOP, $n_{i,j}$ denotes the j th frame in the i th GOP, $B_c(n_{i,j})$ denotes the occupancy of encoder buffer, $A(n_{i,j})$ denotes number of bits generated by the j th frame in the i th GOP, F_r denotes the target frame rate and $u(n_{i,j})$ denotes the available channel bandwidth.

$$B_c(n_{i,j}) = B_c(n_{i,j-1}) + (A(n_{i,j-1}) - (\frac{u(n_{i,j-1})}{F_r})) \quad (2.12)$$

For the first GOP, $QP_1(1)$ is predefined based on the available channel bandwidth as in [15].

- **Frame layer rate control** : This layer rate control composes of 2 stages.

(1) Pre-encoding stage : The objective of this stage is to determination of a target bit and compute a quantization parameter for each P frame consists of two steps.

Step 1: Budget allocation among pictures. The bit allocation is implemented by predefining a target buffer level, $Tbl(n_{i,j+1})$, for each P picture, as shown in eq. (2.13), where N_p is the number of P frames in GOP.

$$T_{bl}(n_{i,j+1}) = T_{bl}(n_{i,j}) - \frac{B_c(n_{i,2}) - B_s / 8}{N_p - 1} \quad (2.13)$$

The target bit rate, $f(n_{i,j})$, for the j th P frame in the i th GOP is scaled based on the target buffer level, current buffer level, frame rate, and channel bandwidth. It is given in eq. (2.14),

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + \gamma(T_{bl}(n_{i,j}) - B_c(n_{i,j})) \quad (2.14)$$

,where γ is a constant weighting factor. Further adjustment by a weighted combination of the average number of remaining bits for each frame is given, as shown in eq. (2.15),

$$f(n_{i,j}) = \beta * \frac{T_r(n_{i,j})}{N_p - j} + (1 - \beta) * \tilde{f}(n_{i,j}) \quad (2.15)$$

,where $f(n_{i,j})$ is the total number of remaining bits left to encode the j th frame onwards in the i th GOP, and β is a constant weighting factor.

Step 2 : Compute the quantization parameter. The MAD of the current frame, MAD_f , is predicted by a linear model using the actual MAD of the previous frame, MAD_{f-1} , as in eq.(2.16).

$$MAD_f = a_1 * MAD_{f-1} + a_2 \quad (2.16)$$

, where a_1 and a_2 are two coefficients. The initial value of a_1 and a_2 are set to 1 and 0, respectively. They are updated by a linear regression method similar to that MPEG-4 after coding each picture or each basic unit. The quantization step size corresponding to the target bits is then computed by using the following quadratic model as in eq.(2.17).

$$f(n_{i,j}) = b_1 \times \frac{MAD_f}{Q_i} + b_2 \times \frac{MAD_f}{Q_i^2} - m_h(n_{i,j}) \quad (2.17)$$

, where $m_h(n_{i,j})$ is the total number of header bits and motion vector bits, b_1 and b_2 are two coefficients.

To maintain the smoothness of visual quality among successive frames, the quantization parameter $QP(n_{i,j})$ is adjusted by following eq.(2.18).

$$QP(n_{i,j}) = \min\{QP(n_{i,j-1}) + 2, \max\{QP(n_{i,j-1}) - 2, QP(n_{i,j})\}\} \quad (2.18)$$

- **Basic unit layer rate control** : Suppose that a frame is composed of N_{mbpic} MBs. A basic unit is defined to be a group of continuous MBs, and consists of N_{mbunit} MBs, where N_{mbunit} is a fraction of N_{mbpic} . If N_{mbunit} equals to N_{mbpic} , it would be a frame layer rate control. The total number of basic units in a frame, N_{unit} , is computed by eq.(2.19).

$$N_{unit} = \frac{N_{mbpic}}{N_{mbunit}} \quad (2.19)$$

If the basic unit is not selected as a frame, an additional basic unit layer rate control for the stored picture should be added. Same as the frame layer rate control, the quantization parameter for I-frame and first P-frame are the same for all basic unit in the same frame. It is computed the similar way as that a frame layer. The basic unit layer rate control selects the values of quantization parameters of all basic unit in a frame, so that the sum of generated bits is close to the frame target, $f(n_{i,j})$. The following is a step by step description of this method.

Step 1 : Predict the MAD of current basic unit $MAD_{cb}(l)$ in the current frame by eq.(2.16) using the actual MAD of the co-located basic unit in previous frame.

Step 2 : Compute the number of texture bits $f_l(n_{i,j})$ for the l^{th} basic unit.

Step 2.1 : Compute the target bits for the l^{th} basic unit as in eq.(2.20).

$$f_l(n_{i,j}) = f_{rb}(n_{i,j}) \times \frac{MAD_{cb}^2(l)}{\sum_{k=l}^{N_{unit}} MAD_{cb}^2(k)} \quad (2.20)$$

, where $f_{rb}(n_{i,j})$ denote the number of remaining bits for the current frame and its initial value is set to $f(n_{i,j})$.

Step 2.2 : Compute the average number of header bits generated by all coded basic units as in eq.(2.21).

$$\begin{aligned}\tilde{m}_{hdr,l} &= \left(\tilde{m}_{hdr,l-1} \times \left(1 - \frac{1}{l}\right) \right) + \frac{\hat{m}_{hdr,l}}{l} \\ m_{hdr} &= \left(\tilde{m}_{hdr,l} \times \frac{l}{N_{unit}} \right) + \left(m_{hdr,1} \times \left(1 - \frac{l}{N_{unit}}\right) \right)\end{aligned}\quad (2.21)$$

, where $\hat{m}_{hdr,l}$ is the actual number of header bits generated by the l^{th} basic unit in the current frame. $m_{hdr,1}$ is the estimation from all basic units in the previous frame.

Step 2.3 : Compute the number of texture bits , $R_{t,l}$, for the l^{th} basic unit as in eq.(2.22).

$$R_{t,l} = f_l(n_{i,j}) - m_{hdr} \quad (2.22)$$

Step 3 : Compute the quantization step size for the l^{th} basic unit by using the quadratic rate-distortion model as in eq.(2.17). To quantization parameter is bounded by eq.(2.23) in order to smoothness of visual quality.

$$Q_{cb} = \max\{1, Q_{apf} - 6, \min\{51, Q_{apf} + 6, \tilde{Q}_{cb}\}\} \quad (2.23)$$

Note that the detail information of H.264 rate control can be found in [15]. Rate control process of H.264 can be shown in Fig.2.25.

2.5 Objective quality measurement

Because of the problems of subjective measurement, developers of digital video systems rely on objective measures of visual quality. The most widely used objective measure is peak signal to noise ratio (PSNR), calculated using eq.(2.24). PSNR is measured on a logarithmic scale and based on the mean squared error (MSE) between an original and an impaired image or video frame, relative to $(2^n - 1)^2$, the square of the highest possible signal value in the image.

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (2.24)$$

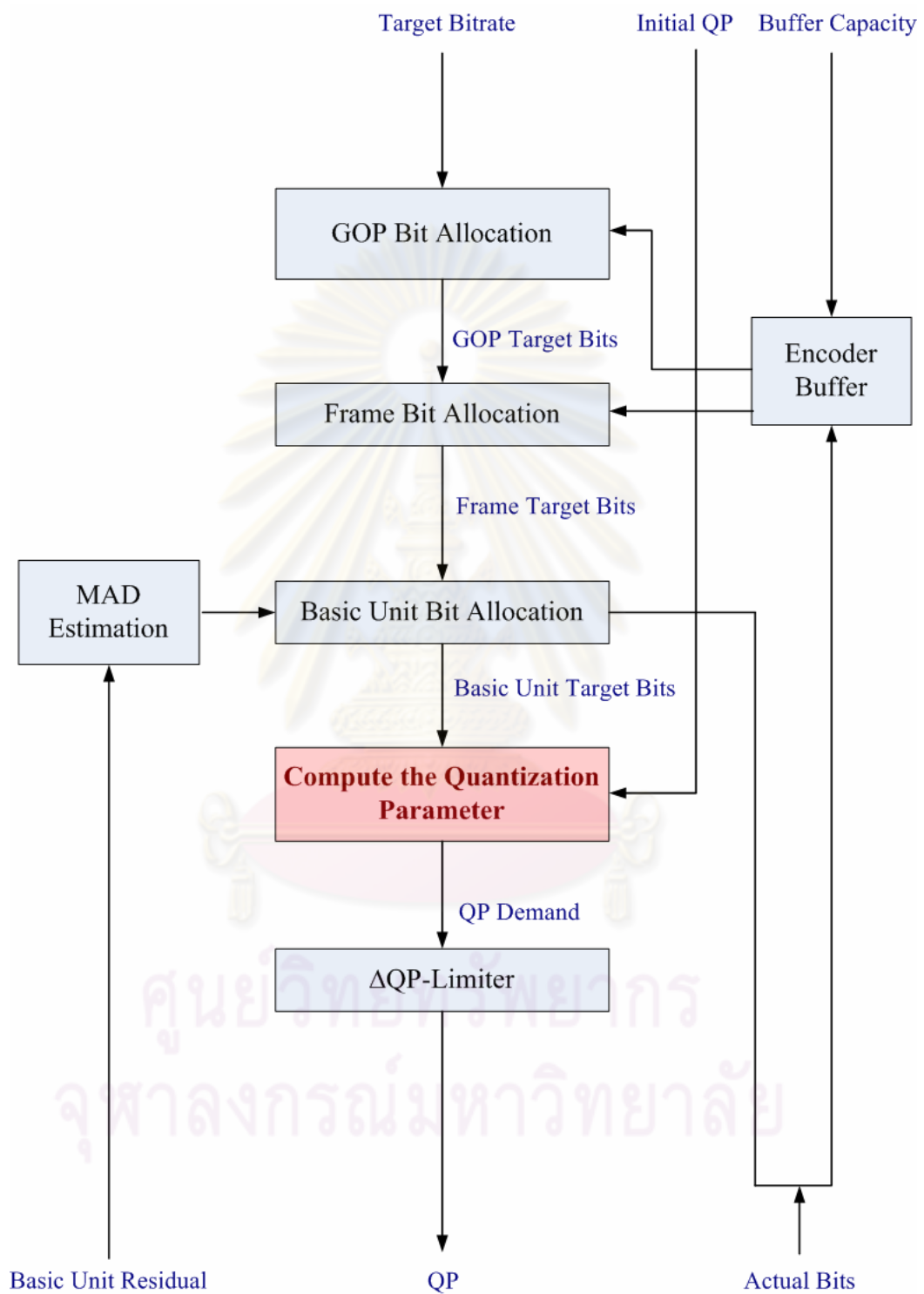


Figure 2.25 Block diagram of rate control process for H.264

Chapter 3

An improved rate-distortion optimization model based on Cauchy density function for H.264 video coding

In this chapter, background on analyzing DCT coefficients and rate-distortion model based Cauchy distribution are introduced. As stated in the background and significance of the research problem in chapter 1, efficient rate control strategy is employed to obtain the best picture quality subject to bit rate constraint. To develop efficient rate-control algorithm, it requires knowledge of information theory, in particular, Rate-Distortion theory [16]. This theory addresses the problem of determining the minimal amount of entropy (or information) R that should be transmitted over a channel. It originates from Shannon's paper [27-28] and forms a basic foundation part of information theory [16] and lossy source coding [30-32].

From the background given in section 2.3, it is clear that the knowledge of the statistical behavior in term of distribution of the transform coefficients is important to the design the rate-distortion model. This implies that, we can find rate model and distortion model based on input statistical distribution. From there, the rate model is proposed to find optimal or sub-optimal choice of quantization parameter. Later on, without losing generality, we will use DCT as an example of block-based transform coding, as it is widely used in standard video coding. For DCT, the transform coefficients are categorized into DC and AC coefficients. Since DC coefficient represents an average of energy in the signal, its values are uniformly distributed. Several studies on the statistical probability distribution then focus on the distribution of AC-coefficients.

In earlier studies, the AC coefficients were conjectured to have Gaussian distributions [38-39]. Later, several other distribution models were reported, including generalized Gaussian and Laplacian distributions [17-22]. Other studies modeled the statistical distributions of DCT coefficients using more complex probability density functions such as Gaussian Mixture Models (GMM). In [19], Muller used a generalized Gaussian function that includes Gaussian and Laplacian probability density functions as special cases. Eude et al. reported that the statistics of the DCT

coefficients can be modeled as a linear combination of a number of Laplacian and Gaussian probability density functions [20]. Comparing their models with Laplacian, Gaussian and Cauchy probability density functions, they claimed that the distribution of DCT coefficients follow neither a Cauchy nor a Laplacian distribution only but can be accurately modeled as a mixture of Gaussian distributions. Although a generalized Gaussian density function can model the statistics of the DCT coefficients more accurately, it is not widely used in practice because it is mathematically difficult to analyze. Cubic spline models [24] in MPEG-2 have shown to be more accurate in estimating the rate characteristics of video sequence, but it is computationally complex. Nevertheless, previous works on the rate control of video coding standards are based on the assumption that AC coefficients follow a Laplacian distribution. Recently the work in [25], N.Kamaci and Y.Altunbasak observed that in most cases Cauchy distributions provide more accurate estimates of the statistical distribution of DCT coefficients in typical video sequences compared to that of Laplacian distributions. Then in the next section we will study the video input source statistics in term of the actual distributions of DCT coefficients of Cauchy distribution compared with the Laplacian distribution.

3.1 Analysis of DCT-coefficients by curve fitting

In this section, we experimented the two curve fitting types to find a curve which matches a series of the AC coefficient data points in different video input source. Based on the observation of [25] that Cauchy distribution provides more accurate estimates of the statistical distribution of DCT coefficient than the Laplacian distributions. Therefore, in our research, Cauchy curve fitting and Laplacian curve fitting are used in DCT-coefficient analysis. In DCT-coefficient analysis, we are plots the histogram of 8×8 DCT-coefficients block. Fig.3.1 shows a typical plot of the histograms of the DCT Coefficients. The upper left most of coefficient is called the DC coefficient while the rest are AC coefficients. The scaling of the histogram is kept the same for all AC coefficients in this plot. Fig.3.1 shows a typical plot of the histograms of the DCT coefficients. Curve fitting is used to find a curve which matches a series of data points or probability density functions of DCT-coefficients.

In this problem, we consider the two curve fitting types, i.e., Laplacian distribution and Cauchy distribution to find a curve which matches a series of the AC coefficient data points. The distributions of Laplacian and Cauchy distributions are described as follows.

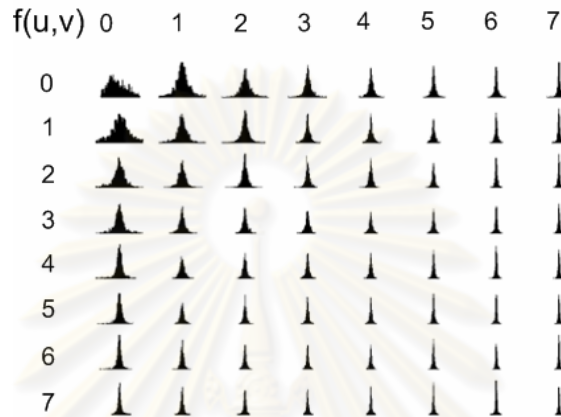


Figure 3.1 Probability density functions of each 8x8 DCT-coefficients [18]

- Laplacian probability density function

The general formula for the Laplacian probability density function, $p(x)$, with parameter λ and b can be defined as shown in eq.(3.1),

$$p(x) = \frac{1}{2b} \exp\left\{-\frac{|x-\lambda|}{b}\right\}, \quad x \in R \quad (3.1)$$

,where λ is location parameter, specifying the location of the peak of the distribution and b is scale parameter. If b is large, then the distribution will be more spread out, if b is small then it will be more concentrated. Variance of Laplacian probability density function can be written in term of $2b^2$.

The Laplacian probability density function has an exponential form, leading to the property that the tail of the density decays very fast. Fig.3.2 shows the plot of the Laplacian probability density function in different values of b and λ .

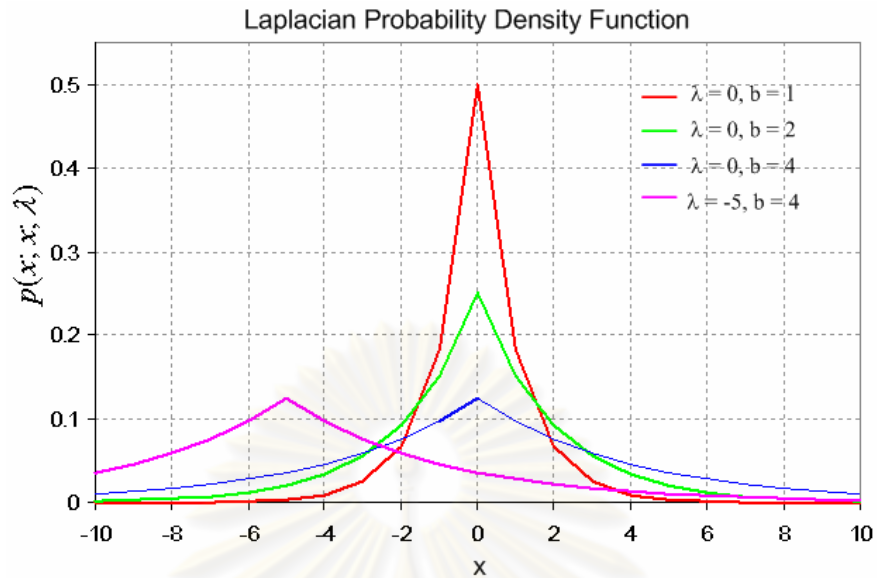


Figure 3.2 Laplacian probability Density Function at different b and λ

- Cauchy probability density function

The general formula for the Cauchy probability density function can be written as shown in eq.(3.2),

$$p(x; x_0, \mu) = \frac{1}{\pi} \left[\frac{\mu}{(x - x_0)^2 + \mu^2} \right] \quad (3.2)$$

, where x_0 is the location parameter, specifying the location of the peak of the distribution, and μ is the scale parameter which specifies the width of probability density function. A special case where $x_0 = 0$ and $\mu = 1$ is called a standard Cauchy distribution. The equation for a standard Cauchy distribution can be simplified as shown in eq.(3.3).

$$p(x; 0, 1) = \frac{1}{\pi} \left[\frac{1}{x^2 + 1} \right] \quad (3.3)$$

Fig.3.3 shows the plot of the Cauchy probability density function in different values of x_0 and μ .

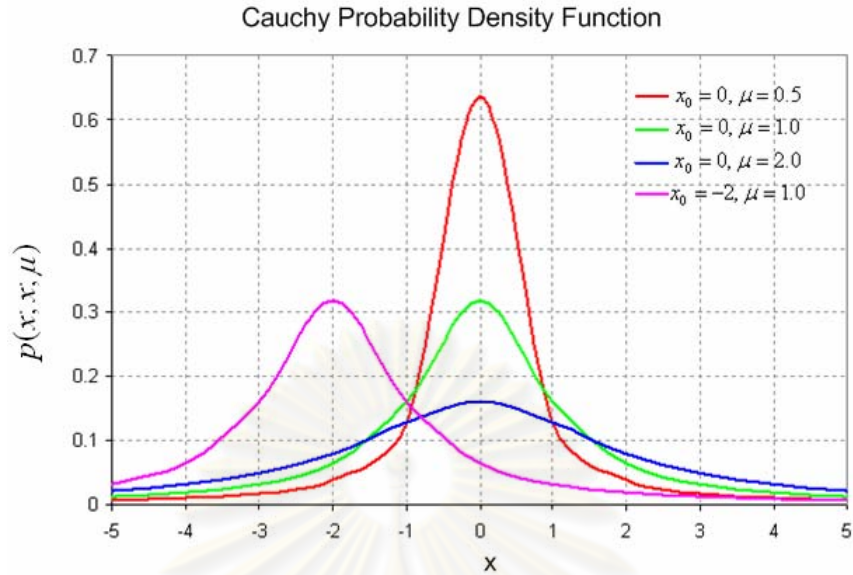


Figure 3.3 Cauchy probability density function at different values of x_0 and μ

Based on 8×8 DCT-coefficients block, we consider the probability density functions of 8×8 DCT-coefficients block only AC coefficients. Fig.3.4 shows the plot of the histogram of 8×8 DCT coefficients block size. The image used here is the Akiyo test video sequence. The red line denotes Cauchy probability density curve fitting while the blue line is the Laplacian probability density curve fitting.

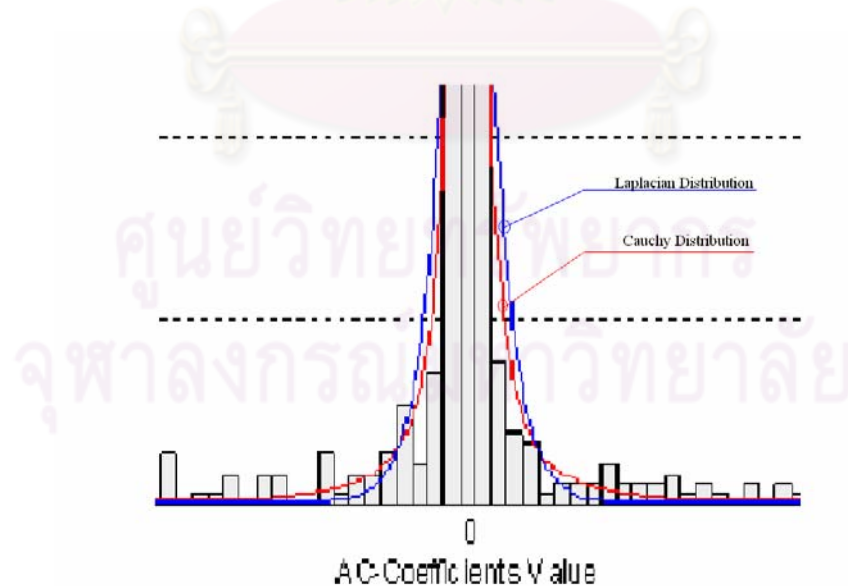


Figure 3.4 Histogram of DCT coefficients of Akiyo sequence

It can be shown that, the Laplacian distribution has an exponential form, leading to the property that the tail of the density decays very quickly. Furthermore, we have verified that Cauchy distribution is better fit to the signals than Laplacian distribution. To find the error plot between video input data points with Cauchy curve fitting and Laplacian curve fitting, the error curve fitting is used in our experiment.

Table 3.1 shows the error curve fitting between AC coefficients data points of different fourteen video input sources. In this table, the percentages in error reduction is the percentage reduction of Cauchy curve fitting compared to the Laplacian curve fitting. Error curve fitting can be calculated by eq.(3.4)

$$\text{Error curve fitting} = \sum_{i=1}^N \frac{|x_i - \hat{x}_i|}{x_i} \quad (3.4)$$

, where x_i is the histogram of AC coefficients distribution, \hat{x}_i is the PDF of curve fitting and N is the number of the histogram bin.

It can be shown that Cauchy distribution is nearly fitted with the histogram of 8×8 DCT coefficients block size all types of the video input source rather than the Laplacian distribution. Correspondingly, we then investigate rate model and distortion model based on Cauchy density function to find the solution of R-D optimization model that will be applied to H.264 video coding as later described in the next sections.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Table 3.1: Comparison of the error curve fitting between Cauchy and Laplacian curve fitting, with the histogram of DCT coefficients.

Sequence	Description	% error curve fitting		% Error Reduction
		Cauchy	Laplacian	
Akiyo	Still camera on human subject with synthetic background	5.3 %	15.6 %	10.3 %
Missam	Still camera on slow human head moving	12.3 %	21.0 %	8.7 %
Claire	Still camera on fast human head and shoulder moving	12.1 %	20.4 %	8.3 %
News	Still camera on slow moving scene change	4.2 %	14.7 %	10.5 %
Suzie	Still camera on medium human moving	5.8 %	13.5 %	7.7 %
Silent	Still camera on slow hand moving	3.6 %	9.8 %	6.2 %
Trevor	Fast human moving with background change	2.5 %	7.2 %	4.7 %
Salesman	Still camera on medium human head and hand moving	4.9 %	13.6 %	8.7 %
Container	Still camera with slow object moving	3.5 %	10.5 %	7.0 %
Coastguard	Medium camera with medium object motion	4.8 %	12.2 %	7.4 %
Foreman	Fast camera and content motion with pan at the end	2.3 %	11.3 %	9.0 %
Carphone	Fast camera and content motion with landscape passing	3.6 %	8.5 %	4.9 %
Mobile	Fast object moving with background change	6.2 %	12.8 %	6.6 %
Walk	Fast object moving with landscape passing	3.9 %	13.4 %	9.5 %

3.2 Background of rate and distortion model based on Cauchy density function

In this section, background of rate-distortion model based on Cauchy distribution is introduced. As previously mentioned, knowledge of AC coefficients probability density function is particularly important in the design of rate control for video coding. To solve the problem of bit allocation and quantization parameter selection requires the knowledge of the rate-distortion model as a function of the encoder parameters and the video input source statistics. As stated above, we found that most of the AC coefficients are distributed more closer to Cauchy than Laplacian

distribution. Therefore, in our study, we focus on the study of adopting Cauchy probability density function to find the rate-distortion model for the applications of rate control.

3.2.1 Cauchy based rate model

As we discussed in chapter 2, after transformation of the pixels, quantization is performed independently on each transformed coefficients then the result is entropy coded. In order to find rate model, we consider use uniform quantization of transform coefficients with optimal bit allocation. We have also assumed perfect entropy coding is possible for transform coefficients. As in section 3.1, transform coefficients of video input source are assumed to be modeled more accurately as Cauchy distribution with scale parameter of μ . Thus, if using uniform quantizer, one should be able to find rate and distortion of any quantization step size and input signal distribution of DCT coefficients input signal as in [55]. For example, the simplest quantizer divides the transform coefficients by quantization step size (Q) and rounds to the nearest integer (i). Reconstructed points (\hat{x}_i) are at the midpoints of each quantization range, that is, $\hat{x}_i = iQ$ for all integers i .

The Cauchy probability density function $p(x;0;\mu)$ with location parameter $x_0 = 0$ is given by eq.(3.5).

$$p(x;0;\mu) = \frac{1}{\pi} \left[\frac{\mu}{x^2 + \mu^2} \right] \quad (3.5)$$

, and the probability distribution that a sample will be quantized to $\hat{x}_i = iQ$ is simply the probability P_i that the sample is between $Q(i-1/2)$ and $Q(i+1/2)$. This is given by eq.(3.6).

$$P_i = \int_{Q\left(i-\frac{1}{2}\right)}^{Q\left(i+\frac{1}{2}\right)} p(x) dx \quad (3.6)$$

If the perfect entropy coding is used, the resulting entropy rate $H(Q)$ to encode the quantized coefficient is given by eq.(3.7).

$$H(Q) = - \sum_{i=-\infty}^{\infty} P_i \log_2 P_i \quad (3.7)$$

A closed form expression for the rate as a function of the quantization step size (Q) for the Cauchy distribution is given by eq.(3.8).

$$H(Q) = \frac{-2}{\pi} \tan^{-1} \left(\frac{Q}{2\mu} \right) \log_2 \left(\frac{2}{\pi} \tan^{-1} \left(\frac{Q}{2\mu} \right) \right) - 2 \sum \left\{ \frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4)Q^2} \right) \times \log_2 \left(\frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4)Q^2} \right) \right) \right\} \quad (3.8)$$

Eq.(3.8) give a parametric description of the lowest rate possible using an entropy coded uniform quantizer. Note that, eq. (3.8) can be used to approximate the entropy function of quantization step size in term of linear equation function, as shown in eq.(3.9) [25].

$$R = H(Q) = aQ^{-\alpha} \quad (3.9)$$

, where a and α are Cauchy rate model parameters. Their values need to be estimated and depend on the value of μ .

The examples of the entropy plot of each video sequence compared with the approximate entropy function in eq.(3.9) are shown in Fig.3.5–3.8. As shown in the figure, we plot the entropy of Akiyo sequence, Foreman, Carphone and Tempete sequences (P-frame) with the approximate entropy function at different quantization step sizes. It can be shown that the approximation of entropy function is very close to an accurate estimate.

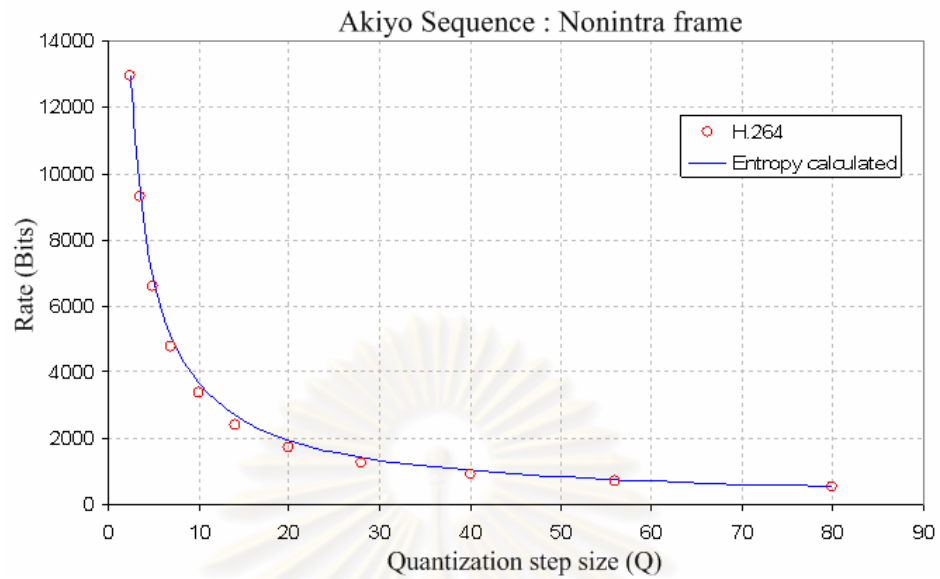


Figure.3.5 Entropy of Akiyo sequence versus approximated entropy functions at different quantization step sizes (Q)

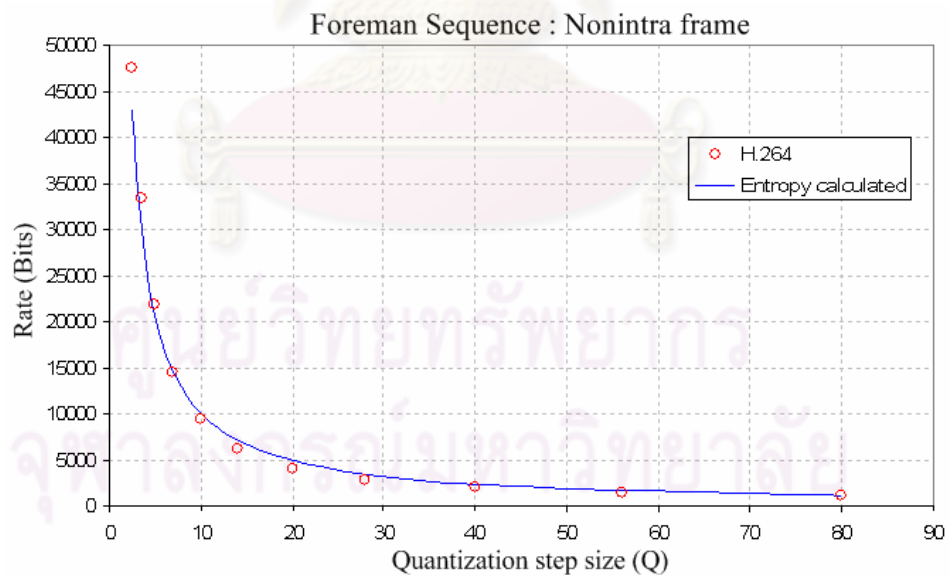


Figure.3.6 Entropy of Foreman sequence versus approximated entropy functions at different quantization step sizes (Q)

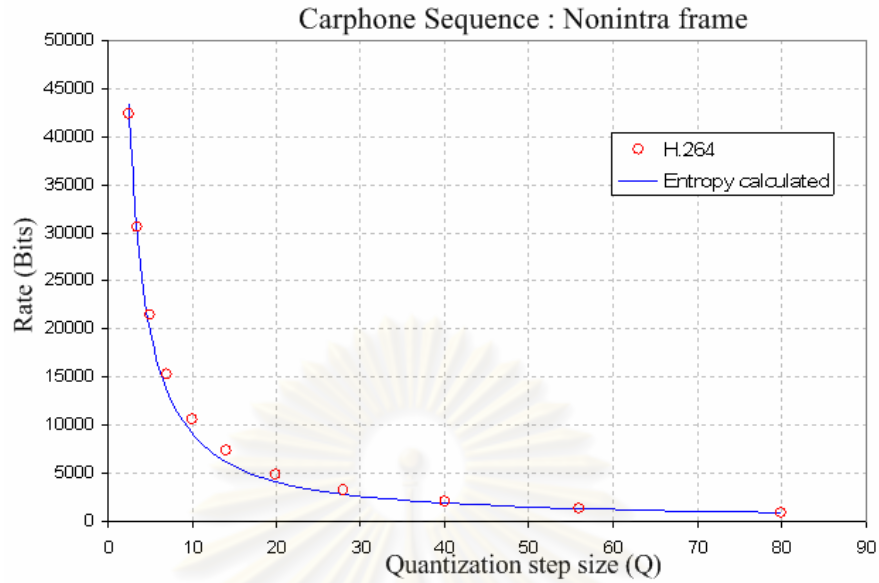


Figure.3.7 Entropy of Carphone sequence versus approximated entropy functions at different quantization step sizes (Q)

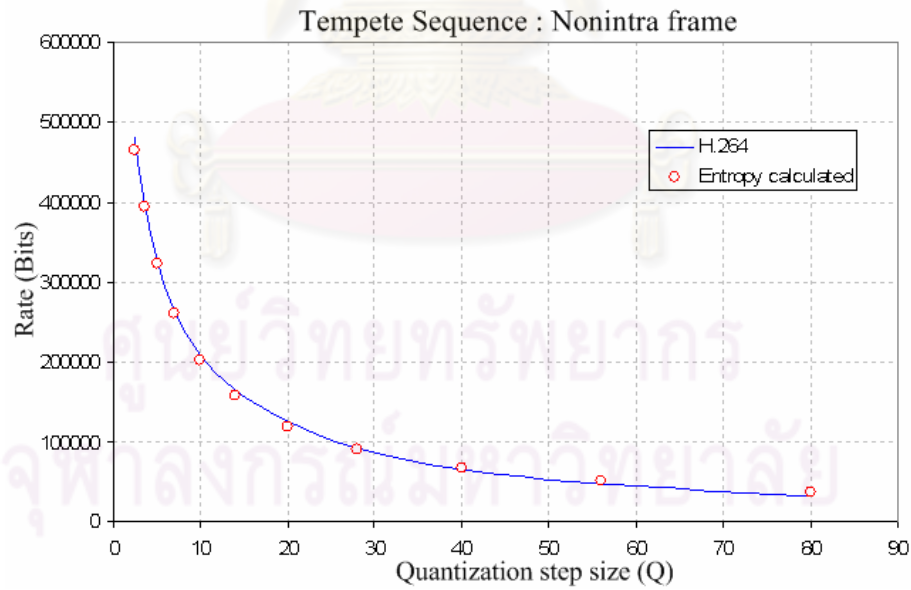


Figure.3.8 Entropy of Tempete sequence versus approximated entropy functions at different quantization step sizes (Q)

3.2.2 Cauchy based distortion model

If considering each quantization range separately, it is therefore possible to compute the expected distortion as a function of Q . For the midpoint reconstruction, $\hat{x}_i = iQ$, the expected distortion $D(Q)$ for given any distribution is shown in eq.(3.10),

$$D(Q) = \sum_{i=-\infty}^{\infty} \int_{Q\left(i-\frac{1}{2}\right)}^{Q\left(i+\frac{1}{2}\right)} (x - Qi)^2 p(x) dx \quad (3.10)$$

The closed form of expected distortion in eq.(3.10) can be shown as in eq.(3.11).

$$D(Q) = \frac{1}{\pi} \sum_{i=1}^{\infty} \left\{ -\mu Q + i\mu Q \ln \left(\frac{\mu^2 + \left(i + \frac{1}{2}\right)^2 Q^2}{\mu^2 + \left(i - \frac{1}{2}\right)^2 Q^2} \right) \right\} + \left(\mu^2 - i^2 Q^2 \left(i^2 - \frac{1}{4} \right) \right) \quad (3.11)$$

Eq.(3.11) shows a parametric description of the distortion using an entropy coded uniform quantizer. Note that, eq.(3.11) can be used to approximate the distortion function of quantization step size in term of linear equation function, as shown in eq.(3.12) [25],

$$D(Q) = bQ^\beta \quad (3.12)$$

, where b and β are Cauchy distortion model parameters. Their values need to be estimated and depend on the value of μ .

As shown in Fig.3.9, we plot the distortion of Akiyo sequence (P-frame) and the approximated distortion in eq.(3.12) at different quantization step sizes. It can be shown that the approximation of distortion function is also close to an accurate estimate.

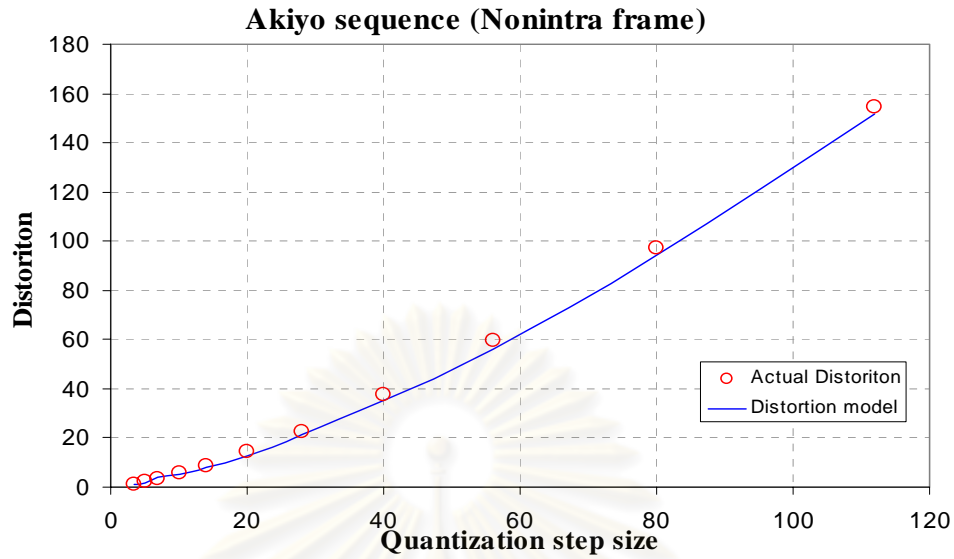


Figure.3.9 Distortion of Akiyo sequence versus approximated distortion functions at different Quantization step sizes (Q)

3.3 Proposed generalized Cauchy R-D optimization model using Lagrange multiplier technique

Based on the finding from the previous section, it is confirmed that the actual distributions of the DCT coefficients in video applications is better fit Cauchy distribution than Laplacian distribution. In our research, we propose a new rate-distortion optimization model base on Cauchy probability density function and optimize to find the optimum choice of quantization step size in video coding by using Lagrange multiplier technique.

In this section, we derive an expression for the quantization step sizes that minimize distortion subject to bit rate constraint. From eq.(3.9) , the expected number of bits for each basic unit l^{th} in a frame is defined, as shown in eq.(3.13),

$$R_l = Ca_l Q_l^{-\alpha_l} + H_l \quad (3.13)$$

, where C is the number of pixel in each basic unit, H_l is the actual number of header bits generated by each basic unit in the current frame, R_l is the number of target bits, and Q_l is the quantization step size for each basic unit.

The distortion measure, D , for encoding every basic unit in each frame is defined by eq.(3.14),

$$D = \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} \quad (3.14)$$

, where N_{unit} is number of basic unit in each frame.

An expression for the quantization step sizes $Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*$ that minimizes the distortion subject to the constraint that the total number of bits, i.e., the sum of the total number of bits of every basic unit in each frame must be equal to $R_{MAX}(j)$, is shown in eqs.(3.15)-(3.16),

$$Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^* = \arg \min_{Q_1, \dots, Q_N} \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l}$$

subject to $\sum_{l=1}^{N_{unit}} R_l = R_{MAX}(j)$ (3.15)

$$Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^* = \arg \min_{Q_1, \dots, Q_N} \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} + \lambda \left[\sum_{l=1}^{N_{unit}} R_l - R_{MAX}(j) \right] \quad (3.16)$$

, where $R_{MAX}(j)$ is defined as the target bit budget in each frame.

By using Lagrange multiplier technique, the expression for the optimum choice of quantization step size, can be formulated and the expression of the cost function $J(Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^*)$ can be shown in eq.(3.17).

$$J(Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^*) = \frac{1}{N} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} + \lambda \left[\sum_{l=1}^{N_{unit}} (C a_l Q_l^{-\alpha_l} + H_l) - R_{MAX}(j) \right] \quad (3.17)$$

Assume that the number of header bits in each basic unit are equal, i.e., $H_l = H$, eq.(3.17) can be rewritten as shown in eq.(3.18).

$$J(Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^*) = \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} + \lambda \sum_{l=1}^{N_{unit}} (C a_l Q_l^{-\alpha_l}) + \lambda (N_{unit} H - R_{MAX}(j)) \quad (3.18)$$

The minimum values of eq.(3.18) can be solved by using partial derivatives, i.e., $\frac{\partial J}{\partial Q_l^*} = 0$ and $\frac{\partial J}{\partial \lambda^*} = 0$ as in eq.(3.19).

Case 1: Find $\frac{\partial J}{\partial Q_l^*}$,

$$\frac{\partial J}{\partial Q_1^*} = \frac{\beta_1}{N} b_1 (Q_1^*)^{(\beta_1-1)} - (\lambda C a_1 \alpha_1) (Q_1^*)^{-(\alpha_1+1)}$$

$$\frac{\partial J}{\partial Q_2^*} = \frac{\beta_2}{N} b_2 (Q_2^*)^{(\beta_2-1)} - (\lambda C a_2 \alpha_2) (Q_2^*)^{-(\alpha_2+1)}$$

$$\frac{\partial J}{\partial Q_N^*} = \frac{\beta_N}{N} b_N (Q_N^*)^{(\beta_N-1)} - (\lambda C a_N \alpha_N) (Q_N^*)^{-(\alpha_N+1)}$$

Then,
$$\frac{\partial J}{\partial Q_l^*} = \frac{1}{N_{unit}} \left(b_l b_l Q_l^{(\beta_l-1)} \right) - \lambda \left(C a_l \alpha_l Q_l^{-(\alpha_l+1)} \right) \quad (3.19)$$

Let $\frac{\partial J}{\partial Q_l^*} = 0$ from eq. (3.19).

$$\frac{1}{N_{unit}} \left(\beta_l b_l Q_l^{(\beta_l-1)} \right) = \lambda \left(C a_l \alpha_l Q_l^{-(\alpha_l+1)} \right)$$

$$\frac{Q_l^{(\beta_l-1)}}{Q_l^{-(\alpha_l+1)}} = (\lambda C N_{unit}) \left(\frac{a_l \alpha_l}{\beta_l b_l} \right)$$

Then,
$$Q_l^{(\alpha_l+\beta_l)} = (\lambda C N_{unit}) \left(\frac{a_l \alpha_l}{\beta_l b_l} \right) \quad (3.20)$$

From eq.(3.20) , we take logarithmic function on both sides of this equation as shown in eq.(3.21),

$$\ln Q_l^{(\alpha_l+\beta_l)} = \ln \left((\lambda C N_{unit}) \left(\frac{a_l \alpha_l}{\beta_l b_l} \right) \right)$$

$$\ln Q_l = \frac{1}{(\alpha_l + \beta_l)} \ln \left((\lambda C N_{unit}) \left(\frac{a_l \alpha_l}{\beta_l b_l} \right) \right)$$

$$Q_l = \left\{ (\lambda C N_{unit}) \left(\frac{a_l \alpha_l}{\beta_l b_l} \right) \right\}^{\frac{1}{(\alpha_l + \beta_l)}} \quad (3.21)$$

We set , $\frac{1}{(\alpha_l + \beta_l)} = s_l$ and $\left(\frac{a_l \alpha_l}{\beta_l b_l} \right) = \varepsilon_l$

Then,
$$Q_l = ((\lambda C N_{unit}) (\varepsilon_l))^{s_l} \quad (3.22)$$

Case 2 : Find $\frac{\partial J}{\partial \lambda^*} = 0$ from eq. (3.18) ,

$$\frac{\partial J}{\partial \lambda} = \sum_{l=1}^{N_{unit}} \left(C a_l Q_l^{-\alpha_l} \right) + (N_{unit} H - R_{MAX}(j)) = 0$$

Then,
$$\sum_{l=1}^{N_{unit}} (Ca_l Q_l^{-\alpha_l}) = (R_{MAX}(j) - N_{unit}H) \quad (3.23)$$

Replace eq.(3.22) in eq. (3.23).

$$\sum_{l=1}^{N_{unit}} (Ca_l (\lambda CN_{unit} \varepsilon_l)^{-s_l \alpha_l}) = (R_{MAX}(j) - N_{unit}H) \quad (3.24)$$

In this case we set, G is constant value $= \frac{\alpha_l}{\alpha_l + \beta_l}$.

$$\sum_{l=1}^{N_{unit}} (Ca_l (\lambda CN_{unit} \varepsilon_l)^{-G}) = (R_{MAX}(j) - N_{unit}H)$$

$$C^{(1-G)} (\lambda N_{unit})^{-G} \sum_{l=1}^{N_{unit}} (a_l (\varepsilon_l)^{-G}) = (R_{MAX}(j) - N_{unit}H)$$

$$(\lambda)^{-G} = N_{unit}^G C^{(G-1)} \frac{(R_{MAX}(j) - N_{unit}H)}{\sum_{l=1}^{N_{unit}} (a_l (\varepsilon_l)^{-G})}$$

Then,
$$\lambda = \left(\frac{\sum_{l=1}^{N_{unit}} (a_l (\varepsilon_l)^{-G})}{N_{unit}^G C^{(G-1)} (R_{MAX}(j) - N_{unit}H)} \right)^{\frac{1}{G}} \quad (3.25)$$

Replace eq.(3.25) in eq.(3.22), we can get the optimized quantization step sizes, as shown in eq.(3.26).

$$Q_l = \left(\left(\frac{\sum_{l=1}^{N_{unit}} (a_l(\varepsilon_l))^{-G}}{N_{unit}^G C^{(G-1)} (R_{MAX}(j) - N_{unit}H)} \right)^{\frac{1}{G}} CN_{unit}(\varepsilon_l) \right)^{s_l}$$

$$\therefore Q_l = \left(\varepsilon_l \left(\frac{\sum_{l=1}^{N_{unit}} (a_l(\varepsilon_l))^{-G}}{R_{MAX}(j) - N_{unit}H} \right)^{\frac{1}{G}} \right)^{s_l} \quad (3.26)$$

, where a_l , α_l and b_l , β_l are model parameters of the basic unit l th, respectively.

3.4 Proposed a linear prediction rate and distortion model parameter by using linear regression analysis

In this section, we relate the expression from eqs.(3.9) and (3.12) in term of linear equation to find the Cauchy rate-distortion model parameters (a_{j-l} , α_{j-l} and b_{j-l} , β_{j-l}) in eq. (3.9) and (3.12) by using statistical linear regression analysis [56].

3.4.1 The linear regression model

A regression model is an application of the linear model where populations of the response variable are identified with numeric values of one or more quantitative variables that are called factor or independent variables. A regression model specifies that the mean of the independent variable y is related to the independent variable x . It means that, linear regression analyzes the relationship between two variables, x and y . For each subject or experimental unit, both x and y are known and we want to find the best straight line through the data. In some situations, the slope and/or intercept have a scientific meaning, as shown in Fig.3.10, the plot of linear regression line.

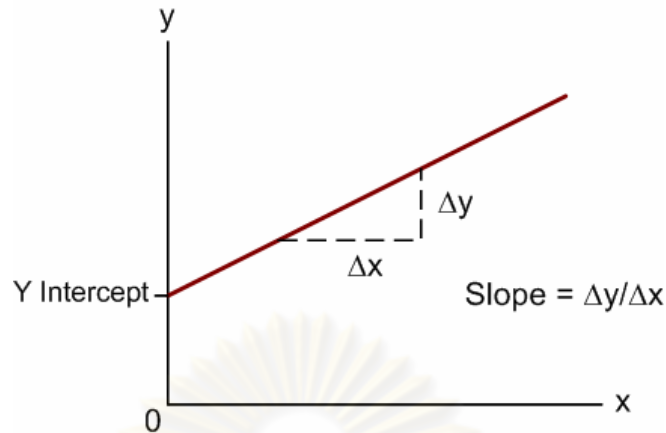


Figure 3.10 Plot of regression line

, where given two points (x_1, y_1) and (x_2, y_2) , the difference between x from one to the other is $\Delta x = x_2 - x_1$, while the difference between y is $\Delta y = y_2 - y_1$.

In other cases, we use the linear regression line as a standard curve to find the new values of x from y , or y from x . In general, the goal of linear regression is to find the line that best predicts y from x .

The simplest regression model is the simple linear regression model, which is defined as in eq.(3.27).

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (3.27)$$

, where y is the dependent variable and x is the independent variable. β_0 , the intercept point, is the value of the mean of the dependent variable when x is zero. β_1 , the slope, is the change in the mean of the dependent variable associated with a unit change in x . ε is the statistical portion or random part of the model.

This model is called a linear model that consists of a deterministic or functional relationship among the variables part and a statistical part. The statistical portion of the model is the random error component. The deterministic portion of the model, $\beta_0 + \beta_1 x$, specifies that for any value of the independent variable, x , or response variable, y , is described by the straight line function $(\beta_0 + \beta_1 x)$.

The random part of the model explains the variability of the responses about the mean. We assume that this term, i.e., the error terms, have a zero mean and a constant variance, σ^2 . A regression analysis is a set of procedures, based on a sample of n ordered pairs, $(x_i, y_i), i = 1, 2, \dots, n$, for estimating and making inferences on the parameters, β_0 and β_1 . These estimates can then be used to estimate the mean values of the dependent variable for specified values of x . It can be seen that the data from these samples is used to construct the estimates of the coefficients which are used in the following eq.(3.28) for estimating the mean of y .

$$\hat{\mu}_{y|x} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3.28)$$

, where $\hat{\mu}_{y|x}$ is the estimate of the mean of the dependent variable, y , for any specified value of x . We can find those values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squared deviation, as shown in eq.(3.29).

$$SS = \sum (y - \hat{\mu}_{y|x})^2 = \sum (y - \hat{\beta}_0 + \hat{\beta}_1 x)^2 \quad (3.29)$$

The values of the coefficients that minimize the sum of squared deviations for any particular set of sample data are given by the solutions of the following equations, which are called the normal equations, as shown in eq.(3.30).

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_1 \sum x &= \sum y \\ \hat{\beta}_0 \sum x + \hat{\beta}_1 \sum x^2 &= \sum xy \end{aligned} \quad (3.30)$$

This solutions are obtained through the solution explain in [56].

The solution for two linear equations which have two unknown variables is obtained and thus provides the estimators of these parameters, as shown in eqs.(3.31)-(3.32).

$$\hat{\beta}_1 = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (3.31)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.32)$$

The estimator of β_1 can also be written, as shown in eq.(3.33).

$$\hat{\beta}_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (3.33)$$

From eq.(3.33), it is the sum of cross products of the deviations of observed values from the means of x and y divided by the sum of squared deviations of the x values. For practical uses, the term, $\sum (x - \bar{x})^2$, denote the corrected sums of squares and can be alternately expressed by S_{xx} , as shown in eq.(3.34). The term, $\sum (x - \bar{x})(y - \bar{y})$, denotes the corrected cross products, and can be alternately expressed by S_{xy} , as shown in eq.(3.35).

$$S_{xx} = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n} \quad (3.34)$$

$$S_{xy} = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \sum x \frac{\sum y}{n} \quad (3.35)$$

The corrected sum of squares of the dependent variable, , is shown in eq.(3.36).

$$S_{yy} = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n} \quad (3.36)$$

Using the notation as in the aboved equations, we can express $\hat{\beta}_1$ in term of S_{xx} and S_{xy} , as shown in eq.(3.37). The example of plot of the data points and estimated line can be shown in Fig.3.11 and shows how the regression line fits the data.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (3.37)$$

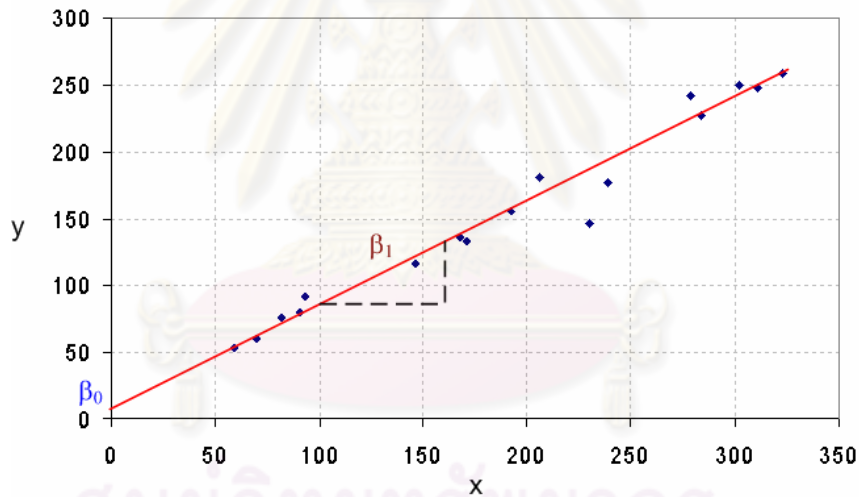


Figure 3.11 Plot of data and regression line

3.4.2 Cauchy rate model parameter

Based on knowledge of linear regression in the previous section. In this section we proposed to use linear regression analysis to find Cauchy rate distortion optimization model parameters. The encoder collects the bit rate and quantization step size for each type of picture at the end of encoding each basic unit. Then, the Cauchy

rate model parameters a_{j-l} , α_{j-l} can be found. From eq.(3.9) , we use logarithmic linear function, as shown in eq.(3.38)

$$\ln(R_l(n_{i,j})) = \ln(a_{j-l}) - \alpha_{j-l} \ln(Q_l(n_{i,j})) \quad (3.38)$$

In our research, we use linear regression analysis that models the relationship between two variables by fitting a linear equation to the observed data. We use the formula below to find the Cauchy rate model parameters, as shown in eqs.(3.39)-(3.40),

$$\ln(a_{j-l}) = \frac{\left(\sum_k^{l-1} \ln(R_k)\right) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)^2} \quad (3.39)$$

$$\alpha_{j-l} = \frac{\left(\sum_k^{l-1} \ln(Q_k)\right) \left(\sum_k^{l-1} (\ln(R_k))\right) - (l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)^2} \quad (3.40)$$

, where R_k and Q_k denote the actual number of bits used for coding and quantization step size in the previously encoded basic unit k to basic unit $l-1$, respectively.

3.4.3 Cauchy distortion model parameter

To derive distortion model parameter, we take logarithmic of linear function from eq.(3.12) and use on the term of distortion of mean square error in each basic unit as shown in eq.(3.41),

$$\ln(MSE_l(n_{i,j})) = \ln(b_{j-l}) + \beta_{j-l} \ln(Q_l(n_{i,j})) \quad (3.41)$$

,where $MSE_l(n_{i,j})$ is mean square error of the basic unit l th in each frame.

As in Cauchy rate model parameters, we use linear regression analysis [56] to update the model parameters after we encoded in each basic unit as in eqs. (3.42)-(3.43),

$$\ln(\hat{b}_{j,l}) = \frac{\left(\sum_k^{l-1} \ln(MSE_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (3.42)$$

$$\beta_{j,l} = \frac{(l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right) - \left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (3.43)$$

, where MSE_k denotes the actual mean square error for the previous coded basic unit k to basic unit $l-1$.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Chapter 4

An improved rate control based on Cauchy rate-distortion optimization model for low bit-rate H.264 video coding

Based on knowledges stated in the previous chapters, we proposed rate-distortion optimization model based on Cauchy probability density function where the optimum choice of quantization step size would be achieved. In this chapter, we applied Cauchy rate-distortion optimization model to an application of rate control. Accordingly, we proposed new rate control for H.264 video coding.

Our proposed rate control scheme with Cauchy rate-distortion optimization model composes of three layers: Group of picture layer (GOP layer) rate control, frame layer rate control and basic unit layer rate control.

- In GOP layer rate control, the computation of the total number of bits in each GOP and the remaining number of bits for all noncoded P-frame are proposed.

- In frame layer rate control, the objective of this stage is to determine the number of target bits budget for each P-frame.

- In basic unit layer rate control, we proposed the computation of the quantization step size of current basic unit based on Cauchy rate-distortion optimization model and basic unit complexity in term of residual variance. The computed quantization parameter (QP) is then adjusted to prevent the fluctuation of PSNR.

4.1 GOP layer rate control

First, the first I-frame and the first P-frame of the GOP are coded by $QP_{initial}$. $QP_{initial}$ is based on the available channel bandwidth as in H.264 rate control [15] by eq. (4.1).

$$QP_{initial} = \begin{cases} 40 & \text{if } bpp \leq l_1 \\ 30 & \text{if } l_1 < bpp \leq l_2 \\ 20 & \text{if } l_2 < bpp \leq l_3 \\ 10 & \text{if } bpp > l_3 \end{cases} \quad (4.1)$$

bpp is defined as the average number of bits per pixel in a video frame can be computed by eq. (4.2),

$$bpp = \frac{u(n_{i,j})}{F_r \times N_{pixel}} \quad (4.2)$$

, where N_{pixel} is the number of pixel in a picture. $l_1 = 0.15$, $l_2 = 0.45$ and $l_3 = 0.9$ are recommended for QCIF/CIF and $l_1 = 0.6$, $l_2 = 1.4$ and $l_3 = 2.4$ are recommended for the picture size larger than CIF.

The total number of bits allocated for the GOP is computed as, shown in eq.(4.3). The remaining bits, $T_r(n_{i,j})$, for all noncoded P-frame are updated after the $(j-1)^{th}$ frame is encoded as in eq.(4.4),

$$T_{GOP} = \frac{u(n_{i,j})}{F_r} \times N_{GOP} \quad (4.3)$$

$$T_r(n_{i,j}) = T_{GOP} - \sum_{j=0}^{j-1} b(n_{i,j}) \quad (4.4)$$

, where $u(n_{i,j})$ is the target channel bandwidth, F_r is the frame rate, N_{GOP} is the number of frames in each GOP and $b(n_{i,j})$ is the actual number of bits generated by each j^{th} frame.

4.2 Frame layer rate control

The objective of this stage is to determine the number of target bits budget before coding of the current j^{th} frame. The algorithm computes the target bits for

each P-frame, $f(n_{i,j})$, according to the number of bits used in the previous P-frame, target channel bandwidth and frame rate, as explained in the following steps.

Step 1 : The number of bits used in the previous P-frame is used to compute the target bits, $\hat{f}(n_{i,j})$, as shown in eq.(4.5).

$$\hat{f}(n_{i,j}) = \frac{T_r(n_{i,j})}{N_{GOP} - N_c} \quad (4.5)$$

, where N_c is the total number of already coded frame.

Step 2 : The target bits budget before coding of the current j th frame, $f(n_{i,j})$ can be computed by eq.(4.6),

$$f(n_{i,j}) = \zeta \times \hat{f}(n_{i,j}) + (1 - \zeta) \times \frac{u(n_{i,j})}{F_r} \quad (4.6)$$

, where ζ is constant value. Note that ζ is empirically set to be 0.6 in our simulations.

4.3 Basic unit layer rate control

In basic unit layer, the suitable quantization step sizes are obtained using Cauchy R-D optimization model as, described in Section 3. The quantization parameter is further adjusted to keep bit rate under the given constraints, and to prevent the fluctuation of the quality. Model parameters are updated after encoding each basic unit. The layer is divided into pre-encoding and post-encoding stages.

1) Pre-encoding stage

In this stage, we compute the quantization step size of current basic unit in three cases.

Case 1 : For the first basic unit in current frame, quantization parameter is obtained, as shown in eq. (4.7),

$$QP_{1,i}(j) = \text{Avg}QP_i(j-1) \quad (4.7)$$

, where $QP_{1,i}(j)$ is the quantization parameter of the first basic unit and $\text{Avg}QP_i(j-1)$ is the average of quantization parameter for all basic unit in previous frame.

Case 2 : When the number of remaining bits, $T_r(n_{i,j})$, is lower than zero, quantization parameter is obtained as shown in eq. (4.8).

$$QP_{l,i}(j) = QP_{l-1,i}(j) + 2 \quad (4.8)$$

Case 3 : The quantization step sizes of current basic unit is computed by Cauchy R-D optimization model using the formula in eq.(3.24), where a_{j-l} , α_{j-l} and b_{j-l} , β_{j-l} are model parameters of the basic unit l th of frame j , respectively. The algorithms are according to the following steps.

Step 1 : To get a better target bits estimation for each frame, we need to consider the complexity factor of each basic unit, $\gamma(j,l)$. It is defined in eq.(4.9),

$$\gamma(j,l) = \frac{\text{Res}_{\text{var}}(j,l)}{\text{Ave}_{\text{Res}_{\text{var}}}(j-1)} \quad (4.9)$$

, where $\text{Res}_{\text{var}}(j,l)$ is the residual variance of basic unit l^{th} and $\text{Ave}_{\text{Res}_{\text{var}}}(j-1)$ is the average of residual variance of all basic unit in the previous frame $(j-1)^{\text{th}}$. In order to achieve the suitable target bits estimation for each P-frame, the frame target, $f_v(j)$ is thus adjusted, as shown in eq.(4.10),

$$f_v(n_{i,j}) = \begin{cases} f(n_{i,j}) & \text{if } \gamma(j,l) \leq 1 \\ 1.1 \times f(n_{i,j}) & \text{if } \gamma(j,l) > 1 \end{cases} \quad (4.10)$$

Step 2 : Compute the quantization step size ($Q_{l,i}$) of the current basic unit by using the parameters, a_{j-l} , α_{j-l} and b_{j-l} , β_{j-l} , $R_{\text{MAX}}(j)$ of the basic unit l th of frame j .

To avoid the fluctuation of the quality, the quantization parameter of each basic unit are bounded by a lower and upper bounds, as shown in eq.(4.11),

$$\begin{aligned} QP_{l,i} &= \text{MAX}\{QP_{l,i}, QP_{Dist} - 6\} \\ QP_{l,i} &= \text{MIN}\{QP_{l,i}, QP_{Dist} + 6\} \end{aligned} \quad (4.11)$$

,where QP_{Dist} is computed based on the average distortion in previous coded frames, $D_{ave}(j)$, as shown in eq.(4.12),

$$QP_{Step_Dist} = \left(\frac{D_{ave}(j)}{b_{j-1}} \right)^{\frac{1}{\beta_{j-1}}} \quad (4.12)$$

, where QP_{Step_Dist} is the quantization step sizes of QP_{Dist} .

2) Post-encoding stage

After encoding each basic unit, only Cauchy rate and distortion model parameters, a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} , are updated by linear regression analysis in eqs. (3.36)-(3.37) and eqs.(3.39)-(3.40), as shown in Chapter 3, to find the optimal quantization step size of each basic unit.

4.4 Simulation results

For our simulation, we compare the performance between our proposed a rate control scheme using Cauchy rate-distortion optimization model with H.264 JM 8.6 rate control in terms of the average PSNR, PSNR standard deviation, bit rate used, and processing time. The parameters of video coding and tested video sequences are set as follows.

- 1) We encoded seven video sequences covering every aspects of characteristics, as shown in Fig.4.1 :
 - Carphone sequence which has fast camera and content motion with landscape passing.
 - Foreman sequence which has fast camera and content motion with pan at the end.

- Silent sequence which has still camera on slow hand moving, News sequence with still camera on slow moving scene change.
- Akiyo sequence which has still camera on human subject with synthetic background.
- Claire sequence which has still camera on fast human head and shoulder moving.
- Missam_ Suzie which has the scene change with connecting of two video sequence (Miss America and Suzie).



(a)



(b)



(c)



(d)



(e)



(f)



(g)

Figure 4.1 Test video sequence uses in simulation : (a) Carphone sequence, (b) Foreman sequence, (c) Silent sequence, (d) News sequence, (e) Akiyo sequence and (f) Claire sequence (g) Missam_ Suzie

- 2) Input video sequences are in the QCIF format with a resolution of 176x144 pixels.
- 3) Structure of encoded video sequence consists of only one I-frame and the other are P-frames.
- 4) The target bit rate is varied from 16-256 kbps. The frame rate is set at 10 fps.
- 5) H.264 reference software version JM 8.6 with main profile is used in the simulation for performance comparison purpose. The parameters of simulation can be shown in Table 4.1.

Table 4.1 Parameters of simulation H.264 software version JM 8.6 with main profile

MV resolution	1/4 pel
Hadamard	On
RD optimization	Off
Search Range (pixels)	± 16
Reference frames	1
Restrict Search Range (pixels)	2
Symbol mode	UVLC
GOP structure	IPPP

Table 4.2 shows the average PSNR, standard deviation of PSNR, processing times and PSNR gain between H.264 JM 8.6 rate control and our proposed scheme at bit rate of 16-256 kbps. On average, the proposed scheme can achieve average PSNR improvement up to 0.45 dB, 0.42 dB, 0.40 dB, 0.34 dB and 0.36 dB for 16, 32, 64, 128 and 256 kbps, respectively compared to H.264 JM 8.6. On average, our proposed scheme can encode video with more uniform quality, as can be seen from lower standard deviation of PSNR compared to JM 8.6. In term of bit rate used, our proposed scheme which is based on Cauchy model can achieve more accurate bit rate, i.e., closer to target bit rate, than that of JM 8.6. Our proposed scheme also achieves lower processing time on average compared to JM 8.6. Figures 4.2 - 4.8 shows the average PSNR in each frame of Akiyo, Carphone, Claire, Foreman, News, Silent and Missam_Suzie sequence at 16 -256 kbps as target bit rates, respectively. In Fig.4.9

(a)- (g), show the average PSNR versus bit rate in each frame for our proposed and JM8.6 rate control of seven video sequences (Akiyo, Carphone, Claire, Foreman, News, Silent and Missam_Suzie sequence).

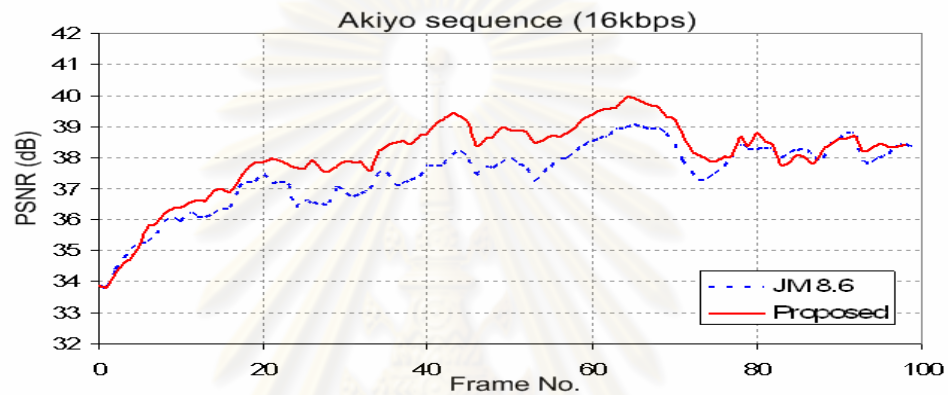
Figure 4.10, shown the subjective quality of rate control for Miss_Suz sequence of our proposed compared with the JM 8.6 rate control. In Fig.4.11, shown the subjective quality of rate control for News sequence of our proposed compared with the JM 8.6 rate control. From these results, it can be seen that our proposed scheme has better performance than the rate control algorithm in JM 8.6 both in terms of objective quality and the subjective quality.



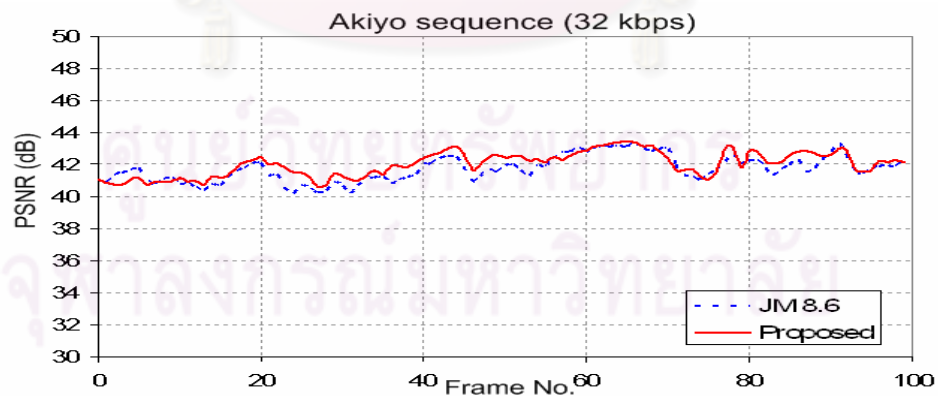
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Simulation results of Akiyo sequence

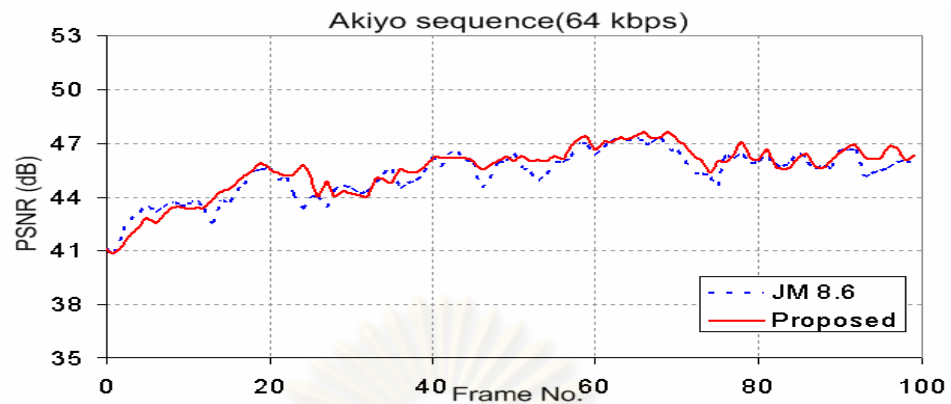
Number of frames to be encoded is 100. The simulation results as shown in Fig.4.2 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.63 dB with smoother video quality than that of H.264 JM 8.6 rate control.



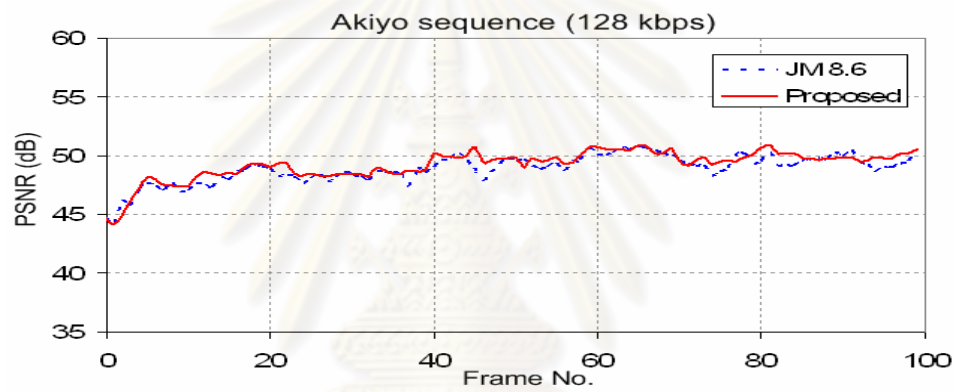
(a) Simulation results at 16 kbps



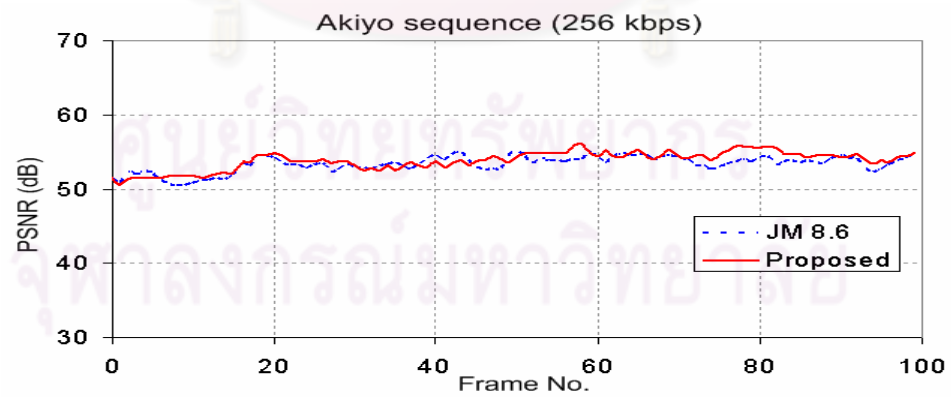
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

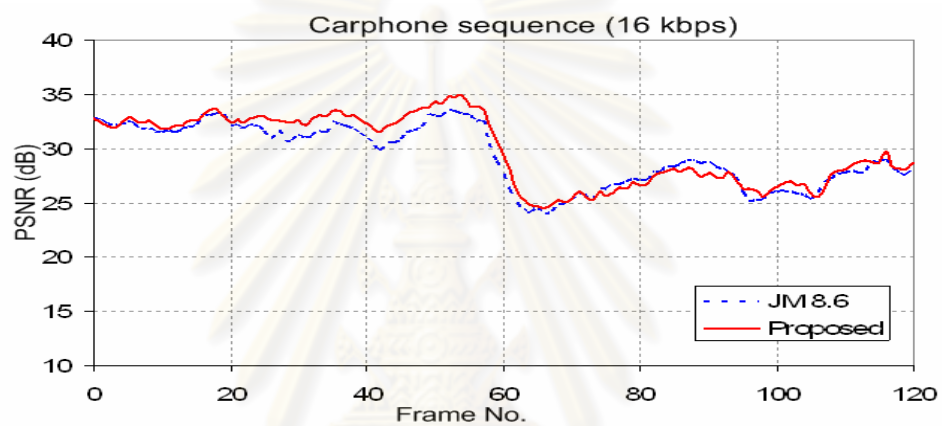


(e) Simulation results at 256 kbps

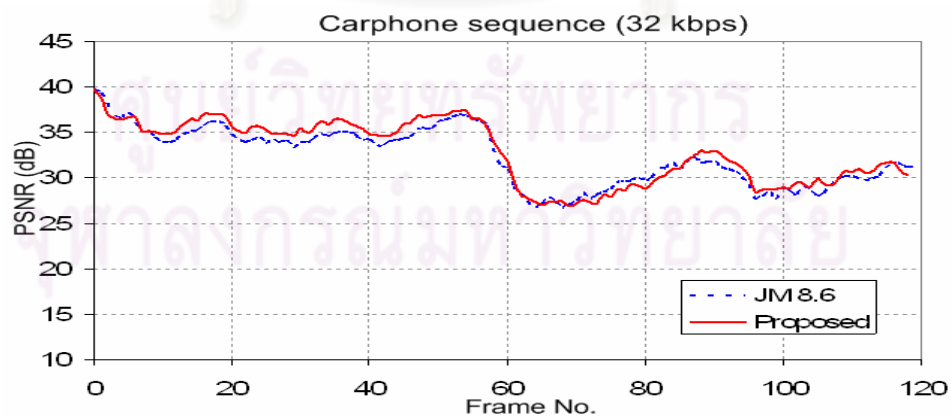
Figure 4.2 PSNR versus frame for our proposed and JM8.6 rate control of Akiyo sequence

Simulation results of Carphone sequence

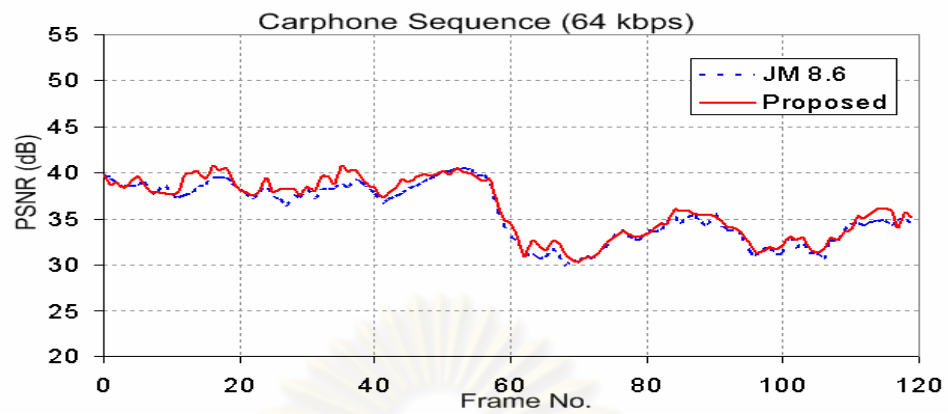
Number of frames to be encoded is 127. The simulation results as shown in Fig.4.3 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.54 dB with smoother video quality than that of H.264 JM 8.6 rate control.



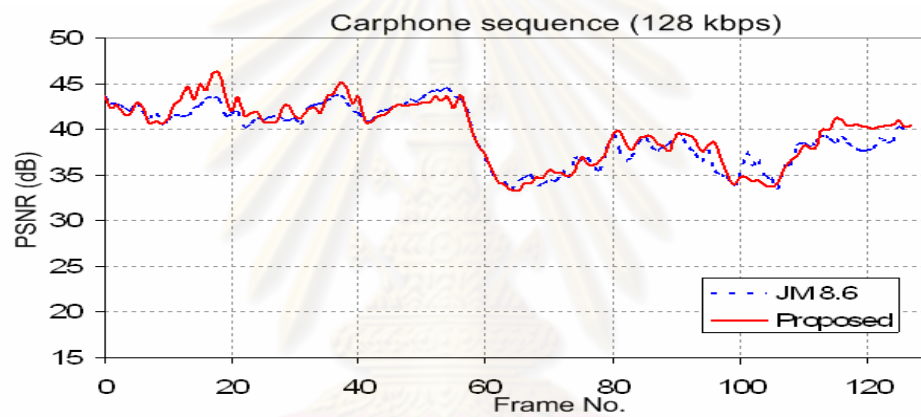
(a) Simulation results at 16 kbps



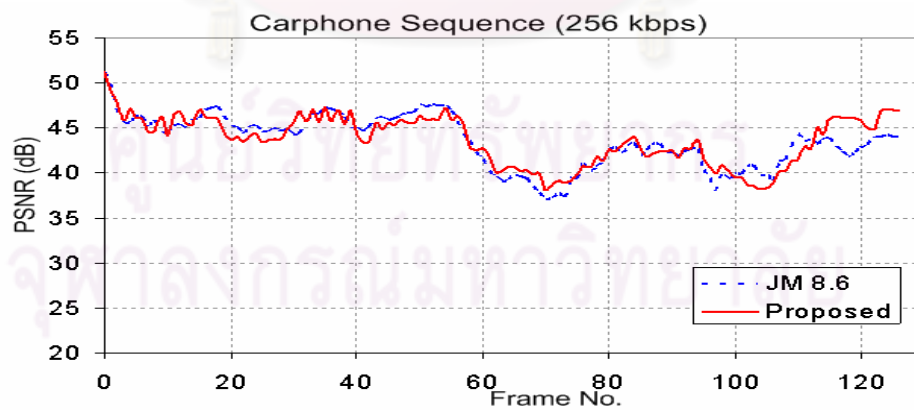
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

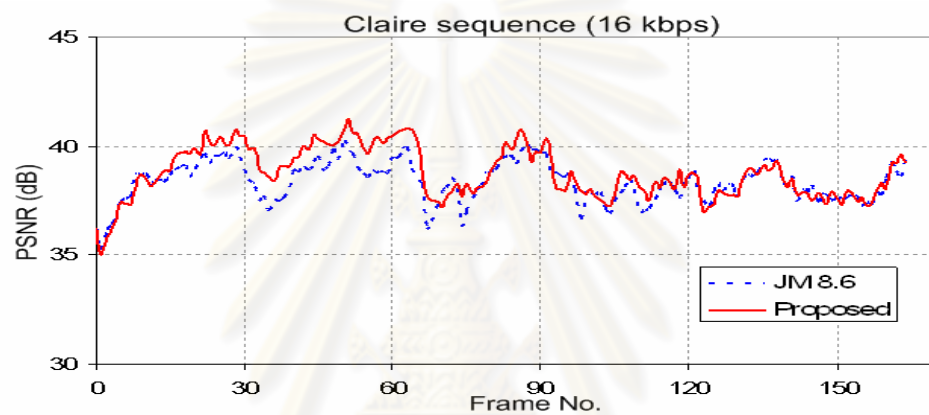


(e) Simulation results at 256 kbps

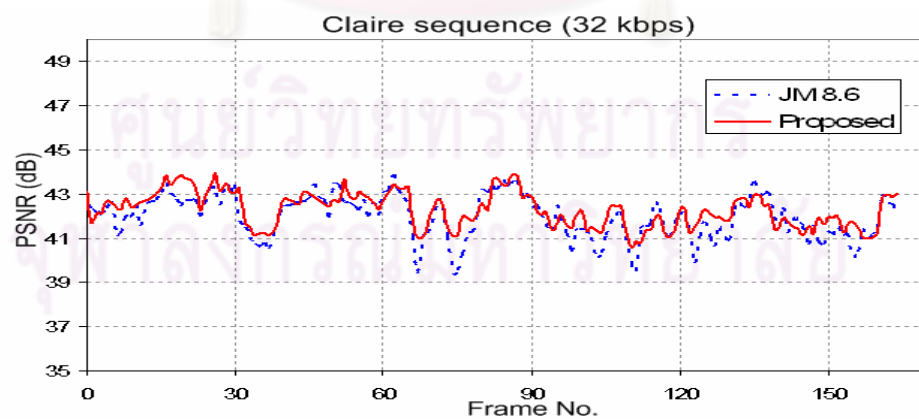
Figure 4.3 PSNR versus frame for our proposed and JM8.6 rate control of Carphone sequence

Simulation results of Claire sequence

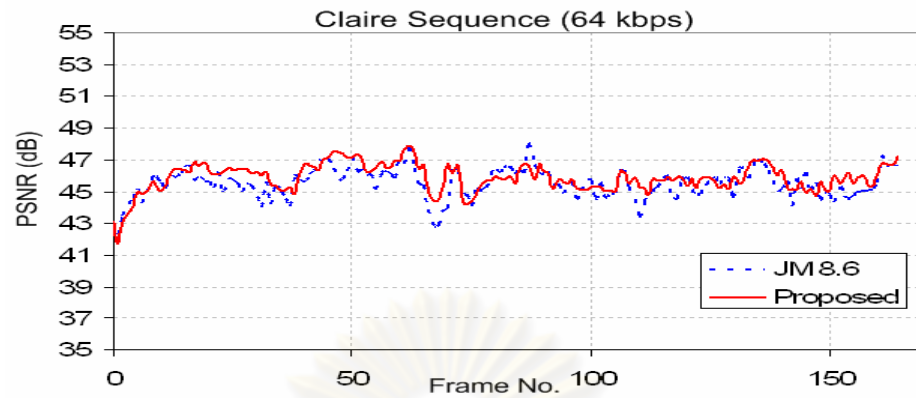
Number of frames to be encoded is 164. The simulation results as shown in Fig.4.4 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.45 dB with smoother video quality than that of H.264 JM 8.6 rate control.



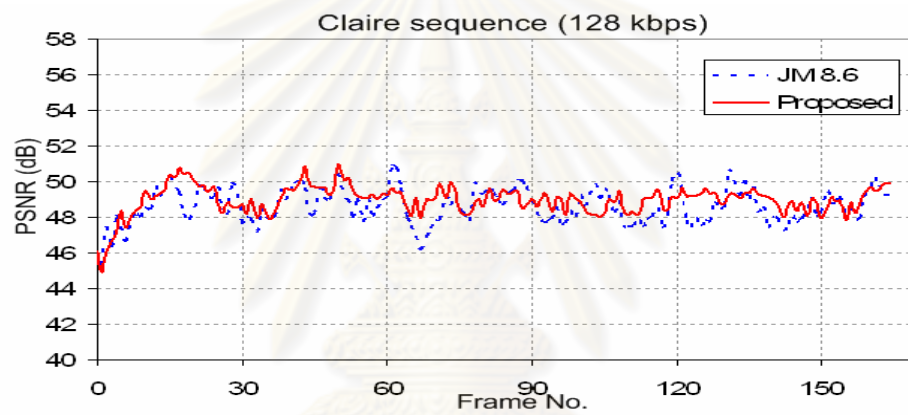
(a) Simulation results at 16 kbps



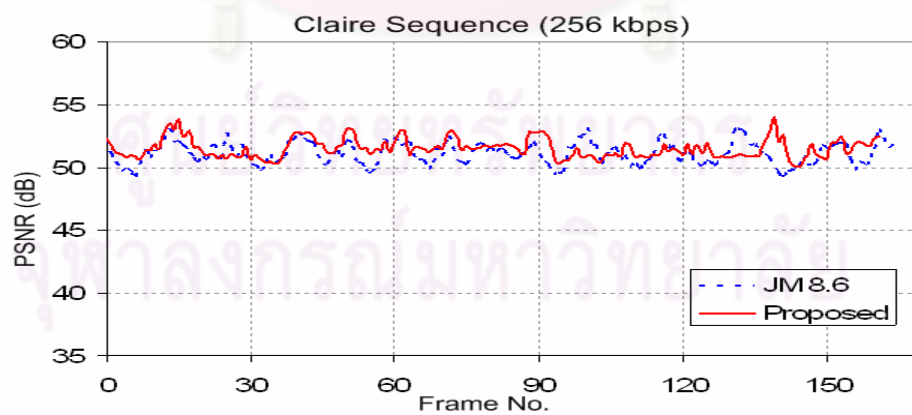
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

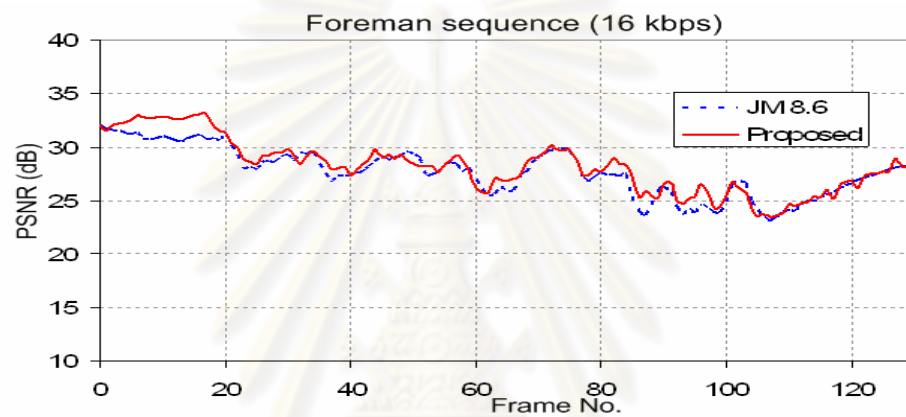


(e) Simulation results at 256 kbps

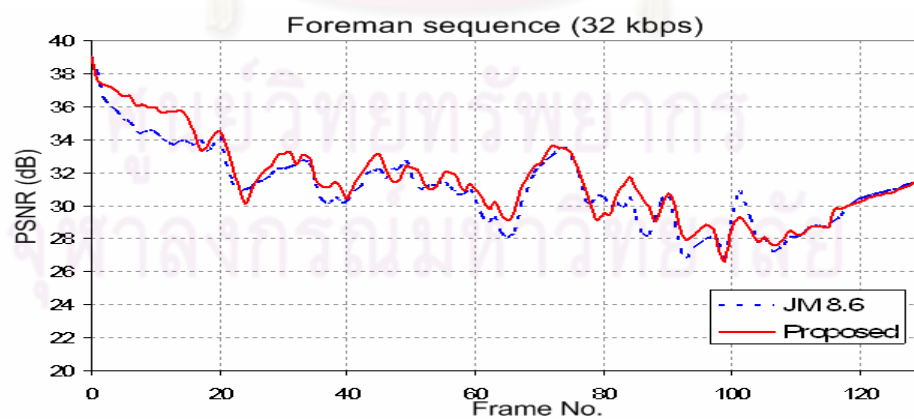
Figure 4.4 PSNR versus frame for our proposed and JM8.6 rate control of Claire sequence

Simulation results of Foreman sequence

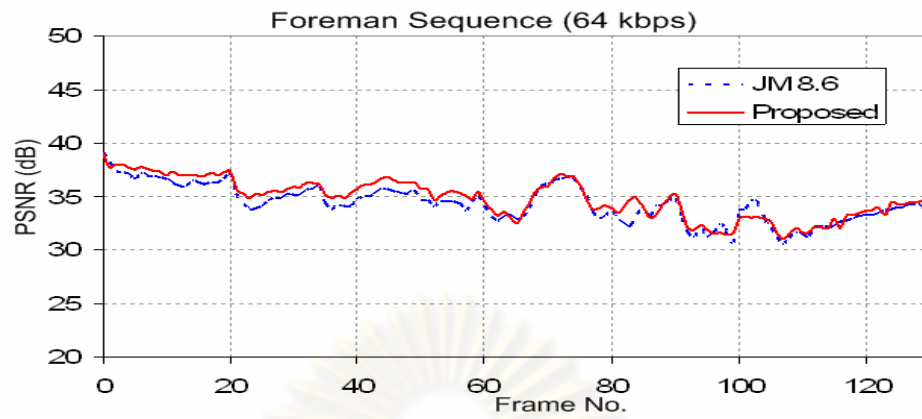
Number of frames to be encoded is 133. The simulation results as shown in Fig.4.5 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.53 dB with smoother video quality than that of H.264 JM 8.6 rate control.



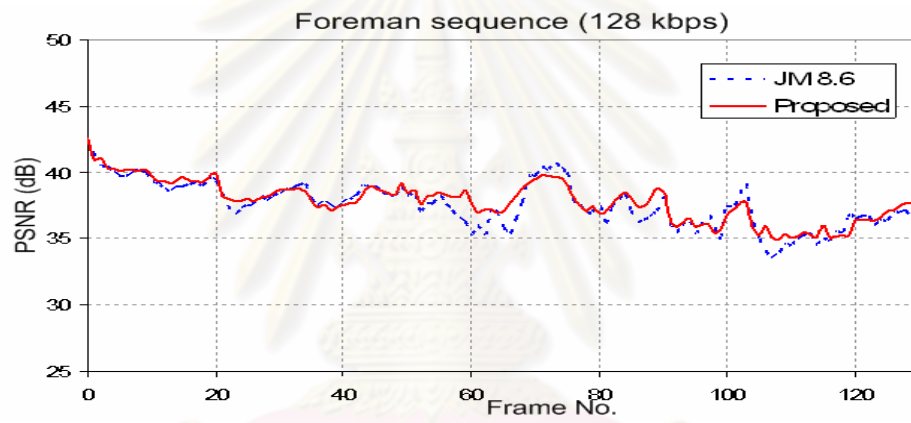
(a) Simulation results at 16 kbps



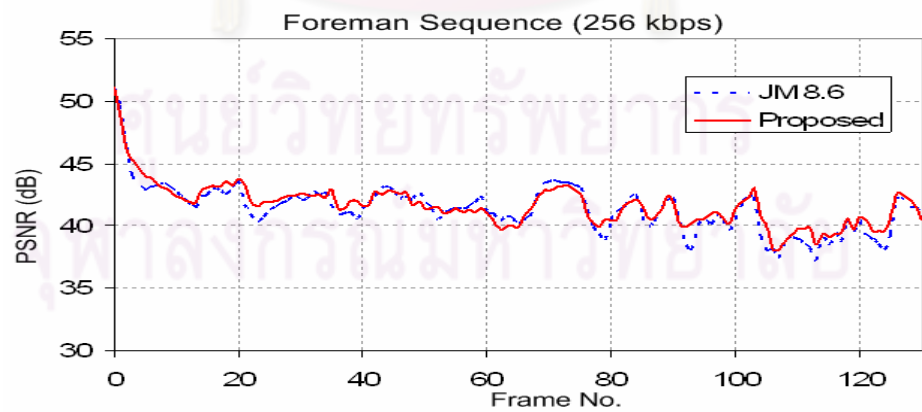
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

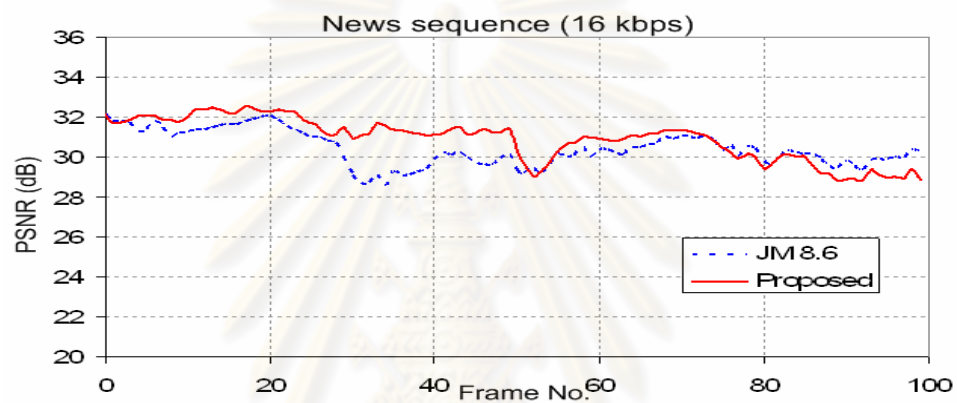


(e) Simulation results at 256 kbps

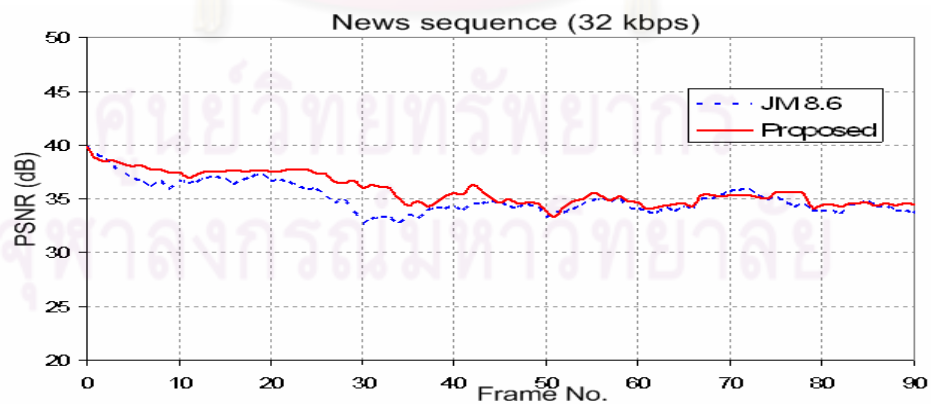
Figure 4.5 PSNR versus frame for our proposed and JM8.6 rate control of Foreman sequence

Simulation results of News sequence

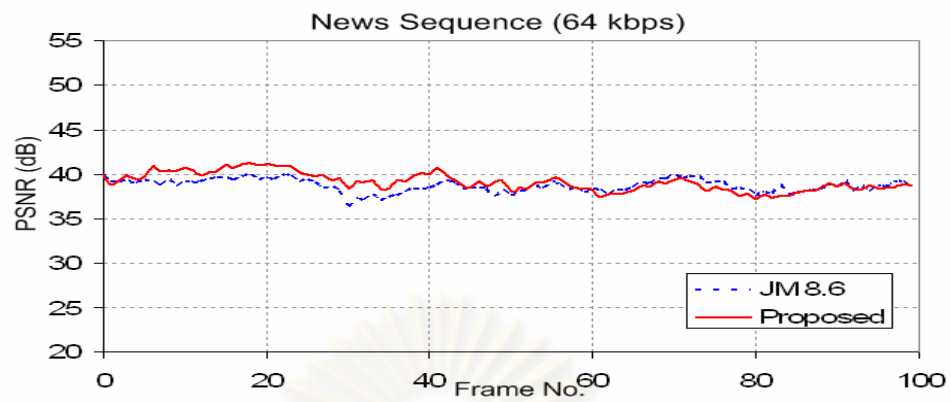
Number of frames to be encoded is 100. The simulation results as shown in Fig.4.6 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.69 dB with smoother video quality than that of H.264 JM 8.6 rate control.



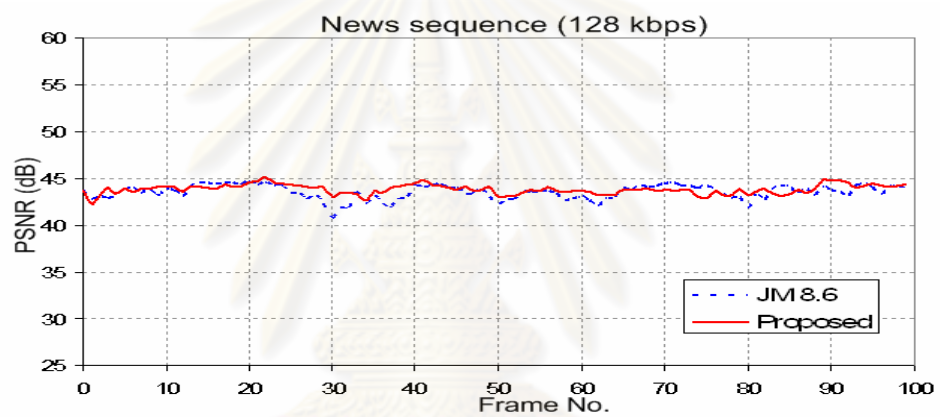
(a) Simulation results at 16 kbps



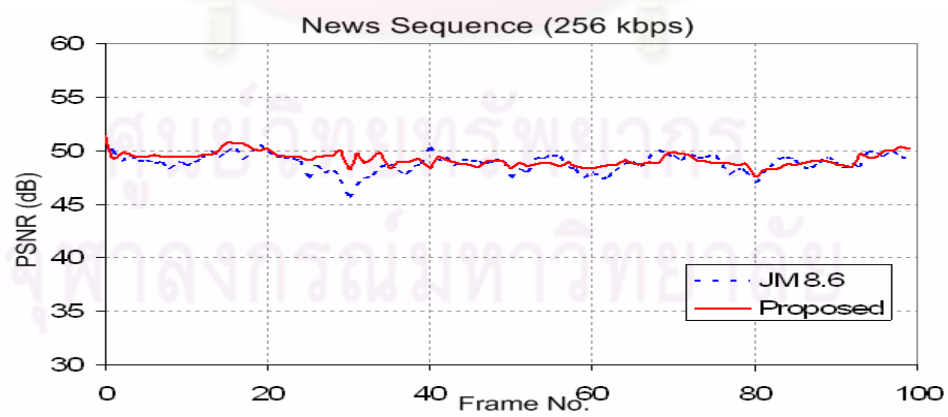
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

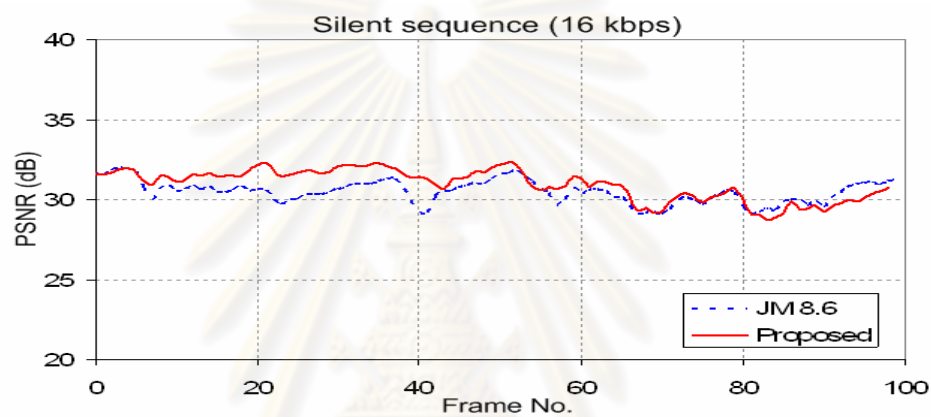


(e) Simulation results at 256 kbps

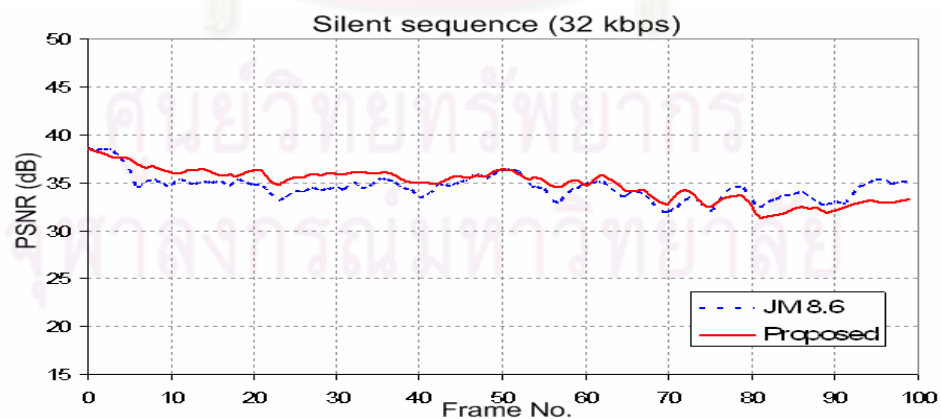
Figure 4.6 PSNR versus frame for our proposed and JM8.6 rate control of News sequence

Simulation results of Silent sequence

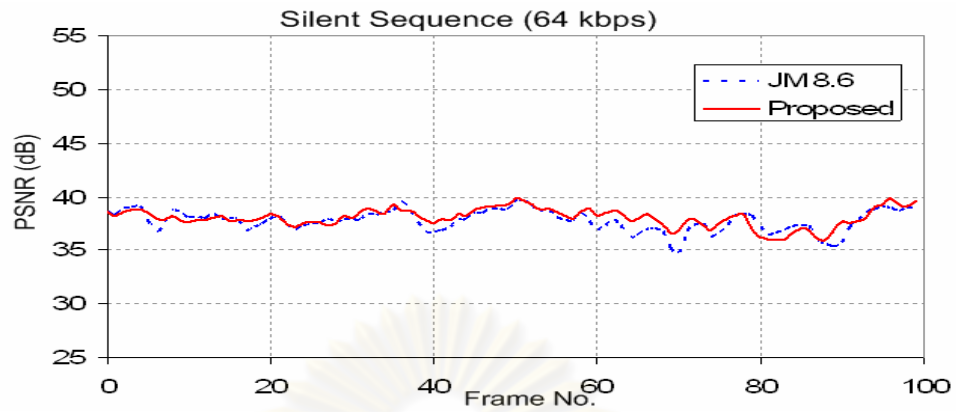
Number of frames to be encoded is 100. The simulation results as shown in Fig.4.7 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.48 dB with smoother video quality than that of H.264 JM 8.6 rate control.



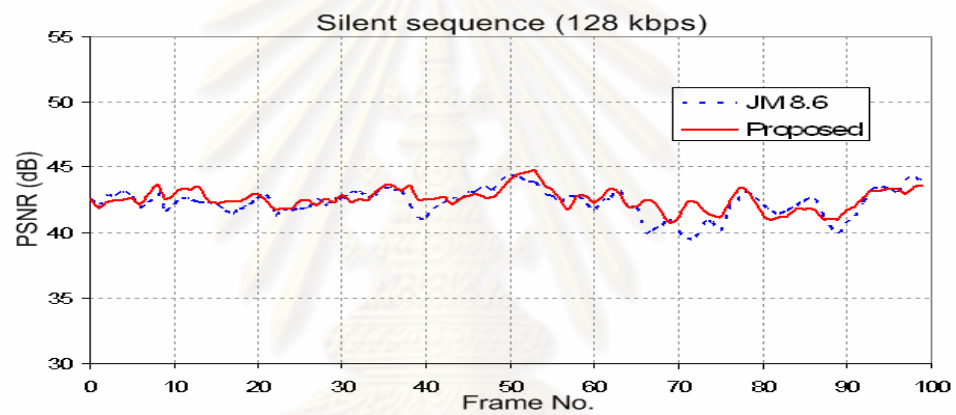
(a) Simulation results at 16 kbps



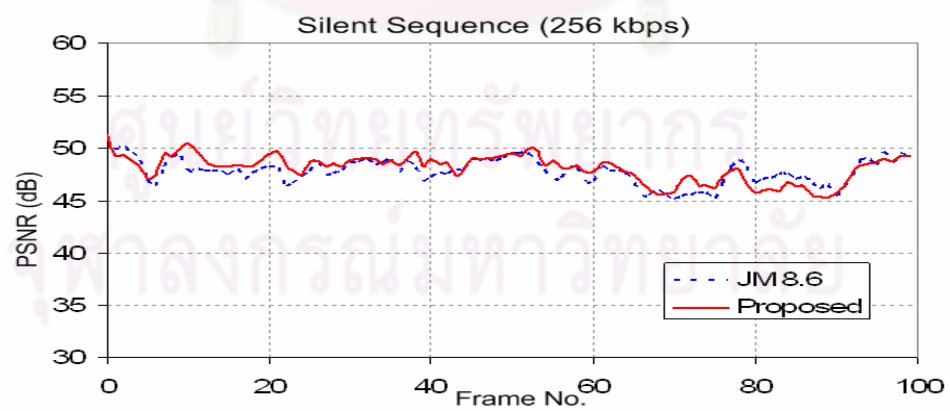
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps



(d) Simulation results at 128 kbps

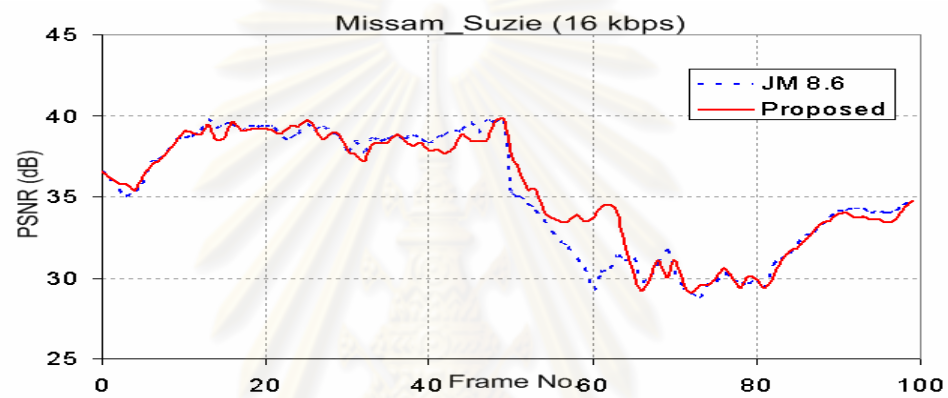


(e) Simulation results at 256 kbps

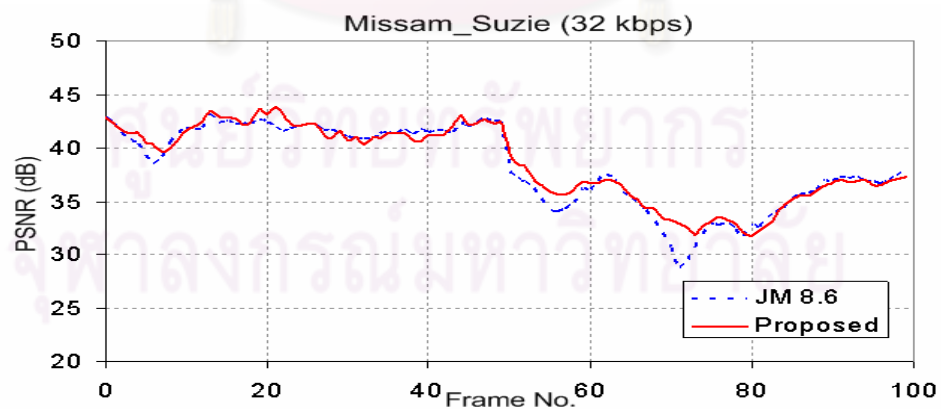
Figure 4.7 PSNR versus frame for our proposed and JM8.6 rate control of Silent sequence

Simulation results of Missam_Suzie sequence

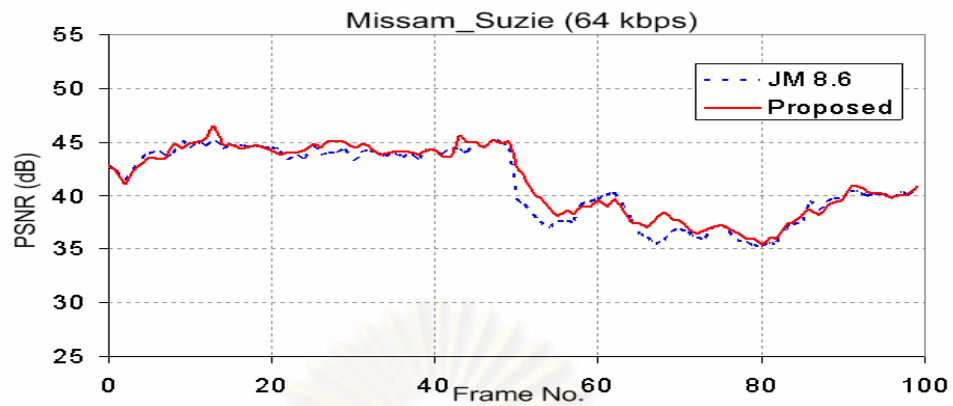
Number of frames to be encoded is 100. The simulation results as shown in Fig.4.8 (a) - (e), showed that the average PSNR at 16-256 kbps. From the experimental result, it can be seen that the proposed scheme can achieve average PSNR improvement up to 0.37 dB with smoother video quality than that of H.264 JM 8.6 rate control.



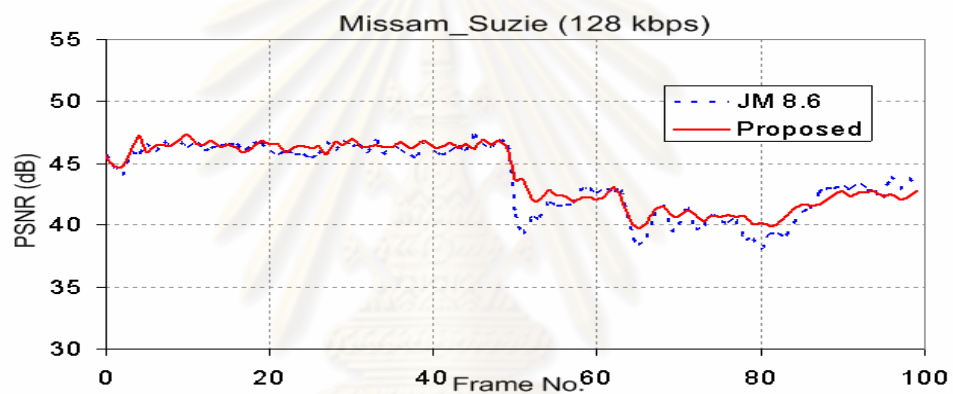
(a) Simulation results at 16 kbps



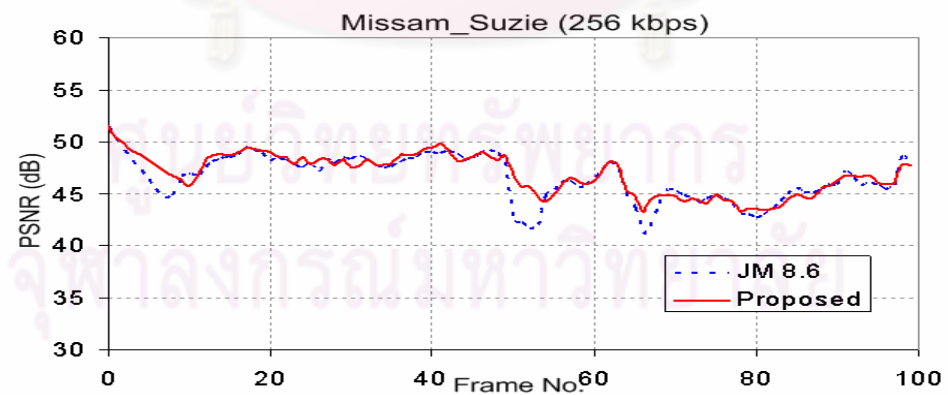
(b) Simulation results at 32 kbps



(c) Simulation results at 64 kbps

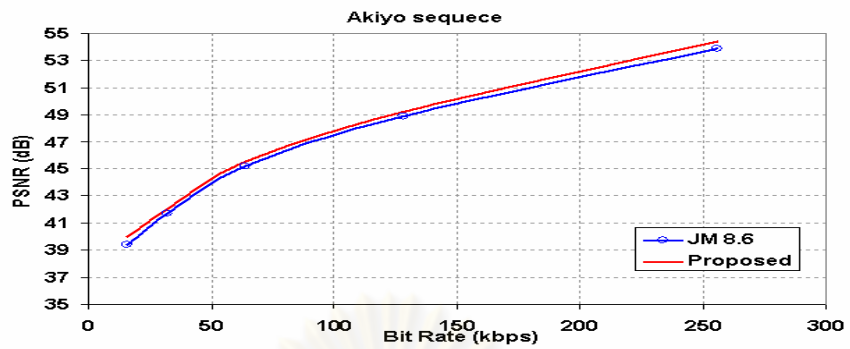


(d) Simulation results at 128 kbps

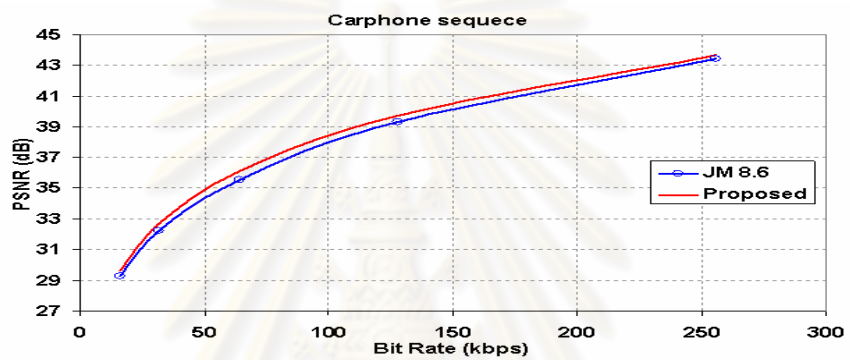


(e) Simulation results at 256 kbps

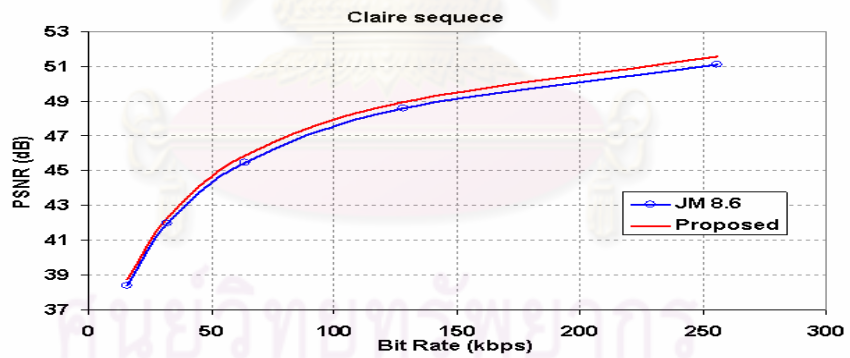
Figure 4.8 PSNR versus frame for our proposed and JM8.6 rate control of Missam_Suzie sequence



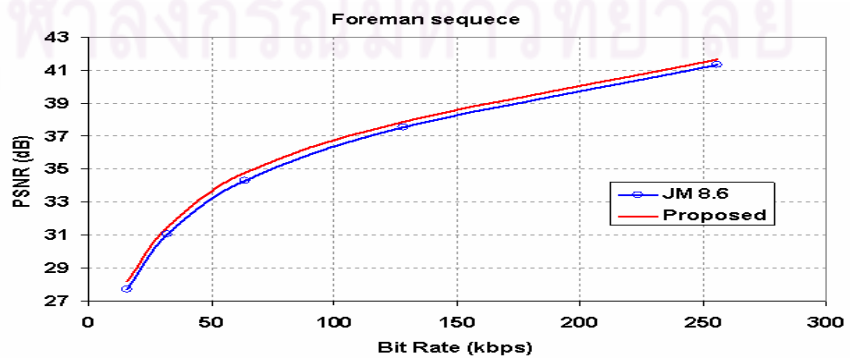
(a) Average PSNR versus bit rate in each frame for Akiyo sequence



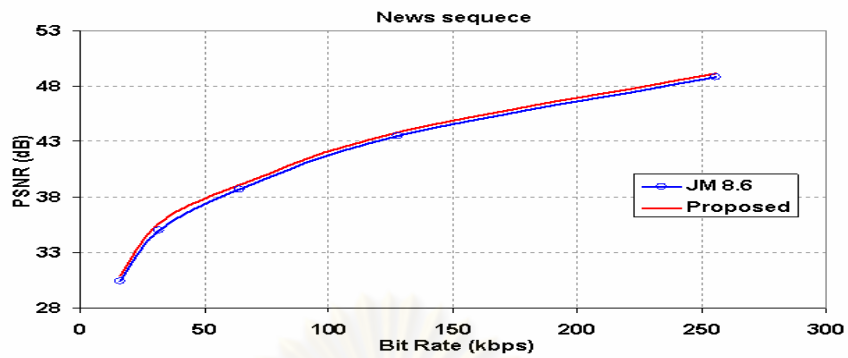
(b) Average PSNR versus bit rate in each frame for Carphone sequence



(c) Average PSNR versus bit rate in each frame for Claire sequence



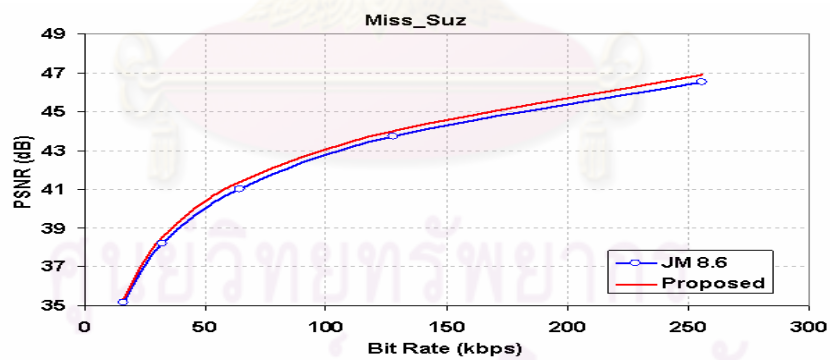
(d) Average PSNR versus bit rate in each frame for Foreman sequence



(e) Average PSNR versus bit rate in each frame for News sequence



(f) Average PSNR versus bit rate in each frame for Silent sequence



(g) Average PSNR versus bit rate in each frame for Missam_Suzie sequence

Figure.4.9 PSNR versus bit rate in each frame for our proposed and JM8.6 rate control of seven video sequence.



Figure 4.10 Decoded frame of rate control algorithm for Missam_Suzie sequence of our proposed compared with the H.264 JM 8.6 rate control.

(a1) 62 th frame of the proposed (PSNR= 34.56 dB) of 16 kbps

(a2) 62 th frame of the JM 8.6 (PSNR= 30.55 dB) of 16 kbps

(b1) 71 th frame of the proposed (PSNR= 32.90 dB) of 32 kbps

(b2) 71 th frame of the JM 8.6 (PSNR= 28.86 dB) of 32 kbps

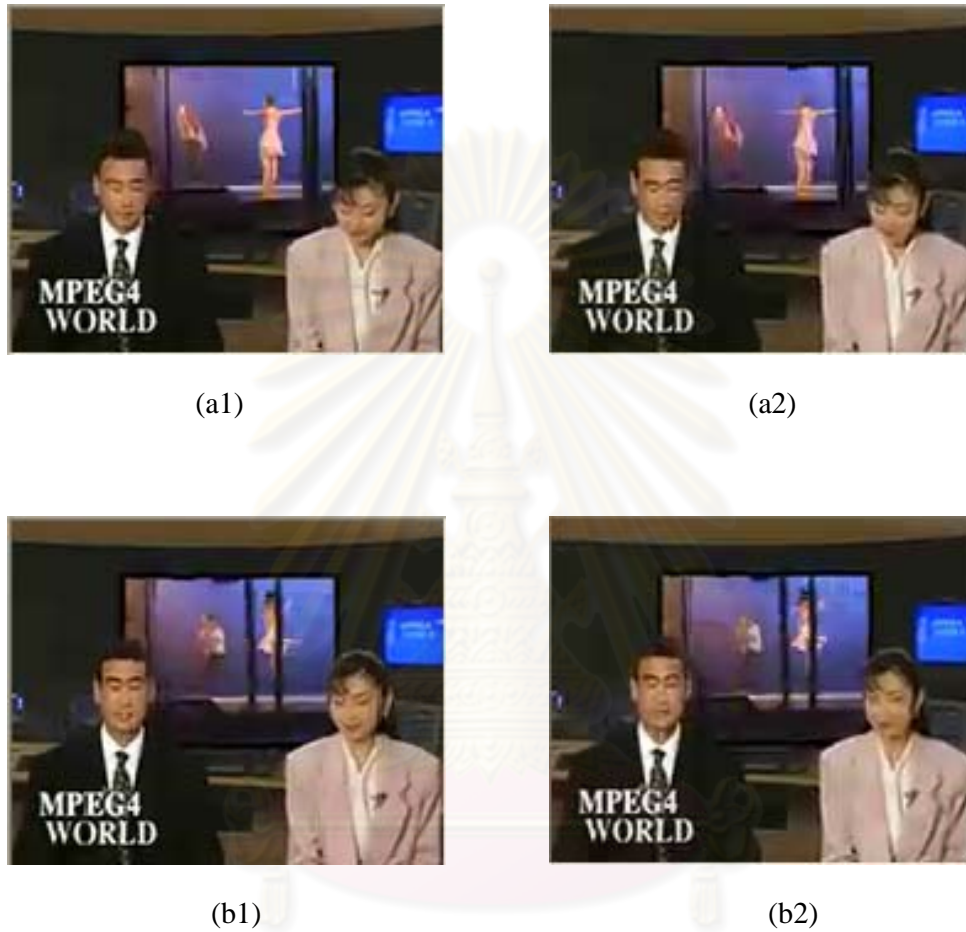


Figure 4.11 Decoded frame of rate control algorithm for News sequence of our proposed compared with the H.264 JM 8.6 rate control at 16 kbps.

(a1) 50 th frame of the proposed (PSNR= 38.67 dB)

(a2) 50 th frame of the JM 8.6 (PSNR= 37.67 dB)

(b1) 63 th frame of the proposed (PSNR= 39.91 dB)

(b2) 63 th frame of the JM 8.6 (PSNR= 38.92 dB)

Table 4.2 Performance of proposed scheme compared with H.264 JM8.6 rate control at 16, 32 and 64 kbps as target bit rate

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.		Processing Times (ms)		Bit Rate (kbps)		PSNR Gain (dB)
		JM 8.6	Proposed	JM8.6	Proposed	JM8.6	Proposed	JM8.6	Proposed	
16 kbps	Akiyo	39.40	40.03	1.12	1.26	233.25	224.83	16.08	16.10	+0.63
	Claire	38.36	38.76	0.95	0.92	222.15	223.88	16.08	16.07	+0.40
	Silent	30.45	30.93	0.71	0.75	223.59	224.02	16.07	16.18	+0.48
	News	30.41	30.90	0.87	0.90	219.43	212.81	16.07	16.03	+0.49
	Carphone	29.27	29.62	2.73	2.42	232.34	235.41	16.02	16.13	+0.35
	Foreman	27.68	28.21	2.24	2.06	229.94	230.07	16.11	16.07	+0.53
	Miss_Suz	35.17	35.37	3.67	3.40	272.06	230.34	15.99	16.15	+0.20
32 kbps	Akiyo	41.71	42.02	0.81	0.77	233.88	212.77	31.98	32.06	+0.31
	Claire	41.96	42.27	0.98	0.80	215.52	215.50	31.92	31.97	+0.31
	Silent	34.50	34.85	1.30	1.25	213.88	212.19	32.08	32.05	+0.35
	News	34.97	35.66	1.39	1.37	212.75	213.90	32.39	32.12	+0.69
	Carphone	32.24	32.69	3.10	2.98	219.49	220.15	32.32	32.11	+0.45
	Foreman	31.04	31.48	2.29	2.18	225.80	226.10	32.07	32.03	+0.44
	Miss_Suz	38.23	38.54	3.81	3.56	255.56	248.26	31.95	32.23	+0.31
64 kbps	Akiyo	45.22	45.50	1.33	1.37	231.79	211.99	63.87	64.11	+0.28
	Claire	45.48	45.85	0.95	0.89	219.53	220.08	63.85	64.05	+0.37
	Silent	37.74	38.05	0.98	0.87	219.01	218.95	63.98	64.10	+0.31
	News	38.68	39.13	0.73	0.72	215.22	215.43	64.31	64.23	+0.45
	Carphone	35.56	36.10	3.01	3.04	223.25	223.81	64.98	64.12	+0.54
	Foreman	34.32	34.80	1.71	1.62	227.60	227.69	63.98	64.08	+0.48
	Miss_Suz	41.02	41.39	3.28	3.16	232.85	230.14	63.63	64.19	+0.37

Table 4.3 Performance of proposed scheme compared with H.264 JM8.6 rate control at 128 and 256 as target bitrate

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.		Processing Times (ms)		Bit Rate (kbps)		PSNR Gain (dB)
		JM 8.6	Proposed	JM8.6	Proposed	JM8.6	Proposed	JM8.6	Proposed	
128 kbps	Akiyo	48.89	49.24	1.22	1.21	231.24	212.45	127.91	127.86	+0.35
	Claire	48.60	48.96	0.99	0.81	221.19	221.30	127.32	127.88	+0.36
	Silent	42.26	42.55	1.05	0.88	219.68	220.06	127.58	128.03	+0.29
	News	43.52	43.83	0.77	0.50	218.94	217.16	128.14	128.19	+0.31
	Carphone	39.29	39.73	3.10	3.28	221.20	220.48	127.96	128.06	+0.44
	Foreman	37.55	37.84	1.69	1.50	227.96	227.52	127.98	128.03	+0.29
	Miss_Suz	43.70	44.01	2.72	2.48	234.97	232.62	127.59	128.05	++0.31
256 kbps	Akiyo	53.37	53.87	1.09	1.12	239.29	207.69	256.16	256.08	+0.50
	Claire	51.09	51.54	0.90	0.77	221.76	221.54	254.93	255.27	+0.45
	Silent	47.73	48.05	1.27	1.23	212.14	212.15	255.37	256.04	+0.32
	News	48.80	49.16	0.87	0.64	219.55	218.34	256.71	256.33	+0.36
	Carphone	43.44	43.64	2.91	2.75	226.18	226.47	255.93	256.06	+0.22
	Foreman	41.33	41.64	1.96	1.73	226.34	226.73	254.28	255.37	+0.31
	Miss_Suz	46.53	46.89	2.16	1.93	236.72	235.41	253.25	256.69	+0.36

4.5 Summary

In this Chapter, we investigate and analyze the use of Cauchy distribution as a model to estimate the rate and distortion characteristics for video coding on the application of rate control. Based on Cauchy R-D optimization model, we derive an expression for optimal quantization step size and a linear prediction rate and distortion model parameters as shown in Chapter 3. We then propose a rate control scheme using Cauchy R-D optimization model. The consideration of bit allocation involves the number of bits used in previous frame, the complexity of basic unit in term of residual variance of each basic unit. The simulation results show that our proposed scheme achieves better performance in terms of better PSNR, lower PSNR standard deviation, more accurate bit rate used, and almost the same processing time compared to that of H.264 JM 8.6 for the ranges of low bit rate constraint indicated.



Chapter 5

An improved rate control based on Cauchy rate-distortion optimization model for low bit-rate H.264 video coding under low delay constraint

In this chapter, an improved rate control based on Cauchy rate-distortion optimization model for H.264 low bit-rate video coding under low delay constraint is presented. This portion of work is an extension of the proposed rate control in chapter 4. In real time video transmission, delay is very critical because video frames must be displayed to the viewer at constant intervals. One factor that contributes to the delay is the encoding process. Buffering the data prior to transmission in a constant bit rate (CBR) encoder poses a considerable trade off as a larger buffer size implies a larger delay. Nevertheless, in low delay video transmission scenario, the buffer size will be small. Thus, under normal operation of rate control, bits accumulated in the encoder buffer will result in the higher number of bits than speculated, thus frame will be skipped to reduce the buffer delay and to avoid buffer overflow. We also show the impact of low delay constraint based on the H.264 rate control algorithm, and show that the degraded performance of H.264 rate control under low delay constraint. Then taking into account the low delay factor, we propose a new rate control based on Cauchy-based rate-distortion (R-D) optimization model for H.264 low delay video transmission.

5.1 Impact of low delay constraint of H.264 JM 8.6 rate control

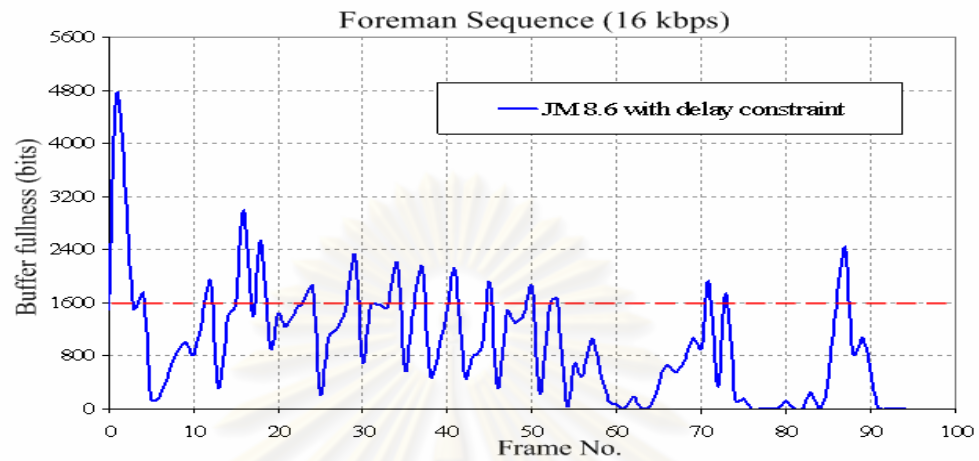
In this section, we study the impact of the low delay constraint of H.264 rate control. In encoding process, the bits left from previously encoded frame that has not been transmitted produce a delay in an order of few milliseconds. To reduce such delay, the encoder buffer size should be kept small. In our study, we impose a low delay constraint that the number of bits in encoder buffer $B_c(n_{i,j})$ must not be greater than the maximum buffer size M , otherwise a frame will be skipped, where,

$M = B_s = u(n_{i,j}) / F_r$. Hence, the maximum buffer delay is $M / u(n_{i,j}) = 1 / F_r$ seconds. Where $u(n_{i,j})$ and F_r is target channel bandwidth and frame rate, respectively.

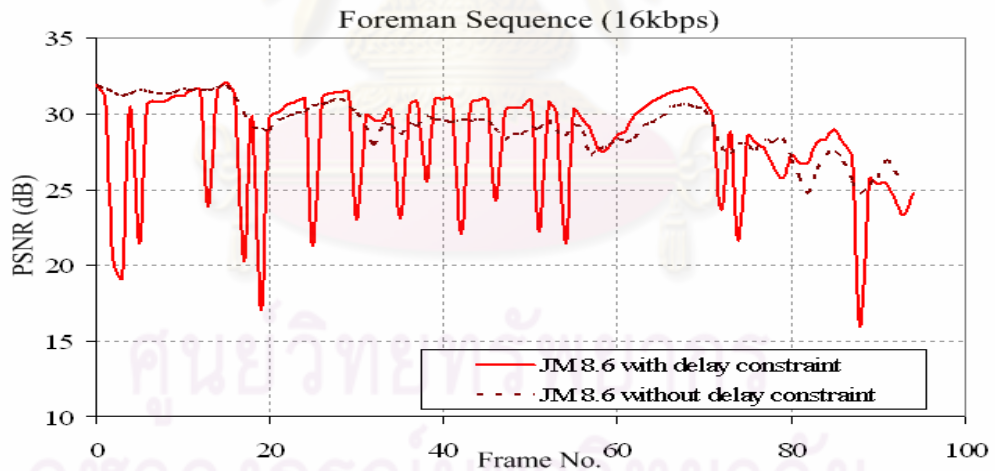
To show the impact of the low delay constraint of JM8.6 rate control [15], we encoded an I-frame followed by P-frame of various video sequences at 16 kbps with the frame rate of 10 fps. The maximum buffer delay in our simulation is set to 100 ms. Example results of impacts low delay constraint for the impact toward H.264 JM 8.6 rate control mechanism can be seen in Fig.5.1-5.3 for Foreman, Carphone and Silent sequences, respectively. As shown in Fig.5.1-5.3 (a) and (b), when the buffer fullness level is higher than maximum delay constraint, frame will be skipped. Thus, without the proposed algorithm described in our proposed rate control scheme, there are several sharp drops in PSNR and additional frames skipped compared to the case with no delay constraint.

- Fig.5.1 (b), shows PSNR degradation of Foreman sequence. On the average, PSNR degrades 0.63 dB with 17 frames skipped, when compared to the case with no delay constraint.
- Fig.5.2 (b), shows PSNR degradation of Carphone sequence. On the average, PSNR degrades 0.3 dB with 10 frames skipped, when compared to the case with no delay constraint.
- Fig.5.3 (b), shows PSNR degradation of Silent sequence. On the average, PSNR degrades 3.0 dB with 10 frames skipped, when compared to the case with no delay constraint.

The simulation results show that, when the buffer fullness level is higher than the maximum delay constraint, i.e., the maximum buffer size, frame will be skipped. This results in the motion discontinuity and low visual quality in video communication system.

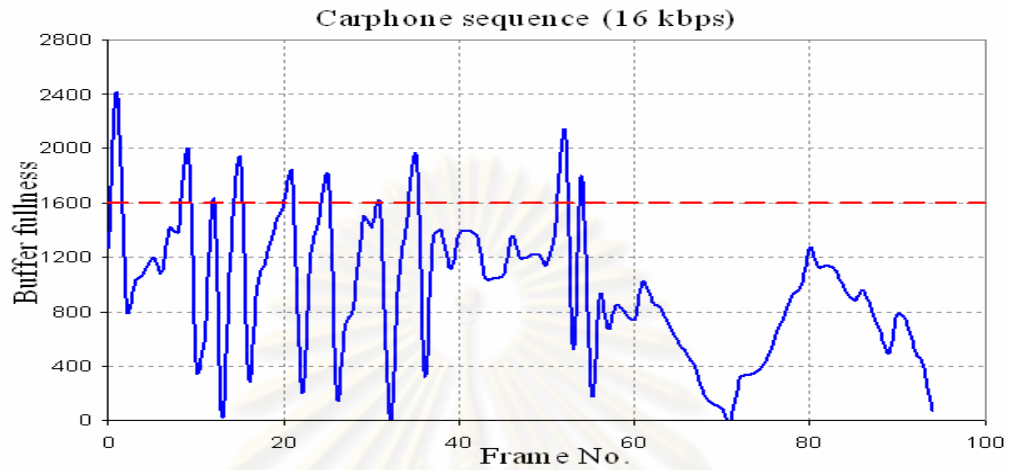


(a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).

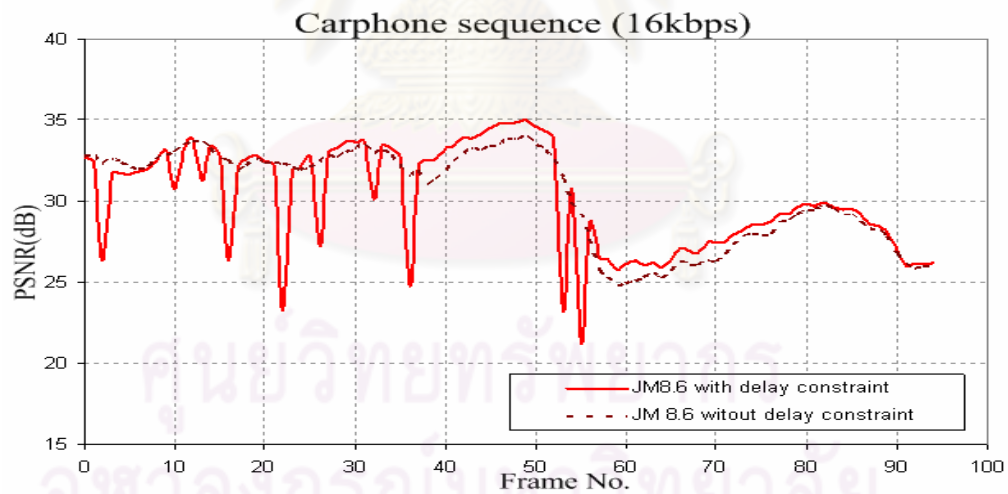


(b) Average PSNR per frame.

Figure 5.1 Simulation using Foreman sequence encoded by H.264 JM8.6 rate control at 16 kbps with delay constraint of 100 ms.

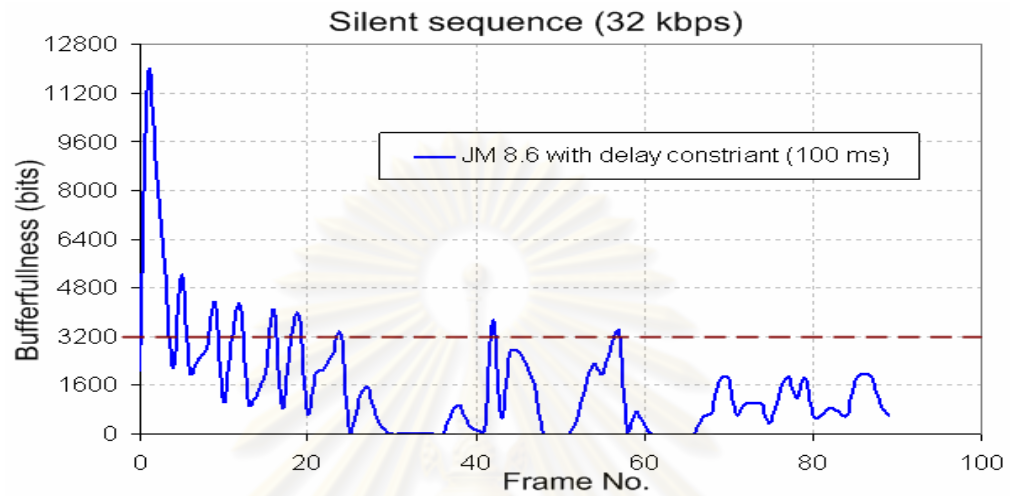


(a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).

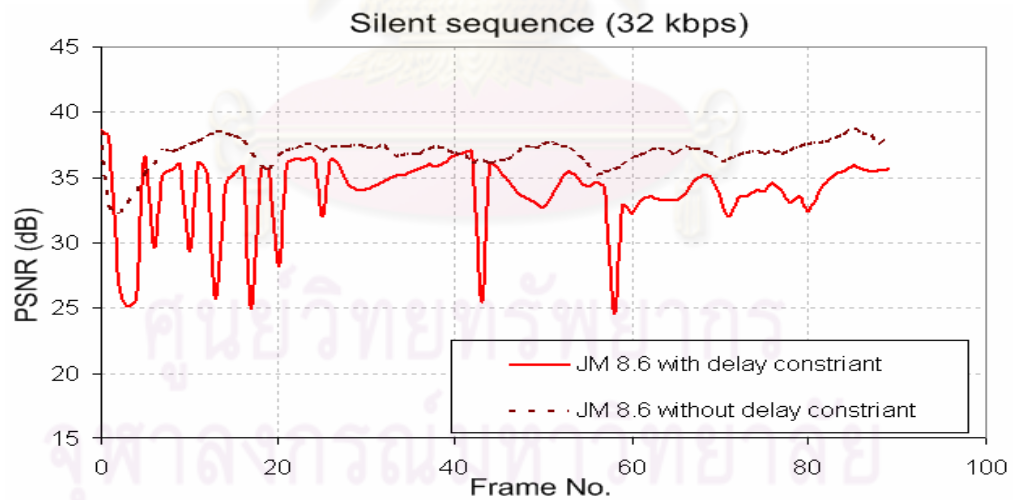


(b) Average PSNR per frame.

Figure 5.2 Simulation using Carphone sequence encoded by H.264 JM8.6 rate control at 16 kbps with delay constraint of 100 ms.



(a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line).



(b) Average PSNR per frame.

Figure 5.3 Simulation using Silent sequence encoded by H.264 JM8.6 rate control at 32 kbps with delay constraint of 100 ms.

5.2 Proposed rate control scheme using Cauchy R-D optimization model under low delay constraint

As shown in the previous section, the impact of low delay constraint for H.264 JM 8.6 is the problem in real time video transmission. In this chapter, we use the expression in eq.(3.26) to design a new rate control scheme to reduce the number of frame skipped with an improvement in average PSNR and smoother video quality. Cauchy R-D models presented in the previous section can be used to find the optimal quantization step size in each basic unit of each frame.

Our proposed rate control scheme with Cauchy R-D optimization model composes of three layers: Group of picture layer (GOP layer), frame layer and basic unit layer rate control.

- In GOP layer rate control, the computation of the total number of bits in each GOP and the remaining number of bits for all noncoded P-frame are proposed and compute the occupancy of virtual buffer after each frame is encoded.
- In frame layer rate control, the objective of this stage is to determine the number of target bits budget for each P-frame.
- In basic unit layer rate control, we proposed the computation of quantization step size of current basic unit based on Cauchy R-D optimization model and basic unit complexity in term of residual variance. The computed quantization parameter (QP) is then adjusted to prevent buffer from overflow and underflow.

5.2.1 GOP layer rate control for low delay constraint

First, the first I-frame and the first P-frame of the GOP are coded by using $QP_{initial}$. The computation of $QP_{initial}$ is based on the available channel bandwidth as used in H.264 rate control [15]. The total number of bits allocated for the GOP is computed as shown in eq.(5.1) and the remaining bits, $T_r(n_i, j)$, for all non-coded P-frames are updated after the $(j-1)^{th}$ frame is encoded, as shown in eq.(5.2),

$$T_{GOP} = \frac{u(n_{i,j})}{F_r} \times N_{GOP} \quad (5.1)$$

$$T_r(n_{i,j}) = T_{GOP} - \sum_{j=0}^{j-1} b(n_{i,j}) \quad (5.2)$$

, where $u(n_{i,j})$ is the target channel bandwidth, F_r is the frame rate, N_{GOP} is the number of frames in each GOP, and $b(n_{i,j})$ is the actual number of bits generated by each j^{th} frame. After each frame are encoded, the occupancy of virtual buffer, $B_c(n_{i,j})$, are updated, as shown in eq.(2.11) in Chapter 2.

5.2.2 Frame layer rate control for low delay constraint

The objective of this stage is to determine the number of target bits budget before coding of the current j^{th} frame. The algorithm computes the target bits for each P-frame, $f(n_{i,j})$, according to the current buffer occupancy and the number of bits used in the previous P-frame are explained in the following steps.

Step 1 : Number of bits used in the previous P-frame is used to compute the target bits, $\hat{f}(n_{i,j})$, as shown in eq.(5.3).

$$\hat{f}(n_{i,j}) = \frac{T_r(n_{i,j})}{N_{GOP} - N_c} \quad (5.3)$$

Step 2 : Current buffer occupancy and buffer size information are used to compute the target bits for each P-frame, $f(n_{i,j})$, as shown in eq.(5.4),

$$f(n_{i,j}) = \hat{f}(n_{i,j}) + \omega V_S - B_c(n_{i,j}) \quad (5.4)$$

, where ω is constant value. Note that ω is empirically set to be 0.8 in our simulations. V_S is the maximum buffer size.

5.2.3 Basic unit layer rate control for low delay constraint

In basic unit layer, the suitable quantization step sizes are obtained using Cauchy R-D optimization model as described in Chapter 3. The quantization parameter is further adjusted to keep bit rate under the given constraints, and to prevent the buffer from overflow and underflow. Model parameters are updated after encoding each basic unit. The layer is divided into pre-encoding and post-encoding stages.

1) Pre-encoding stage

In this stage, we compute the quantization step size of current basic unit in three cases.

Case 1 : For the first basic unit in the current frame, quantization parameter is obtained, as shown in eq. (5.5),

$$QP_{1,i}(j) = AvgQP_i(j-1) \quad (5.5)$$

,where $QP_{1,i}(j)$ is the quantization parameter of the first basic unit and $AvgQP_i(j-1)$ is the average of quantization parameter for all basic units in the previous frame.

Case 2 : When the number of remaining bits, $T_r(n_{i,j})$, is lower than zero, quantization parameter is obtained, as shown in eq. (5.6).

$$QP_{l,i}(j) = QP_{l-1,i}(j) + 2 \quad (5.6)$$

Case 3 : The quantization step size of current basic unit is computed by Cauchy R-D optimization model using the formula in eq. (3.24), where a_{j_l} , α_{j_l} , b_{j_l} , and β_{j_l} are model parameters of the basic unit l th of frame j , respectively. The algorithms are according to the following steps.

Step 1: To get a better target bits estimation for each frame, we need to consider the complexity factor of each basic unit, $\gamma(j,l)$. It is defined in eq. (5.7),

$$\gamma(j,l) = \frac{\text{Re}_{s_{\text{var}}}(j,l)}{\text{Ave_Re}_{s_{\text{var}}}(j-1)} \quad (5.7)$$

, where $\text{Re}_{s_{\text{var}}}(j,l)$ is the residual variance of the basic unit l^{th} and $\text{Ave_Re}_{s_{\text{var}}}(j-1)$ is the average of residual variance of all basic units in the previous frame $(j-1)^{\text{th}}$. Note that the value of $\gamma(j,l)$ is bounded in the range of [0.8,1.2] to prevent too much fluctuations in target bit estimation. The new frame target, $f_v(n_{i,j})$, is thus adjusted, as shown in eq.(5.8).

$$f_v(n_{i,j}) = \gamma(j,l) \times f(n_{i,j}) \quad (5.8)$$

Step 2: To reduce the number of frame skipped when the buffer occupancy level is high, i.e., if the level reaches 80%, the target number of bits obtained from eq. (5.8) will be reduced by 10%. Also to prevent buffer underflow, i.e., the buffer occupancy level is less than or equal to zero, the target number of bits obtained from eq. (5.8) will be increased by 10%. The adjustment conditions are shown in eqs. (5.9)-(5.10),

$$R_{\text{MAX}}(j) = \eta \times f_v(n_{i,j}) \quad (5.9)$$

, where $R_{\text{MAX}}(j)$ is defined as the maximum target bit budget in the current frame, η is constant value. In our simulations, we set the value of η according to the current buffer occupancy conditions, as shown in eq.(5.10).

$$\eta = \begin{cases} 1.10 & \text{if } B_c(n_{i,j}) \leq 0 \\ 0.90 & \text{if } B_c(n_{i,j}) \geq 0.8 \times V_S \\ 1.00 & \text{otherwise} \end{cases} \quad (5.10)$$

Step 3: To avoid the fluctuation of PSNR and buffer overflow, the target bit budget estimation of current frame are bounded by a lower and upper bounds, as shown in eqs. (5.11) and (5.12),

$$Lower_bound_R_{MAX} = MAX \left\{ R_{MAX}(j), \left(\phi \times \frac{u(n_{i,j})}{F_r} \right) \right\} \quad (5.11)$$

$$Upper_bound_R_{max} = MIN \left\{ R_{MAX}(j), \left(\gamma \times \frac{u(n_{i,j})}{F_r} \right) \right\} \quad (5.12)$$

, where ϕ and γ are constant values and are set empirically to 0.5 and 3, respectively.

Step 4: Compute the quantization step size (Q_l) of the current basic unit by using the parameters, a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} , $R_{MAX}(j)$ of the basic unit l th of frame j .

The computed quantization parameter of each basic unit is then further adjusted, as shown in eq.(5.13),

$$\begin{aligned} \text{If } B_s(n_{i,j}) &\geq 1.2 \times B_s \\ Q_{P_{l,i}}(j) &= MAX(QP_{initial} + 5, QP_{l,i}(j) + 3) \\ \text{else if } B_s(n_{i,j}) &\leq 0.2 \\ Q_{P_{l,i}}(j) &= MAX(QP_{initial} - 2, QP_{l,i}(j)) \\ \text{else} \\ Q_{P_{l,i}}(j) &= MAX(QP_{initial} - 1, QP_{l,i}(j)) \end{aligned} \quad (5.13)$$

, where $B_S(n_{i,j})$ is defined as the occupancy of virtual buffer after encoding each basic unit. At this stage, $B_S(n_{i,j})$ are updated by eq.(5.14).

$$B_S(n_{i,j}) = B_c(n_{i,j}) + \sum_{k=1}^l b(n_{i,j-k}) \quad (5.14)$$

, where $b(n_{i,j-k})$ is the actual number of bits use in the l^{th} basic unit. In H.264 standard, the possible quantization parameter is specified between range 0 - 51.

2) Post-encoding stage

After encoding each basic unit, only Cauchy rate and distortion model parameters, a_{j-l} , α_{j-l} and b_{j-l} , β_{j-l} , are updated by linear regression analysis in eqs. (3.36)-(3.37) and eqs. (3.39)-(3.40), as shown in Chapter 3 to find the optimal quantization step size of each basic unit.

5.2.4 Simulation results

For our simulations, we compare the performance between our proposed rate control scheme by Cauchy rate-distortion optimization model with H.264 JM 8.6 rate control and rate control algorithm proposed by *N.Kamaci* and *Y.Altunbasak* [25], in terms of the average PSNR, PSNR standard deviation, the number of frames skipped, bit rate used, and processing times. The parameters of video coding and tested video sequences are as follows.

- 1) We encode six video sequences covering every aspect of characteristics as in chapter 4.
- 2) Input video sequences are in the QCIF format with a resolution of 176x144 pixels.
- 3) Structure of encoded low-delay video sequence consists of only one I-frame and the rest are P-frames.
- 4) The target bit rate is varied from 16-256 kbps. The frame rate is set at 10 fps.
- 5) Four maximum buffer size (B_s) is used in our simulations :
 - Maximum buffer size $B_s = u(n_{i,j})/F_r$ represents a delay time of 100 ms.
 - Maximum buffer size $B_s = 1.5 \times u(n_{i,j})/(F_r)$ represents a delay time of 150 ms.

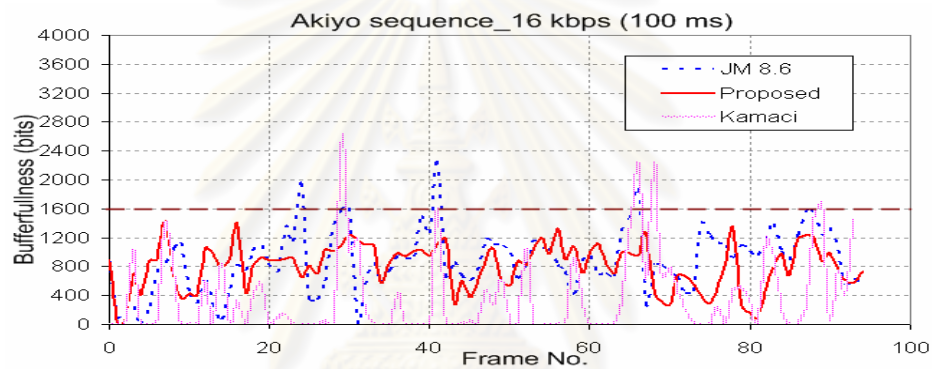
- Maximum buffer size $B_s = 2.0 \times u(n_{i,j}) / (F_r)$ represents a delay time of 200 ms.
 - Maximum buffer size $B_s = 4.0 \times u(n_{i,j}) / (F_r)$ represents a delay time of 400 ms.
- 6) H.264 reference software version JM 8.6 with main profile is used in the simulation for performance comparison purpose.

Table 5.1-5.5 shows the average PSNR, standard deviation of PSNR, number of frames skipped, bit rate, processing time, PSNR gain, and percentage of frames skipped reduction between H.264 JM 8.6 rate control, rate control algorithm in [25] and our proposed scheme encoded at bit rate of 16 , 32, 64, 128 and 256 kbps with delay constraint of 100 ms, respectively. On average the proposed scheme can achieve average PSNR improvement up to 0.43 dB, 0.66 dB, 0.50 dB, 0.44 dB and 0.70 dB, respectively for the bit rate of 16, 32, 64, 128 and 256 kbps compared to H.264 JM 8.6. The proposed scheme can achieve average PSNR improvement up to 1.13 dB, 2.48 dB, 0.75 dB, 1.42 dB and 1.72 dB, respectively for the bit rate of 16, 32, 64, 128 and 256 kbps compared to rate control in [25]. On average, our proposed scheme can encode video with more uniform quality, as can be seen from lower standard deviation of PSNR compared with the H.264 JM 8.6 and rate control algorithm in [25]. Our proposed scheme can also encode video with better motion continuity, as can be seen from the reduction of the numbers of frames skipped for up to 100%. In term of bit rate used, our proposed scheme which is based on Cauchy model can achieve more accurate bit rate, i.e., closer to target bit rate, than that of JM 8.6 and rate control in [25]. Our proposed scheme also achieves lower processing time on average compared to JM 8.6.

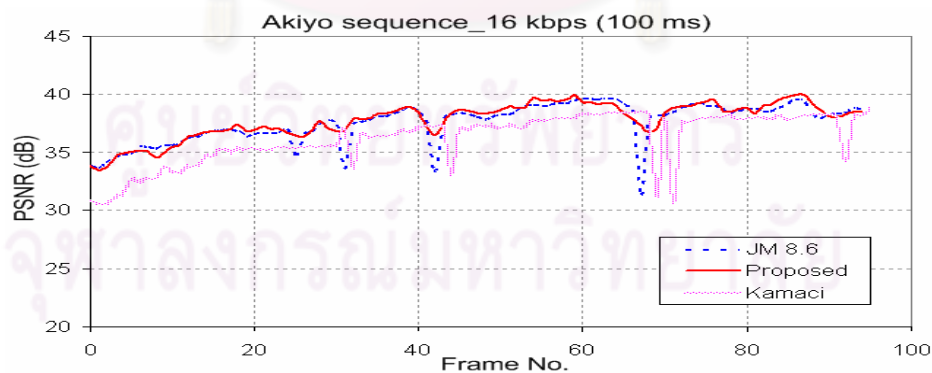
Table 5.6-5.8 shows the average PSNR, standard deviation of PSNR, and number of frames skipped, bit rate, processing time, PSNR gain, and percentage of frames skipped reduction between H.264 JM 8.6 rate control, rate control algorithm in [25] and our proposed scheme encoded at the bit rate of 16 kbps with delay constraints of 150, 200 and 400 ms, respectively. It can be shown that our proposed scheme also outperforms JM 8.6 and rate control in [25] in every aspect.

Simulation results with delay constraint of 100 ms.

- **Akiyo sequence** : Number of frames to be encoded is 100. The simulation results as shown in Fig.5.4. Fig.5.4 (a) showed that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.4 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.21 dB and 1.57 dB compared to the H.264 JM 8.6 and rate control in [25], respectively.



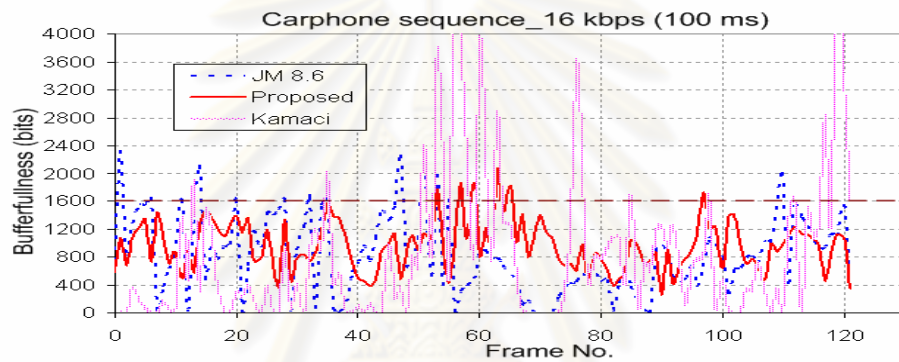
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) .



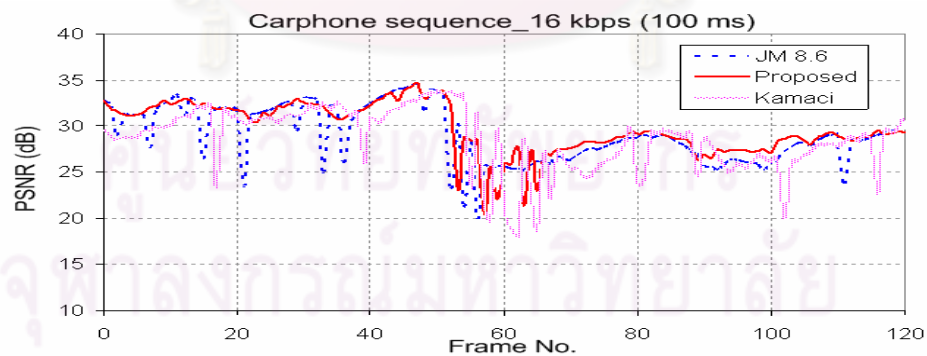
(b) Average PSNR of each frame with delay constraint.

Figure 5.4 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay of 100 ms.

- **Carphone sequence** : Number of frames to be encoded is 127. The simulation results as shown in Fig5.5. Fig.5.5 (a) showed that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.5 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.69 dB and 1.28 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



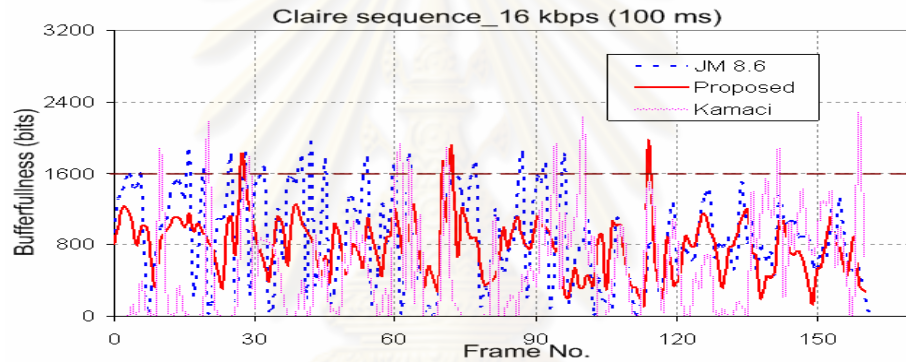
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) .



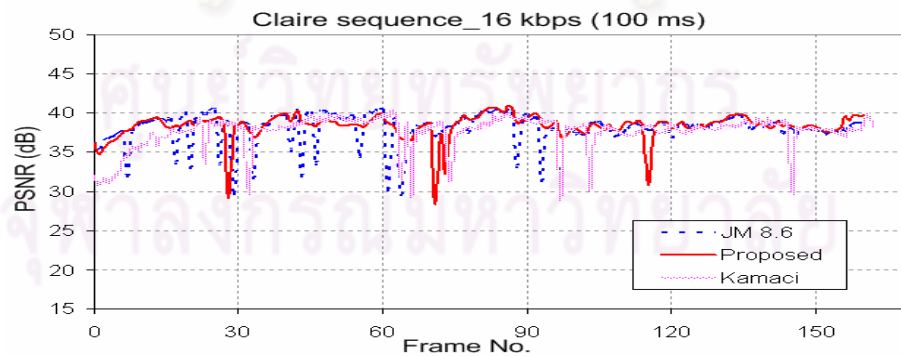
(b) Average PSNR of each frame with delay constraint.

Figure 5.5 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay of 100 ms.

- **Claire sequence** : Number of frames to be encoded is 164. The simulation results as shown in Fig.5.6. Fig.5.6 (a) show that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.6 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.35 dB and 0.79 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



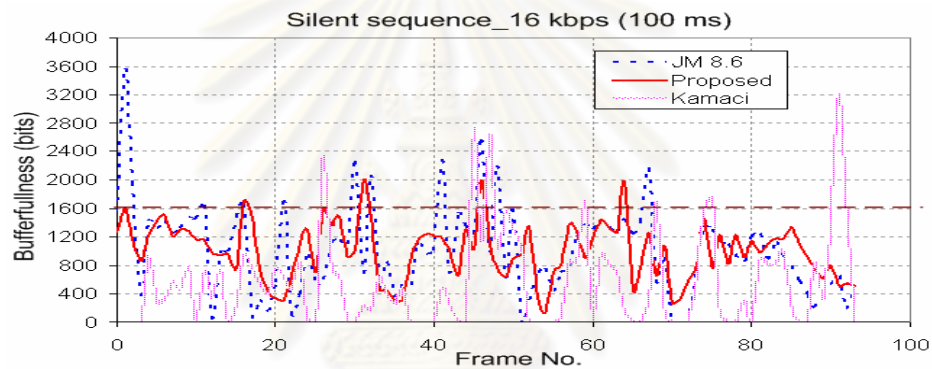
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line).



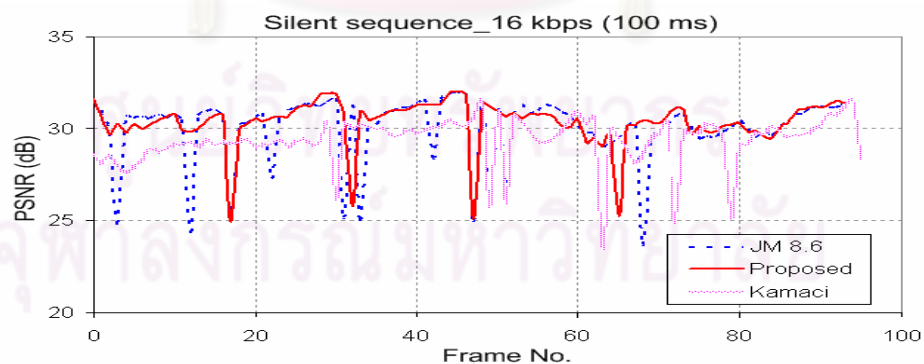
(c) Average PSNR of each frame with delay of constraint.

Figure.5.6 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay of 100 ms.

- **Silent sequence** : Number of frames to be encoded is 99. The simulation results as shown in Fig.5.7. Fig.5.7 (a) show that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.7 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.34 dB and 1.04 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



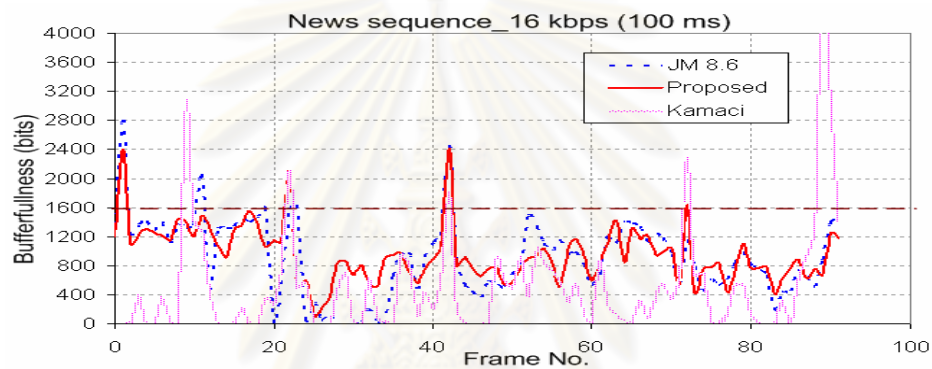
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line).



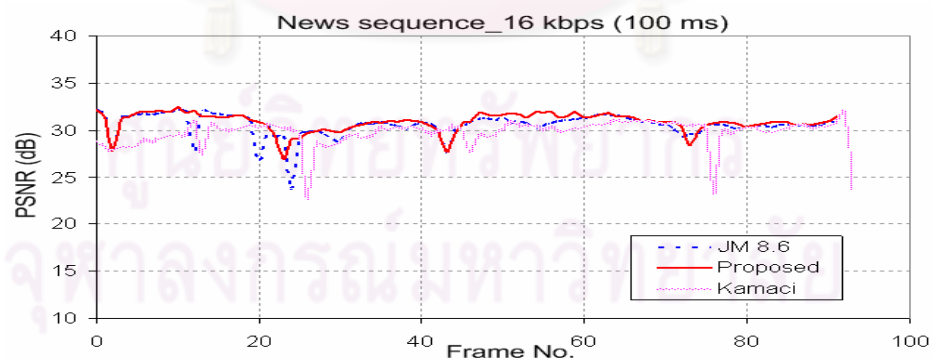
(b) Average PSNR of each frame with delay constraint.

Figure.5.7 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay of 100 ms.

- **News sequence** : Number of frames to be encoded is 99. The simulation results as shown in Fig.5.8. Fig.5.8 (a) show that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.8 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.34 dB and 1.17 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



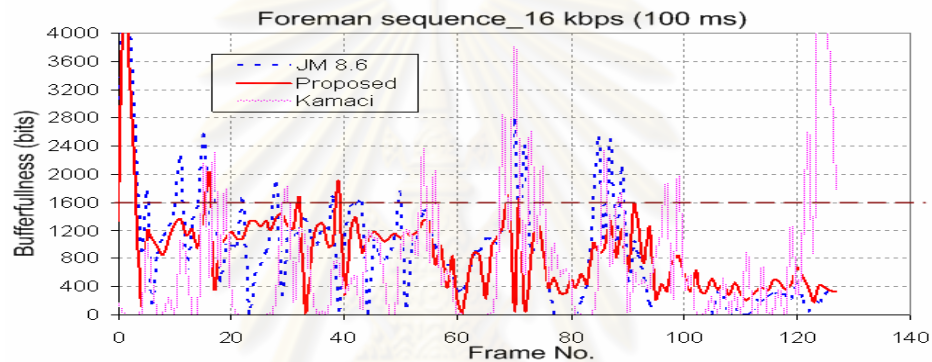
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line).



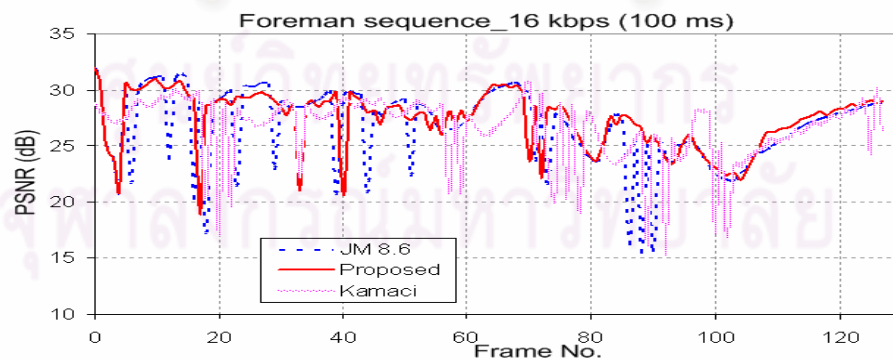
(b) Average PSNR of each frame with delay constraint.

Figure.5.8 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay of 100 ms.

- **Foreman sequence** : Number of frames to be encoded is 133. The simulation results as shown in Fig.5.9. Fig.5.9 (a) show that the buffer fullness level in bits of each frame with maximum delay constraint 100 ms (maximum buffer size is 1600 bits) between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.9 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.67 dB and 0.94 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line).



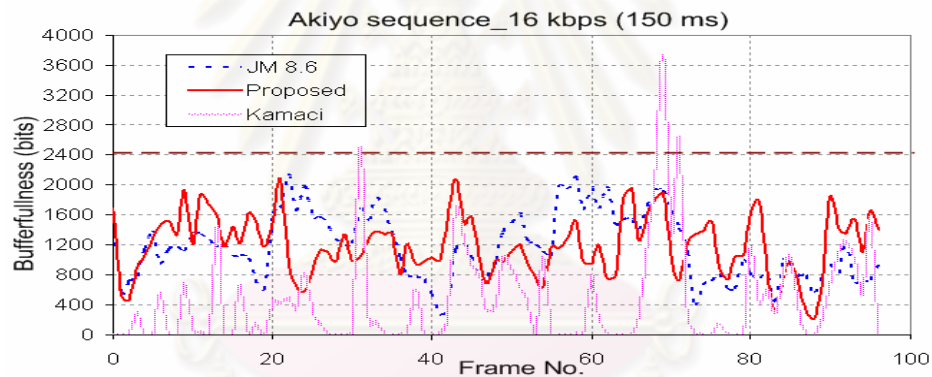
(b) Average PSNR of each frame with delay constraint.

Figure.5.9 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay of 100 ms.

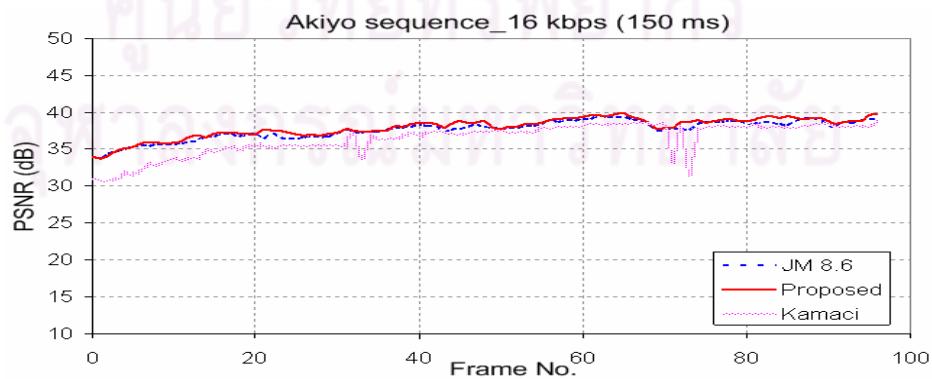
Simulation results with delay constraint of 150 ms.

This simulation results shows the effective of low delay constraint at delay time 150 ms of six video test sequences : Akiyo, Carphone, Claire, Foreman, News and Silent sequence. Target bit rate use in this simulation is 16 kbps. Maximum buffer size is set to 2400 bits represent delay time 150 ms.

Fig.5.10 show the simulation result of Akiyo test sequence. In Fig.5.10 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.10 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.31 dB and 1.58 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



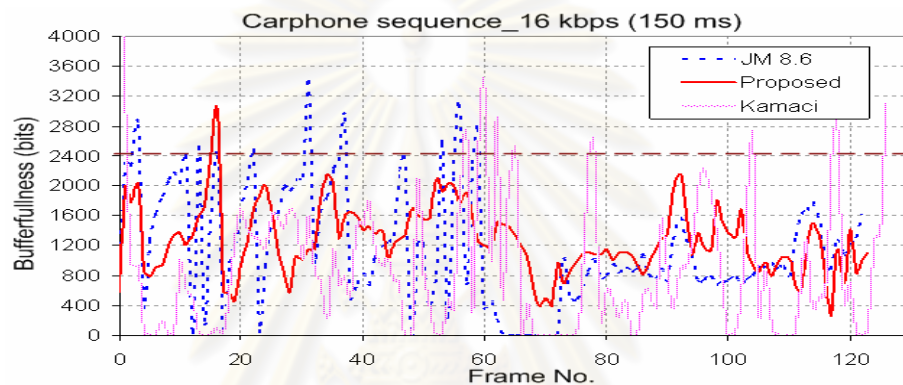
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Akiyo sequence.



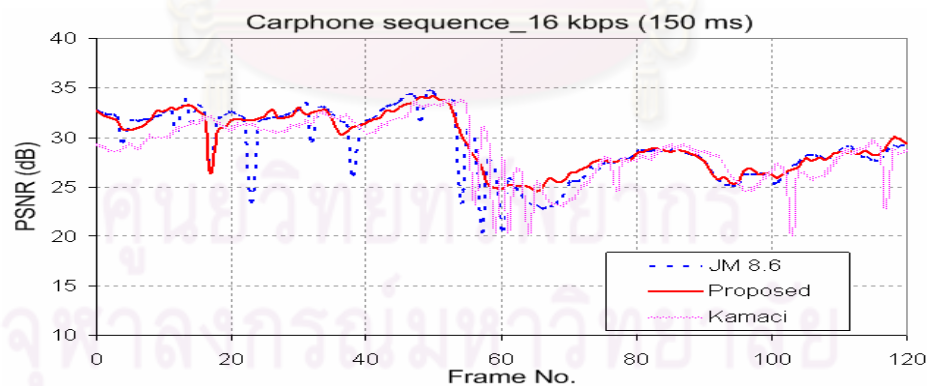
(b) Average PSNR of each frame with delay constraint of Akiyo sequence.

Figure 5.10 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay of 150 ms.

Fig.5.11 show the simulation result of Carphone test sequence. In Fig.5.11 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.11 (b) shown the average PSNR per frame with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.34 dB and 0.74 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



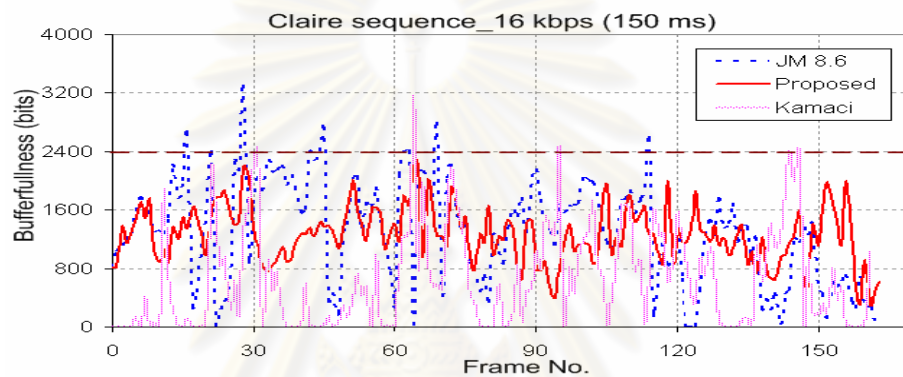
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Carphone sequence.



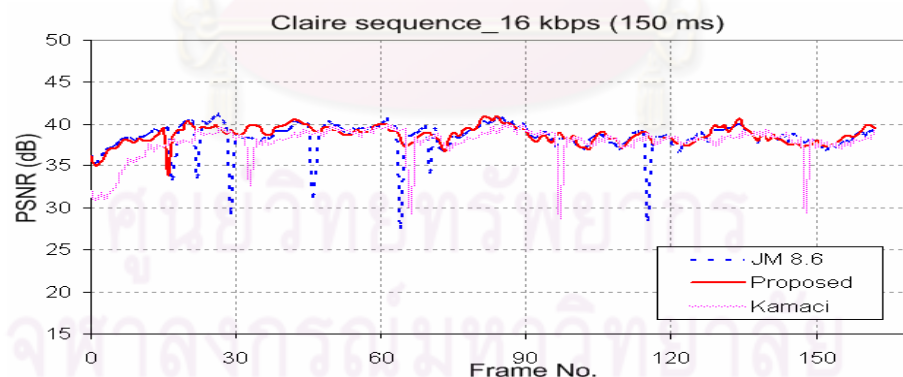
(b) Average PSNR of each frame with delay constraint of Carphone sequence.

Figure 5.11 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay of 150 ms.

Fig.5.12 show the simulation result of Claire test sequence. Fig.5.12 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.12(b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.30 dB and 0.88 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



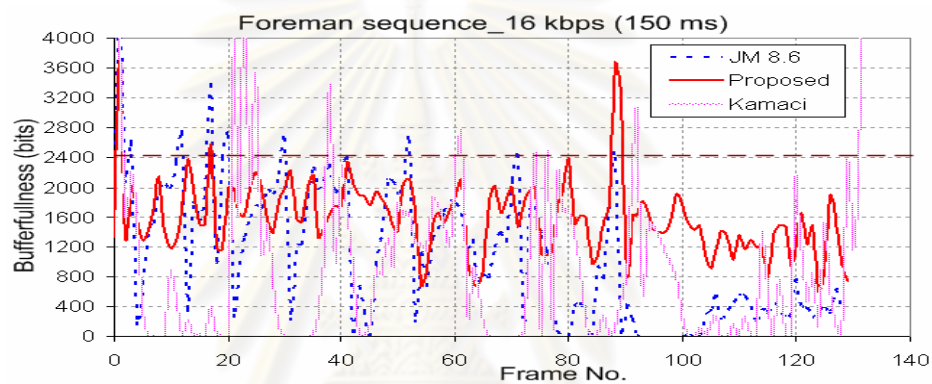
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Claire sequence.



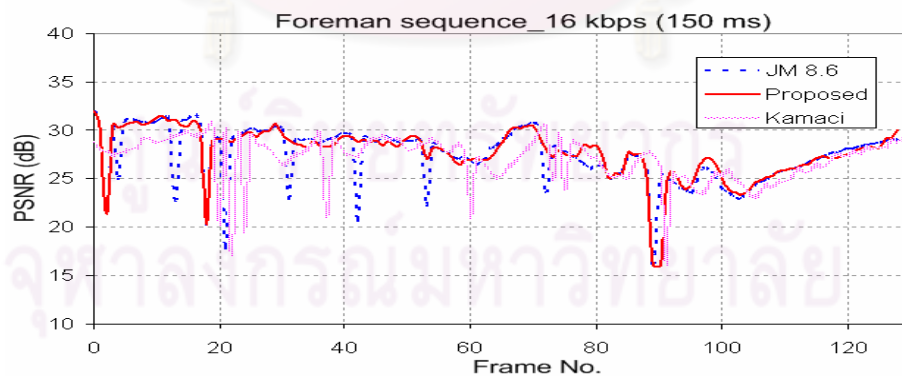
(b) Average PSNR of each frame with delay constraint of Claire sequence.

Figure 5.12 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay of 150 ms.

Fig.5.13 show the simulation result of Foreman test sequence. In Fig.5.13 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.13 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.31 dB and 0.78 dB to the H.264 JM 8.6 and rate control in [25], respectively.



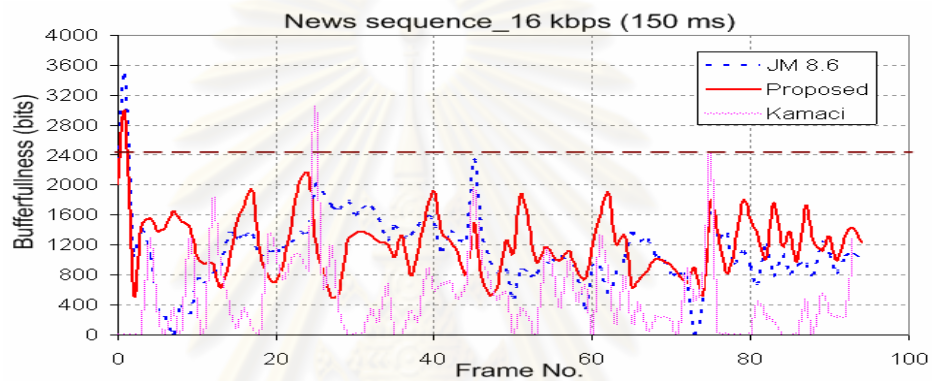
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Foreman sequence.



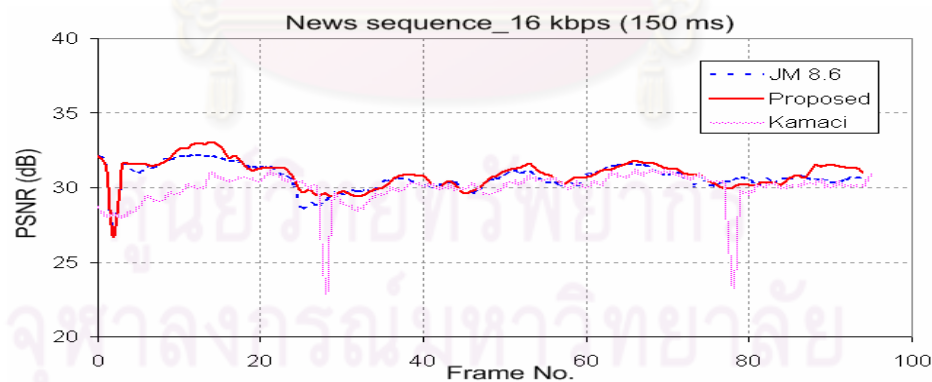
(b) Average PSNR of each frame with delay constraint of Foreman sequence.

Figure 5.13 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay of 150 ms.

Fig.5.14 show the simulation result of News test sequence. Fig.5.14 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.14 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.23 dB and 0.99 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



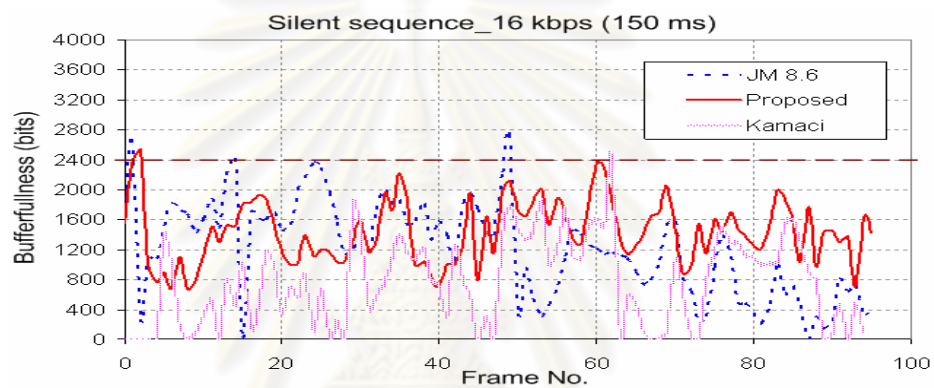
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of News sequence.



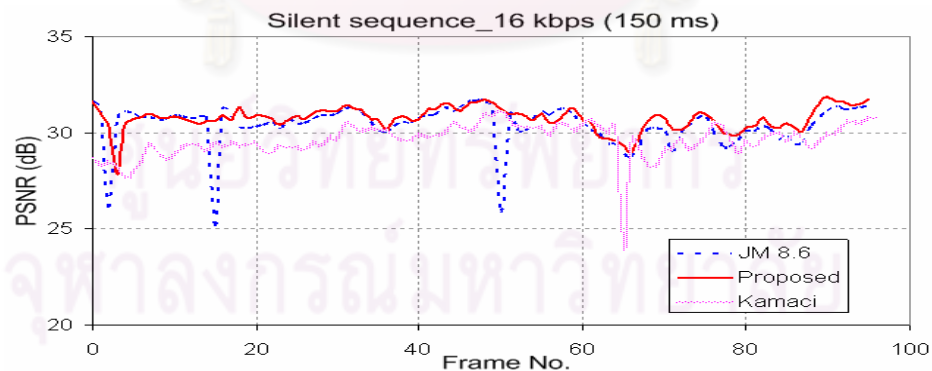
(b) Average PSNR of each frame with delay constraint of News sequence.

Figure 5.14 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay of 150 ms.

Fig.5.15 show the simulation result of Silent test sequence. In Fig.5.15 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 150 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.15 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.35 dB and 1.14 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Silent sequence.



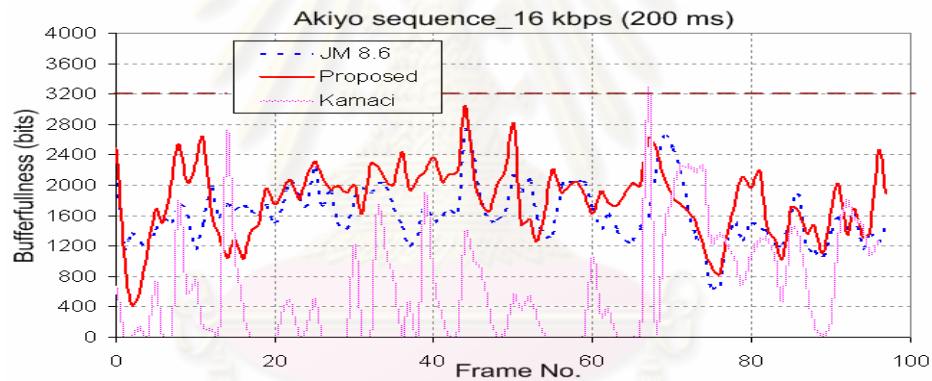
(b) Average PSNR of each frame with delay constraint of Silent sequence.

Figure 5.15 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay of 150 ms.

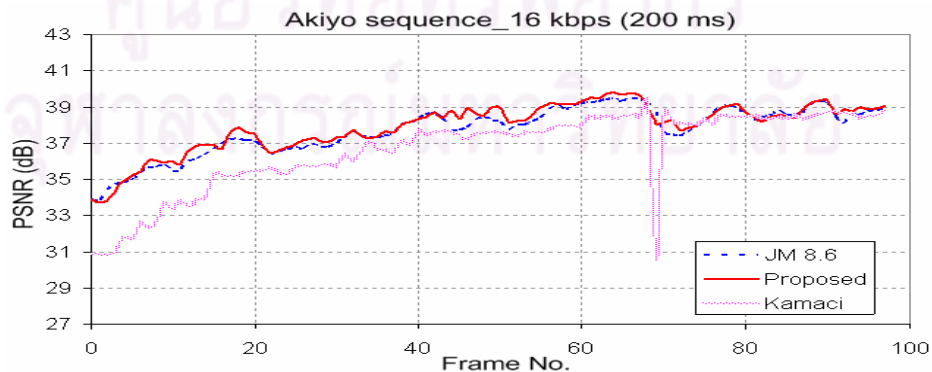
Simulation results with delay constraint of 200 ms.

This simulation results shows the effective of low delay constraint at delay time 200 ms of six video test sequences : Akiyo, Carphone, Claire, Foreman, News and Silent sequence. Target bit rate use in this simulation is 16 kbps. Maximum buffer size is set to 3200 bits represent delay time 200 ms.

Fig.5.16 show the simulation result of Akiyo test sequence. In Fig.5.16 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.16 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.20 dB and 1.27 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



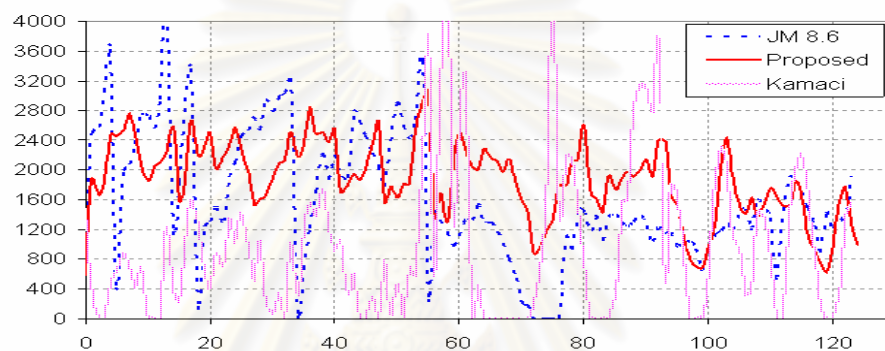
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Akiyo sequence.



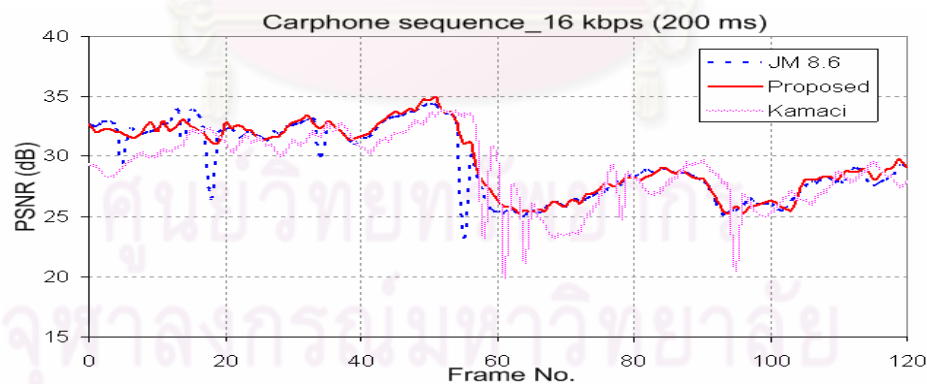
(b) Average PSNR of each frame with delay constraint of Akiyo sequence.

Figure 5.16 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay of 200 ms.

Fig.5.17 show the simulation result of Carphone test sequence. In Fig.5.17 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.17 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.26 dB and 0.83 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



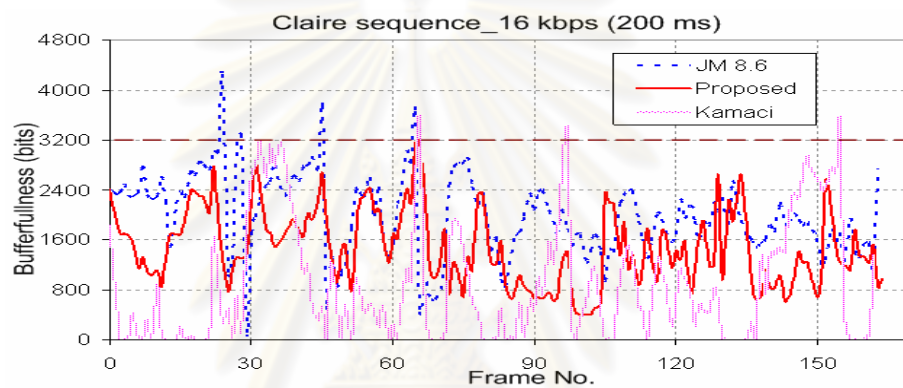
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Carphone sequence.



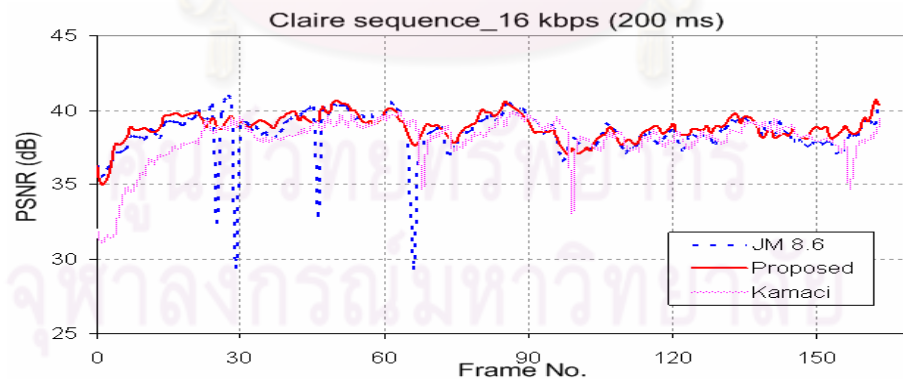
(b) Average PSNR of each frame with delay constraint of Carphone sequence.

Figure 5.17 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay of 200 ms.

Fig.5.18 show the simulation result of Claire test sequence. In Fig.5.18 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.18(b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.40 dB and 0.88 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



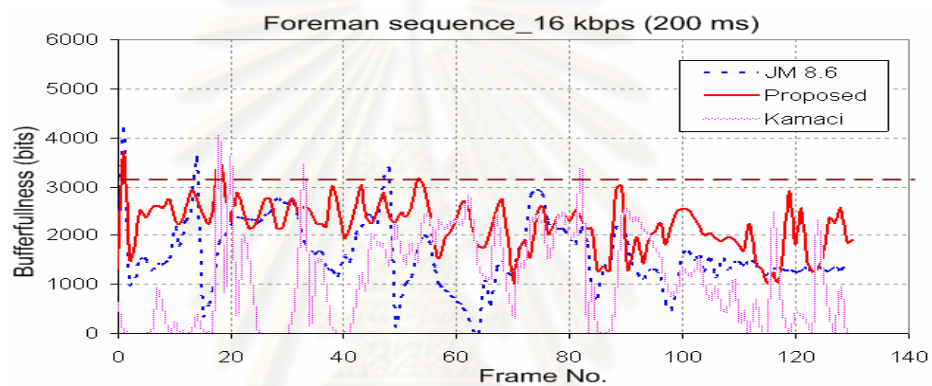
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Claire sequence.



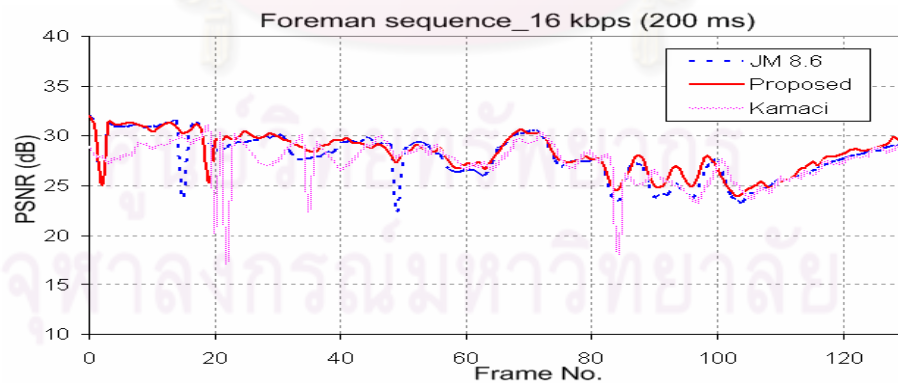
(b) Average PSNR of each frame with delay constraint of Claire sequence.

Figure 5.18 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay of 200 ms.

Fig.5.19 show the simulation result of Foreman test sequence. In Fig. 5.19 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.19 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.41 dB and 0.91 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



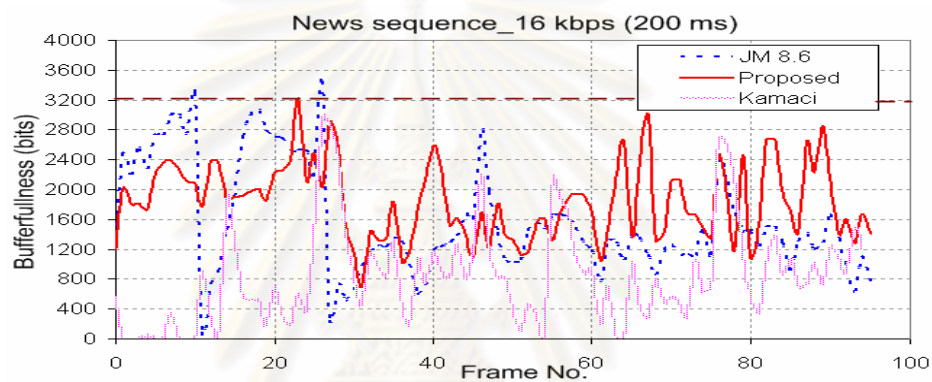
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Foreman sequence.



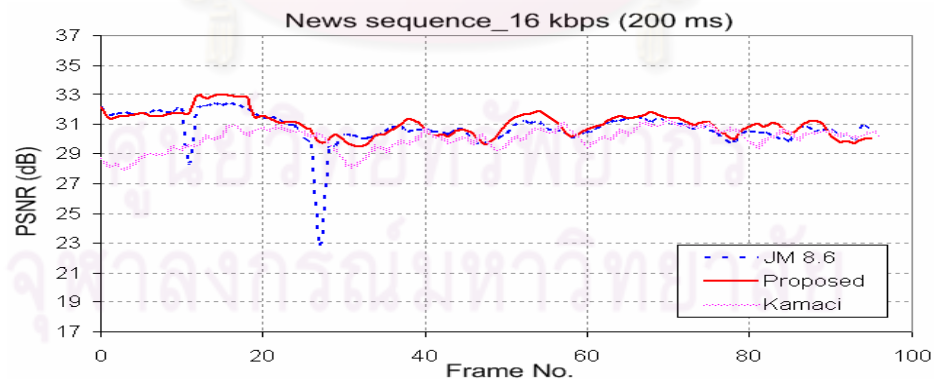
(b) Average PSNR of each frame with delay constraint of Foreman sequence.

Figure 5.19 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay of 200 ms.

Fig.5.20 show the simulation result of News test sequence. In Fig.5.20 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.20 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.30 dB and 1.04 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



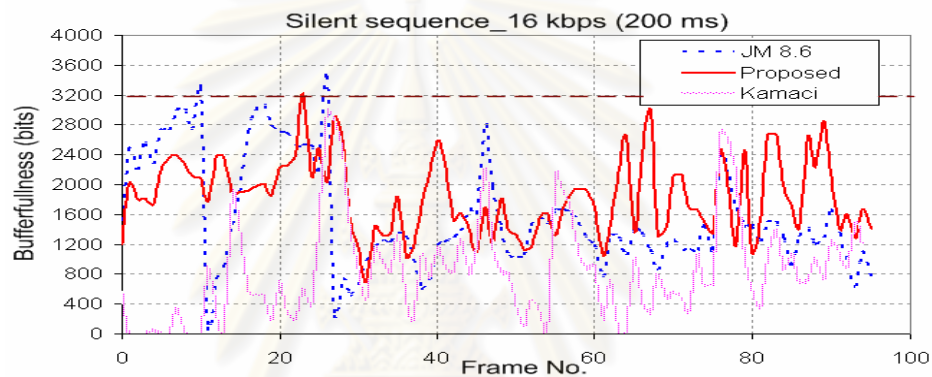
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of News sequence.



(b) Average PSNR of each frame with delay constraint of News sequence.

Figure 5.20 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay of 200 ms.

Fig.5.21 show the simulation result of Silent test sequence. In Fig.5.21 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 200 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.21 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.41 dB and 1.25 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Silent sequence.



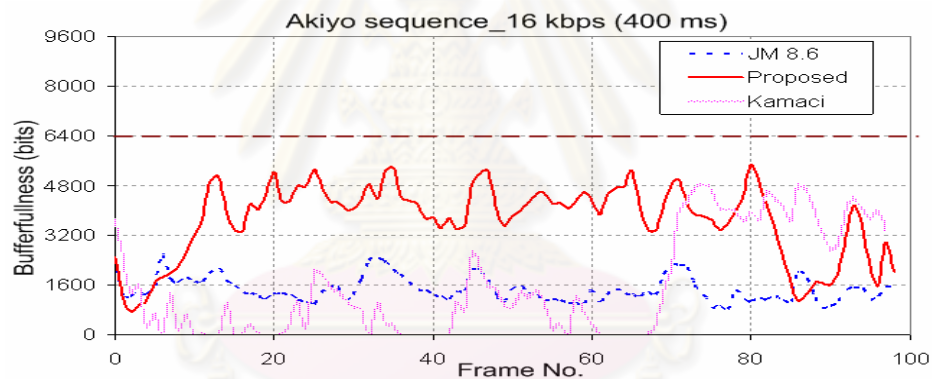
(b) Average PSNR of each frame with delay constraint of Silent sequence.

Figure 5.21 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay of 200 ms.

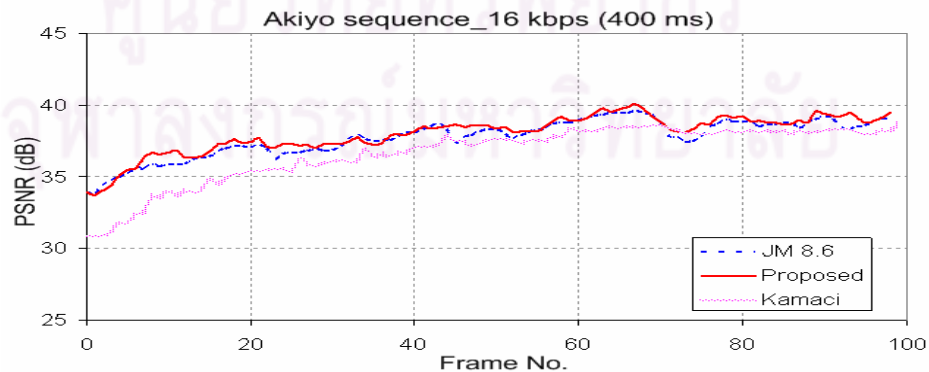
Simulation results with delay constraint of 400 ms.

This simulation results shows the effective of low delay constraint at delay time 400 ms of six video test sequences : Akiyo, Carphone, Claire, Foreman, News and Silent sequence. Target bit rate use in this simulation is 16 kbps. Maximum buffer size is set to 6400 bits represent delay time 400 ms.

Fig.5.22 show the simulation result of Akiyo test sequence. In Fig.5.22 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.22 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.30 dB and 1.38 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



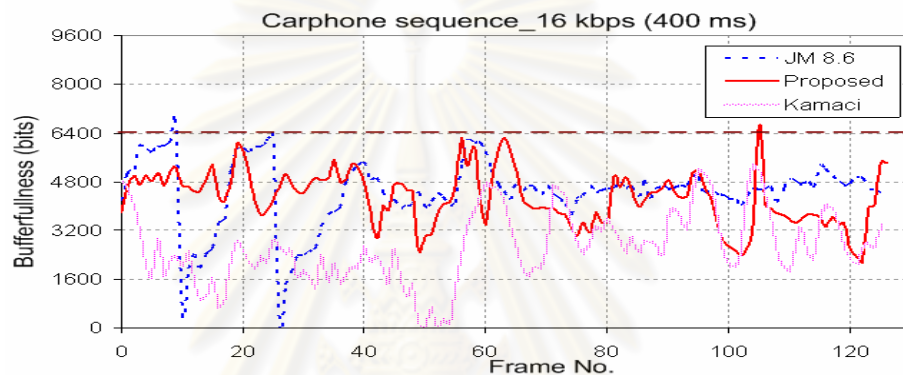
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Akiyo sequence.



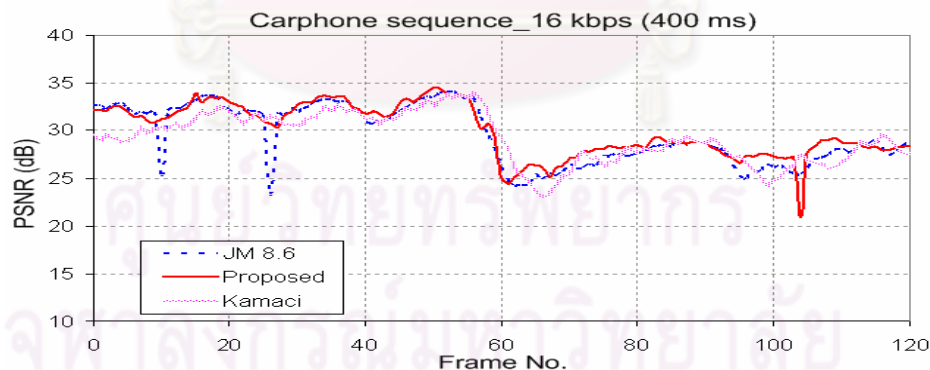
(b) Average PSNR of each frame with delay constraint of Akiyo sequence.

Figure 5.22 Simulation results coded by H.264 JM8.6 rate control and our proposed for Akiyo sequence at 16 kbps as target bit rate with time delay of 400ms.

Fig.5.23 show the simulation result of Carphone test sequence. In Fig.5.23 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.23 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.47 dB and 0.79 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



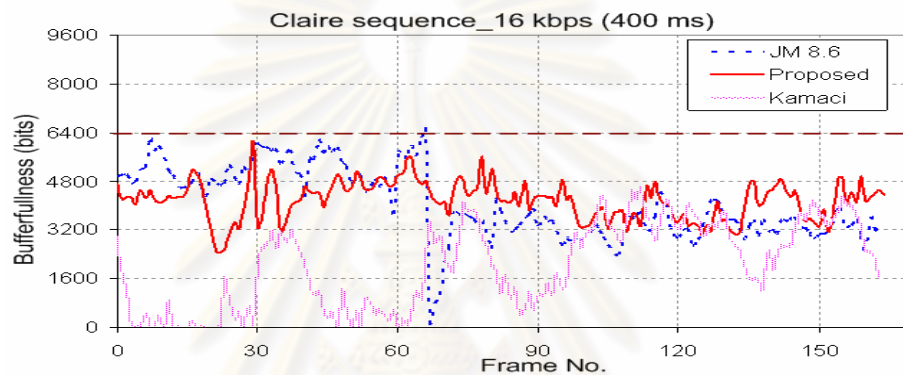
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Carphone sequence.



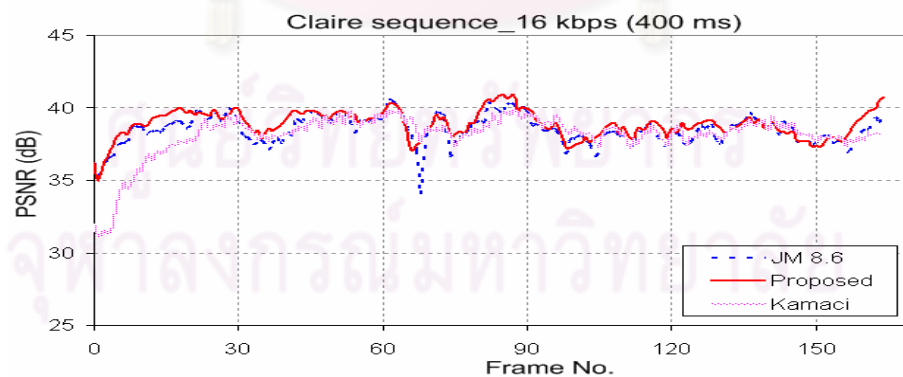
(b) Average PSNR of each frame with delay constraint of Carphone sequence.

Figure 5.23 Simulation results coded by H.264 JM8.6 rate control and our proposed for Carphone sequence at 16 kbps as target bit rate with time delay of 400 ms.

Fig.5.24 show the simulation result of Claire test sequence. In Fig.5.24 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.24 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.41 dB and 0.78 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



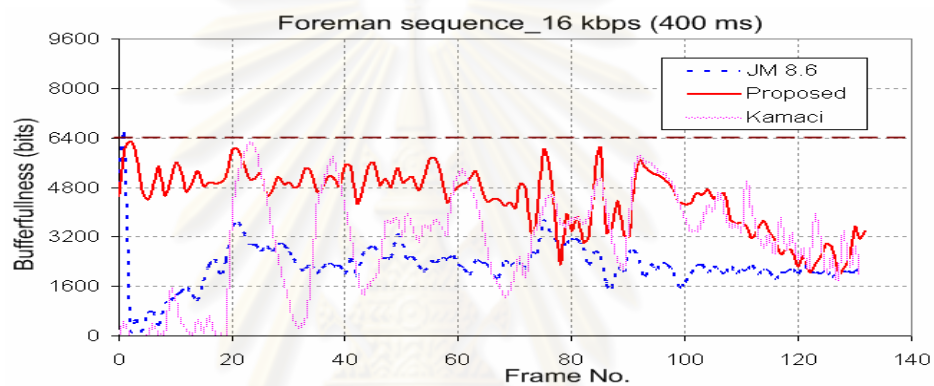
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Claire sequence.



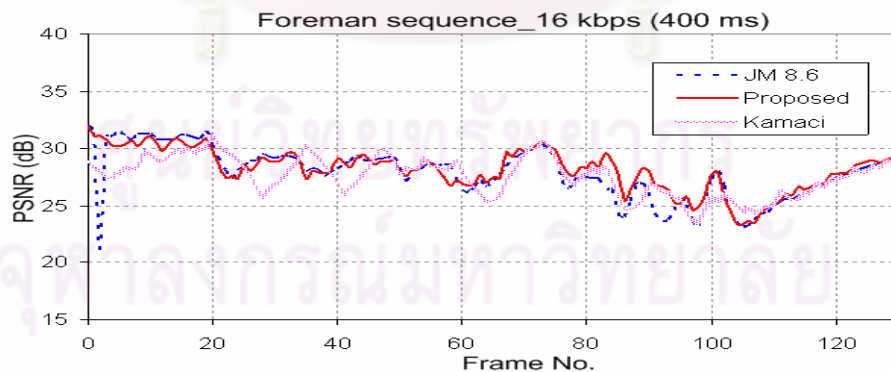
(b) Average PSNR of each frame with delay constraint of Claire sequence.

Figure 5.24 Simulation results coded by H.264 JM8.6 rate control and our proposed for Claire sequence at 16 kbps as target bit rate with time delay of 400 ms.

Fig.5.25 show the simulation result of Foreman test sequence. In Fig.5.25 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.25 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.31 dB and 0.62 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



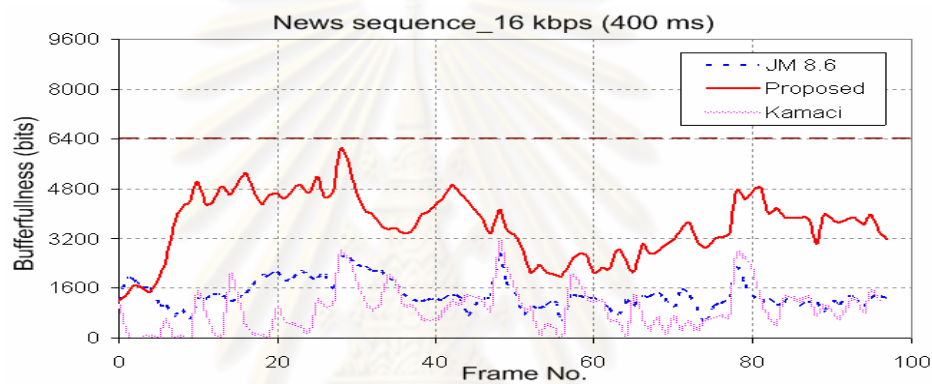
(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Foreman sequence.



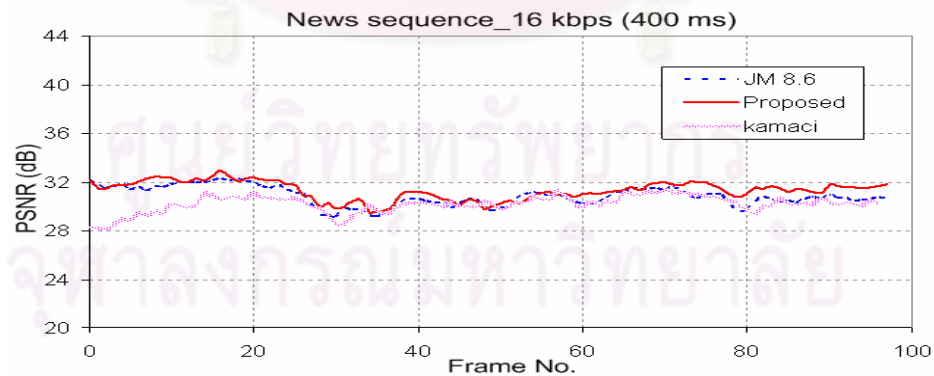
(b) Average PSNR of each frame with delay constraint of Foreman sequence.

Figure 5.25 Simulation results coded by H.264 JM8.6 rate control and our proposed for Foreman sequence at 16 kbps as target bit rate with time delay of 400 ms.

Fig.5.26 show the simulation result of News test sequence. In Fig.5.26 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.26 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.48 dB and 1.45 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of News sequence.



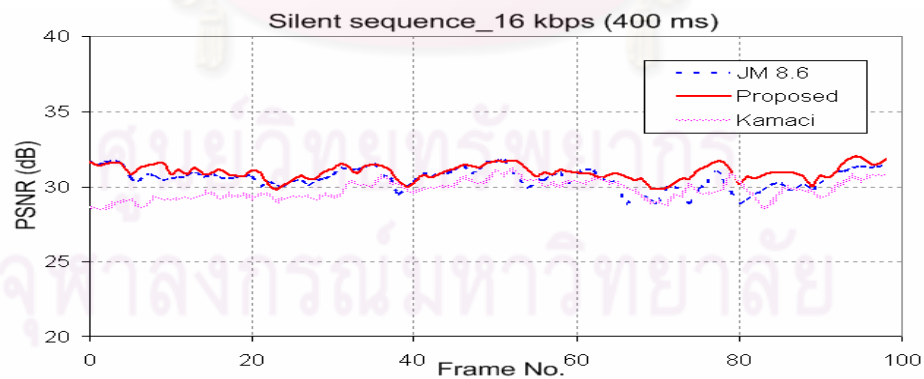
(b) Average PSNR of each frame with delay constraint of News sequence.

Figure 5.26 Simulation results coded by H.264 JM8.6 rate control and our proposed for News sequence at 16 kbps as target bit rate with time delay of 400 ms.

Fig.5.27 show the simulation result of Silent test sequence. In Fig.5.27 (a) show the buffer fullness level in bits of each frame with maximum delay constraint 400 ms between our proposed rate control, H.264 JM8.6 rate control and rate control in [25]. Fig.5.27 (b) shown the average PSNR with delay constraint. From the experimental result, it can be shown that the proposed scheme can achieve average PSNR improvement up to 0.45 dB and 1.22 dB compared to H.264 JM 8.6 and rate control in [25], respectively.



(a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) of Silent sequence.



(b) Average PSNR of each frame with delay constraint of Silent sequence.

Figure 5.27 Simulation results coded by H.264 JM8.6 rate control and our proposed for Silent sequence at 16 kbps as target bit rate with time delay of 400 ms.

Fig.5.28, we illustrate the effect of frame skipping by showing three successive coded video frames of Akiyo sequence at 32kbps under delay time 100 ms. When frame skipping, It can be shown that it is not consecutive in successive frames for H.264 JM 8.6 when buffer fullness was exceed than buffer size but our proposed did not skip this frame with smooth visual video quality.



Figure 5.28 (a1)-(c1) show frames 40,41,42, respectively, of test sequence “Akiyo” coded with our proposed rate control scheme at 32 kbps , (a2)-(c2) are the same frames but coded with H.264 JM 8.6 rate control. Frame (b2) are identical with frame (a2) , because JM 8.6 rate control skipped one frame after coding frame (a2) since encoder buffer fullness was higher than maximum buffer size and (a2) was repeated at the encoder side to generated next frame.

Fig.5.29, we illustrate the effect of frame skipping by showing three successive coded video frames of Claire sequence at 32kbps under delay time 100 ms. When frame skipping, It can be shown that it is not consecutive in successive frames for H.264 JM 8.6 when buffer fullness was exceed than buffer size but our proposed did not skip this frame with smooth visual video quality.

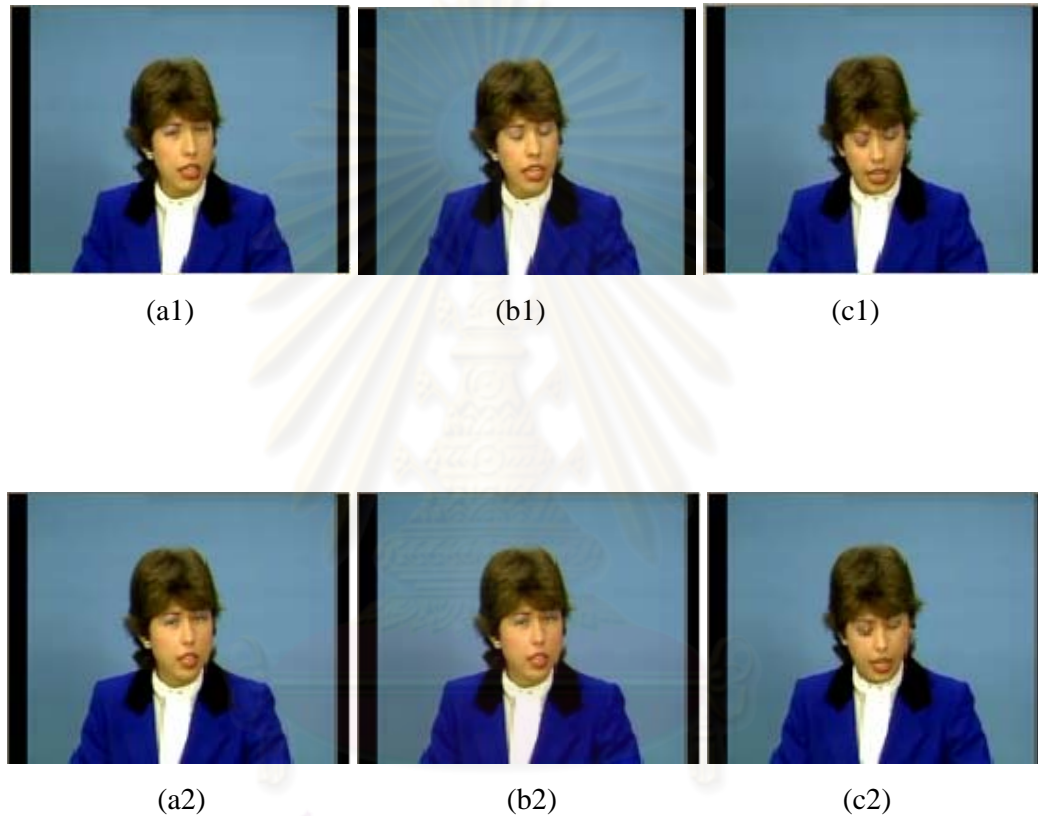


Figure 5.29 (a1)-(c1) show frames 59,60,61, respectively, of test sequence “Claire” coded with our proposed rate control scheme at 32 kbps , (a2)-(c2) are the same frames but coded with. H.264 JM 8.6 rate control. Frame (b2) are identical with frame (a2) , because JM 8.6 rate control skipped one frame after coding frame (a2) since encoder buffer fullness was higher than maximum buffer size and (a2) was repeated at the encoder side to generated next frame.

Table 5.1: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 16 kbps for delay time 100 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	37.59	1.69	4	16.40	271.02
	Kamaci	36.23	2.22	5	14.30	216.40
	Proposed	37.80	1.53	0	16.04	227.37
Carphone	JM 8.6	29.01	3.07	13	16.02	219.74
	Kamaci	28.42	3.49	19	15.07	210.62
	Proposed	29.70	2.71	5	16.01	220.45
Claire	JM 8.6	38.02	2.24	16	15.98	218.89
	Kamaci	37.58	2.26	10	14.77	219.99
	Proposed	38.37	1.65	4	16.04	216.43
Foreman	JM 8.6	26.73	3.45	17	15.53	252.00
	Kamaci	26.46	3.12	19	14.81	237.04
	Proposed	27.40	2.55	9	16.03	240.30
News	JM 8.6	30.52	1.22	5	15.97	288.47
	Kamaci	29.69	1.44	6	15.29	218.77
	Proposed	30.86	0.98	4	16.06	222.68
Silent	JM 8.6	30.05	1.82	12	16.06	239.15
	Kamaci	29.35	1.33	8	14.88	209.62
	Proposed	30.39	1.25	5	16.05	219.51

Table 5.2: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 32 kbps for delay time 100 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	41.73	1.81	5	32.00	218.92
	Kamaci	41.70	2.40	6	28.77	201.13
	Proposed	42.08	1.37	2	32.09	222.70
Carphone	JM 8.6	31.54	3.98	10	31.91	220.50
	Kamaci	29.24	4.49	27	26.72	216.73
	Proposed	32.20	3.65	6	32.25	209.87
Claire	JM 8.6	41.86	2.01	4	31.90	219.35
	Kamaci	41.43	2.55	10	29.72	206.75
	Proposed	42.32	1.36	1	32.07	215.04
Foreman	JM 8.6	30.14	3.88	11	30.70	226.85
	Kamaci	25.83	5.87	32	24.92	212.01
	Proposed	31.04	3.39	6	31.04	227.46
News	JM 8.6	34.89	2.41	5	31.87	213.38
	Kamaci	32.72	4.72	17	27.57	215.73
	Proposed	35.29	2.01	3	32.28	214.75
Silent	JM 8.6	33.89	2.95	11	31.32	222.93
	Kamaci	32.20	4.60	16	26.58	216.40
	Proposed	35.07	1.66	2	32.04	220.48

Table 5.3: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 64 kbps for delay time 100 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	45.23	1.36	0	63.95	228.78
	Kamaci	44.63	1.60	1	56.84	223.46
	Proposed	45.53	1.44	0	63.52	222.31
Carphone	JM 8.6	35.11	3.79	7	63.40	221.85
	Kamaci	34.67	4.54	12	61.84	216.07
	Proposed	35.69	3.13	3	64.16	223.10
Claire	JM 8.6	45.37	1.58	1	63.46	219.37
	Kamaci	44.75	2.63	7	59.30	206.01
	Proposed	45.81	1.18	0	63.69	218.53
Foreman	JM 8.6	33.84	3.09	7	63.95	231.13
	Kamaci	33.13	4.43	12	62.82	217.20
	Proposed	34.55	1.89	1	64.09	225.69
News	JM 8.6	38.77	1.22	1	63.70	249.58
	Kamaci	38.73	2.98	4	61.74	212.05
	Proposed	39.24	0.83	0	63.73	230.12
Silent	JM 8.6	36.96	3.34	9	64.09	223.89
	Kamaci	37.83	2.95	5	60.83	214.03
	Proposed	37.39	2.47	4	64.14	221.53

Table 5.4: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 128 kbps for delay time 100 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	48.99	1.25	0	127.66	225.53
	Kamaci	47.53	1.94	1	110.68	225.68
	Proposed	49.27	1.26	0	127.35	212.12
Carphone	JM 8.6	38.97	3.82	3	126.53	212.21
	Kamaci	38.41	4.79	7	121.16	206.79
	Proposed	39.34	3.16	1	128.90	210.45
Claire	JM 8.6	48.24	2.47	4	124.24	220.16
	Kamaci	48.06	2.58	4	119.06	217.69
	Proposed	48.64	0.79	0	127.84	222.63
Foreman	JM 8.6	37.50	1.95	1	127.57	222.47
	Kamaci	37.03	3.51	5	123.46	219.26
	Proposed	37.88	1.66	0	127.66	209.83
News	JM 8.6	44.94	1.73	1	127.40	237.98
	Kamaci	42.54	3.47	3	120.24	225.27
	Proposed	45.50	0.68	0	128.01	236.49
Silent	JM 8.6	42.08	2.48	2	127.92	213.40
	Kamaci	41.26	2.81	3	119.62	202.13
	Proposed	42.73	0.82	0	128.70	215.02

Table 5.5: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 256 kbps for delay time 100 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	52.92	2.28	2	253.45	213.00
	Kamaci	50.90	2.01	0	212.21	214.45
	Proposed	53.72	1.09	0	256.26	214.14
Carphone	JM 8.6	42.82	4.78	5	253.50	222.10
	Kamaci	42.91	4.58	5	216.46	217.00
	Proposed	43.61	3.47	2	257.99	219.54
Claire	JM 8.6	50.41	3.49	6	245.20	222.60
	Kamaci	50.58	1.06	0	239.69	211.37
	Proposed	51.13	0.47	0	253.59	220.77
Foreman	JM 8.6	41.07	3.33	3	254.92	212.83
	Kamaci	39.82	4.12	5	248.62	205.85
	Proposed	41.86	2.50	1	255.91	218.67
News	JM 8.6	48.67	2.13	1	255.86	236.46
	Kamaci	47.42	2.78	1	236.18	222.75
	Proposed	49.29	2.17	1	255.15	230.41
Silent	JM 8.6	47.58	2.62	1	256.33	221.62
	Kamaci	45.75	3.94	3	234.08	218.16
	Proposed	48.06	1.11	0	256.45	215.63

Table 5.6: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 16 kbps for delay time 150 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	37.62	1.36	0	16.35	267.72
	Kamaci	36.35	2.16	3	14.69	224.77
	Proposed	37.93	1.29	0	16.18	242.04
Carphone	JM 8.6	29.10	3.29	11	16.57	209.07
	Kamaci	28.70	3.14	8	15.77	208.58
	Proposed	29.44	2.74	1	16.26	210.75
Claire	JM 8.6	38.39	1.93	7	16.44	223.56
	Kamaci	37.81	1.97	4	15.26	218.84
	Proposed	38.69	1.08	1	16.52	216.86
Foreman	JM 8.6	27.47	2.87	10	16.60	279.03
	Kamaci	27.00	2.49	9	15.94	233.42
	Proposed	27.78	2.58	3	16.35	245.92
News	JM 8.6	30.68	0.87	1	16.59	279.87
	Kamaci	29.92	1.22	2	15.78	227.41
	Proposed	30.91	0.90	1	16.27	261.80
Silent	JM 8.6	30.40	1.06	3	16.59	230.68
	Kamaci	29.61	0.91	1	15.65	213.05
	Proposed	30.75	0.62	1	16.25	215.41

Table 5.7: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 16 kbps for delay time 200 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	37.69	1.34	0	16.49	261.19
	Kamaci	36.62	2.23	1	15.10	244.67
	Proposed	37.89	1.36	0	16.31	256.44
Carphone	JM 8.6	29.50	2.90	5	16.67	238.47
	Kamaci	28.93	3.14	5	15.88	226.41
	Proposed	29.76	2.80	0	16.37	242.85
Claire	JM 8.6	38.46	1.58	4	16.26	219.55
	Kamaci	37.98	1.60	3	15.50	213.72
	Proposed	38.86	0.90	0	15.99	220.84
Foreman	JM 8.6	27.84	2.25	3	16.79	288.65
	Kamaci	27.34	2.17	4	15.77	259.50
	Proposed	28.25	1.95	1	16.56	265.81
News	JM 8.6	30.75	1.12	2	16.58	217.04
	Kamaci	30.01	0.74	0	16.18	215.21
	Proposed	31.05	0.83	0	16.53	220.12
Silent	JM 8.6	30.49	1.00	3	16.87	245.90
	Kamaci	29.65	0.97	2	15.68	221.01
	Proposed	30.90	0.67	1	16.39	231.24

Table 5.8: Performance of proposed scheme compared with the H.264 JM8.6 rate control and rate control in [25] at 16 kbps for delay time 400 ms

Sequence	Rate Control	Average PSNR (dB)	PSNR Std.	No. frame skipped	Output bit rate (kbps)	Processing times (ms)
Akiyo	JM 8.6	37.72	1.31	0	16.66	229.63
	Kamaci	36.64	2.04	0	15.35	212.76
	Proposed	38.02	1.32	0	16.49	217.89
Carphone	JM 8.6	29.50	2.95	2	16.87	214.87
	Kamaci	29.18	2.53	0	16.28	219.31
	Proposed	29.97	2.68	1	16.66	215.56
Claire	JM 8.6	38.47	1.00	1	16.51	254.75
	Kamaci	38.10	1.51	0	15.65	230.75
	Proposed	38.88	0.95	0	16.29	241.08
Foreman	JM 8.6	27.84	2.23	1	16.90	289.11
	Kamaci	27.53	1.69	0	16.22	267.03
	Proposed	28.15	1.82	0	16.86	290.82
News	JM 8.6	30.81	0.79	0	16.94	256.84
	Kamaci	29.84	0.70	0	16.35	222.20
	Proposed	31.29	0.76	0	16.52	239.09
Silent	JM 8.6	30.51	0.69	0	16.95	246.90
	Kamaci	29.74	0.62	0	16.11	210.23
	Proposed	30.96	0.50	0	16.56	225.23

5.2.5 Summary

In this chapter, we investigate and analyze the use of Cauchy distribution as a model to estimate the rate and distortion characteristics for video coding on the application of rate control. Based on Cauchy R-D optimization model, we derive an expression for optimal quantization step size and a linear prediction rate and distortion model parameters. We then propose a rate control scheme using Cauchy R-D optimization model under low delay constraint. Under low delay constraint, it has been shown that the effect of encoder buffer fill-up causes frame skipping and lower PSNR. The consideration of bit allocation involves the buffer status, the number of bits used in previous frame, the complexity of basic unit. The simulation results show that our proposed scheme achieves better performance in terms of better PSNR, lower PSNR standard deviation, less frame skipping, more accurate bit rate used, and lower processing time compared to that of H.264 JM 8.6 for the ranges of low bit rate and low delay constraints indicated.

Chapter 6

Conclusion

Based on the observation that Cauchy distributions provide more accurate estimates of the statistical distribution of DCT coefficients than the Laplacian distributions. Therefore, in this dissertation we propose a mathematical model for rate and distortion optimization algorithm by Lagrange multiplier technique $J = D + \lambda R$ as the cost function to find the rate and distortion model subject to the target bit rate constraint resulting in the optimum choice of quantization step sizes based on Cauchy probability density function as shown in Chapter 3. The proposed Cauchy rate-distortion optimization model is used to compute the optimal quantization step size of each basic unit in each P-frame. Linear regression analysis is used to find the Cauchy rate distortion optimization model parameters.

In Chapter 4, we propose new rate control scheme that applied Cauchy rate-distortion optimization model and show that proposed algorithm can work under low bit rate for H.264 video coding. We use the expression in eq.(3.26) to design a new rate control scheme to improvement in average PSNR Cauchy rate-distortion optimization models presented in the Chapter 3 can be used to find the optimal quantization step size in each basic unit of each frame for H.264 video coding. Our proposed rate control scheme with Cauchy rate-distortion optimization model composes of three layers: Group of picture layer (GOP layer) rate control, frame layer rate control and basic unit layer rate control. For our simulation, we compare the performance between our proposed a rate control scheme with Cauchy rate-distortion optimization model and H.264 JM 8.6 rate control in terms of the average PSNR, PSNR standard deviation, bit rate used, and processing times. The simulation results show that our proposed scheme achieves better performance in terms of better PSNR, lower PSNR standard deviation, more accurate bit rate used, and lower processing time compared to that of H.264 JM 8.6 for the ranges of low bit rate constraints indicated.

In Chapter 5, we present the impact of delay constraint for H.264 video transmission. In real time video transmission, delay is very critical because video frames must be presented to the viewer at constant intervals. One factor that

contributes to the delay is the encoding process. Buffering the data prior to transmission in a constant bit rate (CBR) encoder is an added tradeoff because a larger buffer size implies a larger delay. But if the buffer size is small, additional number of bits accumulated in the encoder buffer will result in the higher number of bits than speculated, thus frame will be skipped to reduce the buffer delay and to avoid buffer overflow. That is introduced high fluctuated and low visual quality in video communication system. Then taking into account the low delay factor, we propose a new rate control based on Cauchy-based rate-distortion (R-D) optimization model for H.264 low delay video transmission.

In this Chapter, we use the expression in eq.(3.26) to design a new rate control scheme to reduce the number of frame skipped with an improvement in average PSNR and smoother video quality. Cauchy R-D models presented in the previous section can be used to find the optimal quantization step size in each basic unit of each frame. Our proposed rate control scheme with Cauchy R-D optimization model composes of three layers: Group of picture layer (GOP layer), frame layer and basic unit layer rate control. In GOP layer rate control, with consideration of the total number of bits for all non-coded P-frames, we compute the occupancy of virtual buffer after each frame is encoded. In frame layer rate control, the objective of this stage is to determine the number of target bits budget for each P-frame. In basic unit layer rate control is to compute the quantization step size of current basic unit based on Cauchy R-D optimization model and basic unit complexity in term of residual variance. The computed quantization parameter (QP) is then adjusted to prevent buffer from overflow and underflow. For our simulation, we compare the performance between our proposed a rate control scheme with Cauchy rate-distortion optimization model, H.264 JM 8.6 rate control and rate control in [25]. The simulation results show that our proposed scheme achieves better performance in terms of better PSNR, lower PSNR standard deviation, less frame skipping, more accurate bit rate used, and lower processing time compared to that of H.264 JM 8.6 for the ranges of low bit rate and low delay constraints indicated.

References

- [1] ISO/IEC JTCl/SC29/WG11, Test Model 5,(1993).
- [2] ISO/IEC 13818-2:1994. Information technology-generic coding of moving pictures and associated audio. Part s: Visual, (1994).
- [3] The MPEG-2 International Standard, ISO/IEC, Reference Number ISO/IEC 13818-2, (1996).
- [4] Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, (1995).
- [5] ITU-T/SG15, Video codec test model, TMN8, Portland, (June 1997)
- [6] J. Ribas-Corbera and D. Lei. Rate control in DCT video coding for low delay communications. IEEE Trans. Circuits Syst. Video Technol. : 9(Feb.1999) :172-185.
- [7] ISO/IEC 14496-2: 1999. Information technology-coding of audio/visual objects. Part 2: Visual. (1999).
- [8] N. Brady. MPEG-4 standardized methods for the compression of arbitrarily shaped video objects. IEEE Trans. Circuits Syst. Video Technol.: 9 (Dec. 1999) : 1170-1189.
- [9] H.-J. Lee, T. Chiang and Y.-Q.Zhang. Scalable rate control for MPEG-4 video. IEEE Trans. Circuits Syst. Video Technol. : 10 (Sept. 2000) : 878-894.
- [10] ISO/IEC JTCl/SC29/AWG11. MPEG-4 video verification model v180.0. Pisa, Italy, (Jan. 2001).
- [11] T. Sikora. The MPEG-4 video standard verification model. IEEE Trans. Circuits Syst. Video Technol. : 7 (Feb. 1997) : 19–31.
- [12] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. : 13 (Jul. 2003) : 560–576.
- [13] IEEE. Trans. Circuit Syst. Video Techno. Special issue on the h.264/AVC video coding standard. : 13 (July 2003).
- [14] G. Sullivan, T. Weigand, and K.-P.Lim. Joint model reference encoding methoda and decoding concealment methods. JVT-I049, San Diego, (Sept. 2003)

- [15] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264/ISO/IEC 14 496-10 AVC. in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTG050, (2003).
- [16] T.Berger. Rate Distortion Theory. Englewood Cliffs, NJ: Prentice-Hall, (1984).
- [17] R. C. Reininger and J. D. Gibson. Distributions of the two-dimensional DCT coefficients for images. IEEE Trans. Commun. : 31(Jun. 1983) : 835–839.
- [18] E.Y.Lam and J.W.Goodman. A mathematical analysis of the DCT coefficient distributions for images. IEEE Trans. Image Process. : 9 (Oct. 2000) : 1661–1666.
- [19] F. Muller. Distribution shape of two-dimensional DCT coefficients natural images. Electron. Lett. : 29 (Oct. 1993) : 1935–1936.
- [20] T.Eude, R. Grisel, H. Cherifi, and R. Debrie. On the distribution of the DCT coefficients. in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing : 5 (Apr. 1994) : 365-368.
- [21] S. R.Smooth and R.A.Lowe. Study of DCT coefficients distributions. in Proc. SPIE. : (Jan. 1996) : 403-411.
- [22] M. Barni, F. Bartolini, A. Piva, and F. Rigacci. Statistical Modeling of full frame DCT coefficients. in Proc. Eur. Signal Processing Conf. EUSIPCO 98. : 3 (Sep. 1998) : 1513-1516.
- [23] T. Chiang and Y.-Q. Zhang. A new rate control scheme using quadratic rate distortion model. IEEE Trans. Circuits Syst. Video Technol. : 7 (Feb. 1997) : 246–250,.
- [24] L.-J. Lin, A. Ortega, and C.-C. J. Kuo. Cubic spline approximation of rate and distortion functions for MPEG video. Vis. Commun. Image Process. : (Mar. 1996).
- [25] Y.Altunbasak and N.Kamaci. Frame bit allocation for the H.264/AVC video coder via Cauchy-Density –Based rate and Distortion models. IEEE Trans. Circuits Syst. Video Technol. : 15 (Aug. 2005).
- [26] Stanislaw H. Zak and Edwin K.P. Chong, An introduction to optimization. New York: Wiley, (2001).
- [27] C.E.Shannon. A mathematical theory of communications. Bell System Tech. Journal. : 27 (July 1948) : 397-423.

- [28] C.E.Shannon. Coding theorems for a discrete source with a fidelity criterion. in IRE Nat. Conv. Rec. , New York, NY. : 4 (Mar. 1959) : 142-163.
- [29] T.M.Cover and J.A.Thomas. Elements of Information Theory. New York: Wiley, (1991).
- [30] R.M.Gray. Source Coding Theory. Norwell, MA: Kluwer, 1990.
- [31] T.Berger and J.D.Gibson. Lossy source coding. IEEE Trans. On Inform. Theory. : 44 (Oct. 1998) : 2693-2723.
- [32] L.D.Davisson. Rate-distortion theory and applications. Proc. IEEE. :60 (July.1972) : 800-808.
- [33] J.J.Y.Huang and P.M.Schultheiss. Block quantization of correlated Gaussian random variables IEEE Trans. Commun. Syst. : 11 (Sept. 1963) : 289-296.
- [34] R.E.Bellman. *Dynamic programming*. Princeton, NJ: Princeton University Press, (1957).
- [35] A.Otega, K.Ramchandran, and M.Vetterli. Optimal trellis-based buffered compression and last approximations. IEEE Trans. On Image Processing. : 3 (Jan. 1994) : 26-40.
- [36] G.D.Forney. The Viterbi algorithm Proc. IEEE. : 61 (Mar.1973) : 268-278.
- [37] T.Commen, C. leiserson, and R.Rivest. Introduction to algorithms. Cambridge, MA: MIT press, (1990).
- [38] W.K.Pratt, Digital Image Processing. New York: Wiley, (1978).
- [39] A.N.Netravali and J.O.Limb. Picture coding: A review. Proc. IEEE. : 68 (Mar.1960) : 7-12.
- [40] ITU-T Recommendation H.261 version 2. Video codec for audiovisual services at $p \times 64$ kbit/s., (1990).
- [41] CCIT SG XV WP/1/Q4. Description of reference model 8 (RM8).,(June 1989).
- [42] ISO/IEC 11172-2:1991. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbps. Part 2: Visual, (1991).
- [43] E.Viscito and C.Gonzales. A video compression algorithm with adaptive bit allocation and quantization. in Proc. Visual Commun. Image Processing (VCIP'91). Boston, MA:SPIE : 1605 (Nov.1991) : 58-72.
- [44] A. Vetro, H.Sun, and Y.Wang. MPEG-4 rate control for multiple video objects. IEEE Trans. Circuits Syst. Video Technol. : 9(Feb.1999) : 186-199.

- [45] W.Ding and B.Liu. Rate control of MPEG video coding and recording by rate – quantization modeling. IEEE Trans. Circuit Syst. Video Technol. : 6 (Feb.1996) : 12-20.
- [46] H.M.Hang and J.J.Chen. Source model for transform video coder and its application –Part I fundamental theory. IEEE Trans. Circuit Syst. Video Technol. : 7 (Apr.1997) : 287-298.
- [47] B. Tao., B.W.Dickinson, and H.A. Peterson. Adaptive model-driven bit allocation for MPEG video coding. IEEE Trans. Circuit Syst. Video Technol. : 10 (Jan.2000) : 147-157.
- [48] J.B.Cheng and H.M.Hang Adaptive piecewise linear bits estimation model for MPEG based video coding. Journal of Visual Communication and Image Representation. : 8 (Mar.1997) : 51-67.
- [49] Z.He, Y.K.Kim, and S.K.Mitra. Low delay rate control for DCT video coding via ρ domain source modeling. IEEE Trans. Circuit Syst. Video Technol., : 11 (Jan.2001) : 928-940.
- [50] Z.He. ρ - domain rate-distortion analysis and rate control for visual coding and communication. Phd. Dissertation, Univ. of California, Santa Barbara, (2001).
- [51] Pan Feng, Z.G.Li., Lim Keng Pang and G.N.Feng. Reducing frame skipping in MPEG-4 rate control scheme. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). : 4 (May.2002) : 3409-3412.
- [52] H.Song and C.C.J.Kuo. Rate control for low bit rate video via variable – encoding frame rates. IEEE Trans. Image Processing technol. : 11(Apr.2001) : 512-521.
- [53] M.Jiang , X.Yi, and N.Ling . Frame bit allocations scheme for constant quality video. IEEE ICME, (2004).
- [54] M.Jiang, X.Yi, and N.Ling. On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation. IEEE Trans. Consumer Elec. : 51 (Feb. 2005) : 533–545.
- [55] A.Gersho and R.M. Gray. Vector quantization and signal compression. Kluwer Academic, (1992)

- [56] R.J.Freund and W.J.Wilson. Regression Analysis: Statistical Modeling of a Response Variable. New York: Academic, (1998) : 39–41.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Publications

International Conferences

1. N.Eiamjumrus and S.Aramvith, "Cauchy based rate-distortion optimization model for H.264 rate control", Circuits and Systems, 2006 (APCCAS' 2006). IEEE Asia Pacific Conference, Dec 2006, pp.77 – 80.
2. N.Eiamjumrus and S.Aramvith, "New rate control Scheme based on Cauchy Rate-Distortion Optimization Model for H.264 Video Coding," Intelligent Signal Processing and Communications, 2006 (ISPACS' 06). IEEE International Symposium, Dec. 2006, pp.143 – 146.
3. N.Eiamjumrus and S.Aramvith, "Rate Control Scheme Based on Cauchy R-D Optimization Model for H.264/AVC under Low Delay Constraint" Intelligent Information Hiding and Multimedia Signal Processing, 2006 (IIH-MSP' 06). IEEE International Conference, Dec. 2006, pp. 205 – 210.

National Conference

1. N.Eiamjumrus and S.Aramvith, "Improvement of Rate Control Scheme using Cauchy Rate-Distortion Optimization Model for H.264 Video Coding Standard", 29th Electrical Engineering Conference 2006 (EECON-29)., Volume 2, Nov. 2006, pp. 921-924.

International Journal

1. N.Eiamjumrus and S.Aramvith, "Rate Control Scheme based on Cauchy R-D Optimization Model for H.264/AVC under Low Delay Constraint", accepted for publication, Journal of Digital Information Management, 2007.

Rate Control Scheme based on Cauchy R-D Optimization Model for H.264/AVC under Low Delay Constraint

Nongluk Eiamjumrus and Supavadee Aramvith

Department of Electrical Engineering, Faculty of Engineering
Chulalongkorn University, Bangkok 10330, Thailand
Tel : (66-2)218-6909, Fax: (66-2)218-6912
Email: supavadee.a@chula.ac.th

Abstract : *The motivation of this work is based on the observation that Cauchy distribution provides more accurate estimates of rate and distortion characteristics of video sequences than the previously distribution used such as Laplacian distribution. In this paper, we propose a new rate control scheme for H.264 low-delay video transmission using Cauchy rate-distortion optimization model. Our proposed method uses the Lagrange Multiplier technique as the cost function to find the rate and distortion model subject to the target bit rate constraint resulting in the optimum choice of quantization step sizes. Model parameters are estimated using statistic linear regression analysis. The target number of bits for each frame is determined according to their buffer status, the number of bits use in the previous frame, and basic unit complexity. The proposed scheme has been implemented on H.264 video coder. Simulation results show that the proposed rate control algorithm achieves an improvement of average PSNR for up to 1.68dB with less number of frames skipped compared to the H.264 JM8.6 rate control.*

Keywords : Cauchy distribution, low delay, rate control, Lagrange multiplier technique, linear regression analysis

1. Introduction

Multimedia communications have experienced rapid growth and commercial success in the last decades. Many multimedia applications, such as digital television broadcasting, video streaming, video conferencing and video-on-demand, require video coding schemes that can provide acceptable quality of service to the end users. Depending on the type of services and available bandwidth of the channel some video applications, such as real-time video transmission over wireless channel, might, put constraints on the video coding scheme that can affect the visual quality and motion continuity of video. In this paper, we are particularly interested in applications where low-bit rate and low-delay constraints are of great concern.

To provide effective and reliable video communication, rate control plays a key role in assigning optimal number of bits for each video frame and ensures the generation of a constant bit rate video stream to the channel. Due to its importance, rate control algorithms have always been a challenging research area. Several rate control schemes in various international video coding standards such as TM5 [1] for MPEG-2, TMN8 [2] for H.263, VM-8 [3] for MPEG-4 and JM for H.264/AVC [4] have been proposed. These video coding standards employ efficient compression techniques to remove the spatial and temporal redundancy within and between frames. Due to limited storage size and limited communication bandwidth, quantization step size is introduced to compress the bit rate of the video signal such that storage and/or bandwidth constraints can be satisfied. Generally, the rate control part is a normative part in video coding standards. Standards are flexible enough to allow designers to develop suitable rate control schemes for specific applications.

To achieve the best video quality, rate-distortion theory [5] is the theoretical foundation of rate control. It originates from Shannon's paper [6] and forms a central part of information theory [7] and lossy source coding [8] which is directly related to lossy image compression. Lossy source coding schemes, such as video coding, concentrates on the tradeoffs between distortion and bit rate. Rate-distortion optimization techniques can be used to find the optimal quantizer. Lagrange optimization [9] and dynamic programming technique [10-11] are two popular techniques to find the optimal or nearly optimal bit-allocation for each Macroblock. One difficulty is that the optimization scheme requires huge amounts of computation especially in dynamic programming.

Another type of approach to find the optimal quantizer for rate control is through mathematical modeling. Formulas are derived from optimizations based on the input statistical distribution of DCT coefficients. On the percentage and/or number of zeros in the quantized DCT coefficients, i.e., ρ -domain [20]. Several studies on the statistical probability distribution of the AC-coefficients have been investigated. In earlier studies, the AC coefficients were conjectured to have Gaussian distributions [12]. Later, several other distribution models were reported, including generalized Gaussian and Laplacian distributions [13-15]. Other studies modeled the statistical distributions of DCT coefficients using more complex probability density functions such as Gaussian Mixture Models (GMM). In [13], Muller used a generalized Gaussian function that includes Gaussian and Laplacian probability density functions as special cases. *Eude et al.* reported that the statistics of the DCT coefficients can be modeled as a linear combination of a number of Laplacian and Gaussian probability density

functions [15]. Comparing their models with Laplacian, Gaussian and Cauchy probability density functions, they claimed that the distribution of DCT coefficients follow neither a Cauchy nor a Laplacian distribution only but can be accurately modeled as a mixture of Gaussian distributions. Although a generalized Gaussian density function can model the statistics of the DCT coefficients more accurately, it is not widely used in practice because it is mathematically difficult to analyze. Cubic spline models [16] in MPEG-2 have shown to be more accurate in estimating the rate characteristics of video sequence, but it is computationally complex. Nevertheless, previous works on the rate control of video coding standards are based on the assumption that AC coefficients follow a Laplacian distribution. Examples are TMN8 R-D model [2], H.264's Quadratic R-D model [4], and Logarithm R-D model [19].

Recently the work in [17], N.Kamaci and Y.Altunbasak observed that in most cases Cauchy distributions provide more accurate estimates of the statistical distribution of DCT coefficients in typical video sequences compared to that of Laplacian distributions. Nevertheless the rate and distortion characteristics are considered separately. The parameters of the rate model are updated based on the previously encoded frame and a constant factor with the assumption that the distortion is constant regardless of the frames. In this connection, our work is based on the assumption that AC coefficients follow a Cauchy distribution. We model rate and distortion based on Cauchy density function, with approximations as a function of the quantization step size. We further use Lagrange multiplier techniques to obtain optimal solutions of the quantization step sizes with minimum distortion for each basic unit subject to rate constraint. The work has been summarized in [21,22].

In real time video transmission, delay is very critical because video frames must be presented to the viewer at constant intervals. One factor that contributes to the delay is the encoding process. Buffering the data prior to transmission in a constant bit rate (CBR) encoder is an added tradeoff because a larger buffer size implies a larger delay. But if the buffer size is small, additional number of bits accumulated in the encoder buffer will result in the higher number of bits than speculated, thus frame will be skipped to reduce the buffer delay and to avoid buffer overflow. This work also shows the impact of low delay constraint based on the H.264/AVC rate control algorithm, and show that the degraded performance of H.264 rate control under low delay constraint. Then taking into account the low delay factor, we propose a new rate control based on Cauchy-based rate-distortion (R-D) optimization model for H.264 low delay video transmission. The target number of bits for each frame is determined according to their buffer status, and the number of bits used in the previous frame. In the basic unit layer, we consider the complexity on basic unit based on the residual variance of each basic unit to find the optimal target bit rate for each frame. The required model parameters are updated in real – time using linear regression analysis and update based on the complexity of each basic unit. The experimental results show that our proposed scheme achieves better average PSNR with smoother video quality and less frame skipping compared with the H.264 JM8.6 rate control.

This paper is organized as follows. In section 2, background information about Cauchy rate and distortion model are shown. In section 3, we derive the Cauchy rate-distortion models by using Lagrange multiplier technique to find the optimal quantization step sizes that minimizes the distortion subject to the target bit budget. In section 4, linear prediction rate and distortion model parameters are explained. In section 5, we show the impact of low delay constraint based on H.264 rate control algorithm. In section 6, the proposed Cauchy rate – distortion model implemented on rate control scheme under low delay constraint is presented. In section 7 we show the simulation results compared with the JM8.6 rate control. Section 8 concludes our work.

2. Background of rate and distortion model based on Cauchy density function

In this section, we briefly explain the Cauchy rate model and distortion models based on the assumption of Cauchy density function.

2.1 Cauchy rate model

Assume that, the DCT-coefficients of the motion compensated difference frame are Cauchy distribution with Cauchy parameter, μ . A closed form expression of the rate as a function of the quantization step size for the Cauchy distribution is given by the empirical entropy $H(Q)$ of the quantized DCT coefficients, as shown in eq.(1),

$$\begin{aligned}
 H(Q) = & -\frac{2}{\pi} \tan^{-1}\left(\frac{Q}{2\mu}\right) \log_2\left(\frac{2}{\pi} \tan^{-1}\left(\frac{Q}{2\mu}\right)\right) \\
 & - 2 \sum \frac{1}{\pi} \tan^{-1}\left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4)Q^2}\right) \\
 & \times \log_2\left(\frac{1}{\pi} \tan^{-1}\left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4)Q^2}\right)\right)
 \end{aligned} \tag{1}$$

, where Q is the quantization step size.

Eq.(1) gives a parametric description of the rate using an entropy coded uniform quantizer. Note that the entropy function of quantization step size, Q , can be approximated as a linear equation, as shown in eq.(2)

$$R = H(Q) = aQ^{-\alpha} \quad (2)$$

, where a and α are model parameters. The values of these two parameters need to be estimated and their values depend on the value of μ .

2.2 Cauchy distortion model

If considering each quantization range is independent, it is therefore possible to compute the expected distortion as a function of Q . The closed form of expected distortion can be shown as in eq.(3),

$$D(Q) = \frac{1}{\pi} \sum_{i=1}^{\infty} \left\{ -\mu Q + i\mu Q \ln \left(\frac{\mu^2 + \left(i + \frac{1}{2}\right)^2 Q^2}{\mu^2 + \left(i - \frac{1}{2}\right)^2 Q^2} \right) \right. \\ \left. + \left(\mu^2 - i^2 Q^2 \left(i^2 - \frac{1}{4} \right) \right) \right\} \quad (3)$$

Eq.(3) shows a parametric description of the distortion using an entropy coded uniform quantizer. Note that the distortion function in eq.(3) can be approximated as a linear equation, as shown in eq.(4),

$$D(Q) = bQ^\beta \quad (4)$$

, where b and β are model parameters. The values of these two parameters need to be estimated and their values depend on the value of μ .

3. Generalized Cauchy R-D optimization model using Lagrange multiplier technique

In this section, we derive an expression for the quantization step sizes that minimize distortion subject to bit rate constraint. From the distortion shown in eq.(2), the expected number of bits for each basic unit l^{th} in a frame is defined by eq.(5),

$$R_l = Ca_l Q_l^{-\alpha_l} + H_l \quad (5)$$

, where C is the number of pixel in each basic unit, H_l is the actual number of header bits generated by each basic unit in the current frame, R_l is the number of target bits and Q_l is the quantization step size for each basic unit. The basic unit in H.264 video coding is defined as a group of contiguous macroblock in a frame ,i.e., basic unit can be a macroblock, a slice or a frame.

The distortion measure, D , for encoding every basic unit in each frame is thus defined by eq.(6),

$$D = \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} \quad (6)$$

, where N_{unit} is the number of basic unit in each frame.

An expression for the quantization step sizes $Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*$ is formulated as shown in eqs. (7)-(8). It is to find the minimum distortion subject to the constraint that the total number of bits, i.e., the sum of the total number of bits of every basic unit in each j^{th} frame must be equal to $R_{MAX}(j)$,

$$Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^* = \arg \min_{Q_1, \dots, Q_{N_{unit}}} \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} \\ \text{Subject to } \sum_{l=1}^N R_l = R_{MAX}(j) \quad (7)$$

$$Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^* = \arg \min_{Q_1, \dots, Q_{N_{unit}}} \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} + \lambda \left(\sum_{l=1}^N R_l - R_{MAX}(j) \right) \quad (8)$$

, where λ is the Lagrange multiplier and $R_{MAX}(j)$ is defined the target bit budget in each frame.

By using Lagrange multiplier technique, the expression for the optimum choice of quantization step sizes, $J(Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^*)$, can be formulated as shown in eq. (9). We assume that the number of header bits in each basic unit are equal ($H_l = H$).

$$J(Q_1^*, Q_2^*, \dots, Q_{N_{unit}}^*, \lambda^*) = \frac{1}{N_{unit}} \sum_{l=1}^{N_{unit}} b_l Q_l^{\beta_l} + \lambda \sum_{l=1}^{N_{unit}} (C a_l Q_l^{-\alpha_l}) + \lambda (N_{unit} \times H - R_{MAX}(j)) \quad (9)$$

The minimum values can be solved by using partial derivatives, i.e., $\partial J / \partial Q_l^* = 0$ and $\partial J / \partial \lambda^* = 0$. As a result, the optimum quantization step size for each basic unit l^{th} , Q_l^* , can be found in term of Cauchy rate model parameters, (a_l, α_l) , and Cauchy distortion model parameters (b_l, β_l) , as shown in eq.(10),

$$Q_l^* = \left(\varepsilon_l \left(\frac{\sum_{l=1}^{N_{unit}} (a_l (\varepsilon_l)^{-k})}{R_{MAX}(j) - N_{unit} \times H} \right) \frac{C}{G} \right)^{s_l} \quad (10)$$

, where a_l , α_l , b_l , β_l are model parameters of the basic unit l^{th} , ε_l denotes $a_l \alpha_l / b_l \beta_l$, s_l denotes $1 / (\alpha_l + \beta_l)$, and $G = \alpha_l / (\alpha_l + \beta_l)$. Note that, in our study, the expression of G represented in terms of α_l and β_l has approximately a constant value of 0.45.

4. A linear prediction rate and distortion model parameters

In this section, we relate the expression from Cauchy rate model eq.(2) and Cauchy distortion model eq.(4) in term of linear equation to find the Cauchy rate-distortion model parameters of the basic unit l^{th} in the frame j^{th} (a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1}) by using statistical linear regression analysis [18].

4.1 Cauchy rate model parameters

the Cauchy rate model parameters a_{j-1} , α_{j-1} , can be found when the encoder collects the bit rate and quantization step for each type of picture at the end of encoding each basic unit. From eq.(2), we use logarithmic linear function, as shown in eq.(11).

$$\ln(R) = \ln(a) - \alpha \ln(Q) \quad (11)$$

In our work, we use linear regression analysis that models the relationship between two variables by fitting a linear equation to the observed data. In this work, we use the formula below to find the Cauchy rate model parameters as shown in eqs.(12)-(13),

$$\ln(a_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(R_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (12)$$

$$\alpha_{j-1} = \frac{\left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} (\ln(R_k)) \right) - (l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (13)$$

, where R_k and Q_k is the actual number of bits used and quantization step size for coding in previously encoded basic unit k to basic unit $l-1$.

4.2 Cauchy distortion model parameters

In our work, the expression of distortion function in eq. (4) is represented in terms of mean square error (MSE) of each basic unit and logarithmic function, as shown in eq.(14).

$$\ln(MSE) = \ln(b) + \beta \ln(Q) \quad (14)$$

Follow the same derivation as in the case of Cauchy rate model parameters, we use linear regression analysis to update the model parameters after we encoded each basic unit as in eqs. (15)-(16),

$$\ln(b_{j-l}) = \frac{\left(\sum_k^{l-1} \ln(MSE_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (15)$$

$$\beta_{j-l} = \frac{(l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right) - \left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (16)$$

, where MSE_k denotes the actual mean square error for the previous coded basic unit k to basic unit $l-1$.

5. Impact of low delay constraint of H.264 rate control

5.1 H.264 rate control

In H.264 rate control, a quantization parameter is determined by using linear and quadratic rate – distortion models. The rate control in H.264 [4] is composed Group of picture (GOP) layer, frame layer and basic unit layer. When encoded the current frame, rate control will compute the occupancy of encoder buffer by using fluid traffic model as shown in eq. (17). The initial buffer fullness is set to zero. The N_{gop} denotes the total number of GOP, $n_{i,j}$ denotes the j th frame in the i th GOP, $B_c(n_{i,j})$ denotes the occupancy of encoder buffer, $A(n_{i,j})$ denotes number of bits generated by the j th frame in the i th GOP, F_r denotes the target frame rate and $u(n_{i,j})$ denotes the available channel bandwidth.

$$B_c(n_{i,j}) = B_c(n_{i,j-1}) + (A(n_{i,j-1}) - \left(\frac{u(n_{i,j-1})}{F_r} \right)) \quad (17)$$

If the occupancy of encoder buffer is larger than the maximum encoder buffer size, B_s , rate control will skip encoding frame and release accumulated bit in the encoder buffer to the channel. The determination of a target bit for each P frame composes of 2 steps.

Step 1 Budget allocation among pictures. The bit allocation is implemented by predefining a target buffer level, $Tbl(n_{i,j+1})$, for each P picture, as shown in eq. (18), where N_p is the number of P frames in GOP.

$$Tbl(n_{i,j+1}) = Tbl(n_{i,j}) - \frac{B_c(n_{i,2}) - B_s / 8}{N_p - 1} \quad (18)$$

Step 2 Target bit rate computation. The target bit rate, $f(n_{i,j})$, for the j th P frame in the i th GOP is scaled based on the target buffer level, current buffer level, frame rate, and channel bandwidth. It is given in eq. (19),

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + \gamma (Tbl(n_{i,j}) - B_c(n_{i,j})) \quad (19)$$

,where γ is a constant weighting factor. Further adjustment by a weighted combination of the average number of remaining bits for each frame is given, as shown in eq. (20),

$$f(n_{i,j}) = \beta * \frac{T_r(n_{i,j})}{N_p - j} + (1 - \beta) * \tilde{f}(n_{i,j}) \quad (20)$$

,where $T_r(n_{i,j})$ is the total number of remaining bits left to encode the j th frame onwards in the i th GOP, and β is a constant weighting factor. Note that the detail information of H.264 rate control can be found in [4].

5.2 Simulation results on the impact of low delay constraint of H.264 rate control

In encoding process, the bits left from previously encoded frame that has not been transmitted produce a delay in an order of few milliseconds. To reduce such delay, the encoder buffer size should be kept small. In our study, we impose a low delay constraint that the number of bits in encoder buffer $B_c(n_{i,j})$ must not be greater than the maximum buffer size M , otherwise a frame will be skipped, where, $M = B_s = u(n_{i,j})/F_r$. Hence, the maximum buffer delay is $M/u(n_{i,j})=1/F_r$ seconds. Where $u(n_{i,j})$ and F_r is target channel bandwidth and frame rate, respectively.

To show the impact of the low delay constraint of JM8.6 rate control [4], we encoded an I-frame followed by P-frame of Foreman sequence at 16 kbps with the frame rate of 10 fps. The maximum buffer delay in our simulation is set to 100 ms. As shown in Fig.1 (a) and (b), when the buffer fullness level higher than maximum delay constraint, frame will be skipped. Thus, without the proposed algorithm described in Section 6, there are additional 17 frames skipped compared to the case with no delay constraint. Fig. 1(b) shows PSNR degradation. There are several sharp drops in PSNR which can be as bad as 15 dB drop. On the average, PSNR degrades 0.63 dB, when compared to the case with no delay constraint.

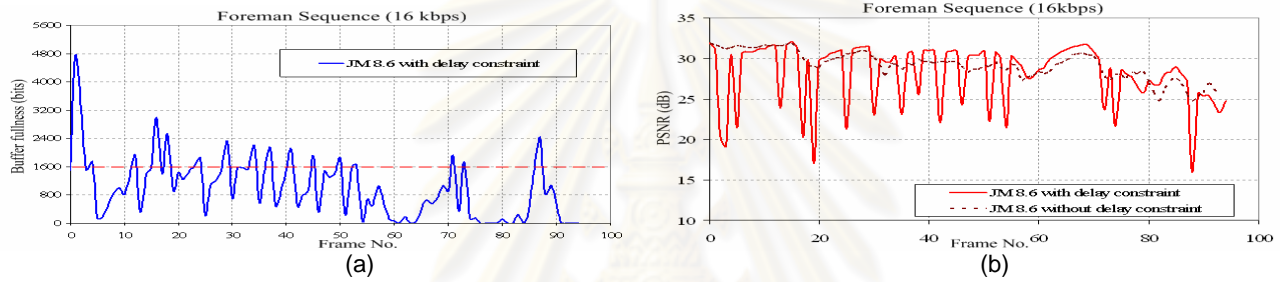


Figure.1 Simulation using Foreman sequence encoded by H.264 JM8.6 rate control at 16 kbps with delay constraint of 100 ms. (a) Buffer fullness level (bits) per frame with maximum buffer size (dashed line). (b) Average PSNR per frame.

6. Proposed rate control scheme using Cauchy R-D optimization model under low delay constraint

In this section, we use the expression in eq.(10) to design a new rate control scheme to reduce the number of frame skipped with an improvement in average PSNR and smoother video quality. Cauchy R-D models presented in the previous section can be used to find the optimal quantization step size in each basic unit of each frame.

Our proposed rate control scheme with Cauchy R-D optimization model composes of three layers: Group of picture layer (GOP layer), frame layer and basic unit layer rate control. In GOP layer rate control, with consideration of the total number of bits for all non-coded P-frames, we compute the occupancy of virtual buffer after each frame is encoded. In frame layer rate control, the objective of this stage is to determine the number of target bits budget for each P-frame. The task of basic unit layer rate control is to compute the quantization step size of current basic unit based on Cauchy R-D optimization model and basic unit complexity in term of residual variance. The computed quantization parameter (QP) is then adjusted to prevent buffer from overflow and underflow.

6.1 GOP layer rate control

First, the total number of bits allocated for the GOP is computed as shown in eq.(21). The first I-frame and the first P-frame of the GOP are coded by using $QP_{initial}$. The computation of $QP_{initial}$ is based on the available channel bandwidth as used in H.264 rate control [4]. The remaining bits, $T_r(n_{i,j})$, for all non-coded P-frames are updated after the $(j-1)^{th}$ frame is encoded, as shown in eq.(22),

$$T_{GOP} = \frac{u(n_{i,j})}{F_r} \times N_{GOP} \quad (21)$$

$$T_r(n_{i,j}) = T_{GOP} - \sum_{j=0}^{j-1} b(n_{i,j}) \quad (22)$$

, where $u(n_{i,j})$ is the target channel bandwidth, F_r is the frame rate, N_{GOP} is the number of frames in each GOP, and $b(n_{i,j})$ is the actual number of bits generated by each j^{th} frame. After each frame are encoded, the occupancy of virtual buffer are updated as in eq.(17) in section 5.1.

6.2 Frame layer rate control

The objective of this stage is to determine the number of target bits budget before coding of the current j^{th} frame. The algorithm computes the target bits for each P-frame, $f(n_{i,j})$, according to the current buffer occupancy and the number of bits used in the previous P-frame, as explained in the following steps.

Step 1 : Number of bits used in the previous P-frame is used to compute the target bits, $\hat{f}(n_{i,j})$, as shown in eq.(23).

$$\hat{f}(n_{i,j}) = \frac{T_r(n_{i,j})}{N_{GOP} - j} \quad (23)$$

Step 2 : Current buffer occupancy and buffer size information are used to compute the target bits for each P-frame, $f(n_{i,j})$, as shown in eq.(24),

$$f(n_{i,j}) = \hat{f}(n_{i,j}) + \omega V_S - B_c(n_{i,j}) \quad (24)$$

, where ω is constant value. Note that ω is empirically set to be 0.8 in our simulations.

6.3 Basic unit layer rate control

In basic unit layer, the suitable quantization step sizes are obtained using Cauchy R-D optimization model, described in Section 3. The quantization parameter is further adjusted to keep bit rate under the given constraints, and to prevent the buffer from overflow and underflow. Model parameters will be updated after encoding each basic unit. The layer is divided into pre-encoding and post-encoding stages.

1. Pre-encoding stage

In this stage, we compute the quantization step size of current basic unit in three cases.

Case 1 : For the first basic unit in the current frame, quantization parameter is obtained, as shown in eq. (25),

$$QP_{1,i}(j) = AvgQP_i(j-1) \quad (25)$$

,where $QP_{1,i}(j)$ is the quantization parameter of the first basic unit and $AvgQP_i(j-1)$ is the average of quantization parameter for all basic units in the previous frame.

Case 2 : When the number of remaining bits, $T_r(n_{i,j})$, is lower than zero, quantization parameter is obtained, as shown in eq. (26).

$$QP_{1,i}(j) = QP_{1-1,i}(j) + 2 \quad (26)$$

Case 3 : The quantization step size of current basic unit is computed by Cauchy R-D optimization model using the formula in eq. (10), where a_{j-1} , α_{j-1} , b_{j-1} , and β_{j-1} are model parameters of the basic unit l^{th} of frame j , respectively. The algorithms are according to the following steps.

Step 1: To get a better target bits estimation for each frame, we need to consider the complexity factor of each basic unit, $\gamma(j,l)$. It is defined in eq. (27),

$$\gamma(j,l) = \frac{Re s_{var}(j,l)}{Ave_Re s_{var}(j-1)} \quad (27)$$

, where $Re s_{var}(j,l)$ is the residual variance of the basic unit l^{th} and $Ave_Re s_{var}(j-1)$ is the average of residual variance of all basic units in the previous frame $(j-1)^{th}$. Note that the value of $\gamma(j,l)$ is bounded in the range of [0.8,1.2] to prevent too much fluctuations in target bit estimation. The frame target, $f_v(j)$, is thus adjusted, as shown in eq.(28).

$$f_v(n_{i,j}) = \gamma(j,l) \times f(n_{i,j}) \quad (28)$$

Step 2: To reduce the number of frame skipped when the buffer occupancy level is high, i.e., if the level reaches 80%, the target number of bits obtained from eq. (28) will be reduced by 10%. Also to prevent buffer underflow, i.e., the buffer occupancy level is less than or equal to zero, the target number of bits obtained from eq. (28) will be increased by 10%. The adjustment conditions are shown in eqs. (29)-(30),

$$R_{MAX}(j) = \eta \times f_v(n_{i,j}) \quad (29)$$

, where $R_{MAX}(j)$ is defined as the maximum target bit budget in the current frame, η is constant value. In our simulations, we set the value of η according to the current buffer occupancy conditions, as shown in eq.(30).

$$\eta = \begin{cases} 1.10 & \text{if } B_c(n_{i,j}) \leq 0 \\ 0.90 & \text{if } B_c(n_{i,j}) \geq 0.8 \times V_S \\ 1.00 & \text{otherwise} \end{cases} \quad (30)$$

Step 3: To avoid the fluctuation of PSNR and buffer overflow, the target bit budget estimation of current frame are bounded by a lower and upper bounds, as shown in eqs. (31) and (32),

$$Lower_bound_R_{MAX} = MAX \left\{ R_{MAX}(j), \left(\phi \times \frac{u(n_{i,j})}{F_r} \right) \right\} \quad (31)$$

$$Upper_bound_R_{max} = MIN \left\{ R_{MAX}(j), \left(\gamma \times \frac{u(n_{i,j})}{F_r} \right) \right\} \quad (32)$$

, where ϕ and γ are constant values and are set empirically to 0.5 and 3, respectively.

Step 4: Compute the quantization step size (Q_l) of the current basic unit by using the parameters, a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} , $R_{MAX}(j)$ of the basic unit l th of frame j .

The computed quantization parameter of each basic unit is then further adjusted, as shown in eq.(33),

$$\begin{aligned} & \text{if } B_S(n_{i,j}) \geq 1.2 \times B_S \\ & \quad QP_{l,i}(j) = MAX(QP_{initial} + 5, QP_{l,i}(j) + 3) \\ & \text{else if } B_S(n_{i,j}) \leq 0.2 \\ & \quad QP_{l,i}(j) = MAX(QP_{initial} - 2, QP_{l,i}(j)) \\ & \text{else} \\ & \quad QP_{l,i}(j) = MAX(QP_{initial} - 1, QP_{l,i}(j)) \end{aligned} \quad (33)$$

, where $B_S(n_{i,j})$ is defined as the occupancy of virtual buffer after encoding each basic unit. At this stage, $B_S(n_{i,j})$ are updated by eq.(34).

$$B_S(n_{i,j}) = B_c(n_{i,j}) + \sum_{k=1}^l b(n_{i,j-k}) \quad (34)$$

, where $b(n_{i,j-k})$ is the actual number of bits use in the l^{th} basic unit. In H.264 standard, the possible quantization parameter is specified between range 0 - 51.

2. Post-encoding stage

After encoding each basic unit, only Cauchy rate and distortion model parameters, a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} , are updated by linear regression analysis in eqs. (12)-(13) and eqs. (15)-(16), as shown in section 4 to find the optimal quantization step size of each basic unit.

7. Simulation results

For our simulation, we compare the performance between our proposed a rate control scheme with Cauchy rate-distortion optimization model and H.264 JM 8.6 rate control in terms of the average PSNR, PSNR standard deviation, the number of frames skipped, bit rate used, and processing times. The parameters of video coding and tested video sequences are as follows.

- 1) We encode six video sequences covering every aspects of characteristics: Carphone sequence with fast camera and content motion with landscape passing, Foreman sequence with fast camera and content motion with pan at the end, Silent sequence with still camera on slow hand moving, News sequence with still camera on slow moving scene change, Akiyo sequence with still camera on human subject with synthetic background and Claire sequence with still camera on fast human head and shoulder moving.
- 2) Input video signal is QCIF format with a resolution of 176x144 pixels.
- 3) Structure of encoded low-delay video sequence consists of only one I-frame and the rest are P-frames.
- 4) We set 16-256 kbps as target bit rates, 10 fps as frame rate.
- 5) Four maximum buffer size (B_s) is used in our simulations : maximum buffer size $B_s = u(n_{i,j})/F_r$ represents a delay time of 100 ms., maximum buffer size $B_s = 1.5 \times u(n_{i,j})/(F_r)$ represents a delay time of 150 ms., maximum buffer size $B_s = 2.0 \times u(n_{i,j})/(F_r)$ represents a delay time of 200 ms. and maximum buffer size $B_s = 4.0 \times u(n_{i,j})/(F_r)$ represents a delay time of 400 ms.
- 6) H.264 reference software version JM 8.6 with main profile is used in the simulation for performance comparison purpose.

Table 1 shows the average PSNR, standard deviation of PSNR, number of frames skipped, bit rate, processing time, PSNR gain, and percentage of frames skipped reduction between H.264 JM 8.6 rate control and our proposed scheme encoded at the bit rate of 16-256 kbps with delay constraint of 100 ms. On average, the proposed scheme can achieve average PSNR improvement up to 0.69 dB, 1.36 dB, 0.71 dB, 1.68 dB and 0.80 dB, respectively for the bit rate of 16, 32, 64, 128 and 256 kbps. On average, our proposed scheme can encode video with more uniform quality, as can be seen from lower standard deviation of PSNR compared to JM 8.6. Our proposed scheme can also encode video with better motion continuity, as can be seen from the reduction of the numbers of frames skipped for up to 100%. In term of bit rate used, our proposed scheme which is based on Cauchy model can achieve more accurate bit rate, i.e., closer to target bit rate, than that of JM 8.6. Our proposed scheme also achieves lower processing time on average compared to JM 8.6. Table 2 shows the average PSNR, standard deviation of PSNR, and number of frames skipped, bit rate, processing time, PSNR gain, and percentage of frames skipped reduction between H.264 JM 8.6 rate control and our proposed scheme encoded at the bit rate of 16kbps with delay constraints of 100, 150, 200, and 400 ms, respectively. In the case of varying delay times, it can be shown obviously that our proposed scheme also outperforms JM 8.6 in every aspect. Figure 2 shows the buffer fullness level and corresponding PSNR of selected test video sequences at selected bit rates ranging from Akiyo (16 kbps), Akiyo (32 kbps), Foreman (64 kbps), Claire (128 kbps), and Claire (256 kbps) at the delay constraint of 100 ms. With better bit allocation and better control of buffer occupancy level, our proposed scheme can encode videos with much less frame skipping and with better PSNR.

8. Conclusions

In this paper, we investigate and analyze the use of Cauchy distribution as a model estimate for rate and distortion characteristics for video coding on the application of rate control. Based on Cauchy R-D optimization model, we derive an expression for optimal quantization step size and a linear prediction rate and distortion model parameters. We then propose a rate control scheme using Cauchy R-D optimization model under low delay constraint. Under low delay constraint, it has been shown that the effect of encoder buffer fill-up causes frame skipping and lower PSNR. Our proposed rate control scheme also takes this factor into account. The consideration of bit allocation involves the buffer status, the number of bits used in previous frame, the complexity of basic unit. The simulation results show that our proposed scheme achieves better performance in terms of better PSNR, lower PSNR standard deviation, less frame skipping, more accurate bit rate used, and lower processing time compared to that of H.264 JM 8.6 for the ranges of low bit rate and low delay constraints indicated.

Acknowledgments

This research is in part supported by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University.

References

1. ISO/IEC JTC1/SC29/WG11, Test Model 5, 1993
2. Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, 1995.
3. T. Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 19-31, Feb. 1997.
4. "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264/ISO/IEC 14 496-10 AVC," in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.
5. T.Berger, "Rate Distortion Theory," Englewood Cliffs, NJ: Prentice-Hall, 1984.

6. C.E.Shannon, "A mathematical theory of communications," *Bell System Tech. Journal*, vol.27, pp.397-423; 623-656, July and Oct. 1948.
7. T.M.Cover and J.A.Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
8. R.M.Gray, *Source Coding Theory*. Norwell, MA: Kluwer, 1990.
9. Stanislaw H. Zak and Edwin K.P. Chong, *An introduction to optimization*. New York: Wiley, 2001.
10. R.E.Bellman, *Dynamic programming*. Princeton, NJ: Princeton University Press, 1957.
11. G.D.Forney, "The Viterbi algorithm," *Proc. IEEE*, vol.61, pp.268-278, Mar.1973.
12. W.K.Pratt, *Digital Image Processing*. New York: Wiley, 1978, ch.10.
13. F. Muller, "Distribution shape of two-dimensional DCT coefficients natural images," *Electron. Lett.*, vol. 29, no. 22, pp. 1935-1936, Oct. 1993.
14. S. R.Smooth and R.A.Lowe, "Study of DCT coefficients distributions," in *Proc. SPIE*, Jan. 1996, pp.403-411.
15. T.Eude, R. Grisel, H. Cherifi, and R. Debrie, "On the distribution of the DCT coefficients," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol.5, Apr. 1994, pp. 365-368.
16. L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "Cubic spline approximation of rate and distortion functions for MPEG video," *Vis. Commun. Image Process.*, Mar. 1996.
17. Y.Altunbasak and N.Kamaci, "Frame bit allocation for the H.264/AVC video coder via Cauchy-Density -Based rate and Distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, No. 8, Aug. 2005.
18. R.J.Freund and W.J.Wilson, "Regression Analysis: Statistical Modeling of a Response Variable," New York: Academic, 1998, pp. 39-41.
19. J.Weï and B.H.Soong, "A new Rate - Distortion Model for Video Transmission Using Multiple Logarithmic Functions," *IEEE Signal Processing Letters*, vol.11, No. 8, Aug, 2004
20. I.H.Shin and Y.L.Lee, "Rate control using linear rate ρ -model for H.264," *Signal processing: Image Communication* 19(2004) 341-352.
21. N. Eiamjumrus and S. Aramvith, "Cauchy Based Rate-Distortion Optimization Model for H.264 Rate Control," *Proceeding of IEEE Asia Pacific Conference on Circuits and Systems (APCCAS) 2006*, Singapore, December 2006.
22. N. Eiamjumrus and S. Aramvith, "New Rate-Control Scheme based on Cauchy Rate-Distortion Optimization Model for H.264 Video Coding," *Proceeding of IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2006*, Japan, December 2006.



Nongluk Eiamjumrus received the B.S. degree in Electrical Engineering from Songkla University, Thailand in 1998. She received the M.S degree in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand in 2002. She is currently working to toward the Ph.D. degree in electrical engineering at Chulalongkorn University. Her research interests is in the field of rate control video coding.



Supavadee Aramvith received the B.S. (first class honors) degree in computer science from Mahidol University, Bangkok, Thailand, in 1993. She received the M.S. and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, USA, in 1996 and 2001, respectively. She is currently an Assistant Professor at Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand. Her research interests include video object segmentation, detection, and tracking, rate-control for video coding, joint source-channel coding for wireless video transmissions, and image/video retrieval techniques. Dr. Aramvith is a Senior Member of IEEE.

จุฬาลงกรณ์มหาวิทยาลัย

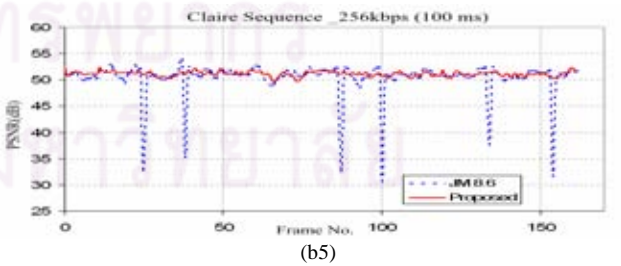
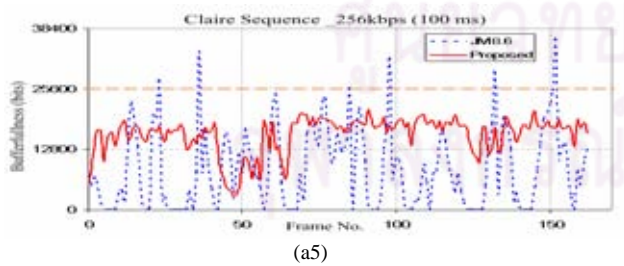
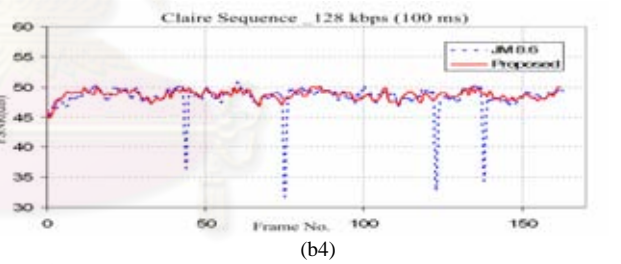
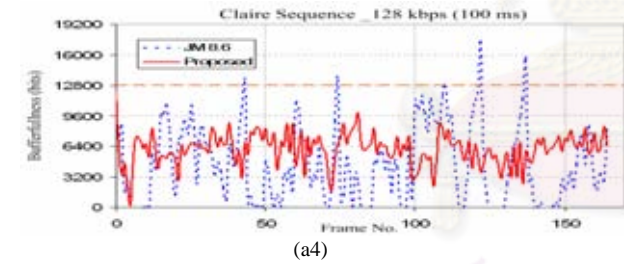
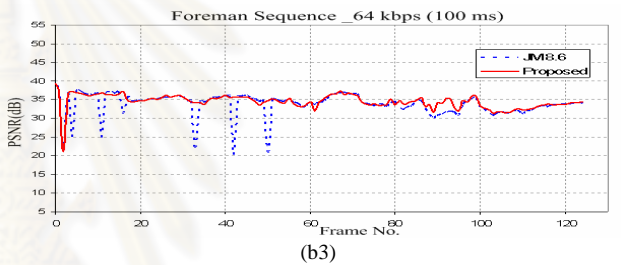
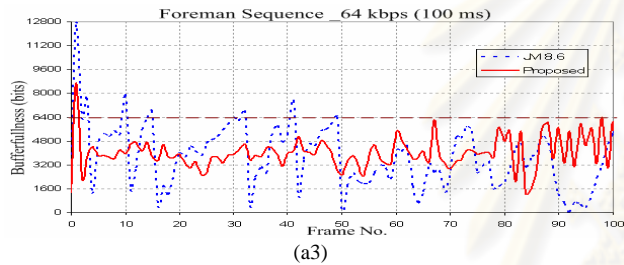
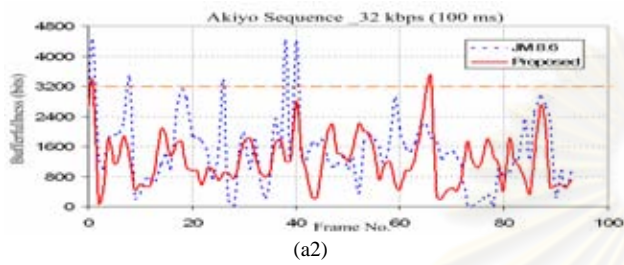
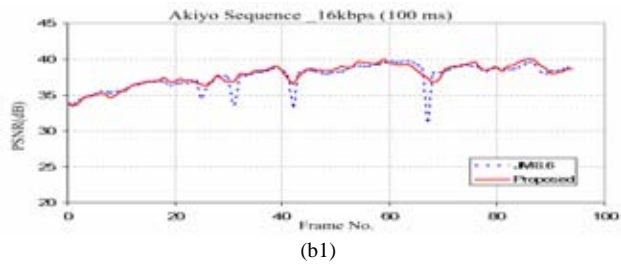
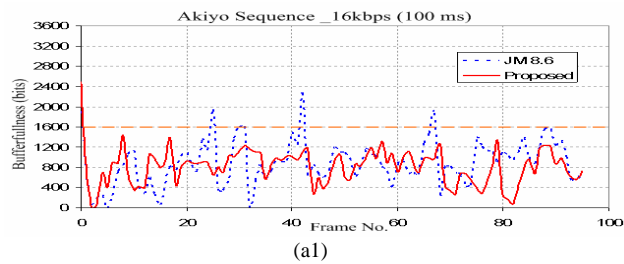


Figure.2 Simulation using test video sequences coded by H.264 JM8.6 rate control and our proposed scheme at 16-256 kbps with a delay constraint of 100 ms.
 (a) Buffer fullness level (bits) at each frame with maximum buffer size (dashed line).
 (b) Average PSNR of each frame.

Table 1: Performance comparison between our proposed scheme and H.264 JM8.6 rate control with a delay constraint of 100 ms.

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.		No. frame skipped		Bit Rate (kbps)		Processing Times (ms)		PSNR Gain (dB)	% of frame skipped reduction
		JM 8.6	Proposed	JM8.6	Proposed	JM 8.6	Proposed	JM 8.6	Proposed	JM 8.6	Proposed		
16 kbps	Akiyo	37.59	37.80	1.69	1.53	4	0	16.04	16.04	271.02	227.37	+0.21	100%
	Claire	38.02	38.37	2.24	1.65	16	4	15.98	16.04	218.89	216.43	+0.35	75%
	Silent	32.47	32.83	1.82	1.25	12	5	16.06	16.05	239.15	219.51	+0.36	58%
	News	30.52	30.86	1.22	0.98	5	4	15.97	16.06	288.47	222.68	+0.34	20%
	Carphone	29.01	29.70	3.07	2.71	13	5	16.02	16.01	219.74	220.45	+0.69	62%
	Foreman	26.73	27.40	3.45	2.55	17	9	15.53	16.03	252.00	240.30	+0.67	47%
32 kbps	Akiyo	41.73	42.08	1.81	1.37	5	2	32.00	32.09	218.92	222.70	+0.35	60%
	Claire	41.86	42.32	2.01	1.36	4	1	31.90	32.07	219.35	215.04	+0.46	75%
	Silent	38.60	39.96	2.95	1.66	11	2	31.32	32.04	222.93	220.48	+1.36	82%
	News	39.25	39.70	2.41	2.01	5	3	31.87	32.28	213.38	214.75	+0.45	40%
	Carphone	31.87	32.51	3.98	3.65	10	6	31.91	32.25	220.50	209.87	+0.64	40%
	Foreman	30.14	31.04	3.88	3.39	11	6	30.70	31.04	226.85	227.46	+0.90	45%
64 kbps	Akiyo	45.23	45.52	1.36	1.44	0	0	63.95	63.52	228.78	222.31	+0.29	-
	Claire	45.37	45.81	1.58	1.18	1	0	63.46	63.69	219.37	218.89	+0.44	100%
	Silent	36.96	37.39	3.34	2.47	9	4	64.09	64.14	223.89	221.53	+0.43	56%
	News	38.77	39.24	1.22	0.83	1	0	63.70	63.73	249.58	230.12	+0.47	100%
	Carphone	36.28	36.88	3.79	3.13	7	3	63.40	64.16	221.85	223.10	+0.60	57%
	Foreman	33.84	34.55	3.09	1.89	7	1	63.95	64.09	231.13	225.69	+0.71	86%
128 kbps	Akiyo	48.98	49.27	1.25	1.26	0	0	127.66	127.35	225.53	212.12	+0.29	-
	Claire	48.24	48.94	2.47	0.79	4	0	124.24	127.84	220.16	222.63	+0.70	100%
	Silent	43.89	44.57	2.48	0.82	2	0	127.92	128.70	213.40	215.02	+1.68	100%
	News	44.94	45.50	1.73	0.68	1	0	127.40	128.01	237.98	236.49	+0.69	100%
	Carphone	38.94	39.31	3.82	3.16	3	1	126.53	128.90	212.21	210.45	+0.37	67%
	Foreman	37.50	37.88	1.95	1.66	1	0	127.57	127.66	222.47	209.83	+0.38	100%
256 kbps	Akiyo	52.92	53.72	2.28	1.09	2	0	253.45	256.26	213.00	214.14	+0.80	100%
	Claire	50.41	51.13	3.49	0.47	6	0	245.20	253.59	222.60	220.77	+0.72	100%
	Silent	47.58	48.06	2.62	1.11	1	0	256.33	256.45	221.62	215.63	+0.48	100%
	News	48.67	49.29	2.13	2.17	1	1	255.86	255.15	236.46	230.41	+0.62	-
	Carphone	42.82	43.61	4.78	3.47	5	2	253.5	257.99	222.10	219.54	+0.79	60%
	Foreman	41.07	41.86	3.33	2.50	3	1	254.92	255.91	212.83	218.67	+0.79	67%

Table 2: Performance comparison between our proposed scheme and H.264 JM8.6 rate control at bit rate of 16 kbps for delay constraints of 100, 150, 200, and 400 ms.

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.		No. of frame skipped		Bit Rate (kbps)		Processing Times (ms)		PSNR Gain (dB)	% of frame skipped reduction
		JM 8.6	Proposed	JM8.6	Proposed	JM 8.6	Proposed	JM 8.6	Proposed	JM 8.6	Proposed		
100 ms	Akiyo	37.59	37.80	1.69	1.53	4	0	16.04	16.04	271.02	227.37	+0.21	100%
	Claire	38.02	38.37	2.24	1.65	16	4	15.98	16.04	218.89	216.43	+0.35	75%
	Silent	32.47	32.83	1.82	1.25	12	5	16.06	16.05	239.15	219.51	+0.36	58%
	News	30.52	30.86	1.22	0.98	5	4	15.97	16.06	288.47	222.68	+0.34	20%
	Carphone	29.01	29.70	3.07	2.71	13	5	16.02	16.01	219.74	220.45	+0.69	62%
	Foreman	26.73	27.40	3.45	2.55	17	9	15.53	16.03	252.00	240.30	+0.67	47%
150 ms	Akiyo	37.62	37.93	1.36	1.29	0	0	16.35	16.18	267.72	242.04	+0.31	-
	Claire	38.39	38.69	1.93	1.08	7	1	16.44	16.52	223.56	216.86	+0.30	86%
	Silent	30.40	30.75	1.06	0.62	3	1	16.59	16.25	230.68	215.41	+0.35	67%
	News	30.68	30.91	0.87	0.90	1	1	16.59	16.27	279.87	261.80	+0.23	-
	Carphone	29.10	29.44	3.29	2.74	11	1	16.57	16.26	209.07	210.75	+0.34	90%
	Foreman	27.47	27.78	2.87	2.58	10	3	16.60	16.35	279.03	245.92	+0.31	70%
200 ms	Akiyo	37.69	37.89	1.34	1.36	0	0	16.49	16.31	261.19	256.44	+0.20	-
	Claire	38.46	38.86	1.58	0.90	4	0	16.26	15.99	219.55	220.84	+0.40	100%
	Silent	30.49	30.90	1.00	0.67	3	1	16.87	16.39	245.90	231.24	+0.41	67%
	News	30.75	31.05	1.12	0.83	2	0	16.58	16.53	217.04	220.12	+0.30	100%
	Carphone	29.50	29.76	2.90	2.80	5	0	16.67	16.37	238.47	242.85	+0.26	100%
	Foreman	27.84	28.25	2.25	1.95	3	1	16.79	16.56	288.65	265.81	+0.41	67%
400 ms	Akiyo	37.72	38.02	1.31	1.32	0	0	16.66	16.49	229.63	217.89	+0.30	-
	Claire	38.47	38.88	1.00	0.95	1	0	16.51	16.29	254.75	241.08	+0.41	100%
	Silent	30.51	30.96	0.69	0.50	0	0	16.95	16.56	246.90	225.23	+0.45	-
	News	30.81	31.29	0.79	0.76	0	0	16.94	16.52	256.84	239.09	+0.48	-
	Carphone	29.50	29.97	2.95	2.68	2	1	16.87	16.66	214.87	215.56	+0.47	50%
	Foreman	27.84	28.15	2.23	1.82	1	0	16.90	16.86	289.11	290.82	+0.31	100%

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

New rate control Scheme based on Cauchy Rate-Distortion Optimization Model for H.264 Video Coding

Nongluk Eiamjumrus and Supavadee Aramvith

Department of Electrical Engineering, Faculty of Engineering
Chulalongkorn University, Bangkok 10330, Thailand
Tel : (66-2)218-6909, Fax: (66-2)218-12
Email: Supavadee.A@chula.ac.th

Abstract – Base on the observation that Cauchy distribution provides accurate estimates of rate and distortion characteristics of video sequences, in this paper, we propose a new rate control scheme based on Cauchy based rate-distortion optimization model for the application of H.264 bit allocation. One solution which has been proposed in this paper uses the Langrange Multiplier technique as the cost function to find the rate and distortion model subject to the target bit rate constraint resulting in the optimum choice of quantization step sizes. Model parameters are estimated using statistical linear regression analysis. Accordingly we then propose a simple rate control scheme using this new Cauchy rate-distortion model. The target number of bit for each frame is determined according to their buffer status, combined with the number of bits use in the previous frame. The technique proposed has been implemented in H.264 video encoder. Experimental results showed that the proposed rate control algorithm achieves an improvement of average PSNR with smooth video quality compared with the H.264 JM8.6 rate control.

I. INTRODUCTION

In real time video coding and transmission, rate control is a vital component in video encoder to assign suitable number of bits to each video frame and to ensure the generation of constant bit rate stream into channel. Due to its importance, rate control algorithms have always been challenging research issues. Several rate control schemes in various international video coding standards such as TM5[1] for MPEG-2, TMN8[2] for H.263, VM-8 [3] for MPEG-4 and JM for H.264 [4] have been proposed.

All of video coding standards use discrete cosine transform (DCT) or Integer transform in H.264, motion compensated prediction, quantization, zig-zag scan and variable length coding (VLC) as the building blocks. The knowledge of the statistical behavior in term of distribution of the DCT-coefficients is important to the design of the encoder algorithms. From there, the rate model is proposed to find optimal or sub-optimal choice of quantization parameter. Several studies on the statistical probability distribution of the AC-coefficients have been studied. Among several, Laplacian distribution [5-6] is widely used in practice with different rate models for each video coding standard, for example, TMN8 rate model in H.263 [7] and quadratic model in H.264 [4]. Cubic spline model [8] in MPEG-2 has shown more accurate

estimation rate characteristics of video sequence, but it requires very high computational complexity. Recently the work in [9] presented the observation that Cauchy distribution provides more accurate estimates of the statistical distribution of DCT coefficients in typical video sequence than that of Laplacian distribution. They also proposed frame bit allocation using Cauchy rate and distortion models for H.264 encoder.

In this paper, we propose to use Cauchy based rate-distortion (R-D) optimization model for H.264/AVC bit allocation. We model rate and distortion based on Cauchy density with approximated version as a function of quantization step sizes. We further use Langrange multiplier technique to obtain the optimal solution of quantization step sizes with minimum distortion subject to rate constraint. In this work, we propose a simple rate control scheme using this new Cauchy rate-distortion model. The target number of bit for each frame is determined according to their buffer status, combined with the number of bits use in the previous frame. Required model parameters are updated in real-time using linear regression analysis.

The paper is organized as follows. In section II, background information about Cauchy rate and distortion model are shown. In section III, we derive Cauchy rate-distortion models by using Langrange multiplier technique to find the optimal quantization step sizes that minimize distortion subject to the target bit budget. In section IV, a linear prediction rate and distortion model parameters are explained by linear regression analysis [10]. In section V, the proposed rate control using Cauchy rate-distortion model is presented. In section VI, we show the experimental results in compared to JM8.6 rate control. Section VII, is the conclusions.

II. R-D MODEL BASED ON CAUCHY DENSITY FUNCTION

Assume that, the DCT-coefficients of the motion compensated difference frame are Cauchy distribution with Cauchy parameter μ , the empirical entropy of the quantized DCT coefficients is given by eq.(1) and distortion due to quantization can be estimated accurately for the Cauchy pdf assumption as shown in eq.(3) [9],

$$H(Q) = -2 \sum_{i=1}^{\infty} \frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4) Q^2} \right) \times \log_2 \left(\frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4) Q^2} \right) \right) \quad (1)$$

$$D(Q) = \frac{1}{\pi} \sum_{i=1}^{\infty} \left\{ -\mu Q + i\mu Q \ln \left(\frac{\mu^2 + (i + \frac{1}{2})^2 Q^2}{\mu^2 + (i - \frac{1}{2})^2 Q^2} \right) + \left(\mu^2 - i^2 Q^2 \left(i^2 - \frac{1}{4} \right) \right) \right\} \quad (2)$$

where $H(Q)$ is the empirical entropy, Q is the quantization step size.

Entropy and distortion model can be approximated in term of linear expression, as shown in eqs.(3) and (4),

$$R = H(Q) = aQ^{-\alpha} \quad (3)$$

$$D(Q) = bQ^{\beta} \quad (4)$$

, where a, α and b, β are model parameters. Their values need to be estimated and depend on value of μ .

III. GENERALIZED CAUCHY R-D OPTIMIZATION MODEL USING LAGRANGE MULTIPLIER TECHNIQUE

In this section, we derive an expression for the quantization step sizes that minimize distortion subject to bit rate constraint. From eq.(3), the expected number of bits for each basic unit l^{th} in a frame is defined as in eq. (5),

$$R_l = Ca_l Q_l^{-\alpha_l} + H_l \quad (5)$$

, where C is the number of pixel in each basic unit, H_l is the actual number of header bits generated by each basic unit in the current frame, R_l is the number of target bits, and Q_l is the quantization step size for each basic unit.

The distortion measure for encoding every basic unit in each frame is defined by eq.(6),

$$D = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} \quad (6)$$

, where N is number of basic unit in each frame.

An expression for the quantization step sizes $Q_1^*, Q_2^*, \dots, Q_N^*$ that minimizes the distortion subject to the constraint that the total number of bits, i.e., the sum of the total number of bits of every basic unit in each frame must be equal to R_{MAX} , is shown in eq.(7).

$$Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^* = \arg \min_{\substack{Q_1, \dots, Q_N \\ \sum_{l=1}^N R_l = R_{MAX}}} \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} \quad (7)$$

By using Lagrange multiplier technique, the expression for the optimum choice of quantization step size $J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*)$, can be formulated as shown in eq.(8). The minimum values can be solved by using partial derivatives,

i.e., $\frac{\partial J}{\partial Q_l^*} = 0$ and $\frac{\partial J}{\partial \lambda^*} = 0$. As a result, the expression for the

optimum quantization step size for each basic unit l^{th} , Q_l^* can be found in terms of Cauchy rate and Cauchy distortion model parameters (a_l, α_l and b_l, β_l), as shown in Eq.(9).

$$J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*) = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} + \lambda \sum_{l=1}^N (Ca_l Q_l^{-\alpha_l}) + \lambda (NH - R_{MAX}) \quad (8)$$

$$Q_l = \left(\varepsilon_l \left(\sum_{l=1}^N (a_l (\varepsilon_l)^{-k}) \frac{C}{R_{MAX} - NH} \right)^{\frac{1}{k}} \right)^{s_l} \quad (9)$$

, where $\varepsilon_l = \left(\frac{a_l \alpha_l}{\beta_l b_l} \right)$, $s_l = \frac{1}{(\alpha_l + \beta_l)}$ and k is constant

value = $\frac{\alpha_{P_initial}}{\alpha_{P_initial} + \beta_{P_initial}}$ (See $\alpha_{P_initial}, \beta_{P_initial}$ in

section V).

IV. A LINEAR PREDICTION RATE AND DISTORTION MODEL PARAMETER

In this section, we relate Eqs.(3) and (4) expression in term of linear equation to find the Cauchy rate-distortion model parameters (a_{j-1}, α_{j-1} and b_{j-1}, β_{j-1}) of each basic unit by using statistic linear regression analysis.

At the end of encoding each basic unit, the encoder collects the bit rate, mean square error (MSE) and quantization step size for each type of picture. Then, the Cauchy rate model parameters can be found from eqs.(3)-(4) by using logarithmic linear function as shown in eqs.(10), (11)

$$\ln(R_l(n_{i,j})) = \ln(a_{j-1}) - \alpha_{j-1} \ln(Q_l(n_{i,j})) \quad (10)$$

$$\ln(MSE_l(n_{i,j})) = \ln(b_{j-1}) + \beta_{j-1} \ln(Q_l(n_{i,j})) \quad (11)$$

In our experiment, we use linear regression analysis to model the relationship between two variables by fitting a linear equation to the observed data. In this work, we use the formula shown in eqs. (12)-(15) to find the Cauchy rate - distortion model parameters,

$$\ln(a_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(R_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (12)$$

$$\alpha_{j-1} = \frac{\left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} (\ln(R_k)) \right) - (l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (13)$$

$$\ln(b_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(MSE_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (14)$$

$$\beta_{j-1} = \frac{(l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right) - \left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} (\ln(MSE_k)) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (15)$$

, where R_k , Q_k and MSE_k denotes the actual number of bits used for coding, quantization step size and the actual mean square error in the previously encoded basic unit k to basic unit $l-1$.

V. PROPOSED RATE CONTROL SCHEME FOR CAUCHY R-D OPTIMIZATION MODEL

In this section, the quantization step size derived previously shown in eq.(9) and Cauchy R-D model are used to design a new rate control scheme.

A. Initialization

We consider initialization of the Cauchy rate – distortion model parameter to find the optimal quantization parameter for the first intra picture $QP_{I_initial}$ as follows,

Step 1) We encode the first P-frame at different quantization parameters (QP = 3,9,15,21,27,33,39,45,51) and collect the number of bits used and mean square error (MSE) for each QP.

Step 2) Use the linear regression analysis in eqs.(12)-(15) with data acquired from the first step to compute the initial Cauchy rate–distortion model parameters ($\alpha_{P_initial}$, $\alpha_{P_initial}$ and $\beta_{P_initial}$, $b_{P_initial}$).

Step 3) Compute the initial quantization parameter for the first intra picture $QP_{I_initial}$, as shown in eq.(16),

$$Q = \left(\frac{R / F_r}{\alpha_{P_initial}} \right)^{\frac{-1}{\alpha_{P_initial}}} \quad (16)$$

,where R , F_r denotes bit rate (kbps) and frame rate (fps)

B. GOP layer

We assume that video sequence is encoded the first frame as an I-frame and subsequent P-frames in each GOP. After each frame are encoded, the occupancy of virtual buffer are updated before coding the j^{th} frame, as shown in eq.(17) [4],

$$B_c(n_{i,j+1}) = B_c(n_{i,j}) + b(n_{i,j}) - \frac{u(n_{i,j})}{F_r} \quad (17)$$

,where $u(n_{i,j})$ is the target channel bandwidth, $b(n_{i,j})$ is the number of bits generated by the j th frame in i th GOP, F_r is the predefined frame rate and $B_c(n_{i,j})$ denotes the buffer occupancy after coding the j th frame

C. Frame layer

The objective of this stage is to determine the number of target bits for each P-frame in the two following steps.

Step1) The bit allocation is implemented by considering the current buffer fullness. In this case, we can rewrite Eq.(18) in new form as follows,

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + \eta V_s - B_c(n_{i,j}) \quad (18)$$

, where $f(n_{i,j})$ and V_s are the target bit rate and the buffer size, i.e., $2.56 \times \text{bit_rate}$, respectively. η denotes the constant value (from the experiments, we set $\eta = 0.2$).

Step 2) Compute the target bit rate by taking into consideration the number of bits use in the previous frame, as in Eq.(19)

$$\hat{f}(n_{i,j}) = \left(\frac{u(n_{i,j})}{F_r} \times N_p \right) - \sum_{j=1}^{j-1} b(n_{i,j}) \quad (19)$$

, where N_p is the number of encoded P-frame.

The target number of bit is then updated as in eq. (20).

$$f(n_{i,j}) = (\tilde{f}(n_{i,j}) + \hat{f}(n_{i,j})) \quad (20)$$

D. Basic unit layer

In H.264, a basic unit is defined as a group of consecutive number of macroblocks. In this paper, we define a basic unit as a macroblock.

1) Pre-encoding stage

The objective is to compute the quantization step size of current basic unit in three cases :

Case 1 For the first basic unit in current frame, quantization parameter is obtained, as shown in eq. (21),

$$QP_{1,j}(j) = \text{Avg}QP_i(j-1) \quad (21)$$

, where $QP_{1,j}(j)$ is the quantization parameter of the first basic unit and $\text{Avg}QP_i(j-1)$ is the average of quantization parameter for all basic unit in previous frame

Case 2 When number of remaining bits is lower than zero, quantization parameter is obtained as shown in eq. (22),

$$QP_{1,j}(j) = QP_{l-1,j}(j) + 2 \quad (22)$$

Case 3 Compute the quantization step size of the current basic unit, $Q_i(n_{i,j})$, as shown in eq. (9), where α_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} are model parameters of the basic unit l th of frame j , respectively. C is the number of pixel in basic unit and N is the number of basic unit in each frame. $m_{hdr,j}$ is the number of header bits that are generated from l th basic unit.

2) Post-encoding stage

After encoding each frame, only Cauchy rate and distortion model parameters, α_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} are updated by linear regression analysis by eqs. (12)-(15).

VI. EXPERIMENTAL RESULTS

We encoded five video sequences: Foreman, News, Carphone, Akiyo, and Coastguard using the proposed scheme

compared with the H.264 JM8.6 rate control at bit rates ranging from 16 - 256 kbps. GOP structure is used with the format IPP... We set 10 fps as frame rate. Main profile for encoding parameters was used in the experiment.

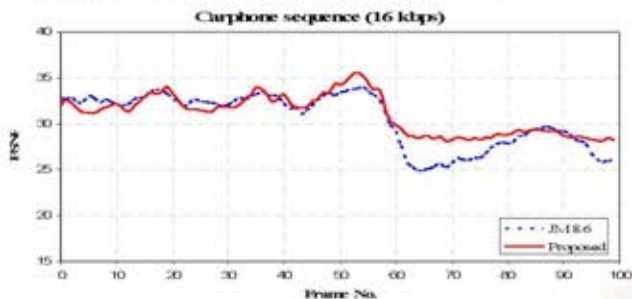


Fig. 1. PSNR (dB) result versus frame for our proposed rate control and JM8.6 in basic unit layer at bit rate 16 kbps of Carphone sequence

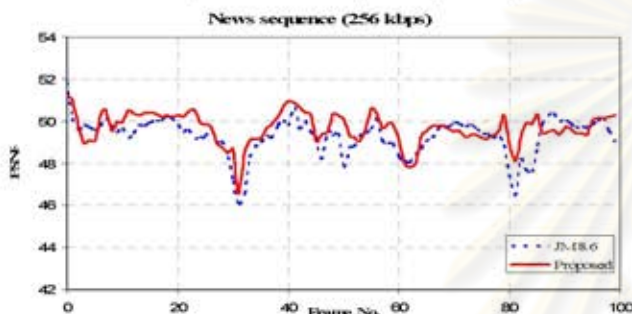


Fig.2 PSNR (dB) result versus frame for our proposed rate control and JM8.6 in basic unit layer at bit rate 256kbps of News sequence

From the results shown in Figs 1-2 and in Table 1 for our proposed scheme and JM. 8.6, the proposed scheme can achieve average PSNR improvement with more uniformly encoded, i.e. lower PSNR standard deviation, while maintain the average bit rate as the same as H.264 JM 8.6.

VII. CONCLUSION

In this paper, new rate control scheme for H.264 video coding is proposed. The Cauchy rate distortion model is realized to obtain optimum quantization step size for each video frame subject to bit rate constraint. Experimental results indicate our proposed scheme can improve overall video quality and can encode video with more uniform quality compared to H.264 JM 8.6.

ACKNOWLEDGMENT

This research is in part supported by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University and Thailand Research Fund.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, Test Model 5, 1993.
 [2] Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, 1995.
 [3] T. Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 19-31, Feb. 1997.

[4] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264/ISO/IEC 14 496-10 AVC," in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.
 [5] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," IEEE Trans. Commun., vol. COM-31, no. 6, pp. 835-839, Jun. 1983.
 [6] E.Y.Lam and J.W.Goodman, "A mathematical analysis of the DCT coefficient distributions for images," IEEE Trans. Image Process., vol. 9, no. 10, pp. 1661-1666, Oct. 2000.
 [7] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," IEEE Trans. Circuits Syst. Video Technol., vol. 9, pp. 172-185, Feb. 1999.
 [8] L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "Cubic spline approximation of rate and distortion functions for MPEG video," Vis. Commun. Image Process., Mar. 1996.
 [9] Y. Altunbasak and N. Kamaci, "Frame bit allocation for the H.264/AVC video coder via Cauchy-Density -Based rate and Distortion models," IEEE Trans. Circuits Syst. Video Technol., vol. 15, No. 8, Aug. 2005.
 [10] R.J.Freund and W.J.Wilson, "Regression Analysis: Statistical Modeling of a Response Variable," New York: Academic, 1998, pp. 39-41.

Table Performance of proposed scheme compared with JM8.6

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.	
		JM8.6	Proposed	JM8.6	Proposed
16 kbps	Foreman	29.21	29.67 (+0.46 dB)	1.85	0.80
	Coastguard	27.23	27.39 (+0.16 dB)	0.85	0.72
	Carphone	30.45	31.06 (+0.61 dB)	2.88	2.11
	Akiyo	38.00	38.27 (+0.27 dB)	1.40	0.98
	News	30.90	31.27 (+0.37dB)	0.68	0.52
32 kbps	Foreman	32.55	32.86 (+0.31 dB)	2.12	1.00
	Coastguard	29.19	29.56 (+0.37 dB)	1.53	0.83
	Carphone	33.66	34.30 (+0.64 dB)	3.25	2.21
	Akiyo	42.25	42.46 (+0.21 dB)	0.82	0.80
	News	35.62	35.91 (+0.29 dB)	1.25	0.55
64 kbps	Foreman	35.75	36.10 (+0.35 dB)	1.60	1.17
	Coastguard	31.91	32.15 (+0.24 dB)	1.09	0.92
	Carphone	37.07	37.57(+0.50 dB)	3.09	2.52
	Akiyo	45.76	46.34 (+0.58 dB)	1.30	0.98
	News	39.24	39.93(+0.69 dB)	0.74	0.57
128 kbps	Foreman	39.17	39.66 (+0.49 dB)	1.43	1.29
	Coastguard	34.97	35.10 (+0.13 dB)	1.37	0.81
	Carphone	40.93	41.40 (+0.47 dB)	2.98	2.53
	Akiyo	49.64	50.37 (+0.73 dB)	1.32	0.90
	News	43.90	44.47 (+0.57 dB)	0.72	0.73
256 kbps	Foreman	43.10	43.66 (+0.56 dB)	1.72	1.19
	Coastguard	38.86	39.39 (+0.53dB)	2.47	0.95
	Carphone	44.99	45.69 (+0.70 dB)	2.75	1.98
	Akiyo	54.40	55.27 (+0.87 dB)	1.00	0.91
	News	49.24	49.69 (+0.45 dB)	0.89	0.73

Rate control Scheme based on Cauchy R-D Optimization Model for H.264/AVC under Low Delay Constraint

N. Eiamjumrus and S. Aramvith

Department of Electrical Engineering, Faculty of Engineering
Chulalongkorn University, Bangkok 10330, Thailand
Tel : (66-2)218-6909, Fax: (66-2)218-6912
Email: supavadee.a@chula.ac.th

Abstract

Base on the observation that Cauchy distribution provides an accurate estimates of rate and distortion characteristics of video sequences. In this paper, we propose to use a new rate control scheme based on Cauchy R-D optimization model for bit allocation of H.264 for low delay constraint. One solution which has been proposed in this paper use the Lagrange Multiplier technique as the cost function to find rate and distortion model subject to target bit rate constraint resulting in optimum of quantization step sizes. Model parameters are estimated using linear regression analysis. We propose a simple rate control scheme that achieves for new Cauchy R-D model in basic unit layer under low delay constraint. The target bits for each frame are determined according to their buffer status, combined with number of bits use in the previous frame. The technique proposed has been implemented on H.264 video coder. Experimental showed that the proposed can be achieves an improvement of average PSNR with smooth video quality and less frame skipping compared to the JM8.6 rate control.

1. Introduction

In real time video coding, rate control is a vital component in video encoder to assign suitable number of bits to each frame and to ensure the generation of constant bit rate stream into channel. Due to its importance, rate control has always been challenging research issues. Several rate control schemes in various international video coding standards such as TM5[1] for MPEG-2, TMN8 [2] for H.263, VM-8 [3] for MPEG-4 and JM for H.264/AVC [4] have been proposed.

The knowledge of the statistical behavior in term of distribution of the DCT-coefficients is important to the design of the encoder algorithms. Laplacian pdf [5] is widely used in practice with different rate models for each video coding standard, for example, TMN8 rate model in H.263 [2] and quadratic model in H.264 [4]. Cubic spline model [6] requires very high

computational complexity. Recently the work in [7] presented the observation that Cauchy pdf provides more accurate estimate of statistical distribution of DCT coefficients in typical video sequence than Laplacian pdf. They also proposed frame bit allocation using Cauchy rate models for H.264 encoder but did not applicable for low delay applications.

Low delay is an important factor in real-time video transmission. In this paper, we adopt Cauchy based rate-distortion (R-D) optimization model for H.264 low delay video transmission. We model rate and distortion based on Cauchy pdf with approximated version as a function of quantization step sizes by used Lagrange multiplier technique to obtain the optimal solution with a simple Cauchy R-D based rate control scheme under low delay constraint.

This paper is organized as follows. Section 2, background information about Cauchy rate and distortion model is shown. Section 3, we derive Cauchy rate-distortion models by using Lagrange multiplier technique to find the optimal quantization step sizes. Section 4 presents the prediction of R-D model parameters. Section 5 shows impact of low delay constraint based on JM8.6 rate control. Section 6 shows the proposed Cauchy R-D based rate control scheme under low delay constraint. Section 7 shows the experimental results compared to JM8.6 rate control. Section 8 concludes this paper.

2. R-D Model based on Cauchy Density function

Based on the observation of [7], DCT-coefficients of motion compensated difference frame are Cauchy distribution. In addition, entropy and distortion model can be estimated accurately by the Cauchy pdf assumption in term of linear expression as follows,

$$R = H(Q) = aQ^{-\alpha} \quad (1)$$

$$D(Q) = bQ^{\beta} \quad (2)$$

,where $H(Q)$, $D(Q)$ is the empirical entropy and distortion due to quantization step size (Q) . a, α and b, β are model parameters depend on the Cauchy pdf.

3. Generalized Cauchy R-D Optimization model using Lagrange multiplier

In this section, we derive an expression for the quantization step sizes that minimize distortion subject to bit rate constraint. From eq.(1), the expected number of bits for each basic unit l^{th} in a frame is defined as $R_l = Ca_l Q_l^{-\alpha_l} + H_l$, where C is the number of pixel in each basic unit, H_l is the actual number of header bits generated by each basic unit in the current frame, R_l is the number of target bits, and Q_l is the quantization step size for basic unit l^{th} .

The distortion measure for encoding every basic unit in each frame is defined by $D = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l}$, where N is number of basic unit in each frame. An expression for the quantization step sizes $Q_1^*, Q_2^*, \dots, Q_N^*$ that minimizes the distortion subject to the total number of bits, i.e., sum of the total number of bits of every basic unit in each frame must be equal to R_{MAX} , is shown in eq.(3).

$$J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*) = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} + \lambda \sum_{l=1}^N (Ca_l Q_l^{-\alpha_l} + H_l - R_{MAX}) \quad (3)$$

By using Lagrange multiplier technique, the expression for the optimum choice of quantization step size $J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*)$, can be formulated by minimum values. That can be solved by using partial derivatives, i.e., $\partial J / \partial Q_l^* = 0$ and $\partial J / \partial \lambda^* = 0$. The expression for the optimum Q_l^* can be found in term of Cauchy R-D model parameters (a_l, α_l and b_l, β_l),

$$Q_l = \left(\varepsilon_l \left(\sum_{l=1}^N (a_l (\varepsilon_l)^{-k}) \frac{C}{R_{MAX} - NH} \right)^{\frac{1}{k}} \right)^{s_l} \quad (4)$$

, where $\varepsilon_l = (a_l \alpha_l) / (b_l \beta_l)$, $s_l = 1 / (\alpha_l + \beta_l)$. k is constant $= \alpha_{P_initial} / (\alpha_{P_initial} + \beta_{P_initial})$ (See detail in section 5)

4. A linear prediction rate and distortion model parameters

In this section, we use the eq.(1),(2) expression in term of linear equation to find Cauchy R-D model parameter (a_{j-1}, α_{j-1} and b_{j-1}, β_{j-1}) of each basic unit by using statistic linear regression analysis [8]. In our experiment, we use linear regression analysis that models the relationship between two variables by fitting a linear equation to observed data. We use the

formula below to find the Cauchy R-D optimization model parameters.

$$\ln(a_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(R_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (5)$$

$$\alpha_{j-1} = \frac{\left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} (\ln(R_k)) \right) - (l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (6)$$

$$\ln(b_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(MSE_k) \right) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (7)$$

$$\beta_{j-1} = \frac{(l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k) \right) - \left(\sum_k^{l-1} \ln(Q_k) \right) \left(\sum_k^{l-1} \ln(MSE_k) \right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2 \right) \left(\sum_k^{l-1} \ln(Q_k) \right)^2} \quad (8)$$

, where R_k , Q_k and MSE_k denotes the actual number of bits used for coding, quantization step size and the actual mean square error (MSE) in the previously encoded basic unit k to basic unit $l-1$.

5. Impact of low delay constraint of H.264 video quality

This section, we show the impact of low delay constraint of JM8.6 rate control. In encoding process, the bits left from previously encoded frame that has not been transmitted produce a delay in an order of few milliseconds. To reduce such delay, the encoder buffer size should be kept small. In our study, we impose a low delay constraint that the number of bits in encoder buffer $B_c(n_{i,j})$ must not be greater than the maximum buffer size M , otherwise a frame will be skipped, where, $M = V_s = u(n_{i,j}) / F_r$. Hence, the maximum buffer delay is $M / u(n_{i,j}) = 1 / F_r$ seconds. Where $u(n_{i,j})$ and F_r is target channel bandwidth and frame rate, respectively.

To show the impact of the low delay constraint of JM8.6 rate control [4], we encoded an I-frame followed by P-frame in basic unit layer for Carphone sequence at 64 kbps with frame rate of 10 fps. Then, the maximum buffer delay in our simulation is 100 ms. As shown in Fig.1 (a) and (b), when the buffer fullness level higher than maximum delay constraint, frame will be skipped.

6. Rate control scheme for Cauchy R-D model under low delay constraint

In this section, we use the expression in eq.(4) to design a new rate control scheme to operate H.264/AVC to reduces number of frame skipped with

an improvement in average PSNR and smooth video quality. Cauchy R-D models presented in the previous section can be used to find the optimal quantization step size in each basic unit of each frame.

A. Initialization

First the Cauchy R-D model parameters are initialized such that the optimal of initial quantization parameter for the first intra picture $QP_{I_initial}$ could be obtained.

Step 1) We encode the first P-frame at different QP (QP=3,9,15,21,27,33,39,45,51) and collect the number of bits used, and MSE as information of each QP.

Step 2) Apply the Least square equation in eq. (5-8). The information collected, i.e., bits used and MSE from the first step, are used to compute the initial Cauchy R-D model parameters ($\alpha_{P_initial}$, $\alpha_{P_initial}$ and $\beta_{P_initial}, b_{P_initial}$).

Step 3) Compute the initial quantization parameter for the first intra picture $QP_{I_initial}$ as in eq.(9),

$$Q_{I_initial} = \left((R / F_r) / a_{P_initial} \right)^{-1 / \alpha_{P_initial}} \quad (9)$$

,where R denotes the bit rate (kbps)

B. GOP layer

We assume that the video sequence is encoded first frame as an I-frame and follow by P-frames in each GOP. After each frame are encoded, the occupancy of virtual buffer are updated before coding the j^{th} frame as in [4] by, $B_c(n_{i,j+1}) = B_c(n_{i,j}) + b(n_{i,j}) - u(n_{i,j}) / F_r$. (where $b(n_{i,j})$ is number of bits generated by j th frame in i th GOP, F_r is predefined frame rate, $B_c(n_{i,j})$ is the buffer occupancy after coding the j th frame)

C. Frame layer

The objective of this stage is to determine number of target bits budget for P-frame $f(n_{i,j})$ in follow steps.

Step1) The bit allocation is implemented by consider the current buffer fullness. We can find target bit rate by $\tilde{f}(n_{i,j}) = u(n_{i,j}) / F_r + (\sigma \times V_s) - B_c(n_{i,j})$, where $\tilde{f}(n_{i,j})$ is defined the target bits in each frame and σ is constant (in our simulation we set $\sigma = 0.6$).

Step 2) Compute the target bit rate $\hat{f}(n_{i,j})$ by consider the number of bits use in the previous frame by

$$\hat{f}(n_{i,j}) = \left((u(n_{i,j}) / F_r) \times N_p \right) - \sum_{j=1}^{j-1} b(n_{i,j}), \text{ where } N_p \text{ is the}$$

number of encoded P-frame. Target bit budget are computed by combined with $f(n_{i,j}) = (\tilde{f}(n_{i,j}) + \hat{f}(n_{i,j}))$

D. Basic unit layer

In H.264, a basic unit is defined as a group of consecutive number of macroblocks.

1) *Pre-encoding stage*: The objective is to compute quantization step size of each basic unit in three cases:

Case1: For the first basic unit in current frame, QP is obtained by $QP_{1,i}(j) = AvgQP_i(j-1)$ (where $QP_{1,i}(j)$ is the quantization parameter of the first basic unit and $AvgQP_i(j-1)$ is the average of QP for all basic unit in previous frame).

Case2: When the number of remaining bit is lower than zero, QP is obtained by $QP_{1,i}(j) = QP_{1,i}(j) + 2$

Case3: Compute the quantization step size $Q_i(n_{i,j})$ of the current basic unit by eq.(4). To reduces the number of frame skipped, the buffer overflow should be avoided by decreasing 10 % of target number of bits estimation in each frame when buffer fullness level reaches 80%. On the other hand, the target number of bits is increased by 10% if buffer is underflow, i.e., buffer fullness are less than or equal to zero. The condition should require by $R_{MAX} = \eta \times f(n_{i,j})$, where R_{MAX} is target bit budget in current frame, η is constant value. In our simulation, we set value of η depending on the buffer fullness conditions.

$$\eta = \begin{cases} 1.10 & \text{if } B_c(n_{i,j}) \leq 0 \\ 0.90 & \text{if } B_c(n_{i,j}) \geq 0.8 \times V_s \\ 1.00 & \text{otherwise} \end{cases} \quad (10)$$

In eq.(4), we compute the quantization step size (Q_i) of the current basic unit by using the parameters a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} . To avoid the fluctuation of PSNR and buffer overflow, the target bit budget estimation of current frame are bounded by a lower and upper bounds shown in eq. (11)

$$Upper_bound_R_{MAX} = MIN \left\{ f(n_{i,j}), \left(\phi \times \frac{u(n_{i,j})}{F_r} \right) \right\} \quad (11)$$

$$Lower_bound_R_{MAX} = MAX \left\{ f(n_{i,j}), \left(\gamma \times \frac{u(n_{i,j})}{F_r} \right) \right\}$$

, where ϕ and γ are constant value and its typical value is 3 and 0.5, respectively. The computed quantization parameter of each basic unit are adjusted as in eq.(12),

$$QP_i = \begin{cases} MAX(QP_{I_initial} + 5, QP_i + 3) & , \text{if } (B_s(n_{i,j}) \geq 0.2 \times V_s) \\ MAX(QP_{I_initial} - 2, QP_i) & , \text{if } (B_s(n_{i,j}) \leq 0) \\ MAX(QP_{I_initial} - 1, QP_i) & , \text{Otherwise} \end{cases} \quad (12)$$

,where $B_s(n_{i,j})$ is defined the buffer fullness after encoding the l^{th} basic unit, that is computed from the number of bits used to encode basic unit i (B_{b-i}) as,

$$B_s(n_{i,j}) = B_c(n_{i,j}) + \sum_{i=1}^l B_b. \text{ All of constant value in eq.(12)}$$

are set from the experiment.

2) *Post-encoding stage* : After encoding each frame, only Cauchy rate and distortion model parameters a_{j-1} , α_{j-1} and b_{j-1} , β_{j-1} are updated by linear regression analysis by eqs. (5)-(8) as show in section 4 to find the optimal quantization step size of each basic unit.

7. Experimental Results

We encode four video sequences: Forman, Carphone, Akiyo and News sequence using the proposed scheme compared with the H.264 JM8.6 rate control. GOP structure are consists of only one I-frame and the other are P-frames. We set 32,64 and 128 kbps as target bit rate, 10 fps as frame rate, and buffer delay of 100 ms. Main profile for encoding parameters was used in the experiment.

Table 1 show the average PSNR, standard deviation of PSNR, number of frame skipping between H.264 JM 8.6 and our proposed scheme at 32, 64 and 128 kbps. On average, the proposed scheme can achieve average PSNR improvement 0.29, 0.33, and 0.30 dB for 32, 64 and 128 kbps with reduces the number of frame skipping than that of JM 8.6. The videos are more uniformly encoded as can be seen from reduced variation of PSNR. As shown in Fig.1, we encoded Carphone sequence at 64 kbps. It can be seen that our proposed scheme can reduce the number of frames skipped with smoother visual video quality.

8. Conclusion

In this paper, a new rate control scheme for H.264 video coding is proposed with Cauchy R-D optimization model under low delay constraint. The Cauchy rate distortion model is realized to obtain optimum quantization step size for each video frame subject to bit rate constraint by used Lagrange multiplier technique. Experimental results indicate our proposed scheme can improve overall video quality and can encode video with more uniform quality with less frame skipping compared to JM 8.6 rate control.

9. Acknowledgment

This research is in part supported by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University.

References

- [1] ISO/IEC JTCl/SC29/WG11, Test Model 5, 1993.
- [2] Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, 1995.
- [3] T. Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 19–31, Feb. 1997.
- [4] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264/ISO/IEC 14 496-10 AVC)," in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.
- [5] E.Y.Lam and J.W.Goodman, "A mathematical analysis of the DCT coefficient distributions for images," IEEE Trans. Image Process., vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
- [6] L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "Cubic spline approximation of rate and distortion functions for MPEG video," Vis. Commun. Image Process., Mar. 1996.
- [7] Y.Altunbasak and N.Kamaci, "Frame bit allocation for the H.264/AVC video coder via Cauchy-Density –Based rate and Distortion models," IEEE Trans. Circuits Syst. Video Technol., vol. 15, No. 8, Aug. 2005.
- [8] R.J.Freund and W.J.Wilson, "Regression Analysis: Statistical Modeling of a Response Variable," New York: Academic, 1998, pp. 39–41.

Table 1. Performance of proposed scheme compared with JM 8.6 rate control at 32, 64 and 128 kbps

Bit rate	Sequence	Average PSNR (dB)		PSNR Std.		No. frame skipped	
		JM	propose	JM	Propose	JM	Propose
32 kbp	Foreman	31.58	31.94	4.49	1.49	8	0
	Carphone	33.09	33.34	4.05	3.34	7	2
	Akiyo	42.38	42.68	1.49	1.18	2	1
	News	35.31	35.54	2.80	1.54	7	2
64 kbps	Foreman	35.41	35.61	2.91	1.38	3	0
	Carphone	36.98	37.12	3.38	3.20	2	0
	Akiyo	45.80	46.15	1.39	1.22	0	0
	News	38.95	39.56	2.97	0.97	4	0
128 kbps	Foreman	38.80	39.01	3.19	1.39	3	0
	Carphone	40.83	41.06	3.34	3.05	1	0
	Akiyo	49.63	49.86	1.20	0.55	0	0
	News	44.11	44.60	2.29	0.89	1	0

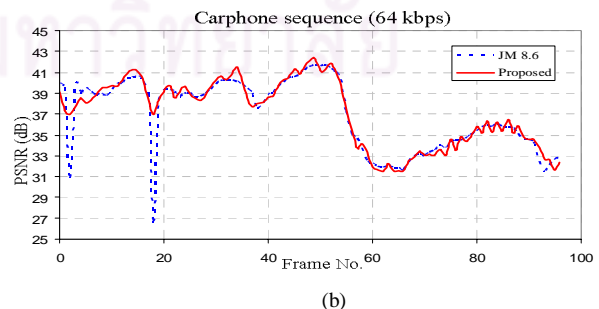
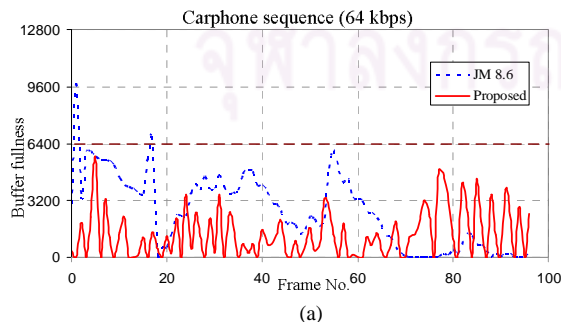


Fig.1 Simulation using the Carphone sequence coded by H.264 JM8.6 rate control at 64 kbps with delay constraint. (a) Buffer fullness level (bits) at each frame with maximum delay constraint (dashed line) . (b) Average PSNR of each frame with delay constraint.

การปรับปรุงระเบียบวิธีควบคุมอัตราสำหรับมาตรฐานการเข้ารหัสสัญญาณวิดีโอ

H.264 โดยใช้แบบจำลองอัตราบิดและความเพี้ยนแบบโคชี

Improvement of Rate Control Scheme using Cauchy Rate-Distortion

Optimization Model for H.264 Video Coding Standard

นงลักษณ์ เอี่ยมจรัส และ สุภาวดี อ่วมวิทย์

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ 10330

โทร. 0-2218-6909 โทรสาร 0-2218-6912

E-mail : Nongluk.e@student.chula.ac.th , Supavadee.A@chula.ac.th

บทคัดย่อ

จากงานวิจัยที่ได้วิเคราะห์ลักษณะการแจกแจงความน่าจะเป็นของสัมประสิทธิ์ DCT ของสัญญาณวิดีโอ พบว่าลักษณะการแจกแจงความน่าจะเป็นแบบโคชีมีความแม่นยำในการประมาณค่าอัตราบิด ความเพี้ยนของสัญญาณวิดีโอ ในบทความนี้ได้นำเสนอการควบคุมอัตราสำหรับการเข้ารหัสสัญญาณวิดีโอมาตรฐาน H.264 โดยใช้แบบจำลองทางคณิตศาสตร์ของอัตราบิดและความเพี้ยนแบบโคชี เพื่อปรับปรุงคุณภาพของสัญญาณวิดีโอให้สูงขึ้น แบบจำลองอัตราบิดและความเพี้ยนแบบโคชีที่นำเสนอ ใช้ในการคำนวณหาค่าระดับขั้นของการควอนไทซ์ที่เหมาะสม ภายใต้เงื่อนไขของอัตราบิดที่กำหนด โดยใช้เทคนิคการหาค่าเหมาะที่สุดแบบตัวคูณลากรองจ์ $J = D + \lambda R$ และใช้เทคนิคการวิเคราะห์การถดถอยแบบเชิงเส้นสำหรับคำนวณหาค่าพารามิเตอร์ของแบบจำลอง จากผลการทดลองพบว่าระเบียบวิธีที่นำเสนอสามารถปรับปรุงคุณภาพวิดีโอเฉลี่ยเพิ่มขึ้นและคุณภาพสัญญาณวิดีโอมีความราบเรียบมากขึ้นในทุกกรณีทดสอบ เมื่อเทียบกับมาตรฐานการเข้ารหัสสัญญาณวิดีโอ H.264

คำสำคัญ: การควบคุมอัตรา, เทคนิคตัวคูณลากรองจ์, เทคนิควิเคราะห์ความถดถอย, H.264

Abstract

Base on the observation that Cauchy distribution provide accurate estimates of rate and distortion characteristics of video sequences, in this paper, we propose a new rate control scheme based on Cauchy based rate-distortion optimization model for the application of H.264 bit allocation. Lagrange Multiplier technique $J = D + \lambda R$ is used to find the rate and distortion model subject to the target bit rate constraint resulting in the optimum choice of quantization step sizes. Model parameters are estimated using statistic linear regression analysis. The technique proposed has been implemented on H.264 video

encoder. Experimental results showed that the proposed scheme achieves an average PSNR improvement with more uniform video quality compared to that of H.264 JM8.6 rate control.

Keywords: Rate control, Lagrange multiplier technique, Linear regression technique, H.264

1. บทนำ

การควบคุมอัตรา (Rate control) เป็นส่วนสำคัญของตัวเข้ารหัสวิดีโอ โดยเฉพาะอย่างยิ่งในการส่งสัญญาณวิดีโอแบบระบบเวลาจริงผ่านช่องสัญญาณที่มีอัตราบิดคงที่ การจัดสรรจำนวนบิตที่เหมาะสมในแต่ละเฟรมเป็นสิ่งจำเป็นเพื่อให้อัตราบิดที่ถูกต้องมีความสอดคล้องกับแบนด์วิดท์ของช่องสัญญาณ ดังนั้นระเบียบวิธีการควบคุมอัตราได้มีการพัฒนาอย่างต่อเนื่อง โดยเฉพาะอย่างยิ่งการพัฒนาการควบคุมอัตราในมาตรฐานการเข้ารหัสสัญญาณวิดีโอต่างๆ เช่น TM5 [1] สำหรับมาตรฐาน MPEG-2 TMN8 [2] สำหรับมาตรฐาน H.263 VM-8 [3] สำหรับมาตรฐาน MPEG-4 และ JM สำหรับมาตรฐานการเข้ารหัสสัญญาณวิดีโอ H.264 [4] ซึ่งเป็นมาตรฐานที่มีประสิทธิภาพในการเข้ารหัส (Coding efficiency) สูงเหมาะสำหรับการนำไปใช้งานบนระบบโครงข่ายไร้สาย

ซึ่งงานวิจัยทางด้านระเบียบวิธีการควบคุมอัตราในอดีตที่ผ่านมา มีการศึกษาแบบจำลองทางคณิตศาสตร์ของอัตราบิด-ความเพี้ยนของสัญญาณ โดยพิจารณาจากลักษณะการแจกแจงความน่าจะเป็นของสัมประสิทธิ์ DCT ของสัญญาณวิดีโอ ซึ่งโดยส่วนใหญ่แล้วจะพิจารณาลักษณะการแจกแจงความน่าจะเป็นของสัมประสิทธิ์ DCT เป็นแบบลาปลาเซียน (Laplacian distribution) [5-6] เช่นเดียวกับในมาตรฐาน H.264 ซึ่งได้นำเสนอแบบจำลองกำลังสอง (Quadratic model) ส่วนในบทความที่ [7] ได้ทำการวิเคราะห์ลักษณะการแจกแจงความน่าจะเป็นของสัมประสิทธิ์ DCT ของสัญญาณวิดีโอ แล้วพบว่ามีลักษณะใกล้เคียงกับการแจกแจงความน่าจะเป็นแบบโคชี (Cauchy distribution) มากกว่าแบบลาปลาเซียน ดังนั้นการสร้างแบบจำลองอัตราบิด และ

แบบจำลองความเพี้ยนโดยพิจารณาจากการแจกแจงความน่าจะเป็นแบบโคชี่ จึงให้ความแม่นยำในการประมาณค่าอัตราบิดและความเพี้ยนของสัญญาณวิตทัศน์มากกว่าแบบจำลองที่อาศัยพื้นฐานของการแจกแจงความน่าจะเป็นแบบลาปลาเซียน

ดังนั้นในบทความนี้ จึงได้นำเสนอแบบแผนการควบคุมอัตราสำหรับการเข้ารหัสสัญญาณวิตทัศน์ H.264 โดยใช้แบบจำลองอัตราบิดและความเพี้ยนแบบโคชี่ เพื่อปรับปรุงคุณภาพของสัญญาณวิตทัศน์ให้สูงขึ้น ซึ่งใช้เทคนิคการหาค่าเหมาะสมที่สุดแบบตัวคูณลากรองจ์ ในการหาแบบจำลองอัตราบิดและความเพี้ยนแบบโคชี่ และคำนวณค่าพารามิเตอร์ของแบบจำลองโดยใช้เทคนิคการวิเคราะห์การถดถอยแบบเชิงเส้น เพื่อปรับปรุงคุณภาพ และความราบเรียบของคุณภาพสัญญาณวิตทัศน์ให้สูงขึ้น

2. แบบจำลองทางคณิตศาสตร์ของอัตราบิดและความเพี้ยนแบบโคชี่

ในส่วนนี้เป็นการอธิบายเทคนิควิธีการหาแบบจำลองทางคณิตศาสตร์ของอัตราบิดและความเพี้ยนแบบโคชี่ โดยใช้เทคนิคการหาค่าเหมาะสมที่สุดแบบตัวคูณลากรองจ์ เพื่อหาค่าระดับขั้นของการควอนไทซ์ที่เหมาะสม และนำเสนอวิธีการหาค่าพารามิเตอร์ของแบบจำลองโดยใช้เทคนิคการวิเคราะห์การถดถอยแบบเชิงเส้น

บทความที่ [7] ได้นำเสนอแบบจำลองอัตราบิดแบบโคชี่ และแบบจำลองความเพี้ยนแบบโคชี่ โดยประยุกต์ใช้กับการเข้ารหัสสัญญาณวิตทัศน์ มาตรฐาน H.264 สำหรับการควบคุมอัตราในระดับเฟรม โดยมีรูปแบบสมการของแบบจำลองดังสมการที่ (1) และ (2)

$$H(Q) = -2 \sum_{i=1}^{\infty} \frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4) Q^2} \right) \times \log_2 \left(\frac{1}{\pi} \tan^{-1} \left(\frac{\mu Q}{\mu^2 + (i^2 - 1/4) Q^2} \right) \right) \quad (1)$$

$$D(Q) = \frac{1}{\pi} \sum_{i=1}^{\infty} \left[-\mu Q + i \mu Q \ln \left(\frac{\mu^2 + (i + \frac{1}{2})^2 Q^2}{\mu^2 + (i - \frac{1}{2})^2 Q^2} \right) \right] + \left(\mu^2 - i^2 Q^2 \left(i^2 - \frac{1}{4} \right) \right) \quad (2)$$

โดยที่ $R(Q), D(Q)$ หมายถึงอัตราบิด และความเพี้ยนของสัญญาณในรูปฟังก์ชันของระดับขั้นของการควอนไทซ์ (Q) และ μ หมายถึงค่าพารามิเตอร์ของการแจกแจงความน่าจะเป็นแบบโคชี่

จากสมการที่ (1) และ (2) สามารถทำการประมาณค่าโดยเขียนให้อยู่ในรูปแบบความสัมพันธ์แบบฟังก์ชันเชิงเส้น ได้ดังสมการ

$$R = H(Q) = aQ^{-\alpha} \quad (3)$$

$$D(Q) = bQ^\beta \quad (4)$$

โดยที่ R หมายถึงจำนวนบิตที่ใช้ในการเข้ารหัส a, α และ b, β หมายถึงค่าพารามิเตอร์ของแบบจำลองอัตราบิด และแบบจำลองความเพี้ยนตามลำดับ ซึ่งเป็นค่าที่สอดคล้องกับค่า μ

บทความนี้นำเสนอแบบจำลองอัตราบิดและความเพี้ยนแบบโคชี่ ซึ่งเป็นการนำแบบจำลองที่ได้จากสมการที่ (3) และ (4) มาหาความสัมพันธ์ร่วมกัน โดยใช้เทคนิคการหาค่าเหมาะสมที่สุดแบบตัวคูณลากรองจ์ ซึ่งจากสมการที่ (3) จะสามารถคำนวณหาจำนวนบิตที่ใช้ในการเข้ารหัสของหน่วยพื้นฐานย่อยที่ I ได้ดังสมการที่ (5) และสามารถคำนวณหาค่าความเพี้ยนของสัญญาณได้ดังสมการที่ (6)

$$R_I = Ca_I Q_I^{-\alpha_I} + H_I \quad (5)$$

$$D = \frac{1}{N} \sum_{I=1}^N b_I Q_I^{\beta_I} \quad (6)$$

โดยที่ C และ H_I คือจำนวนจุดภาพ และจำนวนบิตสำหรับข้อมูลส่วนหัวของแต่ละหน่วยพื้นฐานย่อย ส่วน N หมายถึงจำนวนหน่วยพื้นฐานย่อยในแต่ละเฟรม

ดังนั้นในการคำนวณหาค่าระดับขั้นของการควอนไทซ์ที่เหมาะสม $J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*)$ และมีค่าความเพี้ยนของสัญญาณน้อยที่สุด ภายใต้เงื่อนไขอัตราบิตที่กำหนด โดยใช้การหาค่าเหมาะสมที่สุดแบบตัวคูณลากรองจ์ สามารถคำนวณได้จากสมการที่ (7) และ (8)

$$Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^* = \arg \min_{Q_1, \dots, Q_N} \frac{1}{N} \sum_{I=1}^N b_I Q_I^{\beta_I} \quad (7)$$

$$J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*) = \frac{1}{N} \sum_{I=1}^N b_I Q_I^{\beta_I} + \lambda \sum_{I=1}^N (Ca_I Q_I^{-\alpha_I}) + \lambda (NH - R_{MAX}) \quad (8)$$

โดยที่ λ หมายถึงตัวคูณลากรองจ์ จากนั้นทำการหาอนุพันธ์สมการที่ (8) เทียบกับค่าระดับขั้นของการควอนไทซ์ และ ตัวคูณลากรองจ์ เพื่อหาค่าระดับขั้นของการควอนไทซ์ที่เหมาะสม กล่าวคือกำหนดให้ $\frac{\partial J}{\partial Q_I^*} = 0$ และ $\frac{\partial J}{\partial \lambda^*} = 0$ ซึ่งสามารถคำนวณหาค่าระดับขั้นของการควอนไทซ์ที่เหมาะสมได้จากสมการที่ (9)

$$Q_i = \left[\varepsilon_i \left(\sum_{l=1}^N (a_l (\varepsilon_l)^{-k}) \frac{C}{R_{MAX} - NH} \right)^{\frac{1}{k}} \right]^{s_i} \quad (9)$$

โดยที่ $\varepsilon_i = \left(\frac{a_i \alpha_i}{\beta_i b_i} \right)$, $s_i = \frac{1}{(\alpha_i + \beta_i)}$ และกำหนดให้ค่าคงที่

$k = \frac{\alpha_{P_initial}}{\alpha_{P_initial} + \beta_{P_initial}}$ ซึ่งค่า $\alpha_{P_initial}, \beta_{P_initial}$ จะ

กล่าวถึงรายละเอียดในส่วนต่อไป

3. การหาพารามิเตอร์ของแบบจำลองอัตราบิดและความเพี้ยนแบบโคชี่

ในส่วนนี้เป็นการอธิบายการหาค่าพารามิเตอร์ของแบบจำลองอัตราบิดและความเพี้ยนแบบโคชี่ (a_{j-1}, α_{j-1} และ b_{j-1}, β_{j-1}) โดยใช้ข้อมูลทางสถิติของอัตราบิด และความเพี้ยนของหน่วยพื้นฐานย่อยที่ถูกเข้ารหัสไปแล้ว ซึ่งเมื่อพิจารณาสมการที่ (3) และ (4) สามารถเขียนให้อยู่ในรูปแบบฟังก์ชันเชิงเส้นได้ดังสมการที่ (9) และ (10)

$$\ln(R_i(n_{i,j})) = \ln(a_{j-1}) - \alpha_{j-1} \ln(Q_i(n_{i,j})) \quad (9)$$

$$\ln(MSE_i(n_{i,j})) = \ln(b_{j-1}) + \beta_{j-1} \ln(Q_i(n_{i,j})) \quad (10)$$

และจากเทคนิคการวิเคราะห์การถดถอยแบบเชิงเส้น [8] จะสามารถคำนวณหาค่าพารามิเตอร์ของแบบจำลองในหน่วยพื้นฐานย่อยที่ l ได้ด้วยสมการที่ (11)-(14)

$$\ln(a_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(R_k)\right) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)} \quad (11)$$

$$\alpha_{j-1} = \frac{\left(\sum_k^{l-1} \ln(Q_k)\right) \left(\sum_k^{l-1} (\ln(R_k))\right) - (l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(R_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)} \quad (12)$$

$$\ln(b_{j-1}) = \frac{\left(\sum_k^{l-1} \ln(MSE_k)\right) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) - \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)} \quad (13)$$

$$\beta_{j-1} = \frac{(l-k) \left(\sum_k^{l-1} \ln(Q_k) \ln(MSE_k)\right) - \left(\sum_k^{l-1} \ln(Q_k)\right) \left(\sum_k^{l-1} \ln(MSE_k)\right)}{(l-k) \left(\sum_k^{l-1} (\ln(Q_k))^2\right) \left(\sum_k^{l-1} \ln(Q_k)\right)} \quad (14)$$

โดยที่ R_k , Q_k และ MSE_k หมายถึง จำนวนบิตที่ใช้เข้ารหัส, ระดับชั้นของการควอนไทซ์ และ ค่าความเพี้ยนของสัญญาณในหน่วยพื้นฐานย่อยที่ถูกเข้ารหัสไปแล้ว ตั้งแต่หน่วยย่อยพื้นฐานที่ k ถึง $l-1$

4. เปรียบวิธีการควบคุมอัตราที่นำเสนอรวมกับการใช้แบบจำลองอัตราบิตและความเพี้ยนแบบโคชี

ในส่วนนี้เป็นการอธิบายระเบียบวิธีการควบคุมอัตราที่นำเสนอรวมกับการใช้แบบจำลองอัตราบิตและความเพี้ยนแบบโคชี บนมาตรฐานการเข้ารหัสสัญญาณวิดีโอ H.264

4.1 การกำหนดค่าเริ่มต้น

เป็นการคำนวณหาค่าเริ่มต้นของพารามิเตอร์แบบจำลองอัตราบิต และความเพี้ยนแบบโคชี ($\alpha_{p_initial}$, $\alpha_{p_initial}$ และ $\beta_{p_initial}$, $b_{p_initial}$) และสำหรับการเข้ารหัสภาพชนิด I เฟรมแรกของลำดับภาพ ซึ่งมีขั้นตอนการคำนวณดังนี้

ขั้นตอนที่ 1) ทำการเข้ารหัสภาพชนิด P เฟรมที่ค่าควอนไทซ์เซชันต่างๆกัน ($QP = 3,9,15,21,27,33,39,45,51$) และเก็บค่าจำนวนบิตที่ใช้ในการเข้ารหัส และค่าความเพี้ยนที่ได้จากการเข้ารหัสที่ค่าควอนไทซ์เซชันต่างๆกันไว้เป็นข้อมูลทางสถิติ

ขั้นตอนที่ 2) ใช้การวิเคราะห์การถดถอยแบบเชิงเส้น มาคำนวณหาค่าเริ่มต้นของพารามิเตอร์แบบจำลองอัตราบิตและความเพี้ยนแบบโคชี โดยใช้สมการที่ (11)-(14)

ขั้นตอนที่ 3) คำนวณหาค่าระดับชั้นของการควอนไทซ์สำหรับภาพชนิด I เฟรม จากสมการที่ (15)

$$Q = \left(\frac{R/F_r}{\alpha_{p_initial}}\right)^{\frac{-1}{\alpha_{p_initial}}} \quad (15)$$

โดยที่ R และ F_r หมายถึงอัตราบิต (kbps) และอัตราเฟรม (fps)

4.2 การควบคุมอัตราในระดับกลุ่มภาพ

กำหนดให้การเข้ารหัสเป็นแบบกลุ่มภาพที่มีภาพ I เป็นเฟรมแรก และตามด้วยภาพชนิด P และเมื่อมีการเข้ารหัสครบทุกหน่วยพื้นฐานย่อย ตัวเข้ารหัสจะทำการคำนวณหาจำนวนบิตสะสมคงเหลือในบัพเฟอร์หลังการเข้ารหัสแต่ละเฟรม $B_c(n_{i,j})$ ตามสมการที่ (16)

$$B_c(n_{i,j+1}) = B_c(n_{i,j}) + b(n_{i,j}) - \frac{u(n_{i,j})}{F_r} \quad (16)$$

โดยที่ $u(n_{i,j})$, $b(n_{i,j})$ คืออัตราบิตของช่องสัญญาณ และจำนวนบิตที่ใช้ในการเข้ารหัสเฟรมที่ j

4.3 การควบคุมอัตราในระดับเฟรม

ในขั้นตอนนี้เป็นการคำนวณหาอัตราบิตเป้าหมาย $f(n_{i,j})$ สำหรับการเข้ารหัสภาพชนิด P ซึ่งในบทความนี้จะพิจารณาจาก 2 ส่วนด้วยกันคือ พิจารณาจากจำนวนบิตสะสมคงเหลือในบัพเฟอร์หลังการเข้ารหัสแต่ละเฟรม และ จำนวนบิตที่ใช้เข้ารหัสเฟรมก่อนหน้า ดังสมการที่ (17)-(19)

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + 0.2V_s - B_c(n_{i,j}) \quad (17)$$

$$\hat{f}(n_{i,j}) = \left(\frac{u(n_{i,j})}{F_r} \times N_p\right) - \sum_{j=1}^{j-1} b(n_{i,j}) \quad (18)$$

$$f(n_{i,j}) = (\tilde{f}(n_{i,j}) + \hat{f}(n_{i,j})) \quad (19)$$

โดยที่ V_s และ N_p หมายถึงขนาดของบัพเฟอร์ และจำนวนภาพ P เฟรมที่ถูกเข้ารหัสแล้ว

4.4 การควบคุมอัตราในระดับหน่วยพื้นฐานย่อย

ในส่วนนี้เป็นการหาค่าระดับชั้นของการควอนไทซ์ของแต่ละหน่วยพื้นฐานย่อย โดยแบ่งการพิจารณาออกเป็น 3 กรณีคือ

กรณีที่ 1) สำหรับหน่วยพื้นฐานย่อยแรกของแต่ละเฟรม ค่าพารามิเตอร์การควอนไทซ์จะคำนวณหาได้ดังนี้ [4]

$$QP_{1,j}(j) = AvgQP(j-1) \quad (20)$$

โดยที่ $QP_{1,j}(j)$, $AvgQP(j-1)$ หมายถึงค่าพารามิเตอร์การควอนไทซ์ของหน่วยย่อยพื้นฐานแรกในแต่ละเฟรม และ ค่าพารามิเตอร์การควอนไทซ์เฉลี่ยของทุกๆหน่วยพื้นฐานย่อยในเฟรมที่ถูกเข้ารหัสก่อนหน้า

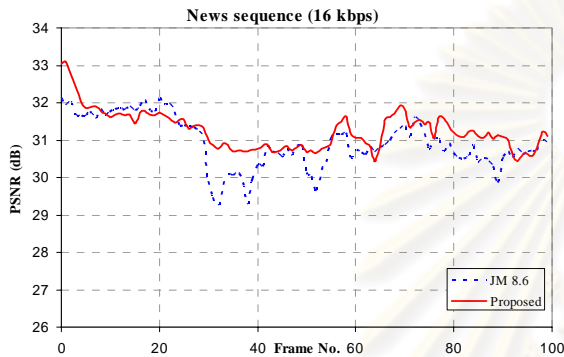
กรณีที่ 2) ถ้าจำนวนบิตเป้าหมายที่เหลืออยู่ มีน้อยกว่าศูนย์ [4] ค่าควอนไทซ์เซชันจะคำนวณหาได้จาก

$$QP_{1,j}(j) = QP_{l-1,j}(j) + 2 \quad (21)$$

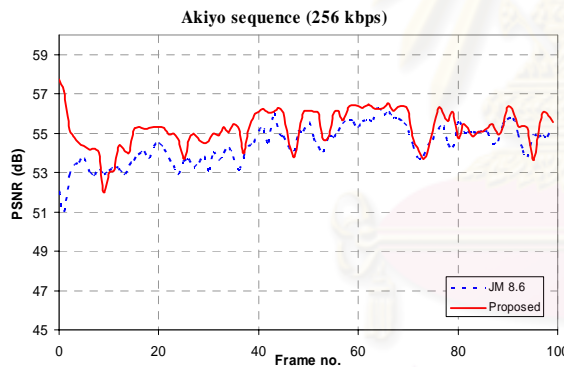
กรณีที่ 3) คำนวณหาค่าระดับชั้นของการควอนไทซ์ของแต่ละหน่วยพื้นฐานย่อยได้จากสมการที่ (9) และคำนวณหาค่าพารามิเตอร์ของแบบจำลองสำหรับแต่ละหน่วยพื้นฐานย่อยได้จากสมการที่ (11)-(14)

5. ผลการทดลอง

ทำการเข้ารหัสสัญญาณวิดีโอที่ส่ง โดยทดลองกับ 5 ลำดับภาพทดสอบ Foreman, Akiyo, Carphone, Coastguard และ News โดยมีอัตราบิตที่ 32-256 กิโลบิตต่อวินาที อัตราเฟรม 10 เฟรมต่อวินาที ซึ่งใช้การควบคุมอัตราด้วยวิธีที่นำเสนอร่วมกับการใช้แบบจำลองอัตราบิตและความเพี้ยนแบบโคซึ่ บนมาตรฐาน JM 8.6 ของ H.264 จากผลการทดลองในตารางที่ 1 แสดงให้เห็นว่าระเบียบวิธีควบคุมอัตราที่นำเสนอสามารถปรับปรุงคุณภาพสัญญาณวิดีโอที่ส่งได้สูงขึ้นโดยพิจารณาจากค่า PSNR เฉลี่ยที่สูงขึ้นในทุกลำดับภาพ รวมถึงคุณภาพสัญญาณวิดีโอที่ส่งมีความราบเรียบกว่า เมื่อพิจารณาจากค่าเบี่ยงเบนมาตรฐานของค่า PSNR ที่น้อยกว่า เมื่อเทียบกับวิธีมาตรฐาน JM 8.6



(ก) ค่า PSNR ต่อเฟรมที่อัตราบิต 16 kbps ของลำดับภาพ News



(ข) ค่า PSNR ต่อเฟรมที่อัตราบิต 256 kbps ของลำดับภาพ Akiyo รูปที่ 1 อัตราบิตต่อเฟรมของการควบคุมอัตราที่นำเสนอเทียบกับ JM 8.6

6. สรุป

บทความนี้นำเสนอระเบียบวิธีการควบคุมอัตราใหม่ของการเข้ารหัสสัญญาณวิดีโอ H.264 โดยใช้แบบจำลองอัตราบิตและความเพี้ยนแบบโคซึ่ ผลการทดลองพบว่าสามารถเพิ่มคุณภาพของสัญญาณวิดีโอที่ส่งได้สูงขึ้น และคุณภาพของสัญญาณยังมีความแปรปรวนน้อยกว่าเมื่อเปรียบเทียบกับมาตรฐานการเข้ารหัสสัญญาณวิดีโอ H.264

เอกสารอ้างอิง

- [1] ISO/IEC JTC1/SC29/WG11, Test Model 5, 1993.
- [2] Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, 1995.

- [3] T. Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp.19–31, Feb. 1997.
- [4] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264/ISO/IEC 14 496-10 AVC," in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.
- [5] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," IEEE Trans. Commun., vol. COM-31, no. 6, pp. 835–839, Jun. 1983.
- [6] E.Y.Lam and J.W.Goodman, "A mathematical analysis of the DCT coefficient distributions for images," IEEE Trans. Image Process., vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
- [7] Y.Altunbasak and N.Kamaci, "Frame bit allocation for the H.264/AVC video coder via Cauchy-Density-Based rate and Distortion models," IEEE Trans. Circuits Syst. Video Technol., vol. 15, Aug. 2005.
- [8] R.J.Freund and W.J.Wilson, "Regression Analysis: Statistical Modeling of a Response Variable," New York: Academic, 1998

ตารางที่ 1 สรุปผลการทดลอง

Bit rate	Sequence	Average PSNR (dB)		PSNR Std .	
		JM8.6	Proposed	JM8.6	Proposed
16 kbps	Foreman	29.21	29.67 (+0.46 dB)	1.85	0.80
	Coastguard	27.23	27.39 (+0.16 dB)	0.85	0.72
	Carphone	30.45	31.06 (+0.61 dB)	2.88	2.11
	Akiyo	38.00	38.27 (+0.27 dB)	1.40	0.98
	News	30.90	31.27 (+0.37dB)	0.68	0.52
32 kbps	Foreman	32.55	32.86 (+0.31 dB)	2.12	1.00
	Coastguard	29.19	29.56 (+0.37 dB)	1.53	0.83
	Carphone	33.66	34.30 (+0.64 dB)	3.25	2.21
	Akiyo	42.25	42.46 (+0.21 dB)	0.82	0.80
	News	35.62	35.91 (+0.29 dB)	1.25	0.55
64 kbps	Foreman	35.75	36.10 (+0.35 dB)	1.60	1.17
	Coastguard	31.91	32.15 (+0.24 dB)	1.09	0.92
	Carphone	37.07	37.57(+0.50 dB)	3.09	2.52
	Akiyo	45.76	46.34 (+0.58 dB)	1.30	0.98
	News	39.24	39.93(+0.69 dB)	0.74	0.57
128 kbps	Foreman	39.17	39.66 (+0.49 dB)	1.43	1.29
	Coastguard	34.97	35.10 (+0.13 dB)	1.37	0.81
	Carphone	40.93	41.40 (+0.47 dB)	2.98	2.53
	Akiyo	49.64	50.37 (+0.73 dB)	1.32	0.90
	News	43.90	44.47 (+0.57 dB)	0.72	0.73
256 kbps	Foreman	43.10	43.66 (+0.56 dB)	1.72	1.19
	Coastguard	38.86	39.39 (+0.53dB)	2.47	0.95
	Carphone	44.99	45.69 (+0.70 dB)	2.75	1.98
	Akiyo	54.40	55.27 (+0.87 dB)	1.00	0.91
	News	49.24	49.69 (+0.45 dB)	0.89	0.73

Cauchy based Rate-Distortion Optimization Model for H.264 Rate Control

N. Eiamjumrus

Department of Electrical Engineering,
Faculty of Engineering Chulalongkorn University
Bangkok 10330, Thailand, Tel: (66-2) 218-6909
Email : Nongluk.E@Student.chula.ac.th

S. Aramvith

Department of Electrical Engineering,
Faculty of Engineering Chulalongkorn University
Bangkok 10330, Thailand, Tel: (66-2) 218-6909
Email : Supavadee.A@chula.ac.th

Abstract— Base on the observation that Cauchy distribution provide accurate estimates of rate and distortion characteristics of video sequences, in this paper, we propose a Cauchy based rate-distortion optimization model for the application of H.264 bit allocation. Lagrange Optimization is used as a tool to derive rate and distortion models subject to the target bit rate constraint resulting in optimum choice of quantization step sizes. Linear regression analysis is then used to update model parameters on-line. The technique proposed has been implemented in H.264 video encoder. Experimental results showed that the proposed rate control algorithm achieves an improvement of average PSNR of up to 0.68 dB with less PSNR variation compared to H.264 JM 8.6 rate control.

Keywords—Cauchy distribution, Lagrange Multiplier, Rate Control

I. INTRODUCTION

In real time video coding and transmission, rate control is a vital component in video encoder to assign suitable number of bits to each video frame and to ensure the generation of constant bit rate stream into channel. Due to its importance, rate control algorithms have always been challenging research issues. Several rate control schemes in various international video coding standards such as TM5 [1-2] for MPEG-2, TMN8 [3] for H.263, VM-8 [4] for MPEG-4 and JM for H.264/MPEG-4 Part 10 [5-6] have been proposed.

All of video coding standards use discrete cosine transform (DCT) or Integer transform in H.264, motion compensated prediction, quantization, zig-zag scan and variable length coding (VLC) as the building blocks. The knowledge of the statistical behavior in term of distribution of the DCT-coefficients is important to the design of the encoder algorithms. From there, the rate model is proposed to find optimal or sub-optimal choice of quantization parameter. Several studies on the statistical probability distribution of the AC coefficients have been studied. Among several, Laplacian distribution [7-8] is widely used in practice with different rate models for each video coding standard, for example, TMN8 rate model in H.263 [10] and quadratic model in H.264 [11]. Cubic spline model [12] in MPEG-2 has shown more accurate estimation of rate characteristics of video sequence, but it requires very high computational complexity.

Recently the work in [9] presented the observation that Cauchy distribution provide more accurate estimate of the statistical distribution of the DCT coefficients in typical video sequence than that of Laplacian distribution. They also proposed frame bit allocation using Cauchy rate and distortion models for H.264 encoder. Nevertheless, the process to estimate model parameters has to be done offline which makes it impractical for real time video coding.

In this paper, we propose to use Cauchy based rate-distortion (R-D) optimization model for H.264 bit allocation. We model rate and distortion based on Cauchy density with approximated version as a function of quantization step sizes. We further use Lagrange optimization to obtain the optimal solution of quantization step sizes with minimum distortion for each basic unit subject to rate constraint. We implement our proposed model on H.264 rate control. Required model parameters are updated in real-time using linear regression analysis technique.

The paper is organized as follows. In Section II, background information about Cauchy rate and distortion model is shown. In section III, we derive Cauchy rate-distortion models by using Lagrange Optimization to find the optimal quantization step sizes that minimize distortion subject to the target bit budget. In section IV, the proposed Cauchy rate-distortion model based frame and basic unit layer rate control are presented. In section V, we show the experimental results in compared to JM 8.6 rate control. Section VI is the conclusions.

II. R-D MODEL BASED ON CAUCHY DENSITY FUNCTION

Assume that, the DCT-coefficients of the motion compensated difference frame are Cauchy Distributed μ , the entropy function $H(Q)$ can approximated in term of linear relation between $\ln(Q)$ and $\ln(H(Q))$ in bit per pixel given by eq.(1) [9],

$$R = H(Q) = aQ^{-\alpha} \quad (1)$$

,where a and α are model parameters. Q is the quantization stepsize.

From [9], the distortion due to quantization can be estimated accurately for the Cauchy pdf assumption. For Cauchy source, the distortion caused by quantization is given by approximation the distortion model as shown in eq. (2),

$$D(Q) = bQ^\beta \quad (2)$$

,where b and β is the model parameters.

III. GENERALIZED CACHY R-D OPTIMIZATION MODEL USING LAGRANGE MULTIPLIER

In this section, we derive an expression for the quantization step sizes that minimize distortion subject to bit rate constraint. From eq. (1), the expected number of bits for each basic unit l th in a frame is defined as follows,

$$R_l = Ca_l Q_l^{-\alpha_l} + H_l \quad (3)$$

,where C is the of pixel in each basic unit, H_l is the actual number of header bits generated by each basic unit in the current frame, R_l is the number of target bits, and Q_l is the quantization step size for each basic unit.

The distortion measure for encoded every basic unit in each frame is defined by eq. (4),

$$D = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} \quad (4)$$

, where N is the number pf basic unit in each frame.

An expression for the quantization step sizes $Q_1^*, Q_2^*, \dots, Q_N^*$ that minimizes the distortion subject to the constraint that the total number of bits, i.e., the sum of the total number of bits of every macroblock in each frame must be equal to R_{MAX} , is shown in eq. (5).

$$Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^* = \arg \min_{\substack{Q_1, \dots, Q_N \\ \sum_{l=1}^N R_l = R_{MAX}}} \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} \quad (5)$$

By using Lagrange multiplier technique, the expression for the optimum choice of quantization step size, can be formulated as shown in eq. (6) and the expression of the cost function $J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*)$ can be shown in eq. (7). The minimum value can be solve by using partial derivative, i.e., $\frac{\partial J}{\partial Q_l^*} = 0$ and $\frac{\partial J}{\partial \lambda^*} = 0$.

$$Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^* = \arg \min_{\substack{Q_1, \dots, Q_N \\ \sum_{l=1}^N R_l = R_{MAX}}} \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} + \lambda \left[\sum_{l=1}^N R_l - R_{MAX} \right] \quad (6)$$

$$J(Q_1^*, Q_2^*, \dots, Q_N^*, \lambda^*) = \frac{1}{N} \sum_{l=1}^N b_l Q_l^{\beta_l} + (\lambda C) \sum_{l=1}^N (a_l Q_l^{-\alpha_l}) + \lambda \sum_{l=1}^N H_l - \lambda R_{MAX} \quad (7)$$

As a result, the expression for the optimum quantization step size for each basic unit l^{th} , Q_l^* , can be found in terms of Cauchy rate and Cauchy distortion parameters model (a_l, α_l and b_l, β_l), as shown in eq.(8),

$$Q_l^* = \left\{ \left(\sum_{l=1}^N a_l \left(\frac{a_l \alpha_l}{\beta_l b_l} \right)^{-k} \left(\frac{C}{R_{MAX} - NH} \right)^{1/k} \left(\frac{a_l \alpha_l}{\beta_l b_l} \right)^{\frac{1}{(\beta_l + \alpha_l)}} \right)^{\frac{1}{\beta_l + \alpha_l}} \right\} \quad (8)$$

IV. H.264 RATE CONTROL USING CAUCHY R-D OPTIMIZATION MODEL

In H.264 framework [5], bit allocation process is done in three layers: Group of Picture (GOP), frame, and basic unit layers. We adopt our bit allocation framework according to H.264 with the integration of the Cauchy rate and distortion models, mentioned in Section II and III, to find the optimum quantization step size and on-line model parameter updates. We assume constant bit rate (CBR) output.

In GOP layer, the total number of bits allocated for the i th GOP and initial Quantization parameter (QP_0) for the first I-frame and the first P-frame are computed. In frame layer, the target number of bits is computed based on buffer occupancy and the remaining number of bits to encode the rest of the frame in a GOP. In basic unit layer, Cauchy R-D model is used to select the quantization step size of all basic units in each frame such that the sum of generated bits is close to the frame target. Model parameters are updated on-line using linear regression analysis.

A. GOP layer rate control

We assume that the video sequence is encoded the first frame as an I-frame and subsequent P-frames in each GOP. The total number of bits allocated for the i th GOP, $T_r(j)$, is computed as H.264 rate control [5], as shown in eq. (9),

$$T_r(j) = \begin{cases} \frac{u(n_{i,j})}{F_r} \times N_{GOP} - B_c(n_{i-1, N_{GOP}}) & \text{for } j=1 \\ T_r(n_{i,j-1}) - b(n_{i,j-1}) & \text{for } j=2, \dots, N_{GOP} + 1 \end{cases} \quad (9)$$

,where $u(n_{i,j})$ is the target channel bandwidth, $b(n_{i,j})$ is the number of bits generated by the j th frame in i th GOP, F_r is the predefined frame rate, N_{GOP} denotes the total number of frames in a GOP and $B_c(n_{i,j+1})$ denotes the buffer occupancy after coding the j th frame, as shown in eq. (10).

$$B_c(n_{i,j+1}) = B_c(n_{i,j}) + b(n_{i,j}) - \frac{u(n_{i,j})}{F_r} \quad (10)$$

The I-frame and the first P-frame of the GOP are coded by using initial quantization parameter (QP_0). This value is computed based on the channel bandwidth and GOP length, more details in [5].

B. Frame layer rate control

The objective of this stage is to determine the number of target bits for each P-frame in the following steps.

Step 1 Budget allocation among pictures. The bit allocation is implemented by predefining target buffer level, $Tb(n_{i,j+1})$ for each P-frame, as shown in eq. (11), where N_p is the number of P frames in GOP.

$$TbI(n_{i,j+1}) = TbI(n_{i,j}) - \frac{TbI(n_{i,j})}{N_p - 1} \quad (11)$$

Step 2 The target bit rate, $f(n_{i,j})$ for the j th P frame in the i th GOP is determined based on two factors. The first one, $\tilde{f}(n_{i,j})$, is scaled base on target buffer level, frame rate, channel bandwidth, and actual buffer occupancy, as shown in eq.(12), where γ is a constant weighting factor.

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + \gamma(TbI(n_{i,j}) - B_i(n_{i,j})) \quad (12)$$

The second one, $\hat{f}(n_{i,j})$, is computed based on average number of remaining bits for each frame as in eq. (13),

$$\hat{f}(n_{i,j}) = \frac{T_r(n_{i,j})}{N_p - j} \quad (13)$$

, where $T_r(n_{i,j})$ is the total number of remaining bits left to encode the j th frame onwards in the i th GOP. Finally, the target bit rate, $f(n_{i,j})$, is a weighted combination of $\tilde{f}(n_{i,j})$ and $\hat{f}(n_{i,j})$, as shown in eq. (14), where β is a constant weighting factor.

$$f(n_{i,j}) = \beta \tilde{f}(n_{i,j}) + (1 - \beta) \hat{f}(n_{i,j}) \quad (14)$$

C. Basic Unit layer rate control

In H.264, a basic unit is defined as a group of consecutive number of macroblocks. In this paper, we define a basic unit as a macroblock.

1) Pre-encoding stage

The objective is to compute the quantization step size of current basic unit in three cases :

Case 1 For the first basic unit in current frame, quantization parameter is obtained as shown in eq. (15),

$$QP_{1,j}(j) = AvgQP_1(j-1) \quad (15)$$

, where $QP_{1,j}(j)$ is the quantization parameter of the first basic unit and $AvgQP_1(j-1)$ is the average of quantization parameter for all basic unit in previous frame

Case 2 When the number of remaining bit is lower than zero, quantization parameter is obtained as shown in eq.(16),

$$QP_{1,j}(j) = QP_{1,j}(j) + 2 \quad (16)$$

Case 3 Compute the quantization step size of the current basic unit, $Q_i(n_{i,j})$, as shown in eq. (17), where α_{j-1} , β_{j-1} and β_{j-1} are model parameters of the basic unit l th of frame j th, respectively. C is the number of pixel in basic unit and N is the number of basic unit in each frame. $m_{hd,l}$ is the number of header bit s that are generated from l th basic unit. In simulation results, we set $k = 0.45$.

$$Q_i(n_{i,j}) = \left(\sum_{l=1}^N \left(\frac{\alpha_{j-1} \alpha_{j-1}}{\beta_{j-1} \beta_{j-1}} \right)^k \left(\frac{C}{f(n_{i,j}) - N m_{hd,l}} \right)^{1/k} \left(\frac{\alpha_{j-1} \alpha_{j-1}}{\beta_{j-1} \beta_{j-1}} \right)^{\frac{1}{\beta_{j-1} \alpha_{j-1}}} \right)^{1/k} \quad (17)$$

2) Post-encoding stage

After encoding each frame, only Cauchy rate and distortion model parameters α_{j-1} , α_{j-1} and β_{j-1} , β_{j-1} are updated by linear regression analysis.

Case 1 Cauchy Rate model parameters

From eq. (1), we use logarithmic linear function as shown in eq. (18).

$$\ln(R_i(n_{i,j})) = \ln(\alpha_{j-1}) - \alpha_{j-1} \ln(Q_i(n_{i,j})) \quad (18)$$

We proposed to use linear regression analysis [13] to update the model parameters after we encoded in each frame, as in eqs. (19)-(20),

$$\ln(\alpha_{j-1}) = \frac{\left(\sum_k \ln(R_k) \right) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \right)^2} \quad (19)$$

$$\alpha_{j-1} = \frac{\left(\sum_k \ln(Q_k) \right) \left(\sum_k (\ln(R_k)) \right) - (l-k) \left(\sum_k \ln(Q_k) \ln(R_k) \right)}{(l-k) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \right)^2} \quad (20)$$

, where R_k and Q_k denotes the actual number of bits used for coding and quantization step size in the previously encoded basic unit k to basic unit $l-1$.

Case 2 Cauchy Distortion model parameters

We take logarithmic linear function on eq. (3) and use distortion in terms of mean square error in each basic unit as shown in eq. (21),

$$\ln(MSE_i(n_{i,j})) = \ln(\beta_{j-1}) + \beta_{j-1} \ln(Q_i(n_{i,j})) \quad (21)$$

, where $MSE_i(n_{i,j})$ is mean square error of the basic unit l th in each frame.

As in case 1, we use linear regression analysis [13] to update the model parameters after we encoded in each frame as in eqs. (22)-(23),

$$\ln(\beta_{j-1}) = \frac{\left(\sum_k \ln(MSE_k) \right) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \ln(MSE_k) \right)}{(l-k) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \right)^2} \quad (22)$$

$$\beta_{j-1} = \frac{(l-k) \left(\sum_k \ln(Q_k) \ln(MSE_k) \right) - \left(\sum_k \ln(Q_k) \right) \left(\sum_k \ln(MSE_k) \right)}{(l-k) \left(\sum_k (\ln(Q_k))^2 \right) - \left(\sum_k \ln(Q_k) \right)^2} \quad (23)$$

, where MSE_k denotes the actual mean square error for the previously encoded basic unit k to basic unit $l-1$.

V. EXPERIMENTAL RESULT

We encoded five video sequences: Foreman, News, Carphone, Akiyo, and Coastguard using the proposed scheme compared with the H.264 JM8.6 rate control at bit rates ranging from 16 - 256 kbps. GOP structure of size 12 is used and consists of one I-frame and eleven P-frames. We set 10 fps as frame rate. Main profile for encoding parameters was used in the experiment.

Table 1 show the average PSNR, standard deviation of PSNR and bit rate comparison between H.264 JM 8.6 and our proposed scheme. On average, the proposed scheme can achieve PSNR improvement of up to 0.65 dB, 0.68 dB, 0.73 dB, 0.74 dB and 1.16 dB for the bit rates of 16, 32, 64, 128 and 256 kbps, respectively. The video is more uniformly encoded as can be seen from the reduced variation of PSNR while maintain average bit rate as the same as H.264 JM 8.6.

Fig. 1 shows PSNR of Foreman sequences at 32 kbps. It can also be seen from the results that our proposed scheme can encode video with more uniform quality.

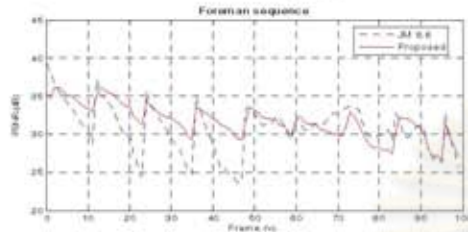


Fig. 1. PSNR (dB) result versus frame for our proposed and JM 8.6 in basic unit layer of Foreman sequence at 32 kbps

In fig.2, PSNR of foreman sequences versus bit rates are shown, it can also be seen from the results that our proposed scheme can encode video at each bit rate with higher quality in term of PSNR than that of JM 8.6.

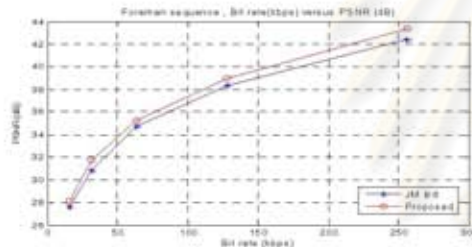


Fig.2 PSNR (dB) result versus frame for our proposed rate control and JM 8.6 in basic unit layer at each bit rate of foreman sequence.

VI. CONCLUSION

In this paper we proposed Cauchy rate - distortion optimization model for rate control for H.264 video coding. We derive models for bit rate and distortion in this type of coder by using Lagrange Optimization. The algorithm proposed is practical as all model parameters related to rate control can be estimated on-line using linear regression analysis. Experimental results indicate our proposed scheme can encode video with more uniform quality with PSNR improvement compared to H.264 JM 8.6.

ACKNOWLEDGMENT

This research is in part supported by the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University and Thailand Research Fund.

TABLE I PERFORMANCE OF PROPOSED SCHEME COMPARED WITH JM8.6

Bit rate	Sequence	PSNR (dB)		PSNR Std	
		JM 8.6	Proposed	JM 8.6	Proposed
16 kbps	Foreman	27.58	28.11	1.81	1.70
	Coastguard	26.43	26.78	1.30	1.21
	Carphone	29.10	29.63	2.70	2.70
	Alayo	35.20	36.20	1.54	1.42
	News	28.83	29.69	1.42	1.13
32 kbps	Foreman	30.81	31.71	3.16	2.27
	Coastguard	27.88	28.87	3.17	1.94
	Carphone	32.42	33.06	3.93	3.23
	Alayo	40.24	40.52	0.79	1.34
	sNew	33.42	34.00	2.66	1.78
64 kbps	Foreman	34.73	35.25	1.72	1.72
	Coastguard	31.23	31.57	2.20	2.07
	Carphone	36.11	37.07	3.45	3.11
	Alayo	43.72	44.78	1.90	2.01
	News	37.67	38.46	1.76	1.71
128 kbps	Foreman	38.31	38.96	2.04	2.35
	Coastguard	34.46	34.74	2.70	2.85
	Carphone	40.14	40.75	3.56	3.49
	Alayo	48.10	49.16	1.90	1.81
	News	42.63	43.74	1.97	1.95
256 kbps	Foreman	30.81	31.71	3.01	2.91
	Coastguard	38.12	39.52	4.90	4.90
	Carphone	44.36	45.22	3.54	3.31
	Alayo	53.31	54.84	2.22	2.35
	News	48.06	49.19	2.42	2.50

REFERENCES

- [1] ISO/IEC JTC1/SC29/WG11, Test Model 5, 1993.
- [2] The MPEG-2 International Standard, ISO/IEC, Reference Number ISO/IEC 13 818-2, 1996.
- [3] Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Recommendation H.263, ver. 1, 1995.
- [4] T. Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 19-31, Feb. 1997.
- [5] "Proposed draft of adaptive rate control," in JVT of ISO/IEC MPEG and ITU-T VCEG, JVTH014, 2003.
- [6] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 598-603, Jul. 2003.
- [7] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," IEEE Trans. Image Process., vol. 9, no. 10, pp. 1661-1666, Oct. 2000.
- [8] F. Muller, "Distribution shape of two-dimensional DCT coefficients natural images," Electron. Lett., vol. 29, no. 22, pp. 1935-1936, Oct. 1993.
- [9] Y. Altunbasak and N. Kamaci, "Frame bit allocation for the H.264/AVC video coder via Cauchy-Density-Based rate and Distortion models," IEEE Trans. Circuits Syst. Video Technol., vol. 15, No. 8, Aug. 2005.
- [10] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," IEEE Trans. Circuits Syst. Video Technol., vol. 9, pp. 172-185, Feb. 1999.
- [11] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quad-ratic rate distortion model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 246-250, Feb. 1997.
- [12] L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "Cubic spline approximation of rate and distortion functions for MPEG video," Vis. Commun. Image Proc. ess., Mar. 1996.
- [13] R. J. Freund and W. J. Wilson, "Regression Analysis: Statistical Model-ing of a Response Variable," New York: Academic, 1998, pp. 39-41.

Curriculum vitae

Nongluk Eiamjumrus was born in Ratchaburi, Thailand on August 9, 1976. She received the B.S. degree in Electrical Engineering from Songkla University, Thailand in 1998. She received the M.S degree in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand in 2002. And Ph.D. degree in electrical engineering at Chulalongkorn University in 2007. Her research interests are in the field of rate control video coding.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย