

A Spatial Scalable Video Coding with Selective Data Transmission using Wavelet Decomposition

by

Lakshmi Veerapandian

Bachelor of Engineering (Information Technology)
University of Madras, India. 2004.

Thesis submitted for the degree
of

Master of Engineering Science
in

School of Electrical and Electronic Engineering

The University of Adelaide

March 2010



© 2010
Lakshmi Veerapandian
All Rights Reserved



Table of Contents

Table of Contents	i
Abstract	iv
Statement of Originality	v
Acknowledgement	vi
Dedication	vii
Presentation.....	viii
List of Figures	ix
List of Tables	xii
Acronym	xiii
1. Introduction	1
1.1 Motivation and Problem Statement	1
1.1.1. Network Constraints	2
1.1.2. User Constraints	3
1.2. Existing Solutions and its Limitations	3
1.3. Research Objective	5
1.4. Outline of the Thesis	6
2. Background	8
2.1. Fundamentals of Video	8
2.2. Characteristics of Video	10
2.3. Network Variations	13
2.4. Video Standards	14
3. Video Coding	18
3.1. Basics of Video Coding	18
3.2. Spatial Video Transformations	19
3.2.1. Transform Coding	19
3.2.2. Subband Coding	19

3.2.3.	Wavelet Transform	20
3.2.4.	Subband/Wavelet Encoders	20
3.3.	Interframe Coding	20
3.3.1.	Three-Dimensional Coding	21
3.3.2.	Motion Compensated Coding	21
3.3.3.	Model Based Coding	21
4.	Existing Solutions	23
4.1.	Video Transcoding	23
4.2.	Scalable Solution	25
4.2.1.	Scalability in Video Standards	26
4.2.2.	Wavelet Based Scalability	28
4.3.	Overview of Our Approach	32
5.	Methods and Simulations	33
5.1.	Initial Investigations and Trials	33
5.1.1.	Initialisations of General Parameters	35
5.1.2.	Multi-level Content Extraction	35
5.1.2.1	External Source	36
5.1.2.2.	JPEG2000 Coder	36
5.1.2.3.	Wavelet Transform	38
5.1.2.3.1.	Choice of Wavelets	38
5.1.2.3.2.	Wavelet Decomposition	39
5.1.2.3.3.	Multiresolution Representation	40
5.1.2.3.4.	Quality Representations	41
5.1.2.4.	Motion Estimation	42
5.1.2.4.1.	Multiresolution Motion Estimation	44
5.1.2.5	Error Correction	48
5.1.2.5.1.	Error Identification Methods	49
5.1.2.5.2.	Error Handlers	51
5.1.3.	Scalable Framework	52

5.2. Our Approach	54
5.2.1. Simulation Assumptions	55
5.2.2. Wavelet Transform	55
5.2.3. Multiresolution Motion Estimation	55
5.2.4. Error Identification	56
5.2.5. Selective Data Transmission	58
5.2.6. Error Correction	61
6. Results and Discussions	63
6.1. Resolution Scalability	64
6.2. Error Resilience and Selective Data Transmission	67
6.2.1. Quality Threshold Constrained for Low Bandwidth	67
6.2.2. Quality Threshold Constrained for High Bandwidth	76
6.2.3. Quality Threshold Relaxed for High Quality and High Bandwidth..	82
6.3. Bandwidth Estimation	89
6.4. Computational Complexity	92
6.5. Performance Evaluation	93
6.5.1. Block Replacement Vs Motion Vector Replacement	94
6.5.2. Performance of the Three Different Modes of Selective Data Transmission	98
6.5.3. Performance in Different Motion Sequences.....	100
6.5.4. Resolution Scalability.....	102
7. Conclusion	104
Reference	106
Appendix	119
A. Haar Wavelet	119
B. Codes	120

Abstract

In this research a scalable video coding framework is proposed, mainly focusing on spatial scalability, and a subjective data compression algorithm based on: (1) quality, (2) resolution (target output device), and (3) bandwidth. This framework enables the scalable delivery of video based on the output display resolution, and through a congested network or limited bandwidth with an acceptable visual quality.

In order to achieve this scalable framework we have used wavelets, for greater flexibility, and a multiresolution approach. The multiresolution motion estimation (MRME) provides the reusability of motion vectors across different resolution levels. In MRME the motion estimation, which is carried out in the wavelet domain, is initially performed in the lower resolution and the resultant motion vectors are used as a basic motion estimate in other higher resolutions. The translation of motion vectors across different resolution levels results in translation error or mismatches. These mismatches are identified using a novel approach, which uses two thresholds. The first threshold is used to determine the possible occurrence of mismatches in a given video frame subject to the motion in the previous frame. This helps to give a broader location of all the mismatches in general. In order to specifically focus on the worst mismatches among them another threshold is used. This gives a more accurate identification of the mismatches that definitely need to be handled while the others can be waived depending upon the available resources. By varying these two parameters, the quality and resolution of the video can be adjusted to suit the bandwidth requirements. The next step is about handling the identified mismatches. The refinements are handled in any of the following two ways: by using motion vector correction, which gives improved prediction, or by using the directly replacing the error block. We have also presented a brief comparative study of the two error correction methods, discussing their benefits and drawbacks.

The methods used here give a precise motion estimate thereby utilizing the temporal redundancy in an efficient manner and providing an effective scalability solution. This scalable framework is useful to provide a flexible multiresolution adaptation to various network and terminal capabilities, to provide quality degradation during severe network conditions, and to provide better error robustness.

Statement of Originality

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of the thesis, when deposited in the University Library, being available for loan, photocopying, and dissemination through the library digital thesis collection, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Thesis Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

Date

Acknowledgements

I would like to thank God for His constant support and the blessings bestowed on me. I would like to thank the University of Adelaide for giving me this wonderful research opportunity and I am grateful to work under the supervision of Dr. Matthew Sorell and Prof. Langford White. The guidance, support and teaching of Dr. Sorell were immeasurable and exploring video scalability under him was a great experience.

My sincere thanks extend out to my CIR research colleagues – Angela, Limin, Belinda, Siew-Lee, Melissa, Dat, Sarah and Pinaki, for encouraging and supporting me thorough out my studies and making me feel at home. In particular, I would like to thank Angela Wong for her constant cheering, sisterly affection, and guidance. I would thank all my departmental staffs for their support especially Rose-Maire for her amazing support and timely help.

I would like to thank the International Student Centre for their constant support, particularly Patricia Anderson for her immense support and encouragement right till the end. My most grateful to Dr. Millie Vukovic, Dr. Cameron, Dr. Mirza, Dr. Bowman, Dr. Biju, Dr. Venkatesan and his family for helping me through when my health deteriorated. I am thankful to my friends and family for their constant support and encouragement.

I would like to thank Jacinta Poskey and her family for being my second family, and Roger Dangerfield and his family for their support and encouragement during difficult times. I would like to thank my father and grandparents for their sacrifice and enormous support for helping me achieve my goal. Finally I would like to thank my partner, Muthukarthikeyan Santhanam, for his unflinching support, guidance, care and holding me up through turbulent times.

I dedicate this work to my beloved and caring father *Veerapandian Sambandan*, my grandfather *Kasinathan*, my late grandpa *Sambandan*, and to my dear husband *Muthukarthikeyan Santhanam*.

Presentation

Veerapandian, L., Sorell, M., and White, L., (2007), “Scalable video coding using multiresolution motion estimation”, *Australian Communications Theory Workshop*, Adelaide, February.

List of Figures

1.1. Heterogeneous nature of video delivery	2
2.1. Building blocks of digital video	9
2.2. Resolution representation of CIF format and its family	11
2.3. Different user device specifications.....	13
2.4. Transmission capacity of the network	14
3.1. Video encoder architecture	18
4.1. Video transcoding scenario	23
4.2. The transcoder unit	24
4.3. Characteristic of a scalable approach	26
5.1. Fully scalable video coding framework	34
5.2. Quality scaled images using JPEG2000 coder	37
5.3. Resolution scaled images using JPEG2000 coder	37
5.4. Two level wavelet decomposition of video signals	39
5.5. Pyramid structure of wavelet decomposition and reconstruction	40
5.6. Multiresolution representation of wavelet decomposed images	41
5.7. Different quality levels of Resolution 1 (144x176)	41
5.8. Block-matching motion estimation scheme	43
5.9. Search area and current prediction block	44
5.10. Variable block-size motion estimation	46
5.11. Mismatch identification based on local motion	51
5.12. Prediction comparison along different resolution levels	51
5.13. Flow diagram of our spatial scalable framework	54
5.14. Level 2 decomposition using haar wavelet on the sample image	55
5.15. Identification of mismatch blocks	57
5.16. Representation of Threshold and WorstBlock factor	57
5.17. Threshold Plot for selective data transmission	60

6.1. The three extracted resolutions from three sample sequences	65
6.2. Comparison of resolution-scaled images (without error correction) against images formed through direct motion estimation without scaling	66
6.3. Number of blocks identified as mismatches using a nominal Threshold	67
6.4. Quality threshold constrained for low bandwidth (CarYellow sequence)	68
6.5. Mismatch identification during nominal Threshold and in low bandwidth scenario	69
6.6. Updated blocks using block and motion vector replacement method in Threshold mode 1 (CarYellow sequence)	71
6.7. Number of blocks identified as mismatches using a nominal Threshold (SwanAndBird sequence)	73
6.8. Quality threshold constrained for low bandwidth (SwanAndBird sequence) ...	73
6.9. Mismatch identification during nominal Threshold and in low bandwidth scenario (SwanAndBird sequence)	74
6.10. Updated blocks using block and motion vector replacement method in Threshold mode 1 (SwanAndBird sequence)	75
6.11. Quality threshold constrained for high bandwidth (CarYellow sequence)	76
6.12. Mismatch identification during nominal Threshold and in high bandwidth scenario (CarYellow sequence).....	77
6.13. Updated blocks using block and motion vector replacement method in Threshold 2 (CarYellow sequence).....	78
6.14. Quality threshold constrained for high bandwidth (SwanAndBird sequence)...	79
6.15. Mismatch identification during nominal Threshold and in high bandwidth scenario (SwanAndBird sequence).....	80
6.16. Updated blocks using block and motion vector replacement method in Threshold 2 (SwanAndBird sequence).....	81
6.17. Number of blocks identified as mismatches using a low Threshold in both the sequence.....	82
6.18. Quality threshold relaxed for high quality and high bandwidth (CarYellow sequence).....	83

6.19. Quality threshold relaxed for high quality and high bandwidth (SwanAndBird sequence).....	83
6.20. Mismatch identification during low threshold and in high bandwidth scenario (CarYellow sequence).....	84
6.21. Updated blocks using block and motion vector replacement method in Threshold mode 3 (CarYellow sequence).....	85
6.22. Mismatch identification during low threshold and in high bandwidth scenario (SwanAndBird sequence).....	86
6.23. Updated blocks using block and motion vector replacement method in Threshold mode 3 (SwanAndBird sequence).....	87
6.24. Motion vector estimate versus Block estimate in bit representation	91
6.25. Performance comparison of block and motion vector replacement methods across three selective data transmission modes in CarYellow sequence.....	96
6.26. Performance comparison of block and motion vector replacement methods across three selective data transmission modes in SwanAndBird sequence....	97
6.27. Performance of the three modes of selective data transmission in CarYellow sequence.....	99
6.28. Performance of the three modes of selective data transmission in SwanAndBird sequence.....	99
6.29. Performance of selective data transmission of low quality video during low bandwidth (mode 1) in CarYellow and SwanAndBird sequences.....	101
6.30. Performance of selective data transmission of high quality video during high bandwidth (mode 3) in CarYellow and SwanAndBird sequences.....	101
6.31. Performance comparison of resolution scaled images using multiresolution motion estimation (MRME without error correction) Vs our approach.....	103
A.1. Haar wavelet in time representation	119

List of Tables

6.1. Size of macroblocks in different resolutions	89
6.2. Motion vector estimate in bit representation	90
6.3. Macroblock estimate in bit representation	90
6.4. Motion vector and block estimate for different bandwidth requirements	91

Acronym

2D	2 Dimensional
3D	3 Dimensional
4CIF	4 Common Intermediate Format
16CIF	16 Common Intermediate Format
ADSL	Asymmetric Digital Subscriber Line
ADSL 2+	Enhanced version of ADSL
B frame	Bi-directionally predictive coded frame
bmp	bitmap – an image format
bpp	bits/pixel
bps	bit/ s
CCIR	International Consultative Committee for Radio
CCITT	International Telegraph and Telephony Consultative Committee
CD-ROM	Compact Disc Read Only Memory
CIF	Common Intermediate Format
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSL	Digital Subscriber Line
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding with Optimized Truncation
EC	Error Correction
EZW	Embedded Zerotree Wavelet
fps	frames per second
FGS	Fine Granular Scalability
FGST	Fine Granular Scalability with Temporal scalability
G bits/s	Giga bits/second
GOP	Group of Pictures
GPRS	General Packet Radio Service
GSM	Global Systems for Mobile communications
HDTV	High Definition Television
HPC	Handheld Personal Computer

HVS	Human Visual System
I frame	Intra coded frame
IEC	International Electrotechnical Commission
ISDN	Integrated Services Digital Network
ITU-R	International Telecommunications Union – Radio Sector
ITU-T	International Telecommunications Union – Telecommunication Standardisation Sector
ISO	International Organisation for Standardisation
JPEG	Joint Photographic Experts Group
k bps	kilo bits per second (kilo = 1000)
KLTF	Karhunen-Loeve Transform
LAN	Local Area Network
M bps	Mega bits/second
MATLAB	Matrix Laboratory
MC	Motion Compensation
MC-EZBC	Motion Compensated – Embedded Zero Block Coding
MCP	Motion Compensated Prediction
MCTF	Motion Compensated Temporal Filtering
ME	Motion Estimation
MPEG	Moving Pictures Experts Group
MRME	Multiresolution Motion Estimation
MSE	Mean Square Error
NTSC	National Television System Committee
P frame	Predictive coded frame
PAL	Phase Alteration Line
PC	Personal Computer
PDA	Personal Digital Assistant
PSNR	Peak Signal-to-Noise Ratio
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QMF	Quadrature Mirror Filter

QSIF	Quarter Source Input Format
RGB	Red-Green-Blue color components
SAD	Sum of Absolute Difference
SECAM	Sequentiel Couleur Aavec Memoire
SDT	Selective Data Transmission
SIF	Source Output Format
SNR	Signal-to-Noise Ratio
SPIHT	Set Partitioning in Hierarchical Trees
SQCIF	Sub-Quarter Common Intermediate Format (also known as Sub-QCIF)
Sub-QCIF	see SQCIF
T1 line	A dedicated phone connection supporting a 1.522 Mbps data rate
TV	Television
UMTS	Universal Mobile Telecommunication System
VHS	Video Home System
WLAN	Wireless Local Area Network
YCbCr	Luminance and Chrominance color components
YUV	see YCbCr