



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

The Hong Kong Polytechnic University

**Department of
Electronic and Information Engineering**

**Efficient Coding
Techniques for Video with
Various Brightness
Variations**

by
TSANG Sik Ho

A thesis submitted in partial fulfillment of the requirement
for the degree of Doctor of Philosophy

December 2012

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

TSANG SIK HO (Name of student)

Abstract

Interframe prediction is a critical component in a video encoder, where the current frame is predicted from a reference frame and only the prediction error is encoded. It assumes that the brightness keeps constant between the current and reference frames. However, some video scenes contain serious brightness variations caused by fade-in/out effects, camera iris adjustment, etc. Whenever brightness variation happens, an encoder might fail to find the true motion vectors. Consequently, it increases the number of bits to encode the prediction error. To improve coding efficiency, the H.264/AVC standard has adopted a weighted prediction (WP) tool for efficiently coding video scenes with global brightness variations (GBVs). Unfortunately, the WP tool in H.264 has not been fully examined in coding video scenes with local brightness variations (LBVs) such as illumination changes caused by a flash being fired during a press conference, a sport match, etc. Therefore, in this thesis, some novel techniques are suggested for the efficient implementation of WP in a digital video system to reduce the bitrates of encoded videos with various types of brightness variations.

With the proliferation of webcams, phone cameras and video editing tools, video effects such as synthetic fading can be added to digital videos easily. The fading effects result in GBVs. Various WP models to estimate the WP parameter set have been discussed in the literature. However, no single WP model works well for diverse fading effects. In this thesis, a single reference frame multiple WP models (SRefMWP) scheme is proposed to facilitate the use of multiple WP models in a single reference frame. The proposed scheme makes a new arrangement of the frame buffers in multiple reference frame motion

estimation. It enables different macroblocks in the same frame to use different WP models even when they are predicted from the same reference frame. A remarkable improvement of coding efficiency can then be achieved without modifying the H.264/AVC bitstream syntax.

Afterwards, we provide a novel solution for coding scenes with flashlight. The salient characteristic of flashlight effect is the abrupt luminance change across frames of the same scene within a very short period of time, which is caused by sudden appearance of the illumination source. The proposed solution then suggests an adaptive coding order technique for increasing the efficiency of video coding by taking account of characteristics of flash scenes in video contents. The use of the adaptive coding order technique also benefits to enhance the accuracy of derived motion vectors for determination of weighting parameter sets. Coding efficiency is thus substantially improved for flash scene with different WP parameter sets applying to different MBs.

Last but not least, we propose a new region-based scheme for the estimation of WP parameter sets for encoders of the H.264/AVC standard. This region-based scheme is specifically designed for handling local brightness variations in video scenes. It is achieved by making use of multiple WP parameter sets for various regions and assigning them to the same reference frame. An accurate estimation of multiple WP parameter sets is accomplished by (1) partitioning regions with a simple WP parameter estimator, (2) selecting regions where WP should be applied, and (3) estimating accurate WP parameter sets with a quasi-optimal WP parameter estimator. Similar to the SRefMWP scheme, the multiple WP parameter sets of different regions are encoded using the framework of multiple reference frames in the H.264/AVC standard. With this arrangement, the

proposed scheme is compliant with the H.264/AVC standard. Results show that the region-based scheme can efficiently handle scenes with global and local brightness variations, and achieve significant coding gain over other WP schemes.

By employing our proposed techniques, the work in this thesis achieves significant coding gain over the state-of-the-art WP tools. Undoubtedly, the results of our work will certainly be useful for the future development of coding videos with diverse brightness variations.

List of Publications

International Journal Papers

1. Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu, “Flashlight Scene Video Coding using Weighted Prediction,” *Journal of Visual Communication and Image Representation*, Volume 23, Issue 2, pp. 264-270, February 2012.
2. Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu, “Multiple Weighted Prediction Models for Video Coding with Brightness Variations”, *IET Image Processing*, vol. 6, issue 4, pp. 434-443, June 2012.
3. Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu, “Efficient intra prediction algorithm for smooth regions in depth coding,” *Electronics Letters*, vol.48, no.18, pp.1117-1119, August 2012.
4. Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu, “Region-based Weighted Prediction for Coding Video with Local Brightness Variations”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 3, pp. 549-561, March 2013.

International Conference Papers

1. Sik-Ho Tsang, Yui-Lam Chan and Wan-Chi Siu, “New Weighted Prediction Architecture for Coding Scenes with Various Fading Effects – Image and Video Processing,” in Proceedings of International Conference on Signal Processing and Multimedia Applications (SIGMAP 2010), pp.118-123, July 26-28, 2010, Athens, Greece.
2. Sik-Ho Tsang and Yui-Lam Chan, “H.264 video coding with multiple weighted prediction models,” in Proceedings of IEEE International Conference on Image Processing (ICIP 2010), pp.2069-2072, September 26-29, 2010, Hong Kong, China.
3. Sik-Ho Tsang, Tsz-Kwan Lee, Yui-Lam Chan and Wan-Chi Siu, “H.264 Region-based Weighted Prediction for Scenes with Local Brightness Variations,” in Proceedings of Constantinides International Workshop on Signal Processing (CIWSP 2013), pp. 1-4, January 25, 2013, London, UK.
4. Sik-Ho Tsang, Tsz-Kwan Lee, Yui-Lam Chan and Wan-Chi Siu, “Region-based Weighted Prediction Algorithm for H.264/AVC Video Coding,” in Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2013), pp. 269-272, May 19-23, 2013, Beijing, China.

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to my supervisor, Dr. Y.L. Chan, and co-supervisor, Professor W.C. Siu, for their continuous encouragement, guidance and care during the period that I worked on this thesis. They gave me information, suggestions and valuable advice contributing to every success of my research. More importantly, I am deeply impressed with their hard working style and their willingness to devote to the advancement of science and research. This gives me a clear image of a great researcher should be and have inspired me to work hard on the thesis. It is beyond doubt that this will continuously influence my future research and career.

Also, I would like to express my sincere thanks to Dr. Bonnie Law, Dr. Rainbow Fu, Dr. Victor Lai, Ms. Glory Lee, Dr. Apple Deng, Mr. W.L. Hui, Mr. W.H. Wong, Mr. Y.H. Kam, Dr. T. C. Hsung, Dr. Calvin Cheung, Dr. K. H. Chung, Mr. W. P. Choi, Mr. K. W. Wong, Dr. X. Jing, Dr. Y. Zhang, Mr. Alan Lun and Ms. Paula Wu. The sharing of ideas and experience with them has greatly contributed to make every success of my work.

Meanwhile, I am glad to express my gratitude to the Department of Electronic and Information Engineering and the Centre of Multimedia Signal Processing for providing me a comfortable working environment and their generous support for my research.

Table of Contents

CERTIFICATE OF ORIGINALITY.....	I
ABSTRACT.....	II
LIST OF PUBLICATIONS	V
INTERNATIONAL JOURNAL PAPERS	V
INTERNATIONAL CONFERENCE PAPERS.....	VI
ACKNOWLEDGMENTS	VII
TABLE OF CONTENTS	VIII
LIST OF FIGURES	X
LIST OF TABLES	XII
ABBREVIATIONS	XIII
CHAPTER 1 INTRODUCTION	1
1.1 DIGITAL VIDEO CODING.....	1
1.2 BLOCK-BASED HYBRID VIDEO CODING.....	3
1.3 BRIGHTNESS VARIATIONS IN HYBRID VIDEO CODING.....	4
1.4 MOTIVATION AND OBJECTIVES.....	6
1.5 ORGANIZATION OF THIS THESIS.....	7
CHAPTER 2 LITERATURE REVIEW	10
2.1 INTRODUCTION	10
2.2 VIDEO COMPRESSION FUNDAMENTALS AND H.264/AVC	11
2.2.1 <i>Video compression principles</i>	11
2.2.2 <i>Overview of H.264/AVC</i>	13
2.2.3 <i>Intra prediction</i>	14
2.2.4 <i>Inter prediction</i>	15
2.2.5 <i>Rate distortion optimization in H.264/AVC</i>	17
2.3 PROBLEM FORMULATION OF CODING SCENES WITH BRIGHTNESS VARIATIONS	19
2.4 CONVENTIONAL WEIGHTED PREDICTION IN H.264/AVC FOR CODING SCENES WITH GLOBAL BRIGHTNESS VARIATIONS.....	24
2.4.1 <i>DC model</i>	28
2.4.2 <i>Offset model</i>	29
2.4.3 <i>LS model</i>	29
2.4.4 <i>LMS model</i>	30
2.5 PREVIOUS ALGORITHMS FOR CODING SCENES WITH LOCAL BRIGHTNESS VARIATIONS.....	31
2.5.1 <i>Adaptive weighted prediction approach</i>	33
2.5.2 <i>Localized weighted prediction approach</i>	34
2.6 CHAPTER SUMMARY	35
CHAPTER 3 MULTIPLE WEIGHTED PREDICTION MODELS FOR SCENES WITH GLOBAL BRIGHTNESS VARIATIONS	37
3.1 INTRODUCTION	37
3.2 THE CONVENTIONAL WP IN H.264/AVC WITH THE SUPPORT OF MULTIPLE REFERENCE FRAME MOTION ESTIMATION.....	38

3.3	THE PROPOSED SINGLE REFERENCE MULTIPLE WP MODELS	40
3.4	SREFMWP USING PRE-CALCULATED LOOK-UP TABLES (LUTs)	43
3.5	EXPERIMENTAL RESULTS.....	45
3.5.1	<i>Rate-Distortion Performance of the Proposed Algorithm</i>	47
3.5.2	<i>Comparison of Encoding Complexity</i>	56
3.6	CHAPTER SUMMARY.....	59
CHAPTER 4 FLASH SCENE VIDEO CODING USING WEIGHTED PREDICTION ...		60
4.1	INTRODUCTION	60
4.2	PROPOSED CODING SCHEME FOR FLASH SCENES	61
4.2.1	<i>Adaptive Coding Order based on Flash</i>	63
4.2.2	<i>MB-based WP with Derived Motion Vectors</i>	65
4.3	EXPERIMENTAL RESULTS	67
4.3.1	<i>Sequences with real flash scenes</i>	69
4.3.2	<i>Sequences with synthetic flash scenes of different motion activities, flash durations, and intensity</i>	77
4.3.3	<i>Comparison of Encoding Time Complexity</i>	80
4.4	CHAPTER SUMMARY.....	81
CHAPTER 5 REGION-BASED WEIGHTED PREDICTION FOR CODING SCENES WITH LOCAL BRIGHTNESS VARIATIONS.....		82
5.1	INTRODUCTION	82
5.2	ANALYSIS OF WP PARAMETERS IN SCENES WITH LBV AND GBV	83
5.3	PROPOSED REGION-BASED WEIGHTED PREDICTION SCHEME.....	85
5.3.1	REGION PARTITIONING	85
5.3.2	DETERMINATION OF REGION-BASED WP PARAMETERS	87
5.3.3	EMBEDDING MULTIPLE REGION-BASED WP PARAMETER SETS INTO THE MRF FRAMEWORK OF H.264/AVC.....	89
5.3.4	THE FLOWCHART OF THE PROPOSED SCHEME	92
5.4	REDUCTION OF MEMORY REQUIREMENT USING LOOK-UP TABLES.....	93
5.5	EXPERIMENTAL RESULTS	95
5.5.1	RATE DISTORTION PERFORMANCES OF THE PROPOSED SCHEME	97
5.5.2	ANALYSIS OF MEMORY REQUIREMENT.....	103
5.5.3	COMPARISON OF ENCODING COMPLEXITY.....	104
5.5.4	IMPACT OF THE GOP STRUCTURE WITH B-FRAMES ON REGION-WP/REGION-WP+LUT	105
5.6	CHAPTER SUMMARY.....	109
CHAPTER 6 CONCLUSIONS AND FUTURE WORK.....		111
6.1	CONTRIBUTIONS OF THE THESIS	111
6.2	FUTURE DIRECTIONS	114
6.2.1	<i>Detection of brightness variations in H.264/AVC</i>	114
6.2.2	<i>Weighted prediction for inter-view inconsistency in depth coding</i>	115
6.2.3	<i>Weighted prediction for zooming effect in depth maps</i>	116
REFERENCES.....		118

List of Figures

Figure 2.1. Block diagrams of H.264/AVC (a) encoder, and (b) decoder.	12
Figure 2.2. Four prediction modes for intra 16x16.	14
Figure 2.3. Nine prediction modes for intra 4x4.	15
Figure 2.4. Variable block sizes for inter-coded blocks.	18
Figure 2.5. Video segments with and without brightness variations: (a) <i>NormalForeman</i> , and (b) <i>FadeForeman</i>	20
Figure 2.6. Average luma value of each frame for <i>NormalForeman</i> and <i>FadeForeman</i>	21
Figure 2.7. Encoding Bits per frame for <i>NormalForeman</i> and <i>FadeForeman</i> at QP24.	22
Figure 2.8. Mode distribution from 17 th to 46 th frames for <i>NormalForeman</i> and <i>FadeForeman</i> at QP24.	23
Figure 2.9. Motion vectors estimated at 29 th frame for <i>NormalForeman</i> and <i>FadeForeman</i> at QP24.	24
Figure 2.10. Illustration for implicit weighted prediction.	26
Figure 2.11. Illustration for localized weighted prediction approach.	35
Figure 3.1. Conventional weighted prediction in H.264/AVC – MRefSWP.	39
Figure 3.2. Proposed weighted prediction scheme – SRefMWP.	41
Figure 3.3. Look-up tables used in proposed SRefMWP scheme.	44
Figure 3.4. RD performances of different schemes for “Football” with (a) fade-in from black effect, and (b) fade-out to black effect.	50
Figure 3.5. Statistics of selected WP models in percentage of 8x8 blocks for “Football” with (a) fade-in from black effect, and (b) fade-out to black effect at QP 20 using SRefMWP/SRefMWP+LUTs.	51
Figure 3.6. RD performances of different schemes for “Mobisode2”.	52
Figure 3.7. RD performances of different schemes for “Foreman” from frame 170 to frame 229.	55
Figure 3.8. Statistics of selected WP models in percentage of 8x8 blocks for “Foreman” from frame 170 to frame 229 at QP 20 using SRefMWP/SRefMWP+LUTs.	55
Figure 4.1. Adaptive coding order for a FL scene.	62
Figure 4.2. Histogram and proposed prediction structure of a FL scene.	62
Figure 4.3. (a) Derivation of MVs for FL frames, (b) Crew, (c) BallSeq.	65
Figure 5.1. Original picture from the scene with GBV inside the “Mobisode2” sequence, (a) 234 th frame, (b) 235 th frame, and (c) estimated $W_i^{DC}(MB_n)$ using MB-based DC model.	84
Figure 5.2. Original picture from the scene with LBV inside the “Mobisode2” sequence, (a) 46 th frame, (b) 47 th frame, and (c) estimated $W_i^{DC}(MB_n)$ using MB-based DC model.	84
Figure 5.3. Distribution of $W_i^{DC}(MB_n)$ using the MB-based DC model in scenes with (a) GBV, and (b) LBV.	86
Figure 5.4. (a) $W_i^{DC}(MB_n)$ after quantization with $Q=4$, and (b) its corresponding histogram.	87
Figure 5.5. Conventional weighted prediction with MRF-ME in the H.264/AVC encoder.	90
Figure 5.6. Region-based weighted prediction adopted in the H.264/AVC MRF-ME architecture.	91
Figure 5.7. The flowchart of the proposed region-based WP scheme.	93
Figure 5.8. Entries of LUT _k , where $k=1, 2, \dots, N_R$	94
Figure 5.9. RD performances of different approaches for “Mo2_s1 (LBV)”.	98
Figure 5.10. Statistics of selected references and intra modes using REGION-WP or REGION-WP+LUT for frame 5 of “Mo2_s1 (LBV)” at QP 28.	99

Figure 5.11. Typical hierarchical-B coding structure.	107
Figure 6.1. Illustration of switching MRF structure for video coding.	115
Figure 6.2. Illustration of (a) texture image of left view, (b) texture image of right view, (c) depth image of left view, and (d) depth image of right view.	116
Figure 6.3. Illustration of depth images of (a) preceding frame, and (b) following frame, with zooming effect.	117

List of Tables

Table 3.1. Memory requirements (bytes).....	47
Table 3.2. BDBR (%), BDPSNR (dB) and encoding time reduction (%) of various schemes compared to ‘Without WP’.....	53
Table 3.3. Average percentages of 8x8 blocks using intra prediction (%) for various schemes.....	58
Table 4.1. Summary of tools used in various WP algorithms.....	69
Table 4.2. BDBR (%) and BDPSNR (dB) compared to “NoWP” for sequences with real flash scenes.....	76
Table 4.3. Average percentage (%) of intra, skip/direct and inter modes for FL frames only.....	76
Table 4.4. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different motion activities.....	79
Table 4.5. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different flash durations.....	79
Table 4.6. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different flash intensity.....	80
Table 4.7. Average percentage (%) of total encoding time increased compared with the conventional WP approaches.....	81
Table 5.1. Details of various video segments of “Mobisode1” and “Mobisode2” used for simulation	95
Table 5.2. Details of HD movie trailer downloaded from Apple iTunes [78] for simulation.....	96
Table 5.3. Average percentage (%) of inter and intra modes for different schemes.....	100
Table 5.4. BD-Bitrate (%) and BD-PSNR (dB) of various schemes compared to WITHOUT WP.....	102
Table 5.5. Memory requirements for various schemes with 5 reference frames for coding WVGA video.....	103
Table 5.6. Total encoding time increment of various schemes compared to WITHOUT WP ..	105
Table 5.7. BD-Bitrate (%) and BD-PSNR (dB) of various algorithms compared to WITHOUT WP using the hierarchical-B encoding structure.....	107
Table 5.8. Average encoding time increment of various schemes compared to WITHOUT WP when the hierarchical-B encoding structure is used.....	108

Abbreviations

3DTV	3-dimensional television
AVC	Advanced Video Coding
AVS	Audio and Video coding Standard
B-frame	Bi-directional predictive frame
CABAC	Context adaptive binary arithmetic coding
CAVLC	Context adaptive variable length coding
DCT	Discrete cosine transform
DVD	Digital Versatile Disc
FL	Flashlight
FTV	Free viewpoint television
GBV	Global brightness variation
GOP	Group of pictures
HDTV	High definition television
I-frame	Intra frame
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IT	Integer transform
ITU	International Telecommunication Union
JM	Joint Model
JPEG	Joint Picture Experts Group
JVT	Joint Video Team
LBV	Local brightness variation

MB	Macroblock
MC	Motion compensation
ME	Motion estimation
MPEG	Moving Picture Experts Group
MRF	Multiple reference frames
MV	Motion vector
MVC	Multiview video coding
MVD	Multiview video plus depth
P-frame	Predictive frame
PPS	Picture parameter set
QP	Quantization parameter
RD	Rate distortion
RDO	Rate distortion optimization
SAD	Sum of absolute difference
SPS	Sequence parameter set
SSD	Sum of squared difference
TV	Television
UHDTV	Ultra high definition television
VCD	Video compact disc
VCEG	Video Coding Experts Group
VLC	Variable Length Coding
WP	Weighted prediction

Chapter 1 Introduction

1.1 Digital video coding

Video has become one of the major mediums for content representation and distribution. From movies and broadcast TV to pre-captured digital video, and then to live digital high definition (HD) TV and 3-dimensional (3D) TV, we have been experiencing a digital video revolution in the last couple of decades. Apart from the more robust form of the digital video signal, the main benefit of digital representation and transmission is easier to provide a diverse range of services. Nevertheless, the bottleneck preventing the use of digital HD video is the huge storage and transmission bandwidth requirements. For instance, a single-side single-layer Digital Versatile Disk (DVD) cannot store more than one minute of raw HD video. Hence, researches are aspired to find the best way for video compression. The interest in video compression has been shown in international efforts for video compression at various applications. Developing an international standard requires collaboration among many parties with different commercial interests, and an organization that can support the standardization process and enforce the standards. As a result, different video coding standards have been established by various international standardization organizations. They include the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC) and the International Telecommunication Union (ITU). In these organizations, there are two major teams of developing video coding standards. They are the ISO/IEC MPEG (Moving Picture Experts Group) and ITU-T VCEG (Video Coding Experts Group). They have been responsible for the successful H.261, MPEG-1 Part 2, H.262/MPEG-2 Part 2,

H.263, MPEG-4 Part 2 [1-7], and H.264/AVC [8-10], which have given rise to widely adopted commercial video delivery applications and services, such as video conferencing, Video-CD (VCD), DVD, digital television, blu-ray disc, 3DTV, etc.

All the developed standards utilize the redundancy inherent in digital video information in order to achieve a significant reduction in data rate. In general natural video, scene and objects in the video content changes smoothly and gently over time, so successive frames contain a large amount of temporal redundancy. To achieve compression of video data, the temporal redundancy can be removed by block-based motion estimation and compensation in the modern video coding standards. For simple video sequences, the existing motion estimation and compensation algorithms have been well-developed with the superior performance and reasonable computational complexity. However, the superior performance can only be guaranteed in video sequences usually acquired in indoor or controlled environments due to the nature of motion estimation and compensation. It assumes that brightness of an object in a video scene keeps constant during motion. When a video sequence contains brightness variations, the existing motion estimation and compensation algorithms do not yield desirable results. Nowadays, most video sequences acquired by general users or in uncontrolled environments can often have remarkable brightness variations such as abrupt illumination changes or camera operations including fade-in/out effects, camera flashes, etc. Besides, to fulfill people's demanding on much higher requirement of visual enjoyment such as free viewpoint television (FTV) [11] and ultra HDTV (UHDTV) which always contain scenes with complex brightness variations, it is essential to solve the brightness variation problem of digital video coding. Therefore, in this thesis, we explore ways to efficiently encode video with various kinds of brightness variations.

In this chapter, the fundamentals of digital video and the need of video compression are introduced. An overview of block-based hybrid video coding is then described briefly. Afterwards, the problem of scenes containing brightness variations is discussed. Last but not least, the motivation, objectives and the organization of this thesis are presented.

1.2 Block-based hybrid video coding

Digital video is formed by a sequence of frames that are created in the form of a two dimensional matrix of individual picture elements known as pixels. Owing to the spatial similarity, a single frame within a video sequence always has a significant amount of spatial redundancy. Apart from the spatial redundancy, the similarity existing between successive frames is called temporal redundancy.

The block-based hybrid video coding approach is essentially the core of all the international video coding standards to reduce the aforementioned redundancies. In this block-based approach, each video frame is divided into blocks of a specific size and each block is coded more or less independently. The "hybrid" indicates that each block is processed using a combination of temporal prediction and transform coding. In temporal prediction, a block in the current frame is predicted from a previously coded reference frame using block-based motion estimation, which computes the motion vector of the current block. The motion vector is the displacement between the current block and the best matching block. The predicted block can then be obtained from the previously coded reference frame based on the motion vector using motion compensation. The difference between the current and predicted block is referred to as a prediction error block. The

more accurate the prediction process, the less energy is contained in the prediction error block. In transform coding, the prediction error block is undergone transformation, quantization and entropy coding processes before it is stored or sent to the decoder. In block-based video coding, the temporal prediction method is successful to reduce temporal redundancy between successive frames by forming a predicted block and subtracting this from the current block while transform coding is used to remove the spatial redundancy.

In natural video, temporal prediction is very successful since the prediction error block always requires fewer bits to code than the original block. This method of coding is referred to as inter-mode. When this is not the case, the original block instead of the prediction error block is coded directly using transform coding. This is called intra-mode.

1.3 Brightness variations in hybrid video coding

As mentioned in the previous sub-section, the block-based motion estimation and compensation are among the most popular approach to reduce temporal redundancy in video coding by estimating motion vectors between successive frames. It assumes that brightness between frames is constant during motion estimation and compensation, changes between video frames may be caused by object movements or camera motions rather than brightness changes between frames. When brightness variations such as fade-in/out effects and camera flashes occur between successive frames, the motion estimation cannot be accurately performed. True motion vectors cannot be obtained which may increase the amount of the prediction errors. Inter-modes with large distortion or even

intra-modes would be chosen mostly through rate-distortion optimization (RDO). Consequently, coding efficiency may be reduced in the presence of brightness variations.

For some early video coding standards such as MPEG-1, MPEG-2, and H.263, they just prejudge that the difference between frames is due to motion only. They are unable to deal with scenes with brightness variations. H.264/AVC is currently one of the most common international video coding standard developed by the ITU-T/ISO/IEC Joint Video Team (JVT), and has been shown to achieve superior coding efficiency. This coding gain mainly comes from new coding tools. Among these tools, weighted prediction (WP) firstly defined in the Main and Extended Profiles is an indispensable tool for coding video scenes containing brightness variations caused by fade in/out effects, camera flashes, camera iris adjustment, etc. WP can be used to enhance motion compensation by modifying the original reference frame with WP parameters including a multiplicative weighting factor W and an additive offset O . In a video sequence with brightness variations, the current frame being encoded is more strongly correlated to the modified reference frame which is scaled and shifted by W and O than to the reference picture itself. This coding mechanism ensures that the difference between the current frame and the modified reference frame becomes smaller, resulting in the fewer bits that are needed to encode the current frame.

1.4 Motivation and objectives

It is interesting to note that the H.264/AVC standard does not define the way to derive WP parameters. Various algorithms for estimating WP parameters have been studied in the literature [12-15], and some of them have also been adopted in the H.264 Joint Model (JM) [16-18]. The basic concept of [12-15] is to use global brightness compensation, and it assumes that the brightness variation is uniformly applied across an entire picture. In this situation, one set of WP parameters is sufficient to code all macroblocks (MBs) in the picture that are predicted from the same reference frame. However, there are various kinds of brightness variations due to the proliferation of webcam, phone cameras, and video editing tools. In addition, brightness variations may be non-uniform in a picture. Firing of flashes in a scene can cause the non-uniform intensity change distributed over the entire picture. It is therefore very difficult to find a single set of WP parameters to estimate the change of intensity within the picture. In other words, the WP scheme adopted in H.264/AVC cannot handle scenes with various brightness variations well. To cope with the aforementioned problems, some efficient WP schemes are in great demand.

Various schemes for estimating WP parameters have been studied in the literature [12-15]. We have done some careful investigation of these WP schemes. In summary, these schemes always focus on how to estimate a more accurate set of WP parameters, or equivalently, how to find a better WP model, in order to improve the effectiveness of video coding. The results reveal that these schemes are still primitive, and there is plenty of room for improvement. For instance, single WP model is not sufficient for diverse fading effects, and this limits the efficiency of WP in this type of video sequences. In order to bolster the practicality of using WP in video coding, our study is to design an H.264/AVC standard-compliant scheme in consideration of multiple WP models. In this

thesis, we perform a detailed analysis and provide a practical solution for adopting multiple WP models in the structure of multiple reference frames (MRF) of H.264/AVC. Some new components are integrated into the conventional H.264/AVC coding architecture so as to obtain remarkable improvement of coding efficiency.

Results of our investigation indicate that the proposed scheme using multiple WP models can compensate for coding scenes with different fading effects. Nevertheless, it only performs well for the scenes with global brightness variations (GBVs) where brightness variation is uniformly applied across an entire picture. Unfortunately, different degrees of brightness variations may be applied to different regions at the same time instant, and it is known as local brightness variations (LBVs). Example includes scenes with a flash being fired during a press conference, a sport match, a news interview, etc. Therefore, another objective of this research is to further extend our scheme using multiple WP models and propose region-based/MB-based WP parameter estimation schemes for the H.264/AVC standard. Several novel techniques are investigated for implementing the required components and addressing these challenging issues. We believe that our proposed techniques will play a vital role in the future video coding standards for improving the coding performance of scenes with GBVs and LBVs.

1.5 Organization of this thesis

This thesis is divided into six chapters. Prior to embarking on the description of the main research in this thesis, Chapter 2 gives a broad overview of video compression techniques. The overview covers some general introductory video compression materials for the purpose of clarifying certain definitions used in later chapters. The impact of brightness

variations in video coding is then addressed. Afterwards, a review of the conventional WP in the H.264/AVC standard and different WP models for estimating the WP parameters are also given. At the end of this chapter, some current WP schemes including frame-based techniques and MB-based techniques for solving the problems of brightness variations are presented.

In Chapter 3, a novel WP scheme utilizing the multiple reference frames (MRF) architecture in H.264/AVC is presented. Techniques including estimation and retrieving of WP parameters by reference indexing are proposed. By merging the existing WP models into the MRF architecture, a new WP scheme for solving the problems of complex GBV is proposed without modification of the bitstream syntax.

An MB-based WP scheme is then proposed in Chapter 4. With the utilization of adaptive coding order (ACO) and the derivation of motion vectors (DMV), MB-based estimation of WP parameters is designed in order to solve the problems of coding flashlight (FL) scenes, which are typical examples of LBVs. MBs with different degrees of brightness variations are motion-compensated by accurate MB-based WP parameters.

Chapter 5 extends the proposed WP scheme utilizing the MRF architecture concept to formulate a region-based WP approach. A region partitioning process is designed to divide the current frame into different regions where each one has some degree of uniformity in its brightness variation. This can assist in estimating multiple sets of region-based WP parameters accurately. By making use of the MRF architecture and reference reordering in H.264/AVC, different MBs in the current frame can then use different WP parameter sets even when they are predicted from the same reference frame.

Consequently, the proposed region-based scheme can improve prediction in scenes with different degrees of LBVs in different regions of the same picture.

Chapter 6 is devoted to a summary of the work herein and the conclusions reached as a result. Suggestions are also included for further research in this area.

Chapter 2 Literature Review

2.1 Introduction

With the recent advances in network technology, video applications and services are being our part of daily life. Examples include digital TV, DVD, video streaming, video surveillance, video conferencing, 3DTV, etc. A number of video coding standards such as MPEG-1, MPEG-2, MPEG-4, H.26L and H.264/AVC [3-10] have been developed to define standard video formats for the purpose of efficient storage and transmission. To reduce the file size of digital video, numerous video compression techniques exploiting the spatial and temporal redundancy are employed. The motion compensated predictive coding is based on the underlying assumption that the difference between two successive frames is due to motion only. This assumption overlooks other possible causes of the frame difference in the presence of temporal brightness variations (synthetic fading effects, camera flashlights and local illumination changes, etc.), which reduce the correlations between successive frames. It causes higher bitrate to code the scenes with brightness variations and finally results in relatively low compression ratio. To encode video sequences with brightness variations, some research efforts have been conducted in different ways.

This chapter is organized as follows. In the first section, we start with the brief description of some fundamental concepts about digital video representation. We then present the generic encoder structure and the hierarchical structure, with an emphasis on the techniques that are related to this research. Next, the problems of coding video scenes

with brightness variations are discussed. Afterwards, the conventional H.264/AVC video coding with the use of weighted prediction (WP) is given. This conventional WP coding technique helps to improve the coding efficiency of videos in the presence of global brightness variations (GBVs) in the scene. Several WP models are then introduced to describe their usages for different kinds of fading effects. The last section reviews several existing algorithms for coding scenes with local brightness variations (LBVs) in the H.264/AVC video coding system.

2.2 Video Compression Fundamentals and H.264/AVC

2.2.1 Video compression principles

Digital video has a substantial amount of data. There is a need to reduce the data rate of digital video. Compression of video data without noticeable degradation of the visual quality is achievable because video has a high degree of redundancy. They are spatial and temporal redundancy. Spatial redundancy always exists within a frame due to the correlation between neighboring pixels. On the other hand, objects in the natural video scene always move gently. Thus, temporally adjacent frames are often highly correlated in a moving video sequence, and the high correlation between frames of a video sequence results in temporal redundancy. By exploiting the redundancy in a video sequence, many video coding techniques has been developed for the past decades.

Most video coding standards assume a model that employs transformation, quantization, entropy coding, and block-based motion estimation and compensation. An H.264/AVC video coding system which reduces both spatial and temporal redundancy is shown in Figure 2.1.

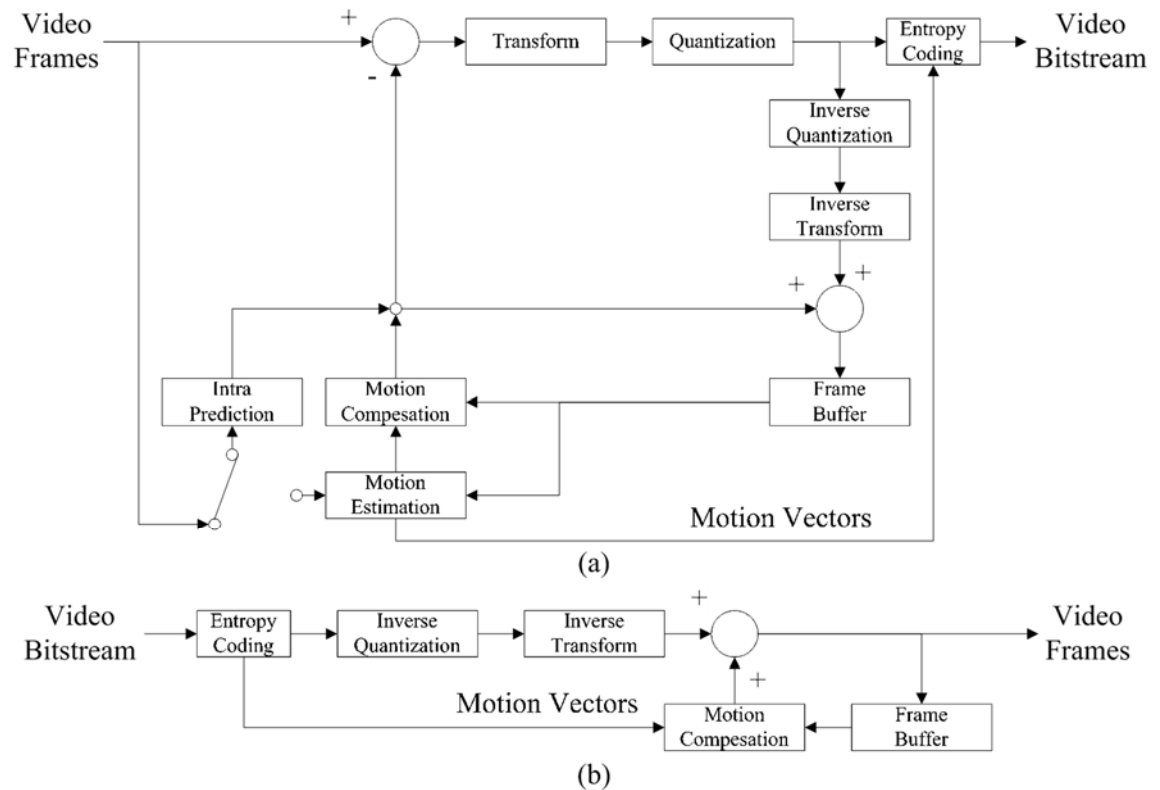


Figure 2.1. Block diagrams of H.264/AVC (a) encoder, and (b) decoder.

Basically, there are three frame types adopted in various video coding standards. They are I-frame (intra frame), P-frame (predictive frame) and B-frame (bi-directional predictive frame). Each frame, regardless of its frame type, is divided into 16×16 -pixel blocks called macroblocks (MBs), and all MBs are usually encoded one by one in raster scan order during the whole encoding process. For I-frame, each MB is encoded without making use of any temporal redundancy. In other words, all MBs within I-frame are encoded without predicting from other frames. We can also say that all MBs are intra coded. Thus, the compression ratio of I-frame is smallest. P-frame is the predictive frame in which it only uses previous frames as references for prediction whereas B-frame uses both previous and future frames as references. Thereby, MBs are said to be inter coded if they are predicted from other frames. For both P-frame and B-frame, MBs can be an intra-coded MB or an inter-coded MB. Notice that the prediction direction for inter-coded MBs

in P-frame can only be forward direction whilst the prediction direction for inter-coded MBs in B-frame can either be forward, backward or bi-directional directions.

2.2.2 Overview of H.264/AVC

H.264/AVC is the most popular video coding standard developed by the Joint Video Team (JVT), which is formed by the ITU-T VCEG and ISO/IEC MPEG standard committees [8-10]. It is targeted for providing similar functionality to the previous standards with significantly better compression performance. Besides, H.264/AVC aims at having provision of a network-friendly video representation which addresses storage, broadcast and streaming applications.

As depicted in Figure 2.1, H.264/AVC is a block-based motion-compensated hybrid video codec. Firstly, the current frame being encoded is divided into MBs with the size of 16×16 as the basic coding unit. Each MB is then predicted by either intra prediction or inter prediction, which is determined by the mode decision. In particular, all the intra and inter prediction modes are checked one by one based on the rate distortion optimization, and the one with the minimum rate distortion cost is selected to be the optimal mode. Secondly, the resulted prediction errors are transformed and quantized. Finally, the headers, motion vectors, and quantized prediction errors are entropy coded to generate the H.264/AVC bitstream.

Though the coding mechanism of H.264/AVC is very similar to that of the previously video coding standards, H.264/AVC achieves much higher coding efficiency contributed by a number of sophisticated tools [9,10,19-29]. The following sub-sections will briefly describe some of these tools that are relevant to this work.

2.2.3 Intra prediction

In H.264/AVC, intra prediction [9,10] is an efficient tool to reduce the spatial redundancy with each frame. It predicts the current pixel based on the spatially neighboring reconstructed pixels. Then, only the residual block, which is the prediction error computed between the current block and its predicted block, is encoded using transformation, quantization, and entropy coding, as shown in Figure 2.1.

To achieve superior performance, a predicted block is formed for each 4×4 block, 8×8 block, or 16×16 macroblock [9,10] in H.264/AVC. They are referred to as intra 4×4 , intra 8×8 or intra 16×16 , respectively. For intra 16×16 shown in Figure 2.2, H.264/AVC offers 4 types of prediction modes (i.e., vertical, horizontal, DC and plane modes) to form the prediction block. For intra 8×8 and intra 4×4 , there are 9 types of prediction directions, including one DC mode and eight directional modes as illustrated in Figure 2.3. The arrows in Figure 2.2 and Figure 2.3 indicate the direction of prediction in each mode. The predicted pixels are formed from a weighted average of the neighboring pixels. It is noted that there are totally four 8×8 blocks and 16 4×4 blocks within an MB for intra 8×8 and intra 4×4 respectively. Typically, intra 16×16 is always used in smooth regions while intra 8×8 and intra 4×4 are selected in complex regions.

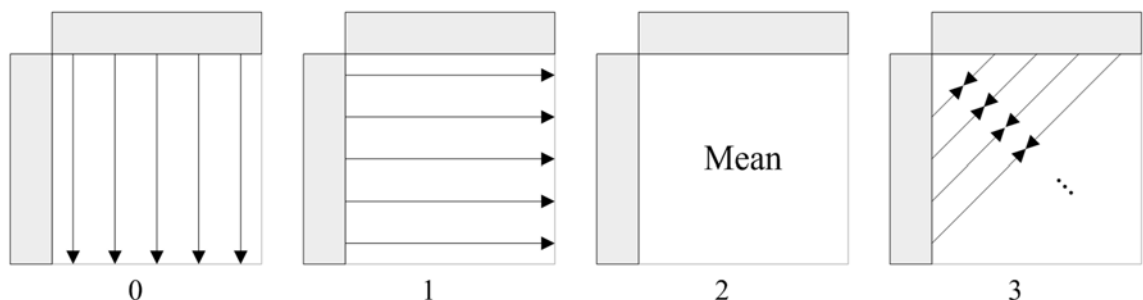


Figure 2.2. Four prediction modes for intra 16×16 .

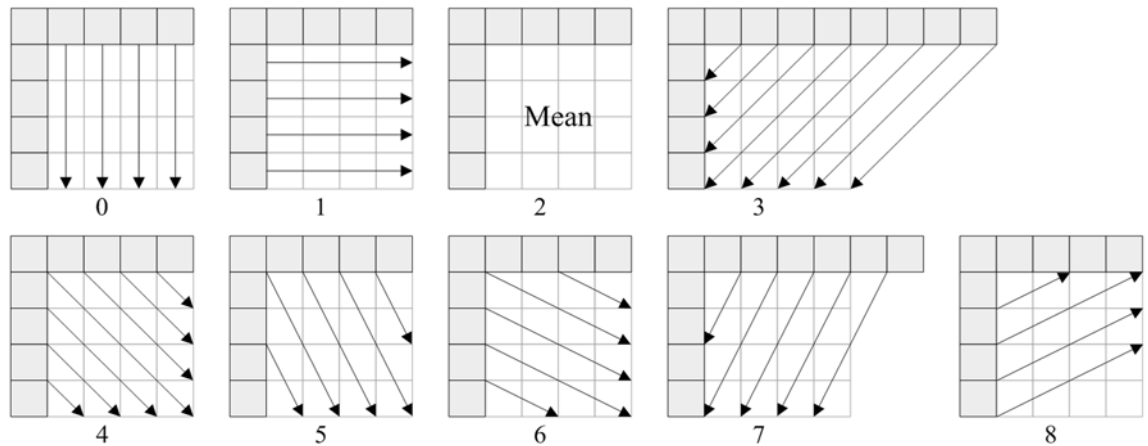


Figure 2.3. Nine prediction modes for intra 4x4.

2.2.4 Inter prediction

Inter prediction is the key to the success of the video coding standards on the removal of temporal redundancy. Instead of predicting pixels within a frame, it uses previously encoded frames as the predictor for the current frame. This technique is known as motion estimation (ME) and motion compensation (MC), and is essentially the core of most video coding standards. In the video coding standards, the motion estimation process uses the rectangular block of $M \times N$ pixels as a basic unit in which all pixels of each block in the current frame are compared on a pixel-by-pixel basis with pixels of the candidate block in the reference frame within a pre-defined search range. The aim is to find the block in the reference frame that gives the best rate-distortion. In H.264/AVC, it supports motion estimation using different block sizes such as 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , and 4×4 . To compute the coding modes and motion vectors for each inter block, motion estimation is firstly performed for all modes and submodes independently by minimizing the Lagrangian cost function J_{motion} .

$$J_{motion} = SAD(b_c, b_i) + \lambda_{motion} \cdot R_{motion}(MV - PMV) \quad (2.1)$$

where MV is the motion vector used for prediction, λ_{motion} is the Lagrangian multiplier for motion estimation, $R_{motion}(MV - PMV)$ is the rate or estimated number of bits for coding MV by subtraction from motion vector predictor PMV , and $SAD(b_c, b_i)$ is the sum of absolute differences between the block b_c in the current frame f_c and the reference block b_i in the i^{th} reference frame f_i , which is computed by

$$SAD(b_c, b_i) = \sum_{\substack{p_c \in b_c \\ p_i \in b_i}} |p_c - p_i| \quad (2.2)$$

where p_c and p_i are the pixels in b_c of f_c and b_i of f_i . It is noted that multiple reference frame motion estimation is supported in H.264/AVC such that several frames can be used as references. Therefore, i is the reference frame number and i is equal to 0 for the nearest reference frame. The candidate MB that has smallest J_{motion} is chosen as the best match. The relative displacement between the current MB and the best-matched MB in the reference frame is encoded. This is known as a motion vector. The predicted MB is obtained from the reference frame based on the motion vector using motion compensation, and is subtracted from the current MB to form the residual block. Then, the residual block is coded by transforming it, quantizing the DCT coefficients and converting them into variable length code words using entropy coding, as shown in Figure 2.1. This procedure is similar to the process of intra coding. However, the consumption of computing power for inter-coded MB is much higher than that for intra-coded MB since exhaustively checking all the possible candidates within the search range for locating the optimal one greatly increases computational complexity.

For a bi-directional inter-coded block, the residue becomes the difference between an interpolated block and the current block. The interpolated block is obtained by interpolating the best matched blocks from the forward and backward reference frames.

So the SAD between the current block b_c and the average of reference blocks from the i -th forward and the j -th backward reference frame, b_i and b_j respectively, is given as

$$SAD(b_c, b_i, b_j) = \sum_{\substack{p_c \in b_c \\ p_i \in b_i \\ p_j \in b_j}} \left| p_c - (p_i + p_j)/2 \right| \quad (2.3)$$

As there are two best matched block candidates from forward and backward references, two MVs, one forward MV and one backward MV, are required to be coded into the bitstream.

Besides, many new tools in H.264/AVC such as quarter-pixel motion estimation, multiple reference frame motion estimation, variable block size motion estimation, etc., are made available for improving coding efficiency in inter prediction. Quarter-pixel and multiple reference frame motion estimation facilitate an H.264/AVC encoder to identify the best-matched MB with higher accuracy and among a larger number of reference frames. In addition, variable block sizes are used to more accurately obtain the true motion of video objects. More details related to variable block sizes can be found in the next section.

2.2.5 Rate distortion optimization in H.264/AVC

Unlike earlier video coding standards, H.264/AVC includes the support for seven block sizes or modes (16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , and 4×4), as shown in Figure 2.4, to carry out motion estimation in order to increase coding efficiency [19,20]. Each 16×16 MB can be divided into 16×8 , 8×16 and 8×8 block partitions. Within an 8×8 block, it can be further divided into 8×4 , 4×8 and 4×4 block partitions. Hence, there are seven kinds of sizes for inter-coded blocks. Using a smaller block size for motion estimation

may give a smaller residual error with the expense of higher bitrate to code the MVs of all partitions within the same MB as each partition has its own MV. Choosing a larger block size for motion estimation may require fewer bits but give a higher energy residual after motion compensation. Generally, a large block size is always suitable for homogeneous areas of the frame and a small block size may be appropriate for detailed areas.

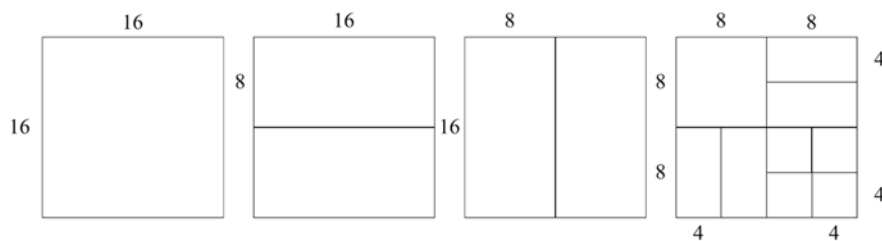


Figure 2.4. Variable block sizes for inter-coded blocks

For the inter-frame coding, there are mainly 11 candidate modes for each MB. They are SKIP, inter-16×16, inter-16×8, inter-8×16, inter-8×8, inter-8×4, inter-4×8, inter-4×4, intra-16×16, intra-8×8, and intra-4×4. For the intra-frame coding, only intra-16×16, intra-8×8, and inter-4×4 are applicable. Note that the residual error is set to zero and its MVs can be generated from the MVs of its neighboring MBs if an MB is coded as SKIP. In this case, no residual error and MVs are required to be transmitted. SKIP is thus highly beneficial to code texture-less MB with fairly small motion.

To determine the optimal mode, H.264/AVC adopts the Lagrangian rate-distortion optimization (RDO) technique [20-24] as its mode decision criterion, and all the modes are exhaustively investigated to find the one with the minimum RD cost as the optimal mode. The function of Lagrangian RDO, J_{mode} , is given by

$$J_{\text{mode}} = SSD(b_c, b_{ri}, m) + \lambda_{\text{mode}} \cdot R_{\text{mode}}(b_c, b_{ri}, m) \quad (2.4)$$

where λ_{mode} is the Lagrangian multiplier for mode decision, m is one of the candidate modes, SSD is the sum of squared differences between the original block b_c and its reconstructed block b_{ri} , and $R_{\text{mode}}(b_c, b_{ri}, m)$ represents the number of coding bits associated with the chosen mode.

2.3 Problem Formulation of Coding Scenes with Brightness

Variations

From the early video coding standards, such as MPEG-1 and MPEG-2, to the latest international video coding standard, H.264/AVC [8-10], motion estimation (ME) process has been playing an indispensable role. After years of development, H.264/AVC has been proven to obtain remarkable coding efficiency. This coding gain is mainly contributed from new prediction tools. For instance, variable block size motion estimation [19,20] described in the preceding section that allows to estimate video sequences with rich local motion and relatively small motion objects. Another example is multiple reference frame motion estimation (MRF-ME) [25-27] which provides several previous frames as references such that motion can be estimated not only from one single reference frame. Moreover, sub-pixel ME [28,29], which scales up the reference frames by interpolating the integer pixels for more fine-grained ME, is another new tool for H.264/AVC. These new prediction tools in H.264/AVC improve the coding efficiency with the expense of high computational complexity.

Notwithstanding the help of variable block size ME, MRF-ME and sub-pixel ME in the video encoding process, the video encoder still fails to estimate accurate motions in scenes with brightness variations. The presence of brightness variations may be caused by fade-in/out effects, camera flashes, camera iris adjustment and local illumination changes. Whenever brightness variation happens, it induces large differences between the current frame and the reference frames. The SADs in (2.2) or (2.3) between the current and reference blocks pointed by MVs might become large due to the large illumination change. Since J_{motion} in (2.1) considers both of the SAD and the amount of bits required to encode the block, the computed MVs might be estimated wrongly which cannot reflect the true object motions. It increases the number of bits to encode the residues. As a consequence, inter modes are unlikely to be selected as the optimal mode due to the large value of J_{mode} . In other words, intra modes are more preferred than inter modes in coding MBs with brightness variations. Note that choosing more intra mode usually reduces coding efficiency.



Figure 2.5. Video segments with and without brightness variations: (a) *NormalForeman*, and (b) *FadeForeman*.

In the following discussion, two video segments in Figure 2.5 are used to illustrate the impact of brightness variations to video coding. The video segment in Figure 2.5(a) is 0th

to 59th frames of the Foreman sequence. It is referred to as *NormalForeman*. Figure 2.5(b) shows another video segment when 30-frame synthetic fade-out-to-black effect is applied into 17th to 46th frames of *NormalForeman*, named *FadeForeman*. The synthetic fade-out-to-black effect is formulated as follows:

$$F_t = \alpha_t f_t + (1 - \alpha_t) C \quad (2.5)$$

where f_t and F_t are the original signal and the modified signal with fading respectively, C is the target fading color value, and α_t is a weighting coefficient from 0 to 1 which is decreasing (increasing) with time for fade-out (fade-in) effects. The corresponding brightness levels of both sequences, represented by the average luminance values of whole frames, are plotted in Figure 2.6. It can be seen that the brightness level is more or less the same for every frame in *NormalForeman* while the brightness level in *FadeForeman* gradually decreases starting from 17th to 46th frames and becomes totally dark until the end of the sequence.

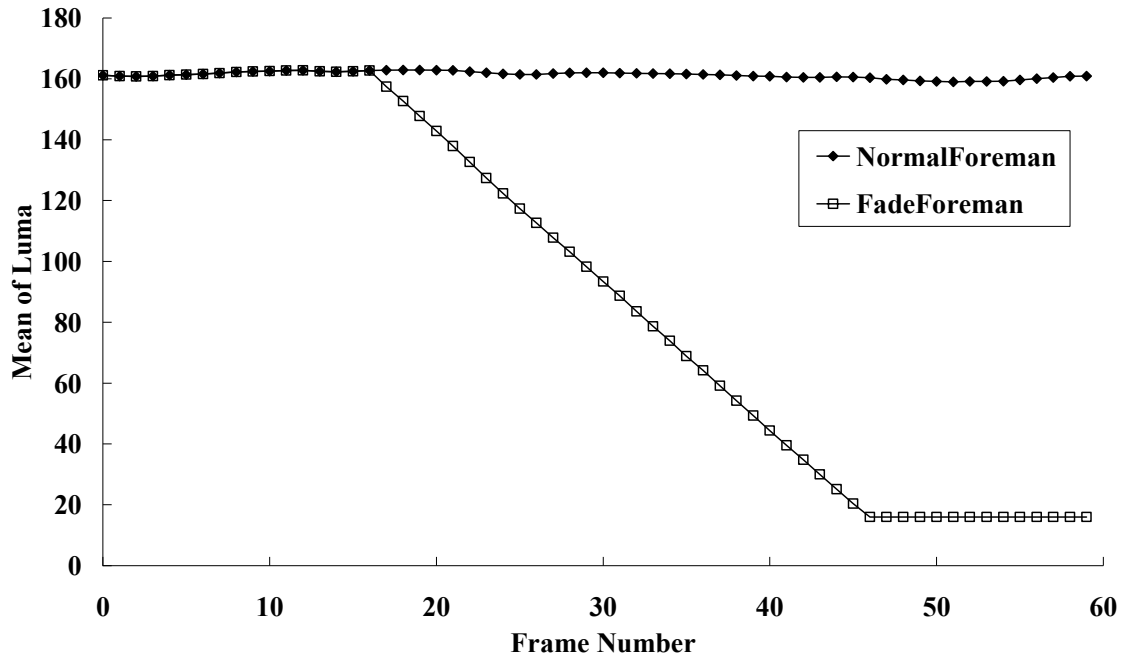


Figure 2.6. Average luma value of each frame for *NormalForeman* and *FadeForeman*.

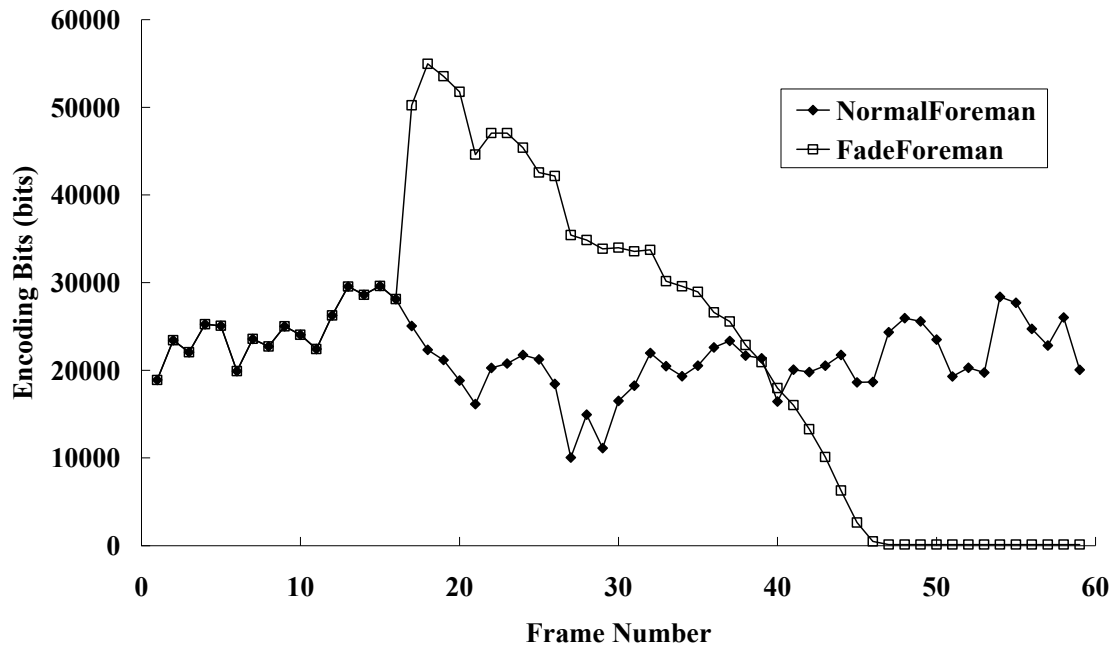


Figure 2.7. Encoding bits per frame for *NormalForeman* and *FadeForeman* at QP24.

Figure 2.7 then shows the coding rates of each frame for *NormalForeman* and *FadeForeman* with QPs (Quantization parameters) of 24 for both I and P frames. The IPPP coding pattern was used and the size of GOP was 60. Variable block size ME with quarter-pel was enabled in which both of the inter and intra modes were selected for RDO. MRF-ME with five reference frames was also used. As can be seen, the coding rates from 0th to 16th frames for both video segments are identical since there is no fading applied. At 17th frame of *FadeForeman*, the coding rate is suddenly boosted up due to the significant change in brightness level caused by fade-out-to-black effect. It in turn reduces the correlation between frames and motion estimation is no longer efficient within the period of fading. The evidence is shown in Figure 2.8 where the mode distributions for *NormalForeman* and *FadeForeman* from 17th to 46th frames (within the period of fading effect for *FadeForeman*) are shown. It can be observed that, for *NormalForeman*, the number of intra modes only possesses 0.73% while number of inter modes including the skip mode contributes over 99%. It is because temporal correlation

between frames is very high in which inter prediction can be utilized very efficiently. For *FadeForeman*, with the synthetic fade-out-to-black effect, the inter mode cannot dominate within the fading period. There are about 44% of inter modes and 56% of intra modes for coding the scenes with the fading effect. The reason behind is that temporal correlation decreases due to the fading effect and inter prediction fails to estimate object motions. Figure 2.9 further shows the MVs estimated at 29th frame of *NormalForeman* and *FadeForeman*. After applying the fading effect, MVs have been estimated wrongly in *FadeForeman*. Lengthy MVs are resulted even at static region where they are supposed to be zero-length MVs. This means that the occurrence of brightness variations upsets the ME in the sense that it reduces the competitiveness of inter prediction against intra prediction. From Figure 2.7, it is interesting to note that the required coding rates for *FadeForeman* are dropped. It is because *FadeForeman* is dimmer and finally becomes totally dark which only needs few bits to code those ‘black frames’.

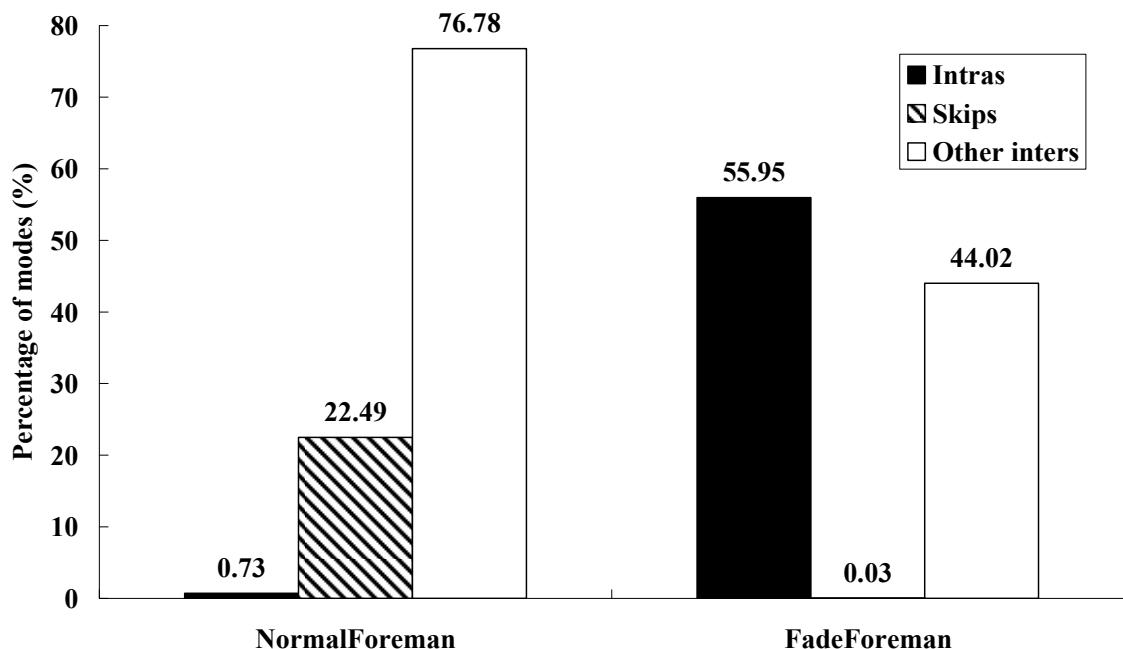


Figure 2.8. Mode distribution from 17th to 46th frames of *NormalForeman* and *FadeForeman* at QP24.



Figure 2.9. Motion vectors estimated at 29th frame of *NormalForeman* and *FadeForeman* at QP24.

From the above experiment, it can be concluded that if brightness level changes along the scene, temporal correlation between frames is reduced and coding efficiency is likely to be reduced. It is necessary to explore ways for enhancing ME and MC with the view to increasing coding efficiency even there are brightness variations in video sequences.

2.4 Conventional Weighted Prediction in H.264/AVC for Coding Scenes with Global Brightness Variations

Before the establishment of H.264/AVC, there have been many researches on video coding using block-based ME and MC to increase coding efficiency of scenes with different kinds of brightness variations [30-36]. In addition, there are other approaches for improving coding efficiency such as retinex based coding [37-39], wavelet based coding [40], inpainting based coding [41], and intra based coding [42-45]. For intra based coding which has no ME and MC, brightness compensation is done by using information of neighboring pixels. In this thesis, we focus on a weighted prediction (WP) tool in H.264/AVC [12-18,46-67], which is defined in the Main and Extended Profiles, for the

efficient coding of video scenes containing brightness variations caused by synthetic fade-in/out effects, camera flashes, camera iris adjustment, and local illumination change, etc.

In MC, the current frame is predicted from a reference frame and then only the prediction error is encoded. As aforementioned above, this simple motion compensation scheme assumes that the brightness of an object in a video scene keeps constant during motion. It then fails to detect true motion vectors when brightness variation happens, and increases the number of bits to encode prediction errors. To improve the coding efficiency for video scenes with brightness variations, WP can be used to enhance ME and MC by introducing a set of WP parameters which includes one multiplicative weighting factor W_i and one additive offset O_i . They can be assigned to each of the i^{th} reference frame f_i (where i is the associated reference frame index number) which are used to predict the current frame f_c and stored in the slice header of f_c . In other words, the reference index is to indicate which set of WP parameters being used. In ME with WP, the commonly used criterion in determining the temporal prediction block for a given block is the SAD between the current block b_c and the reference block b_i being weighted and shifted. It is defined as

$$SAD(b_c, b_i) = \sum_{\substack{p_c \in b_c \\ p_i \in b_i}} \left| p_c - \frac{W_i}{W_D} p_i - O_i \right| \quad (2.6)$$

where W_D is the weight denominator in which higher value of W_D would allow more fine-grained weighting factors W_i with the expense of more bits for coding the weighting factors W_i (the default value of W_D equals to 32) in the slice header. With this equation, the estimation of SAD is more complicated than that in (2.2) since there are additional operations including multiplication, addition and shift operations for division purpose. To speed up the process, each weighted pixel p_i^{wp} , which belongs to the weighted reference frame f_i^{wp} , is pre-calculated before ME by

$$p_i^{WP} = \frac{W_i}{W_D} \times p_i + O_i \quad (2.7)$$

The pre-calculated weighted reference frame f_i^{WP} is used instead of f_i in ME and MC to alleviate the problem of brightness variations. It is noted that p_i^{WP} is clipped between 0 and 255 since 8-bit-depth is generally used for digital videos. The SAD between current block b_c and the weighted reference block b_i^{WP} is then rewritten as

$$SAD(b_c, b_i^{WP}) = \sum_{\substack{p_c \in b_c \\ p_i^{WP} \in b_i^{WP}}} |p_c - p_i^{WP}| \quad (2.8)$$

In a video sequence with brightness variations, f_c is more strongly correlated to f_i^{WP} than to the reference picture itself. Thus it results in fewer bits to encode f_c .

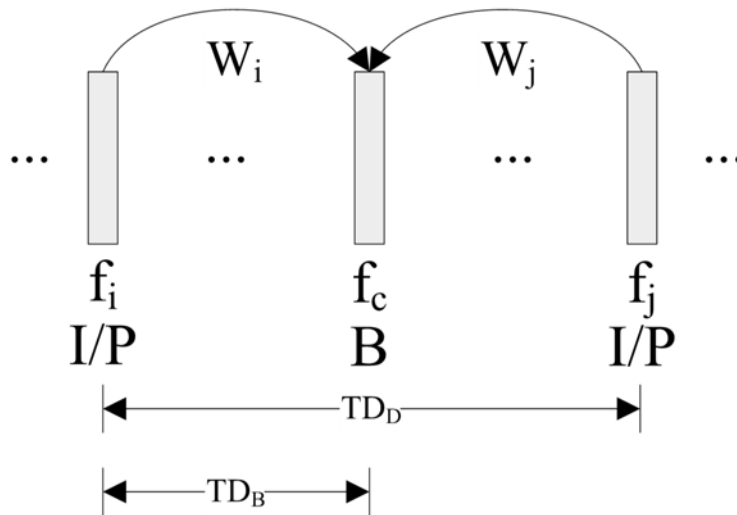


Figure 2.10. Illustration for implicit weighted prediction.

In H.264/AVC, there are two types of WP - explicit mode [12-16,46-54,56] and implicit mode [12,16,55]. In Figure 2.10, implicit WP estimates the weighting factors W_i and W_j

based on the relative temporal distances between the current frame and the reference frames and sets the offsets O_i and O_j as zero, which are represented by

$$\begin{aligned} W_j &= 64 \cdot TD_B / TD_D \\ W_i &= 64 - W_j \\ O_i &= O_j = 0 \end{aligned} \quad (2.9)$$

where TD_B is the temporal distance between the forward reference frame and the current frame whereas TD_D is the temporal distance between the forward reference frame and the backward reference frame. With bi-directional ME, the SAD between the current block b_c , and the weighted average of the forward reference block b_i and the backward reference block b_j is defined as

$$SAD(b_c, b_i, b_j) = \sum_{\substack{p_c \in b_c \\ p_i \in b_i \\ p_j \in b_j}} \left| p_c - \frac{1}{2} \left[\frac{W_i}{W_D} \cdot p_i + \frac{W_j}{W_D} \cdot p_j + O_i + O_j \right] \right| \quad (2.10)$$

where (W_i, O_i) , and (W_j, O_j) are the WP parameters associated with the i^{th} forward and j^{th} backward reference frames, respectively. Implicit WP can be applied in bi-directional prediction only and it is mainly used in encoding dissolve scenes [49] as the equation (2.10) uses the temporal distances between frames instead of using pixel information to estimate WP parameters. In contrast, explicit WP can be applied on both uni-directional and bi-directional ME.

In this research, we only focus on explicit WP and how this technique could be used to improve the motion compensation performance for video sequences with brightness variation. For explicit WP, the standard does not define the way to derive WP parameters. Various models [12-15] for estimating WP parameters have been studied, and some of

them have also been adopted in the H.264 Joint Model (JM) reference software [16,17]. The basic concept is to use global brightness compensation and it assumes that brightness variation is uniformly applied across an entire picture or slice. In this situation, one set of WP parameters is sufficiently enough to code efficiently all MBs in the picture or slice that are predicted from the same reference frame. For the sake of simplicity, we assume that each picture contains one slice only in this thesis. There are mainly four WP models to estimate the WP parameters - DC model [12,16], offset model [16], LS model [13] and LMS model [14-16]. They can handle diverse fading scenarios in video and a detailed description will be given in the following sub-sections.

2.4.1 DC model

In the DC model [12,16], the multiplicative weighting factor W_i^{DC} for the i^{th} reference frame is calculated as the ratio of the mean pixel value of the current frame ($\overline{f_c}$) and the mean pixel value of the i^{th} reference frame ($\overline{f_i}$) whereas the additive offset O_i^{DC} is set to zero. They are represented as follows:

$$\begin{aligned} W_i^{DC} &= W_D \cdot \overline{f_c} / \overline{f_i} \\ O_i^{DC} &= 0 \end{aligned} \quad (2.11)$$

Note here that $\overline{f_n}$ denotes the pixel mean over the n^{th} frame with $W \times H$ pixels where W and H are the width and height of the frame respectively. $\overline{f_n}$ is then given by

$$\overline{f_n} = \frac{1}{W \times H} \sum_{p_n \in f_n} p_n \quad (2.12)$$

This simple model has been adopted in the H.264 JM reference software [16,17] in early time. The authors in [13] proved that the DC model is more efficient for coding the scenes

with black fades (fade-in-from-black or fade-out-to-black) than that with white fades (fade-in-from-white or fade-out-to-white). It is because the estimated weighting factor W_i would be equal to W_D after rounding when both of $\overline{f_c}$ and $\overline{f_i}$ are very large in scenes with white fades. When W_i is equal to W_D , SAD estimation with WP using (2.6) or (2.8) is exactly identical to SAD estimation without WP using (2.2). The use of WP becomes meaningless.

2.4.2 Offset model

In contrast, the offset model [16] simply calculates the offset O_i^{OFF} as the difference between $\overline{f_c}$ and $\overline{f_i}$, and sets the weighting factor W_i^{OFF} as W_D , which can be written as

$$\begin{aligned} W_i^{OFF} &= W_D \\ O_i^{OFF} &= \overline{f_c} - \overline{f_i} \end{aligned} \tag{2.13}$$

This offset model has also been adopted in the H.264 JM reference software [16,17]. It can estimate the additive offset from the slight difference between $\overline{f_c}$ and $\overline{f_i}$ in coding the scenes with white fades in which the DC model cannot. In addition, the offset model is more suitable for the flashlight scenes [41,69,71].

2.4.3 LS model

Recently, some quasi-optimal WP parameter estimators [13-16] have been derived to estimate more accurate WP parameters compared with the DC and offset models. For instance, a determination model that uses the least square (LS) technique to optimize the SAD function in (2.6) is proposed in [13]. Its WP parameters can be computed by:

$$\begin{aligned}
W_i^{LS} &= W_D \frac{\overline{f_c \cdot f_i} - \overline{f_c} \cdot \overline{f_i}}{\overline{f_i^2} - \overline{f_i}^2} \\
O_i^{LS} &= \frac{\overline{f_c \cdot f_i^2} - \overline{f_c} \cdot \overline{f_i} \cdot \overline{f_i}}{\overline{f_i^2} - \overline{f_i}^2}
\end{aligned} \tag{2.14}$$

with $\overline{f_c \cdot f_i}$ given by

$$\overline{f_c \cdot f_i} = \frac{1}{W \times H} \sum_{\substack{p_c \in f_c \\ p_i \in f_i}} (p_c \cdot p_i) \tag{2.15}$$

The LS model depends on the term $\overline{f_c \cdot f_i}$ which is the mean of the product of the pixel values in the current frame and the pixel values at the same position in the reference frame. It implies that it is highly sensitive to object motion or camera movement. Consequently, the LS model has the problem of being sufficiently accurate only when true motion vectors are estimated prior to WP parameter estimation [14]. In coding fading scenes, true motion vectors cannot be obtained without true WP parameters. Therefore, (2.14) solely works well in the scene with low motion activity.

2.4.4 LMS model

Another quasi-optimal WP parameter estimator is a least mean square (LMS) [14-16] model. For this model, the WP parameters W_i^{LMS} and O_i^{LMS} are derived based on the equation of applying fading effect, which are modelled as:

$$\begin{aligned}
W_i^{LMS} &= W_D \frac{\sum_{p_c \in f_c} |p_c - \overline{f_c}|}{\sum_{p_i \in f_i} |p_i - \overline{f_i}|} \\
O_i^{LMS} &= \overline{f_c} - \overline{f_i} \cdot W_i^{LMS} / W_D
\end{aligned} \tag{2.16}$$

This model is effective in video scenes with an artificial fading effect since the derivation

of WP parameters is theoretically derived from formulae of applying fading from/to a fixed color only, i.e. equation (2.5). Otherwise, it might not work well. The LMS model has then been adopted in the H.264 JM reference software [16,17] in recent years.

To conclude, WP in H.264/AVC is useful for coding scenes with GBVs as WP parameters evaluate the brightness differences between the entire frames and compensate the brightness changes during ME and MC which can help to increase the temporal correlations between frames. Unfortunately, brightness variations might be non-uniform in a picture or different types of fading effects may be applied to different video segments. It implies that a single set of WP parameters might not be sufficient, and this limits the efficiency of WP for coding scenes with non-uniform brightness variations. Moreover, the performance is not satisfactory for any single WP model that is needed to support sequences with diverse fading effects.

2.5 Previous Algorithms for Coding Scenes with Local Brightness Variations

Obviously, weighted prediction in H.264/AVC can only code the scenes with GBVs since there is only one set of WP parameters for the purpose of illumination compensation. Video scenes captured by amateurs or in outdoor environments can often have local brightness variations (LBVs) which cause the non-uniform intensity change distributed over the entire picture. It is then difficult to estimate accurate motions even the conventional WP is enabled. It is because WP uses the information of the entire frames to estimate the WP parameters which only can model the global brightness changes between frames, i.e. GBVs. Thus, some WP algorithms have been proposed to code the scenes

with LBVs in the literature [32,35,46-66].

To tackle this problem, WP parameters could be derived and assigned in every MB [35,64]. However, it is difficult to estimate accurate WP parameters for each MB since MB is a small region with only the size of 16×16 pixels. Inaccurate WP parameters are computed based on the information of the current MB and its co-located MB only. It is due to the fact that the WP parameters for MBs with fast object motions cannot be estimated accurately without its true motion vector, and the true motion vectors cannot be obtained without the accurate the WP parameters. This forms a chicken-egg dilemma that some inaccurate WP parameters are obtained in MBs with object motions. To resolve this chicken-egg dilemma, previous MB-based WP algorithms suggested that WP parameters should be estimated for every searching points during ME. Unfortunately, computational complexity is highly increased [64]. In [63], a two-pass search algorithm for illumination compensation in multi-view video was proposed. In the first pass, it uses a mean-removed search to compute W and O for each MB candidate in the search window to find the disparity vectors. Based on the disparity vectors, depth levels are found and new filtered reference frames are generated for the second mean-removed search to find the best match for each MB. This two-pass algorithm can also decrease computational complexity, but is only well suited to multi-view video coding. In the following sub-sections, two approaches for coding scenes with LBVs, which are used for making comparison with our proposed algorithms, are discussed. One is called adaptive weighted prediction and the other one is called localized weighted prediction.

2.5.1 Adaptive weighted prediction approach

In MB-based schemes [35,64], different MBs in the current frame can use different WP parameters in order to solve the problem of LBV. However, WP parameters need to be assigned and encoded in every MB, which induces excessive overhead bits in the encoded bitstream. Shen *et al.* [57] then proposed an adaptive WP approach, and it has been adopted in the AVS (Audio and Video coding Standard), which is the standard initiated by AVS Workgroup of China [58]. In this approach, if there is only a partial region of a frame having brightness variation, WP parameters are estimated by only using the pixel information of that region.

Firstly, WP parameters are estimated in MB basis by using LS model [13] in (2.14). Each pair of MB-based WP parameters $(W_i^{LS}(MB_c), O_i^{LS}(MB_c))$ is calculated depending on MB in the current frame, MB_c , and the co-located MB in the reference frame, MB_i , which is given by

$$\begin{aligned} W_i^{LS}(MB_c) &= W_D \frac{\overline{MB_c \cdot MB_i} - \overline{MB_c} \cdot \overline{MB_i}}{\overline{MB_i^2} - \overline{MB_i}^2} \\ O_i^{LS}(MB_c) &= \frac{\overline{MB_c \cdot MB_i^2} - \overline{MB_c} \cdot \overline{MB_i} \cdot \overline{MB_i}}{\overline{MB_i^2} - \overline{MB_i}^2} \end{aligned} \quad (2.17)$$

with

$$\begin{aligned} \overline{MB_n} &= \frac{1}{16 \times 16} \sum_{p_n \in MB_n} p_n \\ \overline{MB_m \cdot MB_n} &= \frac{1}{16 \times 16} \sum_{\substack{p_m \in MB_m \\ p_n \in MB_n}} (p_m \cdot p_n) \end{aligned} \quad (2.18)$$

where m and n are arbitrary frame indices.

After estimating the WP parameter set for every MB, the most common set of $(W_i^{LS}(MB_c), O_i^{LS}(MB_c))$ is chosen for motion estimation and compensation. An MB-based ON/OFF controller is then designed to classify MBs into two groups. One group uses WP in motion estimation and compensation whereas another group employs the normal motion estimation and compensation without WP. Each MB then needs one extra bit to indicate the usage of WP. This scheme is efficient only for coding scenes with simple LBV. For scenes containing complex LBVs, this scheme cannot work well.

2.5.2 Localized weighted prediction approach

Another approach [60] to avoid the transmission of excessive overhead bits for WP parameters in every MB is to estimate MB-based WP parameters by using the neighboring pixel values of the MB to be encoded and those of MB in the reference frame such that no additional bits are required for WP parameters. Figure 2.11 illustrates this localized WP approach. In this figure, n_c and n_i are the neighboring pixels of the block b_c in f_c and the reference block b_i in f_i , respectively. The WP parameters are then computed based on n_c and n_i , which are given by

$$\begin{aligned} W_i^{OFF}(b_c) &= W_D \\ O_i^{OFF}(b_c) &= \overline{n_c} - \overline{n_i} \end{aligned} \tag{2.19}$$

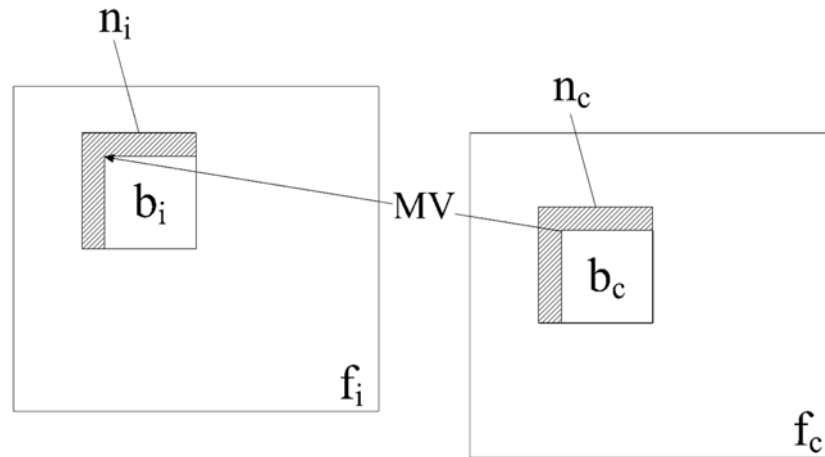


Figure 2.11. Illustration for localized weighted prediction.

where \bar{n}_c and \bar{n}_i denotes the mean of n_c and n_i . Note that these WP parameters are estimated according to offset model [16] defined in equation (2.13), and the offset is equal to the difference of the average pixel values of neighboring areas n_c and n_i . This approach does not need to explicitly code and transmit the WP parameters used to predict the current block. Instead, the weighting parameters are obtained on the fly in the decoder for each MB based on previously decoded spatial neighboring samples and its motion compensated samples. These samples are identical in both of the encoder and decoder. It implies no additional bits are required to be coded and transmitted. Nevertheless, noisy and irrelevant pixels in the neighbouring MBs cause inaccurate WP parameters.

2.6 Chapter summary

In order to solve the difficulties of coding scenes with GBVs and LBVs, many recently proposed algorithms have been reviewed in this chapter. We started this chapter by reviewing the compression techniques employed in current video standards. The redundancy exploited by these compression techniques provides good compression

effects. These compression techniques assume brightness levels between frames are constant. By using the structure of the conventional system, we systematically analyzed the impact of brightness variations. When brightness variation occurs, it results in much higher bit rate for encoding the bitstream due to ineffectiveness of inter-prediction. Next, we introduced weighted prediction tool which is newly employed in H.264/AVC video coding standard. We also reviewed several frame-based WP models for estimating WP parameters with the view to improving coding efficiency when there are scenes with GBVs. We found that different WP models are dedicated to different kinds of fading effects and are only restricted to solve the scenes with GBVs. In addition, we briefly reviewed some previous research approaches for solving the problems of LBVs with the expense of high computational complexity. The reviews of various algorithms in this chapter indicate that these methods are primitive, and there is a plenty of room for improvement. Therefore, in the following chapters, we examine the possibility of improving coding efficiency for the scenes with GBVs as well as LBVs.

Chapter 3 Multiple Weighted Prediction Models for Scenes with Global Brightness Variations

3.1 Introduction

In the last chapter, we have shown the impact of coding scenes with GBVs in the H.264/AVC standard. Enabling a weighted prediction tool in H.264/AVC is the straightforward approach. As aforementioned in the previous chapter, there are four WP models for computing the WP parameters. They are called the DC model [12,16] by (2.11), the offset model [16] by (2.13), the LS model [13] by (2.14), and the LMS model [14,15] by (2.16) accordingly. There are pros and cons to all the different models described in Chapter 2.4. It is therefore expected that no single WP model can cope with all situations of brightness variations. In this chapter, a single reference frame multiple WP models (SRefMWP) scheme is proposed to facilitate the use of multiple WP models in a single reference frame. The proposed scheme makes a new arrangement of the multiple frame buffers in multiple reference frame motion estimation. It enables different MBs in the same frame using different WP models even when they are predicted from the same reference frame. Furthermore, a new re-ordering mechanism for SRefMWP is also proposed to guarantee that the list of the reference picture is in the best order for further decreasing the bit rate. To reduce the implementation cost, a reduction of the memory requirement is achieved via look-up tables (LUTs).

Contents of this chapter have been published in references [52,54].

3.2 The Conventional WP in H.264/AVC with the Support of Multiple Reference Frame Motion Estimation

Multiple reference frame motion estimation (MRF-ME) is a new feature introduced in H.264/AVC to enhance coding performance by searching more than one reference frames [9-10]. In MRF-ME, a reference picture index (ref_idx) is coded to indicate which multiple reference frames are used [25-27]. In WP, a single WP parameter set is associated with each ref_idx , which is encoded in the slice header. If MRF-ME is enabled, more than one WP parameter set is sent per slice. Figure 3.1 shows a block diagram to illustrate the use of WP with MRF-ME in the H.264 encoder. A single WP model, say DC model, is applied to multiple reference frames from f_0 to f_4 for motion estimation. Different weighting factors, W_0^{DC} , W_1^{DC} , W_2^{DC} , W_3^{DC} , and W_4^{DC} , are computed according to (2.11) for f_0 to f_4 , and their weighted reference frames (f_i^{WP} , where $i=0,1,\dots,4$) are placed in the multiple frame buffers for ME, as depicted in Figure 3.1. By doing so, the decoder can recognize the set of WP parameters correctly. Since in MRF-ME [25-27], ref_idx is already available in the bitstream, the use of this index to indicate which set of WP parameters for each MB can avoid the need of additional bits, which results in increasing coding efficiency.

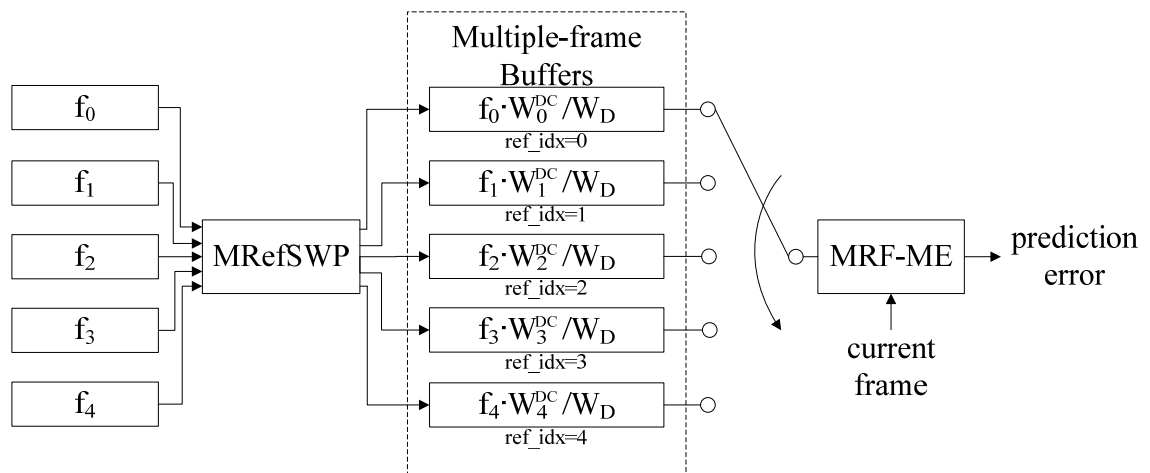


Figure 3.1. Conventional weighted prediction in H.264/AVC – MRefSWP. Nevertheless, this default arrangement is based on the hypothesis that brightness variations are global to the image and they affect pixels of equal greyscale value in the same manner by one particular WP model. It is hereafter referred to as the multiple reference frames single WP model (MRefSWP). In practice, however, brightness variations may be concentrated in different regions of the image and vary spatially. For instance, one region of the picture may be better coded with the DC model while another region may be better coded by using the LMS model. If the WP parameter set is not accurate enough for the current MB, coding efficiency would be reduced due to the surge of prediction errors between the current frame and the reference frame. Thus, it is a key issue to have an appropriate WP model to estimate an accurate set of WP parameters in order to deal with the brightness variation problems. But as aforementioned, different WP models are suitable for different kinds of brightness variations or fading effects. The selection of WP models in advance of WP parameter estimation could not be practical for a variety of reasons. First, a complicated process is needed to detect the existence and types of brightness variations in order to select the most appropriate WP model. For example, fading effects may consist of fade-in from black/white and fade-out to black/white, the coding performance is very sensitive to adoption of WP models in various fading effects. To the best of our knowledge, no method in the literature has been proposed to determine the multiple WP models that can be used in H.264/AVC. Second, most of the brightness variation detection algorithms in the literature depend on a relatively long window of frames to analyze enough statistics for an accurate detection [68-71]. For instance, a method in [68] employs the average luminance changes and semi-parabolic behavior of the variance curve to distinguish various fading effects. A method of using accumulating histogram difference was proposed to detect various fading effects and flashlight [71]. All these methods necessitate the availability of the statistics of the

entire fading duration, which induces long delay and is impractical for encoding. Third, even the detector is smart enough to select a particular WP model in the encoded frame, due to localized brightness variations, uncovered objects, object movements, and camera operations, only a single WP model for a frame cannot perform satisfactorily. For instance, the LS method may be well compensated for most MBs, but the performance is poor and sensitive to some MBs with high motion activity.

3.3 The Proposed Single Reference Multiple WP Models

In this section, we examine the way to joint use of multiple frame buffers and weighted prediction such that more than one ref_idx can be associated with a particular reference picture, and this allows different MBs in the current frame to use different WP parameter sets even they are predicted from the same reference picture. Figure 3.2 shows the new arrangement of the multiple frame buffers for the proposed single reference frame multiple WP models (SRefMWP) scheme. In this Figure, instead of using a single model for multiple reference frames from f_0 to f_4 , different WP models are applied to a single reference frame, f_0 , for motion estimation and compensation. As DC, offset, LS and LMS models are the most common WP models used for coding scenes with brightness variations and can handle different types of brightness variations, these four mentioned WP models are chosen for compensating each other. Different parameter sets, $(W_{f_0}^{DC}, O_{f_0}^{DC})$, $(W_{f_0}^{OFF}, O_{f_0}^{OFF})$, $(W_{f_0}^{LS}, O_{f_0}^{LS})$, and $(W_{f_0}^{LMS}, O_{f_0}^{LMS})$ are estimated between the current frame and f_0 . Then, their weighted reference frames are stored in the multiple frame buffers for motion estimation and compensation, as shown in Figure 3.2. These weighted reference frames are associated with different WP parameters. This arrangement allows different MBs in the same frame to employ different WP parameters even when they are predicted from

the same reference frame. Again, ref_idx is used to select which WP model to be used for each MB. An encoder that uses the proposed buffer arrangement can select WP parameters from different models through rate-distortion optimization (RDO). It can handle various kinds of brightness variations or fading effects in the same picture without the need of the brightness variation detector. Since the proposed algorithm is able to apply multiple WP models to the same reference frame for generating multiple weighted reference frames, it can also potentially improve the prediction for the scenes with local brightness variation. It is interesting to note that the reference frame without WP is also kept in the multiple frame buffers for handling scenes in which the brightness variations are mainly due to sudden camera motion, but not due to fading. For example, the salient character of a fast camera panning shot induces the luminance change which is caused by abrupt appearance or disappearance of video objects. This scenario always misleads the encoder in its use of WP, but the original reference frame reserved in the multiple frame buffers can avoid this problem.

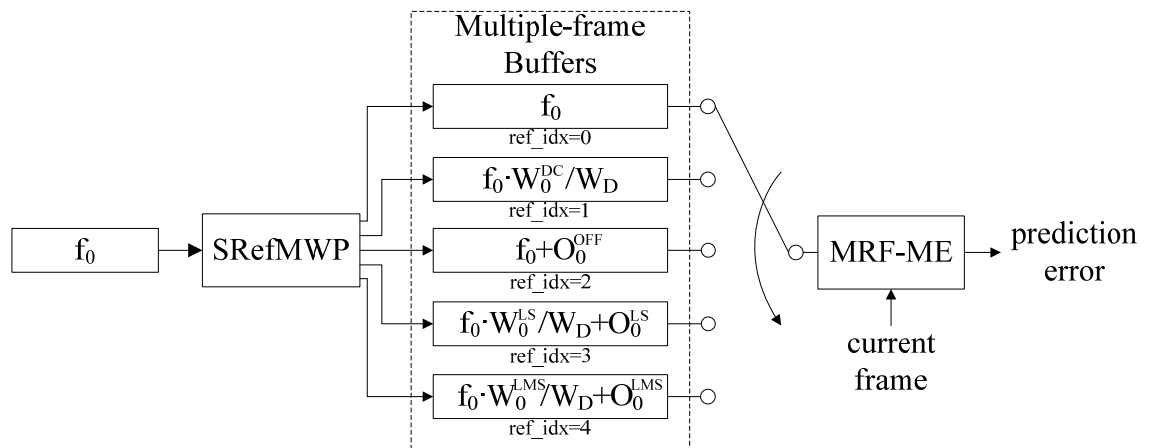


Figure 3.2. Proposed weighted prediction scheme – SRefMWP.

We also consider arranging the reference list of the multiple frame buffers so that the weighted reference frame giving the best prediction are placed first in the list. In

H.264/AVC, the default list of the reference frames is based on display order. The order of this default list is very reasonable due to temporal proximity. In SRefMWP, the encoding results can be improved significantly if more than one WP model is available for motion compensation. In this case, different WP models use the same reference frame, and the default order starting with the most recent frame is no longer applied to the reference list. To solve this, SRefMWP can work with the mechanism of reference list re-ordering defined in H.264/AVC. This mechanism allows the encoder and decoder to re-order the reference list in the best order. In order to determine the best order of SRefMWP, an algorithm is needed to shuffle the reference list. Since fading is always applied over a few seconds, the correlation between frames that uses a particular WP model remains reasonably high. By making use of this property, the proposed scheme determines which WP model is likely to be used in the 8x8 blocks of the previously encoded frame, and the probabilities of using various WP models are computed. The reference picture list used in the current frame is re-ordered based on these probabilities so that the most likely WP model to be used is placed at the top of the list. In other words, the placement of weighted reference frames in the reference list is sorted according to these occurrence probabilities of the 8x8 blocks of the previously encoded frame, starting with the most frequently occurring WP models. This allows using shorter codes for *ref_idx* in the encoded bitstream, which results in further reduction in the bit rate. It is noted that the statistics of using various WP models of each 8x8 block is collected from the previously encoded frame. It means that the decoder can also compute the statistics in order to re-order the reference picture list as the encoder. In this case, it can maintain the consistency between the encoder and decoder without requiring additional signaling.

3.4 SRefMWP using Pre-calculated Look-up Tables (LUTs)

For MRefSWP and SRefMWP as shown in Figure 3.1 and Figure 3.2, five frame buffers are used. Given a frame size of $W \times H$ where W and H are the width and height of the frame respectively, the size of memory requirement is $5 \times 4W \times 4H$, where a factor of 4 is due to the use of quarter-pel motion estimation. Consequently, the use of multiple reference frames in MRefSWP and SRefMWP consumes a significant amount of the memory. However, portable consumer devices such as camera phones have limited system memory due to cost constraints. Reducing memory requirement is of great important for handheld video devices.

SRefMWP has additional benefit by using multiple pre-calculated look-up tables (LUTs) to replace the multiple frame buffers such that the amount of memory required to store the reference frames can be greatly reduced. In SRefMWP as shown in Figure 3.2, the weighted reference frames in the frame buffers are formed by different WP parameters - (W_0^{DC}, O_0^{DC}) , (W_0^{OFF}, O_0^{OFF}) , (W_0^{LS}, O_0^{LS}) , and (W_0^{LMS}, O_0^{LMS}) . To locate the best matched candidate of the current MB, a number of candidate MBs for each weighted reference frame has to be searched during motion estimation. However, the weighted reference frames are all based on f_0 , but only in a modified form with different WP parameters. To avoid the use of the frame buffers, the LUT is generated once per weighted reference frame as shown in Figure 3.3. The entries of each generated LUT store the pixels of the weighted reference frame and can be computed by modifying f_0 with its associated WP parameters, as defined in (2.11), (2.13), (2.14), and (2.16) respectively. Then, four LUTs are formed to support multiple WP models. For an 8-bit-depth video signal, the range is limited to $[0, 255]$. Hence the LUT of 256 entries is sufficient. These look-up tables can

supersede the frame buffers during the processing of every MB candidate in the weighted reference frames in motion estimation and compensation. When a pixel of the weighted reference frame is needed, instead of accessing the actual weighted reference frame by (2.8), the encoder can simply look up the corrected pixels on the table by the index as shown in Figure 3.3. Then, for pixels in the weighted reference frames during motion estimation, SAD calculation in (2.6) can be replaced by

$$SAD(b_c, b_i) = \sum_{\substack{p_c \in b_c \\ p_i \in b_i}} |p_c - LUT_k[p_i]| \quad (3.1)$$

where $LUT_k[.]$ represents one entry of the four LUTs generated, and k represents the multiple WP models including DC, OFF, LS, and LMS as illustrated in Figure 3.3. With the help of this arrangement, the encoder only needs to keep reference frame f_0 as well as the four LUTs. The pixel values of f_0 are also used to retrieve the corresponding pixels of the other weighed reference frame in the LUT. The benefit of using LUTs to implement SRefMWP is to reduce the memory requirement in both the encoder and decoder. Note that each LUT consists of 256 bytes, which is negligible as compared with the size of frame memory. However, MRefSWP is not applicable since it needs multiple reference frames, and their weighted forms are not based on f_0 only.

	LUT ₀	LUT ₁	LUT ₂	LUT ₃
0	0	O_0^{OFF}	O_0^{LS}	O_0^{LMS}
1	W_0^{DC}/W_D	$1+O_0^{OFF}$	$W_0^{LS}/W_D+O_0^{LS}$	$W_0^{LMS}/W_D+O_0^{LMS}$
2	$2W_0^{DC}/W_D$	$2+O_0^{OFF}$	$2W_0^{LS}/W_D+O_0^{LS}$	$2W_0^{LMS}/W_D+O_0^{LMS}$
3	$3W_0^{DC}/W_D$	$3+O_0^{OFF}$	$3W_0^{LS}/W_D+O_0^{LS}$	$3W_0^{LMS}/W_D+O_0^{LMS}$
⋮	⋮	⋮	⋮	⋮
255	$255W_0^{DC}/W_D$	$255+O_0^{OFF}$	$255W_0^{LS}/W_D+O_0^{LS}$	$255W_0^{LMS}/W_D+O_0^{LMS}$

Figure 3.3. Look-up tables used in proposed SRefMWP scheme.

3.5 Experimental results

Various sequences with different characteristics were used for the experiment. These sequences include “Akiyo” (CIF, 352×288), “Football” (CIF, 352×288), “Foreman” (CIF, 352×288), “M&D” (CIF, 352×288), “Silent” (CIF, 352×288), “Flamenco2” (VGA, 640×480), “Mobisode1” (WVGA, 832×480), and “ShuttleStart” (HD 720p, 1280×720). For all testing sequences, the frame-rate was 30 frames/s. To simulate various fading effects, the H.264 JM 15.1 [16] was used to encode two-second long 60 frames with four kinds of synthetic fading effects applied to the CIF sequences [12]. These synthetic effects include fade-in from/fade-out to black/white. For “Flamenco2”, it is a centre view extracted from the multi-view sequence where frame 220 to frame 299 is a shot of lights spotting all around a stage with dancers. For “Mobisode2”, frame 42 to frame 51 is a shot of a guy turning on a light in a room which causes natural brightness variation. For “ShuttleStart”, frame 560 to frame 599 is a shot with a rocket getting off the earth which induces camera iris adjustment. The bitstreams were encoded with IPPP... structure based on the simulation conditions defined in [72] by different algorithms. All experiments were conducted using a GOP length of 60, Main profile, five reference pictures, quarter-pel full search motion estimation with search range of ± 32 pixels, RDO with all seven inter-modes as well as intra-modes, and context-adaptive binary arithmetic coding (CABAC). The encoded bitstreams were encoded by different schemes with a set of four different QPs (i.e. QP=20, 24, 28, and 32). It is noted that other settings such as fast motion estimation technique or RDO is off can also be applied for evaluating the performance.

We incorporated the proposed single reference frame multiple WP models (SRefMWP) schemes with and without LUTs into the H.264 JM 15.1 [16], and let us call them SRefMWP and SRefMWP+LUTs. They are used to compare with the conventional multiple reference frames single WP model (MRefSWP) schemes. Different models such as DC, Offset, LS, and LMS were adopted in MRefSWP for performance comparison, and they are denoted by MRefSWP-DC, MRefSWP-OFF, MRefSWP-LS, and MRefSWP-LMS, respectively. All schemes, except SRefMWP+LUTs, access the weighted reference frames stored in the multiple-frame buffers as shown in Figure 3.1 and Figure 3.2, and fetch the necessary data without further weighting calculation. They both require large memory requirement for both of the encoder and decoder since the multiple weighted reference frames must be maintained in memory. Given a frame of size $W \times H$ and N_{ref} weighted reference frames, the memory size requirement is:

$$N_{ref} \times 4W \times 4H \quad (3.2)$$

where a factor of 4 is due to the use of quarter-pel motion estimation. On the other hand, only one picture memory of $4W \times 4H$ for the reference frame f_0 is needed in SRefMWP+LUTs. In addition, an extra 256-byte LUT for each weighted reference frame is pre-computed for implementing (3.1) and stored in memory. The total size of the memory requirement for SRefMWP+LUTs is:

$$4W \times 4H + (N_{ref} - 1) \times 256 \quad (3.3)$$

Table 3.1 then lists the memory requirements for different schemes. It is observed that SRefMWP+LUTs has significant savings compared to other schemes, especially in “Mobisode2”. It is due to the fact that the frame size of “Mobisode2” is larger than that of other sequences, and SRefMWP+LUTs can still use a 256-byte LUT to replace the buffer of such a large frame size.

Table 3.1. Memory requirements (bytes).

Video Sequences	MRefSWP-DC/OFF/LS/LMS and SRefMWP	SRefMWP+LUTs
Akiyo		
Football		
Foreman	8,110,080	1,623,040
M&D		
Silent		
Flamenco2	24,576,000	4,916,224
Mobisode2	31,948,800	6,390,784
ShuttleStart	73,728,000	14,746,624

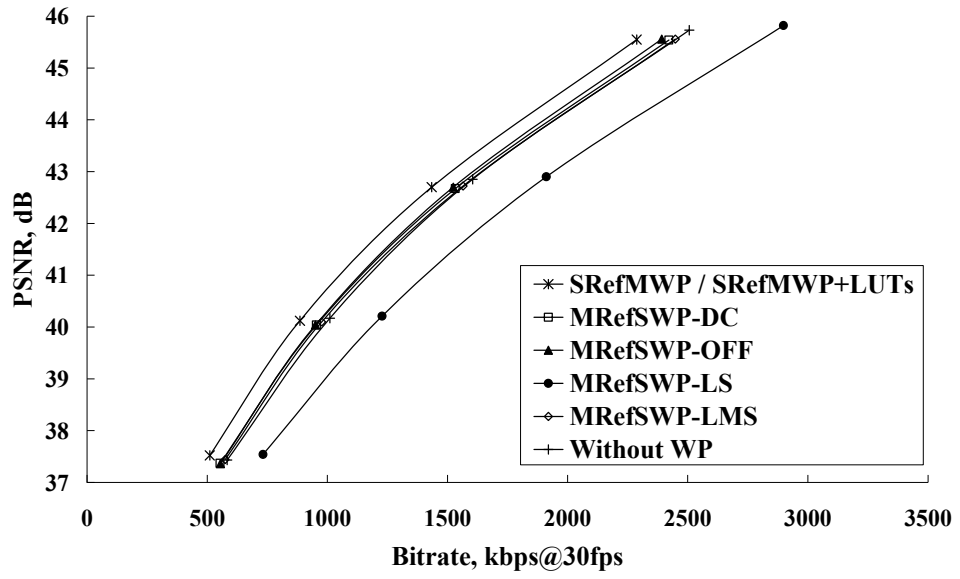
3.5.1 Rate-Distortion Performance of the Proposed Algorithm

Figure 3.4(a) and (b) show the rate-distortion (RD) performances of different schemes for “Football” with fade-in from black effect and fade-out to black effect, respectively. These figures also include the results when WP is not used, and it is referred to as ‘Without WP’. Note that the RD performances of SRefMWP and SRefMWP+LUTs are the same since the purpose of SRefMWP+LUTs is to reduce the memory usage of SRefMWP without affecting its coding efficiency. For simplicity, let us use the same curve to show their performances. From Figure 3.4(a) and (b), MRefSWP-LS performs the worst, and is even inferior to ‘Without WP’ as the LS model cannot work well in scenes with high motion activity such as “Football”. Among all MRefSWP schemes, MRefSWP-LMS achieves better results in Figure 3.4(b). However, there is no clearer winner in Figure 3.4(a). It is obvious from Figure 3.4(a) and (b) that by using the proposed SRefMWP and SRefMWP+LUTs, larger coding gains are obtained compared with all MRefSWP schemes for different fading effects. This gain of SRefMWP and SRefMWP+LUTs is due to the benefit of utilizing the architecture of multiple reference frames, they have successfully chosen the most effective WP model for handling various types of fading

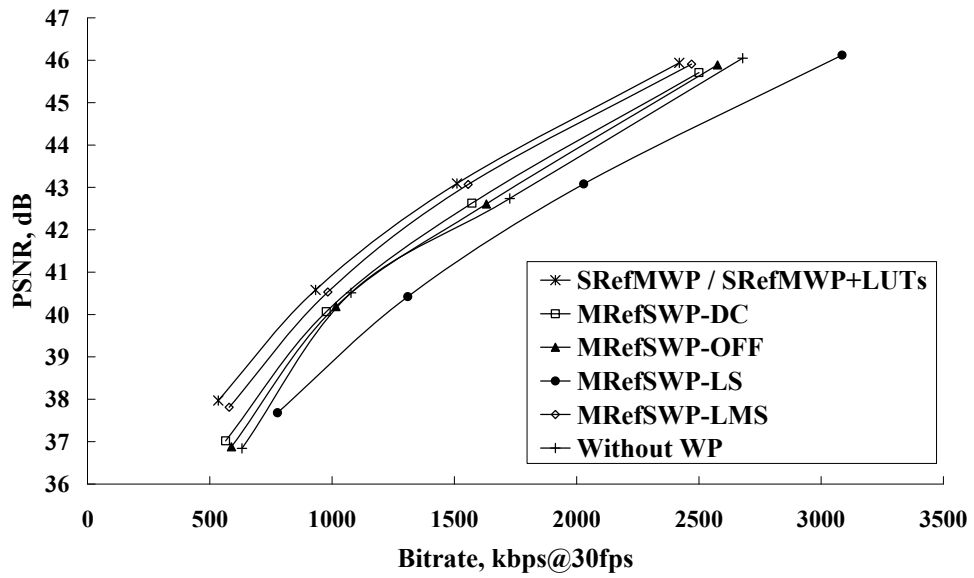
effects. It can also be explained by Figure 3.5(a) and (b) in which the percentage selection of different WP models of 8×8 blocks for “Football” with different fading effects at QP=20 is shown when SRefMWP or SRefMWP+LUTs is adopted. Note that “Intra” in this figure means the percentage of 8×8 blocks selecting intra-modes without predicting from any reference frames. The intra-modes are used when the inter-correlation between the current frame and the reference frame in which all WP models cannot work well. As expected, appropriate WP models are selected for different frames through our proposed SRefMWP and SRefMWP+LUTs. It means that not only one WP model can get an absolute advantage along the fading period, but every WP model would have the probability to get the smallest cost via RDO. Therefore, SRefMWP and SRefMWP+LUTs have the advantage over MRefSWP. For different fading effects, a considerable discrepancy between Figure 3.5(a) and (b) of selected WP models in 8×8 blocks is also shown. It further implies that SRefMWP and SRefMWP+LUTs are capable of selecting an appropriate WP model for each block to improve the coding efficiency. It is contrast to the case of MRefSWP where only one model is adopted for different fading effects. Experimental results for different sequences with various fading effects using the Bjontegaard delta bitrate (BDBR) and Bjontegaard delta PSNR (BDPSNR) [73] are summarized in Table 3.2. Again, the values of BDBR and BDPSNR are the same for SRefMWP and SRefMWP+LUTs, and they are put in the same column. In Table 3.2, it is obvious that SRefMWP and SRefMWP+LUTs can overwhelmingly outperform other MRefSWP schemes as well as ‘Without WP’.

Table 3.3 also shows the average percentages of intra-modes being used for all sequences. From this table, we can see that SRefMWP and SRefMWP+LUTs can substantially reduce the number of intra-modes in comparison with other MRefSWP schemes. This can

be explained as follows. In SRefMWP and SRefMWP+LUTs, various brightness variations can be compensated by weighted prediction which then increases temporal correlation between frames. Consequently, the temporal redundancy can be exploited more effectively during motion estimation and the coding process is favorable to inter-modes. Hence the inter-modes are more preferable than the intra modes, resulting in smaller number of intra-modes of the encoded bitstream. As a result, SRefMWP and SRefMWP+LUTs can achieve higher coding efficiency as shown in Table 3.2. The last column of Table 3.2 also shows the results of SRefMWP and SRefMWP+LUTs without the re-ordering mechanism. With the help of the re-ordering mechanism, it can be seen that SRefMWP and SRefMWP+LUTs can further provide a decrease of BDBR or an increase of BDPSNR. This means that the re-ordering mechanism can help the attainment of shorter codes for coding the reference indices. The additional re-ordering mechanism seems to be very effective to improve coding efficiency for various kinds of fading effects.

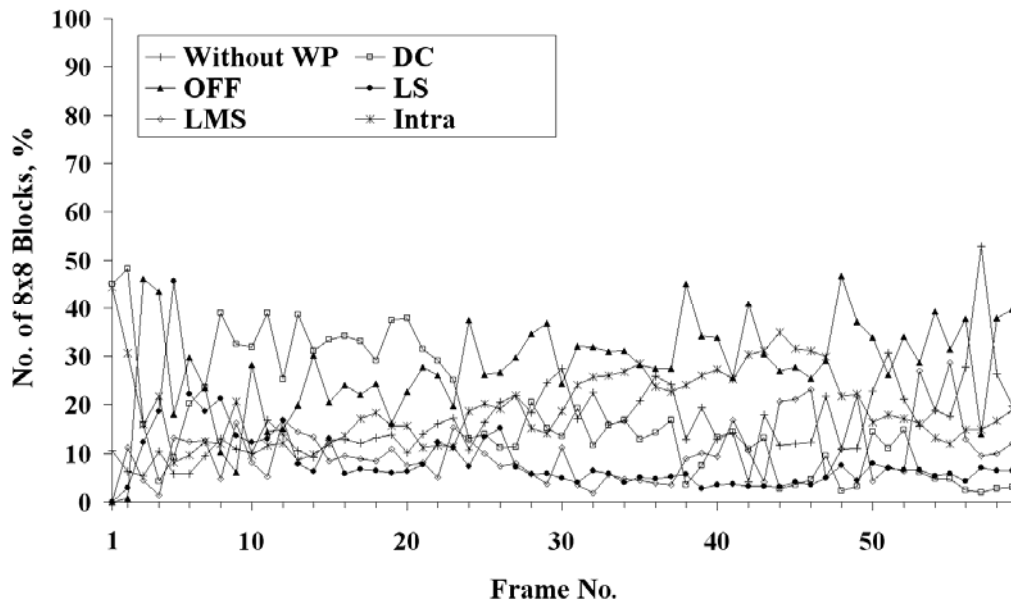


(a)

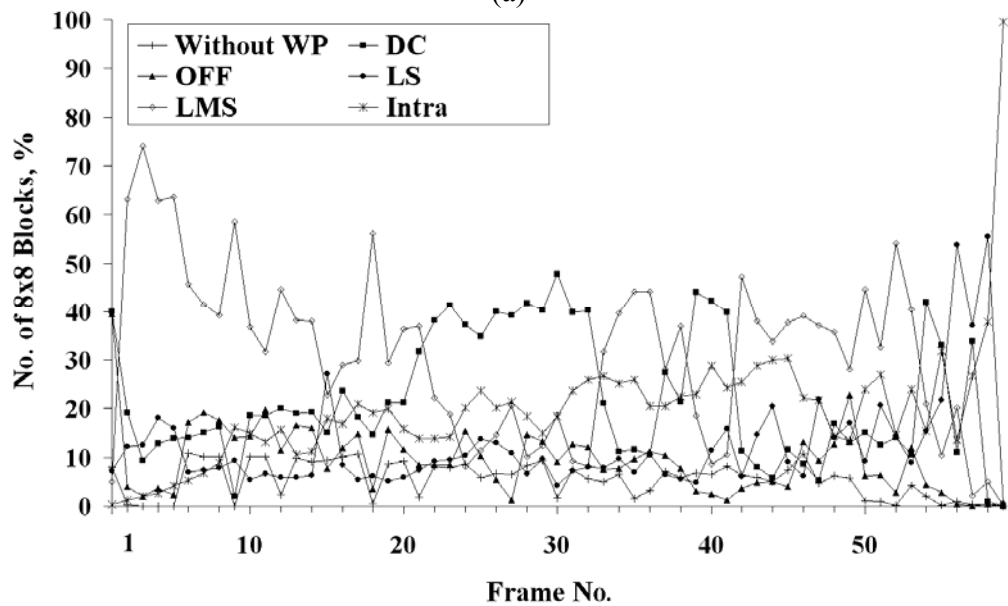


(b)

Figure 3.4. RD performances of different schemes for “Football” with (a) fade-in from black effect, and (b) fade-out to black effect.



(a)



(b)

Figure 3.5. Statistics of selected WP models in percentage of 8×8 blocks for “Football” with (a) fade-in from black effect, and (b) fade-out to black effect at QP 20 using SRefMWP/SRefMWP+LUTs.

For “Mobisode2”, all MRefSWP-DC, MRefSWP-OFF, MRefSWP-LS, and MRefSWP-LMS cannot obtain noticeable performance. At most 0.4 dB PSNR improvement over ‘Without WP’ is obtained, as depicted in Figure 3.6, whereas the proposed SRefMWP and SRefMWP+LUTs can achieve a remarkable gain, about 1.1dB over the ‘Without WP’. In this sequence, the brightness variation does not come from the artificial fading effect. In contrast, it includes a scene with natural brightness variation in which a single WP model is not sufficient to achieve good coding performance. Again, the merit of our proposed schemes is to allow the adaptation of different WP models in the scenario of natural brightness variation. Results of other sequences with natural brightness variations including “Flamenco2” and “ShuttleStart” are also shown in Table 3.2. It is observed that an average BDBR saving of 33.52% or a corresponding BDPSNR gain of 1.17dB is achieved by the proposed SRefMWP and SRefMWP+LUTs. Using MRefSWP-DC, MRefSWP-OFF, MRefSWP-LS, and MRefSWP-LMS can provide at most an average BDBR saving of 22.86% or a corresponding BDPSNR gain of 0.8dB. As a result, SRefMWP and SRefMWP+LUTs have superior performance for the sequences with natural brightness variations.

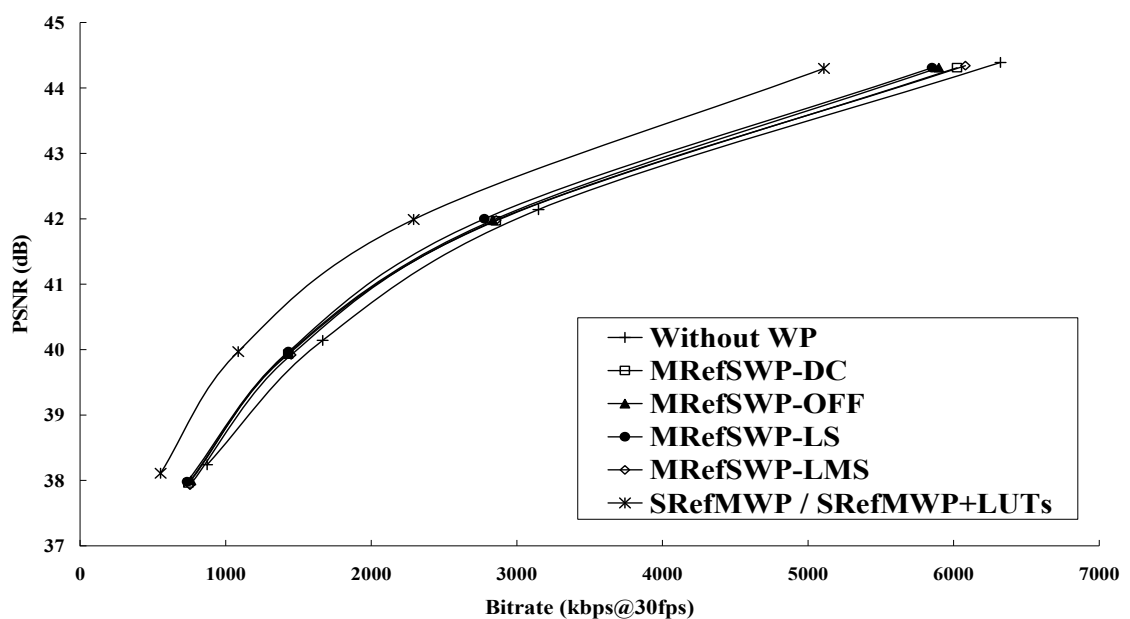


Figure 3.6. RD performances of different schemes for “Mobisode2”.

Table 3.2. BDBR (%), BDPSNR (dB) and encoding time reduction (%) of various schemes compared to ‘Without WP’.

Video Sequences	MRefSWP-DC			MRefSWP-OFF			MRefSWP-LS			MRefSWP-LMS			SRefMWP / SRefMWP+LUTs			SRefMWP / SRefMWP+LUTs (without re-ordering)		
	BD BR	BD PSNR	Time	BD BR	BD PSNR	Time	BD BR	BD PSNR	Time	BD BR	BD PSNR	Time	BD BR	BD PSNR	Time	BD BR	BD PSNR	Time
Fade-in from black																		
Akiyo	-39.02	2.61	-35.92	-16.81	1.07	-16.33	-40.99	2.69	-43.44	-38.35	2.25	-42.07	-53.94	3.98	-46.15/-38.85	-51.91	3.82	-46.65/-38.71
Football	-2.34	0.13	-12.52	-3.04	0.18	-10.4	19.04	-1.03	13.26	-1.09	0.07	-12.67	-9.94	0.58	-25.76/-13.53	-10	0.58	-24.36/-9.84
Foreman	-32.58	1.76	-40.47	-20.69	1.08	-27.63	-29.7	1.46	-34.4	-36	1.89	-44.08	-46.28	2.54	-47.23/-38.64	-44.95	2.46	-44.75/-35.98
M&D	-38.68	2.27	-47.03	-28.9	1.72	-38.35	-38.71	2.12	-47.46	-37.13	1.93	-48.02	-53.29	3.47	-50.88/-43.57	-51.62	3.35	-49.28/-41.69
Silent	-35	1.9	-34.83	-21.35	1.14	-23.73	-34.84	1.86	-35.91	-35.48	1.91	-37.88	-46.61	2.68	-41.92/-32.71	-45.09	2.61	-40.55/-30.67
Average	-29.52	1.73	-34.15	-18.16	1.04	-23.29	-25.04	1.42	-29.59	-29.61	1.61	-36.94	-42.01	2.65	-42.39/-33.46	-40.71	2.56	-41.12/-31.38
Fade-out to black																		
Akiyo	-57.43	4.52	-41.22	-19.05	1.15	-19.58	-68.62	5.41	-51.36	-72.48	6.13	-56.01	-73.23	5.83	-47.48/-41.24	-69.45	5.58	-47.19/-40.79
Football	-5.14	0.31	-14.98	-2.3	0.15	-11.82	15.86	-0.91	11.69	-12.18	0.77	-16.06	-16.57	1.03	-23.87/-9.99	-13.65	0.84	-23.51/-9.48
Foreman	-48.77	2.89	-40.07	-23.13	1.25	-22.37	-48.46	2.64	-33.08	-61.5	4.04	-48.05	-61.82	3.75	-43.78/-35.26	-61.89	3.86	-43.41/-33.48
M&D	-61.7	4.62	-52.07	-29.85	1.97	-34.02	-68.18	5.39	-54.76	-72.66	6.11	-59.22	-72.98	6.17	-52.80/-46.58	-70.41	5.6	-52.65/-46.56
Silent	-56.99	3.87	-44.46	-23.53	1.23	-26.01	-64.81	4.73	-47.78	-69.11	5.44	-52.35	-69.89	5.41	-47.73/-40.63	-69.87	5.31	-47.95/-40.20
Average	-46.01	3.24	-38.56	-19.57	1.15	-22.76	-46.84	3.45	-35.06	-57.59	4.5	-46.34	-58.9	4.44	-43.13/-34.74	-57.05	4.24	-42.94/-34.10
Fade-in from white																		
Akiyo	-13.94	0.79	-26.36	-29.24	1.84	-34.39	-45.75	3.07	-37.15	-48.78	3.04	-42.29	-61.35	4.47	-47.49/-38.63	-57.98	4.23	-47.45/-38.31
Football	-2.53	0.14	-11.92	-4	0.23	-15.2	14.37	-0.79	11.15	-4.01	0.23	-5.82	-10.14	0.59	-23.27/-9.58	-10.17	0.59	-23.22/-8.28
Foreman	-3.43	0.16	6.34	-6.87	0.33	-1.5	-11.83	0.53	7.74	-22.7	1.08	-4.65	-32.76	1.58	-25.14/-12.78	-31.57	1.53	-22.18/-9.58
M&D	-15.25	0.76	-28.85	-25.98	1.33	-41.08	-37.79	2.08	-28.25	-39.44	2.08	-33.11	-54.15	3.19	-43.52/-33.89	-51.27	3.05	-41.77/-31.24
Silent	-8.61	0.38	-14.92	-15.82	0.77	-20.68	-30.28	1.51	-16.38	-35.54	1.83	-18.82	-45.5	2.5	-32.98/-21.27	-43.35	2.36	-31.50/-19.51
Average	-8.75	0.45	-15.14	-16.38	0.9	-22.57	-22.26	1.28	-12.58	-30.09	1.65	-20.94	-40.78	2.47	-34.48/-23.23	-38.87	2.35	-33.23/-21.38
Fade-out to white																		
Akiyo	-2.22	0.11	-24.34	-24.25	1.63	-35.06	-58.73	4.18	-41.15	-73.34	6.21	-55.21	-75.59	6.3	-50.66/-43.43	-73.47	6.08	-50.32/-42.35
Football	0.87	-0.05	-14.72	-1.59	0.09	-19.94	19.35	-1.05	35.08	-10.15	0.61	-15.48	-10.24	0.63	-25.07/-10.83	-10.98	0.66	-24.57/-10.18
Foreman	2.76	-0.1	1.17	-1.85	0.12	-2.62	-25.44	1.22	4.77	-50	2.85	-20.9	-49.92	2.69	-29.77/-19.05	-49.08	2.77	-29.37/-18.13
M&D	-10.72	0.53	-33.72	-34.63	2.33	-44.5	-57.66	3.98	-39.5	-71.79	5.66	-53.83	-73.53	5.72	-51.04/-44.54	-70.26	5.55	-51.49/-44.27
Silent	-9.07	0.46	-17.18	-26.28	1.59	-24.73	-58.08	4.2	-25.69	-69	5.6	-39.63	-68.45	5.64	-42.40/-33.85	-66.9	5.42	-42.62/-33.94
Average	-3.68	0.19	-17.76	-17.72	1.15	-25.37	-36.11	2.51	-13.3	-54.86	4.19	-37.01	-55.55	4.2	-39.79/-30.34	-54.14	4.1	-39.67/-29.78
Natural brightness variations																		
Flamenco2	-5.33	0.26	-7.62	-1.57	0.08	-6.12	-1.11	0.06	2.58	-4.16	0.2	-7.76	-10.64	0.51	-36.88/-28.35	-8.41	0.4	-22.29/-12.10
Mobisode2	-5.41	0.17	-11.14	-6.21	0.2	4.34	-7.9	0.26	-0.59	-4.76	0.15	-10.78	-26.32	0.88	-15.01/2.21	-25.93	0.87	-14.88/2.76
ShuttleStart	-28.48	0.77	-18.36	-21.91	0.56	-18.42	-28.4	0.77	-13.67	-31.14	0.83	-20.34	-41.32	1.07	-27.55/-16.42	-39.3	1.04	-27.56/-16.39
Average	-17.27	0.58	-15.32	-19.75	0.69	-16.37	-21.96	0.76	-14.05	-22.86	0.8	-21.47	-33.52	1.17	-30.64/-18.41	-32.07	1.11	-26.97/-14.20
Other																		
Foreman (Panning)	14.86	-0.67	16.31	9.96	-0.46	12.29	63.11	-2.46	46.06	12.25	-0.56	9.01	-4.17	0.2	-14.75/-0.02	-4.13	0.2	-14.72/0.91

It is interesting to note that SRefMWP and SRefMWP+LUTs have additional advantage of coding a scene with fast camera panning motion. To demonstrate this, a fast camera panning motion without brightness variation in “Foreman” (from frame 170 to frame 229) was encoded by different schemes. Figure 3.7 shows the RD performances of different schemes. MRefSWP-LS gets the worst performance due to its motion-sensitive characteristics. It has at most 2.7dB drop compared to ‘Without WP’ while other single model schemes such as MRefSWP-DC, MRefSWP-OFF, and MRefSWP-LMS also perform unsatisfactory with PSNR drop of about 0.9dB. It is due to the fact that large luminance difference induced by object motions may mislead the encoder in its use of WP. In this situation, irrelevant WP parameter sets may then be computed, which are not based on brightness variations from fading. Using wrong weighted parameters in motion estimation is likely to get a larger RD cost comparing with that of ‘Without WP’. It results in lower coding efficiency. On the other hand, our proposed SRefMWP and SRefMWP+LUTs can prevent this situation and obtain about 0.2dB PSNR gain in comparison with that of ‘Without-WP’, as depicted in Figure 3.7. In Figure 3.8, the percentage selection of different WP models of 8×8 blocks for this video segment of “Foreman” accounts for the results in Figure 3.7. It can be easily seen from Figure 3.8 that the original reference frame without WP is dominant. In fact, the original reference frame is essential for object movement or camera motion scenes whereas the weighted reference frames are good for brightness variation scenes. The proposed buffer management provides the mechanism to avoid the misuse of weighted prediction in the scene without brightness variation.

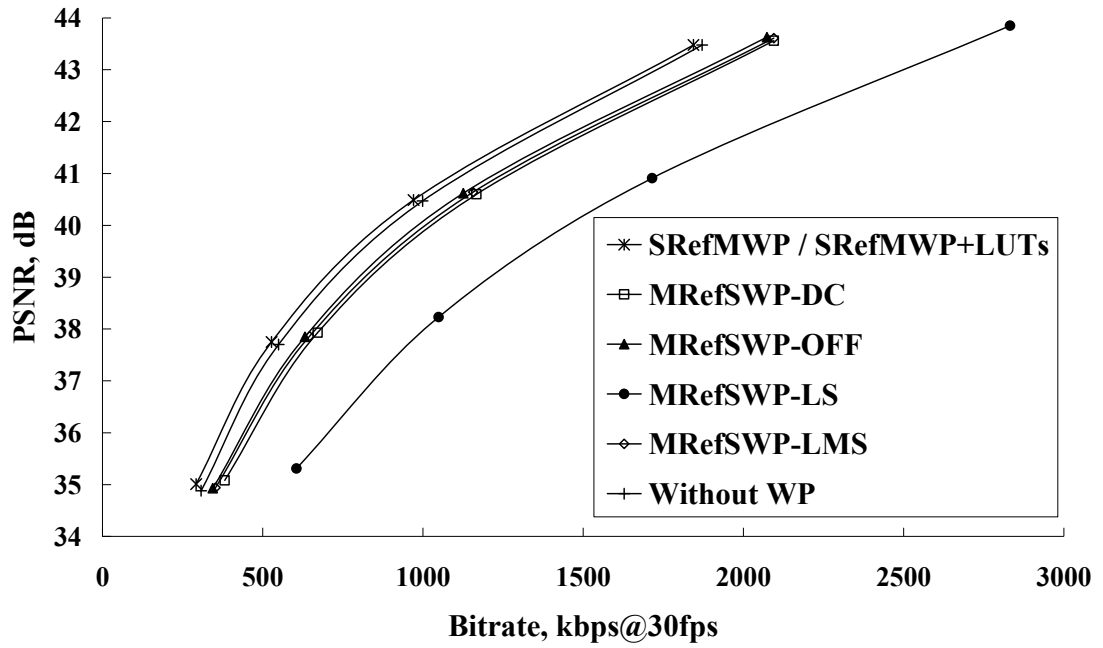


Figure 3.7. RD performances of different schemes for “Foreman” from frame 170 to frame 229.

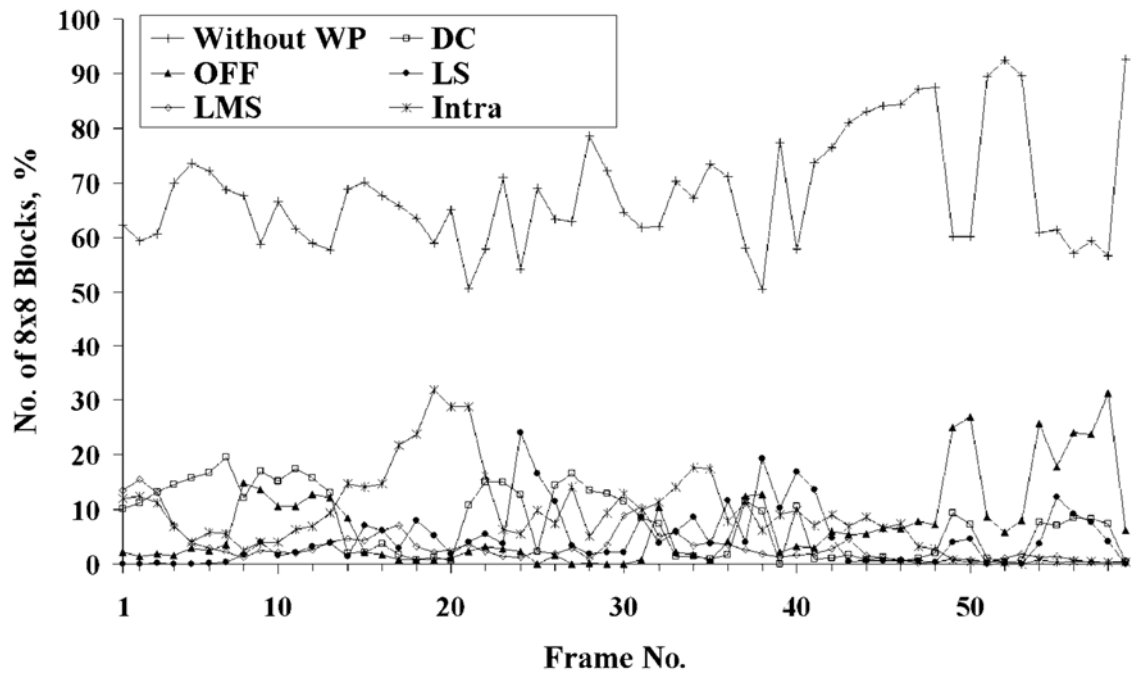


Figure 3.8. Statistics of selected WP models in percentage of 8x8 blocks for “Foreman” from frame 170 to frame 229 at QP 20 using SRefMWP/SRefMWP+LUTs.

3.5.2 Comparison of Encoding Complexity

In JM15.1 [16], a partial distortion search strategy is used in nearly all motion estimation options. The partial distortion search provides the optimal result equal to full search with reduced complexity [74-76]. It rejects impossible candidate motion vectors by means of a half-way stopping technique with partial distortion comparison to the current minimum distortion in a pixel-wise basis. If the current minimum distortion is computed sooner, the impossible candidates will be eliminated faster, which results in decreasing encoding time. In fact, the re-ordering mechanism used in the proposed SRefMWP is also beneficial from partial distortion search since the most likely reference frames are put first in the list. Consequently, the minimum distortion may be computed sooner. To show the complexity of the proposed schemes, the encoding time reduction of various schemes in comparison with ‘Without WP’ for all test sequences are measured and tabulated in Table 3.2. The experiments were performed on an Intel Xeon X5550 2.67GHz computer with 12GB memory. It is observed that the savings of encoding time of SRefMWP are mostly larger than those of other schemes. For the SRefMWP+LUTs, except “Mobisode2”, the savings still exist although the look-up process causes an increase in computational complexity. For “Mobisode2”, the encoding time of SRefMWP+LUTs is increased by 2.21% compared to ‘Without-WP’. This is due to the property of partial distortion search. It is known that the partial distortion search cannot work well when the minimum distortion gets a smaller value later [74-76]. This is a scenario with a lower temporal correlation between frames. From

Table 3.3, this happens in “Mobisode2” where 29.76% of 8x8 blocks being encoded as intra-modes by using SRefMWP and SRefMWP+LUTs. The percentage is relatively large as compared with other sequences, about 4.77% to 21.76% only. It puts a limit on the encoding time reduction of SRefMWP for “Mobisode2”, only 15.01% reduction in Table 3.2, but it is still the best among all the tested schemes. However, the encoding time reduction from SRefMWP cannot compensate for the additional complexity due to the look-up process of SRefMWP+LUTs, as shown in Table 3.2.

Note that all the single model schemes need longer encoding time for coding “Foreman” from frame 170 to frame 229 in which the scene contains fast camera panning motion. As the partial distortion search strategy is used in the full search inside JM15.1, longer encoding time is expected when irrelevant weighted parameters in the single model schemes are likely to get the minimum distortion later. The proposed SRefMWP and SRefMWP+LUTs can provide remarkable improvement, as shown in Table 3.2. According to these results, the proposed schemes obtain the best RD performance among all schemes with reduced computational complexity.

Table 3.3. Average percentages of 8x8 blocks using intra prediction (%) for various schemes.

Video Sequences	Without WP	MRefSWP-DC	MRefSWP-OFF	MRefSWP-LS	MRefSWP-LMS	SRefMWP / SRefMWP+LUTs	SRefMWP / SRefMWP+LUTs (without re-ordering)
Fade-in from black							
Akiyo	36.83	16.68	24.74	11.04	16.32	4.77	4.79
Football	38.18	26.03	25.66	47.31	29.83	17.31	17.21
Foreman	40.9	13.13	23.14	14.46	14.88	6.91	7.26
M&D	48.34	15.81	18.6	15.86	20.13	6.19	5.87
Silent	31.57	12.64	18.83	10.89	13.6	5.7	5.69
Average	39.16	16.86	22.19	19.91	18.95	8.18	8.16
Fade-out to black							
Akiyo	39.32	15.58	25.93	8.85	7.27	6.3	6.15
Football	41.14	24.94	31.2	41.89	24.1	17.28	17.21
Foreman	43.2	11.22	27.74	10.62	5.6	5.15	5.82
M&D	50.08	13.68	32.55	12.03	7.94	6.59	6.24
Silent	32.63	11.7	20.96	8.55	7.25	6.62	6.92
Average	41.27	15.42	27.67	16.39	10.43	8.39	8.47
Fade-in from white							
Akiyo	59.1	37.33	38.27	23.53	17.8	12.86	12.99
Football	47.04	37.76	32.76	52.46	32.7	21.76	21.57
Foreman	23.54	19.63	19.31	22.07	14.93	8.48	8.27
M&D	54.71	39.29	29.13	28.29	23.03	15.57	15.47
Silent	30.92	21.32	22.04	18.57	15.57	9.96	10.5
Average	43.06	31.06	28.3	28.99	20.8	13.72	13.76
Fade-out to white							
Akiyo	58.11	39.16	32.73	19.42	8.31	7.12	7.96
Football	45.51	36.31	27.3	51.02	24.78	19.54	19.73
Foreman	23.51	21.03	23.29	16.57	6.08	5.38	5.23
M&D	54.47	41.84	22.7	23.6	8.79	7.09	7.45
Silent	30.71	22.91	19.01	14.86	7.67	6.93	7.21
Average	42.46	32.25	25.01	25.1	11.13	9.21	9.52
Natural brightness variations							
Flamenco2	13.79	12.2	12.94	16.9	12.68	8.76	9.7
Mobisode2	51.91	52.92	55.15	52.88	54.92	29.76	29.82
ShuttleStart	30.05	13.57	15.43	16.58	10.61	10.14	10.12
Average	31.92	26.23	27.84	28.78	26.07	16.22	16.54
Other							
Forman (Panning)	15.09	28.01	26.81	60.8	23.83	12.47	12.56

3.6 Chapter Summary

In this work, we have proposed an enhanced weighted prediction scheme that utilizes the concept of MRF-ME and a new reference re-ordering mechanism to improve motion-compensation performances for video sequences with various types of brightness variations. Based on the novel arrangement of multiple frame buffers, the proposed scheme facilitates the use of multiple WP models in a single reference frame. It is concluded from the experimental results that the proposed scheme can outperform any conventional WP models in sequences with different types of fading effects and even in scenes with natural brightness variations. In addition, we have also found that our proposed scheme can work with the LUT technique to reduce the memory requirement. Experimental results show that the additional memory is effectively removed by 80%.

Chapter 4 Flash Scene Video Coding using Weighted Prediction

4.1 Introduction

A flash being fired during a press conference, a sport match, a news interview, etc., can cause the non-uniform intensity change distributed over the entire picture. Weighted prediction in H.264/AVC is a frame-based approach which can only code the scenes with global brightness variations (GBV) such as fade-in and fade-out efficiently but not the scenes with local brightness variations (LBV) including flashlight (FL) scenes. It is therefore very difficult to find an accurate frame-based model to estimate the change of the intensity within the picture. Macroblock (MB)-based approaches [35,64] in which different MBs in the same frame can use different W and O were proposed to solve the problem of LBVs. Unfortunately, this may lead to increased computational complexity, considering that it would be necessary to perform ME using all possible sets of W and O [35]. In [41], a human vision system based scheme was designed to solve the problems of coding FL scenes by interpolating and inpainting non-FL frames as FL frames. Nevertheless, the objective quality of FL frames is dropped a lot. In this chapter, a novel scheme for coding flash scenes is proposed. In principle, flash scenes can be detected by analyzing the histogram differences between frames. The proposed scheme then suggests an adaptive coding order technique for increasing coding efficiency by taking account of characteristics of flash scenes in video contents. The use of the adaptive coding technique also benefits to enhance the accuracy of derived motion vectors for determination of

weighting parameter sets. Experimental results show that a significant improvement of coding performance in terms of bitrate and PSNR can be achieved in comparison with the conventional weighted prediction algorithms.

Some results in this chapter have been reported in the reference [67].

4.2 Proposed Coding Scheme for Flash Scenes

The salient characteristic of flashlight effect is the abrupt luminance change across frames of the same scene within a very short period of time, which is caused by sudden appearance of the illumination source. Normally, we assume that a flash scene cannot last more than 0.15-0.2 second [69,71]. In other words, the number of FL frames should be smaller than 5 if the video frame rate is 25 fps. Notice that the FL frames have a much stronger intensity than the previous and the later non-FL frames in the video sequence. When flashlight occurs, it lowers the correlation between a non-FL frame and a FL frame due to the abrupt change of brightness. Figure 4.1(a) illustrates three consecutive frames of flashlight effect lasting only one frame. The blocks in the FL frame, f_{t+1} , cannot locate well-matching blocks in its previous non-FL frame, f_t , as there is the extremely great intensity difference between these two frames. Consequently, a large number of blocks is coded as intra modes instead of inter modes since the intra coding can achieve better rate-distortion performance. In general, the FL frame, f_{t+1} , needs more bits than its previous non-FL frame, f_t . Similarly, the next non-FL frame, f_{t+2} , in Figure 4.1(a) encounters the same problem that the blocks in f_{t+2} cannot find well-matching blocks in f_{t+1} owing to their great difference in pixel intensity. It is observed that the bit rate burst induced is mainly due to the coding order of the prediction structure. The fixed coding order ignores the nature of scenes with flashlight. In this work, we take into consideration the nature of

video content with respect to the existence of flashlight. Therefore, the proposed algorithm changes the coding order dynamically according to flashlight in input scenes in order to improve coding efficiency. Figure 4.1(b) then shows the new coding order that is made adaptive based on the existence of flashlight. In this example, the non-FL frame f_{t+2} can be predicted from the other non-FL frame f_t and hence f_{t+2} will not produce large amount of bits. Besides, the FL frame is encoded by using the proposed MB-based WP.

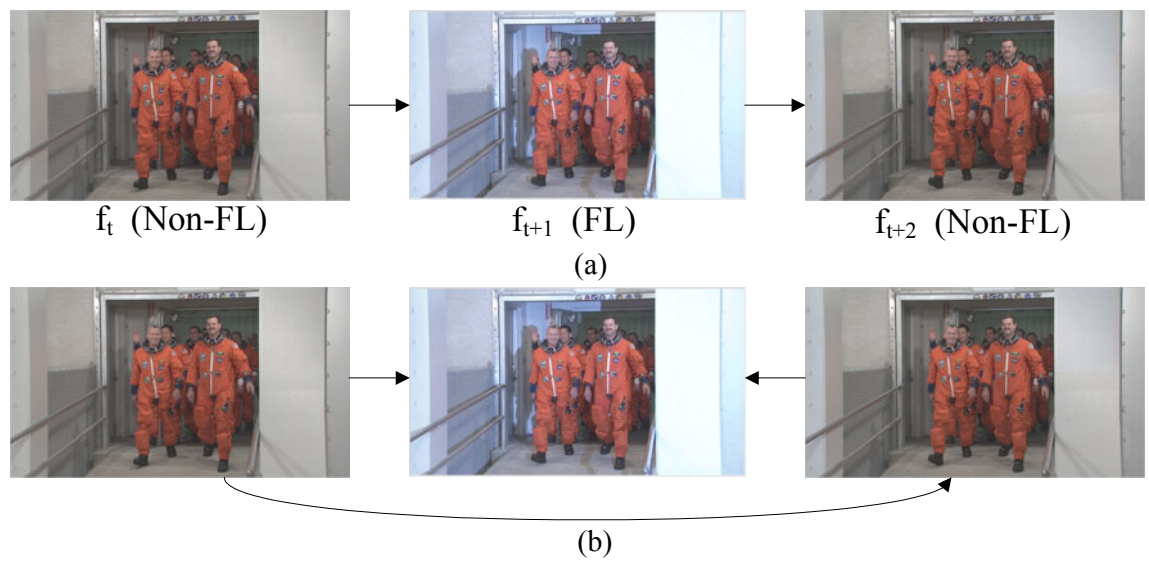


Figure 4.1. Adaptive coding order for a FL scene.

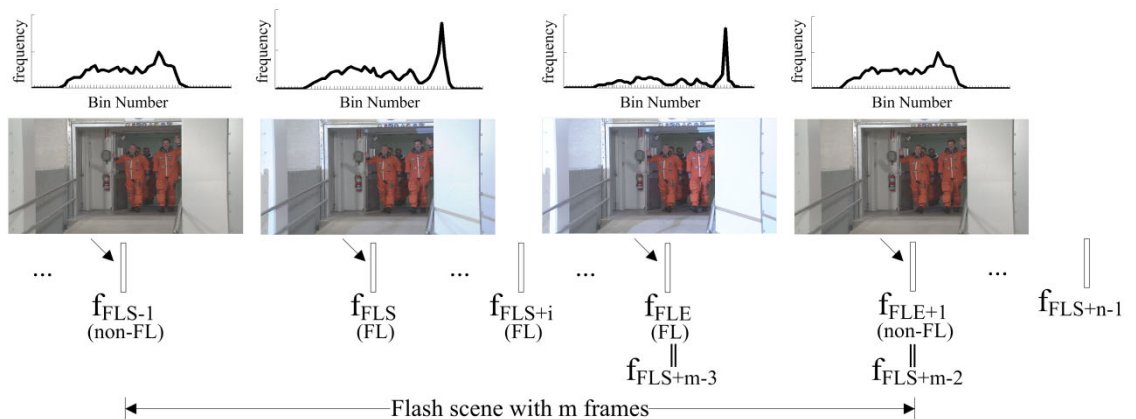


Figure 4.2. Histogram and proposed prediction structure of a FL scene.

4.2.1 Adaptive Coding Order based on Flash

To do so, detection of flash scenes is necessary such that the coding order in the prediction structure can be changed more appropriately to achieve coding gain in video sequences with flashlight effect. Flashlight detection methods [69,71] have been widely studied for automatic video indexing, browsing, and retrieval. A FL-frame can be located by computing its average intensity [69] or histogram [71] and comparing it with the values of neighboring pictures [69,71]. All of these methods are tailor-made for video indexing applications. In this chapter, we modify the histogram difference (HD)-based method in [71] such that it is more applicable to video coding. For coding a flashlight scene with m frames, two non-FL frames and $m-2$ FL frames should be taken into consideration ($m \geq 3$), as depicted in Figure 4.2. Let f_{FLS} and f_{FLE} represent the starting and ending frames of a flashlight scene, respectively. From Figure 4.2, the relationship between f_{FLS} and f_{FLE} can be written as

$$\overline{f_{FLE}} \approx \overline{f_{FLS}} \quad (4.1)$$

where \overline{f} denotes the average luminance value of all pixels as in equation (2.12).

As shown in Figure 4.2, a flashlight starts with a large increase of luminance level from f_{FLS-1} to f_{FLS} , i.e. $\overline{f_{FLE-1}} \ll \overline{f_{FLS}}$. It is then followed by a period of constant high luminance level from f_{FLS} to $f_{FLS+m-3}$, and ends with a large decrease in luminance level from $f_{FLS+m-3}$ to $f_{FLS+m-2}$, i.e. $\overline{f_{FLS+m-3}} \ll \overline{f_{FLS+m-2}}$. Accordingly, the problem of flash scene detection can then be converted to identify f_{FLS-1} and f_{FLE+1} ($=f_{FLS+m-2}$ in Figure 4.2) according to the histogram difference (HD) among frames. HD between the frames f_a and f_b can be written as

$$HD(f_a, f_b) = \sum_{x=0}^{63} |Hist(f_a)_x - Hist(f_b)_x| \quad (4.2)$$

where $Hist(f_a)_x$ and $Hist(f_b)_x$ denote the x -th bin of the normalized luminance histograms of f_a and f_b respectively. Each histogram is quantized into 64 bins for noise suppression and fast calculation. Figure 4.2 also shows several frames of flashlight effect with the corresponding histograms. Luminance values at each location increase due to the brightness effect of flashlight (i.e. f_{FLS} to $f_{FLS+m-3}$) when flashlight appears. Hence, the histogram of f_{FLS} , $Hist(f_{FLS})$, shifts from left to right compared to the histogram of f_{FLS-1} , $Hist(f_{FLS-1})$. In other words, the HD between f_{FLS-1} and the other frames inside the flash scene consistently become larger. The luminance level of $f_{FLS+m-2}$, $\overline{f_{FLS+m-2}}$, returns to the level which is similar to the luminance level of f_{FLS-1} , $\overline{f_{FLS-1}}$, after flashlight effect, i.e. $\overline{f_{FLS+m-2}} \approx \overline{f_{FLS-1}}$. As a consequence, the HD between f_{FLS-1} and $f_{FLS+m-2}$, $HD(f_{FLS-1}, f_{FLS+m-2})$, is very small, and then the following relationship can be derived:

$$HD(f_{FLS-1}, f_{FLS+i-3}) > k \cdot HD(f_{FLS-1}, f_{FLS+m-2}) \quad \text{where } 3 \leq i \leq m \quad (4.3)$$

k is a positive quantity to control the sensitivity of the flash scene detector.

To detect flash scenes, we adopt a sliding window of n frames whose first frame is the current frame being encoded, f_c . The HDs between f_c and all the other frames in the sliding windows, $HD(f_c, f_{c+j})$ where $1 \leq j < n$, are computed. f_c is expected to be f_{FLS-1} if there exist $f_{FLS+m-2}$ ($m < n$) that satisfies the relationship in (4.3). Based on the results from [69,71], flashlight events do not, generally, last longer than 5 frames. As a result, the size of the sliding window, n , is set to 6. Since f_{FLS-1} and $f_{FLS+m-2}$ (or f_{FLE+1}) are both non-FL frames, the correlation is high and it is efficient to encode $f_{FLS+m-2}$ as P-frame by using f_{FLS-1} as the reference. Moreover, for those FL frames between f_{FLS-1} and $f_{FLS+m-2}$, they will be encoded as B-frames with the help of weighted prediction as discussed in the

next section. It is noted that the insertion of B-frames will introduce decoding delay which depends on the duration of the flash scene.

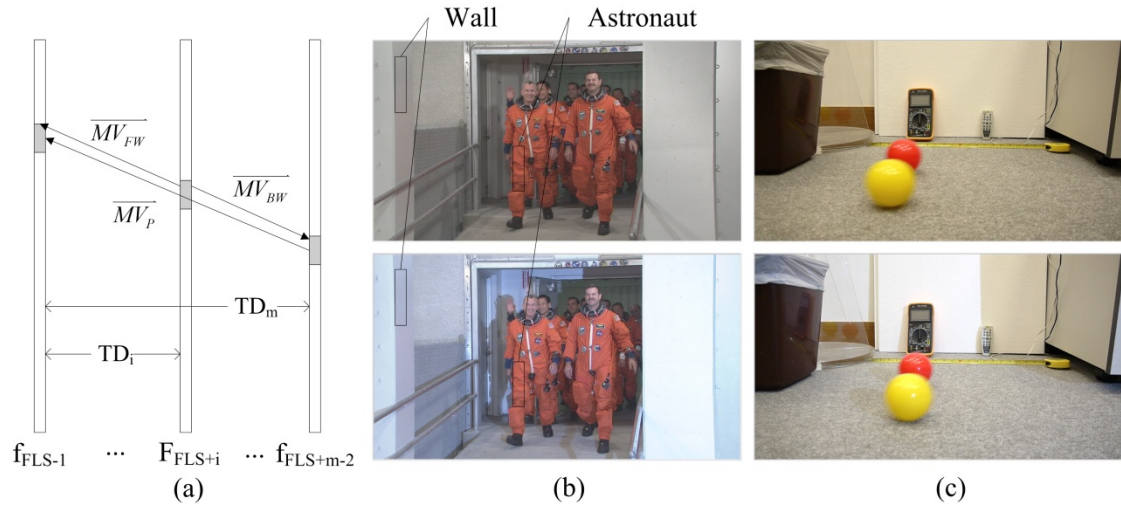


Figure 4.3. (a) Derivation of MVs for FL frames, (b) Crew, (c) BallSeq.

4.2.2 MB-based WP with Derived Motion Vectors

After coding of $f_{FLS+m-2}$, all FL frames within the detected flash scene (f_{FLS+i} where $0 < i < m-2$) are coded using MB-based WP rather than using the conventional frame-based WP in H.264/AVC. A typical characteristic of FL frames is that the intensity change is non-uniformly distributed over the entire picture. In other words, some regions of the picture may have a greater intensity increase than other regions. However, the frame-based WP only estimates one set of W and O for each reference frame and is only able to handle the scenes with global brightness variations. But, for the FL frame shown in Figure 4.3, W and O in the wall region are from 40 to 46 and from -6 to -24 respectively whereas W and O in the astronaut region are from -2 to 24 and from 39 to 105 respectively. Since the MB-based WP technique allows each MB has its own W and O , it is more suitable for coding FL frames in order to handle different amounts of brightness changes in different

regions. Nevertheless, in the MB-based WP, the WP parameter set for each MB cannot be estimated accurately without its true motion vector (MV), and the true motion vectors cannot be obtained without the accurate WP parameter set. Although this chicken-egg dilemma can be resolved through an exhaustive evaluation of all possible weighting parameter sets for each searching point during ME by minimizing the SAD function in (2.6), this brings out the problem of impractical high computational complexity. Therefore we suggest deriving the initial forward and backward MVs (\overline{MV}_{FW} and \overline{MV}_{BW} , respectively) of the FL frames for computing W and O , as depicted in Figure 4.3. These initial MVs are based on the forward MVs of $f_{FLS+m-2}$ pointing to f_{FLS-1} . Note that there may be more than one MVs tracking through the current MB of the FL frame. In order to choose the most representative one for deriving MVs for estimation of weighting parameter sets, the forward MV of $f_{FLS+m-2}$, \overline{MV}_P , tracks through the current encoded MB of the FL frames with the largest overlapping area is then used to derive \overline{MV}_{FW} and \overline{MV}_{BW} , which can be written as

$$\begin{aligned}\overline{MV}_{FW} &= \frac{TD_i}{TD_m} \overline{MV}_P, \\ \overline{MV}_{BW} &= \frac{TD_i - TD_m}{TD_m} \overline{MV}_P\end{aligned}\tag{4.4}$$

where TD_i is the temporal distance between the current B frame and f_{FLS-1} , and TD_m is the temporal distance between the two non-FL frames, f_{FLS-1} and $f_{FLS+m-2}$. The areas pointed by these derived \overline{MV}_{FW} and \overline{MV}_{BW} are used for determining W and O accurately such that object motions are taken into consideration without introducing huge complexity. It is noted that the reliability of \overline{MV}_P is vital for the proposed scheme since both \overline{MV}_{FW} and \overline{MV}_{BW} affect the accuracy of the weighting parameter sets in FL frames. With the help of the mechanism of adaptive coding order, the correlation between f_{FLS-1} and $f_{FLS+m-2}$ is high

and \overline{MV}_p of $f_{FLS+m-2}$ is more reliable. That is to say, \overline{MV}_p with proper scaling can derive \overline{MV}_{FW} and \overline{MV}_{BW} for each MB of the FL frames within the flash scene. After the determination of W and O , motion estimation using SAD in (2.6) is performed to obtain the final \overline{MV}_{FW} and \overline{MV}_{BW} . Similar to other MB-based WP algorithms [35], only one bit/MB of additional information is needed for the indication of utilizing WP, and predictive coding is used for W and O due to the high correlation of weighting parameter sets between the current MB and neighboring MBs. Thus the median of the weighting parameter set are computed from the neighboring MBs and subtracted from the weighting parameter set of the current MB. The difference is then encoded using zero-order Exp-Golomb codes defined in H.264/AVC [9,16] to achieve higher coding efficiency.

4.3 Experimental Results

Experiments have been conducted over two 720p video sequences with flash scenes to evaluate the overall efficiency of various weighted prediction algorithms. A standard sequence "Crew" contains NASA crews leaving a building with flashlight while "Ballseq" is a self-recorded flashlight sequence of two balls rolling from left to right, as shown in Figure 4.3. The test sequences were all encoded at 25 frames/s. Besides, three standard sequences "Shields", "Exit" and "Sunflower", with synthetic flashlight applied at the frame centre, have also been conducted to show the impact of motion activities, flash durations, and flashlight intensity on different algorithms. "Shields" is a sequence with a person walking slowly from left to right. "Exit" is a sequence with people walking through the door in the office while "Sunflower" is a sequence with a bee collecting nectar on a flower which contains fast local and global motion. To simulate flashlight effect,

synthetic flashlight is applied between f_{FLS} and f_{FLE} based on [71], and a general mathematical model for the synthetic flashlight can be expressed as follows:

$$\bar{f} = \begin{cases} \bar{f}, & f \leq f_{FLS-1} \\ \bar{f} + \alpha, & f_{FLS} \leq f \leq f_{FLE} = f_{FLS+m-3} \\ \bar{f}, & f_{FLS+m-2} = f_{FLE+1} \leq f \end{cases} \quad (4.5)$$

where α is the constant controlling the synthetic flashlight energy. All experiments were conducted using Main profile, two reference pictures, quarter-pel full search motion estimation with search range of ± 32 pixels, RDO with all seven inter modes as well as intra modes, and context-adaptive binary arithmetic coding (CABAC). The interval between two I-frames was set to 60. The encoded bitstreams were encoded by different algorithms with a set of four different QPs (i.e. QP=20, 24, 28, and 32). It is noted that other settings such as fast motion estimation technique or RDO is off can also be applied for evaluating the performance.

We incorporated the proposed adaptive coding order (ACO) into both of the frame-based and MB-based WP algorithms while the technique of using derived motion vectors (DMV) is only applicable to the MB-based algorithms. For the flashlight detection process, as aforementioned, n was set to 6. For the value of k , it is easy to have false detections if k is too small. If k is too large, flash scenes with only slight luminance changes cannot be detected. By considering this tradeoff, k was experimentally set to 25. Two most popular WP models - LS and LMS were adopted in all WP algorithms. It is noted that LS depends on the mean of the product of the current frame and the pixels in the same position in the reference frame. It implies that it only works well in the scene with low motion activity as compared with LMS [14,15]. The use of the two models in this chapter is intended to show the flexibility of the proposed ACO and DMV in which they can be applied to

different WP models. The proposed MB-based algorithms using different WP models - MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV were implemented based on the JM15.1 encoder [16], which are used to compare the performance of the conventional frame-based WP algorithms by employing different models, named as WP(LS) and WP(LMS). In order to demonstrate the contributions from the proposed two techniques (ACO and DMV), different combinations have been also included in the experiments. Table 4.1 summaries the tools used in various algorithms.

Table 4.1. Summary of tools used in various WP algorithms.

		Algorithm	LS model	LMS model	ACO	DMV
Frame-based WP algorithm	WP(LS)		✓			
	WP(LMS)			✓		
	WP(LS)+ACO		✓		✓	
	WP(LMS)+ACO			✓	✓	
MB-based WP algorithm	MBWP(LS)+ACO		✓		✓	
	MBWP(LMS)+ACO			✓	✓	
	MBWP(LS)+DMV+ACO		✓		✓	✓
	MBWP(LMS)+DMV+ACO			✓	✓	✓

4.3.1 Sequences with real flash scenes

The detailed comparisons of sequences "Crew" and "Ballseq" with real flash scenes for different algorithms using the Bjontegaard delta bitrate (BDBR) and Bjontegaard delta PSNR (BDPSNR) [73] compared to H.264/AVC coding without WP, denoted by 'Without WP', for all frames and only for FL frames are tabulated in

Table 4.2. This table shows that WP(LS) and WP(LMS) cannot achieve noticeable improvement over NoWP, and it is even worse than ‘Without WP’ in "Crew". It is due to the low correlation between a non-FL frame and a FL frame when a fixed coding order is used to code sequences with flashlight.

Table 4.2 also involves the results of the frame-based algorithms using ACO, denoted by WP(LS)+ACO and WP(LMS)+ACO, and it shows the merit of our proposed ACO in coding flash scenes. Despite the time delay introduced by the backward prediction in ACO, in which a very small time delay might be brought to real-time video conferencing applications, RD performance has been improved. However, it is clear from

Table 4.2 that only a slight improvement has been demonstrated by using WP(LS)+ACO and WP(LMS)+ACO. It is not unexpected since the intensity change is non-uniformly distributed over the entire picture. This can be solved by the proposed MB-based algorithms, MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV. With the help of MB-based WP parameter sets, MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV work better than WP(LS)+ACO and WP(LMS)+ACO due to the reason that W and O are estimated in MB basis which are more favor to local brightness variations. It is noted that the gains of MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV for all frames are small as the number of FL frames is very small compared with the number of non-FL frames for the whole sequence. In order to evaluate the merit of deriving MVs for estimation of WP parameter sets, MBWP(LS)+ACO and MBWP(LMS)+ACO, in which MB-based WP parameter sets W and O are estimated using the current MB and its co-located MB, are also included in

Table 4.2. For the sequence "Crew", the performance of MBWP(LS)+ACO is even worse than that of WP(LS)+ACO. This is because LS model is a WP model which is sensitive to motion activity which would induce inaccurate W and O easily [15]. Without the proposed DMV, MBWP(LS)+ACO cannot ensure the coding performance due to the nature of LS. Our proposed MBWP(LS)+ACO+DMV can solve this problem since object motions or camera movements can be compensated by deriving MVs before estimation of WP parameter sets which makes W and O more accurate. It concludes that ACO should work with DMV to secure better coding efficiency. From

Table 4.2, it can also be found that the gain of "Crew" is smaller than that of "BallSeq". It is because "BallSeq" has simpler motion than "Crew" and the flash intensity in "BallSeq" is much stronger than that in "Crew". In the next section, we further show that the proposed techniques are more appropriate for sequences with low motion activity and strong flashlight. Table 4.3 then shows the average distributions of intra, skips/direct, and inter modes in which only FL frames of each sequence are included. It can be seen that MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV successfully increase the number of skips/direct and inter modes since inter-frame correlations increase due to more accurate WP parameter sets, which make better predicted frames as references for coding FL frames. It is well known that the increase in skips/direct and inter modes benefits the coding gain, but it only roughly shows the tendency. For instance, from Table 4.3, it shows that MBWP(LS)+ACO+DMV has small percentage of intra MBs and large percentage of skip/direct modes as compared with MBWP(LMS)+ACO+DMV, but its coding gain is smaller as shown in

Table 4.2. It is because the use of LMS can achieve better coding gain when brightness variation exists, as compared with LS. Though MBWP(LMS)+ACO+DMV only has small percentage of skip/direct modes, it still outperforms MBWP(LS)+ACO+DMV.

Table 4.2. BDBR (%) and BDPSNR (dB) compared to “NoWP” for sequences with real flash scenes.

Algorithm	All frames				FL frames			
	Crew		BallSeq		Crew		BallSeq	
	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR
LS model								
WP(LS)	1.794	-0.05	-3.595	0.124	3.356	-0.133	-9.937	0.571
WP(LS) +ACO	-0.723	0.021	-7.76	0.285	1.746	-0.07	-17.432	1.065
MBWP(LS) +ACO	-0.376	0.004	-21.171	0.768	-3.041	0.112	-46.856	3.012
MBWP(LS) +ACO+DMV	-0.928	0.031	-23.489	0.88	-6.517	0.28	-52.664	3.427
LMS model								
WP(LMS)	1.839	-0.051	-3.442	0.12	3.442	-0.137	-9.424	0.537
WP(LMS) +ACO	-0.704	0.021	-6.821	0.249	2.142	-0.085	-15.393	0.93
MBWP(LMS) +ACO	-0.889	0.029	-20.028	0.743	-6.811	0.283	-45.005	2.977
MBWP(LMS) +ACO+DMV	-1.276	0.037	-21.297	0.774	-8.584	0.336	-47.173	3.104

Table 4.3. Average percentage (%) of intra, skip/direct and inter modes for FL frames only.

Algorithm	Crew			BallSeq		
	Intras	Skips/Directs	Inters	Intras	Skips/Directs	Inters
NoWP	66.38	4.94	28.68	76.19	10.31	13.51
WP(LS)	71.46	1.19	27.34	71.13	0.22	28.65
WP(LMS)	70.82	1.14	28.05	71.49	0.31	28.2
WP(LS)+ACO	69.57	1.13	29.3	53.78	21.16	25.06
WP(LMS)+ACO	69.42	1.09	29.48	54.05	13.71	32.24
MBWP(LS)+ACO	48.11	31.4	20.49	30.18	60.39	9.43
MBWP(LMS)+ACO	56.51	20.71	22.78	35.77	57.4	6.83
MBWP(LS)+ACO+DMV	43.9	38.31	17.8	20.85	69.22	9.93
MBWP(LMS)+ACO+DMV	53.39	24.87	21.74	32.65	57.49	9.86

4.3.2 Sequences with synthetic flash scenes of different motion activities, flash durations, and intensity

In order to evaluate the performances of the proposed techniques under different test conditions, various synthetic flashlight effects were applied to "Shields", "Exit" and "Sunflower" which are considered as low, medium, and high motion sequences, respectively. Table 4.4 shows the BDBR and BDPSNR results of all tested algorithms for the above sequences with different motion activities in which m and α of the synthetic flashlight in (4.5) were fixed at 4 and 50, respectively. It implies that the flashlight strength keeps constant and the duration of each synthetic flash scene is two-frame long that repeats every four frames. As expected, the frame-based approaches, WP(LS), WP(LS)+ACO, WP(LMS) and WP(LMS)+ACO, cannot improve the coding efficiency. For the MB-based WP approaches, they achieve remarkable coding gains for all sequences, as shown in Table 4.4. From this table, we found that the coding gain is smaller for sequences with high motion activity. It can also be seen from this table that MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV provide significant improvement over MBWP(LS)+ACO and MBWP(LMS)+ACO, especially in "Exit" and "Sunflower" which contain fast motion activity. In this situation, DMV is necessary to derive the initial forward and backward MVs for the estimation of accurate weighting parameter sets.

We also demonstrate the effects of different flash durations ($m=3, 4, \text{ and } 5$ in (4.5)). The BDBR and BDPSNR results of all tested algorithms for "Exit" are shown in Table 4.5. Again, α was fixed at 50. Similarly, WP(LS), WP(LS)+ACO, WP(LMS) and WP(LMS)+ACO obtain lower coding efficiency compared with NoWP as they are unable to handle local brightness variations. For the proposed MB-based WP approaches, coding

gains for different flash durations are encouraging. It is interesting to note that the improvement of coding gain of MBWP(LS)+ACO+DMV/MBWP(LMS)+ACO+DMV is smaller for a long flash duration. It can be explained by the nature of DMV. As mentioned in Section 4.2.2, the use of DMV assumes linear motion within a flash scene. If the flash duration become longer, this assumption is easy to be invalid. Though DMV is less effective with a long flashlight duration, a typical flash duration is very short which can always be handled by DMV.

As the flash intensity can be adjusted according to the user preference, we also evaluate the coding performances of all tested algorithms with different flash intensity. Synthetic flashlight was applied to "Exit" with $m=4$ and $\alpha=30, 50, \text{ and } 70$. With the same duration of flash scenes but varying flash intensity, Table 4.6 shows the BDBR and BDPSNR results of various algorithms. For WP(LS), WP(LS)+ACO, WP(LMS) and WP(LMS)+ACO, the coding performances get worse as flash intensity increases. It is due to the fact that an increase in flash intensity lowers the correlation between a non-FL frame and a FL frame. Again, larger coding gains can be achieved by the proposed MB-based WP algorithms. MBWP(LS)+ACO and MBWP(LMS)+ACO can handle flash scenes with different flash intensity well with BDBR improvement from 14.894% to 23.211%. With the use of DMV, MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV further improve the BDBR from 17.140% to 26.925%. Therefore the larger intensity differences between non-FL and FL frames, the larger the need of the proposed ACO and DMV.

Table 4.4. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different motion activities.

Algorithm	Shields ($m=4, \alpha=50$)		Exit ($m=4, \alpha=50$)		Sunflower ($m=4, \alpha=50$)	
	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR
LS Model						
WP(LS)	38.489	-1.19	39.685	-1.238	48.837	-1.588
WP(LS)+ACO	37.8	-1.186	36.458	-1.052	47.927	-1.561
MBWP(LS)+ACO	- 22.061	0.879	-19.133	0.761	-10.975	0.614
MBWP(LS)+ACO+DMV	- 26.121	0.965	-26.592	0.918	-17.036	0.895
LMS Model						
WP(LMS)	34.519	-1.179	38.588	-1.211	52.78	-1.659
WP(LMS)+ACO	33.786	-1.156	34.998	-0.989	51.44	-1.639
MBWP(LMS)+ACO	- 28.786	1.028	-18.427	0.734	-12.474	0.702
MBWP(LMS)+ACO+DMV	- 33.208	1.338	-25.169	0.886	-20.354	0.992

Table 4.5. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different flash durations.

Algorithm	Exit ($m=3, \alpha=50$)		Exit ($m=4, \alpha=50$)		Exit ($m=5, \alpha=50$)	
	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR
LS Model						
WP(LS)	32.011	-0.938	39.685	-1.238	55.841	-1.494
WP(LS)+ACO	26.993	-0.695	36.458	-1.052	50.342	-1.429
MBWP(LS)+ACO	- 20.853	0.813	-19.133	0.761	-13.153	0.604
MBWP(LS)+ACO+DMV	- 29.027	0.971	-26.592	0.918	-18.202	0.743
LMS Model						
WP(LMS)	33.363	-0.945	38.588	-1.211	54.964	-1.477
WP(LMS)+ACO	25.724	-0.664	34.998	-0.989	49.640	-1.401
MBWP(LMS)+ACO	- 21.299	0.816	-18.427	0.734	-12.800	0.598
MBWP(LMS)+ACO+DMV	- 29.813	0.980	-25.169	0.886	-16.849	0.701

Table 4.6. BDBR (%) and BDPSNR (dB) of FL frames compared to “NoWP” for synthetic flash scenes with different flash intensity.

Algorithm	Exit ($m=4, \alpha=30$)		Exit ($m=4, \alpha=50$)		Exit ($m=4, \alpha=70$)	
	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR
LS Model						
WP(LS)	31.793	-0.720	39.685	-1.238	50.846	-1.455
WP(LS)+ACO	27.122	-0.928	36.458	-1.052	42.001	-1.314
MBWP(LS)+ACO	- 15.685	0.565	-19.133	0.761	-22.378	0.766
MBWP(LS)+ACO+DMV	- 17.794	0.689	-26.592	0.918	-26.925	0.943
LMS Model						
WP(LMS)	32.376	-0.736	38.588	-1.211	49.974	-1.447
WP(LMS)+ACO	28.092	-0.955	34.998	-0.989	41.096	-1.312
MBWP(LMS)+ACO	- 14.894	0.533	-18.427	0.734	-23.211	0.807
MBWP(LMS)+ACO+DMV	- 17.140	0.608	-25.169	0.886	-26.123	0.913

4.3.3 Comparison of Encoding Time Complexity

To compare the computational complexity of the proposed techniques, the frame-based algorithm, WP(LS) or WP(LMS), is used as a reference method, and all simulations were performed on an Intel Xeon X5550 at 2.67GHz computer with 12GB memory. The percentage of total encoding time increased as compared with WP(LS)/WP(LMS) is tabulated in Table 4.7. From Table 4.7, it can be seen that the use of ACO increases the required time for motion estimation and compensation since some P-frames are changed to B-frames in flash scenes, and bi-directional motion estimation for both forward and backward references inevitably increases the encoding time. It is observed that an increase in complexity of MBWP(LS)+ACO and MBWP(LMS)+ACO as shown in Table 4.7. The additional complexity comes from the MB-based estimation of W and O . Due to the adoption of DMV in MBWP(LS)+ACO+DMV and MBWP(LMS)+ACO+DMV, a further complexity requirement for deriving the motion vectors to estimate the

weighting parameter sets is induced. According to these results, we conclude that the proposed techniques provide higher coding efficiency with increased computational complexity.

Table 4.7. Average percentage (%) of total encoding time increased compared with the conventional WP approaches.

Algorithm	Crew	BallSeq	Shields	Exit	Sunflower
			(m=4, $\alpha=50$)	(m=4, $\alpha=50$)	(m=4, $\alpha=50$)
LS Model					
WP(LS)+ACO	4.32	6.07	6.04	7.32	6.03
MBWP(LS)+ACO	9.11	8.93	18.17	19.36	14.72
MBWP(LS)+ACO+DMV	11.79	11.34	21.75	24.77	19.81
LMS Model					
WP(LMS)+ACO	6.86	5.23	4.24	5.68	7.71
MBWP(LMS)+ACO	12.09	8.49	17.21	20.46	13.93
MBWP(LMS)+ACO+DMV	13.37	11.44	20.61	25.28	21

4.4 Chapter Summary

In this chapter, we have proposed an adaptive coding order technique for video coding based on flash scene, which extracts FL and non-FL frames according to histogram differences, and assigns appropriate coding type to each frame correspondingly. Motion vector derivation is then adopted instead of using co-located block in the determination of WP parameter sets. Experimental results show that the proposed scheme with the adaptive coding order and motion vector derivation techniques achieves significant performance gain over the conventional WP schemes for coding flash scenes.

Chapter 5 Region-based Weighted Prediction for Coding Scenes with Local Brightness Variations

5.1 Introduction

In Chapter 3, we proposed an H.264/ AVC standard-compliant scheme for multiple WP models. To achieve this, the structure of multiple reference frames (MRF) in H.264/AVC is utilized to facilitate the use of multiple WP models. This scheme can compensate for scenes with different fading effects. Nevertheless, the use of multiple WP models cannot work well for scenes with LBV. In this chapter, we further extend our work in [54] and propose a novel region-based WP parameter estimation scheme for encoders of the H.264/AVC standard. It can handle complex LBV scenes with more than one region. Motivated by [50,54], the proposed scheme can embed multiple WP parameter sets into the MRF architecture of H.264/AVC. This arrangement ensures that the proposed scheme is in compliance with the H.264/AVC standard. In this chapter, we start by introducing an in-depth study of WP parameter sets in scenes with different types of brightness variations as the basis for our scheme. We then proceed to present our novel region-based WP scheme. Finally, experimental results using scenes with LBVs as well as GBVs are presented.

Details of the scheme are shown in the following sections, while results of this chapter have been published in the reference [56].

5.2 Analysis of WP Parameters in Scenes with LBV and GBV

To make analysis of how GBV and LBV affect the WP parameters, scenes with GBV and LBV in Figure 5.1 and Figure 5.2 are encoded by an MB-based WP scheme using the DC model in (2.11), where O_i^{DC} and W_D are set to 0 and 32 respectively. For the sake of discussion, let $w_i^{DC}(MB_n)$ be W_i^{DC} at MB_n of the current frame being encoded. It can be computed as the ratio of the mean value of MB_n in the current frame, $\overline{f_c(MB_n)}$, to the mean value of the co-located MB in the i^{th} reference frame, $\overline{f_i(MB_n)}$, given by

$$w_i^{DC}(MB_n) = W_D \cdot \frac{\overline{f_c(MB_n)}}{\overline{f_i(MB_n)}} \quad (5.1)$$

Two frames in the GBV scene are shown in Figure 5.1(a) and (b), and the values of $w_i^{DC}(MB_n)$ are illustrated in Figure 5.1(c). For this uniform brightness variation, it can be seen that the values of $w_i^{DC}(MB_n)$ are nearly constant. Therefore, the frame-based WP scheme is appropriate for this GBV scene. On the other hand, the values of $w_i^{DC}(MB_n)$ for two frames (Figure 5.2(a) and (b)) in the LBV scene is depicted in Figure 5.2(c). From this figure, it can be observed that the values of $w_i^{DC}(MB_n)$ are fluctuated significantly, but the values are relatively close among neighbouring MBs. It means that the degree of brightness variation may be kept nearly constant in a group of MBs. Based on this observation, it can be concluded that one group of MBs may be better coded with the use of one WP parameter set while another group may be coded more efficiently by using another WP parameter set. It is noted that a group of MBs is referred to as a region in the following discussions. If the WP parameter set is not accurate enough for the current MB, coding efficiency would be reduced. Thus, it is a key issue to have a number of accurate WP parameter sets for several regions in order to deal with the problem of LBVs.

Motivated by this, a region-based WP scheme is proposed to encode the LBV scenes in this chapter.

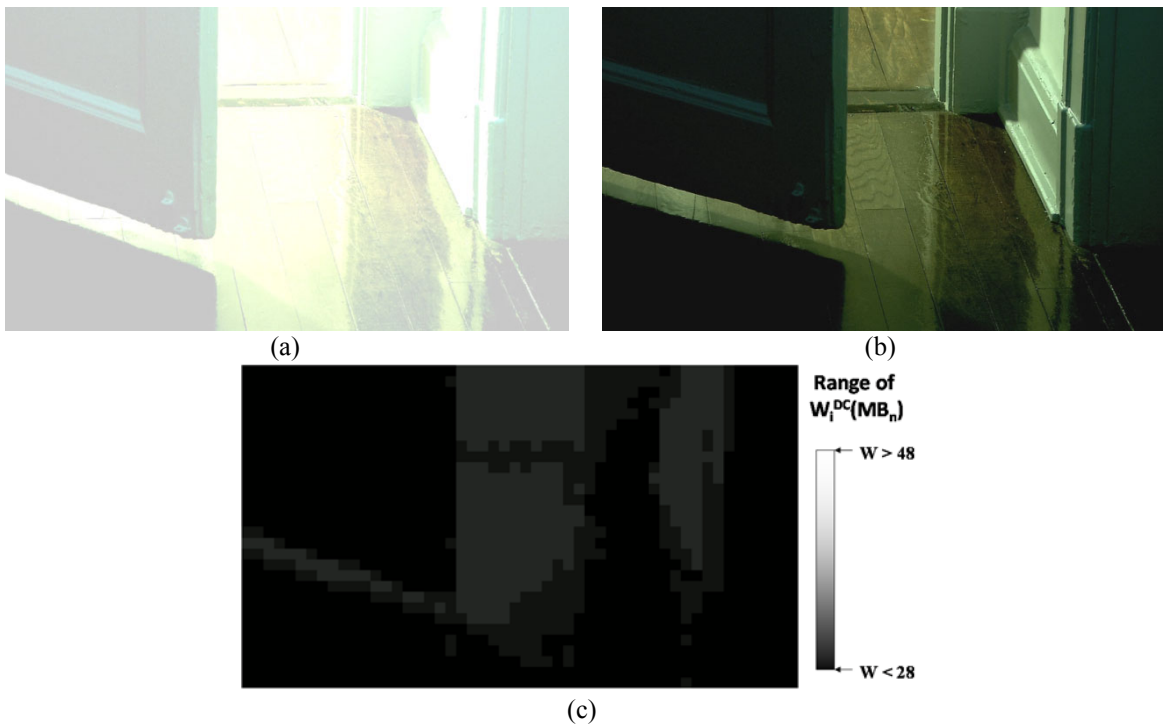


Figure 5.1. Original picture from the scene with GBV inside the "Mobisode2" sequence, (a) 234th frame, (b) 235th frame, and (c) estimated $W_i^{DC}(MB_n)$ using MB-based DC model.

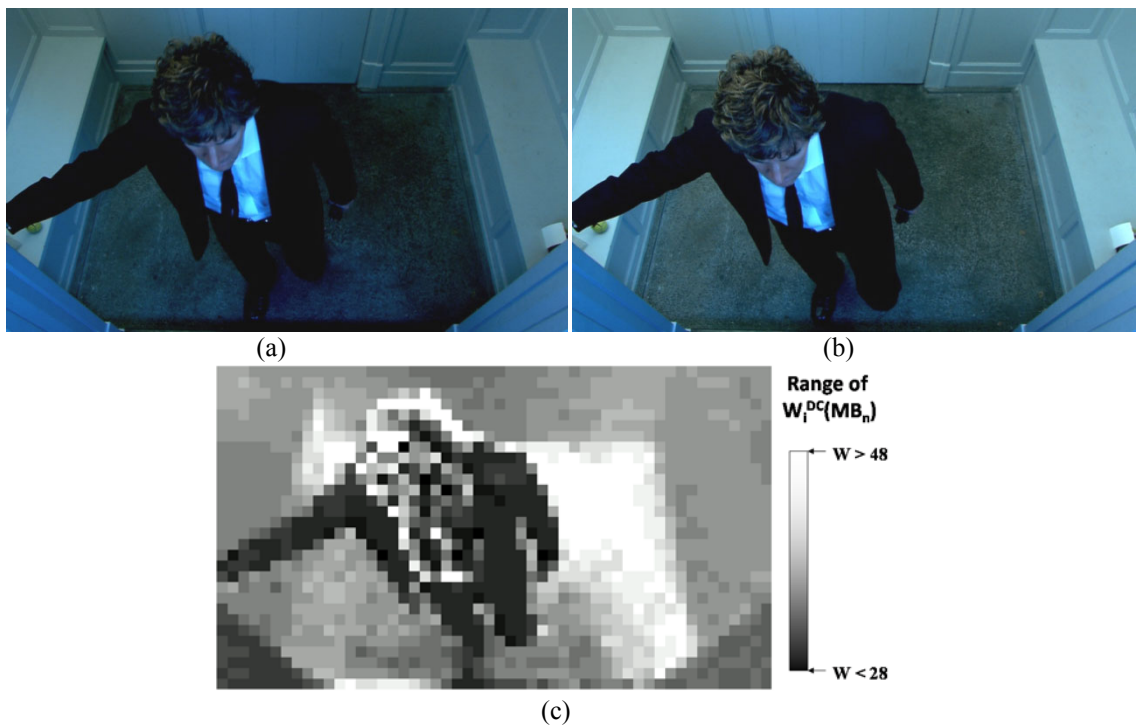


Figure 5.2. Original picture from the scene with LBV inside the "Mobisode2" sequence, (a) 46th frame, (b) 47th frame, and (c) estimated $W_i^{DC}(MB_n)$ using MB-based DC model.

5.3 Proposed Region-based Weighted Prediction Scheme

The work in this region-based coding scheme focuses on identifying regions with different brightness changes which gets benefit from the use of multiple WP parameter sets. The new scheme has three new features:

- 1) region partitioning based on brightness variation;
- 2) accurate estimation of region-based WP parameter sets;
- 3) adoption of a MRF architecture in WP that supports the coding of multiple WP parameter sets.

5.3.1 Region partitioning

The proposed region-based WP scheme attempts to partition or group regions according to the degrees of brightness variations. The scheme starts at MB level. In the first step, a simple DC model is used for the computation of $w_i^{DC}(MB_n)$ in (5.1). As discussed in Section 5.2, Figure 5.2(c) visually demonstrates $w_i^{DC}(MB_n)$ in the scene with LBV and its brightness variation histogram is depicted in Figure 5.3(b). The histogram is a representation of the distribution of $w_i^{DC}(MB_n)$ computed between the current frame and its i^{th} reference frame. For the current frame, the brightness variation histogram in Figure 5.3(b) represents the number of MBs that have brightness variation in each of a fixed list of brightness variation ranges. It is noted that the brightness variation histogram can be built for any kind of WP models in (2.11), (2.13), (2.14) and (2.16). The DC model or the offset model generates a one-dimensional histogram. On the other hand, the histogram of the LMS model is two dimensional, as it includes two variables - a weighting factor and an offset. It implies that the use of the LMS model involves two parameters for region

partitioning, which makes the process be complex. For simplicity, a one-dimensional histogram using the DC model is good enough for region partitioning. The histogram of Figure 5.3(b) shows the distribution of $W_i^{DC}(MB_n)$ in the LBV scene, and its broad histogram reflects that a single $W_i^{DC}(MB_n)$ is not sufficient for coding the LBV scene. It is contrast to the narrow histogram in the GBV scene, as depicted in Figure 5.3(a), where a single $W_i^{DC}(MB_n)$ is adequate.

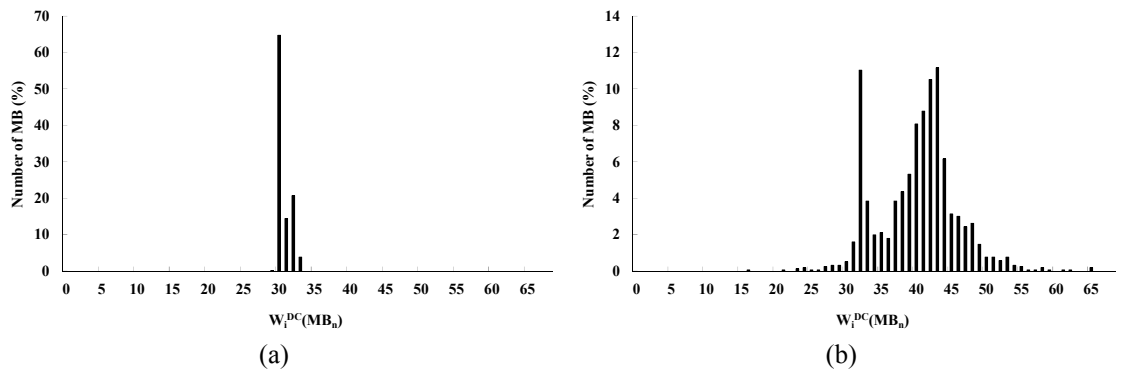


Figure 5.3. Distribution of $W_i^{DC}(MB_n)$ using the MB-based DC model in scenes with (a) GBV, and (b) LBV.

Region partitioning in the proposed scheme involves dividing the current frame into different regions where each one has some degree of uniformity in its brightness variation. If the possible values of $W_i^{DC}(MB_n)$ are sufficiently small, then each of those may be placed on a range by itself. For LBV scenes, the range of $W_i^{DC}(MB_n)$ is too large, as shown in the histogram of Figure 5.3(b), and it becomes inefficient to encode so many possible values of $W_i^{DC}(MB_n)$ by making use of the MRF architecture, as will be discussed later. To reduce the possible values of $W_i^{DC}(MB_n)$, the brightness variation histogram is divided into an appropriate number of ranges, often referred to as histogram bin, each containing many similar values of $W_i^{DC}(MB_n)$. To do that, $W_i^{DC}(MB_n)$ is quantized uniformly by a quantization factor Q into N -bin-histogram, as visually depicted in Figure 5.4(a) and (b).

From this figure, the quantization process leads to more compact region representations that have some degree of uniformity in their brightness variation. Basically, this N -bin-histogram forms N regions for the subsequent steps of the region-based WP scheme. However, the percentage of MBs in some histogram bins shown Figure 5.4(b) is insignificant. It is due to the fact that $w_i^{DC}(MB_n)$ in MBs with fast object motions cannot be estimated accurately without its true motion vector, and the true motion vectors cannot be obtained without the accurate $w_i^{DC}(MB_n)$. This forms a chicken-egg dilemma that some inaccurate $w_i^{DC}(MB_n)$ are obtained in MBs with object motions. To remove these unreliable MBs during the computation of the region-based WP parameter sets, only the most representative regions with N_R largest areas are selected among N regions. After this process, N_R regions (denoted as R_k , where $k=1, \dots, N_R$) are formed for the proposed region-based WP scheme, and these regions are characterized by their homogeneity in brightness variation.

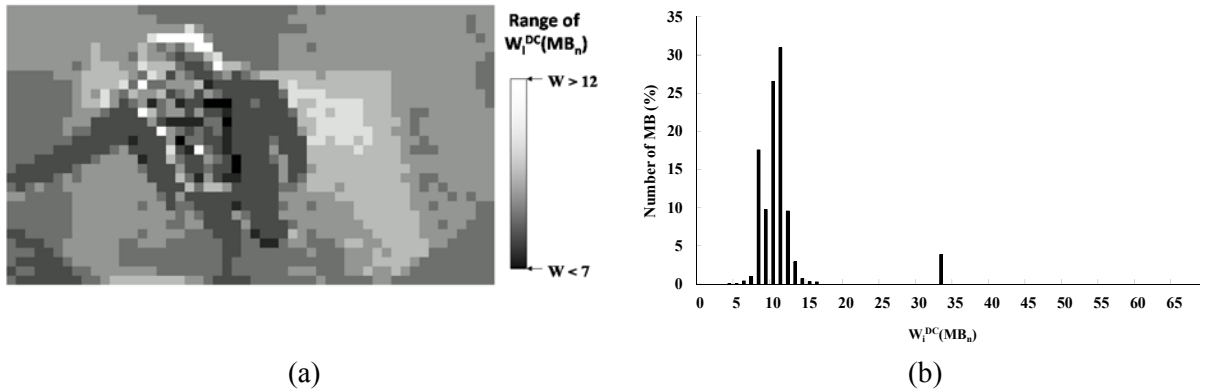


Figure 5.4. (a) $w_i^{DC}(MB_n)$ after quantization with $Q=4$, and (b) its corresponding histogram.

5.3.2 Determination of region-based WP parameters

A crucial component of the proposed region-based WP scheme is the estimation of the multiple WP parameter sets. Instead of using the simple DC model, a quasi-optimal estimator using the LMS model in (2.16) is adopted to compute the WP parameter sets

for R_k . Let us define them as $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$. (2.16) can then be modified as region basis, and can be written as:

$$W_i^{LMS}(R_k) = W_D \cdot \frac{\sum_{p_c \in f_c(R_k)} |p_c - \overline{f_c(R_k)}|}{\sum_{p_i \in f_i(R_k)} |p_i - \overline{f_i(R_k)}|} \quad \text{where } k = 1, 2, \dots, N_R \quad (5.2)$$

$$O_i^{LMS}(R_k) = \overline{f_c(R_k)} - \overline{f_i(R_k)} \cdot W_i^{LMS}(R_k) / W_D$$

where $f_c(R_k)$ and $f_i(R_k)$ are the pixels in the region R_k of the current frame f_c and the i^{th} reference frame f_i , respectively. $\overline{f_c(R_k)}$ denotes the mean pixel value within R_k in f_c , and $\overline{f_i(R_k)}$ is the mean pixel value of the co-located region in f_i .

In the conventional frame-based WP, all pixels in f_c and f_i are involved to calculate a single set of W_i^{LMS} and O_i^{LMS} , as formulated in (2.16). Since different regions in f_c exhibit different degrees of brightness variations in the LBV scene, the single set of W_i^{LMS} and O_i^{LMS} cannot reflect the real brightness variation in the scene. In comparison to frame-based WP, only pixels in R_k contribute to the computation of $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$. This way, we prevent $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$ of one region contaminating those in another region. The accuracy of the multiple sets of $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$ depends on the region partitioning. As mentioned in Section 5.3.1, the removal of unreliable MBs is able to reduce the effect of brightness variation due to object motions. In this case, only reliable pixels are used for computing $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$, which accurately reflect the degrees of LBV in different regions. This process offers a way to increase the accuracy of $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$.

5.3.3 Embedding multiple region-based WP parameter sets into the MRF framework of H.264/AVC

In [50, 54], the concept of assigning multiple WP parameters to the same reference was suggested by using multiple reference frame motion estimation (MRF-ME) [25-27]. It motivates us to use the MRF-ME framework in our proposed region-based scheme such that the multiple sets of $W_i^{LMS}(R_k)$ and $O_i^{LMS}(R_k)$ can be embedded into the H.264/AVC compliant bitstream.

In MRF-ME mentioned in section 3.2, a reference picture index (ref_idx) is embedded into the H.264/AVC bitstream to indicate which reference frame is used. In the conventional H.264/AVC WP scheme, a single WP parameter set is associated with each ref_idx . If MRF-ME is activated, more than one WP parameter sets are transmitted. Similar to Figure 3.1, Figure 5.5 depicts a block diagram to show the adoption of WP with MRF-ME in the H.264/AVC encoder using the LMS model. Assume that there are K_{ref} reference frames. K_{ref} sets of WP parameters, (W_i^{LMS}, O_i^{LMS}) , where $i=0, 1, \dots, K_{ref}-1$, are computed based on f_c and the reference frames from f_0 to $f_{K_{ref}-1}$. Their weighted reference frames, f_0^{WP} to $f_{K_{ref}-1}^{WP}$, are generated and put into the multiple frame buffers for motion estimation and compensation, as shown in Figure 5.5. The WP parameter set applied to the current MB is indicated by ref_idx . This arrangement allows the decoder to recognise the WP parameter set correctly without the need of additional bits. Even though multiple WP parameter sets are supported in H.264/AVC with MRF-ME, it only handles GBV scenes. It is because each calculation of (W_i^{LMS}, O_i^{LMS}) is still based on all pixels between the current and reference frames.

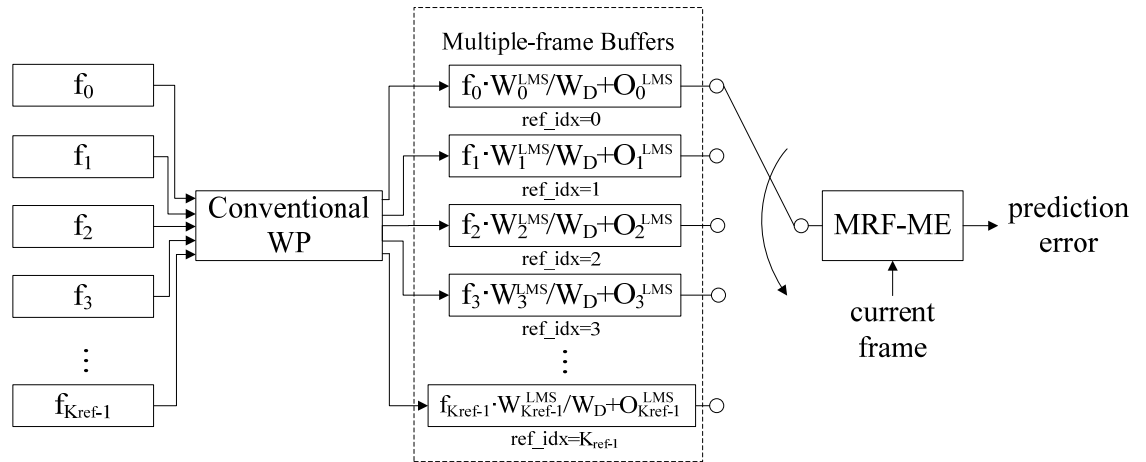


Figure 5.5. Conventional weighted prediction with MRF-ME in the H.264/AVC encoder.

Therefore, in this work, we further explore the technique in Chapter 3 [52,54] to the jointly use of multiple frame buffers and weighted prediction. In this case, more than one ref_idx can be associated with a particular reference picture. Figure 5.6 demonstrates the novel arrangement of the multiple frame buffers for the proposed region-based scheme. Instead of applying WP to multiple reference frames from f_0 to $f_{K_{ref}-1}$, different region-based WP parameter sets - $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$, where $k=1, 2, \dots, N_R$, are applied only to f_0 . Each $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ represents one single LBV within a certain region in the same reference frame, f_0 . The weighted reference frames associated with different value pairs of $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$, as shown in Figure 5.6, are stored in the multiple frame buffers for motion estimation and compensation. This arrangement allows different MBs in the current frame to employ different value pairs of $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ even they are predicted from the same reference frame. Again, ref_idx can be used to indicate which $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ to be used for each MB. An encoder that uses the proposed buffer arrangement can select the best $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ through rate-distortion optimization (RDO), resulting in handling various LBVs within the same frame. From Figure 5.6, the proposed scheme also keeps the reference frame without WP in the multiple frame buffers

for handling scenes in which the large luminance difference is mainly due to sudden camera motion, but not due to brightness variations, as discussed in section 3.3.

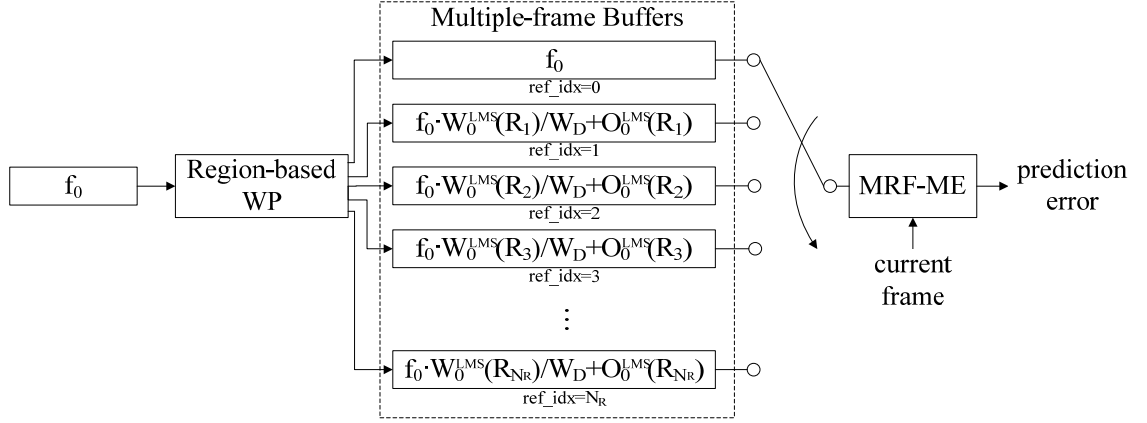


Figure 5.6. Region-based weighted prediction adopted in the H.264/AVC MRF-ME architecture.

To further improve the coding efficiency, we also re-organize the reference list of the multiple frame buffers so that the weighted reference frame giving the best prediction is the first one in the list. The default list of the reference frames in H.264/AVC is sorted according to display order. The order of this default list is sensible due to temporal proximity. In the proposed region-based WP scheme, different value pairs of $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ modify the same reference frame, and the default order starting with the most recent frame is no longer applied to the reference list. To re-solve this, the proposed scheme can work with the mechanism of reference list to re-order defined in H.264/AVC. This mechanism allows the encoder and decoder re-ordering the reference list in the best order. In order to determine the best order, the proposed scheme examines which $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ is likely to be used in the current frame being encoded. The reference picture list used in the current frame is then re-ordered based on the areas of different R_k so that $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ with the largest area is the first one in the list. This re-ordering strategy can ensure more MBs in the current frame being pointed to the first

reference frame in the list. This costs fewer bits to code ref_idx , which results in a decrease in bitrate of the encoded bitstream.

5.3.4 The flowchart of the proposed scheme

The flowchart of the proposed region-based WP is shown in Figure 5.7. It is mainly divided into three steps including region partitioning, determination of region-based WP parameter sets and embedding multiple WP parameter sets into the framework of MRF-ME in H.264/AVC. Let us provide a summary of the proposed scheme in the following:

- 1) Compute $w_i^{DC}(MB_n)$ using the DC model for each MB in the current frame being encoded.
- 2) Uniformly quantize $w_i^{DC}(MB_n)$ by a quantization factor Q into N -bin-histogram to form N regions.
- 3) Select R_k with the N_R largest areas among N regions in step (2).
- 4) Determine region-based WP parameter sets, $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ for $k=1, 2, \dots, N_R$, according to (5.2).
- 5) Generate N_R weighted reference frames according to $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ obtained in step (4).
- 6) Re-order the list of the reference frames in the frame buffer based on the areas of R_k so that $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$ with the largest area is the first one in the list.
- 7) Perform MRF-ME on the reference frames in the frame buffer and obtain ref_idx and motion vector for each MB. Note that each ref_idx is associated with a single $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$.

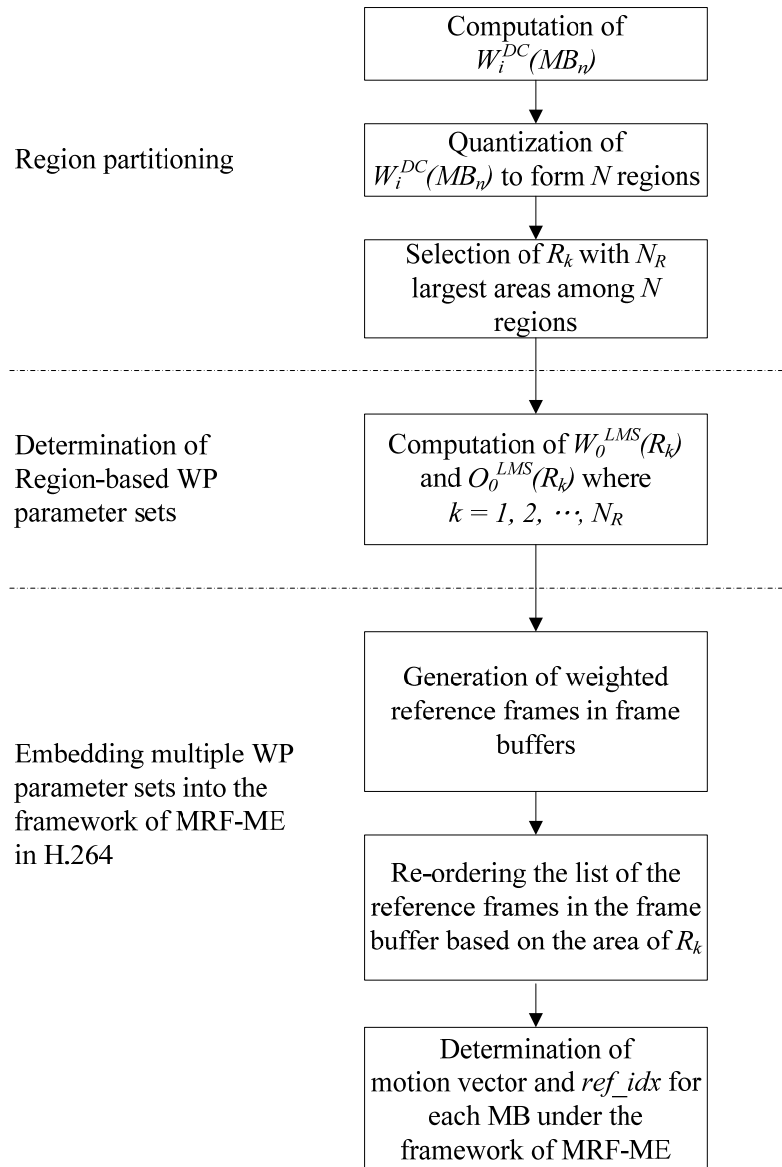


Figure 5.7. The flowchart of the proposed region-based WP scheme.

5.4 Reduction of Memory Requirement using Look-up Tables

The inherent nature of the proposed region-based scheme also provides benefit by using multiple pre-calculated look-up tables (LUTs) mentioned in section 3.4 to replace the multiple frame buffers. In the proposed scheme shown in Figure 5.6, N_R+1 frame buffers are required, including the buffer to store the original reference frame, f_0 , and N_R weighted

reference frames, which are formed by different parameter sets - $(W_0^{LMS}(R_1), O_0^{LMS}(R_1))$, $(W_0^{LMS}(R_2), O_0^{LMS}(R_2))$, ..., $(W_0^{LMS}(R_k), O_0^{LMS}(R_k))$. In motion estimation, the candidate MBs for each reference frame in the frame buffer are searched to find the best matching candidate of the current MB. It is noted that all the weighted reference frames are based on f_0 , but only in a modified form with the different WP parameter sets. To exclude the use of the frame buffers for weighted reference frames, the LUT is created once per weighted reference frame, as depicted in Figure 5.8. The entries of each LUT keep the pixels of the weighted reference frame and can be computed by modifying f_0 with its associated WP parameter set. The pixel value of f_0 is treated as the index for these tables. An 8-bit video signal gives a range of possible values from 0 to 255, and therefore the size of each LUT contains 256 entries. Instead of accessing those weighted reference frames of SAD calculation in (2.8), it is computed via LUTs in (3.1). By doing so, the encoder only requires to keep a single frame buffer for f_0 and N_R LUTs. As each LUT contains 256 bytes, the size of N_R LUTs is of little size in comparison with the size of the frame memory.

LUT _k	
0	$O_0^{LMS}(R_k)$
1	$W_0^{LMS}(R_k)/W_D + O_0^{LMS}(R_k)$
2	$2W_0^{LMS}(R_k)/W_D + O_0^{LMS}(R_k)$
3	$3W_0^{LMS}(R_k)/W_D + O_0^{LMS}(R_k)$
⋮	⋮
255	$255W_0^{LMS}(R_k)/W_D + O_0^{LMS}(R_k)$

Figure 5.8. Entries of LUT_k, where $k=1, 2, \dots, N_R$.

Note that for an encoder conforming to the H.264/AVC standard, the encoder must produce a bitstream that meets the requirements of the specified syntax and is capable of being decoded using the decoding process described in the H.264/AVC standard. The standard does not specify the implementation of the encoding processing, in order to give designers the flexibility to choose their own method of encoding. The use of LUTs is only the memory saving implementation of the proposed scheme, and does not affect the syntax of producing a compliant H.264/AVC bitstream.

5.5 Experimental Results

We performed computer simulation on two standard video sequences in WVGA format (832×480) - “Mobisode1” and “Mobisode2”, which contain plentiful scenes with different types of brightness variations. These two sequences were split into various segments for performance evaluation in either GBV or LBV scenes. The details of those video segments are tabulated in Table 5.1. Besides, five high resolution (HD) movie trailers downloaded from Apple iTunes [77] were also used for experimental evaluation. These trailer sequences include “Iamnumber4”, “Inception”, “Ironman2”, “Meninblack3”, and “Tranformer3”, which have lots of LBV scenes, as described in Table 5.2.

Table 5.1. Details of various video segments of “Mobisode1” and “Mobisode2” used for simulation

Segment Name	Segment	Characteristics
Mo1_s1 (GBV)	Scene 1 of Mobisode1 (WVGA, 832x480, Frame 178 to 186)	GBV with fade-in-from-white effect
Mo1_s2 (GBV)	Scene 2 of Mobisode1 (WVGA, 832x480, Frame 230 to 249)	GBV with fade-in-from-white effect
Mo1_s3 (LBV)	Scene 3 of Mobisode1 (WVGA, 832x480, Frame 0 to 29)	LBV with a house at the center
Mo2_s1 (LBV)	Scene 1 of Mobisode2 (WVGA, 832x480, Frame 42 to 51)	LBV with a person turning on a light in a room

Mo2_s2 (LBV)	Scene 2 of Mobisode2 (WVGA, 832x480, Frame 288 to 299)	LBV with a door closing
-----------------	---	-------------------------

Table 5.2. Details of HD movie trailer downloaded from Apple iTunes [77] for simulation

Trailer Name	Sources	Characteristics
Iamnumber4	Trailer (1280x688) from http://trailers.apple.com/movies/dreamworks/iamnumberfour/	LBV with a light source zooming out at the centre
Inception	Trailer 3 (1280x544) from http://trailers.apple.com/trailers/wb/inception/	LBV at both foreground and background with windblown hair of an actress
Ironman2	Trailer 2 (1280x544) from http://trailers.apple.com/trailers/paramount/ironman/	LBV with a person walking towards a shocking camera due to explosion at the background
Meninblack3	Trailer (1280x688) from http://trailers.apple.com/trailers/sony_pictures/meninblack3/	LBV with a light source zooming in followed by zooming out near the centre
Transformer3	Trailer 2 (1280x532) from http://trailers.apple.com/trailers/paramount/transformersdarkofthemoon/	LBV with multiple explosions at different spatial locations in the war scene

We have incorporated the proposed schemes using the LMS model in (5.2) with and without LUTs into the H.264 JM 15.1 [16,17], and let us call them REGION-WP+LUT and REGION-WP, respectively. For REGION-WP+LUT and REGION-WP, Q in step (2) and N_R in step (3) of Section 5.3.4 were both set to 4. That is, four regions are defined. They were used to compare the performances of the conventional frame-based WP scheme using the LMS model [16] (H.264-WP), and two non-standard-complaint schemes - the adaptive weighted prediction in [57] (ADAPTIVE-WP), and localized weighted prediction in [60] (LOCAL-WP). MRF-ME was also enabled for the simulation of H.264-WP, ADAPTIVE-WP and LOCAL-WP approaches because it achieves better predictions than those using just single reference frame in scenes with light changes [27]. Note that, for REGION-WP+LUT and REGION-WP, N_R is equal to 4. It is fair to use five reference pictures in H.264-WP, ADAPTIVE-WP and LOCAL-WP such that all schemes require the same number of frame buffers. For H.264-WP and ADAPTIVE-WP, WP has been employed in the five reference frames, and its implementation is shown in

Figure 5.5. For Local-WP, an offset is estimated for each searching point during motion estimation for the five reference frames. In summary, all experiments were conducted using IPPP... structure, Main profile, five reference pictures, quarter-pel full search motion estimation with search range of ± 32 pixels, RDO with all seven inter modes as well as skip mode and intra modes, and context-adaptive binary arithmetic coding (CABAC). Four QPs (i.e. QP=20, 24, 28, and 32) were used for encoding the bitstreams [72]. It is noted that other settings such as fast motion estimation technique or RDO is off can also be applied for evaluating the performance.

5.5.1 Rate distortion performances of the proposed scheme

Figure 5.9 shows the rate-distortion (RD) performances of different schemes for “Mo2_s1(LBV)”. This sequence is a shot of a person turning on a light in a room that causes local brightness variation. It is noted that the RD performances of REGION-WP and REGION-WP+LUT are the same since the purpose of REGION-WP+LUT is to reduce the memory usage of REGION-WP without affecting their coding efficiency. For the sake of simplicity, the same curve is used for showing their performances. Figure 5.9 also includes the RD curve when WP is not applied, and it is referred to as ‘WITHOUT WP’. From this figure, it can be observed that the coding gains of REGION-WP and REGION-WP+LUT are about 1dB improvement over ‘WITHOUT WP’. The gains are still remarkable in comparison with H.264-WP. It is because H.264-WP is a frame-based method and can only use one WP parameter set for each reference frame. Note that by applying only one WP parameter set to all pixels of whole frame is not efficient for coding a video with LBVs in which brightness variations happen only in partial regions but not in whole frame.

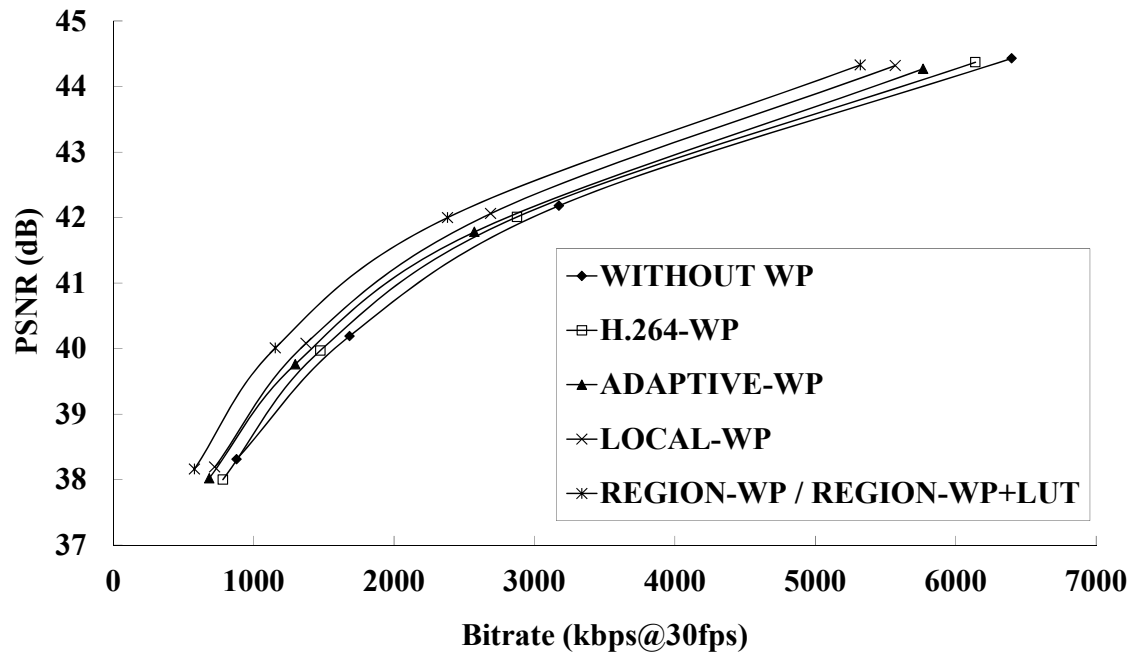


Figure 5.9. RD performances of different approaches for “Mo2_s1 (LBV)”.

Figure 5.9 also shows the comparison of REGION-WP and REGION-WP+LUT with the non-standard-compliant schemes - ADAPTIVE-WP and LOCAL-WP. ADAPTIVE-WP tries to choose the most common WP parameter set for the region with LBVs. However, this common WP parameter set for each reference frame is still not sufficient for the scenes that have multiple LBVs in different regions happening at the same time instant. For LOCAL-WP, it assumes that the spatial correlation is high between pixels of the current MB and pixels of neighboring MBs, the strong spatial correlation may not be always true for every block. It results in lowering the accuracy of estimating the block-based WP parameter sets. As expected and shown in Figure 5.9, they all obtain the inferior RD performances as compared with our proposed REGION-WP and REGION-WP+LUT. The significant improvements of the proposed schemes are due to the benefit of supporting multiple WP parameter sets in the MRF-ME framework. The evidence is further illustrated in Figure 5.10. This figure shows the graphical representation of each

8x8 block selecting which references or being intra-coded (intra modes) using REGION-WP and REGION-WP+LUT for frame 5 of “Mo2_s1(LBV)” at QP 28. It can be seen that those blocks belonging to the black suit of the guy mainly choose the reference with $ref_idx=0$. It is the original reference frame without performing weighted prediction. The use of this reference frame in these blocks is due to the black pigment of the suit which absorbs all light and no brightness variation is resulted within the scene. On the other hand, the blocks inside the carpet and the wall at the right side mainly select the references with $ref_idx=1$ and $ref_idx=2$ as their optimal references since they are undergone different degrees of LBVs. It means that only one WP parameter set cannot get an optimal solution in the encoded frame of LBV scenes, but an appropriate WP parameter set is selected from $(W_0^{LMS}(R_1), O_0^{LMS}(R_1))$, $(W_0^{LMS}(R_2), O_0^{LMS}(R_2))$, ..., $(W_0^{LMS}(R_{N_R}), O_0^{LMS}(R_{N_R}))$ to get the smallest cost via RDO. Therefore, REGION-WP and REGION-WP+LUT have the advantage over other schemes.



Figure 5.10. Statistics of selected references and intra modes using REGION-WP or REGION-WP+LUT for frame 5 of “Mo2_s1 (LBV)” at QP 28.

Table 5.3 shows the average distribution of inter and intra modes for using different schemes to encode scenes with different types of brightness variations. It can be seen that REGION-WP and REGION-WP+LUT successfully suppress the number of intra blocks as well as boost up the number of inter blocks by increasing the temporal correlations between the current and reference frames. It is well known that the increase in inter modes benefits the coding gain and the evidence is shown in Table 5.4. In this table, experimental results for both standard and trailer sequences with various brightness variations using the Bjontegaard delta bitrate (BD-Bitrate) and Bjontegaard delta PSNR (BD-PSNR) [73] compared to ‘WITHOUT WP’ are shown. From Table 5.4, it is obvious that REGION-WP and REGION-WP+LUT can overwhelmingly outperform other schemes for coding the scenes with LBVs. In the meantime, our schemes can still obtain similar performances as the conventional H.264/AVC frame-based scheme, H.264-WP, when dealing with the scenes with GBVs. It is contrast to the performances of ADAPTIVE-WP and LOCAL-WP which obtain lower coding efficiency compared with ‘WITHOUT WP’, as manifested in “Mo1_s1(GBV)” and “Mo1_s2(GBV)” of Table 5.4. From this table, we can conclude that our proposed REGION-WP and REGION-WP+LUT can work well in both LBV and GBV scenes.

Table 5.3. Average percentage (%) of inter and intra modes for different schemes

		WITHOUT WP	H.264-WP	ADAPTIVE-WP	LOCAL-WP	REGION-WP / REGION-WP+LUT
		Standard Sequences				
Mo1_s1 (GBV)	Intra	54.02	44.31	46.50	46.72	28.92
	Inter	45.98	55.69	53.50	53.28	71.08
Mo1_s2 (GBV)	Intra	90.27	26.87	54.72	46.32	29.14
	Inter	9.730	73.13	45.28	53.68	70.86
Mo1_s3 (LBV)	Intra	6.750	15.93	11.86	6.750	4.330
	Inter	93.25	84.07	88.14	93.25	95.67
Mo2_s1 (LBV)	Intra	51.91	54.92	38.75	38.02	30.19
	Inter	48.09	45.08	61.25	61.98	69.81
Mo2_s2 (LBV)	Intra	19.78	17.56	26.70	19.78	17.69
	Inter	80.22	82.44	73.30	80.22	82.31

		Trailer Sequences				
Iamnumber4 (LBV)	Intra	76.99	78.39	76.23	76.27	72.18
	Inter	23.01	21.61	23.77	23.73	27.82
Inception (LBV)	Intra	55.71	70.09	54.31	55.71	47.95
	Inter	44.29	29.91	45.69	44.29	52.05
Ironman2 (LBV)	Intra	65.40	66.99	62.71	64.20	62.09
	Inter	34.60	33.01	37.29	35.80	37.91
Meninblack3 (LBV)	Intra	62.79	72.91	61.88	62.51	60.25
	Inter	37.21	27.09	38.12	37.49	39.75
Transformer3 (LBV)	Intra	41.37	39.60	37.95	32.32	28.41
	Inter	58.63	60.40	62.05	67.68	71.59

Table 5.4. BD-Bitrate (%) and BD-PSNR (dB) of various schemes compared to WITHOUT WP

		H.264-WP	ADAPTIVE-WP	LOCAL-WP	REGION-WP / REGION-WP+LUT
Standard Sequences					
Mo1_s1 (GBV)	BD-Birate	-7.59	-3.20	-3.47	-12.08
	BD-PSNR	0.46	0.20	0.21	0.74
Mo1_s2 (GBV)	BD-Birate	-51.41	-6.93	-47.71	-51.22
	BD-PSNR	2.04	0.23	1.92	1.97
Mo1_s3 (LBV)	BD-Birate	3.27	0.56	0.02	-9.17
	BD-PSNR	-0.13	-0.02	-0.00	0.33
Mo2_s1 (LBV)	BD-Birate	-4.21	-8.38	-13.07	-22.91
	BD-PSNR	0.13	0.25	0.43	0.75
Mo2_s2 (LBV)	BD-Birate	-5.88	0.95	0.07	-9.32
	BD-PSNR	0.06	-0.02	-0.00	0.13
Trailer Sequences					
Iamnumber4 (LBV)	BD-Birate	-2.12	-1.32	-4.55	-5.69
	BD-PSNR	0.09	0.05	0.18	0.22
Inception (LBV)	BD-Birate	6.21	-1.20	0.01	-6.04
	BD-PSNR	-0.21	0.04	0.00	0.20
Ironman2 (LBV)	BD-Birate	0.21	-2.39	-1.04	-2.86
	BD-PSNR	-0.01	0.11	0.05	0.11
Meninblack3 (LBV)	BD-Birate	10.96	-2.29	-0.47	-3.37
	BD-PSNR	-0.50	0.11	0.03	0.18
Transformer3 (LBV)	BD-Birate	-3.95	-6.57	-13.98	-17.07
	BD-PSNR	0.22	0.35	0.80	0.94

In Table 5.4, it is also interesting to note that the coding gain is smaller for “Ironman2”. This sequence is a shot of a person walking towards a shocking camera due to explosion at the background. The shocking camera introduces global motion. From (5.1), the computation of $w_i^{DC}(MB_n)$ is made using the co-located MB in the reference frame. This assumes that the motion between frames is small in general. The global motion might affect the accuracy of computing $w_i^{DC}(MB_n)$ for region partitioning. As a result, the proposed REGION-WP cannot work very well in the sequence with global motion. Further suggestion to compensate for global motion in REGION-WP will be provided in Chapter 5.5.4.

5.5.2 Analysis of memory requirement

For the schemes except REGION-WP+LUT, they access the weighted reference frames stored in the multiple-frame buffers, and fetch the necessary weighted pixel values without further weighting calculation. They require large memory requirement for the encoder since the multiple weighted reference frames must be maintained in memory. Given a video of frame size $W \times H$ and $N_R + 1$ reference frames, the memory size requirement is:

$$(N_R + 1) \times 4W \times 4H \quad (5.4)$$

where a factor of 4 is due to the use of quarter-pel motion estimation. On the other hand, REGION-WP+LUT requires only one picture memory of $4W \times 4H$ for the reference frame F_0 . In addition, an extra 256-byte LUT associated with each weighted reference frame is pre-computed based on Figure 5.8 and stored in memory. The total size of the memory requirement for REGION-WP+LUT is therefore reduced as:

$$4W \times 4H + N_R \times 256 \quad (5.5)$$

Table 5.5 then summaries the required memory of different schemes. It can be seen that REGION-WP+LUT has savings of about 80% in comparison with other schemes for both encoder and decoder. It is due to the fact that REGION-WP+LUT can use four 256-byte LUTs to replace four weighted reference frames in the frame buffers.

Table 5.5. Memory requirements for various schemes with 5 reference frames for coding WVGA video

	WITHOUT WP/H.264-WP/ADAPTIVE-WP/LOCAL-WP/REGION-WP	REGION-WP+LUT
Memory for Reference pictures	5*4*832*4*480	4*832*4*480
Memory for Look-up tables	0	4*256
Total Memory in Megabytes (KB)	31200	6241 (20%)

5.5.3 Comparison of encoding complexity

To show the complexity of the proposed schemes, the encoding time increment as compared to WITHOUT WP for all test sequences is measured and tabulated in Table 5.6. The experiments were performed on an Intel Xeon X5550 2.67GHz computer with 12GB memory. From Table 5.6, all tested schemes require extra complexity to estimate the WP parameter sets and generate the weighted reference frames. The main difference in terms of encoding complexity between H.264-WP and ADAPTIVE-WP is very small since they both use a straightforward WP estimator. On the other hand, LOCAL-WP is an MB-based scheme where the WP parameter set per MB is calculated based on the neighbouring pixel values of the MB being encoded and those of MB in the reference frame. The parameter set of each MB is then necessary to be estimated on-the-fly during motion estimation, thus time elapsed is evidently increased, as shown in Table 5.6. For the proposed REGION-WP, the estimation of region-based WP parameter sets is carried out by two-pass WP estimators. This first pass is to partition regions with a simplified WP estimator while the second pass is to estimate accurately the WP parameter sets with a quasi-optimal WP estimator. From Table 5.6, it can be found that the increase in complexity of our proposed REGION-WP is not remarkable as compared with those in H.264-WP and ADAPTIVE-WP. It also shows significant savings as compared with LOCAL-WP. It is observed that the use of the LUT technique in REGION-WP+LUT causes an increase in computational complexity. It is because the look-up process includes additional memory access in the implementation of (3.1). Nevertheless, the increase in encoding time can achieve memory reduction of 80%.

Table 5.6. Total encoding time increment of various schemes compared to WITHOUT WP

	H.264-WP	ADAPTIVE-WP	LOCAL-WP	REGION-WP	REGION-WP+LUT
Standard Sequences					
Mo1_s1 (GBV)	4.70%	6.45%	17.49%	7.58%	15.26%
Mo1_s2 (GBV)	3.10%	5.48%	18.65%	7.04%	14.22%
Mo1_s3 (LBV)	4.23%	6.44%	14.95%	8.92%	15.73%
Mo2_s1 (LBV)	2.80%	4.95%	20.70%	6.80%	16.31%
Mo2_s2 (LBV)	3.29%	3.29%	18.48%	9.01%	19.01%
Trailer Sequences					
Iamnumber4 (LBV)	2.88%	2.54%	20.12%	6.04%	18.20%
Inception (LBV)	3.55%	4.90%	21.59%	6.31%	19.44%
Ironman2 (LBV)	5.08%	5.59%	23.40%	8.81%	22.83%
Meninblack3 (LBV)	4.86%	4.70%	20.51%	9.25%	22.07%
Transformer3 (LBV)	4.24%	4.04%	23.26%	6.81%	19.13%
Average	3.87%	4.84%	19.92%	7.66%	18.22%

5.5.4 Impact of the GOP Structure with B-frames on REGION-WP/REGION-WP+LUT

In the following, we discuss the impact of the GOP structure with B-frames on the proposed REGION-WP and REGION-WP+LUT. Each 8×8 block in a B-frame may be predicted in one of several ways: motion-compensated prediction from list 0 reference frames, motion-compensated prediction from list 1 reference frames or motion-compensated bi-directional prediction from list 0 and list 1 reference frames. Basically, all techniques proposed in this chapter can be easily extended to B-frame coding. For adopting REGION-WP and REGION-WP+LUT in list 1 prediction of B-frames, the same arrangement of the multiple frame buffers or LUTs in MRF-ME as list 0 prediction can be directly applied. The significant improvement of REGION-WP and REGION-WP+LUT is due to the benefit of region partitioning, which depends on $w_i^{DC}(MB_n)$. From (5.1), the computation of $w_i^{DC}(MB_n)$ uses the co-located MB in the reference frames. This

assumes the motion between frames is small. However, a typical GOP structure with B frames always contains larger temporal distance between the current and reference frames. It potentially induces larger motion vectors. As a result, $W_i^{DC}(MB_n)$ becomes inaccurate if the MB contains object motion along the scene. In principle, the computation of $W_i^{DC}(MB_n)$ can take the motion into account by modifying (5.1) into

$$W_i^{DC}(MB_n(\hat{\delta}x, \hat{\delta}y)) = W_D \cdot \frac{\overline{f_c(MB_n)}}{f_i(MB_n(\hat{\delta}x, \hat{\delta}y))} \quad (5.6)$$

where $\overline{f_c(MB_n(\hat{\delta}x, \hat{\delta}y))}$ is the mean value of MB_n in f_i displaced by $(\hat{\delta}x, \hat{\delta}y)$, which is the initial motion vector compensated for object motion in MB_n . It is estimated by minimizing the SAD in the search range defined as follows:

$$SAD(\delta x, \delta y) = \sum_{\substack{p_c \in f_c(MB_n(x, y)) \\ p_i \in f_i(MB_n(x+\delta x, y+\delta y))}} \left| p_c - \frac{\overline{f_c(MB_n)}}{f_i(MB_n(\delta x, \delta y))} p_i \right| \quad (5.7)$$

Table 5.7 tabulates the BD-Bitrate and BD-PSNR [73] compared to ‘WITHOUT WP’ for different schemes using the hierarchical-B encoding structure (IbBbBbBbP) shown in Figure 5.11. REGION-WP+MV and REGION-WP+LUT+MV are the proposed REGION-WP and REGION-WP+LUT with the use of (5.6) for region partitioning. Except the encoding structure, the same encoding configuration and test sequences in the previous section were employed. It is obvious from Table 5.7 that REGION-WP and REGION-WP+LUT can still obtain better coding efficiency compared with ADAPTIVE-WP, LOCAL-WP, and H.264-WP. This table also shows that further coding gain can be achieved by REGION-WP+MV and REGION-WP+LUT+MV in the hierarchical-B encoding structure since both of the local brightness variation and object motion are taken into consideration of computing $W_i^{DC}(MB_n(\hat{\delta}x, \hat{\delta}y))$. However, this technique includes additional computational complexity of the estimation of $(\hat{\delta}x, \hat{\delta}y)$, as shown in Table 5.8

where the average encoding time increment as compared to WITHOUT WP for all test sequences is measured. In comparison with ‘WITHOUT WP’, the encoding time of REGION-WP+MV/REGION-WP+LUT+MV is increased by 23.36%/36.43% in average whereas the encoding time of REGION-WP/REGION-WP+LUT is only increased by 16.45%/25.80% in average.

Table 5.7. BD-Bitrate (%) and BD-PSNR (dB) of various algorithms compared to WITHOUT WP using the hierarchical-B encoding structure.

		H.264-WP	ADAPTIVE-WP	LOCAL-WP	REGION-WP / REGION-WP+LUT	REGION-WP+MV / REGION-WP+LUT+MV
Standard Sequences						
Mo1_s1 (GBV)	BD-Birate	-11.58	-14.63	-1.48	-10.21	-12.25
	BD-PSNR	0.61	0.83	0.08	0.54	0.66
Mo1_s2 (GBV)	BD-Birate	-54.09	-9.47	-54.96	-55.24	-58.63
	BD-PSNR	2.06	0.26	2.10	2.12	2.29
Mo1_s3 (LBV)	BD-Birate	-1.39	1.36	0.03	-5.74	-12.26
	BD-PSNR	0.04	-0.05	0.00	0.21	0.46
Mo2_s1 (LBV)	BD-Birate	-1.80	-3.14	-3.99	-17.21	-20.43
	BD-PSNR	0.05	0.11	0.12	0.57	0.69
Mo2_s2 (LBV)	BD-Birate	24.72	2.21	0.10	-4.78	-4.42
	BD-PSNR	-0.28	-0.03	0.00	0.07	0.05
Trailer Sequences						
Iamnumber4 (LBV)	BD-Birate	-0.01	2.62	-3.81	-5.15	-9.90
	BD-PSNR	-0.01	-0.12	0.20	0.24	0.50
Inception (LBV)	BD-Birate	4.98	-1.63	0.03	-8.91	-9.60
	BD-PSNR	-0.19	0.07	0.00	0.35	0.38
Ironman2 (LBV)	BD-Birate	2.90	-3.24	-1.55	-4.41	-7.19
	BD-PSNR	-0.12	0.14	0.07	0.19	0.31
Meninblack3 (LBV)	BD-Birate	21.71	-0.66	-0.36	-3.35	-4.48
	BD-PSNR	-1.22	0.03	0.02	0.19	0.27
Transformer3 (LBV)	BD-Birate	-6.57	-4.81	-14.51	-20.12	-24.51
	BD-PSNR	0.31	0.23	0.71	1.01	1.26

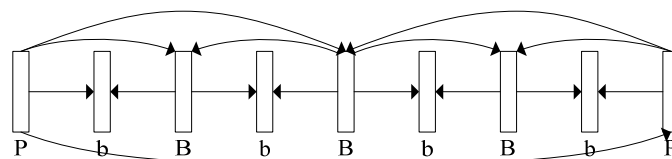


Figure 5.11. Typical hierarchical-B coding structure.

Table 5.8. Average encoding time increment of various schemes compared to WITHOUT WP when the hierarchical-B encoding structure is used.

H.264-WP	ADAPTIVE-WP	LOCAL-WP	REGION-WP	REGION-WP +LUT	REGION-WP +MV	REGION-WP +LUT+MV
5.07%	8.72%	24.10%	16.45%	25.80%	23.36%	36.43%

It is important to note that, as comparing with the result from Table 5.6, the average encoding time of various schemes is augmented when B-frames are encoded. It is because encoders cannot use the weighted reference frames stored in the multiple frame buffers or LUTs in case of bi-directional prediction. In the implementation of multiple frame buffers in Figure 5.5 and Figure 5.6, all weighted reference frames are rounded and clipped to the range of $[0,255]$. For bi-directional prediction, the prediction MB, $f_{BI}^{WP}(MB_n)$, is calculated as a weighted average of the list 0 and list 1 prediction MBs, and is written as

$$f_{BI}^{WP}(MB_n) = Clip / Round \left(\frac{W_{list0} \times f_{list0}(MB_n) + W_{list1} \times f_{list1}(MB_n)}{2 \times W_D} + \frac{O_{list0} + O_{list1}}{2} \right) \quad (5.8)$$

where $f_{list0}(MB_n)$ and $f_{list1}(MB_n)$ are the prediction MBs of the list 0 reference frame and the list 1 reference frame, respectively; (W_{list0}, O_{list0}) and (W_{list1}, O_{list1}) are the WP parameter sets for the list 0 reference frame and the list 1 reference frame, and $Clip/Round()$ is the necessary clipping and rounding operations used in the H.264/AVC standard [12,16]. However, only MBs of the weighted reference frames for uni-directional prediction such as $Clip / Round((W_{list0} \times f_{list0}(MB_n))/W_D + O_{list0})$ and $Clip / Round((W_{list1} \times f_{list1}(MB_n))/W_D + O_{list1})$ are available in the multiple frame buffers of MRF. Owing to the non-linear property of rounding and clipping operations, $f_{BI}^{WP}(MB_n)$ cannot be obtained from $Clip / Round((W_{list0} \times f_{list0}(MB_n))/W_D + O_{list0})$ and

$Clip / Round \left((W_{list1} \times f_{list1}(MB_n)) / W_D + O_{list1} \right)$. In other words, bi-directional prediction needs to estimate predicted samples on the fly for all the tested schemes when B-frames are adopted, resulting in an increase in the computational complexity. This is also the limitation of the proposed schemes and other WP schemes when the weighted reference frames are required.

5.6 Chapter Summary

In this chapter, we found that the distribution of the values of WP parameter sets in the frame of the LBV scene is varied greatly, but they are relatively close among neighbouring MBs. Motivated by this, we have proposed a novel region-based WP scheme to encode scenes with local brightness variation. It consists of three main steps. First, different regions are grouped in accordance with the degree of uniformity in its brightness variation. Second, the precise region-based WP parameter sets are estimated based on the partition. Finally, the framework of MRF-ME is adopted to encode the multiple WP parameter sets without explicitly coding the information of regions. The H.264/AVC framework can then support WP parameter adaptation at a region level. Experimental results show that the proposed region-based WP scheme can improve the coding performance for scenes with local brightness variations while still maintaining the coding performance for scenes with GBVs in comparison with other WP schemes. In addition, we further incorporate the LUT technique with our region-based scheme to reduce the memory requirement effectively. For further development, if there is no constraint of standard compliance, it is possible to share the region partitioning results such as contour to the decoder for enhancing region-based WP in terms of the coding performance.

Chapter 6 Conclusions and Future Work

In this thesis, we have carried out a study on weighted prediction (WP) schemes for coding scenes with global brightness variations (GBVs) and local brightness variations (LBVs). The coding mechanisms of the WP scheme in the H.264/AVC standard and some related techniques in the literature have been studied in details. Results of our study indicated that there is plenty of room for improvement. Therefore, in this research, three different schemes have been employed to provide efficient solutions for coding scenes with GBVs and LBVs. These schemes have been shown to achieve substantial coding gain in scenes with different types of brightness variations. In this chapter, we highlight the main contributions of this thesis and then suggest some possible directions that could be the focus of future research.

6.1 Contributions of the Thesis

Our contributions chiefly include a comprehensive study of WP coding in order to (i) facilitate the use of multiple WP models in coding scenes with diverse brightness variations, (ii) handle changes in illumination caused by a flash being fired during a press conference, a sport match, a news interview, etc., and (iii) contrive a region-based scheme for efficiently coding scene with LBVs. In particular, our conclusions are:

- There are several different WP models available for coding scenes with various types of brightness variations. But no single model can deal with all fading effects. Therefore, the discussions in Chapter 3 result in designing a single reference frame

multiple WP models (SRefMWP) scheme that utilizes the structure of multiple reference frames in H.264/AVC. It facilitates the use of multiple WP models to compensate for non-uniform brightness variations in video scenes or sequences with different fading effects. The proposed technique does not need to explicitly code and transmit the WP parameters to support parameter adaptation at a MB level. Moreover, the use of look-up tables can help to reduce the memory requirement of SRefMWP. Experimental results show that SRefMWP can achieve significant coding gain in scenes with different types of brightness variations. Furthermore, SRefMWP with LUTs can reduce the memory requirement by about 80% while keeping the same coding efficiency as SRefMWP.

- When video sequences are taken in a press conference, a sport match, and a news interview, flash lighting often comes into view due to the photographing by journalists. It is difficult for the existing WP scheme to encode flashlight (FL) scenes since the intensity changes drastically and non-uniformly. In coding of FL scenes, motion estimation cannot locate a well-matched block in reference frames. The MB-based scheme in Chapter 4 is specifically designed for coding FL frames. With the use of an adaptive coding order (ACO) technique, the histogram differences of luminance between frames are measured and the coding gain can be achieved by re-ordering the coding order adaptively. Furthermore, more accurate estimation of WP parameters is performed by making use of derived motion vectors (DMV). Experimental results show that our scheme can efficiently handle FL scenes, which achieves significant coding gain over the conventional WP methods in H.264/AVC.
- The region-based WP approach in Chapter 5 has been designed with the inspiration of the MRF architecture in Chapter 3. We reveal that one group of MBs may be better coded with the use of one WP parameter set while another group may be coded more

efficiently by using another WP parameter set. By partitioning the frame being encoded into regions, several representative weighting parameter sets and their corresponding regions can be obtained. Accurate WP parameter sets can be estimated with a quasi-optimal WP parameter estimator. The multiple WP parameter sets of different regions are encoded using the framework of multiple reference frames in the H.264/AVC standard such that the proposed scheme is compliant with the H.264/AVC standard. Experimental results show that the region-based scheme can efficiently encode scenes with GBVs and LBVs.

In conclusion, we expect that there are varieties of brightness variations in video content with the widespread adoption of digital movie editing and continuous developments of UHD TV, multiview video coding (MVC) and multiview video plus depth (MVD) coding. Also, great impact on the computation complexity would be brought due to extra operations for estimating sum of absolute differences using WP. Parallel processing for WP in hardware should be researched and multi-threading for WP should be considered and optimized for software and mobile applications. A further need may arise for compressing video contents with various brightness variations. In our present work, a number of WP schemes have been investigated that can cope with complex brightness variation scenes. We believe that the results obtained in this work contribute significantly to modern video coding standards for solving the problems of different kinds of brightness variations.

6.2 Future Directions

Based on the successful techniques mentioned in this thesis and proven by a wide range of experimental work, we propose here some directions for future research.

6.2.1 Detection of brightness variations in H.264/AVC

It is noted that our proposed techniques in Chapters 3 and 5 are restricted to the scenes with GBVs and LBVs, but not the scenes without brightness variations. They cannot adaptively change the MRF structure for scenes with GBVs, LBVs, or without brightness variations. For instance, if there is a movie that is composed of scenes without brightness variations, with GBVs as well as LBVs, the encoding system should be switched to an appropriate MRF structure for the best coding performance. For scenes without brightness variations, the conventional MRF structure with different temporal frames should be used. On the other hand, the MRF structures proposed in Chapter 3 and 5, depicted in Figure 3.2 and Figure 5.6 respectively, should be used for scenes with GBVs and LBVs respectively. One way for selecting the MRF structures is to use multi-pass encoding strategy [18]. However, this will cause extremely high encoding complexity because each frame would be encoded for 3 times to test all MRF structures and then the encoding system chooses the best one. Therefore it is expected to contrive a brightness variation detector to select the most appropriate MRF structure, as illustrated in Figure 6.1, based on the information and correlations of frames prior to encoding. One idea is to have a theoretical study or statistical modeling should be applied for brightness variation detection. For example, the spectrum of weight histogram can be studied in order to classify into three different kinds of scenes. That is the scene without brightness variations, scene with global brightness variation, and scene with local brightness variation.

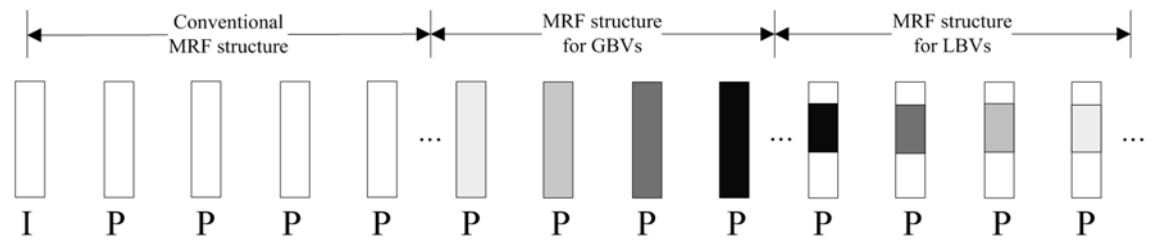


Figure 6.1. Illustration of switching MRF structure for video coding.

6.2.2 Weighted prediction for inter-view inconsistency in depth coding

In multiview video plus depth (MVD) coding [79,80], the depth map is a grayscale image that represents the information related to the relative distance from a camera to an object in the 3D space. A small value represents the far distance while a large value represents the close distance. This depth map enables to generate intermediate or virtual views such that observers can watch the same scene but from another view angle. Figure 6.2 shows an example of texture images for left and right views and their corresponding depth images. Though there is high correlation between two texture images in Figure 6.2(a) and (b), the correlation between two depth images, which depict in Figure 6.2(c) and (d), is low. Inaccurate depth estimation might be one of the reasons causing this inconsistency. Inter-view prediction cannot work well with large differences of depth values. The problem of inter-view inconsistency can be classified as a problem of local brightness variations (LBVs). Hence, with the help of WP, the coding efficiency for MVD coding could probably be improved.

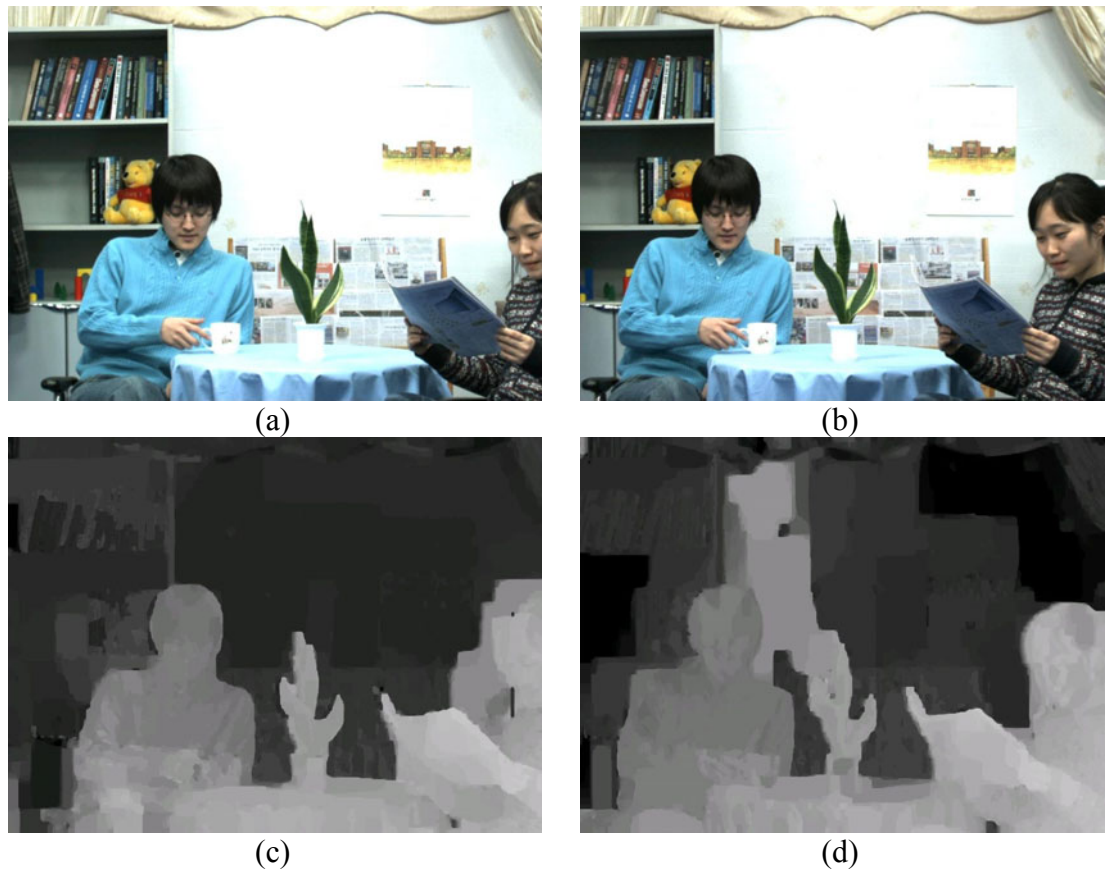


Figure 6.2. Illustration of (a) texture image of left view, (b) texture image of right view, (c) depth image of left view, and (d) depth image of right view.

6.2.3 Weighted prediction for zooming effect in depth maps

In depth map coding [79,80], object movement towards (away from) the camera would increase (decrease) the pixel values which belong to the particular moving object. In contrast, pixel values in the object contour still keep the same. Figure 6.3 shows an example of depth images with zooming effect. It can be seen that the pixel values of the frame along the scenes with zooming effect increase while the contours of the doorframe and fence still keep the same despite the enlargement. The region-based weighted prediction approach could be applied for solving this problem such that different WP parameter sets could be used to different objects locally in order to increase the coding efficiency.



Figure 6.3. Illustration of depth images of (a) preceding frame, and (b) following frame, with zooming effect.

References

- [1] ISO/IEC 10918-1, "Information technology -- Digital Compression and Coding of Continuous-tone Still Images: Requirements and Guidelines", 1994.
- [2] Y. M. Lei and M. Ouhyoung, "Software-based Motion JPEG with Progressive Refinement for Computer Animation," IEEE Transactions on Consumer Electronics, vol. 40, pp. 557-562, Aug. 1994.
- [3] ISO/IEC 11172-2, "Information Technology -- Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1,5 Mbit/s -- Part 2: Video," 1993
- [4] ISO/IEC 13818-2, "Information Technology -- Generic Coding of Moving Pictures and Associated Audio Information: Video," 1996.
- [5] Video Codec for Audiovisual Services at p×64 kbit/s, ITU-T Recommendation H.261., 1993
- [6] Video Coding for Low Bitrate Communication, ITU-T Recommendation H.263, May 1997.
- [7] ISO/IEC 14496-2, "Information Technology – Coding of audio-visual Objects – Part 2: Video," 2001.
- [8] ISO/IEC 14496-10, "Information Technology – Coding of audio-visual Objects – Part 10: Advanced Video Coding," 2003.
- [9] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.7, pp.560-576, Jul. 2003.

- [10] Iain E. G. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia," Wiley, 2003.
- [11] M. Tanimoto., M. P. Tehrani, T. Fujii, and T. Yendo, "Free-Viewpoint TV," IEEE Signal Processing Magazine, vol.28, no.1, pp.67-76, Jan. 2011.
- [12] J. M. Boyce, "Weighted Prediction in the H.264/MPEG AVC Video Coding Standard," Proceedings of International Symposium on Circuits and Systems, vol.3, pp. III- 789-92, May 2004.
- [13] H. Kato, and Y. Nakajima, "Weighting Factor Determination Algorithm for H.264/MPEG-4 AVC Weighted Prediction," IEEE Workshop on Multimedia Signal Processing, pp. 27- 30, Sep.-Oct. 2004.
- [14] R. Zhang, and G. Cote, "Accurate Parameter Estimation and Efficient Fade Detection for Weighted Prediction in H.264 Video Compression," IEEE International Conference on Image Processing, pp.2836-2839, Oct. 2008.
- [15] H. Aoki, and Y. Miyamoto, "An H.264 Weighted Prediction Parameter Estimation Method for Fade Effects in Video Scenes," IEEE International Conference on Image Processing, pp.2112-2115, Oct. 2008.
- [16] JVT H.264/AVC Joint Model (JM) Reference Software Available: <http://iphome.hhi.de/suehring/tml/download/>
- [17] A. M. Tourapis and A. Leontaris, "H.264/14496-10 AVC Reference Software Manual," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-AE010, Jul. 2009.
- [18] A. M. Tourapis, K. Sühring, and G. Sullivan, "H.264/MPEG-4 AVC Reference Software Enhancements," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Docs. JVT-N014, Jan. 2005.

- [19] M. H. Chan, Y. B. Yu, and A. G. Constantinides, "Variable Size Block Matching Motion Compensation with Applications to Video Coding," IEE Proceedings I of Communications, Speech and Vision, vol.137, no.4, pp.205-212, Aug. 1990.
- [20] G. J. Sullivan, and R. L. Baker, "Rate-distortion Optimized Motion Compensation for Video Compression using Fixed or Variable Size Blocks," Global Telecommunications Conference, vol.1, pp.85-90, Dec. 1991.
- [21] M. C. Chen, and A. N. Jr. Willson, "Rate-distortion Optimal Motion Estimation Algorithms for Motion-compensated Transform Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol.8, no.2, pp.147-158, Apr. 1998.
- [22] G. J. Sullivan, and T. Wiegand, "Rate-distortion Optimization for Video Compression," IEEE Signal Processing Magazine, vol.15, no.6, pp.74-90, Nov. 1998.
- [23] A. Ortega, and K. Ramchandran, "Rate-distortion Methods for Image and Video Compression," IEEE Signal Processing Magazine, vol.15, no.6, pp.23-50, Nov. 1998.
- [24] Y. H. Kam, and W. C. Siu, "A Fast Full Search Scheme for Rate-Distortion Optimization of Variable Block Size and Multi-frame Motion Estimation," 49th IEEE International Midwest Symposium on Circuits and Systems, vol.1, pp.183-186, Aug. 2006.
- [25] T. Wiegand, X. Zhang, and B. Girod, "Long-term Memory Motion-compensated Prediction," IEEE Transactions on Circuits and Systems for Video Technology, vol.9, no.1, pp.70-84, Feb. 1999.
- [26] A. Chung, O. C. An, and Y. M. Yeung, "A novel approach to fast multi-frame selection for H.264 video coding," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. III-413-III-416, Apr. 2003.

- [27] Y. Su and M. T. Sun, "Fast multiple reference frame motion estimation for H.264/MPEG-4 AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 447-452, Mar. 2006.
- [28] B. Girod, "Motion-compensating Prediction with Fractional-pel Accuracy," *IEEE Transactions on Communications*, vol.41, no.4, pp.604-612, Apr. 1993.
- [29] T. Wedi, and H. G. Musmann, "Motion- and Aliasing-compensated Prediction for Hybrid Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.13, no.7, pp. 577- 586, Jul. 2003.
- [30] F. J. Hampson, R. E. H. Franich, J.-C. Pesquet, and J. Biemond, "Pel-recursive Motion Estimation in the Presence of Illumination Variations," *Proceedings of International Conference on Image Processing*, vol.1, pp.101-104 vol.1, Sep. 1996.
- [31] J. Wei, and Z.-N. Li, "Motion Compensation in Color Video with Illumination Variations," *Proceedings of International Conference on Image Processing*, vol.3, pp.614-617, Oct. 1997.
- [32] K. Kamikura, H. Watanabe, H. Jozawa, H. Kotera, and S. Ichinose, "Global Brightness-variation Compensation for Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.8, no.8, pp.988-1000, Dec. 1998.
- [33] N. M. M. Rodrigues, V. M. M. da Silva, and S. M. M. de Faria, "Hierarchical motion compensation with spatial and luminance transformations," *Proceeding of International Conference on Image Processing*," vol.3, pp.518-521, 2001.
- [34] D. Tian, M. M. Hannuksela, Y. K. Wang, and M. Gabbouj, "Coding of Faded Scene Transitions," *International Conference on Image Processing*, vol.2, pp. II-505- II-508, 2002.

- [35] S. H. Kim, and R. H. Park, "Fast Local Motion-compensation Algorithm for Video Sequences with Brightness Variations," IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.4, pp. 289- 299, Apr. 2003.
- [36] S. Koto, T. Chujoh, and Y. Kikuchi, "Adaptive Bi-predictive Video Coding using Temporal Extrapolation," Proceedings of International Conference on Image Processing, vol.3, pp. III- 829-32, Sep. 2003.
- [37] H. K. Cheung, W. C. Siu, D. Feng, and Z. Wang, "Retinex Based Motion Estimation for Sequences with Brightness Variations and Its Application to H.264," IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1161-1164, Mar.-Apr. 2008.
- [38] H. K. Cheung, W. C. Siu, D. Feng, and T. Cai, "New Block-Based Motion Estimation for Sequences with Brightness Variation and Its Application to Static Sprite Generation for Video Compression," IEEE Transactions on Circuits and Systems for Video Technology, vol.18, no.4, pp.522-527, Apr. 2008.
- [39] H. K. Cheung, W. C. Siu, D. Feng, and Z. Wang, "Windowing Technique for the DCT Based Retinex Algorithm to Handle Videos with Brightness Variations Coded using the H.264," IEEE International Conference on Image Processing, pp.2860-2863, Oct. 2008.
- [40] L. Song, H. Xiong, J. Xu, F. Wu, and H. Su, "Adaptive Predict Based on Fading Compensation for Lifting-based Motion Compensated Temporal Filtering," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.2, no., pp. ii/909- ii/912, Mar. 2005.
- [41] Q. Chen, Z. Nie, Z. Chen, X. Gu, G. Qiu, and C. Wang, "A Human Vision System based Flash Picture Coding Method for Video Coding," IEEE International Symposium on Circuits and Systems, pp.989-992, May 2007.

- [42] S. Yu, Y. Gao, J. Chen, and J. Zhou, "Distance-based Weighted Prediction for H.264 Intra Coding," International Conference on Audio, Language and Image Processing, pp.1477-1480, Jul. 2008.
- [43] Y. Zheng, P. Yin, O. D. Escoda, X. Li, and C. Gomila, "Intra Prediction using Template Matching with Adaptive Illumination Compensation," IEEE International Conference on Image Processing, pp.125-128, Oct. 2008.
- [44] Y. H. Kim, B. Choi, and J. Paik, "High-Fidelity RGB Video Coding Using Adaptive Inter-Plane Weighted Prediction," IEEE Transactions on Circuits and Systems for Video Technology, vol.19, no.7, pp.1051-1056, Jul. 2009.
- [45] L. M. Po, L. Wang, K. W. Cheung, K. M. Wong, K. H. Ng, S. Li, and C. W. Ting, "Distance-based Weighted Prediction for Adaptive Intra Mode Bit Skip in H.264/AVC," IEEE International Conference on Image Processing, pp.2869-2872, Sep. 2010.
- [46] Y. Zhou, X. Sun, H. Bao, and S. Li, "Weighted Motion Estimation for Efficiently Coding Scene Transition Video," IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing, vol.3, pp. iii- 361-4, May 2004.
- [47] K. Panusopone, X. Fang, and L. Wang, "An Efficient Implementation of Motion Estimation with Weight Prediction for ITU-T H.264 | MPEG-4 AVC," IEEE International Conference on Consumer Electronics, Digest of Technical Papers, pp.1-2, Jan. 2007.
- [48] K. Panusopone, X. Fang, and L. Wang, "An Efficient Implementation of Motion Estimation with Weight Prediction for ITU-T H.264 MPEG-4 AVC," IEEE Transactions on Consumer Electronics, vol.53, no.3, pp.974-978, Aug. 2007.
- [49] F. Kamisli, and D. M. Baylon, "Estimation of Fade and Dissolve Parameters for Weighted Prediction in H.264/AVC," IEEE International Conference on Image Processing, vol.5, no., pp.V-285-V-288, Sept. 2007.

- [50] A. Leontaris, and A. M. Tourapis, "Weighted Prediction Methods for Improved Motion Compensation," IEEE International Conference on Image Processing (ICIP), pp.1029-1032, Nov. 2009.
- [51] S. H. Tsang, Y. L. Chan, and W. C. Siu, "New Weighted Prediction Architecture for Coding Scenes with Various Fading Effects – Image and Video Processing," Proceedings of International Conference on Signal Processing and Multimedia Applications (SIGMAP 2010), pp.118-123, Jul. 2010.
- [52] S. H. Tsang, and Y. L. Chan, "H.264 video coding with multiple weighted prediction models," IEEE International Conference on Image Processing, pp.2069-2072, Sep. 2010.
- [53] D. K. Kwon, and H. J. Kim, "Region-based weighted prediction for real-time H.264 encoder," IEEE International Conference on Consumer Electronics, pp.47-48, Jan. 2011.
- [54] S. H. Tsang, Y. L. Chan, and W. C. Siu, "Multiple Weighted Prediction Models for Video Coding with Brightness Variations," IET Image Processing, vol. 6, issue 4, pp. 434-443, Jun. 2012.
- [55] A. Tanizawa, T. Chujoh, and T. Yamakage, "Multi-directional Implicit Weighted Prediction based on Image Characteristics of Reference Pictures for Inter Coding," IEEE International Conference on Image Processing, vol. 6, issue 4, pp. 1545-1548, Sep. 2012.
- [56] S. H. Tsang, Y. L. Chan, and W. C. Siu, "Region-based Weighted Prediction for Coding Video with Local Brightness Variations", IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 3, pp. 549-561, March 2013.
- [57] Y. Shen, D. Zhang, C. Huang, and J. Li, "Adaptive Weighted Prediction in Video Coding," IEEE International Conference on Multimedia and Expo, vol.1, pp.427-430, Jun. 2004.

- [58] AVS Reference Software (RM) Available:
<http://www.avs.org.cn/fruits/softList.asp>
- [59] D. Liu, Y. He, S. Li, Q. Huang, and W. Gao, "Linear Transform Based Motion Compensated Prediction for Luminance Intensity Changes," IEEE International Symposium on Circuits and Systems, Vol. 1, pp. 304- 307, May 2005.
- [60] P. Yin, A. M. Tourapis, and J. Boyce, "Localized Weighted Prediction for Video Coding," IEEE International Symposium on Circuits and Systems, Vol. 5, pp. 4365- 4368, May 2005.
- [61] K. Hayase, Y. Bandoh, S. Takamura, K. Kamikura, and Y. Yashima, "A Weighted Prediction of Spatial Scalable Video Coding with Inter-layer Information," EURASIP Proceedings of International Picture Coding Symposium (PCS), Nov. 2007.
- [62] J. H. Hur, S. Cho, and Y. L. Lee, "Adaptive Local Illumination Change Compensation Method for H.264/AVC-Based Multiview Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17 no.11, pp. 1496-1501, Nov. 2007.
- [63] J. H. Kim, PoLin Lai, J. Lopez, A. Ortega, Y. Su, P. Yin, and C. Gomila, "New Coding Tools for Illumination and Focus Mismatch Compensation in Multiview Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17 no.11, pp. 1519-1535, Nov. 2007.
- [64] Nikola Sprljan, and L. Cieplinski, "Macroblock-level Weighted Prediction," ISO/IEC JTC1/SC29/WG11, Moving Picture Experts Group (MPEG), Docs. M16969, Oct. 2009.
- [65] D. T. Võ, C. W. Seo, D. Jin, J. K. Han, and T. Q. Nguyen, "Optimal Spatial-temporal Weight Prediction for Inter-frame Coding of H.264/AVC Video Sequences,"

International Conference on Advanced Technologies for Communications, pp.266-269, Oct. 2010.

[66] H. I. Bang, J. H. Choi, and M. H. Sunwoo, "An Efficient Skipping Method of H.264/AVC Weighted Prediction for Various Illuminating Effects," Proceedings of IEEE International Symposium on Circuits and Systems, pp.1177-1180, May-Jun. 2010.

[67] S. H. Tsang, Y. L. Chan, and W. C. Siu, "Flashlight Scene Video Coding using Weighted Prediction," Journal of Visual Communication and Image Representation, vol. 23, Issue 2, pp. 264-270, Feb. 2012.

[68] A. M. Alattar, "Detecting Fade Regions in Uncompressed Video Sequences," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.4, pp.3025-3028, Apr. 1997.

[69] J. Wang, Y. Xu, S. Yu, and Y. Zhou, "Flashlight Scene Detection for MPEG Videos," IEEE 6th Workshop on Multimedia Signal Processing, pp. 1-4, Oct. 2005.

[70] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," IEEE Transactions on Circuits and Systems for Video Technology, vol.16, no.1, pp. 82- 91, Jan. 2006.

[71] X. Qian, G. Liu, and R. Su, "Effective Fades and Flashlight Detection Based on Accumulating Histogram Difference," IEEE Transactions on Circuits and Systems for Video Technology, vol.16, no.10, pp.1245-1258, Oct. 2006.

[72] Tan T.K., Sullivan G., and Wedi T., "Recommended Simulation Conditions for Coding Efficiency Experiments Revision 4," ITU-T SC16/Q6, 36th VCEG Meeting, Document VCEG-AJ10, Jan. 2008.

[73] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-curves". ITU-T Q6/SG16, Video Coding Experts Group (VCEG), Docs. VCEG-M33, Mar. 2001.

- [74] Y. L. Chan, and W. C. Siu, "Search Strategy for Partial Distortion Elimination in Motion Estimation," *Electronics Letters*, vol.38, no.23, pp. 1427- 1428, 7 Nov. 2002.
- [75] Y. L. Chan, K. C. Hui, and W. C. Siu, "Adaptive Partial Distortion Search Algorithm for Block Motion Estimation," *ScienceDirect Journal of Visual Communication and Image Representation*, vol. 15, issue 4, pp.489-506, Dec. 2004.
- [76] K. C. Hui, W. C. Siu, and Y. L. Chan, "New Adaptive Partial Distortion Search using Clustered Pixel Matching Error Characteristic," *IEEE Transactions on Image Processing*, , vol.14, no.5, pp.597-607, May 2005.
- [77] Trailer sequences downloaded from <http://trailers.apple.com/>.
- [78] S. Kamp, and M. Wien, "High Definition Test Sequences for High-performance Video Coding (HVC)," *ISO/IEC JTC1/SC29/WG11, Moving Picture Experts Group (MPEG)*, M16462, Apr. 2009.
- [79] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol.99, no.4, pp.626-642, Apr. 2011.
- [80] K. Müller, P. Merkle, and T. Wiegand, "3-D Video Representation Using Depth Maps," *Proceedings of the IEEE*, vol.99, no.4, pp.643-656, Apr. 2011.