CITY UNIVERSITY OF HONG KONG

香港城市大學

Motion Compensated and Disparity Compensated Prediction Techniques for Video Coding

應用於視頻編碼的運動補償與視差補償預測技術

Submitted to

Department of Electronic Engineering

電子工程學系

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

哲學博士學位

by

Wong Ka Man

黃嘉文

August 2011

二零一一年八月

Abstract

Inter frame prediction plays a vital role in achieving enormous video coding efficiency in advanced video coding standards. It uses reference frames to predict the contents of current frame. Block matching techniques are used to estimate and compensate the spatial difference between frames. In Motion Compensated Prediction (MCP), reference frames are formed from the same scene in different time and block-matching techniques are used for modeling the movement of objects from time to time. In Disparity Compensated Prediction (DCP), however, the reference frames are based on frames captured from other viewing angles. Although the characteristics of these MCP and DCP frames are quite different, they are performed by the same set of coding tools in existing video coding standards. It assumes block based pure translations between frames and does not consider the actual motion and disparity effects. Different strategies for MCP and DCP are proposed in this thesis for improving the prediction accuracy and the coding efficiency in practical video codec implementation.

For MCP, motion can be very complex and normally not just pure translational motion. Attempts toward the general motion model are usually too complex in parameter estimation and too heavy with memory requirement in practical implementation. In this study, a translation and zoom motion model is investigated and the Block-matching Translation and Zoom Motion-Compensated Prediction (BTZMCP) is proposed to extend the assumed model to a more general one with zoom and translation motions in a practical way. A novel subsampling technique applied to the interpolated frame for subpixel motion estimation is used to implement the spatial scaling of temporal reference frames without using additional memory. Although BTZMCP requires block-matching with larger candidate pool, there is no additional zoom-related parameter estimation involved. Experimental results show that BTZMCP can achieve significant improvement on coding efficiency.

For DCP, as the reference frames are usually from the same moment, there are some known correlations among frames from these views. Conventional motion compensated prediction approach can be used to predict the disparity effect among views. However, in multiple views scenario, the same object in different views has deformations of different extents and, thus accurate disparity prediction cannot be achieved with such simple translational motion model. Similar to single video coding case, attempts to achieve more accurate disparity prediction usually involve affine transformations and are too complex for practical implementation. As zooming effect usually not exist in DCP, stretching, compression and shearing (SCSH) effects are investigated for better modeling the disparity effect in this study. Specially designed subsampling patterns for SCSH effects are designed for block-matching among frames from different views in disparity estimation and compensation. No affine parameter estimation or additional frame buffers is required in the proposed disparity compensated prediction algorithm. The overall increase in memory requirement and computational complexity is moderate. Experimental results show that the Rate-Distortion (RD) performance is also improved significantly. As the proposed new techniques for MCP and DCP can be applied to the current video coding standards, the developments on reference software are also reported to show the feasibility of these techniques.

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **iii**

Table of Contents ..... **iv**

List of Figures ..... **vi**

List of Tables ..... **ix**

Chapter 1 Introduction **1**

1.1. Motivation 2
1.2. Contribution 3
1.3. Thesis Organization 5

Chapter 2 Compression Techniques in Existing Video Coding Standards **6**

2.1. Hybrid Video Codec 6
2.2. Block-Matching MCP 8
2.3. Highlights of H.264/AVC Inter Frame Coding 10
2.3.1. Quarter-pixel MCP 10
2.3.2. Multiple Reference Frames MCP 13
2.3.3. Variable Block Size Motion Compensation 13
2.3.4. Entropy Coding 16
2.3.5. Rate-distortion Optimization (RDO) 17
2.4. Fast Algorithms for H.264/AVC 19
2.5. Summary 20

Chapter 3 Advanced Techniques for Motion Compensated and Disparity Compensated Predictions **21**

3.1. Higher Order Motion Models 22
3.1.1. Affine Multi-frame MCP (AMMCP) 22
3.1.2. Affine motion prediction by neighboring translational MVs 24
3.1.3. Mesh based MCP 28
3.1.4. Elastic MCP 29
3.2. Techniques in multiview video coding 32
3.2.1. Stereo vision 32
3.2.2. MVC extension of H.264/AVC 33
3.2.3. Block Matching Disparity Compensated Prediction 36
3.2.4. Disparity Compensated Prediction by higher order motion models 36
3.2.5. View synthesis by computer vision approach 36
3.2.6. View synthesis by Depth Image Base Rendering (DIBR) 37
3.3. Summary 38

Chapter 4 Block-Matching Translation and Zoom Motion-Compensated Prediction **39**

4.1. Extended translation and zoom motion model 40
4.2. Block-Matching Translation and Zoom Motion Compensated Prediction (BTZMCP) 41
4.3. Subsampled Block-Matching on the Subpixel Motion Estimation Interpolated Frame 45

4.4. Applying BTZMCP on practical video codec 51
4.5. BTZMCP implementation in KTA software 55
4.5.1. KTA software 55
4.5.2. Affected components in the encoder 55
4.5.3. Affected components in the decoder 66
4.6. Experimental results 67
4.6.1. RD Performance with different GOP structures 68
4.6.2. RD Performances With and Without BTZMCP 69
4.6.3. Performance On Using T-shaped Motion Model 70
4.6.4. Performance with High-Resolution Sequences 70
4.6.5. Comparison with MCP using other general motion models 78
4.7. Summary 79

Chapter 5 Stretching, Compression and Shearing Disparity Compensated Prediction for Stereo and Multiview Video Coding 80

5.1. Limitations of existing disparity compensated prediction methods 81
5.2. SCSH effects in multiview video 82
5.3. SCSH by subsampled block matching 85
5.4. Analysis on SCSH DCP 91
5.5. Applying SCSH DCP on H.264/AVC 97
5.5.1. RD optimized DCP 97
5.5.2. Parameter representation 97
5.5.3. Memory requirement 98
5.5.4. Computation 98
5.6. SCSH DCP in JM software 99
5.6.1. JM software 99
5.6.2. Affected components for SCSH DCP in JM17.0 100
5.7. Experimental results 101
5.7.1. Experiment setup 101
5.7.2. Direct improvement of SCSH inter-view prediction 101
5.7.3. Overall improvement of SCSH inter-view prediction 102
5.8. Summary 102

Chapter 6 Conclusions 110

References 113

Appendix Publications 120

List of Figures

Fig. 2.1. Architecture of a basic hybrid video coding system 7
Fig. 2.2. (a) Frame 1, (b) Frame 2, (c) The residual picture by subtracting the frame 1 from frame 2 without motion compensation, (d) The residual picture by subtracting the frame 1 from frame 2 with motion compensation. 7
Fig. 2.3. Block-matching motion estimation with integer-pixel motion vector accuracy. 10
Fig. 2.4. (a) Pixel locations of a 4 times interpolated reference frame and (b) Block-matching motion estimation with 1/4-pixel motion vector accuracy..... 12
Fig. 2.5. An example of MRF in case of temporary occlusion. 12
Fig. 2.6. Block diagram of multiple reference frame motion estimation 15
Fig. 2.7. Macroblock partitions and sub-macroblock partitions for motion compensation 15
Fig. 3.1. Block diagram of affine multi-picture motion estimation 23
Fig. 3.2. 8x8 blocks in a video frame and their TMVs 26
Fig. 3.3. Eight regions for an 8x8 block..... 26
Fig. 3.4(a) An example of predicting region 1 of block I by TMVs of blocks A, H, I..... 26
Fig. 3.4(b) An example of predicting region 5 of block I by TMVs of block I, D, E. 27
Fig. 3.5. Mesh based motion estimation 27
Fig. 3.6. Basis functions of elastic MCP for P=8, M=N=4. 31
Fig. 3.7. Simple disparity model. 32
Fig. 3.8. Prediction structure of stereo video coding 34
Fig. 3.9. Prediction structure of multiview video coding with 6 views. View 0 is the base view, views 2,4,5 are P views, views 1,3 are B views. 34
Fig. 3.10. Encoding order of stereo video coding case shown in Fig. 3.2. 35
Fig. 3.11. Block matching of inter-view prediction in MVC 35
Fig. 3.12. An example of DIBR. (a) is the reference view; (b) is the depth of the reference view; (c) is the rendered view without hole filling; (d) is the rendered view with hole filling. 38
Fig. 4.1. 2D Frame based motion model (a) Translation motion model (b) Translation and zoom motion model..... 40
Fig. 4.2. An example of block-matching MCP with multiple zoom reference frames..... 42
Fig. 4.3. PSNR improvement of BTZMCP with various zooming step sizes with bi-linear interpolation 44
Fig. 4.4. PSNR improvement of BTZMCP with various number of zooming levels with

zooming step size of 1/128 44
Fig. 4.5. (a) Pixel locations of a ¼-pixel interpolated reference frame and (b) an example of ¼-pixel accuracy subpixel Motion estimation with MV (1 ¾ , -2 ½). 47
Fig. 4.6. An example of MV (5, -9) with zoom factor *a*=4. 47
Fig. 4.7. Block-matching on a reference frame of zoom factor a=4/3. 48
Fig. 4.8. Block-matching on a reference frame of zoom factor a=4/5. 48
Fig. 4.9. (a) MRF with five temporal reference frames (b) BTZMCP with five zoom reference frames. 54
Fig. 4.10. T-shaped BTZMCP model 54
Fig. 4.11. Search positions of full pixel ME in non-zooming frame, large circles are full pixel locations, small circles are sub pixel locations. The numbers are the search order used in KTA software. 59
Fig. 4.12. Search positions of full pixel ME in zoomed frame with zoom factor of 8/7 60
Fig. 4.13. Example search positions of full pixel ME zoomed frame with zoom factor of 8/9. 60
Fig. 4.14. An example of motion prediction of a 16x16 block in KTA software, black dots are the top left pixels of 4x4 sub blocks. 63
Fig. 4.15. An example of motion prediction of 16x16 block mode with BTZMCP, s is the sub sampling rate defined in TABLE 4.2. 63
Fig. 4.16. An example of motion prediction of 8x16 block mode with BTZMCP, the sub sampling rate of two 8x16 blocks can be different 64
Fig. 4.17. Architecture of a conventional video decoder. 66
Fig. 4.18. Two frames in *stefan* sequence with a short fast zoom out scene. 75
Fig. 4.19. Two frames in *table* sequence with a long fast zoom out scene. 75
Fig. 4.20. RD curve comparison of *foreman* sequence. 76
Fig. 4.21. RD curve comparison of *stefan* sequence. 76
Fig. 4.22. RD curve comparison of *mobile* sequence. 77
Fig. 4.23. RD curve comparison of *table* sequence. 77
Fig. 5.1. A stereo image pair example. The 3-D effect can be viewed by parallel eyes method. 83
Fig. 5.2. Magnification on the wall of Fig. 5.1. 83
Fig. 5.3. Magnification on the ceiling of Fig. 5.1. 83
Fig. 5.4. Frame based SCSH disparity estimation. 84
Fig. 5.5. Block-matching on a reference frame with horizontal scale of 0.75x. 87
Fig. 5.6. Block-matching on a reference frame with horizontal scale of 1.25x. 87
Fig. 5.7. Block-matching on a reference frame of horizontal shearing factor of 1. 88
Fig. 5.8. Block-matching on a reference frame of horizontal shearing factor of -1. 88

Fig. 5.9. Block-matching on a reference frame of horizontal shearing factor of 0.5. 89
Fig. 5.10. Block-matching on a reference frame of horizontal shearing factor of 1 and horizontal scale of 0.75x. 89
Fig. 5.11. Block-matching on a reference frame of horizontal shearing factor of -1 and horizontal scale of 1.25x. 90
Fig. 5.12. Example stereo image pair of sequence *akko_kayo*..... 92
Fig. 5.13. Example stereo image pair of sequence *rena*..... 92
Fig. 5.14. Example stereo image pair of sequence *balloons*. 92
Fig. 5.15. Example stereo image pair of sequence *kendo*. 92
Fig. 5.16. PSNR improvement of SCSH DCP with various scale interpolated frame 94
Fig. 5.17. PSNR improvement of various numbers of SC levels..... 95
Fig. 5.18. PSNR improvement of various numbers of SH levels 95
Fig. 5.19. RD curve comparison of *rena* sequence. 106
Fig. 5.20. RD curve comparison of *akko_kayo* sequence. 106
Fig. 5.21. RD curve comparison of *balloons* sequence..... 107
Fig. 5.22. RD curve comparison of *kendo* sequence. 107
Fig. 5.23. RD curve comparison of *champagne* sequence. 108
Fig. 5.24. RD curve comparison of *dog* sequence. 108

List of Tables

TABLE 4.1. List of zoom factors of corresponding zoom step used in the analysis 43
TABLE 4.2. Definition of Zoom parameter and associated values in the implementation..... 52
TABLE 4.3. Pseudo code of encoding process of a P frame in KTA and JM. 56
TABLE 4.4. Pseudo code of decoding process of a P frame in KTA and JM. 66
TABLE 4.5. BD-Bitrate of BTZMCP compared to the original KTA with the same number of B frames without BTZMCP. 71
TABLE 4.6. BD-Bitrate as compared to original KTA software with GOP of IPPP..... 72
TABLE 4.7. RD comparisons between best configurations of different number of B frames. 72
TABLE 4.8. BD-Bitrate of BTZMCP with T-shaped motion model compared to the original KTA with the same number of B frames..... 73
TABLE 4.9. Encoding time comparison between BTZMCP with T-shaped model and original KTA with 7 hierarchical B frames 73
TABLE 4.10. RD performance comparisons between BTZMCP and original KTA software with distribution of zoom motion..... 74
TABLE 5.1 SCSH parameters used in the analysis 91
TABLE 5.2 Simulation result of SCSH DCP with different scales of interpolated reference frame 93
TABLE 5.3 Simulation result of SCSH DCP with various SC and SH levels with 4 times interpolated reference frame..... 96
TABLE 5.4. Definition of SCSH parameter and associated values 98
TABLE 5.5. RD comparison of inter-view prediction between JM17 and SCSH DCP 104
TABLE 5.6. Mode distribution comparison between JM17 and SCSH DCP 105
TABLE 5.7. Comparison of overall RD performance between JM17 and SCSH DCP..... 109